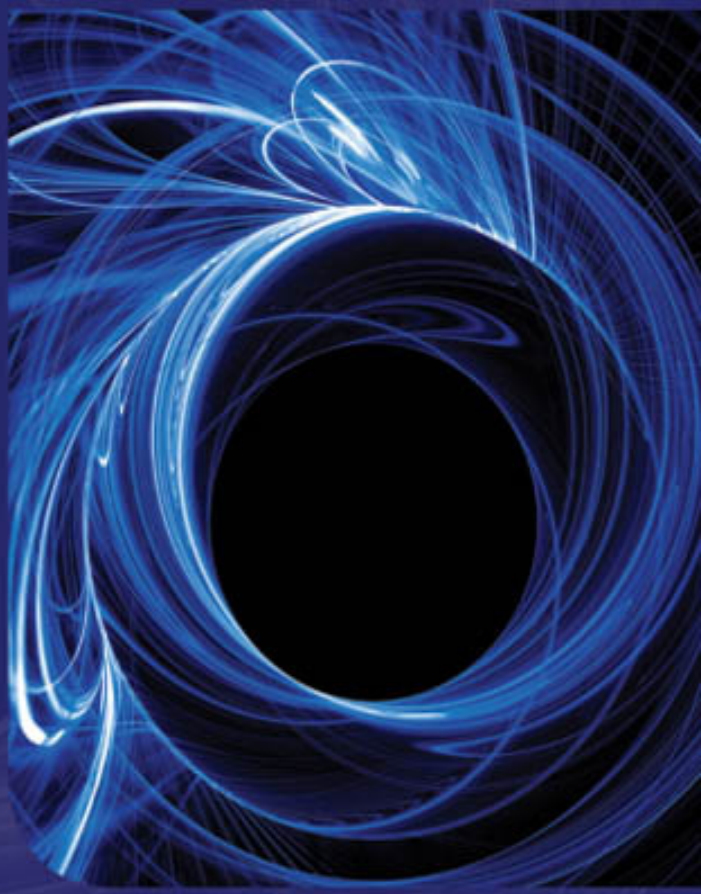
New Oxford Textbook of Psychiatry

SECOND EDITION

Edited by
Michael G Gelder
Nancy C Andreasen
Juan J López-Ibor Jr
John R Geddes



VOLUME 1 & 2

OXFORD

New Oxford Textbook of

Psychiatry

This page intentionally left blank

VOLUME 1

New Oxford Textbook of Psychiatry

SECOND EDITION

Edited by

Michael G. Gelder
Emeritus Professor of Psychiatry,
Warneford Hospital, University of Oxford,
Oxford, UK

Nancy C. Andreasen
Director, Mental Health Clinical Research Centre,
University of Iowa Hospital and Clinic,
Iowa City, USA

Juan J. López-Ibor Jr.
Professor of Psychiatry,
Complutense University, Madrid, Spain

and

John R. Geddes
Professor of Epidemiological Psychiatry
University of Oxford, Warneford Hospital,
Oxford, UK

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford ox2 6dp

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First edition published 2000
Reprinted 2003

This edition published 2009

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloguing in Publication Data

Data available

Typeset in Cepha Imaging Pvt. Ltd., Bangalore, India
Printed in Italy on acid-free paper by Rotolito Lombarda SpA

ISBN 978-0-19-920669-8
ISBN 978-0-19-920669-8 (set)
ISBN 978-0-19-955992-3 (Volume 1)
ISBN 978-0-19-955993-0 (Volume 2)

10 9 8 7 6 5 4 3 2 1

Preface to the second edition

This new edition, like the first, aims to present a comprehensive account of clinical psychiatry with reference to its scientific basis and to the ill person's perspective. As in the first edition, the authors are drawn from many countries, including the UK, the USA, 12 countries in continental Europe, and Australasia. The favourable reception of the first edition has led us to invite many of the original authors to revise their chapters for this second edition but 50 chapters are the work of new authors, many concerned with subjects that appeared in the first edition, while others are completely new. The forensic psychiatry section has the most new chapters, followed by the section on psychology as a scientific basis of psychiatry.

The overall plan of the book resembles that of the first edition (see preface to the 1st edition, reprinted on pages vii and viii). One important feature is that information about treatment appears in more than one place. The commonly used physical and psychological treatments are described in Section 6. Their use in the treatment of any particular disorder is considered in the chapter concerned with that disorder and the account is in two parts. The first part is a review of evidence about the effects of each of the treatments when used for that disorder. The second part, called Management, combines evidence from clinical trials with accumulated clinical experience to produce practical advice about the day to day care of people with the disorder.

Although much information can now be obtained from internet searches, textbooks are still needed to provide the comprehensive account of established knowledge into which new information can be fitted and against which recent findings can be evaluated. As well as seeking to provide an authoritative account of essential knowledge, each chapter in the new edition includes a brief list of sources of further information, including where appropriate, regularly updated web sites.

An essential component of good practice is the need to be aware of patients' perspectives, to respect their wishes, and to work with them, and often their families, as partners. The book opens with an important chapter on the experience of being a patient, and there are chapters on stigma, ethics, and the developing topic of values-based practice.

We are grateful to the following who advised us about parts of the book; Professor John Bancroft (Psychosexual Disorders), Professor Tom Burns (Social and Community Psychiatry), Professor William Fraser (Intellectual Disability), Professor Keith Hawton (Suicide and Deliberate Self Harm), Professor Susan Iversen (Psychology), Professor Robin Jacoby (Old Age Psychiatry), Professor Paul Mullen (Forensic Psychiatry), Sir Michael Rutter (Child and Adolescent Psychiatry), and Professor Gregory Stores (Sleep Disorders).

The editors

This page intentionally left blank

Preface to the first edition

Three themes can be discerned in contemporary psychiatry: the growing unity of the subject, the pace of scientific advance, and the growth of practice in the community. We have sought to reflect these themes in the *New Oxford Textbook of Psychiatry* and to present the state of psychiatry at the start of the new millennium. The book is written for psychiatrists engaged in continuous education and recertification; the previous, shorter, *Oxford Textbook of Psychiatry* remains available for psychiatrists in training. The book is intended to be suitable also as a work of reference for psychiatrists of all levels of experience, and for other professionals whose work involves them in the problems of psychiatry.

The growing unity of psychiatry

The growing unity in psychiatry is evident in several ways. Biological and psychosocial approaches have been largely reconciled with a general recognition that genetic and environmental factors interact, and that psychological processes are based in and can influence neurobiological mechanisms. At the same time, the common ground between the different psychodynamic theories has been recognized, and is widely accepted as more valuable than the differences between them.

The practice of psychiatry is increasingly similar in different countries, with the remaining variations related more to differences between national systems of health care and the resources available to clinicians, than to differences in the aims of the psychiatrists working in these countries. This unity of approach is reflected in this book whose authors practise in many different countries and yet present a common approach. In this respect this textbook differs importantly from others which present the views of authors drawn predominantly from a single country or region.

Greater agreement about diagnosis and nosology has led to a better understanding of how different treatment approaches are effective in different disorders. The relative specificity of psychopharmacological treatments is being matched increasingly by the specificity of some of the recently developed psychological treatments, so that psychological treatment should no longer be applied without reference to diagnosis, as was sometimes done in the past.

The pace of scientific advance

Advances in genetics and in the neurosciences have already increased knowledge of the basic mechanisms of the brain and are beginning to uncover the neurobiological mechanisms involved in psychiatric disorder. Striking progress has been achieved in the understanding of Alzheimer's disease, for example, and there are indications that similar progress will follow in uncovering the causes of mood disorder, schizophrenia, and autism. Knowledge of genetics and the neurosciences is so extensive and the pace of change is so rapid that it is difficult to present a complete account within the limited space available in a textbook of clinical psychiatry. We have selected aspects of these sciences that seem, to us and the authors, to have contributed significantly to psychiatry or to be likely to do so before long.

Psychological and social sciences and epidemiology are essential methods of investigation in psychiatry. Although the pace of advance in these sciences may not be as great as in the neurosciences, the findings generally have a more direct relation to clinical phenomena. Moreover, the mechanisms by which psychological and social factors interact with genetic, biochemical, and structural ones will continue to be important however great the progress in these other sciences. Among the advances in the psychological and social sciences that are relevant to clinical phenomena, we have included accounts of memory, psychological development, research on life events, and the effects of culture. Epidemiological studies continue to be crucial for defining psychiatric disorders, following their course, and identifying their causes.

Psychiatry in the community

In most countries, psychiatry is now practised in the community rather than in institutions, and where this change has yet be completed, it is generally recognized that it should take place. The change has done much more than transfer the locus of care; it has converted patients from passive recipients of care to active participants with individual needs and preferences. Psychiatrists are now involved in the planning, provision, and evaluation of services for whole communities, which may include members of ethnic minorities, homeless people, and refugees. Responsibility for a community has underlined the importance of the prevention as well as the treatment of mental disorder and of the role of agencies other than health services in both. Care in the community has also drawn attention to the many people with psychiatric disorder who are treated in primary care, and has led to new ways of working between psychiatrists and physicians. At the same time, psychiatrists have

worked more in general hospitals, helping patients with both medical and psychiatric problems. Others have provided care for offenders.

The organization of the book

In most ways, the organization of this book is along conventional lines. However, some matters require explanation.

Part 1 contains a variety of diverse topics brought together under the general heading of the subject matter and approach to psychiatry. Phenomenology, assessment, classification, and ethical problems are included, together with the role of the psychiatrist as educator and as manager. Public health aspects of psychiatry are considered together with public attitudes to psychiatry and to psychiatric patients. Part 1 ends with a chapter on the links between science and practice. It begins with a topic that is central to good practice—the understanding of the experience of becoming a psychiatric patient.

Part 2 is concerned with the scientific foundations of psychiatry grouped under the headings neurosciences, genetics, psychological sciences, social sciences, and epidemiology. The chapters contain general information about these sciences; findings specific to a particular disorder are described in the chapter on that disorder. Brain imaging techniques are discussed here because they link basic sciences with clinical research. As explained above, the chapters are selective and, in some, readers who wish to study the subjects in greater detail will find suggestions for further reading.

Part 3 is concerned with dynamic approaches to psychiatry. The principal schools of thought are presented as alternative ways of understanding the influence of life experience on personality and on responses to stressful events and to illness. Some reference is made to dynamic psychotherapy in these accounts, but the main account of these treatments is in Part 6. This arrangement separates the chapters on the practice of dynamic psychotherapy from those on psychodynamic theory, but we consider that this disadvantage is outweighed by the benefit of considering together the commonly used forms of psychotherapy.

Part 4 is long, with chapters on the clinical syndromes of adult psychiatry, with the exception of somatoform disorders which appear in Part 5, Psychiatry and Medicine. This latter contains more than a traditional account of psychosomatic medicine. It also includes a review of psychiatric disorders that may cause medical symptoms unexplained by physical pathology, the medical, surgical, gynaecological, and obstetric conditions most often associated with psychiatric disorder, health psychology, and the treatment of psychiatric disorder in medically ill patients.

Information about treatment appears in more than one part of the book. Part 6 contains descriptions of the physical and psychological treatments in common use in psychiatry. Dynamic psychotherapy and psychoanalysis are described alongside counselling and cognitive behavioural techniques. This part of the book contains general descriptions of the treatments; their use for a particular disorder is considered in the chapter on that disorder. In the latter, the account is generally in two parts: a review of evidence about the efficacy of the treatment, followed by advice on management in which available evidence is supplemented, where necessary, with clinical experience. Treatment methods designed specially for children and adolescents, for people with mental retardation (learning disability), and for patients within the forensic services are considered in Parts 9, 10, and 11 respectively.

Social psychiatry and service provision are described in Part 7. Public policy issues, as well as the planning, delivery, and evaluation of services, are discussed here. Psychiatry in primary care is an important topic in this part of the book. There are chapters on the special problems of members of ethnic minorities, homeless people, and refugees, and the effects of culture on the provision and uptake of services.

Child and adolescent psychiatry, old age psychiatry, and mental retardation are described in Parts 8, 9, and 10. These accounts are less detailed than might be found in textbooks intended for specialists working exclusively in the relevant subspecialty. Rather, they are written for readers experienced in another branch of psychiatry who wish to improve their knowledge of the special subject. We are aware of the controversy surrounding our choice of the title of Part 10. We have selected the term 'mental retardation' because it is used in both ICD-10 and DSM-IV. In some countries this term has been replaced by another that is thought to be less stigmatizing and more acceptable to patients and families. For example, in the United Kingdom the preferred term is 'learning disability'. While we sympathize with the aims of those who adopt this and other alternative terms, the book is intended for an international readership and it seems best to use the term chosen by the World Health Organization as most generally understood. Thus the term mental retardation is used unless there is a special reason to use another.

In Part 11, Forensic Psychiatry, it has been especially difficult to present a general account of the subject that is not tied to practice in a single country. This is because systems of law differ between countries and the practice of forensic psychiatry has to conform with the local legal system. Although many of the examples in this part of the book may at first seem restricted in their relevance because they are described in the context of English law, we hope that readers will be able to transfer the principles described in these chapters to the legal tradition in which they work.

Finally, readers should note that the history of psychiatry is presented in more than one part of the book. The history of psychiatry as a medical specialty is described in Part 1. The history of ideas about the various psychiatric disorders appears, where relevant, in the chapters on these disorders, where they can be considered in relation to present-day concepts. The history of ideas about aetiology is considered in Part 2, which covers the scientific basis of psychiatric aetiology, while the historical development of dynamic psychiatry is described in Part 3.

Michael Gelder
Juan López-Ibor
Nancy Andreasen

Acknowledgements from the first edition

We are grateful to the many colleagues who have advised us about certain parts of the book.

The following helped us to plan specialized parts of the book: Dr Jeremy Holmes (Section 3, Psychodynamic Contributions to Psychiatry); Professor Richard Mayou (Section 5, Psychiatry and Medicine); Professor Robin Jacoby (Section 8, Psychiatry of Old Age); Sir Michael Rutter (Section 9, Child and Adolescent Psychiatry); Professor William Fraser (Section 10, Intellectual Disablity); Professor Robert Bluglass (Section 11, Forensic Psychiatry).

The following helped us to plan certain sections within Section 4, General Psychiatry: Professor Alwyn Lishman (delirium, dementia, amnestic syndrome, and other cognitive disorders); Professor Griffith Edwards (alcohol use disorders); Dr Philip Robson (other substance use disorders); Professor Guy Goodwin (mood disorders); Professor John Bancroft (sexuality, gender identity, and their disorders); Professor Gregory Stores (sleep–wake disorders); Professor Keith Hawton (suicide and attempted suicide). In Section 6, Professor Philip Cowen advised about somatic treatments, Dr Jeremy Holmes about psychodynamic treatments, and Professor David Clark about cognitive behavioural therapy. Dr Max Marshall provided helpful advice about forensic issues for Section 7. We also thank the many other colleagues whose helpful suggestions about specific problems aided the planning of the book.

Finally, we record our special gratitude to the authors and to the staff of Oxford University Press.

This page intentionally left blank

Contents
Volume 1

Contents
Volume 2

This page intentionally left blank

Contributors list

S.E. Abbey Associate Professor of Psychiatry, University of Toronto, Toronto, Canada
Chapter 5.3.6

Henry David Abraham Distinguished Life Fellow, American Psychiatric Association
Chapter 4.2.3.3

Clive E. Adams Cochrane Schizophrenia Group, University of Oxford Department of Psychiatry, Warneford Hospital, Oxford, UK
Chapter 6.1.1

Hagop S. Akiskal Professor of Psychiatry and Director of the International Mood Center, University of California at San Diego, California, USA
Chapter 4.5.9

Nancy C. Andreasen Dept of Psychiatry, University of Iowa Hospitals & Clinics, Iowa City, USA
Chapter 4.3.1

Adrian Angold Associate Professor of Child and Adolescent Psychiatry, Duke University Medical Center, Durham, North Carolina, USA
Chapter 9.1.2

Jules Angst Emeritus Professor of Psychiatry, Zurich University
Chapter 4.5.7

Ella Arensman Director of Research, National Suicide Research Foundation, Ireland
Chapter 4.15.2

J.K. Aronson Reader in Clinical Pharmacology, University Department of Primary Health Care, Headington, Oxford
Chapter 6.2.1

José-Luis Ayuso-Mateos Chairman, Department of Psychiatry, Universidad Autónoma de Madrid, Hospital Universitario de la Princesa, Spain
Chapter 5.3.5

Susan Bailey Consultant Child and Adolescent Forensic Psychiatrist, Salford NHS Trust and Maudsley NHS Trust; Senior Research Fellow, University of Manchester, UK
Chapter 11.7

Robert Baldwin Consultant, Old Age Psychiatrist, and Honorary Senior Lecturer, Manchester Royal Infirmary, UK
Chapter 8.5.4

James C. Ballenger Retired Professor and Chairman, Department of Psychiatry and Behavioral Sciences and Director, Institute of Psychiatry, Medical University of South Carolina
Chapter 4.7.3

John Bancroft, The Kinsey Institute for Research in Sex, Gender, & Reproduction and Department of Psychiatry, University of Oxford
Chapter 4.11.2

David H. Barlow Center for Anxiety and Related Disorders at Boston University, Massachusetts, USA
Chapter 4.7.1

Delwyn J. Bartlett Woolcock Institute of Medical Research, Sydney, Australia
Chapter 4.14.2

Christopher Bass Consultant in Liaison Psychiatry, John Radcliffe Hospital, Oxford, UK
Chapters 5.2.4 and 5.2.9

Antony W. Bateman Halliwick Psychotherapy Dept, St Ann's Hospital, London, UK
Chapter 4.12.6

Per Bech Professor of Psychiatry and Head of Psychiatric Research Unit, WHO Collaborating Centre, Frederiksborg General Hospital, Hillerød, Denmark
Chapter 4.5.2

Sidney Benjamin Senior Lecturer, University of Manchester, UK
Chapter 5.2.6

Thomas P. Berney Consultant Developmental Psychiatrist Honorary Research Associate, University of Newcastle upon Tyne
Chapter 10.6

Jose M. Bertolote Chief, Mental Disorders Control Unit, World Health Organization, Geneva; Associate Professor, Department of Psychogeriatrics, University of Lausanne, Switzerland
Chapter 7.4

Zubin Bhagwagar CT Mental Health Center, Yale University, New Haven CT, USA
Chapter 6.2.3

Mary Beth Connolly Gibbons Assistant Professor of Psychology in Psychiatry Department of Psychiatry, University of Pennsylvania, Pennsylvania, USA
Chapter 6.1.2

Dinesh Bhugra Professor of Mental Health and Cultural Diversity, King's College London, Institute of Psychiatry, London, UK
Chapter 7.10.3

Michel Billiard Professor of Neurology, School of Medicine, Guide Chauliac Hospital, Montpellier, France
Chapter 4.14.3

Max Birchwood Director, Early Intervention Service, Northern Birmingham Mental Health Trust, and University of Birmingham, UK
Chapter 6.3.2.4

Boris Birmaher UPMC Western Psychiatric Institute, Pittsburgh, USA
Chapter 9.2.7

Stella Bitran, Center for Anxiety and Related Disorders, Boston University, Beacon, MA
Chapter 4.7.1

Alan H. Bittles Centre for Comparative Genomics, Murdoch University, Perth, Australia
Chapter 8.1

Dora Black Honorary Consultant, Child and Adolescent Psychiatry, Traumatic Stress Clinic, London; Honorary Lecturer, University of London, UK
Chapters 9.3.2 and 9.3.7

Michelle A. Blackmore, Doctoral Student of Clinical Psychology Adult Anxiety Clinic at Temple University, Philadelphia, Pennsylvania, USA
Chapter 4.7.2

Carlos Blanco New York State Psychiatric Institute, New York,
Chapter 6.3.3

Sidney Bloch Professor of Psychiatry, University of Melbourne; Senior Psychiatrist, St Vincent's Hospital, Melbourne, Australia
Chapters 1.5 and 6.3.8

William V. Bobo Assistant Professor of Psychiatry, Vanderbilt University School of Medicine Nashville, Tennessee (USA)
Chapter 6.2.5

Bernhard Bogerts Department of Psychiatry, University of Magdeburg, Germany
Chapter 2.3.5

Jeff Bostic School of Psychiatry, Harvard Medical School, Cambridge MA, USA
Chapter 9.1.3

Nick Bouras Professor, Institute of Psychiatry - King's College London MHiLD - York Clinic, London, UK
Chapter 10.9

Christopher R. Bowie Department of Psychiatry, Mount Sinai School of Medicine, New York, USA
Chapter 4.3.3

Colin Bradbury Department of Psychological Medicine, Institute of Psychiatry, De Crespigny Park, London, UK
Chapter 4.2.4

Brian P. Brennan Instructor in Psychiatry, Harvard Medical School and Associate Director for Translational Neuroscience Research, Biological Psychiatry Laboratory, McLean Hospital, Belmont, MA
Chapter 6.2.6

David Brent Dept of Psychiatry, University of Pittsburgh Medical School, Pittsburgh PA, USA
Chapter 9.2.7

Chris R. Brewin Research Dept of Clinical, Educational & Health Psychology, University College London, UK
Chapter 4.6.3

Elaine Brohan Institute of Psychiatry, David Goldberg Centre, De Crespigny Park, London, UK
Chapter 1.2

Ian Brockington Professor of Psychiatry, University of Birmingham, UK
Chapter 5.4

George W. Brown Professor of Sociology, Academic Department of Psychiatry, St Thomas's Hospital, London, UK
Chapter 2.6.1

Arne Brun Professor of Neuropathology. Department of Pathology, Lund University Hospital, Lund, Sweden
Chapter 4.1.3

Richard A. Bryant School of Psychology, University of New South Wales, Sydney NSW, Australia
Chapter 4.6.1

E.T. Bullmore Institute of Psychiatry, King's College London, UK
Chapters 2.3.7 and 2.3.8

Tom Burns Professor of social psychiatry, Dept of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK
Chapter 7.5

José Luis Carrasco Professor of Psychiatry, Hospital Fundacion Jimenez Diaz, Universidad Autonoma, Madrid, Spain
Chapter 4.12.3

D.J. Castle University of Western Australia, Fremantle, Australia
Chapter 4.3.6.1

Jonathan Chick Consultant Psychiatrist, NHS Lothian, and Senior Lecturer, Department of Psychiatry, University of Edinburgh
Chapter 4.2.2.4

Daniel Chisholm Department of Health System Financing, Health Systems and Services, World Health Organization, Geneva, Switzerland
Chapter 7.7

David Christmas Dept of Psychiatry, University of Dundee, Dundee, UK
Chapter 6.2.10.4

David M. Clarke Consultant Psychiatrist, Lea Castle Centre, Kidderminster DY10 3PP
Chapters 6.3.3.1 and 10.4

L. Clark Dept of Experimental Psychology, University of Cambridge, Cambridge, UK
Chapter 2.5.5

Judy Clegg Lecturer, Speech and language therapist, HPC, RCSLT Department of Human Communication Sciences University of Sheffield, UK
Chapter 9.2.11

C. Robert Cloninger Dept of Psychiatry, Washington University School of Medicine, St Louis MO, USA
Chapter 1.8.2

John Collinge Head of the Department of Neurodegenerative Disease at the Institute of Neurology, University College London and the Director of the UK Medical Research Council's Prion Unit, London, UK
Chapter 4.1.4

Henry O'Connell Consultant Psychiatrist, Co. Tipperary, Ireland
Chapter 8.5.2

Melissa A. Culhane Harvard Program in Refugee Trauma, Department of Psychiatry, Massachusetts General Hospital, Cambridge, USA
Chapter 7.10.1

John E. Cooper Emeritus Professor of Psychiatry, University of Nottingham, UK
Chapter 1.8.1

Sally-Ann Cooper Professor of Learning Disabilities, Division of Community Based Sciences, Faculty of Medicine, University of Glasgow
Chapter 10.2

Zafra Cooper Principal Research Psychologist, Oxford University Department of Psychiatry, Warneford Hospital, Oxford, UK
Chapters 4.10.2 and 6.3.2.2

E. Jane Costello Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Brightleaf Square, Durham NC
Chapter 9.1.2

Elizabeth Coulthard Institute of Neurology, University College London, UK
Chapter 2.5.2

F.-R. Cousin Psychiatrist, Centre Hospitalier Saint-Anne, Paris, France
Chapter 4.3.10

Philip J. Cowen Professor of Psychopharmacology, Department of Psychiatry, University of Oxford
Chapter 6.2.10.2

Tom K.J. Craig Professor of Social Psychiatry, King's College London, Institute of Psychiatry, London UK
Chapters 7.10.2 and 7.10.3

Paul Crits-Christoph Professor of Psychology in Psychiatry Director, Center for Psychotherapy Research Department of Psychiatry, University of Pennsylvania. Pennsylvania, USA
Chapter 6.1.2

Michael Crowe Consultant Psychiatrist, South London and Maudsley NHS Trust; Honorary Senior Lecturer, Institute of Psychiatry, King's College London, UK
Chapter 6.3.7

D.G. Cunningham Owens Reader in Psychiatry, Department of Psychiatry, University of Edinburgh, UK
Chapter 4.3.8

Dianne Currier Division of Molectular Imaging & Neuropathology, Department of Psychiatry, Columbia University
Chapter 4.15.3

Robert Dantzer Integrative Immunology and Behavior Program, University of Illinois at Urbana-Champaign, Edward R. Madigan Laboratory, West Gregory Drive, Urbana, IL, USA
Chapter 2.3.10

Rajan Darjee Division of Psychiatry, University of Edinburgh, Edinburgh, UK
Chapter 11.3.1

Anthony S. David Professor of Cognitive Neuropsychiatry, Institute of Psychiatry, King's College London, UK
Chapters 4.3.4 and 4.9

Kate Davidson Senior Research Psychologist, Department of Psychological Medicine, University of Glasgow, UK
Chapter 4.12.7

Martin Davies Dept of Experimental Psychology, University of Oxford, Oxford, UK
Chapter 2.1

Giovanni de Girolamo Health Care Research Agency, Emilia-Romagna Region, Bologna, Italy
Chapters 4.12.2 and 4.12.4

Shoumitro Deb Clinical Professor of Neuropsychiatry & Intellectual Disability, Division of Neuroscience, University of Birmingham, UK
Chapter 10.4

R.J. Dolan Institute of Neurology, University College London, UK
Chapter 2.5.4

Anton Došen Emeritus Professor of Psychiatric Aspects of Intellectual Disability at the Radboud University, Nijmegen, The Netherlands
Chapter 10.5.2

D. Colin Drummond Professor of Addiction Psychiatry, Section of Alcohol Research, National Addiction Centre, Division of Psychological Medicine and Psychiatry, Institute of Psychiatry, King's College London
Chapter 4.2.2.5

Emma Dunn School of Medicine, Cardiff University, Cardiff, UK
Chapter 11.17

Graham Dunn Professor of Biomedical Statistics, Health Methodology Research Group, School of Community Based Medicine, University of Manchester
Chapter 2.2

Julie Dunsmore Honorary Clinical Associate, SciMHA Unit, University of Western Sydney, Australia
Chapter 4.6.5

Anke Ehlers Department of Psychiatry, University of Oxford, UK
Chapters 4.6.1 and 4.6.2

Timo Erkinjuntti Professor of Neurology, Head of the University Department of Neurological Sciences, University of Helsinki and Head Physician, Department of Neurology and Memory Research Unit, Helsinki University Central Hospital, Finland
Chapter 4.1.8

Brigette A. Erwin Adult Anxiety Clinic of Temple University, Philadelphia, Pennsylvania, USA
Chapter 4.7.2

Colin A. Espie Professor of Clinical Psychology and Head of Department of Psychological Medicine, University of Glasgow, UK
Chapter 4.14.2

Jonathan J. Evans Section of Psychological Medicine, University of Glasgow, Glasgow, UK
Chapter 4.1.14

Christopher G. Fairburn Wellcome Principal Research Fellow and Professor of Psychiatry, University of Oxford, UK
Chapters 4.10.2 and 6.3.2.2

Peter Falkai Professor of Medical Psychology, Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany
Chapter 2.3.5

Stephen V. Faraone Director, Medical Genetics Research, Professor of Psychiatry and of Neuroscience & Physiology, Director, Child and Adolescent Psychiatry Research, SUNY Upstate Medical University, New York
Chapter 4.3.9

Michael Farrell Senior Lecturer and Consultant Psychiatrist, National Addiction Centre, South London and Maudsley NHS Trust, London, UK
Chapter 4.2.3.4

David P. Farrington Professor of Psychological Criminology, University of Cambridge, UK
Chapter 11.2

J. Paul Fedoroff Director, Sexual Behaviors Clinic Royal Ottawa Mental Health Centre and Director of Forensic Research University of Ottawa Institute of Mental Health Research
Chapter 4.11.3

Howard H. Feldman Professor and Head, Division of Neurology, Department of Medicine, University of British Columbia, Vancouver, BC, Canada
Chapter 8.5.1.1

Melanie J.V. Fennell Consultant Clinical Psychologist; Director, Oxford Diploma in Cognitive Therapy, University of Oxford Department of Psychiatry, Warneford Hospital, Oxford, UK
Chapter 6.3.2.3

Max Fink Emeritus Professor of Psychiatry and Neurology, State University of New York at Stony Brook; Professor of Psychiatry, Albert Einstein College of Medicine; Attending Psychiatrist, Long Island Jewish Medical Center, New York, USA
Chapter 6.2.10.1

Michael B. First Columbia University, New York, USA
Chapter 1.9

Per Fink Director, Research Unit for Functional Disorders, Aarhus University Hospital, Risskov, Denmark
Chapter 5.2.3

Philip A. Fisher Research Scientist, Oregon Social Learning Center, Eugene, Oregon, USA
Chapter 9.5.4

Martine F. Flament Chargée de Récherche INSERM, CNRS UMR 7593, Paris, France
Chapter 9.2.8

Simon Fleminger Consultant Neuropsychiatries, Lishman Brain Injury Unit, Maudsley Hospital, London, UK
Chapters 4.1.10 and 4.1.13

Jonathan Flint Wellcome Trust Centre for Human Genetics Roosevelt Drive, Oxford
Chapter 2.4.2

Susan Folstein Professor of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, USA
Chapter 4.1.7

Peter Fonagy Freud Memorial Professor of Psychoanalysis, University College London; Director of Research, Anna Freud Centre, London, UK; Director, Child and Family Center and Clinical Protocols and Outcomes Center, Menninger Clinic, Topeka, Kansas, USA
Chapters 4.12.6, 6.3.5 and 9.5.2

Gillian C. Forrest Consultant Child and Adolescent Psychiatrist
Chapter 9.5.8

Leah Fostick Department of Psychiatry, Chaim Sheba Medical Centre, Tel Hashomer, Israel
Chapter 4.8

W. Fraser Division of Psychological Medicine, University of Wales College of Medicine, Cardiff, UK
Introduction to Section 10

Laura Fratiglioni Aging Research Centre, Karolinska Institute, Stockholm, Sweden
Chapter 8.3

David M. Fresco Adult Anxiety Clinic of Temple University, Philadelphia, Pennsylvania, USA
Chapter 4.7.2

K.W.M. Fulford Professor of Philosophy and Mental Health, University of Warwick; Honorary Consultant Psychiatrist, University of Oxford, UK
Chapter 1.5.2

Robert Fung, Specialist Registrar in Psychiatry, Sheffield Care NHS Trust, UK
Chapter 4.2.3.2

Pier Maria Furlan Director of Department of Mental Health San Luigi Gonzaga Hospital - University of Torino, Italy
Chapter 5.5

Glen O. Gabbard Bessie Walker Callaway Distinguished Professor of Psychoanalysis and Education in the Kansas School of Psychiatry, Menninger Clinic, Topeka; Clinical Professor of Psychiatry of Kansas School of Medicine, Wichita, Kansas, USA
Chapter 3.3

Jean Garrabé Honorary President of L'Evolution psychiatrique, Paris
Chapter 4.3.10

M. Elena Garralda Professor of Child and Adolescent Psychiatry, Imperial College of Medicine, London, UK
Chapter 9.3.4

Ann Gath Formerly of University College London, UK
Chapter 10.8

John R. Geddes Professor of Epidemiological Psychiatry, Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford
Chapters 1.10 and 4.5.1

David Gill Research Fellow, Department of Psychiatry, University of Oxford, UK
Chapters 5.2.9

Kathryn J. Gill MUHB Addictions Unit, McGill University, Montreal QC, Canada
Chapter 4.2.2.1

Julia Gledhill Clinical Research Fellow, Imperial College of Medicine, London, UK
Chapter 9.3.4

David Goldberg Director of Research and Development, Institute of Psychiatry, King's College London, UK
Chapter 7.8

Terry E. Goldberg The Zucker Hillside Hospital, Glen Oaks NY, USA
Chapter 2.5.3

Guy Goodwin Professor, University Department, Warneford Hospital, Oxford
Chapter 4.5.6

Kevin Gournay Emeritus Professor, Institute of Psychiatry, King's College London
Chapter 6.4.2

Simon G. Gowers Professor of Adolescent Psychiatry, University of Liverpool
Chapter 9.5.9

Cynthia A. Graham, Oxford Doctoral Course in Clinical Psychology Warneford Hospital, Oxford and The Kinsey Institute for Research in Sex, Gender, & Reproduction
Chapter 4.11.2

Philip Graham Emeritus Professor of Child Psychiatry, Institute of Child Health, London
Chapter 9.5.3

P.M. Grasby Senior Lecturer, MRC Cyclotron Unit, Hammersmith Hospital, London, UK
Chapter 2.3.6

Richard Green Head, Gender Identity Clinic, and Visiting Professor of Psychiatry, Imperial College of Medicine at Charing Cross Hospital, London, UK; Emeritus Professor of Psychiatry, University of California, Los Angeles, California, USA
Chapters 4.11.4 and 9.2.12

Stephen Green Clinical Professor of Psychiatry, Georgetown University School of Medicine, Washington, D.C.
Chapter 1.5.1

Michael Gunn Professor of Law and Head of Department, Department of Academic Legal Studies, Nottingham Law School, Nottingham Trent University, UK
Chapter 11.1

Lars Gustafson Professor of Geriatric Psychiatry, Lund University Hospital, Lund, Sweden
Chapter 4.1.3

Francesca Guzzetta Bologna, Italy
Chapter 4.12.4

Jennifer Gutstein Department of Child Psychiatry, College of Physicians and Surgeons, Columbia University, New York, USA
Chapter 9.2.10

Sarah Harper Oxford Institute for Aging, University of Oxford, Oxford, UK
Chapter 8.2

Rex Haigh Project Lead, Community of Communities, Centre for Quality Improvement, Royal College of Psychiatrists, London; Consultant Psychiatrist, Berkshire Healthcare NHS Foundation Trust
Chapter 6.3.9

S.A. Hales Psychiatry Fellow, Princess Margaret Hospital, University Health Network, Toronto, Canada
Chapter 5.3.6

John N. Hall Professor of Mental Health, School of Health and Social Care, Oxford Brookes University, Oxford, UK
Chapter 1.8.3

Wayne Hall Professor of Public Health Policy, University of Queensland, Herston, Australia
Chapter 4.2.3.7

Edwin Harari Consultant Psychiatrist, St Vincent's Hospital, Melbourne, Australia
Chapter 6.3.8

Sarah Harper Oxford Institute for Aging, University of Oxford, Oxford, UK
Chapter 8.2

James C. Harris Director Developmental Neuropsychiatry Clinic, Professor of Psychiatry and Behavioral Sciences, Pediatrics, and Mental Hygiene, The Johns Hopkins University School of Medicine
Chapter 9.2.1

Paul J. Harrison Clinical Reader in Psychiatry, University of Oxford Department of Psychiatry, Warneford Hospital, Oxford, UK
Chapter 4.3.6.2

Allison G. Harvey Department of Experimental Psychology, University of Oxford, UK
Chapter 4.6.1

Philip D. Harvey Professor of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Woodruff Memorial Building, Atlanta, GA, USA
Chapter 4.3.3

Keith Hawton Director, Centre for Suicide Research, University Department of Psychiatry, Warneford Hospital, Oxford
Chapter 4.15.4

Richard G. Heimberg Adult Anxiety Clinic of Temple University, Philadelphia, Pennsylvania, USA
Chapter 4.7.2

Scott Henderson Emeritus Professor, The Australian National University, Canberra, Australia
Chapters 2.7 and 8.3

George R. Heninger Professor, Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut, USA
Chapter 6.2.3

Leslie Hicks, University of York, UK
Chapter 9.5.6

A.J. Holland Lecturer, Department of Psychiatry, University of Cambridge, UK
Chapter 10.1

Jimmie C. Holland Wayne E. Chapman Chair in Psychiatric Oncology, Department of Psychiatry and Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, USA
Chapter 5.3.7

Sheila Hollins Professor of Psychiatry of Learning Disability, Department of Psychiatry and Disability, St George's Hospital Medical School, University of London, UK
Chapter 10.7

Jeremy Holmes Consultant Psychiatrist/Psychotherapist, North Devon District Hospital, Barnstaple; Senior Lecturer, University of Bristol, UK
Chapter 3.2

Suzanne Holroyd Professor, Director of Geriatric Psychiatry, Department of Psychiatry and Neurobehavioral Science, University of Virginia, Charlottesville VA
Chapter 8.5.6

Geraldine Holt Honorary Senior Lecturer in Psychiatry at the Institute of Psychiatry, King's College London, UK
Chapter 10.9

Allan House Professor of Liaison Psychiatry, University of Leeds, UK
Chapter 5.3.1

Daniel H. Hovelson The Harvard program in refugee trauma, Massachusetts general hospital, Dept of psychiatry.
Chapter 7.10.1

Jane Hubert Senior Lecturer in Social Anthropology, Department of Psychiatry and Disability, St George's Hospital Medical School, University of London, UK
Chapter 10.7

Stephen Hucker University of Toronto, Toronto, Canada
Chapter 11.8

T. A. Hughes Consultant Psychiatrist, St Mary's Hospital, Leeds, UK
Chapter 4.1.6

Masud Husain Institute of Neurology & Institute of Cognitive Neuroscience, UCL, London and National Hospital for Neurology & Neurosurgery, London
Chapter 2.5.2

Frits J. Huyse Psychiatrist, Consultant integrated care, Department of General Internal Medicine, University Medical Centre Groningen (UMCG), Groningen, The Netherlands
Chapter 5.7

Matti Iivanainen Professor, Department of Child Neurology, University of Helsinki, Finland
Chapter 10.5.3

Leslie Iversen Visting Professor, Department of Pharmacology, University of Oxford, UK
Chapter 6.2.7

Richard Ives National Children's Bureau, London, UK
Chapter 4.2.3.6

Assen Jablensky Professor of Psychiatry, University of Western Australia, Perth, Australia
Chapters 4.3.5 and 4.3.7

Robin Jacoby Clinical Reader in the Psychiatry of Old Age, University of Oxford, UK
Chapters 8.4 and 8.5.7

Claudia Jacova Assistant Professor, Division of Neurology, Department of Medicine, University of British Columbia, Vancouver, BC, Canada
Chapter 8.5.1.1

David V. James Consultant Forensic Psychiatrist, North London Forensic Service and Fixated Threat Assessment Centre
Chapter 11.6

Kay Redfield Jamison Professor of Psychiatry, Johns Hopkins School of Medicine, Baltimore, Maryland, USA
Chapter 1.1

John G.R. Jefferys Department of Neurophysiology, Division of Neuroscience, University of Birmingham, UK
Chapter 2.3.9

Andrew Johns Consultant Forensic Psychiatry and Honorary Senior Lecturer, Maudsley Hospital, London, UK
Chapter 11.3.2.

E.C. Johnstone Professor of Psychiatry and Head, Department of Psychiatry, University of Edinburgh, UK
Chapter 4.3.8

David P.H. Jones Senior Clinical Lecturer in Child Psychiatry, Park Hospital for Children, University of Oxford, UK
Chapter 9.3.3

Peter R. Joyce Professor, Department of Psychological Medicine, Christchurch School of Medicine, Christchurch, New Zealand
Chapter 4.5.4

Elizabeth Juven-Wetzler Department of Psychiatry, Chaim Sheba Medical Centre, Tel Hashomer, Israel
Chapter 4.8

Horst Kachele Universitätsklinik Psychosomatische Medizin and Psychotherapie Universitätsklinik Ulm, Germany
Chapter 6.3.5

Adam Ian Kaplin Assistant Professor, Departments of Psychiatry and Neurology, Johns Hopkins University School of Medicine, Johns Hopkins Hospital, Baltimore, MD
Chapter 1.1

Markus Kaski Director, Rinnekoti Research Foundation, Director and Chief Physician of Rinnekoti Foundation, Espoo, Finland
Chapter 10.3

Aliya Kassam Institute of Psychiatry, David Glodberg Centre, De Crespigny Park, London, UK
Chapter 1.2

Roger G. Kathol, Adjunct Professor of Internal Medicine and Psychiatry, University of Minnesota, President, Cartesian Solutions, Inc. Burnsville, MN, USA
Chapter 5.7

Paul E. Keck Jr. Lindner Center of HOPE, Mason, and Department of Psychiatry, University of Cincinnati College of Medicine, Cincinnati, OH, USA
Chapter 4.13.1

John Kellett St George's Hospital Medical School, London, UK
Chapter 8.5.8

Keith W. Kelley Department of Animal Sciences, University of Illinois, Urbana-Champaign, USA
Chapter 2.3.10

David Kennard Chair of the UK Network of the International Society for the Psychological Treatments of the Schizophrenias and other psychoses (ISPS UK); former Head of Psychology Services, The Retreat, York, UK
Chapter 6.3.9

Ad.J.F.M. Kerkhof Professor of Clinical Psychology, Vrije Universiteit, Amsterdam, The Netherlands
Chapter 4.15.2

Otto F. Kernberg Professor of Psychiatry, Cornell University Medical College, New York; Training and Supervising Analyst, Columbia University Center for Psychoanalytic Training and Research, New York, USA
Chapter 3.1

Falk Kiefer Professor of Addiction Research, Deputy Director, Department of Addictive Behaviour and Addiction Medicine, Central Institute of Mental Health CIMH, University of Heidelberg, Mannheim, Germany
Chapter 4.2.2.3

Arthur Kleinman Presley Professor of Anthropology and Psychiatry, Harvard University; Chair, Department of Social Medicine, Harvard Medical School, Cambridge, Massachusetts, USA
Chapter 2.6.2

Ami Klin Yale University, New Haven, Connecticut, USA
Chapter 9.2.3

Kimberly Klipstein Department of Psychiatry, Mount Sinai School of Medicine, New York, USA
Chapter 4.6.4

Martin Knapp Institute of Psychiatry, King's College London; London School of Economics and Political Science, University of London, UK
Chapter 7.7

Michael D. Kopelman Professor of Neuropsychiatry at King's College London, Institute of Psychiatry, UK
Chapter 4.1.12

Malcolm Lader Emeritus Professor of Clinical Psychopharmacology, King's College London, Institute of Psychiatry, Denmark Hill, London, UK
Chapter 6.2.2

Fergus D. Law Honorary Senior Registrar and Clinical Lecturer, Psychopharmacology Unit, University of Bristol, UK
Chapters 4.2.1 and 6.2.8

Brian Lawlor Conolly Norman Professor of Old Age Psychiatry, St. James's Hospital & Trinity College, Dublin, Ireland
Chapter 8.5.2

Dusica Lecic-Tosevski Professor of Psychiatry, Institute of Mental Health, School of Medicine, University of Belgrade, Belgrade, Serbia
Chapter 4.12.3

Julian Leff Emeritus Professor, Department of Psychological Medicine, Institute of Psychiatry, King's College London, UK
Chapter 1.3.2

R.J. Levin Department of Biomedical Science, University of Sheffield, UK
Chapter 4.11.1

Rhohel Lenroot Child Psychiatry Branch, NIMH, Bethesda MD, USA
Chapter 9.1.4

Peter F. Liddle Professor of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada
Chapter 4.3.2

Felice Lieh Mak Emeritus Professor, Department of Psychiatry, University of Hong Kong, Hong Kong
Chapter 5.2.10

James Lindesay Professor of Psychiatry for the Elderly, University of Leicester, UK
Chapters 8.5.1 and 8.5.5

Jouko K. Lonnqvist Professor, National Public Health Institute, Helsinki, Finland
Chapter 4.15.1

Juan J. López-Ibor Jr. Chairman, Department of Psychiatry, San Carlos University Hospital, Complutense University, Madrid, Spain
Chapters 1.6 and 4.12.1

Mª Inés López-Ibor Madrid, Spain
Chapter 4.2.3.8

Simon Lovestone Professor of Old Age Psychiatry, NIHR Biomedical Research Centre for Mental Health, MRC Centre for Neurodegeneration Research, Departments of Psychological Medicine and Neuroscience, King's College London, Institute of Psychiatry, London, UK
Chapter 4.1.2

Leanne Magee Temple University, Philadelphia, Pennsylvania, USA
Chapter 4.7.2

Andrew McCulloch The Mental Health Foundation, London, UK
Chapter 7.1

Jane McCarthy Division of Mental Health, St George's Hospital, London, UK
Chapter 10.8

Susan L. McElroy Lindner Center of HOPE, Mason, and Department of Psychiatry, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA
Chapter 4.13.1

Peter McGuffin Director and Professor of Psychiatric Genetics, Institute of Psychiatry, King's College London, UK
Chapter 2.4.1

I.G. McKeith Clinical Director, Institute for Ageing and Health, Newcastle University, Newcastle Upon Tyne, UK
Chapter 4.1.5

Kwame McKenzie Centre for Addictions and Mental Health, Toronto, Canada; University of Toronto, Canada; University of Central Lancashire, UK
Chapter 7.3

Declan McLoughlin Institute of Psychiatry, King's College London, UK
Chapter 6.2.10.3

Mark W. Mahowald Director, Minnesota Regional Sleep Disorders Center, Hennepin County Medical Center; Professor of Neurology, University of Minnesota Medical School, Minneapolis, Minnesota, USA
Chapter 4.14.4

Mario Maj Institute of Psychiatry, University of Naples, Italy
Chapter 4.1.9

Ulrik Fredrik Malt Professor of Psychiatry (Psychosomatic Medicine), National Hospital, University of Oslo, Norway
Chapter 5.3.8

J. John Mann Vice Chair for Research Scientific Director, Kreitchman PET Center, Columbia University and Chief, Division of Molecular Imaging & Neuropathology, New York State Psychiatric Institute, USA
Chapter 4.15.3

Karl F. Mann Professor and Chair in Addiction Research, Deputy Director Central Institute of Mental Health (CIMH), University of Heidelberg, Mannheim, Germany
Chapter 4.2.2.3

Russell L. Margolis Professor of Psychiatry and Neurology Director, Johns Hopkins Schizophrenia Program Director, Laboratory of Genetic Neurobiology Division of Neurobiology, Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, USA
Chapter 4.1.7

John C. Markowitz Associate Professor of Psychiatry, Weill Medical College of Cornell University; Director, Psychotherapy Clinic, Payne Whitney Clinic, New York Presbyterian Hospital, New York, USA
Chapter 6.3.3

John Marsden Lecturer, Institute of Psychiatry, King's College London, UK
Chapter 4.2.4

Jane Marshall Senior Lecturer in the Addictions, National Addiction Centre, Institute of Psychiatry, King's College London, UK
Chapters 4.1.11 and 4.2.2.2

Andrés Martin Professor of Child Psychiatry, Child Study Center Yale University School of Medicine, New Haven, Connecticut, USA
Chapter 9.1.3

Keith Matthews Dept of Psychiatry, University of Dundee, Dundee, UK
Chapter 6.2.10.4

Barbara Maughan MRC Child Psychiatry Unit, Institute of Psychiatry, King's College London, UK
Chapter 9.3.1

Soraya Mayet National Addiction Centre, Institute of Psychiatry, King's College London, UK
Chapter 4.2.3.1

Richard Mayou Emeritus Professor of Psychiatry, University of Oxford, UK
Chapter 5.2.1

Nick Medford Institute of Psychiatry, King's College London, UK
Chapter 4.9

David Meagher Dept of Adult Psychiatry, Midwestern Regional Hospital, Limerick, Ireland
Chapter 4.1.1

Thomas W. Meeks Division of Geriatric Psychiatry, University of California San Diego, La Jolla CA, USA
Chapter 8.5.3

Pamela S. Melding Honorary Senior Lecturer, Department of Psychological Medicine, University of Auckland, New Zealand and Consultant in Psychiatry of Old Age, Mental Health Serviced, North Shore Hospital, Waitemata District Health Board, Takapuna, North Shore City, Auckland, New Zealand
Chapter 8.7

Herbert Y. Meltzer Bixler/May/Johnaon Professor of Psychiatry, Professor of Pharmacoloqy Vanderbilt University School of Medicine, Nashville, Tennessee, USA
Chapter 6.2.5

Julien Mendlewicz Department of Psychiatry, University Clinics of Brussels, Erasme Hospital, Brussels, Belgium
Chapter 4.5.5

Andreas Meyer-Lindenberg Dept of Psychiatry, Central Institute of Mental Health, Mannheim, Germany
Chapter 2.5.3

Gillian C. Mezey Consultant and Senior Lecturer in Forensic Psychiatry, Traumatic Stress Service, St George's Hospital Medical School, London, UK
Chapters 11.12 and 11.13

R.H.S. Mindham Emeritus Professor of Psychiatry, University of Leeds, UK
Chapter 4.1.6

Andrew Mogg Institute of Psychiatry, King's College London, UK
Chapter 6.2.10.3

Richard F. Mollica Director, Harvard Program in Refugee Trauma; Associate Professor of Psychiatry, Harvard Medical School and Harvard School of Public Health, Cambridge, Massachusetts, USA
Chapter 7.10.1

Emanuel Moran Consultant Psychiatrist, Grovelands Priory Hospital, London, UK
Chapter 4.13.2

Stella Morris Dept of Psychological Medicine, Hull Royal Infirmary, Hull, UK
Chapter 5.2.6

Matt Muijen WHO Regional Office for Europe, Copenhagen, Denmark
Chapter 7.1

Paul E. Mullen Professor of Forensic Psychiatry, Monash University; Clinical Director, Victorian Institute of Forensic Mental Health, Monash University, Melbourne, Australia
Chapters 11.10, 11.11 and 11.14

Alistair Munro Emeritus Professor of Psychiatry, Dalhousie University, Halifax, Nova Scotia, Canada
Chapter 4.4

Rebecca Murphy Research Psychologist, Oxford University Department of Psychiatry, Warneford Hospital, Oxford, UK
Chapters 4.10.2 and 6.3.2.2

Lynne Murray Winnicott Research Unit, University of Reading, Reading, UK
Chapter 9.3.6

R.M. Murray Institute of Psychiatry, King's College London, UK
Chapter 4.3.6.1

Norbert Nedopil Professor of Forensic Psychiatry, Head of the Department of Forensic Psychiatry at the Psychiatric Hospital of the University of Munich, Munich, Germany
Chapter 11.3.3

Juan C. Negrete Professor and Head, Addictions Psychiatry Program, University of Toronto, Canada
Chapter 4.2.2.1

Gretchen N. Neigh Dept of Psychiatry and Behavioral Sciences, Emory University, Atlanta GA, USA
Chapter 2.3.3

Charles B. Nemeroff Reunette W. Harris Professor and Chairman, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia, USA
Chapter 2.3.3

Giles Newton-Howes Division of Neurosciences and Mental Health, Imperial College School of Medicine, London, UL
Chapter 4.12.7

Jeffrey Newcorm Mount Sinai School of Medicine, New York, USA
Chapter 4.6.4

Russell Noyes Jr. Department of Psychiatry, University of Iowa College of Medicine, Iowa City, Iowa, USA
Chapter 5.2.5

David J. Nutt Professor of Psychopharmacology and Head of Clinical Medicine, University of Bristol, UK
Chapters 4.2.1 and 6.2.8

Margaret Oates Senior Lecturer in Psychiatry, University of Nottingham, UK
Chapter 1.8.1

James R.P. Ogloff Victorian Institute of Forensic Mental Health, Thomas Embling Hospital, Fairfield VIC, Australia
Chapters 11.4 ,11.14, and 11.16

John O'Grady Knowle Hospital, Fareham, UK
Chapter 11.15

Catherine Oppenheimer Consultant Psychiatrist, Warneford Hospital, Oxford, UK
Chapter 8.5.8 and 8.6

Luca Ostacoli Liaison Psychiatry and Psychosomatic Unit, Department of Mental Health, San Luigi Gonzaga Hospital - University of Torino, Italy
Chapter 5.5

Pierre Oswald Dept of Psychiatry, ULB Erasme, Brussels, Belgium
Chapter 4.5.5

Barton W. Palmer Veterans Affairs Medical Center, University of California, San Diego CA, USA
Chapter 8.5.3

Gordon Parker Professor, University of New South Wales; and Executive Director, Black Dog Institute, Australia
Chapter 4.5.3

E.S. Paykel Emeritus Professor of Psychiatry, Department of Psychiatry, University of Cambridge, UK
Chapter 4.5.8

John B. Pearce Emeritus Professor of Child and Adolescent Psychiatry, University of Nottingham, UK
Chapters 9.4 and 9.5.1

†**R.C.A. Pearson** Department of Biomedical Science, University of Sheffield, UK
Chapter 2.3.1

Keith J. Petrie Associate Professor, School of Medicine,University of Auckland, New Zealand
Chapter 5.6

Cynthia R. Pfeffer Weill Medical College of Cornell University, New York Presbyterian Hospital-Westchester Division, White Plains, New York, USA
Chapter 9.2.10

Katharine A. Phillips Professor of Psychiatry and Human Behavior, The Warren Alpert Medical School of Brown University; Director, Body Dysmorphic Disorder Program, Butler Hospital, Providence, USA
Chapter 5.2.8

Pierre Pichot Académie Nationale de Médecine, Paris, France
Chapter 1.4

Harold Alan Pincus Columbia University, New York, USA
Chapter 1.9

Vanessa Pinfold 'Rethink', London, UK
Chapter 7.9

Daniel S. Pine Division of Intramural Research Programs, National Institutes of Health, Bethesda, USA
Chapter 9.2.6

Malcolm Pines Founding Member, Institute of Group Analysis, London, UK
Chapter 6.3.6

Harrison G. Pope Jr. Professor of Psychiatry, Harvard Medical School, Boston; Chief, Biological Psychiatry Laboratory, McClean Hospital, Belmont, Massachusetts, USA
Chapter 6.2.6

Robert M. Post Chief, Biological Psychiatry Branch, National Institute of Mental Health, Bethesda, Maryland, USA
Chapter 6.2.4

Graham E. Powell Psychology Services, Powell Campbell Edelmann, London, UK
Chapter 1.8.3

Herschel Prins Professor, Midlands Centre for Criminology and Criminal Justice, University of Loughborough, UK
Chapter 11.9

Paul Ramchandani Dept of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK
Chapter 9.3.6

Shulamit Ramon Professor of Interprofessional Health and Social Studies, Anglia Polytechnic University, Cambridge, UK
Chapter 6.4.3

Beverley Raphael University of Western Sydney Medical School, Sydney NSW, Australia
Chapter 4.6.5

James Reich Clinical Professor of Psychiatry, University of California, San Francisco Medical School and Adjunct Associate Professor of Psychiatry, Stanford School of Medicine
Chapter 4.12.2

Helmut Remschmidt Director, Department of Child and Adolescent Psychiatry, Philipps Universität, Marburg, Germany
Chapter 9.2.2

Philippe Robaey Institute of Mental Health Research, Royal Ottawa Hospital, Ottawa, Canada
Chapter 9.2.8

Ian Robbins Consultant Clinical Psychologist, St George's Hospital, London, UK
Chapter 11.13

T.W. Robbins Section of Forensic Psychiatry, St George's Hospital Medical School, London, UK
Chapter 2.5.5

G.M. Rodin Professor of Psychiatry, University of Toronto, Toronto, Canada
Chapter 5.3.6

Maria A. Ron Professor of Neuropsychiatry, Institute of Neurology, University College London, UK
Chapter 5.3.2

Robin Room Professor, School of Population Health, University of Melbourne; and Director, AER Centre for Alcohol Policy Research, Turning Point Alcohol and Drug Centre, Fitzroy, Victoria, Australia
Chapter 4.2.2.6

W. Rössler Professor of Clinical Psychiatry and Psychology, University of Zürich, Switzerland
Chapter 6.4.1

James R. Rundell Department of Psychiatry and Psychology, Mayo Clinic Professor of Psychiatry, Mayo Clinic College of Medicine
Chapter 5.3.4

Gerald Russell Emeritus Professor of Psychiatry, Director of the Eating Disorders Unit, Hayes Grove Priory Hospital, Hayes, Kent, UK
Chapter 4.10.1

B.J. Sahakian Dept of Psychiatry, University of Cambridge, Cambridge, UK
Chapter 2.5.5

Diana Sanders Chartered Counselling Psychologist, working in Psychological Medicine in Oxford, UK
Chapter 6.3.1

Paramala J. Santosh Great Ormond Street Hospital for Sick Children, London, UK
Chapter 9.5.5

Benedetto Saraceno Director of Department of Mental Health and Substance Abuse, World Health Organization WHO
Chapter 1.3.1

Gauri N. Savla, Veterans Affairs Medical Center, University of California, San Diego CA, USA
Chapter 8.5.3

Carlos H. Schenck Staff Psychiatrist, Minnesota Regional Sleep Disorders Center, Hennepin County Medical Center; Associate Professor of Psychiatry, University of Minnesota Medical School, Minneapolis, Minnesota, USA
Chapter 4.14.4

John Schlapobersky Consultant Psychotherapist, Trumatic Stress Clinic Middlesex/University College Hospital, formerly also of The Medical Foundation for the Care of Victims of Torture London, UK
Chapter 6.3.6

Fabrizio Schiffano, Chair in Clinical Pharmacology and Therapeutics Associate Dean, Postgraduate Medical School, Hon Consultant Psychiatrist Addictions, University of Hertfordshire, School of Pharmacy, College Lane Campus, Hatfield, UK
Chapter 4.2.3.5

Gerd Schulte-Körne Director of the Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University of Munich, Pettenkoferstr, München/Germany
Chapter 9.2.2

J. Scott Professor of Psychological Medicine, University of Newcastle & Honorary Professor, Psychological Treatments Research, Institute of Psychiatry, London and University Department of Psychiatry, Leazes Wing, Royal Victoria Infirmary, Newcastle upon Tyne, England
Chapter 4.5.8

Stephen Scott Professor of Child Health & Behaviour, King's College London, Institute of Psychiatry, and Director of Research National Academy for Parenting Practitioners, London, UK
Chapters 9.1.1 and 9.2.5

Nicholas Seivewright Consultant Psychiatrist in Substance Misuse, Community Health Sheffield NHS Trust, Sheffield, UK
Chapter 4.2.3.2

David Shaffer Department of Child Psychiatry, College of Physicians and Surgeons, Columbia University, New York, USA
Chapter 9.2.10

Trevor Sharp Dept of Pharmacology, University of Oxford, Oxford, UK
Chapter 2.3.4

Michael Sharpe Professor of Psychological Medicine & Symptoms Research, University of Edinburgh, UK
Chapters 5.1 and 5.2.7

Jennifer Shaw Centre for Suicide Prevention, The School of Medicine, University of Manchester, UK
Chapter 11.5

Mauricio Sierra Institute of Psychiatry, King's College London, UK
Chapter 4.9

Gregory Simon Investigator, Center for Health Studies, Group Health Cooperative, Seattle, Washington, USA
Chapter 5.2.2

Andrew Sims Professor of Psychiatry, University of Leeds, UK
Chapter 1.7

Ian Sinclair Professor of Social Work, University of York, UK
Chapter 9.5.6

Mike Slade Health Service and Population Research Department and Institute of Psychiatry, King's College London, UK
Chapter 7.2

Elita Smiley Consultant Psychiatrist and Clinical Senior Lecturer, Division of Community Based Sciences, Faculty of Medicine, University of Glasgow, UK
Chapter 10.2

Wolfgang Söllner Department of Psychosomatic Medicine and Psychotherapy General Hospital Nuremberg, Prof.Ernst-Nathan-Str. 1, Nürnberg, Germany
Chapter 5.7

Daniel Souery Department of Psychiatry, University Clinics of Brussels, Erasme Hospital, Brussels, Belgium
Chapter 4.5.5

Elizabeth Spencer Senior Clinical Medical Officer, Early Intervention Service, Northern Birmingham Mental Health Trust, Birmingham, UK
Chapter 6.3.2.4

David A. Spiegel Center for Anxiety and Related Disorders at Boston University, Boston, Massachusetts, USA
Chapter 4.7.1

Costas Stefanis Honorary Professor of Psychiatry, University of Athens, Greece
Chapter 1.6

Alan Stein Royal Free and University College Medical School, University College London, and Tavistock Clinic, London, UK
Chapter 9.3.6

Jessica Stiles Department of Psychiatry and Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, USA
Chapter 5.3.7

William S. Stone Assistant Professor of Psychology, Director of Neuropsychology Training and Clinical Services, Department of Psychiatry, Harvard Medical School, Massachusetts Mental Health Center Public Psychiatry, Division of the Beth Israel Deaconess Medical Center, Boston, USA
Chapter 4.3.9

Gregory Stores Emeritus Professor of Developmental Neuropsychiatry, University of Oxford, UK
Chapters 4.14.1 and 9.2.9

Elizabeth A. Stormshak Assistant Professor, University of Oregon, Eugene, Oregon, USA
Chapter 9.5.4

James J. Strain Professor/Director, Behavioral Medicine and Consultation Psychiatry, Mount Sinai School of Medicine, New York, USA
Chapter 4.6.4

John Strang National Addiction Centre, Institute of Psychiatry, King's College London, UK
Chapters 4.2.3.1 and 4.2.4

J. Suckling Brain Mapping Unit, Department of Psychiatry, University of Cambridge, Addenbrookes Hospital, Cambridge, UK
Chapters 2.3.7 and 2.3.8

Nicola Swinson Centre for Sucide Prevention, The School of Medicine, University of Manchester, UK
Chapter 11.5

Michele Tansella Professor of Psychiatry and Chairman, Department of Medicine and Public Health, Section of Psychiatry, University of Verona, Italy
Chapters 7.2 and 7.6

Mary Target Senior Lecturer in Psychoanalysis, Psychoanalysis Unit, University College London; Deputy Director of Research, Anna Freud Centre, London, UK
Chapter 9.5.2

Eric Taylor Head of Department, Child & Adolescent Psychiatry, King's College London, Institute of Psychiatry
Chapter 9.2.4

John-Paul Taylor Academic Specialist Registrar, Institute for Ageing and Health Newcastle University, Campus for Ageing and Vitality, Newcastle upon Tyne, UK
Chapter 4.1.13

Pamela J. Taylor School of Medicine, Cardiff University, Cardiff, UK
Chapter 11.17

Tatiana Taylor Dept of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK
*Chapter 4.15.*4

Mary Teasdale 'Rethink', London, UK
Chapter 7.9

Anita Thapar Department of Psychological Medicine, School of Medicine, Cardiff University, UK
Chapter 2.4.1

June Thoburn Emeritus Professor of Social Work, University of East Anglia, Norwich, UK
Chapter 9.3.5

Sian Thomas Chester Young People's Centre, Shester, UK
Chapter 9.5.9

Lindsay Thomson Division of Psychiatry, University of Edinburgh, Edinburgh, UK
Chapter 11.3.1

Anne E. Thompson Emeritus Professor Child and Adolescent Psychiatry, University of Nottingham, UK
Chapter 9.4

Graham Thornicroft Professor of Community Psychiatry, Institute of Psychiatry, King's College London, UK
Chapters 1.2, 7.2 and 7.6

Adolf Tobeña Professor of Psychiatry, Director of the Dept. of Psychiatry and Forensic Medicine, Autonomous University of Barcelona, Bellaterra (Barcelona), Spain
Chapter 4.12.5

Bruce J. Tonge Head Monash University School of Psychology Psychiatry & Psychological Medicine, Monash Medical Centre, Clayton, Victoria, Australia
Chapter 10.5.1

Brian Toone Consultant, Maudsley Hospital; Honorary Senior Lecturer, Institute of Psychiatry, King's College London, UK
Chapter 5.3.3

David Trickey Leicester Royal Infirmary, Leicestershire Partnership NHS Trust, UK
Chapters 9.3.2 and 9.3.7

Paula Trzepacz Eli Lilly & Co, USA
Chapter 4.1.1

Wen-Shing Tseng Professor at Department of Psychiatry, University of Hawaii School of Medicine, USA
Chapters 4.16 and 6.5

Ming T. Tsuang Behavioral Genomics Endowed Chair and University Professor, University of California; Distinguished Professor of Psychiatry and Director, Center for Behavioral Genomics, Department of Psychiatry, University of California, San Diego, CA, USA
Chapter 4.3.9

André Tylee Director, Royal College of General Practitioners Unit for Mental Health Education in Primary Care, Institute of Psychiatry, King's College London, UK
Chapter 7.8

Amy M. Ursano Department of Psychiatry, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, North Carolina, USA
Chapter 6.3.4

Robert J. Ursano Professor and Chairman, Department of Psychiatry, Uniformed Services University of the Health Sciences, F. Edward Herbert School of Medicine, Bethesda, Maryland, USA
Chapter 6.3.4

Jim van Os Professor of Psychiatric Epidemiology, Maastricht University, Maastricht, The Netherlands and Visiting Professor of Psychiatric Epidemiology Institute of Psychiatry, London, UK
Chapter 7.3

Fred R. Volkmar Yale University, New Haven, Connecticut, USA
Chapter 9.2.3

Jane Walker Clinical Lecturer and Honorary Specialist Registrar in Liaison Psychiatry, Psychological Medicine & Symptoms Research Group, School of Molecular & Clinical Medicine, University of Edinburgh, UK
Chapter 5.1

Diane Waller Professor of Art Psychotherapy, Goldsmiths, University of London
Chapter 6.4.4

Paul Walters MRC Fellow & Specialist Psychiatrist , Programme Leader MSc in Mental Health Services Research, Section of Primary Care Mental Health, Health Service and Population Research Department, David Goldberg Centre, Institute of Psychiatry, London, UK
Chapter 7.8

John Weinmann Professor of Psychology as applied to Medicine, Institute of Psychiatry, King's College London, UK
Chapter 5.6

Myrna M. Weissman Professor of Epidemiology in Psychiatry, College of Physicians and Surgeons of Columbia University; Chief, Division of Clinical and Genetic Epidemiology, New York State Psychiatric Institute, New York, USA
Chapter 6.3.3

Sarah Welch Gloucestershire Partnership NHS Foundation Trust, UK
Chapter 4.2.3.4

Ursula Werneke Consultant Psychiatrist, Norrkoping, Sweden
Chapter 6.2.9

Simon Wessely Professor of Epidemiological and Liaison Psychiatry, Institute of Psychiatry, King's College London, UK
Chapter 5.2.7

Kay Wheat Senior Lecturer in Law, Department of Academic Legal Studies, Nottingham Law School, Nottingham Trent University, UK
Chapter 11.1

Adam R. Winstock Senior Staff Specialist, Drug Health Services, Conjoint Senior Lecturer, National Drug and Alcohol Research Centre, UNSW
Chapters 4.2.3.1 and 4.2.3.5

Sally Wooding Senior Research Fellow, SciMHA Unit, University of Western Sydney, Australia
Chapter 4.6.5

Matt Woolgar Institute of Psychiatry, King's College London, UK
Chapter 2.5.1

Miranda Wolpert Director of Child and Adolescent Mental Health Services, Evidence Based Practice Unit, University College London and Anna Freud Centre, UK
Chapter 9.5.7

Lawson Wulsin Professor of Psychiatry and Family Medicine, University of Cincinnati, OH, USA
Chapter 5.7

Richard Jed Wyatt† National Institutes of Mental Health, Bethesda, Maryland, USA
Chapter 1.1

William Yule Professor of Applied Child Psychology, Institute of Psychiatry, King's College London, UK
Chapter 2.5.1

Karl Zilles Professor, Institute of Neuroscience and Biophysics, INB-3 Research Centre, Jülich and C.&O. Vogt Institute of Brain Research, University Düsseldorf, Germany
Chapter 2.3.2

Joseph Zohar Psychiatric Medical Center, Sheba Medical Center, Tel Hashomer and Sackler School of Medicine, Tel Aviv University, Israel
Chapter 4.8

SECTION 1

The Subject Matter of and Approach to Psychiatry

This page intentionally left blank

1.1

The patient's perspective

Kay Redfield Jamison, Richard Jed Wyatt,† and Adam Ian Kaplin

It is difficult to be a psychiatric patient, but a good doctor can make it less so. Confusion and fear can be overcome by knowledge and compassion, and resistance to treatment is often, although by no means always, amenable to change by intelligent persuasion. The devil, as the fiery melancholic Byron knew, is in the details.

Patients, when first given a psychiatric diagnosis, are commonly both relieved and frightened—relieved because often they have been in pain and anxiety for a considerable period of time, and frightened because they do not know what the diagnosis means or what the treatment will entail. They do not know if they will return to the way they once were, whether the treatment they have been prescribed will or will not work, and, even if it does work, at what cost it will be to them in terms of their notions of themselves, potentially unpleasant side-effects, and the reactions of their family members, friends, colleagues, and employers. Perhaps most disturbing, they do not know if their depression, psychosis, anxieties, or compulsions will return to become a permanent part of their lives. Caught in a state often characterized by personal anguish, social isolation and confusion, newly diagnosed patients find themselves on a quest to regain a sense of mastery of themselves and their surroundings. One of the main goals of therapies of all types is to empower the patient and give them some control back over their world.

The specifics of what the doctor says, and the manner in which he or she says it, are critically important. Most patients who complain about receiving poor psychiatric care do so on several grounds: their doctors, they feel, spend too little time explaining the nature of their illnesses and treatment; they are reluctant to consult with or actively involve family members; they are patronizing, and do not adequately listen to what the patient has to say; they do not encourage questions or sufficiently address the concerns of the patient; they do not discuss alternative treatments, the risks of treatment, and the risks of no treatment; and they do not thoroughly forewarn about side-effects of medications.

Most of these complaints are avoidable. Time, although difficult to come by, is well spent early on in the course of treatment when confusion and hopelessness are greatest, non-adherence is highest, and the possibility of suicide substantially increased. Hope can be realistically extended to patients and family members, and its explicit extension is vital to those whose illnesses have robbed them not only of hope, but of belief in themselves and their futures. The hope provided needs to be tempered, however, by an explication of possible difficulties yet to be encountered: unpleasant side-effects from medications, a rocky time course to meaningful recovery which will often consist of many discouraging cycles of feeling well, only to become ill again, and the probable personal, professional, and financial repercussions that come in the wake of having a psychiatric illness.

It is terrifying to lose one's sanity or to be seized by a paralysing depression. No medication alone can substitute for a good doctor's clinical expertise and the kindness of a doctor who understands both the medical and psychological sides of mental illness. Nor can any medication alone substitute for a good doctor's capacity to listen to the fears and despair of patients trying to come to terms with what has happened to them. A good doctor is a therapeutic optimist who is able to instill hope and confidence to combat confusion and despair. Great doctors are able to provide the unwavering care to their patients that they would want a member of their own family to receive, blending empathy, and compassion with expertise.

Doctors need to be direct in answering questions, to acknowledge the limits of their understanding, and to encourage specialist consultations when the clinical situation warrants it. They also need to create a therapeutic climate in which patients and their families feel free, when necessary, to express their concerns about treatment or to request a second opinion. Treatment non-adherence, one of the major causes of unnecessary suffering, relapse, hospitalization, and suicide, must be addressed head-on. Young males, early in the course of their illness, are particularly likely to stop medication against medical advice, and the results can be lethal.[1,2] Unfortunately, doctors are notoriously variable in their ability to assess and predict adherence in their patients.[3]

Asking directly and often about medication concerns and side-effects, scheduling frequent follow-up visits after the initial diagnostic evaluation and treatment recommendation, and encouraging adjunctive psychotherapy, or involvement in patient support groups, can make a crucial difference in whether or not a patient takes medication in a way that is most effective. Aggressive treatment of unpleasant or intolerable side-effects, minimizing the dosage and number of doses, and providing ongoing, frequently repetitive

†Deceased.

education about the illness and its treatment are likewise essential, if common-sense, ways to avert or minimize non-adherence.

Education is, of course, integral to the good treatment of any illness, but this is especially true when the illnesses are chronic. The term 'doctor' derives originally from the Latin word for teacher, and it is in their roles as teachers that doctors provide patients with the knowledge and understanding to combat the confusion and unpredictability that surrounds mental illness. Patients and their family members should be encouraged to write down any questions they may have, as many individuals are intimidated once they find themselves in a doctor's office. Any information that is given orally to patients should be repeated as often as necessary (due to the cognitive difficulties experienced by many psychiatric patients, especially when acutely ill or recovering from an acute episode) and, whenever feasible, provided in written form as well. Additional information is available to patients and family members in books and pamphlets obtainable from libraries, bookstores, and patient support groups, as well as from audiotapes, videotapes, and the Internet.[2,4] Visual aids, such as charts portraying the natural course of the treated and untreated illness, or the causes and results of sleep deprivation and medication cessation, are also helpful to many.[5–7] Finally, providing the patients with self-report scales to monitor their daily progress, such as mood charts in affective disorder, not only provides invaluable clinical data, but also teaches patients to better understand their own illness and its response to therapeutic interventions as well as exacerbating stressors. Patients, when they are well, often benefit from a meeting with their family members and their doctor, which focuses upon drawing up contingency plans in case their illness should recur. These meetings also provide an opportunity to shore up the support system the patient has by educating their caregivers about the nature, cause, manifestations, and treatment of their loved one's mental illness. Such meetings may also include what is to be done in the event that hospitalization is required and the patient refuses voluntary admission, a discussion of early warning signs of impending psychotic or depressive episodes, methods for regularizing sleep and activity patterns, techniques to protect patients financially, and ways to manage suicidal behaviour should it occur. Suicide is the major cause of premature death in the severe psychiatric illnesses,[8,9] and its prevention is of first concern. Those illnesses most likely to result in suicide (the mood disorders, comorbid alcohol and drug abuse, and schizophrenia) need to be treated early, aggressively, and often for an indefinite period of time.[2,10] The increasing evidence that treatment early in psychiatric illness may improve the long-term course needs to be considered in light of the reluctance of many patients to stay in treatment.[10,11]

No one who has treated or suffered from mental illness would minimize the difficulties involved in successful treatment. Modern medicine gives options that did not exist even 10 years ago, and there is every reason to expect that improvements in psychopharmacology, psychotherapy, and diagnostic techniques will continue to develop at a galloping pace. Still, the relationship between the patient and doctor will remain central to the treatment, as Morag Coate wrote 35 years ago in *Beyond All Reason*:[12]

> Because the doctors cared, and because one of them still believed in me when I believed in nothing, I have survived to tell the tale. It is not only the doctors who perform hazardous operations or give life-saving drugs in obvious emergencies who hold the scales at times between life and death. To sit quietly in a consulting room and talk to someone would not appear to the general public as a heroic or dramatic thing to do. In medicine there are many different ways of saving lives. This is one of them.

Further information

Non-Governmental Mental Health Websites: US

http://www.nami.org/
http://www.dbsalliance.org/site/PageServer?pagename=home

Governmental Mental Health Websites: US

http://www.nimh.nih.gov/
http://www.hhs.gov/samhsa/mentalhealth/

Non-Governmental Mental Health Websites: UK

http://www.mentalhealth.org.uk/
http://www.depressionalliance.org/index.html

Governmental Mental Health Websites: US

http://www.dh.gov.uk/en/Healthcare/NationalServiceFrameworks/Mentalhealth/index.html
http://www.mentalhealthcare.org.uk/

References

1. Jamison, K.R., Gerner, R.H., and Goodwin, F.K. (1979). Patient and physician attitudes toward lithium: relationship to compliance. *Archives of General Psychiatry*, **36**, 866–9.
2. Goodwin, F.K. and Jamison, K.R. (2007). *Manic-depressive illness* (2nd edn.). Oxford University Press, New York.
3. Osterberg, L. and Blaschke, T. (2005). Adherence to medication. *The New England Journal of Medicine*, **353**, 487–97.
4. Wyatt, R.J. and Chew, R.H. (2005). *Practical psychiatric practice. Forms and protocols for clinical use* (3rd edn). American Psychiatric Association, Washington, DC.
5. Post, R.M., Rubinow, D.R., and Ballenger, J.C. (1986). Conditioning and sensitisation in the longitudinal course of affective illness. *The British Journal of Psychiatry*, **149**, 191–201.
6. Wehr, T.A., Sack, D.A., and Rosenthal, N.E. (1987). Sleep reduction as a final common pathway in the genesis of mania. *The American Journal of Psychiatry*, **144**, 201–4.
7. Baldessarini, R.J., Tondo, L., and Hennen, J. (2003). Lithium treatment and suicide risk in major affective disorders: update and new findings. *The Journal of Clinical Psychiatry*, **64**(Suppl. 5), 44–52.
8. Harris, E.C. and Barraclough, B. (1997). Suicide as an outcome for mental disorders. A meta-analysis. *The British Journal of Psychiatry*, **170**, 205–28.
9. Institute of Medicine (IoM). (2002). *Reducing suicide: a national imperative*. National Academy Press, Washington, DC.
10. Wyatt, R.J. (1995). Early intervention for schizophrenia: can the course of the illness be altered? *Biological Psychiatry*, **38**, 1–3.
11. Berger, G., Dell'Olio, M., Amminger, P., *et al.* (2007). Neuroprotection in emerging psychotic disorders. *Early Intervention in Psychiatry*, **1**, 114–27.
12. Coate, M. (1964). *Beyond all reason*. Constable, London.

1.2

Public attitudes and the challenge of stigma

Graham Thornicroft, Elaine Brohan, and Aliya Kassam

Introduction

The starting point for this discussion is the idea of stigma. This term (plural stigmata) was originally used to refer to an indelible dot left on the skin after stinging with a sharp instrument, sometimes used to identify vagabonds or slaves.[1–4] In modern times stigma has come to mean 'any attribute, trait or disorder that marks an individual as being unacceptably different from the 'normal' people with whom he or she routinely interacts, and that elicits some form of community sanction.'[5–7]

Understanding stigma

There is now a voluminous literature on stigma.[5,8] [9–13,13–19] The most complete model of the component processes of stigmatization has four key components:[20]

i) Labelling, in which personal characteristics, which are signalled or noticed as conveying an important difference.

ii) Stereotyping, which is the linkage of these differences to undesirable characteristics.

iii) Separating, the categorical distinction between the mainstream/normal group and the labelled group as in some respects fundamentally different.

iv) Status loss and discrimination: devaluing, rejecting, and excluding the labelled group. Interestingly, more recently the authors of this model have added a revision to include the emotional reactions which may accompany each of these stages.[21,22]

Shortcomings of work on stigma

Five key features have limited the usefulness of stigma theories. First, while these processes are undoubtedly complex, academic writings on stigma (which in the field of mental health have almost entirely focused upon schizophrenia) have made relatively few connections with legislation concerning disability rights policy[23] or clinical practice. Second, most work on mental illness and stigma has been descriptive, overwhelmingly describing attitude surveys or the portrayal of mental illness by the media. Very little is known about effective interventions to reduce stigma. Third, there have been notably few direct contributions to this literature by service users.[24] Fourth, there has been an underlying pessimism that stigma is deeply historically rooted and difficult to change. This has been one of the reasons for the reluctance to use the results of research in designing and implementing action plans. Fifth, stigma theories have de-emphasized cultural factors and paid little attention to the issues related to human rights and social structures.

Recently there have been early signs of a developing focus upon discrimination. This can be seen as the behavioural consequences of stigma, which act to the disadvantage of people who are stigmatized.[23,25–27] The importance of discriminatory behaviour has been clear for many years in terms of the personal experiences of service users, in terms of devastating effects upon personal relationships, parenting and childcare, education, training, work, and housing.[28] Indeed, these voices have said that the rejecting behaviour of others may bring greater disadvantage than the primary condition itself.

Stigma can therefore be seen as an overarching term that contains three important elements: [29]

- problems of knowledge ignorance
- problems of attitudes prejudice
- problems of behaviour discrimination

Ignorance: the problem of knowledge

At a time when there is an unprecedented volume of information in the public domain, the level of accurate knowledge about mental illnesses (sometime called 'mental health literacy') is meagre.[30] In a population survey in England, for example, most people (55 per cent) believe that the statement 'someone who cannot be held responsible for his or her own actions' describes a person who is mentally ill.[31] Most (63 per cent) thought that fewer than 10 per cent of the population would experience a mental illness at some time in their lives.

There is evidence that deliberate interventions to improve public knowledge about depression can be successful, and can reduce the effects of stigmatization. At the national level, social marketing

campaigns have produced positive changes in public attitudes towards people with mental illness, as shown recently in New Zealand and Scotland.[32,33] In a campaign in Australia to increase knowledge about depression and its treatment, some states and territories received this intensive, co-ordinated programme, while others did not. In the former, people more often recognized the features of depression, were more likely to support help seeking for depression, or to accept treatment with counselling and medication.[34]

Prejudice: the problem of negative attitudes

Although the term prejudice is used to refer to many social groups, which experience disadvantage, for example minority ethnic groups, it is employed rarely in relation to people with mental illness. The reactions of a host majority to act with prejudice in rejecting a minority group usually involve not just negative thoughts but also emotion such as anxiety, anger, resentment, hostility, distaste, or disgust. In fact prejudice may more strongly predict discrimination than do stereotypes. Interestingly, there is almost nothing published about emotional reactions to people with mental illness apart from that which describes a fear of violence.[35]

Discrimination: the problem of rejecting and avoidant behaviour

Surveys of attitude and social distance (unwillingness to have social contact) usually ask either students or members of the general public what they would do in imaginary situations or what they think 'most people' who do, for example, when faced with a neighbour or work colleague with mental illness. Important lessons have flowed from these findings. This work has emphasized what 'normal' people say without exploring the actual experiences of people with mental illness themselves about the behaviour of normal people towards them. Further it has been assumed that such statements (usually on knowledge, attitudes, or behavioural intentions) are congruent with actual behaviour, without assessing such behaviour directly. Such research has usually focussed on hypothetical rather than real situations, neglecting emotions, and the social context, thus producing very little guidance about interventions that could reduce social rejection. In short, most work on stigma has been beside the point.

Global patterns

Do we know if discrimination varies between countries and cultures? The evidence here is stronger, but still frustratingly patchy.[36] Although studies on stigma and mental illness have been carried out in many countries, few have been comparison of two or more places, or have included non-Western nations.[37]

In Africa one study described attitudes to mentally ill people in rural sites in Ethiopia. Among almost 200 relatives of people with diagnoses of schizophrenia or mood disorders, 75 per cent said that they had experienced stigma due to the presence of mental illness in the family, and a third (37 per cent) wanted to conceal the fact that a relative was ill. Most family members (65 per cent) said that praying was their preferred of treating the condition.[38] Among the general population in Ethiopia schizophrenia was judged to be the most severe problem, and talkativeness, aggression, and strange behaviour were rated as the most common symptoms of mental illness.[39] The authors concluded that it was important to work closely with traditional healers.

In South Africa,[40,41] a survey was conducted of over 600 members of the public on their knowledge and attitudes towards people with mental illness.[42] Different vignettes, portraying depression, schizophrenia, panic disorder, or substance misuse were presented to each person. Most thought that these conditions were either related to stress or to a lack of willpower, rather than seeing them as medical disorders.[43] Similar work in Turkey,[44] and in Siberia and Mongolia[45] suggests that people in such countries may be more ready to make the individual responsible for his or her mental illness and less willing to grant the benefits of the sick role.

Most of the published work on stigma is by authors in the USA and Canada,[11,27,46,47] but there are also a few reports from elsewhere in the Americas and in the Caribbean.[48] In a review of studies from Argentina, Brazil, Dominica, Mexico, and Nicaragua, mainly from urban sites, a number of common themes emerged. The conditions most often rated as 'mental illnesses' were the psychotic disorders, especially schizophrenia. People with higher levels of education tended to have more favourable attitudes to people with mental illness. Alcoholism was considered to be the most common type of mental disorder. Most people thought that a health professional needs to be consulted by people with mental illnesses.[49]

A great deal of work has studied the question of stigma towards mentally ill people in Asian countries and cultures.[50–52] Within China,[53] a large scale survey was undertaken of over 600 people with a diagnosis of schizophrenia and over 900 family members.[54] Over half of the family members said that stigma had an important effect on them and their family, and levels of stigma were higher in urban areas and for people who were more highly educated.

In the field of stigma research we find that schizophrenia is the primary focus of interest. It is remarkable that there are almost no studies, for example, on bipolar disorder and stigma. A comparison of attitudes to schizophrenia was undertaken in England and Hong Kong. As predicted, the Chinese respondents expressed more negative attitudes and beliefs about schizophrenia, and preferred a more social model to explain its causation. In both countries most participants, whatever their educational level, showed great ignorance about this condition.[55] This may be why most of population in Hong Kong are very concerned about their mental health and hold rather negative views about mentally ill people.[56] Less favourable attitudes were common in those with less direct personal contact with people with mental illness (as in most Western studies), and by women (the opposite of what has been found in many Western reports).[57]

Little research on stigma has been conducted in India. Among relatives of people with schizophrenia in Chennai (Madras) in Southern India, their main concerns were: effects on marital prospects, fear of rejection by neighbours, and the need to hide the condition from others. Higher levels of stigma were reported by women and by younger people with the condition. [58] Women who have mental illness appear to be at a particular disadvantage in India. If they are divorced, sometimes related to concerns about heredity,[59] then they often receive no financial support from their former husbands, and they and their families experience intense distress from the additional stigma of being separated or divorced.[60]

In Japan mental illnesses are seen to reflect a loss of control, and so are not subject to the force of will power, both of which lead to a sense of shame.[61–63] Although, it is tempting to generalize about the degree of stigma in different countries, reality may not allow such simplifications. A comparison of attitudes to mentally ill people in Japan and Bali, for example showed that views towards people with schizophrenia were less favourable in Japan, but that people with depression and obsessive-compulsive disorder were seen to be less acceptable in Bali.[64]

What different countries do often share is a high level of ignorance and misinformation about mental illnesses. A survey of teachers' opinions in Japan and Taiwan showed that relatively few could describe the main features of schizophrenia with any accuracy. The general profile of knowledge, beliefs, and attitudes was similar to that found in most Western countries, although the degree of social rejection was somewhat greater in Japan.[65]

In a unique move aimed to reduce social rejection, the name for schizophrenia has been changed in Japan. Following a decade of pressure from family member groups, including Zenkaren, the name for this condition was changed from *seishi buntetsu byo* (split-mind disorder) to *togo shiccho sho* (loss of co-ordination disorder).[66,67] The previous term went against the grain of traditional, culturally-valued concepts of personal autonomy, as a result of which only 20 per cent of people with this condition were told the diagnosis by their doctors.[68–70] There are indications from service users and family members that the new term is seen as less stigmatizing and is more often discussed openly.

Little is written in the English language literature on stigma in Islamic communities, but despite earlier indications that the intensity of stigma may be relatively low,[52] detailed studies indicate that on balance, it is no less than we have seen described elsewhere.[71–74] A study of family members in Morocco found that 76 per cent had no knowledge about the condition, and many considered it chronic (80 per cent), handicapping (48 per cent), incurable (39 per cent), or linked with sorcery (25 per cent). Most said that they had 'hard lives' because of the diagnosis.[75] Turning to religious authority figures is reported to be common in some Moslem countries.[76,77] Some studies have found that direct personal contact was not associated with more favourable attitudes to people with mental illness,[78,79] especially where behaviour is seen to threaten the social fabric of the community.[80,44]

What sense can we make of all these fragments of information? Several points are clear. First there is no known country, society, or culture in which people with mental illness are considered to have the same value and to be as acceptable as people who do not have mental illness. Second, the quality of information that we have is relatively poor, with very few comparative studies between countries or over time. Third, there do seem to be clear links between popular understandings of mental illness, if people in mental distress want to seek help, and whether they feel able to disclose their problems.[81] The core experiences of shame (to oneself and for others) and blame (from others) are common everywhere stigma has been studied, but to differing extents. Where comparisons with other conditions have been made, then mental illnesses are more, or far more, stigmatized,[82,83] and have been referred to as the 'ultimate stigma' [9]. Finally, rejection and avoidance of people with mental illness appear to be universal phenomena.

Conclusions

If we deliberately shift focus from stigma to discrimination, there are a number of distinct advantages. First attention moves from attitudes to actual behaviour, not if an employer *would* hire a person with mental illness, but if he or she *does*. Second, interventions can be tried and tested to see if they change behaviour towards people with mental illness, without *necessarily* changing knowledge or feelings. The key candidates as active ingredients to reduce stigma are: (i) at the local level, direct social contact with people with mental illness;[84–86] and (ii) social marketing techniques at the national level. Third, people who have a diagnosis of mental illness can expect to benefit from all the relevant anti-discrimination policies and laws in their country or jurisdiction, on a basis of parity with people with physical disabilities. Fourth, a discrimination perspective requires us to focus not upon the 'stigmatized' but upon the 'stigmatizer'. In sum, this means sharpening our sights upon human rights, upon injustice, and upon discrimination as actually experienced by people with mental illness.[7,24,87,88]

Further information

Thornicroft, G. (2006). *Shunned: Discrimination against people with mental illness*. Oxford University Press, Oxford.

Hinshaw, S. (2007). *The mark of shame: stigma of mental illness and an agenda for change*. Oxford University Press, Oxford.

Corrigan, P. (2005). *On the stigma of mental illness*. American Psychological Association, Washington, DC.

Sartorius, N. and Schulze, H. (2005). *Reducing the stigma of mental illness*. A report from a global programme of the World Psychiatric Association. Cambridge University Press, Cambridge.

References

1. Cannan, E. (1895). The stigma of pauperism. *Economic Review*, 380–91.
2. Thomas Hobbes of Malmesbury. (1657). *Markes of the absurd geometry, rural language, Scottish church politics, and barbarisms of John Wallis professor of geometry and doctor of divinity*. Printed for Andrew Cooke, London.
3. Gilman, S.L. (1982). *Seeing the insane*. Wiley, New York.
4. Gilman, S.L. (1985). *Difference and pathology: stereotypes of sexuality, race and madness*. Cornell University Press, Ithaca.
5. Goffman, I. (1963). *Stigma: notes on the management of spoiled identity*. Penguin Books, Harmondsworth, Middlesex.
6. Scambler, G. (1998). *Stigma and disease: changing paradigms*. Lancet, **352**, 1054–5.
7. Hinshaw, S.P. and Cicchetti, D. (2000). Stigma and mental disorder: conceptions of illness, public attitudes, personal disclosure, and social policy. *Development and Psychopathology*, **12**(4), 555–98.
8. Mason, T. (2001). *Stigma and social exclusion in healthcare*. Routledge, London.
9. Falk, G. (2001). *Stigma: how we treat outsiders*. Prometheus Books, New York.
10. Heatherton, T.F., Kleck, R.E., Hebl, M.R., *et al.* (2003). *The social psychology of stigma*. Guilford Press, New York.
11. Corrigan, P. (2005). *On the stigma of mental illness*. American Psychological Association, Washington, DC.
12. Wahl, O.F. (1999). *Telling is a risky business: mental health consumers confront stigma*. Rutgers University Press, New Jersey.
13. Pickenhagen, A. and Sartorius, N. (2002). *The WPA global programme to reduce stigma and discrimination because of schizophrenia*. World Psychiatric Association, Geneva.

14. Sartorius, N. and Schulze, H. (2005). *Reducing the stigma of mental illness: a report from a global association*. Cambridge University Press, Cambridge.
15. Hayward, P. and Bright, J.A. (1997). Stigma and mental illness: a review and critique. *Journal of Mental Health*, **6**(4), 345–54.
16. Link, B.G., Cullen, F.T., Struening, E.L., *et al.* (1989). A modified labeling theory approach in the area of mental disorders: an empirical assessment. *American Sociological Review*, **54**, 100–23.
17. Link, B.G., Struening, E.L., Rahav, M., *et al.* (1997). On stigma and its consequences: evidence from a longitudinal study of men with dual diagnoses of mental illness and substance abuse. *Journal of Health and Social Behaviour*, **38**(2), 177–90.
18. Link, B.G., Phelan, J.C., Bresnahan, M., *et al.* (1999). Public conceptions of mental illness: labels, causes, dangerousness, and social distance. *American Journal of Public Health*, **89**(9), 1328–33.
19. Smith, M. (2002). Stigma. *Advances in Psychiatric Treatment* **8,** 317–25.
20. Link, B.G. and Phelan, J.C. (2001). Conceptualizing stigma. *Annual Review of Sociology*, **27,** 363–85.
21. Link, B.G., Yang, L.H., Phelan, J.C., *et al.* (2004). Measuring mental illness stigma. *Schizophrenia Bulletin*, **30**(3), 511–41.
22. Jones, E., Farina, A., Hastorf, A., *et al.*(1984). *Social stigma: the psychology of marked to relationships*. W.H. Freeman & Co., New York.
23. Sayce, L. (2000). *From psychiatric patient to citizen. Overcoming discrimination and social exclusion*. Palgrave, Basingstoke.
24. Chamberlin, J. (2005). User/consumer involvement in mental health service delivery. *Epidemiologia e Psichiatria Sociale*, **14**, 10–14.
25. Corrigan, P.W., Larson, J.E., Watson, A.C., *et al.* (2006). Solutions to discrimination in work and housing identified by people with mental illness. *Journal of Nervous and Mental Disease*, **194**(9), 716–18.
26. Angermeyer, M.C. and Matschinger, H. (2005). Labeling—stereotype—discrimination. An investigation of the stigma process. *Social Psychiatry and Psychiatric Epidemiology*, **40**(5), 391–5.
27. Estroff, S.E., Penn, D.L., and Toporek, J.R. (2004). From stigma to discrimination: an analysis of community efforts to reduce the negative consequences of having a psychiatric disorder and label. *Schizophrenia Bulletin*, **30**(3), 493–509.
28. Thornicroft, G. (2006). *Shunned: discrimination against people with mental illness*. Oxford University Press, Oxford.
29. Thornicroft, G., Rose, D., Kassam, A., *et al.* (2007). Stigma: ignorance, prejudice or discrimination? *British Journal of Psychiatry*, **190,** 192–3.
30. Crisp, A., Gelder, M.G., Goddard, E., *et al.* (2005). Stigmatization of people with mental illnesses: a follow-up study within the changing minds campaign of the Royal College of Psychiatrists. *World Psychiatry*, **4,** 106–13.
31. Department of Health. (2003). *Attitudes to mental illness 2003 report*. Department of Health, London.
32. Vaughn, G. (2004). Like minds, like mine. In *Mental health promotion: case studies from countries* (eds. S. Saxena and P. Garrison), pp. 62–6. World Health Organization, Geneva.
33. Dunion, L. and Gordon L. (2005). Tackling the attitude problem. The achievements to date of Scotland's 'see me' anti-stigma campaign. *Mental Health Today*, **4**, 22–5.
34. Jorm, A.F., Christensen, H., and Griffiths, K.M. (2005). The impact of beyondblue: the national depression initiative on the Australian public's recognition of depression and beliefs about treatments. *Australian and New Zealand Journal of Psychiatry*, **39**(4), 248–54.
35. Graves, R.E., Cassisi, J.E., and Penn, D.L. (2005). Psychophysiological evaluation of stigma towards schizophrenia. *Schizophrenia Research*, **76**(2–3), 317–27.
36. Littlewood, R. (2004). Cultural and national aspects of stigmatisation. In *Every family in the land* (ed. A.H. Crisp), pp. 14–17. Royal Society of Medicine, London.
37. Fabrega, H., Jr. (1991). The culture and history of psychiatric stigma in early modern and modern western societies: a review of recent literature. *Comprehensive Psychiatry*, **32**(2), 97–119.
38. Shibre, T., Negash, A., Kullgren, G. *et al.* (2001). Perception of stigma among family members of individuals with schizophrenia and major affective disorders in rural Ethiopia. *Social Psychiatry and Psychiatric Epidemiology*, **36**(6), 299–303.
39. Alem, A., Jacobsson, L., Araya, M., *et al.* (1999). How are mental disorders seen and where is help sought in a rural Ethiopian community? A key informant study in Butajira, Ethiopia. *Acta Psychiatrica Scandinavica*, **397**(Suppl.), 40–7.
40. Stein, D.J., Wessels, C., Van Kradenberg, J., *et al.* (1997). The mental health information centre of South Africa: a report of the first 500 calls. *Central African Journal of Medicine*, **43**(9), 244–6.
41. Minde, M. (1976). History of mental health services in South Africa. Part XIII. The national council for mental health. *South African Medical Journal*, **50**(3F), 1452–6.
42. Hugo, C.J., Boshoff, D.E., Traut, A., *et al.* (2003). Community attitudes toward and knowledge of mental illness in South Africa. *Social Psychiatry and Psychiatric Epidemiology*, **38**(12), 715–19.
43. Cheetham, W.S. and Cheetham, R.J. (1976). Concepts of mental illness amongst the rural Xhosa people in South Africa. *Australian and New Zealand Journal of Psychiatry*, **10**(1), 39–45.
44. Ozmen, E., Ogel, K., Aker, T., *et al.* (2004). Public attitudes to depression in urban Turkey—the influence of perceptions and causal attributions on social distance towards individuals suffering from depression. *Social Psychiatry and Psychiatric Epidemiology*, **39**(12), 1010–16.
45. Dietrich, S., Beck, M., Bujantugs, B., *et al.* (2004). The relationship between public causal beliefs and social distance toward mentally ill people. *Australian and New Zealand Journal of Psychiatry*, **38**(5), 348–54.
46. Link, B.G., Yang, L.H., Phelan, J.C., *et al.* (2004). Measuring mental illness stigma. *Schizophrenia Bulletin*, **30**(3), 511–41.
47. Corrigan, P., Thompson, V., Lambert, D., *et al.* (2003). Perceptions of discrimination among persons with serious mental illness. *Psychiatric Services*, **54**(8), 1105–10.
48. Villares, C. and Sartorius, N. (2003). Challenging the stigma of schizophrenia. *Revista Brasasileira de Psiquiatria*, **25,** 1–2.
49. de Toledo Piza, P.E. and Blay, S.L. (2004). Community perception of mental disorders—a systematic review of Latin American and Caribbean studies. *Social Psychiatry and Psychiatric Epidemiology*, **39**(12), 955–61.
50. Ng, C.H. (1997). The stigma of mental illness in Asian cultures. *Australian New Zealand Journal of Psychiatry*, **31**(3), 382–90.
51. Leong, F.T. and Lau, A.S. (2001). Barriers to providing effective mental health services to Asian Americans. *Mental Health Services Research*, **3**(4), 201–14.
52. Fabrega, H. Jr. (1991). Psychiatric stigma in non-western societies. *Comprehensive Psychiatry*, **32**(6), 534–51.
53. Kleinman, A. and Mechanic, D. (1979). Some observations of mental illness and its treatment in the people's republic of China. *Journal of Nervous and Mental Disease*, **167**(5), 267–74.
54. Phillips, M.R., Pearson, V., Li, F., *et al.* (2002). Stigma and expressed emotion: a study of people with schizophrenia and their family members in China. *British Journal of Psychiatry*, **181,** 488–93.
55. Furnham, A. and Chan, E. (2004). Lay theories of schizophrenia. A cross-cultural comparison of British and Hong Kong Chinese attitudes, attributions and beliefs. *Social Psychiatry and Psychiatric Epidemiology*, **39**(7), 543–52.
56. Chou, K.L., Mak, K.Y., Chung, P.K., *et al.* (1996). Attitudes towards mental patients in Hong Kong. *International Journal of Social Psychiatry*, **42**(3), 213–9.
57. Chung, K.F., Chen, E.Y., and Liu, C.S. (2001). University students' attitudes towards mental patients and psychiatric treatment. *International Journal of Social Psychiatry*, **47**(2), 63–72.

58. Thara, R. and Srinivasan, T.N. (2000). How stigmatising is schizophrenia in India? *International Journal of Social Psychiatry*, **46**(2), 135–41.
59. Raguram, R., Raghu, T.M., Vounatsou, P., *et al.* (2004). Schizophrenia and the cultural epidemiology of stigma in Bangalore, India. *Journal of Nervous and Mental Disease*, **192**(11), 734–44.
60. Thara, R., Kamath, S., and Kumar, S. (2003). Women with schizophrenia and broken marriages—doubly disadvantaged? Part II: family perspective. *International Journal of Social Psychiatry*, **49**(3), 233–40.
61. Desapriya, E.B. and Nobutada, I. (2002). Stigma of mental illness in Japan. *Lancet*, **359**(9320), 1866.
62. Hasui, C., Sakamoto, S., Suguira, B., *et al.* (2000). Stigmatization of mental illness in Japan: images and frequency of encounters with diagnostic categories of mental illness among medical and non-medical university students. *Journal of Psychiatry & Law*, **28**(Summer), 253–66.
63. Sugiura, T., Sakamoto, S., Kijima, N., *et al.* (2000). Stigmatizing perception of mental illness by Japanese students: comparison of different psychiatric disorders. *Journal of Nervous and Mental Disease*, **188**(4), 239–42.
64. Kurihara, T., Kato, M., Sakamoto, S., *et al.* (2000). Public attitudes towards the mentally ill: a cross-cultural study between Bali and Tokyo. *Psychiatry and Clinical Neuroscience*, **54**(5), 547–52.
65. Kurumatani, T., Ukawa, K., Kawaguchi, Y. *et al.* (2004). Teachers' knowledge, beliefs and attitudes concerning schizophrenia—a cross-cultural approach in Japan and Taiwan. *Social Psychiatry and Psychiatric Epidemiology*, **39**(5), 402–9.
66. Desapriya, E.B. and Nobutada, I. (2002). Stigma of mental illness in Japan. *Lancet*, **359**(9320), 1866.
67. Takizawa, T. (1993). Patients and their families in Japanese mental health. *New Directions for Mental Health Service*, (**60**), 25–34.
68. Goto, M. (2003). Family psychoeducation in Japan. *Seishin Shinkeigaku Zasshi*, **105**(2), 243–7.
69. Kim, Y. and Berrios, G.E. (2001). Impact of the term schizophrenia on the culture of ideograph: the Japanese experience. *Schizophrenia Bulletin*, **27**(2), 181–5.
70. Mino, Y., Yasuda, N., Tsuda, T., *et al.* (2001). Effects of a one-hour educational program on medical students' attitudes to mental illness. *Psychiatry and Clinical Neuroscience*, **55**(5), 501–7.
71. Karim, S., Saeed, K., Rana, M.H., *et al.* (2004). Pakistan mental health country profile. *International Review of Psychiatry*, **16**(1–2), 83–92.
72. Al-Krenawi, A., Graham, J.R., and Kandah, J. (2000). Gendered utilization differences of mental health services in Jordan. *Community Mental Health Journal*, **36**(5), 501–11.
73. Al-Krenawi, A., Graham, J.R., Ophir, M., *et al.* (2001). Ethnic and gender differences in mental health utilization: the case of Muslim Jordanian and Moroccan Jewish Israeli out-patient psychiatric patients. *International Journal of Social Psychiatry*, **47**(3), 42–54.
74. Cinnirella, M. and Loewenthal, K.M. (1999). Religious and ethnic group influences on beliefs about mental illness: a qualitative interview study. *British Journal of Medical Psychology*, **72**(Pt 4), 505–24.
75. Kadri, N., Manoudi, F., Berrada, S., *et al.* (2004). Stigma impact on Moroccan families of patients with schizophrenia. *Canadian Journal of Psychiatry*, **49**(9), 625–9.
76. Al-Krenawi, A., Graham, J.R., Dean, Y.Z., *et al.* (2004). Cross-national study of attitudes towards seeking professional help: Jordan, United Arab Emirates (UAE) and Arabs in Israel. *International Journal of Social Psychiatry*, **50**(2), 102–14.
77. Loewenthal, K.M., Cinnirella, M., Evdoka, G., *et al.* (2001). Faith conquers all? Beliefs about the role of religious factors in coping with depression among different cultural-religious groups in the UK. *British Journal of Medical Psychology*, **74**(Pt 3), 293–303.
78. Arkar, H. and Eker, D. (1992). Influence of having a hospitalized mentally ill member in the family on attitudes toward mental patients in Turkey. *Social Psychiatry and Psychiatric Epidemiology*, **27**(3), 151–5.
79. Arkar, H. and Eker, D. (1994). Effect of psychiatric labels on attitudes toward mental illness in a Turkish sample. *International Journal of Social Psychiatry*, **40**(3), 205–13.
80. Coker, E.M. (2005). Selfhood and social distance: toward a cultural understanding of psychiatric stigma in Egypt. *Social Science & Medicine*, **61**(5), 920–30.
81. Littlewood, R. (1998). Cultural variation in the stigmatisation of mental illness. *Lancet*, **352**(9133), 1056–7.
82. Lai, Y.M., Hong, C.P., and Chee C.Y. (2001). Stigma of mental illness. *Singapore Medical Journal*, **42**(3), 111–14.
83. Lee, S., Lee, M.T., Chiu, M.Y., *et al.* (2005). Experience of social stigma by people with schizophrenia in Hong Kong. *British Journal of Psychiatry*, **186**, 153–7.
84. Link, B.G. and Cullen, F.T. (1986). Contact with the mentally ill and perceptions of how dangerous they are. *Journal of Health and Social Behaviour*, **27**(4), 289–302.
85. Pinfold, V., Toulmin, H., Thornicroft, G., *et al.* (2003). Reducing psychiatric stigma and discrimination: evaluation of educational interventions in UK secondary schools. *British Journal of Psychiatry*, **182**, 342–6.
86. Pinfold, V., Huxley, P., Thornicroft, G., *et al.* (2003). Reducing psychiatric stigma and discrimination—evaluating an educational intervention with the police force in England. *Social Psychiatry and Psychiatric Epidemiology*, **38**(6), 337–44.
87. Rose, D. (2001). *Users' voices, the perspectives of mental health service users on community and hospital care*. The Sainsbury Centre, London.
88. Kingdon, D., Jones, R., and Lonnqvist, J. (2004). Protecting the human rights of people with mental disorder: new recommendations emerging from the council of Europe. *British Journal of Psychiatry,* **185,** 277–9.

1.3

Psychiatry as a worldwide public health problem

Contents

1.3.1 Mental disorders as a worldwide public health issue

Benedetto Saraceno

Magnitude and burden of mental disorders

The twentieth century has witnessed significant improvements in somatic health in most countries. A number of key public health threats have been eradicated or brought under control under the leadership of WHO. Priority was given to communicable diseases in view of their inherent potential to spreading.

At the present time, a focus on non-communicable diseases and mental health would now appear as the next natural step in public health priorities. In the case of mental health, this is due to the capacity of mental disorders to proliferate not only as a result of complex and multiple biological, psychological but also social determinants. WHO estimates that at any given time 450 million people suffer from some form of mental or brain disorder, including alcohol and substance use disorders. In other words, one in four of the world's population suffer from different forms of mental, behavioural, and neurological disorders.[1]

The World Development Report: investing in health[2] and the development of the disability-adjusted life-year for estimating the global burden of disease, including years lost because of disability[3,4] and the World Health Report 2001, have all raised the awareness of the global burden of mental disorders. Mental disorders already account for more than 13.46 per cent of the GBD. Furthermore, it is estimated that by the year 2015, the GBD from all neuropsychiatric illnesses will reach 14.14 per cent and by 2030, 14.42 per cent. According to WHO, mental disorders accounted for 6 of the 20 leading causes of disability worldwide for the 15–44 age group, the most productive section of the population.[1] While a greater proportion of the burden is found in high-income countries (21.4 per cent) including those with formerly socialist economies (16.4 per cent), low- and middle-income countries are greatly affected and are likely to see a disproportionately large increase in the burden attributable to mental disorders in the coming decades as infectious diseases are brought under better control and as the population ages. The growing burden of mental, neurological, and substance use disorders is exacerbated in low and middle-income countries due to a projected increase in the number of young people entering the age of risk for the onset of certain mental disorders. An estimated 849 000 people commit suicide every year. This figure represents 1.4 per cent of the global burden of disease as estimated using Disability Adjusted Life Years (DALY) methodology. The proportion of the global disease burden due to suicide varies from 0.2 per cent in Africa up to 2.5 per cent in the Western Pacific region. In the European, South East Asian, and Western Pacific regions, this proportion exceeds the world average. Suicide among young people is of significant concern; in some regions, suicide is the third leading cause of death in the age group of 15–35 years. Suicide is the leading cause of death for this age group in China and the second in the European region. Alcohol consumption alone is responsible for 4 per cent of the global disease burden.[5] In 2000, the global use of alcohol was estimated to have caused 1.8 million deaths or 3.2 per cent of the total deaths from all causes that year. It is estimated that 2.2 million people died from alcohol-related causes in 2005 and increase of 22 per cent from 2000. The population of injecting drug users comprises approximately 10 million people worldwide. Globally, 4–12 per cent of all HIV cases are due to injection drug use, a driving force behind the HIV/AIDS epidemic in many parts of the world.

Economic and social costs of mental disorders

The economic and social costs of mental disorders fall on societies, governments, people with mental disorders, and their carers and families. Given the long-term nature of mental disorders, the most

evident economic burden is that of direct treatment costs. For example, the most important contributor to direct costs of depression is hospitalization, accounting for around half of the total in the United Kingdom and three-quarters in the United States.[6] However a common finding from studies of the economic burden of mental disorders in high-income countries is that the 'indirect' costs of lost productivity and premature mortality outweigh the 'direct' costs of treatment and care.[7] Three recent mental health economic studies carried out in India have likewise shown that lost production and other time costs greatly exceed the costs of targeted clinical intervention.[8–10]

In most countries, families bear a significant proportion of these economic costs because of the absence of publicly funded comprehensive mental health service networks. However, ultimately governments and societies pay a price in terms of reduced national income and increased expenditure on social welfare programmes. Thus, the economic logic for societies and countries is simple: treating mental disorders is expensive but leaving them untreated can be more expensive.

In addition to the obvious suffering caused by mental disorders there is a hidden burden of stigma and discrimination and human rights violations. Rejection, unfair denial of employment opportunities and discrimination in access to services, health insurance, and housing are common as are violations of basic human rights and freedoms, as well as denials of civil, political, economic, and social rights, in both institutions and communities. Much of this goes unreported and therefore the burden remains unquantified. Families and primary care providers also incur social costs, such as the emotional burden of looking after disabled family members, diminished quality of life, social exclusion, stigmatization, and loss of future opportunities for self-improvement.

Global resources for mental health

The WHO survey of mental health resources (Project Atlas) highlighted the huge existent gap between the burden of mental disorders and available resources.[11,12]

Mental health policy and legislation

Mental health services and strategies must be well coordinated with other services, such as social security, education, and public interventions in employment and housing through an adequate mental health policy. In spite of this, only 62 per cent of countries have a policy in the mental health field (see Table 1.3.1.1). Mental health legislation is essential to guarantee the dignity of patients and protect their fundamental human rights, though 22 per cent of countries do not have legislation in the field of mental health.

Table 1.3.1.1 Policy and legislation on mental health in WHO regions and the world—countries (%)

WHO regions	Policy (N: 190)	Legislation (N: 173)
Africa	50%	80%
Americas	73%	75%
Eastern Mediterranean	73%	57%
Europe	71%	92%
South-East Asia	55%	64%
Western Pacific	48%	76%
World	62%	78%

Source: The World health report 2001–Mental Health: New Understanding, New Hope, © 2001, World Health Organization, www.who.int

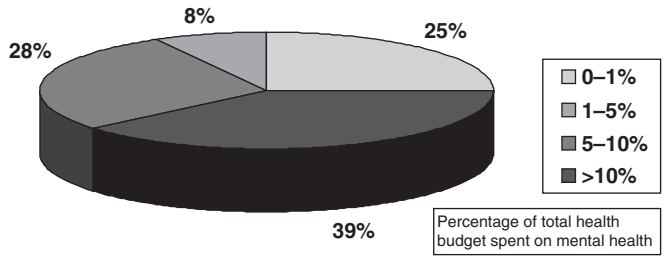


Fig. 1.3.1.1 Percentage of total health budget spent on mental health—countries (%) (*N* = 101). Taken from The World health report 2001–Mental Health: New Understanding, New Hope © 2001, World Health Organization, www.who.int

Mental health budget

In spite of the importance of mental health burden in the world (representing more than 13.46 per cent of global burden of diseases), out of only 101 countries that reported having a specific budget, 25 per cent spend less than 1 per cent of the total health budget on mental health (Fig. 1.3.1.1).

Methods of financing mental health care

The tax-based method is the preferred method for financing mental health care present in 63 per cent of countries (Fig. 1.3.1.2), while all the countries with out-of-pocket financing as the primary method are low- or middle-income countries. However, families of people with severe chronic mental disorders are often among the poorer and, in addition to the family burden, can access to basic mental health care.

Community care for mental health

Community care has a better effect than institutional treatment on the outcome and quality of life of individuals with chronic mental disorders. Globally, 68 per cent of countries reported to have at least some community care facilities for mental health. Community care facilities in mental health are only present in 52 per cent of the low-income countries versus 97 per cent of high-income countries.

Though community-based services are recognized to be most effective, 65 per cent of all psychiatric beds are still in mental hospitals—eating away the already meagre budgets while providing largely custodial care in an environment that violates basic human rights of inmates.[10]

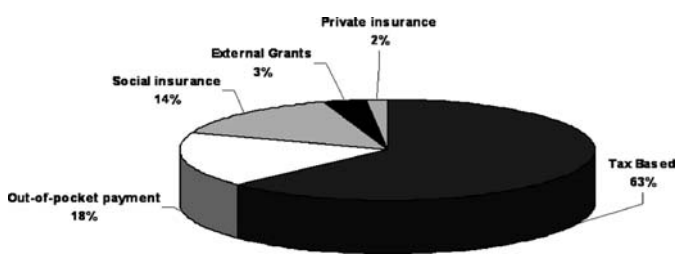


Fig. 1.3.1.2 Methods of financing mental health care in the world—countries (%) (*N* = 180). Taken from The World health report 2001–Mental Health: New Understanding, New Hope, © 2001, World Health Organization, www.who.int

Psychiatric beds

The distribution of psychiatric beds by setting across different income countries also varies. In low-income countries 74 per cent of the psychiatric beds are located in mental hospitals, while in high-income countries only 55 per cent. Across different regions, south-east Asia has 83 per cent of its psychiatric beds in mental hospitals compared with 64 per cent in the European region (see Table 1.3.1.2). The Western Pacific region has the highest proportion of psychiatric beds in general hospitals (35 per cent), followed by Europe with 22 per cent of their total psychiatric beds. In approximately 41 per cent of countries there is less than one psychiatric bed per 10 000 of the population. The proportion of beds which are not located in mental hospitals or in general hospitals includes those in private and military hospitals, hospitals for special groups of population or long-term rehabilitation centres.

Professionals working in mental health

All the countries in the south-east Asia region and most of countries in the African region have less than one psychiatrist per 100 000 population compared to 10 psychiatrist per 100 000 populations in the European region (see Table 1.3.1.3).

The median number of psychiatric nurses per 100 000 population varies from 0.10 in the south-east Asia region to 25 in the European region.

The median number of psychologists in mental health per 100 000 population varies from 0.03 in the south-east Asia and Western Pacific region to 3.10 in the European region and 2.80 in the American region.

In the world there is less than one psychologist per 100 000 population in 61.6 per cent of countries and in low-income countries almost all the population has access to less than one psychologist per 100 000.

The median number of social workers working in mental health per 100 000 population varies from 0.04 in the south-east Asian region to 1.50 in the European region. In about 64 per cent of countries there is less than one social worker per 100 000 population. In the African and Eastern Mediterranean regions more than 85 per cent of the population has access to less than one social worker per 100 000 population.

Table 1.3.1.2 Psychiatric beds per 10 000 population and proportion of psychiatric beds in mental hospitals in WHO regions and the world (*N* = 185)

WHO regions	Median per 10,000 population	Mental hospitals (%)
Africa	0.34	73.0
Americas	2.60	80.6
Eastern Mediterranean	1.07	83.0
Europe	8.00	63.5
South-East Asia	0.33	82.7
Western Pacific	1.06	60.1
World	1.69	68.6

Source: Taken from The World health report 2001–Mental Health: New Understanding, New Hope, © 2001, World Health Organization, www.who.int

Treatment gap for mental disorders

A large proportion of the individuals who suffer from mental disorders do not receive any health care for their condition. The treatment gap for most mental disorders is high. According to a recent review done by WHO from published sources, originating from the United States, Europe, Brazil, Chile, China, India, Zimbabwe, and others,[13] the percentages of people in need for treatment not receiving it are as follows (see Table 1.3.1.4).

Improving mental health care

The mental health infrastructure and services in most countries is grossly insufficient for the large and growing needs. In order to deliver a high standard of mental health treatment and care, WHO emphasizes the adoption of an integrated system of service delivery which attempts to comprehensively address the full range of psychosocial needs of people with mental disorders. A number of policy recommendations for service organizations have been highlighted in the World Health Report 2001.[1] They include (i) shifting care away from large psychiatric hospitals, (ii) developing community mental health services, and (iii) integrating mental health care into general health services.

Table 1.3.1.3 Median number of psychiatrists, psychiatric nurses, psychologists and social workers working in mental health per 100 000 population in WHO regions and the world

WHO regions	Psychiatrists (N=176)	Psychiatric nurses (N=187)	Psychologists in mental health (N=177)	Social workers in mental health (N=161)
Africa	0.04	0.2	0.05	0.05
Americas	2	2.6	2.80	1.00
Eastern Mediterranean	0.95	1.25	0.60	0.40
Europe	9.8	24.8	3.10	1.50
South-East Asia	0.2	0.1	0.03	0.04
Western Pacific	0.32	0.5	0.03	0.05
World	1.2	2	0.60	0.40

Source: Taken from The World health report 2001–Mental Health: New Understanding, New Hope, © 2001, World Health Organization, www.who.int

Table 1.3.1.4 Treatment gap for some mental and substance use disorders

Schizophrenia	32.2%
Depression	56.3%
Bipolar disorder	50.2%
Panic disorder	55.9%
Obsessive compulsive disorder	57.3%
Alcohol abuse and dependence	78.1%

Source: Taken from The World health report 2001–Mental Health: New Understanding, New Hope, © 2001, World Health Organization, www.who.int

Essentially, ethical and scientific considerations have given impetus to the movement to transfer mental health care from mental hospitals to primary health care, general hospitals, and a range of community services in the expectation of enhancing accessibility and acceptability of services, achieving better 'mental' and 'physical' health outcomes, and also a better rationalization of resources.

A large part of mental health care can be self-managed and/or managed by informal community mental health services and low-cost resources can be made available in the community to this effect. Where additional expertise and support is needed a more formalized network of services is required. In ascending order these include primary care services, followed by psychiatric services based in general hospitals and formal community mental health services and lastly by specialist and long stay mental health services.

The mental health field is developing rapidly. There is an evolving information base to guide policy, legislation, service development, and clinical practice. However, there remains a gap between what we know in terms of what works and what is actually occurring in practice in countries around the world. This gap needs to be closed by continued advocacy efforts to raise mental health on the agenda of governments, by continued dissemination of information on effective policies, service development and clinical practice, and the dissemination of international human right standards.

Acknowledgements

M. Garrido Cumbrera has contributed to this publication providing significant comments and technical assistance during the preparation of this manuscript.

Disclaimer

B. Saraceno is a staff member of the World Health Organization. The author alone is responsible for the views expressed in this publication and they do not necessarily represent the decisions, policy or views of the World Health Organization.

Further information

WHO Mental Health website: http://www.who.int/mental_health/en/

References

1. World Health Organization (2001). *The World Health Report 2001: mental health: new understanding, new hope*. World Health Organization, Geneva.
2. World Bank (1993). *World Development Report: investing in health*. Oxford University Press, New York.
3. Murray, C.J.L. and Lopez, A.D. (1996). *The global burden of diseases: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020*. Harvard School of Public Health, World Health Organization and World Bank, Boston.
4. Murray, C.J.L. and Lopez, A.D. (2000). Progress and directions in refining the global burden of disease approach: a response to Williams. *Health Economics*, **9**, 69–82.
5. World Health Organization (2004). *Global status report on alcohol*. World Health Organization, Geneva.
6. Berto, P., D'Ilario, D., Ruffo, P., *et al.* (2000). Depression: cost-of-illness studies in the international literature, a review. *The Journal of Mental Health Policy and Economics*, **3**, 3–10.
7. Hu, T-W. (2004). An International Review of the Economic Costs of Mental Illness. Disease Control Priorities Project. Working Paper No. 31. University of California, Berkeley. Available at: http://www.dcp2.org/file/45/wp31.pdf
8. Chisholm, D., Sekar, K., Kumar, K.K., *et al.* (2000). Integration of mental health care into primary care. Demonstration cost-outcome study in India and Pakistan. *The British Journal of Psychiatry*, **176**, 581–8.
9. Patel, V., Chisholm, D., Rabe-Hesketh, S., *et al.* (2003). Efficacy and cost-effectiveness of drug and psychological treatments for common mental disorders in general health care in Goa, India: a randomised, controlled trial. *Lancet*, **361**(9351), 33–9.
10. Srinivasa Murthy, R., Kishore Kumar, K.V., Chisholm, D., *et al.* (2005). Community outreach for untreated schizophrenia in rural India: a follow-up study of symptoms, disability, family burden and costs. *Psychological Medicine*, **35**(3), 341–51.
11. World Health Organization (2005). *Mental health atlas 2005*. World Health Organization, Geneva.
12. Saxena, S., Sharan, P., Garrido Cumbrera, M., *et al.* (2006). World Health Organization's mental health atlas 2005: implications for policy development. *World Psychiatry*, **5**(3): 179–84.
13. Kohn, R., Saxena, S., Levav, I., *et al.* (2004). The treatment gap in mental health care. *Bulletin of the World Health Organization*, **82**(11): 856–66.

1.3.2 Transcultural psychiatry

Julian Leff

Clinical relevance of transcultural psychiatry

With the mass movements of populations that have characterized the second half of the twentieth century, there can be few psychiatrists who do not encounter members of an ethnic minority group in their practice. The principles of transcultural psychiatry are obviously of relevance to this type of psychiatrist–patient interaction, but they are also of central importance even when the psychiatrist and patient share the same ethnic background. This is because within a particular ethnic group there are invariably many subcultures, for example based on religious affiliation, which encompass a diversity of beliefs. It is essential that the psychiatrist be aware of the common belief systems likely to be encountered, not simply to enhance rapport with patients and relatives, but in order to avoid serious mistakes in ascribing pathology to experiences that are accepted as normal by the subculture. For example, it is important to be aware that between 10 and 17.5 per cent of the

general population report experiencing psychotic symptoms.[1,2] The political repercussions of ignorance of such subcultural phenomena are illustrated by the accusations of misdiagnosis of Black patients by White psychiatrists which have come from both outside and within the profession. It is somewhat reassuring that the only published scientific study of this contention fails to support it.[3]

There are two main streams of thought and enquiry that have shaped the development of transcultural psychiatry: social anthropology and psychiatric epidemiology. In a number of ways these disciplines are opposed; the former is concerned with qualitative data and emphasizes cultural relativity (see Chapter 2.6.2.), while the latter relies on quantitative data and prioritizes a search for universal disease categories (see Chapter 2.7). The tools of the epidemiologist are standardized interview schedules which are linked with definitions of symptoms and signs, and rules for reaching a diagnosis. These have been introduced in an attempt to reduce the subjectivity of the psychiatrist's judgement to a minimum. By contrast, it is the person's subjective experience of illness that is the prime focus of the anthropologist. Consequently the use of standardized psychiatric interviews has been criticized by anthropologists as imposing a western biomedical model of disease on the rich variety of experience of illness and distress. The two approaches are not mutually exclusive and are best viewed as contributing complementary material to our understanding of psychiatric morbidity.[4]

The contribution of psychiatric epidemiology

Cultural influences on the psychoses

Epidemiologists have been keen to discover whether psychiatric conditions are universal and appear with the same incidence across human populations. Universality would minimize the role of culture in shaping the form of a condition, while a uniform incidence would indicate that biological factors played a major role in aetiology. Schizophrenia has been the focus of many epidemiological surveys, especially the cross-national studies conducted by the World Health Organization (WHO). The International Pilot Study of Schizophrenia[5] showed that it was possible to conduct a psychiatric epidemiological study across a wide variety of cultures and languages.[6] The use of standardized assessment and diagnostic techniques revealed that the core symptoms of schizophrenia were subject to few cultural variations. The most striking difference in the form of the illness was that catatonic symptoms were relatively frequent in patients from developing countries, but rare in the other centres.

The success of this study led to an even more ambitious project—the Determinants of the Outcome of Severe Mental Disorders. The main aim was to collect epidemiologically based samples of psychotic patients making a first contact with health services in centres around the world. It was found that the incidence of narrowly defined schizophrenia was remarkably uniform across a diversity of countries.[7] However, when patients with a broad diagnosis of schizophrenia but lacking the core Schneiderian symptoms were considered, the incidence rates across centres showed a threefold difference which was highly significant. This suggests that socio-cultural factors are likely to play a much greater role in the aetiology of non-Schneiderian schizophrenia than in the narrowly defined form, although the nature of these factors remains to be determined.

Dramatic differences in outcome at a 2-year follow-up were found, patients with schizophrenia in developing centres faring considerably better than those in developed centres despite a paucity of psychiatric personnel and facilities. This was not explained by a higher proportion of cases with an acute onset in the developing centres, raising intriguing questions about the beneficial aspects of traditional cultures. Explanations that have been proposed include beliefs that the causes of illness are external to the patient, the low demands for productivity and punctuality in an agrarian economy enabling the employment of disabled patients, and the quality of traditional family life. Only the latter has been investigated and appears to make an important contribution, since family carers in India are far less critical and more tolerant of patients with schizophrenia than their counterparts in Britain.[8]

The existence of relatively large populations of people of ethnic minority status in developed countries has facilitated the study of cultural influences on psychoses. Such research has revealed a remarkably elevated incidence of both schizophrenia and mania in some of these groups.[9,10] Of a number of possible explanations, the most likely lie in the social environment.[11]

Mania has been the focus of much less transcultural research than schizophrenia, but what little there is suggests that psychotic experiences are more common in Nigerian and African–Caribbean patients than in patients from European countries.[12,13]

Cultural influences on the neuroses

(a) Variations in frequency across cultures

Whereas neither the form nor the incidence of psychoses vary much across cultures, neuroses show dramatic variations in both respects. So-called culture-bound syndromes are an extreme example of variation in frequency since it is claimed that they are confined to specific cultural groups (see Chapter 4.16.). It is an error to think of these as exotic manifestations in traditional societies, since eating disorders, while increasingly common in developed countries are infrequent elsewhere.[14] Even with common conditions such as depression, the range of prevalence rates from studies across cultures is extremely wide.[15] This is partly attributable to a greater focus on bodily symptoms in patients in developing countries. The significance of somatic symptoms may well be missed by standardized interviews designed to detect the cognitive experiences of depression.

The emphasis on the measurement of prevalence of neuroses as opposed to incidence is due to the small proportion of new cases of neurosis that present to psychiatric services. In order to detect the majority of new cases of neuroses it is necessary to conduct population surveys, which are costly in terms of time and trained personnel. The few population surveys that have been conducted in both developed and developing countries using the same methods of interviewing and case ascertainment have shown either no difference in the prevalence of neuroses[16,17] or a higher rate in the developing country.[18,19]

(b) Variations in form across cultures

One of the most striking transcultural aspects of the neuroses is the great variation in the frequency of classical conversion hysteria. Whereas this condition is rarely seen in psychiatric and neurological services in developed countries today, it is still a common form

of presentation in developing countries.[20] This is another manifestation of the tendency to present emotional distress in bodily terms that prevails in those cultures. Somatization is by no means uncommon in patients in developed countries, particularly in individuals of lower socio-economic status, but somatic symptoms are more likely to dominate the picture in patients in a developing country. This is determined partly by beliefs about illness (see Chapter 2.6.2.) and partly by mutual expectations of patients and doctors and of traditional healers, who treat the majority of people with neuroses in developing countries.

Contributions of anthropology

Help-seeking behaviour

In general people seek help from healers who hold the same beliefs as they do (see Chapter 7.3). Traditional healers in developing countries have the advantage of sharing the same belief system about illness with their clients, so that they can take for granted a great deal of common ground and do not need to embark on long explanations. Clients of traditional healers often present their distress in terms of somatic symptoms. Skilled healers are adept at understanding the relationship problems that underlie the client's bodily complaints, and their prescription of rituals is aimed at involving the client's social network and regularizing relationships.[21] Problems in communication arise when the patient brings somatic symptoms to the western trained doctor, who may fail to detect the emotional distress generating the symptoms[22] and is unable to recognize the relationship difficulties that have prompted the comlaints.

Traditional healers are by no means confined to developing countries or to ethnic minority groups in developed countries. Alternative medicine flourishes where western biomedicine is perceived by the public to be ineffective, and psychiatry is one of those areas. Patients with psychiatric conditions are very likely to seek help from acupuncture, spiritual healing, homeopathy, or herbal remedies, in addition to consulting the general practitioner or psychiatrist. Sympathetic questioning of psychiatric patients will elicit their use of a number of sources of alternative medicine in their neighbourhood.

The concept of depression

At the same time as the evidence for a biological basis for depression appears to be strengthening, the western concept of depression has been criticized by transcultural researchers. Obeyesekere[23] considers that each culture has developed its own methods for dealing with painful emotions, for example, the Buddhists of Sri Lanka cope with the loss of a loved person by meditating on the illusory nature of the world of sense, pleasure, and domesticity. Obeyesekere[23] refers to these coping measures as 'the work of culture' and views the construction of a disease known as depression as a western cultural resource. Its incorporation into international classifications of diseases could be viewed as 'the imposition of western cultural standards that are presented as universal and inseparable parts of an emerging new world order'.[24] If a biological basis for the neuroses was firmly established such a formulation could be readily dismissed, but the efficacy of non-biological treatments for depression and anxiety, such as cognitive therapy, couple therapy, and behaviour therapy, indicates that Obeyesekere's view deserves serious consideration. It represents a specific example of the general premise that western biomedicine is itself a cultural construction and needs to be seen as one of many different ways of dealing with the experience of illness and distress.[24] The achievement of biomedicine in ridding the world of smallpox and other fatal diseases is undeniable, but in the field of psychiatry in particular we need to remain open to the ways other cultures have developed for helping people with what we would term psychiatric illness.

Further information

Kleinman, A. (1988). *Patients and healers in the context of culture: an exploration of the borderland between anthropology, medicine and psychiatry*. University of California Press, Berkeley.

Swartz, L. (2000). *Culture and mental health: a Southern African view*. Oxford University Press, Oxford.

References

1. Van Os, J., Hanssen, M., Bijl, R.V., *et al.* (2001). Prevalence of psychotic disorder and community level of psychotic symptoms: an urban-rural comparison. *Archives of General Psychiatry*, **58**, 663–8.
2. Wiles, N.J., Zammit, S., Bebbington, P., *et al.* (2006). Self-reported psychotic symptoms in the general population. Results from the longitudinal study of the British National Psychiatric Morbidity Survey. *British Journal of Psychiatry*, **188**, 519–26.
3. Hickling, F.W., McKenzie, K., Mullen, R., *et al.* (1999). A Jamaican psychiatrist evaluates diagnoses at a London psychiatric hospital. *British Journal of Psychiatry*, **175**, 283–5.
4. Morgan, C., Mallett, R., Hutchinson, G., *et al.* (2004). Negative pathways to psychiatric care and ethnicity: the bridge between social science and psychiatry. *Social Science and Medicine*, **58**(4), 739–52.
5. World Health Organization (1973). *The international pilot study of schizophrenia*, Vol. 1. WHO, Geneva.
6. Westermeyer, J. and Janca, A. (1997). Language, culture and psychopathology: conceptual and methodological issues. *Transcultural Psychiatry*, **34**, 291–311.
7. Jablensky, A., Sartorius, N., Ernberg, G., *et al.* (1992). Schizophrenia: manifestations, incidence and course in different cultures: a World Health Organization ten-country study. *Psychological Medicine*, **20**, 1–97.
8. Wig, N., Menon, D.K., Bedi, H., *et al.* (1987). Expressed emotion and schizophrenia in North India. II. Distribution of expressed emotion components among relatives of schizophrenic patients in Aarhus and Chandigarh. *British Journal of Psychiatry*, **151**, 160–5.
9. Cantor-Graae, E., Pedersen, C.B., McNeil, T.F., *et al.* (2003). Migration as a risk factor for schizophrenia: a Danish population-based cohort study. *British Journal of Psychiatry*, **182**, 117–22.
10. Fearon, P., Kirkbride, J. B., Morgan, C., *et al.* (2006) Incidence of schizophrenia and other psychoses in ethnic minority groups: results from the MRC AESOP study. *Psychological Medicine,* **36**, 1541–50.
11. Mallett, R., Leff, J., Bhugra, D., *et al.* (2002). Social environment, ethnicity and schizophrenia: a case control study. *Social Psychiatry and Psychiatric Epidemiology*, **37**, 329–35.
12. Leff, J.P., Fischer, M., and Bertelsen, A. (1976). A cross-national epidemiological study of mania. *British Journal of Psychiatry*, **129**, 428–37.
13. Makanjuola, R.O. (1982). Manic disorder in Nigerians. *British Journal of Psychiatry*, **141**, 459–63.
14. Greenberg, L., Cwikel, J., and Mirsky, J. (2007). Cultural correlates of eating attitudes: a comparison between native-born and immigrant university students in Israel. *International Journal of Eating Disorders*, **40**, 51–8.

15. Üstün, T.B. and Sartorius, N. (eds.) (1995). *Mental illness in general health care*. Wiley, New York.
16. Leighton, A.H., Lambo, T.A., Hughes, C.C., *et al.* (1963). *Psychiatric disorder among the Yoruba*. Cornell University Press, Ithaca, New York.
17. Stanley, D. and Wand, R.R. (1995). Obsessive-compulsive disorder: a review of the cross-cultural epidemiological literature. *Transcultural Psychiatric Research Review*, **32**, 103–36.
18. Orley, J. and Wing, J.K. (1979). Psychiatric disorders in two African villages. *Archives of General Psychiatry*, **36**, 513–20.
19. Hollifield, M., Katon, W., Spain, D., *et al.* (1990). Anxiety and depression in a village in Lesotho, Africa: a comparison with the United States. *British Journal of Psychiatry*, **156**, 343–50.
20. Wig, N.N. and Pershad, D. (1975–1977). *Triennial statistical report*. Department of Psychiatry, Postgraduate Institute of Medical Education and Research, Chandigarh, India.
21. Nichter, M. (1981). Negotiation of the illness experience: ayurvedic therapy and the psychosocial dimension of illness. *Culture, Medicine and Psychiatry*, **5**, 5–24.
22. Jacob, K.S., Bhugra, D., Lloyd, K.R., *et al.* (1998). Common mental disorders, explanatory models and consultation behaviour among Indian women living in the UK. *Journal of the Royal Society of Medicine*, **91**, 66–71.
23. Obeyesekere, G. (1985). Depression, buddhism and the work of culture in Sri Lanka. In *Culture and depression: studies in the anthropology and cross-cultural psychiatry of affect and disorder* (eds. A. Kleinman and B. Good), pp. 134–52. University of California Press, Berkeley, CA.
24. Bibeau, G. (1997). Cultural psychiatry in a creolizing world: questions for a new research agenda. *Transcultural Psychiatry*, **34**, 9–41.
25. Kakar, S. (1995). Clinical work and cultural imagination. *Psychoanalytic Quarterly*, **64**, 265–81.

1.4

The history of psychiatry as a medical specialty

Pierre Pichot

Introduction

In 1918, Emil Kraepelin wrote:[1]

> A hundred years ago, they were practically no alienists. The care of the mental patients was nearly everywhere in the hands of head supervisors, attendants and administrators of the houses for the mentally ill and the role of the physicians was limited to the treatment of the physical illnesses of the patients.

He pointed out that, in the first decades of the 19th century, many of the books dealing with psychiatric themes were still written by medical doctors, such as Reil (who coined the word psychiatry), who had few contacts with mental patients or even by philosophers and theologians, and that only in the great scientific centres had specialists appeared 'who had decided to spend their life in the study and treatment of mental diseases'.

The history of psychiatry as a medical specialty has to be distinguished from the history of psychiatric medical knowledge which began in ancient Greece with the birth of medicine as a science. For more than 2000 years, only physicians observed and treated mental illnesses, and institutions were created in which the 'lunatics' and the 'insane' were received. But, as rightly pointed out by Kraepelin, the truth is that psychiatry was not really a medical specialty. One can argue about the precise date of the appearance of psychiatry as a specific field of medicine and of the psychiatrist as a specialist, devoting his professional competence exclusively to the care of the mentally ill. Denis Leigh recognizes that 'some degree of specialization occurred [in England] among respectable physicians' in the middle of the eighteenth century when the monopoly of Bethlem was broken and new 'lunatic hospitals', such as St Luke's were opened.[2] On the other hand, the American historian Jan Goldstein stresses that in France the language, as an exact reflection of the underlying reality, began to use expressions such as *homme spécial* to describe a physician specializing in a branch of medicine such as psychiatry only around 1830.[3]

Pinel and the birth of psychiatry as a branch of medicine

Despite those divergences, it is generally accepted that the work of Philippe Pinel constitutes a turning point. His role has several aspects. He is known worldwide as the physician who 'liberated the insane from their chains' in a dramatic initiative he started in 1793, at the height of the French revolution, at the Bicêtre asylum, and completed 3 years later at the Salpêtrière asylum. However, the reality is more complex.

Pinel, who was born in 1745, had studied medicine, translated Cullen's books into French, and published scientific papers on various subjects. He acted as a physician in a small Parisian 'madhouse', the Pension Belhomme, in which wealthy lunatics were confined at the request of their families. At that time most of the Parisian insanes were confined for a few weeks in the general hospital—the Hôtel Dieu. If their state did not rapidly improve, they were considered as incurable and send to Bicêtre or the Salpêtrière, built a century before, which also received other social deviants like beggars and prostitutes. Pinel, who was known by his politically influential friends for his progressive scientific ideas, was appointed physician to Bicêtre. The division for the insane was under the direction of an overseer (*surveillant*), Pussin, who had already introduced humanitarian reforms in the care of the patients. Pinel's merit was to approve and systematically develop Pussin's empirical measures and to propose an explicit scientific theory for their mode of action. Inspired by Crichton's views about the nature of the 'passions' by Condillac's psychology, and by the ideas of Jean-Jacques Rousseau, he created the *traitement moral* which he claimed to be effective with patients previously considered as incurably ill.

The improvement of the conditions in which the insane were cared for, supported and expanded by Pinel, was not an isolated French phenomenon. In Tuscany, Chiarugi in 1789 had already asserted that the basis of the extensive reforms he had introduced in the local asylum for the insane was that 'it is a supreme moral duty and a medical obligation to respect the mental patient as a person'. In England, where public had been shocked by the inhuman treatment to which King George III had been submitted during his mental illness; and where, a pious Quaker, William Tuke, deeply affected by the conditions in which the wife of a member of the Society of Friends had died in York lunatic asylum, decided to set up a special institution under the government of the Friends 'for the care and accommodation of their own members'. At the Retreat, opened in 1796 near York, physical restraints were largely abolished, and religious and moral values were emphasized in relations with the patients.

Chiarurgi's reforms did not survive the upheavals caused by subsequent wars and the political divisions of Italy, and Tuke's creation of the Retreat had not been prompted by medical considerations but was the expression of religious humanitarian purposes. The role played by Pinel was decisive, not so much because of the changes he promoted in the conditions of the patients, although they had a profound influence, but because he made the study and treatment of mental disorders a branch of medicine.

In 1801, Pinel published the *Medico-philosophical Treatise on Mental Alienation*. In it, he presented the various clinical manifestations he had observed; proposed a simple nosological system largely borrowed from older authors; examined possible aetiological factors; and described his 'moral treatment' in detail. The book has remained a landmark in the history of psychiatry, even being considered by the philosopher Hegel as a 'moment of capital importance in the history of humanity'. For Pinel, insanity was a disease and the patient affected by it remained, despite the loss of his reason, a human being. Its study, like the rest of medicine, had to be 'a science which consists of carefully observed facts'. Goldstein[3] has shown that Pinel's main preoccupation was to prove this scientific nature of the new medical specialty by repudiating the previous practices of the 'empirics' and 'charlatans'—the two terms being practically synonymous. He had accepted the method Pussin had developed empirically and transformed it in his moral treatment by providing a scientific theory of its mode of action. A curiously premonitory aspect of his emphasis on the necessity of a scientific methodology is to be found in his *Tables to Determine How Probable is the Curability of Alienation*, published in 1808. He provided statistical data on the efficacy of his therapeutic method according to the types of mental disorders and in comparison with spontaneous evolution, and concluded that medicine can only be a true science through the use of the calculus of probability!

Psychiatry as a profession: Esquirol and the clinical approach

If, because of the international influence of the ideas expressed in his book, Pinel is the founder of psychiatry as a medical discipline, he was not a psychiatric specialist in the strict meaning of the term. Although he retained his position at the Salpêtrière until his death in 1826 and is known today for his contributions to mental medicine, he had many other medical interests which gave him, in his time, a leading position among the Paris physicians; his *Philosophical Nosology*, published in 1796 and a classical reference for several decades, deals with general pathology. The case of his pupil and successor, Esquirol, who became the prototype of the psychiatric specialist was very different. At the Salpêtrière he was only in charge of the 'section of the insane'. He was later appointed medical director of the Charenton psychiatric asylum near Paris and owned in addition a small clinic, in which he treated his private patients. All his activities were exclusively dedicated to the study and treatment of mental disorders and the teaching of psychiatry. His book, *On Mental Diseases* published in 1838, in which he collected his previous publications, acquired a fame as great as Pinel's *Treatise*. In 1913, Karl Jaspers recognized that the later great representatives of German psychiatry, such as Griesinger and Kraepelin, were strongly indebted to Esquirol. He, and the school he founded, effectively developed one of the basic tenets of the new medical specialty. For Esquirol, careful objective observation and analysis of the symptoms and the behaviour of the patients were fundamental. He originated the descriptive clinical approach expanded by his pupils. Even more than Pinel, he was suspicious of unproved theories and when he eventually suggested relations between pathogenic factors and syndromes, he remained extremely cautious in his interpretations. Zilboorg, the psychoanalytically oriented historian of psychiatry, has accused this predominantly descriptive approach of creating a 'psychiatry without psychology' because, lacking psychodynamic concepts, its attempted objectivity remained at an allegedly superficial level.[4] The truth is that it laid the foundations of the present description of the mental disorders. The 'atheoretical' descriptive approach adopted in the present nosological systems—both the American *Diagnostic and Statistical Manual* and the *International Classification of Diseases*—whose proclaimed purpose is to emphasize the medical character of psychiatry is, in this respect, a return to Esquirol's principles.

The social aspects of psychiatry and the asylum system

By the end of the eighteenth century it was recognized that the study of mental alienation was part of medicine. However, mental diseases were of such a nature that it was not possible to treat the insane in the same conditions as patients affected by other diseases. Their most obvious manifestations had social consequences. According to the prevailing philosophical view, the mentally ill were deprived of free-will by their illness. In practice, they were unable to participate in the normal life of the society and were often considered as potentially dangerous. Because of this, they had generally been confined in madhouses of various kinds. One of the aspects of the reforms initiated by Pinel had been to make more explicit the difference in nature between the socially deviant behaviour of the insane, which, being the consequence of an illness, belonged exclusively to medicine, and the other deviations which society had to control and eventually to repress. The implementation of this fundamental distinction during the first half of the 19th century helped to give psychiatry its specific shape as a profession by being at the origin of forensic psychiatry and by leading to the formulation of precise rules concerning the commitment of the insane to institutions of a strictly medical character.

The legal code promulgated by Napoleon in 1810 stipulated that 'no crime or delict exists if commited in a state of dementia', with the old term dementia being used as a synonym of Pinel's mental alienation. This legal provision, introduced in similar forms in other countries, opened an important domain of activity to the medical profession of psychiatrist. Because of their now recognized specialized knowledge, the alienists were to help the judges in determining whether the mental state of an individual convicted of a 'crime or delict' was normal or pathological, with decisive consequences on the subsequent decision. The title of Esquirol's *Treatise* mentions explicitly that it describes mental diseases 'in their medical, hygienic and medico-legal aspects'. The conflict (which still exists) between the judges, usually supported by public opinion, who took a restrictive view of the concept of mental disease; and the psychiatrists, who tended to expand it to include new types of deviant behaviour, is illustrated by the violent controversies provoked by Esquirol's description of 'homicidal monomania'. They had an even more famous counterpart in England. J.C. Pritchard, an admirer of

Esquirol, had isolated 'moral insanity' as a specific mental disorder in two books published in 1837 and 1842; in the second, he examined its 'relations to jurisprudence'. Half a century later, in 1897, Henry Maudsley, who was in favour of the use of this diagnosis, recognized that this category, although internationally accepted by the psychiatrists, corresponded to

> . . . a form of mental alienation which has so much the look of vice and crime that may persons regard it as an unfounded medical invention'. Judges have repeatedly denounced it from the bench as a 'most dangerous medical doctrine', 'a dangerous innovation' which, in the interest of society, should be reprobated.

The general acceptance of the new medical concept of mental alienation implied the existence of adequate facilities for the treatment of the patients. The creation of new asylums—the term was retained—and the reorganization of the old ones were the answers. The French law of 1838 that fixed the detailed rules for the expansion of the new system to the whole country and for its functioning and financial support had a model character. Similar results were obtained in, for example, England with the Asylum Act 1828 and the Lunacy Act 1845. Outwardly, the new system was the extension, under more humane conditions, of the previous institutional practices. However, it had radically original features. While recognizing the necessity of protecting society, it stressed the fact that the insane had a fundamental right to be protected and medically treated in a competent way. The deprivation of liberty for the patients which it still implied, was strictly controlled to prevent possible misuse and was anyway justified, according to Esquirol and most contemporary psychiatrists, not only by the loss of free will, which was a consequence of the illness, but also by the therapeutic value of separation from a pathogenic milieu.

The asylum system became the central element of psychiatric care and was both the consequence and determining factor of the emergence of psychiatry as a medical specialty to which it gave, until the end of the 19th century and even beyond, an original character. The asylums acquired a quasi-monopoly in the care of the mentally ill. The few private institutions reserved for the wealthier members of the population, which often belonged to alienists in charge of the asylum, were generally submitted to the same legal rules. Private practice with ambulatory patients, as existing today, was exceptional or dealt with cases which were not then considered to belong to mental alienation. As a result, the study of mental illness was predominantly restricted to the more severe forms of disorder. Another consequence was that the alienists in charge of patients committed to the asylums had a dual function, a fact that differentiated them from other hospital physicians. In addition to their medical duties, they were involved in legal procedures which determined the conditions of admission, stay, and eventually release of the mentally ill. As superintendents, they also often had economic and financial responsibilities, being in charge of the material as well as the medical aspects of the functioning of their institutions.

Despite the fact that the laws now strictly differentiated the nature of the limitations of liberty in asylums and in prisons, the participation of the alienist in a form of social control was eventually perceived negatively by the public, and often by other physicians, and contributed to accentuating the specificity of psychiatry inside medicine. During the third and the fourth decades of the 19th century, which saw the birth of the asylum system, the psychiatrists became really conscious of their identity as a professional group. In England, France, Germany, and the United States they founded societies and began to publish journals with specialized scientific goals. Such a description oversimplifies an evolution which was progressive and in some cases took different directions. The creation and the extension of the asylum system took many years; it did not reach its classical form until the last part of the century, as testified by the famous campaign conducted in the United States during the 1840s by Dorothea Dix who complained that many of the mentally ill were still incarcerated in almshouses and prisons. The moral treatment practized in the institutions was eventually used to justify brutal measures, alleged to be therapeutic, and the behaviour of the attendants, who were not usually medically trained (significantly, they were known as *surveillants* in France), was too often of a purely repressive character. It was a long time before the proposals made in 1856 by the British psychiatrist John Conolly in his book, *The Treatment of the Insane without Mechanical Restraints* were put into practice everywhere.

The biological and the psychological model

The clinical orientation of Pinel, Esquirol, and their followers was basically empirical. By concentrating on describing observable symptoms and abnormal behaviours, it avoided theoretical controversies. However, many believed that if psychiatry was to become a branch of the medical sciences and to progress, it had to adopt models similar to those accepted by the rest of medicine. According to the anatomoclinical perspective, which was now dominant, diseases were distinct entities. Each disease was defined by a characteristic pattern of symptoms provoked by a lesion or eventually, a dysfunction of an organ to be discovered at autopsy. In 1821, Bayle, following this scheme, described the typical clinical symptoms and lesions of the brain in the general paralysis of the insane. Despite the disappointing results of the further anatomopathological studies (brain lesions were observed in only a small proportion of cases), there was increasing conviction that, with better investigation methods, mental disorders, like other diseases, could be explained by somatic causes. The degeneration theory, proposed in 1857 by Morel, which attributed many forms of insanity to the hereditary transmission of dysfunction of the nervous system produced by the noxious effects of environmental factors, and whose influence lasted until Kraepelin, is another expression of this biological orientation whose aim was to give psychiatry an undisputed medical status.

The biological and the purely clinical approaches were concerned with different conceptual levels—the discovery of the causes of insanity and the description of its manifestations respectively. Therefore, they could easily coexist. Even when the followers of Pinel and Esquirol expressed reservations about the applicability of the biological model to every type of mental disorder, they still believed in the medical nature of psychiatry. The situation created in the German-speaking countries by the school of the 'mentalists' (the term *Psychiker* by which they were known means 'psychologically oriented'), who were predominant during the first half of the 19th century, was very different. Influenced by philosophical, religious, and romantic trends, these psychiatrists took a radical dualistic position, postulating the absolute difference between the physical body and the spiritual soul. The soul was the source of the whole psychic life and hence eventually of its abnormal aspect—insanity. A term such as disease, appropriate for the somatic illness,

could only be used metaphorically in psychiatry. The sins of the patients were the origin of the mental disorders, and psychiatry belonged more to moral philosophy than to medicine. These ideas were developed in various related forms by the majority of the German psychiatrists of the period (Heinroth, Ideler, Langerman, and many others). Their ideological position had two consequences: scientific relations with other schools, such as the French and the English who saw in the publications of the mentalists obscure philosophical theories devoid of medical character, were largely cut off, and they provoked a violent reaction in Germany itself. The most extreme representatives of the contending group of 'somatists' (*Somatiker*), such as Jakobi and Friedreich, saw the mental disorders as symptoms of somatic diseases, not necessarily of the brain. In fact for them mental diseases as such did not exist. They defended aggressively their biological and sometimes bizarre hypotheses, such as the aetiological role of intestinal worms, against the mentalists. Finally, around 1850, they gained the upper hand. The publication in 1845 of *Pathology and Therapy of the Nervous Diseases* by Wilhelm Griesinger, an heir to their school who was also influenced by the French alienists, is a landmark in the history of the German psychiatry. With his appointment in 1865 as professor of psychiatry in Berlin, where he succeeded the mentalist Ideler, medical psychiatry was definitely established in Germany as a branch of the natural sciences.

The rise of neuropsychiatry

Romberg's *Lehrbuch der Nervenkrankheiten* symbolizes the birth of neurology as an autonomous medical specialty studying and treating the diseases of the nervous system. It was published 5 years after Griesinger's *Textbook* in which, adopting and expanding Bayle's anatomoclinical model he had affirmed: 'Mental diseases are diseases of the brain'. If both psychiatric and neurological symptoms originated in the nervous system, some form of association between the two specialties was a logical step, at least at the conceptual level. One aspect of their complex relationship was the creation of neuropsychiatry which developed its most characteristic aspects in the German-speaking countries.

The universities acquired considerable power and influence in the second half of the 19th century. From the 1850s on, chairs were created for the teaching of the new common discipline and special institutions, the university clinics, were built with hospital beds for psychiatric patients (if their disorders became chronic they were sent to the nearest asylum), laboratories for research on neurophysiology and neuroanatomy, and special wards for the neurological cases were developed. Griesinger's first move when he took over the chair of psychiatry at Berlin was the creation of neurological wards at the Charité. The leading neuropsychiatrists in charge of these institutions often performed research in both fields with equal competence, as shown by the work of Wernicke and Westphal, and later of Kleist and Bonhöffer, in Germany and of Meynert in Austria.

The concept of neuropsychiatry, appearing at a period during which the German school was progressively gaining influence, had a deep impact on psychiatric thought and on the psychiatric profession, even if its institutional driving force, the university clinic system, was not developed everywhere to the same extent as in Germany. For example, it was conspicuously absent in England, despite the fact that the theoretical position taken by the most important psychiatrist of the time, Henry Maudsley, was very close to that of Griesinger. The National Hospital in Queen's Square, London, founded in 1860, retained a virtual monopoly on the teaching of neurology for many decades, and psychiatry, taught essentially in hospitals, was not represented at university level until the 1930s. However, in most countries, neuropsychiatric institutions coexisted with the asylums where the alienists had the unenviable task of caring for chronic mental patients, often with inadequate means. The concept of neuropsychiatry reflected a basically biological perspective on the aetiology of the mental illnesses, expressed in the creation of a new specialty associating competence in the two previously separated domains of medicine. However, it provoked ideological and professional tension between the 'pure' psychiatrists, mainly those in charge of asylums, and the neuropsychiatrists, predominantly involved in teaching and research. In the long term, this conflict was one of the factors which finally led, in the 1960s, to the almost complete administrative and institutional separation of the two specialties in countries such as France where they had been, at least formally, associated. But many traces of the old situation remain. The most influential scientific journal published in German, *Nervenarzt*, still deals equally with neurology and psychiatry, and the term 'neuropsychiatric' survives in the titles of many teaching and research institutions.

The neuroses and the birth of the psychotherapies

The study of the neuroses, in which the relation between psychiatry and neurology was also involved, resulted in completely different, but equally important changes to psychiatry as a medical specialty. The term neurosis had been coined in 1769 by Cullen to describe a class of diseases he attributed to a dysfunction of the nervous system. In this very heterogenous group, two entities of very ancient origin, hysteria and hypochondriasis, had predominantly psychological manifestations. Since the affected patients were not usually commited to asylums, they were not normally studied by alienists, but by specialists in internal medicine such as Briquet, who in 1859, wrote the classical *Treatise on Hysteria*. Because of the assumed nature of the neuroses, the new discipline of neurology rapidly took an interest in them.

Charcot, the founder of the French neurological school, was responsible for the internal medicine wards at the Salpêtrière—they were not associated with the 'divisions of the insane' at the same hospital, the domain of the alienists. In about 1880, he became interested in hysterical patients who, because of their seizures, were admitted to the same ward as the epileptics. He developed a purely neurological theory of the disease which he described and studied using hypnosis. This was the former 'animal magnetism', long fallen into disrepute, but to which he gave a new scientific status. Charcot's descriptions of the *grande hystérie*, which he demonstrated on selected patients in his famous public lectures, were justly criticized later, but his international fame attracted students from all over the world. One of them, was a young lecturer in neuropathology at the University of Vienna, Sigmund Freud, who, impressed by Charcot's lectures, decided to devote all his energies to the study and the treatment of the neuroses. Another was a French professor of philosophy (psychology was then a branch of philosophy), Pierre Janet, who had become interested in the psychological aspects of the neuroses. He was later to develop, in parallel with Freud, a

psychopathological theory which, despite the traces it has left (the concepts of psychasthenia and the dissociative processes in hysteria) was not to be as internationally successful as Freud's psychoanalysis. Charcot's ideas were opposed by Bernheim, the professor of internal medicine at the Nancy Medical School and also an adept of hypnosis. He attacked the neurological interpretations of the Salpêtrière and claimed that suggestion played a central role in the phenomena described by Charcot.

The general interest in the neuroses, which extended beyond medicine to *fin de siècle* literature, was an international phenomenon. In 1880, Beard, an American neurologist, described a new neurosis, neurasthenia, which soon aroused even more interest than Charcot's hysteria. Psychiatry had played almost no part in this evolution, but this was to change under the influence of three related developments: the changes which took place within the concept of neurosis, the birth of the psychotherapies, and the incorporation in the field of psychiatry of psychopathological manifestations, even if they were of minor intensity.

The transformation of the concept of neurosis is apparent in the position taken by Kraepelin in the 1904 edition of his *Textbook*. He introduced a chapter,'The psychogenic neuroses', on the grounds that 'among the neuroses, to which belong epilepsy and chorea, one must isolate a sub-group characterized by the purely psychological cause of the apparition of the symptoms'. The disintegration of the old concept left to neurology, which from now on abandoned the generic term, diseases (such as epilepsy and chorea) whose somatic manifestations could be shown to express a dysfunction of a precise part of the nervous system. Psychiatry took charge of hysteria, hypochondriasis, neurasthenia, and the related phobic, obsessional, and anxious disorders, which constituted the new neuroses. This concept was justified by the psychological nature of the symptoms and the causes recognized even by a biologically oriented psychiatrist such as Kraepelin. This redrawing of the frontier between the neurological and psychiatric specialties also testified to the extension of the limits of psychiatry. Pinel's insanity, until then defined by the necessity of commitment to special institutions, was replaced by a broader concept. A new class corresponding to our present personality disorders had already appeared in the 1894 edition of Kraepelin's *Textbook*. It had been isolated for the first time in 1872–1874 by the psychiatrist Koch. Like the neuroses, the cases were rarely observed in asylums but, nevertheless, they were now considered as belonging to the psychiatric field of study.

This field was further modified by the birth of the psychotherapies. In fact, they had a long history. In 1803, one of the first German mentalists, Reil, had described under the name of 'psychic therapy' (*psychische Curmethode*) a number of procedures, including very violent somatic ones, which could influence the 'perturbed passions of the soul', and Pinel's moral therapy contained psychotherapeutic elements. However, psychotherapies as techniques whose formal rules were based on an explicit theory about their psychological mechanisms of action, derived mainly from Mesmer's animal magnetism as rehabilitated by Charcot. The emergence of the psychotherapies, characteristic of the last decades of the 19th century, was intimately related to the renewed study of the neuroses. After he had abandoned hypnosis, Freud developed psychoanalysis, but many other techniques evolved during the same period, which were as well or even better known at the time, although they were to have a less lasting success. One of these was the method of Janet, who still occasionally used hypnosis. In 1904, Dubois, a Swiss neuropathologist from Bern, introduced a technique influenced by Bernheim's theory of suggestion in *The Psychoneuroses and their Moral Treatment*, and claimed to produce a 'psychological re-education' by a combination of rational and persuasive elements. His international reputation brought him patients from all over the world. The 'rest cure', proposed in 1877 by the American neurologist S. Weir Mitchell for the treatment of hysteria and later of neurasthenia, was combined with Dubois' method by Dejerine, Charcot's successor as the professor of neurology in Paris.

This very incomplete summary illustrates the striking fact that, because of their intimate connections with the neuroses, the psychotherapies originated inside neurology. When the study and treatment of the neuroses were incorporated into psychiatry, the psychiatrists considered that they were an integral part of their activity and tried to retain the monopoly of their practice. They never completely succeeded. Already Freud had, according to his biographer Jones, 'warmly welcomed the incursion in the therapeutic field of suitable people from another walk of life than medicine'. The problem of the 'lay analysts', a source of conflict within the psychoanalytic movement, is only an aspect of a broader question which was later to involve the relations of the medical specialty of psychiatry with the new professional group of clinical psychologists.

From the beginning of the 20th century to the Second World War

During the first half of the 20th century, psychiatry developed in many directions. Kraepelin's monumental synthesis[5] established around 1900, a nosological system which, in its broad outlines, has remained valid until today. Without being radically altered it was completed, to mention only a few contributions, in 1911, by Bleuler's description of schizophrenia and in 1913, by Jaspers' psychopathological perspective, developed by the Heidelberg school and Kurt Schneider, and by other psychiatrists working in academic institutions. However, the old conflict between the 'mentalists' and the 'somatists' reappeared in a modified form. The mainstream of psychiatry had abandoned the extreme positions of the 'brain pathologists' of the Meynert-Wernicke type but, while recognizing a limited influence of psychological factors, admitted in a general way the biological origin of the more severe mental disorders—the psychoses. The empirical discoveries of biological treatments—of general paralysis by malaria therapy (Wagner von Jauregg in 1917), of schizophrenia by insulin coma (Sakel in 1933) or by chemically induced seizures (von Meduna in 1935), and of depression by electroconvulsive therapy (Cerletti in 1938)—not only helped to dispel the prevailing therapeutic pessimism, but provided supporting arguments. However, an opposing ideological current represented by psychoanalysis had arisen from the study of the neuroses. Its attention was concentrated on the study of complex psychopathological mechanisms postulated to be at the origin of the neurotic, and later also of the psychotic symptoms, favoured psychogenetic aetiological theories, and advocated psychotherapy as the fundamental form of treatment. Psychoanalysis expanded steadily during this period and gained enthusiastic adherents in many countries. However, partly because of the suspicion and even hostility of many members of the psychiatric establishment, they remained isolated in close-knit groups with their own teaching

system independent of the official medical curriculum, and the use of their therapeutic technique was restricted to a small number of mostly neurotic patients seen in outpatient clinics or, more often, in private practice.

The great majority of patients suffering from mental disorders were still confined in asylums, and the enormous increase in their number, mainly related to the social changes accompanying industrialization and urbanization, although other factors have been invoked, was striking. In Great Britain it grew from 16 000 in 1860 to 98 000 in 1910, three times more rapidly than the population. A similar phenomenon was observed in all countries and persisted until the end of the 1940s despite the introduction of the first biological therapies. In the United States, there were already 188 000 patients in mental hospitals in 1910, and by the end of the Second World War, 850 000 were lodged in huge institutions which were overpopulated, understaffed, and could only provide custodial care. This obvious degeneracy of the asylum system, contrasting with the progresses in the scientific field, stimulated efforts to improve the practice of psychiatry and its institutional framework. Most of these improvements took place after 1920 and, although their results remained relatively limited, they were the forerunners of later more drastic changes.

The education of psychiatric specialists, which had varied widely from country to country, was improved and systematized. A convergence of evolution is apparent during this period which can be said, to some extent, to have seen the formal administrative recognition of psychiatry as a medical specialty. Educational programmes and controls of the level of competence were introduced which extended beyond psychiatrists in academic positions. A limited teaching of psychiatry became compulsory even in the general medical curriculum. In France, psychiatrists for public asylums and, in some cases, residents in psychiatry were selected by a competitive examination system. In England, the Board of Control recommended in 1918, that a leading position in a psychiatric institution could only be occupied by a physician who had obtained a Diploma in Psychological Medicine awarded by the Royal College of Physicians and by five universities. In the United States, the moving force behind the reforms was Adolf Meyer, the Director of the Henry Phipps Clinic at Johns Hopkins University from 1913 to 1939, who organized a systematic residency system and promoted the creation of the Board of Neurology and Psychiatry. This Board was established in 1936 and awarded a diploma which it became necessary to hold, to be recognized as a specialist.

The changes were reflected in the vocabulary. The term psychiatry, originating in the German-speaking countries and mostly used there, was adopted everywhere at the beginning of the century. In France, the health authorities officially substituted '*hôpital psychiatrique*' for '*asile d'aliénés*' and '*psychiatre*' for '*aliéniste*' in the 1930s. In England, a Royal Commission used the words 'hospital', 'nurse', and 'patient' instead of 'asylum', 'attendant', and 'lunatic' for the first time between 1924 and 1926. However, efforts were also made to dissociate, when possible, the social protection function of the institutions from their medical role by allowing them to admit patients under the same conditions as the general hospitals. In 1923, a special section was created in the Paris Sainte-Anne asylum which provided treatment to voluntary patients and had both hospital beds and a large outpatient department. In England, the Mental Health Act of 1930 made voluntary admissions to psychiatric hospitals possible; by 1938, they already constituted 35 per cent of all admissions.

Social considerations had always been evident in psychiatry, but their traditional expressions had mainly been of a negative nature, i.e., the confinement of patients in asylums. The new possibility of free admissions reflected an increase in tolerance towards the disturbing character of mental illness. At the same time, a differently oriented and broader social perspective appeared. The concept of mental hygiene originated in the United States in 1919 with the creation by a former patient, Clifford Beers, of an organization whose internationally growing influence was manifested by well-attended congresses held in Washington in 1930 and in Paris in 1937. From its beginning, the movement was not purely medical and was influenced by various humanitarian philosophical trends. It emphasized the role of social factors, such as living conditions or educational practices, in the origin of mental disturbances and promoted their prevention and treatment by the close co-operation of psychiatrists and nurses with non-medical groups in the community. One of the institutional consequences of these ideas was the creation of the profession of social worker. They began their activity in Adolf Meyer's clinic (Adolf Meyer had been an early supporter of the mental hygiene movement whose principles converged with his own ideas) at the Sainte-Anne Hospital in Paris, in England where the London School of Economics opened a special training course in 1929, and elsewhere.

Contemporary with the emergence of psychiatric social work was the expansion of clinical psychology. The Binet-Simon scale for the measurement of intelligence, developed in 1905, was the first application to psychiatry of the new discipline of experimental psychology which had originated at the end of the previous century. This initial contribution led to the creation of a professional class of clinical psychologists who were initially concerned with the development and use of psychological assessment instruments and with theoretical research in a few psychiatric centres. Their number initially remained low; in 1945 the United States, where they were most numerous, had about 4 000 psychiatrists but only 200 clinical psychologists.

The expansion of psychiatry after 1945

The Second World War coincided with a major transformation of the psychiatric specialty. The war had vividly demonstrated the frequency of mental disorders in the United States; they had proved to be the leading cause of medical discharges from the military service and the primary cause of almost 40 per cent of selective service rejections. The previously prevailing view that psychiatry was a minor and often somewhat despised medical discipline, concerned primarily with the custodial care of psychologically deviant and potentially troublesome individuals, was progressively dispelled. The preservation and the restoration of mental health—an expression from now on often used by national and international institutions—began to be considered by governments as an important task. The fundamental changes which took place after 1945 and shaped psychiatry as we know it today were the result of this new atmosphere and of the emergence of new perspectives in the three traditional domains— the psychological, the social, and the biological. Some appeared in slightly different forms at different times, their relative influence was submitted to variations, and eventually they came into conflict. The result has been an impressive expansion and increase of the efficacy of psychiatry, profound institutional transformations, and successive ideological waves

which have had a major impact on the professional position of the psychiatrist.

The demographic data reflect the new importance of psychiatry in medicine. In the United States, the proportion of psychiatrists in the medical profession was 0.7 per cent in 1920, 1.4 per cent in 1940, and 5.5 per cent in 1970, the rate of growth having doubled after the Second World War. In France, at present there are 18 psychiatrists for 100 000 habitants; they constitute 6 per cent of all physicians. Similar levels were reached during the postwar decades in the developed countries and remain relatively stable today. Even before this spectacular increase in numbers, psychiatrists had been becoming conscious of the necessity to affirm the identity of their discipline. The First World Congress of Psychiatry, held in Paris in 1950, has been followed by periodic meetings and by the creation of the World Psychiatric Association to which almost every national society of psychiatry belongs. The health authorities of various countries have become conscious of the necessity to provide adequate financial means to support research and training in the discipline. In 1946, the United States government created the National Institutes of Mental Health for such a purpose, and similar efforts were made in many countries although the structures of the organizations formed were different. To promote the same goals at the international level, the World Health Organization, created immediately after the Second World War, had a Section (later Division) of Mental Health which, among other co-ordinating activities, tried to overcome the difficulties of communication between the national schools by establishing a common nosological language.

While the changes affected almost all countries, they were most spectacular in the United States. From the end of the 19th century until the 1930s, the concepts developed in the German-speaking countries had been the most influential. This disappeared with the advent of the National Socialist regime which, under cover of racist theories, expelled many of the leading psychiatrists from Germany and Austria, introduced compulsory sterilization for several varieties of mental illnesses, and promoted the voluntary killing in psychiatric hospitals of mentally-retarded children and chronic patients. The United States, which had emerged from the Second World War as the most powerful country in the world, began to exert a widespread influence in psychiatry as in the rest of medicine. Because of the prestige of its research and teaching institutions and the worldwide influence of its scientific publications, reinforced by the progressive adoption of English as the language of international scientific communication, American psychiatry became a model in many countries, even though many of the theoretical trends and technical advances it adopted and developed had originated in Europe. However, in the United States, with a local colouring, they took on a special intensity.

The psychodynamic wave

An important factor in the spread of the doctrine of psychoanalysis was the emigration of a relatively large number of German and Austrian psychoanalysts to the United States from 1933 onwards. They had been compelled to leave their home countries for racial reasons—psychoanalysis had been condemned by the National Socialist regime as Jewish and Freud's books had been publicly burned. Many of the young psychiatrists trained in large numbers to answer the demands of the armed forces adopted psychoanalysis under the influence of some of those in charge of the programmes. For a generation, until the end of the 1960s, psychoanalysis became the dominant ideology in American psychiatry.

The American form of psychodynamism often deviated from Freudian orthodoxy, but it emphasized the role of psychogenetic factors, the value of the study of intrapsychic mechanisms, and the basic importance of psychotherapy, while giving little consideration to the traditional clinical approach and to nosology. The domination of this essentially psychological orientation, sometimes compared with the success of the German mentalist school during the first half of the 19th century, had important consequences. Although the disorders of hospitalized psychotics were eventually interpreted according to psychoanalytic theory, psychotherapy was mostly used, as it has been since its beginning, on ambulatory neurotic patients. As early as 1951–1952, 3 000 of the 7 500 American psychiatrists identified private practice as their main activity, and in 1954, the number of private psychiatrists exceeded that of their salaried colleagues for the first time, with a quarter of the former devoted exclusively to psychotherapy. However, with the initial encouragement of official institutions such as the Veterans Administration, the clinical psychologists began to engage in psychotherapeutic activities. The number of members of the Clinical Psychology Section of the American Psychological Association reached 20 000 in 1980, at a time when they were 26 000 psychiatrists in the United States. In public opinion, and to a certain extent in general medical opinion also, psychiatry was assumed to consist only of psychotherapy and psychology.

In most other countries the developments that occurred in the United States were not as intense, generally appeared later, and were modified by local traditions and influences. In the German-speaking countries they were delayed by the still powerful neuropsychiatric perspective and the temporary vogue for existential phenomenology. In the United Kingdom, the eclectic current fostered by the influential London Institute of Psychiatry during the decades following the war restricted the advance of psychodynamism; in 1956, *Time Magazine* could affirm, as a conclusion of a survey, that 'all of Great Britain [had] half as many analysts as New York City'. In France, the psychoanalyst Jacques Lacan gave the doctrine a special colouring. On the whole, however, the rise of psychodynamism was a general phenomenon, except in the communist countries where Freud's doctrine had been condemned on ideological grounds.

A reaction began in the 1960s with the successes of the new pharmacotherapies. Clinical psychologists had developed alternative radically different psychotherapeutic methods based on learning theories, especially the behaviour therapy introduced in 1958 by Wolpe, supported in the United Kingdom by Eysenck, and the cognitive therapy often associated with it. These methods competed successfully with the psychodynamic techniques and conquered a large part of the field. Psychodynamism did not disappear; many of its concepts retained their place in psychiatry and psychotherapeutic methods continued to be practized, but it lost its predominant ideological position. In addition to its theoretical contributions, when its influence on the professional aspect of psychiatry is considered from a historical perspective, it has been an important factor in the further expansion of the activity of psychiatrists in the treatment of relatively minor disorders and has also encouraged clinical psychologists to play an active and independent role in this field.

The social wave

At the end of the Second World War there was a great desire for social change; one of its aspects was the belief that everyone had a 'right to health' or at least the right to receive adequate medical care regardless of the ability to pay. This resulted in the creation of the National Health Service in the United Kingdom in 1948 and the Social Security system in France, together with similar developments in other countries. The social perspective, which was one of the basic principles underlying these developments, initiated major institutional changes in psychiatry. They were the result of a number of factors—the necessity to give to the whole population an easy access to psychiatric care, and also the belief that social elements played an important role in the aetiology of the mental disorders and that they could greatly contribute to the healing process, with the aim of progressively reintegrating the patient in the community.

The most spectacular aspect of the new policy was the decline of the asylum system, still in a dominant position in psychiatry; in fact, the number of patients in psychiatric hospitals in the developed countries reached its peak in 1955. The criticisms of the 'degeneration' of the functioning of psychiatric hospitals and the segregation of patients in institutions, often located far from their homes and families, were not new. However, the previous partial improvements, such as the decrease in the number of compulsory commitments or the creation of outpatient departments, were replaced by the creation of completely new structures. Ideally, the country would be divided into geographical zones or sectors with a population of about 100 000, and each zone would have a multidisciplinary team of psychiatrists, nurses, clinical psychologists, social workers, and occupational therapists responsible for mental health. Visits and therapeutic interventions in the patient's home and easily accessible outpatient departments were to play an increasingly important role. If hospitalization was necessary, it should be as far as possible in small units located in a general hospital where the time of stay was to be reduced to the absolute minimum. Special institutions such as day hospitals, night hospitals, and specially adapted workshops would contribute to the progressive readaptation of the patient to the life in the community. The introduction of this 'community care', which was expected to work in close co-operation with general practitioners and various public and private institutions, would result in the disappearance of the traditional psychiatric hospital and to 'deinstitutionalize' psychiatry. The new system was introduced in various forms in most countries after 1969. In the United States, the Community Mental Health Center Act was promulgated in 1967. In the United Kingdom, which had strong traditions of social psychiatry, plans for the implementation of community care were discussed in the 1960s, and in 1975, the Government White Paper *Better Services for the Mentally Ill* encouraged the formation of multidisciplinary 'primary care teams' which also included general practitioners. In France, an official directive in 1960 created the *psychiatrie de secteur* which was expected to result in the progressive elimination of *hospitalocentrisme*. The World Health Organization (WHO) encouraged all its member countries to adopt similar practices.

Although, in the last 40 years community care has become the official doctrine everywhere, except in Japan where the rate of hospitalization in mostly private hospitals has grown continuously, its implementation has not been easy despite the major therapeutic improvements brought about by pharmacotherapy. In some parts of the United States, the sudden closure of public psychiatric hospitals combined with the inadequacies of the Community Mental Health Centers were for a time at the origin of an appalling lack of care for a number of mentally ill people. The expected 'fading out' of hospitalization has been slow. According to the WHO, in 1976 the number of mental health beds (including beds for the mentally retarded) per 1 000 population was 6.5 in Sweden, 5.5 in the United Kingdom, 3 in France, and 2 in Germany. These figures have since decreased and the types of hospitalization have changed. In 1955, 77 per cent of the 'psychiatric care episodes' in the United States occurred in public psychiatric hospitals, compared with 20 per cent in 1990. In 1994, 1.4 million mental patients were hospitalized, but only 35 per cent in public psychiatric hospitals compared with 43 per cent in general hospitals and 11 per cent in private psychiatric hospitals, which increased in number from 150 in 1970 to 444 in 1988. In France, where the total number of psychiatric patients treated in public institutions (including children) is now about a million, 60 per cent are seen exclusively on an ambulatory basis, but the number of hospital beds has only been reduced by half.

Reflecting the increasing influence of social perspectives, the organizational changes modified psychiatry as a profession. The increase in the number of psychiatrists in private practice was paralleled, in general to a lesser extent, by an increase in the public sector where their role was modified. In the traditional asylum, the authority of the psychiatrist was unchallenged and limited only by the legal provisions related to the procedures of commitment. The nurses, and later the clinical psychologists, social workers, and occupational therapists, were 'paramedical auxiliaries' in a subordinate position. The creation of multidisciplinary teams, working in various settings, gave the psychiatrist a function of co-ordination made increasingly complex by the claims of professional autonomy made by the former auxiliaries. In some cases, such as in the American Mental Health Centers, the psychiatrists who were a small minority in the team and had less and less control over its functioning, resented what they considered to be the loss of their medical status.

The importance given to social factors was not limited to the system by which care was delivered. Sometimes, combined with radical ideological and political attitudes, it took more extreme forms. The criticisms, which first centred on the inadequacies of the existing institutions, extended to the concept of mental disease itself. The antipsychiatric movement claimed that mental diseases were artificial constructs which were not related to diseases in the medical meaning of the term. The allegedly pathological behaviours, such as those conceptualized as schizophrenia, were in fact normal reactions to an inadequate social system. The so-called treatments were techniques used by the ruling classes to preserve the social order of which they were the beneficiaries. The only solution was a drastic reform of society. Such theses varied in their content and in the arguments used. They were developed by authors such as Szasz, Laing, and Cooper in the English-speaking world, the philosopher Foucault in France, and the psychiatrist Basaglia in Italy. They reached their greatest influence in the 1960s and a few attempts were made to put their ideological principles into practice. Although they attracted much attention at the time, they were very limited and short lived. One of the few countries where this movement had a practical impact was Italy. Basaglia's strongly politically oriented theories were influential in the later legal reform of the antiquated asylum system, but, despite the apparently

revolutionary character of some of the new administrative provisions, the changes made were very similar to those taking place in other countries.

The biological wave

Psychotropic drugs, such as opium, had been used since the origin of the medical treatment of psychiatric patients. During the 19th and the first half of the 20th century, synthetic drugs such as the bromides, the barbiturates, and the amphetamines were developed. Some of them, especially the sedatives and hypnotics, had a real but in practice, marginal value in alleviating some symptoms. They had never constituted an effective treatment of mental disorders. Modern psychopharmacology not only initiated what has been rightly called a therapeutic revolution in psychiatry but also gave a powerful new impulse to the biological perspective. Its date of birth is usually considered to be 1952, when the remarkable activity of chlorpromazine on the symptoms of schizophrenia and mania was discovered. This had been preceded in 1949 by the demonstration of the value of lithium salts in manic states. A few years later, it was shown that the continuous administration of lithium salts prevented the recurrence of manic and depressive phases in the mood disorders. This was followed by the introduction of drugs acting on the depressive manifestations (imipramine and the monoamine oxidase inhibitors in 1957) and on anxiety (including chlorediazepoxide, the prototype of the benzodiazepines, in 1960). In one decade, clinicians had empirically discovered the fields of application of the main classes of psychoactive drugs—the neuroleptics, the antidepressants, the anxiolytics, and the mood stabilizers—which had been synthesized by biochemists and previously tested by pharmacologists on animal models. The scale and rapidity of the spread of their use had major repercussions.

The first was a modification of the image of psychiatry. The layman generally expected a physician to prescribe drugs to treat the disease from which he suffered. In part, because it did not conform to the expected therapeutic behaviour, psychiatry had been seen as an atypical and almost non-medical specialty. In addition to the specificity of the institutions in which it was generally practized, psychological techniques were unknown in the rest of medicine, and even the recently introduced biological techniques (the shock therapies and the lobotomy) had a somewhat strange and frightening character. The establishment of pharmacotherapy contributed strongly to modifying this perception, even if it did not completely remove the traditional prejudices.

The second consequence was even more important. There were, at least initially, controversies about the roles of pharmacotherapy and of the new social perspectives in the restructuring of the mental health care system. In fact, the number of inpatients in psychiatric hospitals began to decrease from 1955 on, and it seems obvious that the main cause was the therapeutic efficacy of the drugs. They reduced the mean length of hospitalization and eventually even made it unnecessary. Although some types of patients did not benefit from them and the mental state of others was only improved, many who had previously been condemned to long stays in the hospital were able to return to the community, with their treatment eventually being continued in rehabilitation settings and often on an ambulatory basis. Pharmacotherapy had made possible the practical implementation of social trends. In addition to this basic contribution to the 'deinstitutionalization' movement, pharmacotherapy was an essential factor in the growth of private practice. The success of psychotherapy had been one contribution to this, but the complexity of its techniques, the length of the treatment, its applicability to only a few types of disorders, and the uncertainty of the results limited its use to a relatively small number of selected patients, even in the United States during the period of the greatest popularity of psychodynamism. Pharmacotherapy could be used much more easily, on a much larger number of patients, and did not require a long and complex training. Some of the drugs, such as the anxiolytics, had an immediate symptomatic effect, and others (the antidepressants and the neuroleptics) could attenuate or suppress the pathological manifestations in a few weeks and, outside the acute phase requiring hospitalization, could be used on an ambulatory basis. It was not only private psychiatrists who were able to treat many of their patients successfully; general practitioners also began to prescribe psychotropic drugs on a large scale.

The third consequence was the explosive development of biological research in psychiatry. The first therapeutic discoveries were largely empirical, but new biochemical techniques allowed some of the modes of action of the drugs to be elucidated. From 1960 on, studies of the influence of these drugs on various aspects of neurotransmission in the brain stimulated hypotheses about the abnormal biochemical mechanisms considered to be the physical substrate of the mental disorders. Meanwhile new methods had been introduced for examination of morphological modifications of the living brain and even of the nature and localization of the biochemical processes taking place in its different parts. The discovery by Watson and Crick in 1953, of the chemical basis of heredity and the subsequent spectacular advances in molecular biology gave a fresh impulse to psychiatric genetics, which had been partly discredited by their misuse by the National Socialist regime. Under the name of neurosciences, these new fields of enquiry progressively acquired a dominant role in psychiatric research at the same time as the introduction of an ever-increasing number of drugs, eventually more potent, usually with less inconvenient side-effects, and sometimes with new therapeutic indications.

'Remedicalization' of psychiatry

In 1983, Melvin Sabshin, the Director of the American Psychiatric Association, summarized the overlapping chronologies of the psychodynamic, biological, and social waves as follows:[6]

> Psychoanalysis surged through the United States during the 1940s and the 1950s. During the 1950s, a new psychopharmacological approach emerged which had great impact on psychiatric practice generally . . . The 1960s saw the dawning of a community psychiatric approach which attempted to accomplish a massive desinstitutionalization of patients from public psychiatric hospitals.

Although less radical and not strictly identical, the general picture was similar in other countries. The 1960s saw an often uneasy coexistence of three schools. 'During that decade', wrote Sabshin, 'American psychiatry enlarged its boundaries and its practices so broadly that many critics grew increasingly concerned with the 'bottomless pit' of the field'. The extension of the practice of psychotherapy, frequently to cases with no clear pathological character, tended to blur the limits of the mental disease concept and to neglect the traditional diagnostic approach. Social work was also tempted to concern itself with problems with no obvious medical nature, such as those still described in 1978 in the United States by the President's Commission of Mental Health, which asserted

that 'American mental health cannot be defined only in terms of disabling mental illness and identified mental disorders' and identified as a domain of concern for workers in the field 'unrelenting poverty and unemployment and the institutionalized discrimination that occurs on the basis of race, sex, class, age . . .'. In sharp contrast, the new biological psychiatry recognized only a strictly medical model, stressing the necessity of an accurate diagnosis for the prescription of the drugs and for the testing of their efficacy, and advocated restrictive limits in the definition of the mental diseases.

Around 1970, a profound change took place. Although the institutional modifications of the care system favoured by the generalization of drug therapy continued and expanded under its various forms everywhere, the influence of psychodynamism began to decline within the psychiatric profession. According to the Director of the National Institutes for Mental Health 'it was nearly impossible in 1945 for a non-psychoanalyst to become Chairman of a Department of Psychiatry (in the United States)' but by the mid-1970s the situation was reversed. The publication by the American Psychiatric Association of the Third Revision of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III) is often considered as the symbolic expression of the change. This took place in 1980, but its origins were more than a decade previously, and it was significantly presented by its apologists, such as Klerman, as 'a decisive turning point in the history of American psychiatry . . . an affirmation of its medical identity'. The new nosology, which was categorical in nature and which introduced diagnostic criteria borrowed from experimental psychology in the delimitation of the categories, did not allow any reference to 'unproven' aetiological factors or pathogenic mechanisms, unless 'scientifically demonstrated'. It claimed to be purely descriptive and therefore acceptable as a means of communication by all psychiatrists, whatever their individual orientation may be. It was in fact perceived, not only in its country of origin, as a reaction against the extreme socio-psychological positions—the deletion of the term neurosis because of its usual association with the psychoanalytic theory of intra-psychic conflicts raised violent controversies—and, despite its proclaimed 'a-theorism', as favouring the biological medical model. Although initially exclusively devised for the use of the American psychiatrists, to the surprise of its authors it was rapidly accepted in all countries and the WHO adopted finally its principles in its own nosological system, the International Classification of Diseases. Originally the result of a brutal reversal of trends in the American psychiatry, it expressed a general change of direction in the psychiatric way of thinking towards the affirmation, against the forces believed to threaten it of the medical character of psychiatry.

Crisis in psychiatry?

At first glance, the new status of psychiatry seems to have taken firm root in the last three decades. It rests on the general acceptance of the medical definition of the concept of mental disease and of the progressive realization of a diversified but co-ordinated institutional system of mental health care. The biological perspective, even if it has taken a prominent place in research and therapy, is now combined with psychological and social approaches in the bio-psychosocial model. The psychiatrist, in accordance with his medical professional responsibilities, occupies a central position in a multidisciplinary team whose members contribute their special competences to the common goal.

This idyllic picture is far from a reflection of reality, even in the developed countries, and the existence of a crisis in psychiatry is evoked with increasing frequency. An indication of the loss of prestige of psychiatry in the medical profession is the alarming decrease of the proportion of American medical students choosing a psychiatric residency; it fell to 2 per cent in 1990, a level much too low to ensure the maintenance of the present demography. Under the pressure of economic constraints, efforts are made everywhere to control the rising burden of medical care. They have taken different forms according to the country—from the managed care system in the United States to the *numerus clausus* system in France, in which the number of internships available is determined by the government—but their common aim is to limit the number of psychiatrists and the cost of their activities. Paradoxically, the recognition of the frequency of the mental disorders and the growing demand for psychiatric treatments has been associated with a reduction in the domain of action of psychiatrists, who are now often vastly outnumbered by clinical psychologists and social workers. In the United States, by 1990, 80 000 'clinical' social workers were active in the psychiatric socio-psychological domain, a quarter of them in part- or full-time private practice. The claims of these powerful professional groups are not limited to a completely autonomous status but, in the case of the clinical psychologists, extend to the demand for a legal recognition of such typical 'medical privileges' as the right to hospitalize patients and to prescribe drugs. Even within medicine, psychiatry is under attack. In Germany, a medical psychotherapeutic specialty distinct from psychiatry has been created. The most impressive change has been in the proportion of mental disorders being now treated by general practitioners as a result of the availability of psychotropic drugs with fewer side-effects; in France, 60 per cent of antidepressants are now prescribed by general practitioners. These examples may not be a fair representation of the global picture, but there is undoubtedly a movement towards a limitation of the psychiatric specialty to the care of the most severe cases—in practice, the psychotic cases. However, some neuroscientists raise doubts about the usefulness of maintaining psychiatry as a specialty even in this field. Influential biologically oriented psychiatrists have recently proposed on theoretical and practical grounds, that psychiatry should be absorbed into a new medical discipline, akin to the former neuropsychiatry, and all or most of its socio-psychological aspects should be left to non-medical professions.

Since psychiatry has emerged as a specialty, it has been submitted to conflicting forces. The demands of society, changes in the concept of mental disorder and of its limits, variations in the role played by different theoretical perspectives, and successive scientific discoveries have been responsible for an evolution reflected in the professional status and role of the psychiatrist. Displacements of the centre of gravity of a complex structure in which biological, psychological, and social factors interact have modified the image of psychiatry. The threat of being incorporated in other medical specialties or being deprived of its medical character is but another transitory episode in its history.

Further information

Hunter, R. and Macalpine, I. (1963). *Three hundred years of psychiatry 1535–1860*. Oxford University Press, London.

Pichot, P. (1996). *Un siècle de psychiatrie*. Synthelabo, Le Plessis-Robinson.

Postel, J. and Questel, U. (ed.) (1994). *Nouvelle histoire de la psychiatrie*. Dunod, Paris.

Shorter, E.A. (1997). *History of psychiatry: from the era of the asylum to the age of Prozac*. Wiley, New York.

References

1. Kraepelin, E. (1918). Hundert Jahre Psychiatrie. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, **38**, 161–275.
2. Leigh, D. (1961). *The historical development of British psychiatry*. Vol. I, *Eighteenth and nineteenth centuries*. Pergamon Press, Oxford.
3. Goldstein, J. (1987). *Console and classify: the French psychiatric profession in the nineteenth century*. Cambridge University Press.
4. Zilboorg, G. (1941). *A history of medical psychology*. Norton, New York.
5. Kraepelin, E. (1904). *Psychiatrie* (7th edn). Barth, Leipzig.
6. Sabshin, M. (1983). Preface. In *International perspectives on DSM-III* (ed. R.L. Spitzer, J.B.W. Williams, and A.E. Skodol). American Psychiatric Press, Washington, DC.

1.5

Ethics and values

Contents

1.5.1 Psychiatric ethics

Sidney Bloch and Stephen Green

A myriad of ethical problems pervade clinical practice and research in psychiatry. Yet with few exceptions,[1–3] psychiatric ethics has generally been regarded as an addendum to mainstream bioethics. An assumption has been made that 'tools' developed to deal with issues like assisted reproduction or transplant surgery can be used essentially unmodified in psychiatry. These tools certainly help the psychiatrist but the hand-me-down approach has meant that salient features of psychiatric ethics have been prone to misunderstanding. Psychiatric ethics is concerned with the application of moral rules to situations and relationships specific to the field of mental health practice. We will focus on ethical aspects of diagnosis and treatment that challenge psychiatrists, and on codes of ethics. Resolution of ethical dilemmas requires deliberation grounded in a moral theoretical framework that serves clinical decision-making, and we conclude with our preferred theoretical perspective.

Diagnostic issues

Conferring a diagnosis of mental illness on a person has profound ethical sequelae since the process may embody substantive adverse effects, notably stigma, prejudice, and discrimination (e.g. limited job prospects, inequitable insurance coverage). Furthermore, those deemed at risk to harm themselves or others may have their civil rights abridged. These consequences justify Reich's[4] call for the most thorough ethical examination of what he terms the clinician's 'prerogative to diagnose'.

Psychiatrists strive to diagnose by using as objective criteria as possible and information gained from previous clinical encounters. The process is relatively straightforward when findings such as gross memory impairment and life-threatening social withdrawal strongly suggest severe depression. Other situations are not so obvious. For instance, the distress felt by a bereaved person may incline one clinician towards diagnosing clinical depression whereas another may construe the picture as normal grief. Expertise, peer review, and benevolence combine to protect against arbitrariness and idiosyncrasy. Notwithstanding, psychiatrists must, to some extent, apply what might be termed as 'reasoned subjectivism'. Thus, specified criteria in the American Psychiatric Association's (APA) DSM-IV[5] and the World Health Organization's ICD-10[6] do not preclude debate about the preciseness or legitimacy of syndromes like Attention Deficit Hyperactivity Disorder (ADHD) and sexual orientation disturbance. Concern about the intrusion of value judgements into contemporary classification has led to the contention that some diagnoses reflect pejorative labelling rather than scientific decisions. For example, charges of sexism were leveled against DSM-III[7] on the grounds that masculine-based assumptions shaped criteria, resulting in women receiving unwarranted diagnoses like premenstrual dysphoria.[8]

The issue central to this debate is whether certain mental states are grounded in fact or value judgements. Szasz[9] takes a radical position, arguing that disordered thinking and behaviour are due to objective abnormalities of the brain whereas mental illness *per se* is a 'myth', created by society in tandem with the medical profession in order to exert social control. The 'anti-psychiatrist movement'[10,11] posits that mental illnesses are social constructs, reflecting deviations from societal norms. This argument is supported by the role of values in both defining homosexuality in the past as a psychiatric disorder, then reversing that position, in the case of American psychiatry through a ballot among members of the APA in 1973.[12] Legitimate diagnoses necessarily combine aspects of fact and value, as Wakefield[13] avows in his conception of 'harmful dysfunction'. He views 'dysfunction' as a scientific and factual term, based in biology, which refers to the failure of an internal evolutionary mechanism to perform a natural function for which it is designed and 'harmful' as a value-oriented term which covers the consequences of the dysfunction deemed detrimental in socio-cultural terms. Applying this notion to mental functioning, Wakefield describes beneficial effects of natural mechanisms like those mediating

cognition and emotional regulation, and judges their dysfunction harmful when it yields effects disvalued by society (e.g. self-destructive acts). Diagnosable conditions occur when the inability of an internal mechanism to perform its natural function causes harm to the person. DSM-IV[5] rightly emphasizes that mental disorders should not be diagnosed solely by reference to social norms. The deterioration of functioning by which schizophrenia is (partly) defined under Criterion B, or the norm violations of antisocial personality disorder, must therefore be, in DSM-IVs phrase, 'clinically significant'. What this amounts to, then, is that a negative value judgement is insufficient to diagnose. The repercussions of these issues can be considerable, (e.g. exposing children erroneously labelled as ADHD to long-term medication with its attendant risks).[14] A related matter, so-called 'cosmetic psychopharmacology', involves the use of medication to enhance psychological functioning. As Kramer[15] notes, fluoxetine may modulate emotions like anxiety, guilt, and shame, raising ethical questions regarding a person's capacity to possess 'two senses of self'. Psychiatric diagnosis may also mitigate legal and personal results of one's actions (e.g. interpreting excessive sexual activity as a variant of obsessive-compulsive disorder rather than as wilful).

Some of the worst perversions of psychiatry, in which it has been deployed as a form of social control, have been driven by misuse of its diagnostic concepts. In the former Soviet Union, for example, thousands of political, religious, and other dissidents were committed to psychiatric hospitals on the basis of 'delusions of reformism' and other similar tainted concepts.[16]

Treatment issues

Assessing and treating patients require a working alliance in conjunction with informed consent. Many psychiatric patients are in a position to understand and appreciate the nuances of treatment options, to express an informed preference, and to feel allied with a therapist in the task. When the process of informed consent is responsibly handled, particularly with reference to benefits and risks of therapeutic options, mentally ill people are in a comparable position to their counterparts in general medicine. This comparability is grounded in two concepts—competence[17] and voluntarism. [18] The former satisfies the required criterion that the person facing choices in treatment has the 'critical faculties' to appreciate the implications of each course of action. Voluntarism refers to a state in which the process of consent is devoid of any form of coercion. Obviously, given that the organ of decision-making is the same one that is impaired in many psychiatric conditions, profound ethical complications may ensue when seeking informed consent.

Other issues also present themselves in this context; these have been conveniently examined as a series of rights—to treatment, effective treatment, and refusal of treatment—and involuntary treatment.

The right to treatment

The asylum revealed tragically how this right was never actualized; the overcrowded institution became little more than a warehouse.[19] Its custodial nature persisted even after the advent of psychotropics and psychosocial therapies. It took a plaintiff[20] to determine that a person committed involuntarily had the 'right to receive treatment that would offer him a reasonable opportunity to be cured or to improve his mental condition'. Diagnosed with schizophrenia in 1957, Kenneth Donaldson received minimal treatment for the next decade and a half. The US Supreme Court concluded in 1975 that a patient who does not pose a danger to himself or to others and who is not receiving treatment should be released into the community.

The right to effective treatment

The right to treatment has been revisited in subsequent judgements, predominantly in the United States.[21] However, the right has lacked a guarantee that patients will receive effective treatment, reflected vividly in Osheroff v Chestnut Lodge (a private psychiatric hospital in the United States). In this case, the plaintiff sued the staff for their failure to provide antidepressant treatment in the face of his deteriorating depression. Klerman[22] subsequently argued that the clinician is duty-bound to use only' 'treatments for which there is substantial evidence' 'or seek a second opinion in the absence of a clinical response. Stone[23] countered this position which he averred was tantamount to '. . . promulgating more uniformed scientific standards of treatment in psychiatry, based on . . . opinion about science and clinical practice'. Moreover, he posited that legal standards of care should not be established by one 'school' for the whole profession, even if enveloped in science. Instead, we should depend on 'the collective sense' of psychiatry, as well as apply the 'respectable minority rule', namely that a relatively small group within psychiatry can legitimately devise novel therapies.

The right to refuse treatment

As a voluntary patient, Osheroff could have refused treatment of any type as part of informed consent. He pinpointed the institution's alleged failure to offer him an alternative treatment in the face of his worsening state with the therapy that was administered. If principles of informed consent had been applied correctly, his freedom to choose one treatment over others, and to withdraw consent at any stage thereafter, would have prevailed.

The situation differs radically when the patient is committed involuntarily to hospital or community treatment. The right to refuse treatment then looms large.[24] A key event in this context was another US legal judgement when a court ruled that detained patients have a right to refuse treatment.[25] This coincided with changing commitment laws in many jurisdictions from criteria linked to need for treatment to those highlighting the danger posed to oneself and/or others. The ethical repercussions are profound. If psychiatrists are empowered to detain a patient, is it not a contradiction if they are then powerless to provide treatment should the person refuse? The argument rests on the premise that someone disturbed enough to warrant involuntary admission is axiomatically entitled to treatment, and the psychiatrist well placed to give it. Without this arrangement, the psychiatrist's functions are reduced to the custodial.

A countervailing argument is grounded in constitutional rights. Merely because people are committed does not mean they are incapable of participating in the process of informed consent. In the event they cannot appreciate the rationale for a course of action, a form of substituted judgement should be employed thereby ensuring that their rights remain prominent.

An assortment of legal remedies has emerged in response to this ethical quandary, ranging from a full adversarial process to reliance

on a guardian. Appelbaum[24] has contributed a lucid account of available options and a predilection for a treatment-driven model in which patients are committed because their capacity to decide about their medical care is lacking as part of their disturbed mental functioning. His own research demonstrates that most 'refusing' patients voluntarily accept treatment within 24 h.[26]

In another pragmatically oriented account Stone[27] proposes that presumption of competence should be dealt with before admission to hospital. Dealing with commitment and competence together obviates the problem of compulsory admission without the powers to treat. The snag is the fluidity of the mental state. What patients think about treatment during the maelstrom of being detained may well change once they settle in and are suitably cared for.

Involuntary treatment

A consensus has prevailed universally that a proportion of psychiatric patients lack the capacity for self-determination. They are prone to harming themselves and/or others, acting in ways they will later regret (e.g. a manic patient's sexual indiscretions); and suffer from self-neglect (e.g. malnourished and physically ill schizophrenic patients). What is not universally agreed is how best to deal with such vulnerable people. Society has, generally, devised laws as the vehicle to respond to the thorny issue of how to protect this group. However, variations in legislation and its application are legion, reflecting, in part, the ethical underpinnings of the process. Psychiatrists and society need to establish coherent arguments concerning relevant moral principles. A good start is Mill's[28] contention that the 'only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant'. Mill's caveat that an exception must be made in the case of children and mentally disturbed people (i.e. 'delirious' or in a 'state of excitement or absorption incompatible with the full use of the reflecting faculty') suggests they can legitimately be assisted.

Chodoff[29] has addressed the awesome question of compulsorily treating a person on the grounds of mental illness. He finds classical moral theory wanting and therefore proposes a 'chastened and self-critical' paternalism, one 'willing to commit to strong safeguards against abuse'. This humanism is epitomized in a concluding sentiment: involuntary treatment is not a conflict of right versus wrong but one over the right to remain at liberty against the right 'to be free from dehumanizing disease'.

Our account hitherto has referred to patients as a homogeneous group. Loss of critical faculties may be a unifying feature but ethical factors will vary according to particulars of the clinical state. One noteworthy example is suicidal behaviour. Szasz[30] sees suicide as the act of a moral agent. The State should therefore not assume power to prevent self-killing although it may opt to advise for or against. This argument is a libertarian one, with the corollary that everyone should have the right to end their life. Szasz has, however, neglected Mill's[28] point that when respecting a person's right to liberty, a possible exception is loss of critical faculties. This is not to aver that all suicidal behaviour is the product of a disordered mind. Suicide in the wake of debilitating illness and a long-standing commitment to euthanasia, seems rational—for example, the British author, Arthur Koestler, left a suicide note demonstrating that he arrived at his decision authentically and competently in terms of psychological function.[31]

The suicidal patient epitomizes the psychiatrist's dilemma in having no choice but to impose treatment in various circumstances and having to declare a person's incapacity, by dint of mental illness, to make rational judgements about what is in their best interests.[32] Van Staden and Kruger[33] cover this topic by highlighting its dimensions, namely the failure to understand relevant information, choose decisively between options, and accept that the need for treatment prevails. They refer to the utility of a 'functional approach' in determining capacity, especially the temporal factor, so that a patient incapable of consenting at one point in their illness may become able at another. Ethical arguments to justify detention in a hospital can be extrapolated into the community setting. Similar restrictions on liberty lie at the heart of the moral dilemma and the psychiatrist again has to consider patients' competence. Munetz and his colleagues[34] apply three ethical arguments in a compelling fashion—utilitarian, communitarian, and beneficence—concluding that all three support the use of compulsory treatment in the community setting. Advance directives are a means to obviate some of the ethical complications of compulsory treatment. In summary, patients prone to recurrence of their illness during which they may be too disturbed to provide informed consent reach an agreement with their psychiatrist about what constitutes the best course of action should they suffer an episode in the future and be unable to decide appropriately what treatment is in their best interest. Given that several mental illnesses have a recurring course and are associated with incapacity, advanced directives would appear, on the face of it, to have a useful role. Empirical studies to examine this potential are a clear option; promising results have been achieved in work carried out hitherto.[35]

Codes of ethics

The development of codes of ethics in the history of medicine reflects their possible role in promoting sound clinical practice (and research). Some codes have been a direct response to the collapse of professional standards—the Nuremberg Statement, for instance, was formulated in the aftermath of the Nazi medical crimes—but they obviously also have positive functions.

Promoting professional cohesion is one such function. Despite George Bernard Shaw's depiction of professions as 'conspiracies against the laity' and the risk that professional codes may indeed be self-serving—a charter for restrictive practices, protectionist rather than protective—a profession can only function effectively if it is cohesive and acts in a collegial way. Thus, a code which sets out its members' obligations to one another can contribute substantially to achieving this goal. Most codes have emphasized medicine's tradition of commitment and dedication to society; some even call for 'whistle blowing' in appropriate circumstances.

A second function of codes is to enhance high standards of practice. Professions are characterized, in part, by a corpus of specialized knowledge and skills, not readily available to others, and offered to a dependent, often vulnerable clientele.[36] To the extent, therefore, that it takes an expert to judge relevant expertise, a degree of self-regulation is essential. But this must be balanced by external monitoring. Where bad practice becomes the norm, self-regulation may reinforce it: the abuse of psychiatry to suppress dissent in the former Soviet Union was, in effect, promoted by leaders of the profession.[16] It can be argued that a prescriptive ethical code is

unnecessary, since implicit discipline and a shared ethos suffice to maintain standards. As we have noted, history contradicts this, with codes often appearing as a response to compromised care. What kind of codes best promote sound practice will vary with circumstances. They have therefore differed widely in form and content, ranging from aspirational principles to practice guidelines which are set out in considerable detail. The latter are pertinent, especially for education and training. The code of the Royal Australian and New Zealand College of Psychiatrists[37] and the American Psychiatric Association[38] combine general principles with a series of annotations on specific areas such as confidentiality, respect for professional boundaries, and informed consent. Codes also vary in ethical focus, some are virtue-driven, emphasizing character traits which support best possible practice whereas others are duty-based in that they lay out specific responsibilities and obligations.

Codes thus have several meritorious purposes. Furthermore, these support one other. For instance, their inherent educational quality serves to enhance sound ethical clinical practice and their stipulation of potential hazards in ethical decision-making may prevent compromised care, even misuse of expert knowledge.

Conclusion

We have readily noted how the psychiatrist faces ethical quandaries at several levels, both diagnostically and therapeutically, and the potential role of codes of ethics to grapple with them. Diverse moral theories have also been promulgated to aid the practitioner[1,2] and we conclude with our own preferred approach, a combination of principlism[39] and care ethics.[40] Principlism (or principle-based ethics) relies on a set of well-recognized moral principles to identify and analyse ethical problems: respect for autonomy (literally self-government), non-maleficence (first of all, do no harm), beneficence (acting in peoples' best interests), and justice (treating people fairly). The essence of care ethics revolves around the 'natural' propensity of health professionals to extend care to dependent, vulnerable people and to react with such 'moral' feelings as compassion, sensitivity, and trustworthiness.[40] The approach fits well with psychiatry since its practitioners depend day in, day out, on empathy in order to understand the interests and needs of patients and their families. A synthesis of care ethics and principlism permits sound moral reflection within an emotionally based environment in which connectedness between patient and therapist is paramount. We believe this approach, a complementarity between feeling and reason, acknowledges and best exploits the role of 'moral emotions'[41] when clinicians are presented with the many, nuanced ethical conundrums of psychiatric practice.

Further information

Bloch, S. and Green, S. (eds.) (1999). *Psychiatric ethics* (3rd edn). Oxford University Press, Oxford.

Green, S. and Bloch, S. (2006). *An anthology of psychiatric ethics*. Oxford University Press, Oxford.

Post, S. (ed.) (2004). *Encyclopedia of bioethics* (3rd edn). Macmillan, New York.

References

1. Green, S. and Bloch, S. (2006). *An anthology of psychiatric ethics*. Oxford University Press, Oxford.
2. Bloch, S., Chodoff, P., and Green, S. (1999). *Psychiatric ethics* (3rd edn). Oxford University Press, Oxford.
3. Dickenson, D. and Fulford, W. (2001). *In two minds: a casebook of psychiatric ethics*. Oxford University Press, Oxford.
4. Reich, W. (1999). Psychiatric diagnosis as an ethical problem. In *Psychiatric ethics* (ed. S. Bloch, P. Chodoff, and S. Green) (3rd edn). Oxford University Press, Oxford, pp. 193–224.
5. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). APA, Washington, DC.
6. World Health Organization. (1992). *International statistical classification of diseases and related health problems*, 1989 revision. WHO, Geneva.
7. Kaplan, M. (1983). A woman's view of DSM-III. *American Psychologist* **38**, 786–92.
8. Chesler, P. (1972). *Women and madness*. Avon, New York.
9. Szasz, T. (1960). The myth of mental illness. *American Psychologist*, **15**, 113-8.
10. Scheff, T. (1966). *Being mentally ill: a sociological theory*. Aldine, Chicago.
11. Laing, R. (1969).*The divided self*. Pantheon, New York.
12. Stoller, R., Marmor, J., Bieber, I., *et al.* (1973). A symposium: should homosexuality be in the APA nomenclature. *American Journal of Psychiatry*, **130**,1207–16.
13. Wakefield, J. (1992). The concept of mental disorder: on the boundary between biological facts and social values. *American Psychologist*, **47**, 373–88.
14. Halasz, G. (2002). A symposium of attention deficit hyperactivity disorder (ADHD): an ethical perspective. *Australian and New Zealand Journal of Psychiatry*, **36**, 472–5.
15. Kramer, P. (1993). *Listening to Prozac*. Viking, New York.
16. Bloch, S. and Reddaway, P. (1977). *Russia's political hospitals*. Gollancz, London (Published in the US as *Psychiatric Terror*, New York: Basic Books).
17. Lidz, C.W., Meisel, A., Zerubavel, E., *et al.* (1984). *Informed consent: a study of decision making in psychiatry*. Guilford, New York.
18. Roberts, L. (2002). Informed consent and the capacity for voluntarism. *American Journal of Psychiatry*, **159**, 705–12.
19. Bloch, S. and Pargiter, R. (2002). A history of psychiatric ethics. *Psychiatric Clinics of North America*, **25**, 509–24.
20. Donaldson v O'Connor, F. 2d 5th Cir, decided Apr. 26, 1974; O'Connor v Donaldson 422US. **563**, 1975.
21. Wyatt v Stickney 325F Supp 781 (1971); Wyatt v Stickney 344F Supp 373, 376, 379–85 (1972).
22. Klerman, G. (1990). The psychiatric patient's right to effective treatment: implications of Osheroff v Chestnut Lodge. *American Journal of Psychiatry*, **147**, 419–27.
23. Stone, A. (1990). Law, science, and psychiatric malpractice: a response to Klerman's indictment on psychoanalytic psychiatry. *American Journal of Psychiatry*, **147**, 419–27.
24. Appelbaum, P. (1988). The right to refuse treatment with antipsychotic medications: retrospect and prospect. *American Journal of Psychiatry*, **145**, 413–9.
25. Rogers v Okin, 478F Supp 1342 (D Mass, 1979); Rogers v Commissioner of the Department of Mental Health, 458NE 2d 308 (Mass Sup Jud Ct, 1983).
26. Appelbaum, P. and Gutheil, T. (1979). 'Rotting with their rights on': constitutional theory and clinical reality in drug refusal by psychiatric patients. *Bulletin of the American Academy of Psychiatry and the Law*, **7**, 308–17.
27. Stone, A. (1981). The right to refuse treatment: why psychiatrists should and can make it work. *Archives of General Psychiatry*, **38**, 358–62.
28. Mill, J.S. (1976). On liberty. In *Three essays*. Oxford University Press, Oxford.

29. Chodoff, P. (1984). Involuntary hospitalisation of the mentally ill as a moral issue. *American Journal of Psychiatry*, **141**, 384–9.
30. Szasz, T. (1986). The case against suicide prevention. *American Psychologist*, **41**, 806–12.
31. Cesarani, D. (1998). *Arthur Koestler: the homeless mind*. Heinemann, London.
32. Heyd, D. and Bloch, S. (1998). The ethics of suicide. In: *Psychiatric ethics* (ed. S. Bloch, P. Chodoff, and S. Green) (3rd edn). Oxford University Press, Oxford, pp. 441–60.
33. Van Staden, C. and Kruger, C. (2003). Incapacity to give informed consent owing to mental disorder. *Journal of Medical Ethics*, **29**, 41–3.
34. Munetz, M., Galon, P., and Frese, F. (2003). The ethics of mandatory community treatment. *Journal of the American Academy of Psychiatry and the Law*, **31**,173–83.
35. Green, S. and Bloch, S. (2006). *An anthology of psychiatric ethics*. Oxford University Press, Oxford. pp. 183–84.
36. Fullinwinder, R. (1994). *Codes of ethics and the professions* (ed. M. Coady, S. Bloch). Melbourne University Press, Melbourne, pp. 72–87.
37. Royal Australian and New Zealand College of Psychiatrists. (2001). *Code of ethics* (3rd edn). Melbourne.
38. American Psychiatric Association. (2001). *Principles of medical ethics with annotations especially applicable to psychiatry* (5th edn). APA, Washington, DC.
39. Beauchamp, T. and Childress, J. (2001). *Principles of biomedical ethics* (5th edn). Oxford University Press, New York.
40. Baier, A. (2004). Demoralization, trust and the virtues. In: *Setting the moral compass* (ed. C. Calhoun). Oxford University Press, New York, pp. 176–88.
41. Hume, D. (1969). *A treatise of human nature*. Penguin, Harmondsworth.

1.5.2 Values and values-based practice in clinical psychiatry

K. W. M. Fulford

Introduction

Values-based practice is a new skills-based approach to working more effectively with complex and conflicting values in health and social care. This chapter illustrates some of the ways in which combining values-based with evidence-based approaches supports the day-to-day practice of the clinical psychiatrist, particularly in the context of multidisciplinary teamwork.

What are values?

Perhaps one of the most familiar ways in which values impact medicine is by way of ethics. But values are wider than ethics. Ethical values are indeed one kind of value. But there are many other kinds of values, such as aesthetic and prudential values. Values also extend to needs, wishes, preferences, indeed to any and all of the many different ways in which we express negative or positive evaluations and value judgments. Within each and all of these areas, moreover, there are wide differences in the particular values held by different individuals, by different cultures, and at different historical periods.

Given the breadth and complexity of values, it is small wonder that the term 'values' means different things to different people. This is illustrated by Table 1.5.2.1 which lists the responses of a group of trainee psychiatrists when asked, at the start of a training session on values-based practice, to write down three words or short phrases that they associate with 'values'. As the table shows, although there is some overlap, every member of the group came up with a different set of associations.

If our values are diverse, however, they are not completely idiosyncratic. To the contrary, there are many values that are widely shared, at least within a given group at a particular period. The values of patient autonomy (freedom of choice) and of acting in the patient's best interests, for example, are shared values that underpin contemporary medical ethics, and these two values are indeed among the values evident in Table 1.5.2.1.

It is the diversity of human values, and how this can be linked with the shared ethical values underpinning clinical practice, that is the starting point for values-based practice. There is a sense in which medicine has always been values-based just as there is a sense in which it has always been evidence-based.(1) The need for values-based practice in contemporary practice, again like the need for evidence-based practice, arises from the growing complexity of medicine; the growing complexity of the evidence underpinning medicine has led to the need for the new tools of evidence-based medicine; the growing complexity of the values underpinning medicine has led to the need for the new tools of values-based medicine.

The growing complexity of values, as well as evidence, is particularly evident in psychiatry. 'Autonomy' and 'best interests', for example, although both shared ethical values, are often in tension. In the past, most people were content to allow doctors to decide what is in their best interests and this is still the case in many parts of the world.(2) Increasingly though, at least in Europe and North

Table 1.5.2.1 What are values?

Faith Internalization Acting in best interests	How we treat people Attitudes Principles Autonomy
Integrity Conscience Best interests Autonomy	Love Relationships
Respect Personal Difference . . . diversity	Non-violence Compassion Dialogue
Beliefs Right/wrong to me What I am	Responsibility Accountability Best interests
Belief Principles Things held dear	What I *believe* What makes me tick What I won't compromise
Subjective merits Meanings Person-centred care	'Objective' core Confidentiality Autonomy
A *standard* for the way I conduct *myself* Belief about how things *should* be Things you would not want to change	Significant Standards Truth

America, a growing emphasis on patient autonomy has led to complex interactions between these two values in clinical care. In particular, autonomy and best interests come into direct conflict in relation to issues of compulsory treatment (Chapter 1.5.1). Then again, even considered in isolation, 'best interests' have highly complex applications in practice, in the sense that what is 'best' for one person may be very different from what is 'best' for another, according to differences in their personal values and the values of others concerned. Establishing 'best interests' thus presents particular challenges in areas such as old age psychiatry, for example, where patients may lack the decision-making capacity to exercise genuine autonomy on their own behalf.

One response to the growing complexity of the values bearing on clinical practice is to write ever more detailed rules aimed at fixing in advance the 'right outcomes' for any given clinical situation. It is this response that is driving the growing volume of ethical codes and regulatory bodies concerned with medicine. Values-based practice offers quite a different albeit complementary response. It switches the focus from pre-set *right outcomes* to a reliance on *good process*. Values-based practice, that is to say, focuses not so much on *what* is done but on *how* it is done. Starting from the 'democratic' ethical premise of respect for differences of values, values-based practice relies on good process (in particular good clinical skills, see below) to support balanced decision-making within the framework of shared values defined by codes of ethical practice.[3]

Values-based and evidence-based medicine

As a process-based approach to clinical decision-making, values-based practice is complementary not only to regulatory ethics but also to evidence-based practice. The processes of values-based practice and of evidence-based practice are of course very different. Evidence-based practice, as John Geddes describes (Chapter 1.10), relies on statistical and other methods for combining evidence from methodologically sound research. Values-based practice, by contrast, relies primarily on learnable clinical skills. There are other components of the process of values-based practice, including a number of specific links between values-based and evidence-based practice.[3] But at the heart of values-based practice are four areas of clinical skill. As set out more fully in Table 1.5.2.2, these are, raised awareness of values and of differences of values, reasoning about values, knowledge of values, and communication skills.

The close interdependence of values-based and evidence-based approaches has been well recognized by many of those involved in the development of evidence-based medicine. Indeed, there is perhaps no clearer statement of this inter-dependence than the very definition of evidence-based medicine given by David Sackett and his colleagues in their book, *Evidence-Based Medicine: How to Practice and Teach EBM*.[1] Evidence-based medicine is standardly thought to be concerned only with research evidence, as outlined above. To the contrary, Sackett *et al.* say (p. 1), evidence-based medicine combines three distinct elements. The first element is, certainly, best research evidence. In clinical practice, however, best research evidence has to be combined with the experience and skills of practitioners, and, crucially, with patients' values. 'By patients' values', Sackett *et al.* continue, 'we mean the unique preferences, concerns and expectations each patient brings to a clinical encounter and which must be integrated into clinical decisions if they are to serve the patient'. Furthermore, they conclude, it is only 'when these three elements (best research 'evidence, clinicians' experience and patients' values) are integrated, (that) clinicians and patients form a diagnostic and therapeutic alliance which optimizes clinical outcomes and quality of life.'

Table 1.5.2.2 The four key skills areas underpinning values-based practice

Skills area	Applications in values-based practice
1. Raising Awareness of Values	Values, our own and those of others, are often implicit: thus a first step towards balanced decision-making is to raise awareness, 1) of values as such, 2) of differences of values, (See text)
2. Reasoning about Values	In ethics and law, various methods of reasoning are used to derive ethical conclusions. In values-based practice, the same range of methods is used but primarily to explore and open up the range of values bearing on a given situation. These include principles, casuistry (case-based reasoning), utilitarianism (balancing utilities, used especially in health economics), and deontology (rule-based reasoning, used especially in law). (See Further information)
3. Knowledge of Values	Values-based practice draws on evidence about values derived from, 1) the full range of empirical methods (including qualitative social science methods), 2) a range of philosophical methods, 3) combined methods (see text, also Further Reading).
4. Communication Skills	In values-based practice, communication skills are central to, 1) eliciting and understanding individual values, 2) resolving conflicts of values, for example by negotiation and conflict resolution (see text).

Values in the multidisciplinary team

In many parts of the world, psychiatric services are increasingly delivered through multidisciplinary and multi agency teams (Chapter 1.8.1). The move to multidisciplinary team-working reflects a broadly evidence-based recognition that different professional groups offer different but complementary resources of knowledge and skills. It was realized early on, however, that differences of perspective (which include different value perspectives) between different professional groups may lead to communication and other problems of effective team-working.[4,5] This is where values-based practice can help to support the leadership role of the clinical psychiatrist in the multidisciplinary team. To anticipate a little, values-based practice, as we will see in this section and the next, 1) helps to make differences of perspective between team members more transparent, thus improving communication and shared decision-making; and 2) converts these differences of perspective between team members from a barrier into a positive resource for decision-making that is sensitive to the particular and

often very different values—the needs, wishes, preferences, etc.—of individual patients and their families.

First, then, what are the differences of perspective between different team members? The perhaps surprising extent of these differences is illustrated by Table 1.5.2.3. This is based on a study, led by the British social scientist, Anthony Colombo, of multidisciplinary teams in the UK concerned with the community care of people with long-term schizophrenia.[6] To understand the significance of Table 1.5.2.3, we need to look briefly at the background to Colombo's study and how it was carried out.

Colombo was interested in implicit models of disorder and how such models might influence the processes of decision-making in day-to-day clinical care within multidisciplinary teams. Asked directly, most team members, from whatever professional background, will indicate that they share much the same broadly biopsychosocial model of schizophrenia. This is their shared *explicit* model, then. The hypothesis guiding Colombo's study, however, was that, notwithstanding their explicit commitment to a shared biopsychosocial model, in actual practice different team members would be guided by different *implicit* models. These different implicit models reflected different weightings or priorities (hence values), in turn reflecting differences of professional background and training, that different team members might attach to the different aspects of a given case. Their different implicit models, furthermore, just in being implicit rather than explicit, could help explain the difficulties of communication and other problems of shared decision-making within multidisciplinary teams that had been identified in the literature (as above).

The aim of Colombo's study, therefore, was to access the *implicit* models (including values) guiding different professional groups in their responses to patients with schizophrenia. Colombo's method, correspondingly, was indirect rather than direct. He presented subjects with a standardized case vignette, of a man called 'Tom', with features of schizophrenia (though without using that term as such), and then explored their responses using a semi-structured interview and carefully validated scoring system. In previous work, Colombo had shown how different models of disorder (six of which are represented by the columns in Table 1.5.2.3) could be analysed

Table 1.5.2.3 Comparison of models grids for psychiatrists and social workers (shared elements of models shown highlighted)

	Models - Psychiatrists					
Elements	**Medical (Organic)**	**Social stress**	**Cognitive behaviour**	**Psychotherapeutic**	**Family (interaction)**	**Political**
1. Diagnosis/Description	P					
2. Interpretation of Behaviour	P					
3. Labels	P					
4. Aetiology	P					
5. Treatment	P					
6. Function of the Hospital	**P**	**P**				**P**
7. Hospitality & Community	P					
8. Prognosis	P					
9. Rights of the Patient	**P**					
10. Rights of Society	**P**					
11. Duties of the Patient	P		**P**			
12. Duties of Society	P					
	Models - Social Workers					
1. Diagnosis/Description				S		
2. Interpretation of Behaviour				S		
3. Labels				S		
4. Etiology				S		
5. Treatment		S			S	
6. Function of the hospital	**S**	**S**				**S**
7. Hospitality & Community		S		S		
8. Prognosis				S		
9. Rights of the Patient	**S**	S				S
10. Rights of Society	**S**					
11. Duties of the Patient			**S**			
12. Duties of Society		S				

and compared along 12 key dimensions (diagnosis, causal factors, etc.) as represented by the lines in Table 1.5.2.3. Responses to the semi-structured interview thus allowed a profile to be developed for each subject, and cumulatively for each professional group, of their implicit models. These profiles, or 'models grids', gave an overall picture of the implicit model on which an individual or group was drawing in their responses to Tom.

It is the 'models grids' for psychiatrists and psychiatric social workers respectively that are compared in Table 1.5.2.3. As this shows, notwithstanding their shared explicit commitment to a biopsychosocial model, psychiatrists and social workers, working in similar teams in the same area of the UK, had widely different *implicit* models. Direct comparison of the two models grids shows that psychiatrists and approved social workers coincided on only six out of a total possible of 72 elements! Small wonder, then, given that team members were unaware of these differences of implicit models, that difficulties of communication and of shared decision-making often arose. They all accepted an explicitly biopsychosocial approach. But their different professional perspectives, including value perspectives, led them to attach very different priorities to different aspects of 'Tom', and, by extension, to the real patients with whom they were concerned in the real world of day-to-day multidisciplinary care.

Values-based practice in the multidisciplinary team

Colombo's study illustrates how careful research may improve our knowledge of values (skills area 3, Table 1.5.2.2 above) and his research has subsequently been adapted and developed as part of a training manual for values-based practice.[7] Research is as important in values-based practice as in any other aspect of medicine. There is a widespread assumption that understanding each other's values is a matter of relatively transparent intuition. But in addition to empirical studies, surveys,[8] patient narratives,[9] and other sources all point to the extent to which our perceptions of each other's values are often *mis*perceptions. Colombo's study also illustrates the inter-dependence of the four skills areas. As a contribution to skills area 3 (knowledge), the study also contributes to raising awareness of values and of differences of values (skills area 1) in the specific context of multidisciplinary team working.

Merely raising awareness of differences of implicit models may itself be enough to improve communication between team members and hence, shared decision-making as the basis of effective multidisciplinary care. There are circumstances, however, in which raising awareness may not be sufficient. Where team values are directly conflicting, for example, raising awareness may even have the effect of accentuating rather than reducing difficulties of shared decision-making. So, how should differences of values be managed? Different responses are possible here. One approach is to try to create a homogenized composite model. At a relatively abstract level, this is what the biopsychosocial model offers. A different response is to seek to establish a 'top' model that takes priority over all other models. It is a natural enough assumption of any given professional group that their own particular model should be the 'top' model.

Rather than either a composite model or a 'top' model, values-based practice suggests that differences of value perspective, as incorporated into implicit models of disorder, far from being suppressed should be acknowledged and built on as a positive resource for effective multidisciplinary teamwork. This is essentially because, as the models grids in Table 1.5.2.4 show, differences in implicit models between different team members reflect corresponding differences among *patients themselves*.

The models grids in Table 1.5.2.4 were derived in Colombo's study using precisely the same methods as the models grids for the professional groups (i.e., using the same standardized case vignette, interview schedule, etc.). The patients involved in Colombo's study, however, were not recruited through multidisciplinary teams. Rather, they were recruited as volunteers from local MIND, a mental health NGO in the UK that 'advocates for patients' rights. To volunteer for the study, a patient had to have had a diagnosis of schizophrenia for at least three years. There was no requirement that they should agree with the diagnosis. Rather, the aim was to explore the implicit models of a group of subjects who had had this diagnosis 'willing or no' for an extended period. The expectation in the study was that this group of patients, recruited in this way, would include a significant number with a 'political', or 'anti-psychiatric', model of disorder, represented by the right-hand column of the models grids. In fact, as Table 1.5.2.4 shows, the patient group divided naturally into two sub-groups, one with implicit models very close to those of the psychiatrists in the study, the other with implicit models very close to those of the social workers.

The correlation in Colombo's study between different professional models and different patient models, gives a whole new values-based rationale for multidisciplinary teamwork. From an evidence-based perspective, multidisciplinary team working offers a diversity of knowledge and skills in meeting patients' needs. From a values-based perspective, multidisciplinary team working brings, in addition to different knowledge and skills, different value perspectives. In a well-functioning multidisciplinary team, these different value perspectives can help to ensure that professionals' knowledge and skills are matched appropriately to the different values—the needs, wishes, preferences, etc.—of individual patients and their families.

In helping to bring together the different perspectives of team members in a positive and well-balanced way, values-based practice thus, directly supports the leadership role of the consultant psychiatrist in the multidisciplinary team. Again, all four skills areas of values-based practice are closely interdependent here. In addition to knowledge of values, the reasoning skills noted in Table 1.5.2.2 may be helpful. Good communication skills are also crucial. As Table 1.5.2.2 indicates, these include in particular, skills for eliciting and understanding values, and where values conflict, skills of negotiation and conflict resolution.

Conclusions

This chapter has introduced a number of key points about values-based practice (summarized in Table 1.5.2.5) as a new skills-based approach to working more effectively with complex and conflicting values. The importance of values-based as well as evidence-based approaches has been illustrated particularly by reference to the leadership role of the consultant psychiatrist in the multidisciplinary team. Multidisciplinary team-working is of course not unique in its requirement for values-based as well as evidence-based approaches. As Sackett *et al.*[1] reminded us at the start of this chapter, it is only by combining best research evidence with practitioners' knowledge and skills and with patients' values, that we can

Table 1.5.2.4 Comparison of models grids for two groups of Patients—Group 1 similar to Medical Psychiatrists (Pt-Med), Group 2 similar to Social Workers (Pt-SW) (See also Table 1.5.2.3)

	Models - Group 1 (similar to psychiatrists)					
Elements	**Medical (Organic)**	**Social stress**	**Cognitive behaviour**	**Psycho-therapeutic**	**Family (interaction)**	**Political**
1. Diagnosis/Description	Pt-Med					
2. Interpretation of Behaviour			Pt-Med	Pt-Med		
3. Labels	Pt-Med					
4. Aetiology	Pt-Med					
5. Treatment	Pt-Med					
6. Function of the hospital	Pt-Med	Pt-Med				Pt-Med
7. Hospitality & Community		Pt-Med		Pt-Med		
8. Prognosis	Pt-Med					
9. Rights of the Patient	Pt-Med					Pt-Med
10. Rights of Society		Pt-Med				
11. Duties of the Patient		Pt-Med	Pt-Med			
12. Duties of Society	Pt-Med		Pt-Med			Pt-Med
	Models - Group 2 (similar to social workers)					
1. Diagnosis/Description				Pt-SW		
2. Interpretation of Behaviour				Pt-SW		
3. Labels			Pt-SW	Pt-SW		
4. Aetiology				Pt-SW		
5. Treatment		Pt-SW		Pt-SW		
6. Function of the hospital	Pt-SW					Pt-SW
7. Hospitality & Community		Pt-SW		Pt-SW		
8. Prognosis				Pt-SW		
9. Rights of the Patient	Pt-SW	Pt-SW				Pt-SW
10. Rights of Society		Pt-SW				
11. Duties of the Patient		Pt-SW	Pt-SW			
12. Duties of Society			Pt-SW			Pt-SW

build the 'diagnostic and therapeutic alliance' on which in any area of medicine, good clinical care crucially depends.

Two further points are worth adding in conclusion. The first is that values-based practice as introduced here, as being based primarily on learnable clinical skills, is only one among a number of new disciplinary resources supporting more effective ways of working with complex and conflicting values in healthcare. Further and quite different resources are provided by such empirical disciplines as decision theory,[10] for example, and health economics, an innovative use of which has recently been developed by a group at the Centre for Value-Based Medicine in the States.[11] Values-based practice itself is so underpinned by a branch of analytic philosophy, called philosophical value theory (see Further Reading, below) that, as an analytic discipline, it is a natural partner both of empirical research, as in Colombo's study,[12] and of other philosophical disciplines more familiar in psychiatry, notably phenomenology.[13] As Andreasen[14] has argued, these and other philosophical disciplines, have a growing practical importance not only in clinical psychiatry, as illustrated in this chapter, but also in the new neurosciences.

The second concluding point has to do with the place of psychiatry as a science-led medical discipline in the 21st century. As the most value-laden area of medicine, psychiatry was widely stigmatized in the 20th century as being, at best, scientifically underdeveloped,[15] at worst outwith medicine altogether;[16] and debates about eliminate values from psychiatric diagnostic concepts continue in the context of current revisions of the ICD and DSM classifications.[17,18] Philosophical value theory, as the discipline underpinning values-based practice, shows to the contrary that the value-ladenness of psychiatry, far from reflecting any deficiency in its science, is a direct consequence of the fact that psychiatry, in being concerned with the higher functions of emotion, desire, volition, sexuality, and so forth, is by the same token concerned with areas of experience and behaviour where human value themselves are highly diverse ([19], especially chapters 4 and 5, building on[20]).

Table 1.5.2.5 Key points about values-based practice for clinical psychiatry

1. Values are wider than ethics and include all the many ways in which we express positive and negative evaluations, i.e., preferences, needs, wishes, etc. as well as ethical values
2. Values-Based Practice is a new skills-based approach to working with complex and conflicting values in medicine
3. Ethical principles provide a framework of shared values – such as 'best interests' and 'autonomy of patient choice' - that guide the *outcomes* of clinical decision-making supported by codes of practice and regulatory bodies
4. Values-based practice is complementary to regulatory ethics in focusing on the *process* of clinical decision-making: ethics focuses on 'right outcomes' (reflecting shared values); values-based practice focuses on 'good process' (reflecting complex/conflicting values)
5. In focusing on process rather than outcomes, values-based practice (concerned with complex and conflicting values) is fully complementary to evidence-based practice (concerned with complex and conflicting evidence) in clinical decision-making
6. At the heart of the 'good process' on which values-based practice depends are four key areas of clinical skills: 1) awareness of values and of diversity of values, 2) reasoning about values, 3) knowledge of values, and 4) communication skills (including skills in such areas as negotiation and conflict resolution)
7. Among other applications, values-based practice supports the role of the consultant psychiatrist in multidisciplinary teams by, 1) improving understanding of differences of values between different team members (thus improving communication and shared decision-making), and 2) improving understanding of the particular and often very different values of individual patients and carers (thus improving the extent to which care and treatment are appropriately matched to the particular needs, preferences and wishes of each individual patient and their families)
8. Values-based practice is research-based, drawing in particular on the resources of philosophical value theory and a number of other areas of the philosophy of psychiatry, such as phenomenology, in addition to empirical social science methods, patient narratives and other sources
9. In clinical work, values-based practice is supported by a wide range of training materials and is the basis of a number of both national and international developments in psychiatry aimed at building a strong diagnostic and therapeutic alliance between professionals and patients in mental health and social care
10. The skills-based approach of values-based practice is one of a number of disciplinary resources for working with complex and conflicting values in medicine: in addition to ethics, other important disciplines include health economics and decision theory

Psychiatry, therefore, in addition to being *scientifically* complex, is shown by philosophical value theory to be *evaluatively* complex. This is why it has been appropriate that values-based practice, as a skills-based approach to working with complex and conflicting values, should have developed first in psychiatry. But with the growing complexity of the values bearing on all areas of medicine, it seems likely that, as Sackett *et al.* anticipated (as above), values-based as well as evidence-based approaches will become increasingly crucial in the 21st century not only in psychiatry but across the board. The effect of this will be to reverse the 20th century stigmatization of psychiatry. Instead of being perceived as a scientific also-ran, psychiatry, by being first in the field with values-based practice as an essential partner to evidence-based practice, will be seen to have led the way with a 21st century model of medicine that is not only fully science-based but also genuinely patient-centred.

Acknowledgments

The information on which Tables 1.5.2.3 and 1.5.2.4 are based is derived from the study of models of disorder described in the text and first published in Colombo *et al.* 2003. (6)

Further information

1) Skills training for values-based practice

Woodbridge, K., and Fulford, K.W.M. (2004). *Whose Values? A workbook for values-based practice in mental health care*. London: Sainsbury Centre for Mental Health. (see website www.scmh.org.uk); also by mail from The Sainsbury Centre for Mental Health, 134–8, Borough High Street, London SE1 1LB

2) The theory of values-based practice

Fulford, K.W.M. (1989, reprinted 1995 and 1999). *Moral Theory and Medical Practice*. Cambridge: Cambridge University Press.

Fulford, K.W.M. (2004). Ten Principles of Values-Based Medicine. Ch 14 In *The Philosophy of Psychiatry: A Companion* (ed. J. Radden), pp. 205–34. New York: Oxford University Press. (Describes the links between values-based and evidence-based approaches in clinical decision-making).

3) Values-based practice and the new philosophy of psychiatry

Fulford, K.W.M., Thornton, T., and Graham, G. (2006). *The Oxford Textbook of Philosophy and Psychiatry*. Oxford: Oxford University Press. (see especially chapter 18, setting values-based practice in context with more familiar approaches to ethics; and chapter 21, showing how values-based alongside evidence-based approaches are important in diagnostic assessment as well as in treatment and care planning).

References

1. Sackett, D.L., Straus, S.E., Scott Richardson, W., *et al.* (2000). *Evidence-Based Medicine: How to Practice and Teach EBM* (2nd edn.). Edinburgh and London: Churchill Livingstone.
2. Okasha, A. (2000). Ethics of Psychiatric Practice: Consent, Compulsion and Confidentiality. *Curr Op in Psychiatry*, **13**, 693–8.
3. Fulford, K.W.M. (2004). Ten Principles of Values-Based Medicine. Ch 14 In *The Philosophy of Psychiatry: A Companion* (ed. J. Radden), pp. 205–34. New York: Oxford University Press.
4. Clunis Report: North East Thames and South East Thames Regional Health Authorities. (1994). *The Report of The Enquiry into the Care tnd Treatment of Christopher Clunis.* London: HMSO.
5. Department of Health. (1999). *Still Building Bridges: The Report of a National Inspection of Arrangements for the Integration of Care Programme Approach with Care Management*. London: Department of Health.
6. Colombo, A., Bendelow, G., Fulford, K.W.M., *et al.* (2003). Evaluating the influence of implicit models of mental disorder on processes of shared decision making within community-based multi-disciplinary teams. *Social Science and Medicine*, **56**, 1557–70.
7. Woodbridge, K., and Fulford, K.W.M. (2004). *Whose Values? A workbook for values-based practice in mental health care*. London: Sainsbury Centre for Mental Health.
8. Rogers, A., Pilgrim, D. and Lacey, R., (1993). *Experiencing Psychiatry: Users' Views of Services*. London: The Macmillan Press.

9. Fulford, K.W.M., Dickenson, D. and Murray, T.H. (eds.) (2002) *Healthcare Ethics and Human Values: An Introductory Text with Readings and Case Studies*. Malden, USA, and Oxford, UK: Blackwell Publishers.
10. Hunink, M., Glasziou, P., Siegel, J., *et al.* (2001). *Decision Making in Health and Medicine: Integrating Evidence and Values*. Cambridge: Cambridge University Press.
11. Brown, M. M., Brown, G.C. and Sharma, S. (2005). *Evidence-Based to Value-Based Medicine*. Chicago: American Medical Association Press.
12. Fulford, K.W.M. and Colombo, A. (2004). *Six Models of Mental Disorder: A Study Combining Linguistic-Analytic and Empirical Methods.* Philosophy, Psychiatry, & Psychology, **11/2**, 129–44.
13. Stanghellini, G. (2004). Deanimated bodies and disembodied spirits. In *Essays on the Psychopathology of Common Sense*. Oxford: Oxford University Press.
14. Andreasen, N.C. (2001). *Brave New Brain: Conquering Mental Illness in the Era of the Genome*. Oxford: Oxford University Press.
15. Kendell, R.E. (1975). The concept of disease and its implications for psychiatry. *British Journal of Psychiatry*, **127**, 305–15.
16. Szasz, T.S. (1961; revised 1974). *The Myth of Mental Illness*. Foundations of a Theory of Personal Conduct. New York, USA: Harper and Row.
17. Sadler, J.Z. (2004). *Values and Psychiatric Diagnosis*. Oxford: Oxford University Press.
18. Wakefield, J.C. (2000). Aristotle as Sociobiologist: The 'Function of a Human Being' Argument, Black Box Essentialism, and the Concept of Mental Disorder. *Philosophy, Psychiatry, & Psychology,* **7/1**, 17–44.
19. Fulford, K.W.M. (1989, reprinted 1995 and 1999). *Moral Theory and Medical Practice*. Cambridge: Cambridge University Press.
20. Hare, R.M. (1952). *The Language of Morals*. Oxford: Oxford University Press.
21. Fulford, K.W.M., Thornton, T., and Graham, G. (2006). *The Oxford Textbook of Philosophy and Psychiatry.* Oxford: Oxford University Press.

1.6

The psychiatrist as a manager

Juan J. López-Ibor Jr. and Costas Stefanis

Introduction: integrating two perspectives

The past years have witnessed the introduction and implementation of strict macro and micro economical principles in the planning and delivery of health care. The change is a consequence of the limitation of resources and the obligation to optimize their utilization in the delivery of health care to those in need. This has led to the birth of a new domain of management science dealing with hard choices and with the selection of priorities in the delivery of health care.[1] Even Western European countries, where a tradition of equity presides over health care and developing countries, with a tradition of care based on welfare system are adopting strategies built upon managed care principles. Of course, the private health care present in many forms in the different countries has been a driving force in this movement.

The new management perspective is based on the standards of management discipline but has to adapt to the main ethical concerns of delivery of care to suffering human beings. Because management has oftentimes been imposed as an 'external' mandate to the clinical community, clinicians, lacking understanding for the basics behind management, are reluctant to accept it; they often feel degraded by managers and develop negative attitudes towards them.

The objective of the present chapter is to provide clinicians with some input on how to manage the resources at hand and how to understand and communicate with managers in order to reach priorities closer to the needs of patients.

Health professionals and managers differ in many aspects. They belong to two different cultures which are summarized in Table 1.6.1.

However, many managerial skills are extremely useful to deal with psychiatric diseases, because they are chronic, they are accompanied by high degrees of disability, they require an interdisciplinary perspective, and they have important interactions with the social environment. Nowadays, most psychiatrists work as members of a multidisciplinary team, need to develop collaborative working relationships with other professionals, should have an understanding of the roles and the limits and extent of involvement of other agencies, and lastly, of the lines of accountability. Furthermore, the concerns about the competence of psychiatrists that is the framework for training programmes include some such as: the psychiatrist is a medical expert, a communicator, a collaborator, a manager, a health promoter, a professional and somebody able to tolerate ambiguity and uncertainty. In most of those, managerial skills are essential. Unfortunately, those skills are not taught in most medical schools.

There are three levels at which there is a parallel between a manager and a psychiatrist: a) the psychiatrist as a manager of the interventions needed to implement an individualized treatment plan for his patient, b) the psychiatrist as a manager of the involvement of other professionals in clinical settings, and c) the psychiatrist as a manager of health care resources available to his practice.

The psychiatrist as a manager of his patient's needs is a consequence of the introduction of processes of disease and patient management by most health care organizations. These play an increasing emphasis on prioritizing health care provision on the basis of limited resources and increased sensitivity to specialized patient needs. Therefore, the clinician has to keep a delicate balance between cost containment principles and quality in care provided. It is crucial for the psychiatrist to be able to identify and implement practices that assure quality of care without sacrificing this to any external pressures for containment of cost within his clinical practice. The *Madrid Declaration* of the World Psychiatric Association (WPA)[2] has one item on the rights of psychiatrists that in essence

Table 1.6.1 Two different cultures

	Health professionals	**Managers**
Values	Health and fighting diseases	Economy and administration
Main interest	1. Patients 2. The profession	1. Organization 2. Management
Principal loyalty	The profession	The institution
Main concern	Patients	Health policy
Persons are	Patients	Clients, stakeholders
Terminology	Medical	Business-like
Training in management	No	Yes
Clinical training	Yes	No
Worries about costs	No	Yes
Stability in working places	Long	Short

declares that the first right of the psychiatrist is to be able to practice the profession without external constraints of any kind.

The psychiatrist as a manager of other professionals in clinical settings is consequently working as an element of a multidisciplinary team. Everybody in a team should at least possess an understanding of the essence, the extent and the limitations of cooperation and accountability. Managing other professional also largely refers to managing other psychiatrists within the same clinical settings. Management goes beyond the simple 'coordination' of various roles and steps in the process of providing health care, in the professional education of colleagues, the training of young professionals and the sharing of experience.

The psychiatrist as a manager of health care services has to struggle to reach a satisfactory degree of equity and in order to do so, the clinician has to become familiar with issues such as human resource management, customer satisfaction and change, crisis and conflict management.

What is management?

The word management derives from the Italian *maneggiare* 'to handle' (i.e., a horse), from *mano* 'hand', from Latin *manus*. Management is 'the art of getting things done through people';[3] it is the act (sometimes the art) of conducting or supervizing something (initially a business) and the thoughtful use of means to accomplish an end. Management needs to direct and to control a group of people or entities for the purpose of coordinating and harmonizing that group towards accomplishing a goal and it often encompasses the deployment and handling of human, financial, technological, and other resources.

Management in health care is not new. Physicians have a long tradition of being supervisors or conductors of a team. The simplest form of management is the partnership, an essential model for the doctor–patient relationship.

Management can also refer to the person or people who perform the acts of management; in this sense management has to do with power by position, whereas leadership involves power by influence.

From a functional perspective, management consists of measuring a quantity on a regular basis and of adjusting some initial plan in order to reach an intended goal. This applies even in situations where planning does not take place.

Functions of management

Management has several functions, which are summarized as following:

Planning: deciding what needs to happen in the future and generating plans for action.

Organizing: making optimum use of the resources required to enable the successful carrying out of plans.

Leading and motivating: exhibiting skills in these areas for getting others to play an effective part in achieving plans.

Coordinating: making different people or equipments work together for a goal or effect.

Controlling (**monitoring**): checking progress against plans, which may need modification based on feedback.

Basic managerial concepts

Although in the following paragraphs we will, as often as possible, replace managerial jargon by one more pleasant to clinicians, there some basic concepts which need a definition.

Efficacy is the ability to produce a desired effect and it is measured by the closeness to an achievable goal.

In clinical settings, efficacy is the degree of the benefit for patients, induced by an intervention (treatment, procedure or service) in ideal research conditions (i.e., in a controlled trial). It indicates that the therapeutic effect is acceptable. 'Acceptable' refers to a consensus that it is at least as good as other available interventions to which it will have ideally been compared to in a clinical trial.

Effectiveness, on the contrary, refers to the impact in real world situations.

Efficiency is the achieved results or effects related to the effort invested in terms of money, time and other resources. It is the maximization of some desired output or effect for the least amount of input, means or effort. Usually, the larger the ratio, the greater the efficiency.

Efficiency is not a pure scientific concept as it carries a value judgement. Efficiency is achieved through design, the process by which intelligence is substituted for matter and energy in technological systems.

Productivity is a measure of efficiency; it is the amount of output created (in terms of goods produced or services rendered) per unit input used. For instance, labour productivity is measured as output per worker or output per labour-hour.

Equity is social justice, the way of providing services according to the needs of each individual in a defined population. It is not an equalitarian principle because each individual should not get the same, but what he or she would need in a specific situation.

The light bulb example: the efficacy is the amount of visible light measured in lumens; the efficiency is the ratio of lumens to the amount of energy consumed to produce them, measured in Watts. Equity will measure the reach of the lumens to the needs of, let's say, the passers by on a street.

Ethical aspects of management in clinical settings

Health care and economic management are two different cultures. Cultures are defined by their values and peculiarities, among them ethics. Specific values are part of the property of a culture and belong to the identity of every social group.

There are three stages in the development of medical ethics,[4] each one adding value to the previous one without replacing it totally. In each one of them particular managerial skills are helpful.

The **ethics of welfare** is the traditional medical ethics, first appeared in Hippocratic writings. According to it, the doctor's primary goal and duty are the well-being of the patient and as much as possible, harm avoidance. The doctor is perceived and behaves as a good father, to be fully trusted, convinced that the physician will act adequately to the benefit of the patient. To meet this obligation, the doctor has to increase to the maximum his own medical knowledge and to assume a series of obligations. The scientific advances increase the paternalism of the professional who has to learn how to manage information. The ***Madrid Declaration***

of the WPA expressed this notion in the following way: *Psychiatry is a medical discipline concerned with the provision of the best treatment for mental disorders. Psychiatrists serve patients by providing the best therapy available consistent with accepted scientific knowledge and ethical principles. It is the duty of psychiatrists to keep abreast scientific developments of the specialty.*

The USA influence, with its strong emphasis on autonomy and individualism, has lead to the **ethics of autonomy**. The ethics of autonomy considers the patient as an autonomous human being, adult and free and consequently, able to take his/her own decisions. The values and beliefs of the patient are the background for the moral responsibilities of the doctor. As a consequence, doctors have to truly inform patients about all possible diagnoses and treatments so that patients are able to decide. The basic element of this new way of establishing the doctor-patient relationship is the informed consent.[5] From this perspective, patient-doctor relationship is defined in new terms: *The patient should be accepted as a partner by right in therapeutic process. The therapist-patient relationship must be based on mutual trust and respect to allow the patient to make free and informed decisions. It is the duty of psychiatrists to provide the patient with relevant information so as to empower the patient to come to a rational decision according to his or her personal values and preferences* (WPA *Madrid Declaration*).

The **ethics of equity** is a consequence of the impact of economic factors in medicine. The need of equal access to health care resources for all patients, including those suffering from mental illnesses and the principle of equity in a period of intrinsic and extrinsic limitations to health care cost, is leading to a third stage of bioethics which has also been called the **ethics of management**. The main reasons for the increase or imbalance of the costs are partially due to the successful developments of modern medicine: health care by itself is increasingly expensive (implementation of new and expensive technologies, incorporation of new professions into medicine, financing research in biomedical sciences, and applying resources for the training of physicians and specialists); the better control of acute diseases which increases the proportion of chronic illness requiring care; the increased demand due to ageing of the population and in social security systems, the change in the population pyramid, decreases the population of those paying compared to those making the expenses.

Resources to be invested in health care are limited. The first one to ask for limits was President Carter in the USA, during his first public speech after assuming the presidency, when he claimed for a ceiling of the 7 per cent of the GNP to be devoted to health care. This was in 1977. But, why such a limit? Why could it not be possible for an enlightened society to decide to devote 10, 20 or even 50 per cent of its GNP to health care and less, for instance, to defence? President Carter expressed it very well: too much spending in this area would decrease investments in education and care of the environment, which would lead to a deterioration of health.[8] In Europe, the cost of brain disorders (to be precise brain diseases and mental disorders) is more than the double of the cost of all cancers and diabetes together.[7]

The fact that more is not better is evident when comparing health indexes, and among them the bottom line, which is life expectancy. This is much lower in the US which dedicate over twice the percentage of their GNP to health care, than countries such as Japan, France, Italy, or Spain.

Economic factors can limit the access to health care, either because individuals lack sufficient insurance coverage or because of waiting lists. In the last few years, limitations have been imposed by governments which through different approaches try to control the access of patients to interventions which are not considered economically worthy. Again the WPA has addressed these issues in several documents such as the *Hawaii Declaration* (rev.), the *WPA Statement and Viewpoints on the Rights and Legal Safeguards of the Mentally Ill*, the first official document on the rights of mentally ill, and the *Madrid Declaration* of the WPA. This last code of ethics states *'As members of society, psychiatrists must advocate for fair and equal treatment of the mentally ill, for social justice and equity for all. While doing so, psychiatrists should be aware of and concerned with the equitable allocation of health resources'.*

Therefore, the goal of cost control should not be considered in isolation from other goals, such as quality assurance and equity. Here a managerial approach is useful, when based on three pillars: Information (to know what we physicians do, how patients behave on long-term outcome of medical interventions and on the impact on quality of life), consensus (on the right approaches to decide interventions), and a new social contract on sustainable health care. This approach is not limited to health care, it has a great influence on the culture of modern enterprises and in post-modern perspectives.[9,10]

Nature of managerial activity

Workplace democracy has become more common and advocated. Management is based on classical military type of command-and-control but it should not throw itself to the other extreme here all management functions are distributed among all workers, each of whom takes on a portion of the work, and the institution is run by assemblies of staff and of patients as was common in some anti-psychiatry experiences. Management relies increasingly more on facilitating, promoting and supporting collaborative activity, which is essential in health care. Modern management embraces democratic principles, in that, in the long term, workers must give majority support to management; otherwise they leave to find other work, or go on strike.

In for-profit organizations, the primary function of management is the satisfaction of a range of stakeholders. This typically means making profit (for the shareholders), creating valued products at a reasonable cost (for customers), and providing rewarding employment opportunities (for employees). In non-profit management, other functions are added, such as keeping the faith of donors, attaining social and political goals such as increasing health and fighting diseases.

In most models of management, shareholders vote for the board of directors, and the board then appoints senior management. This model is rarely applied in health care (or in Academic life).

Management also has the task of innovating and improving the functioning of organizations.

Categories of management

There are many categories of management. The most important are: human resource management, production (operations) management, strategic management, marketing management, financial management, and information technology management.

Nevertheless, as more and more processes simultaneously involve several categories, it is better to think in terms of the various processes, tasks, and objects subject to management.

Management science and organizational psychology

Whether management is rightly claiming a scientific status is debatable. It is more appropriate to classify it as a branch of economic sciences often confounded with its practical arm, operations research. Management science is the discipline of using analytical methods, to help make better decisions. Among others methods are decision-making analysis, optimization, simulation, forecasting, game theory[11] (which had a strong impact in psychotherapy[12]), network (transportation) forecasting models, mathematical modelling, data mining, probability and statistics, morphological analysis, resources allocation, and project management.

Industrial and organizational psychology consists of the application of psychological theories, research methods, and intervention strategies to workplace issues in order to hire suitable employees for the job, to reduce absenteeism, to improve communication and to increase job satisfaction.

Information systems

An information system is the array of persons, data records and activities that process the data and information in a given organization, including manual processes or automated processes. Information systems are also social systems whose behaviour is heavily influenced by the goals, values and beliefs of individuals and groups, as well as the performance of the technology. An information system consists of three components: human, technology, and organization.[13]

The systems rely on data from the unit as well as data acquired outside it (such as literature research, scientific meetings, consensus documents and others) and data provided by others (i.e. the Health Care System, partners, suppliers, and customers).

A computer based information system is a technologically implemented medium for recording, storing, and disseminating linguistic expressions, as well as for drawing conclusions from such expressions.[14]

Managerial levels and styles

The management of a large organization usually has three self-evident levels: senior management, middle management, and low-level management.

There are several management styles that can be applied depending on the nature of the activity, the type of the task, the characteristics of the workforce, and the personality and skills of the leaders. As the style of leadership is dependent upon the prevailing circumstance, leaders should exercise a range of leadership styles and should deploy them as appropriate.[15]

An **autocratic** or authoritarian manager makes all the decisions and keeps the information and decision-making among the senior management. Objectives and tasks are set and the workforce is expected to do exactly as required. The communication involved with this method is mainly downward, from the leader to the subordinate. The main advantage of this style is that the direction of the business is stable and the decisions are similar and comparable. This in turn projects the image of a trustworthy and well managed business. However, this method can lead to a decrease in motivation of employees and subordinates who may become highly dependant on the leaders and close supervision may be unavoidable.

A **paternalistic** approach is also dictatorial; however, the decisions tend to be in the best interests of the employees rather than the business. This can help balance out the lack of worker motivation caused by an autocratic management style. Feedback is again generally downward; however, feedback to the management will occur in order for the employees to be kept happy. This style can be highly advantageous, and can engender loyalty from the employees, leading to a lower labour turnover, thanks to the emphasis on social needs. It shares the same disadvantages of the authoritarian style; employees becoming highly dependant on the leader, and if wrong decisions are made, then employees may become dissatisfied with the leader.

In a **democratic** style, the manager allows the employees to take part in decision-making; therefore everything is agreed by the majority. The communication is extensive in both directions (from subordinates to leaders and vice-versa). This style can be particularly useful when complex decisions that require a range of specialized skills need to be made. From the overall business point of view, job satisfaction and quality of work will improve. However, the decision-making process is severely slowed down, and the need of a consensus may avoid taking the 'best' decision for the business. It can go against a better choice of action.

In a **laissez-faire** leadership style, the leader's role is peripheral and the staffs manage their own areas of the business; the leader therefore evades the duties of management leading to an uncoordinated delegation. The communication in this style is horizontal, meaning that it is equal in both directions; however, very little communication occurs in comparison with other styles. The style brings out the best in highly professional and creative groups of employees; however, in many cases it is not deliberate and is simply a result of poor management. This leads to a lack of staff focus and sense of direction, which in turn leads to much dissatisfaction, and a poor company image.

Roles and responsibilities of heads of clinical units and leaders

The head of a clinical unit is a managing director. In business, the principal leader is the Managing Director, who is in charge of the definition, the development and implementation of the strategic plan of their unit or service in the most cost-effective and time-efficient manner.

The managing director is responsible for both the day-to-day running of the company and developing business plans for the long-term future of the organization. In business, the managing director is accountable to the board and the shareholders of the company. It is the board that grants the managing director the authority to run the company. In clinical settings, the head of a unit is accountable to and gets the power from the health authorities.

A head of a unit may or may not have direct clinical responsibilities and is usually burdened by much office-based work, but he or she is the leader of the organization, chairing different sorts of meetings, motivating the workforce and developing the culture and style of the organization.

As the title suggests, the managing director needs to manage everything. This includes the staff, the patients, the budget and the

resources to make the best use of them and increase the company's profitability.

Strategic planning

Strategic planning is the process of defining the goal of an organization and of making decisions on allocating resources to pursue the goal. In order to determine where it is going, the organization needs to know exactly where it stands, then determine where it wants to go (over the next years, typically 3 to 5) and how it will get there.[16] The resulting document is called the 'strategic plan'.

Strategic planning deals with at least one of three key questions: 1 'What do we do?' 2. 'For whom do we do it?' and 3. 'How do we excel?'

There are many approaches to strategic planning. Typically it is done in a stepwise manner:

1 Vision (define the vision and set a mission statement with a hierarchy of goals);

2 SWOT analysis (strengths, weaknesses, opportunities and threats);

3 Formulation (of the actions and processes to be taken to attain these goals);

4 Implementation (of the agreed upon processes);

5 Controlling (to get full control of the operation); and

6 Monitoring (to get feedback from implemented processes).[17]

Situational analysis. When developing strategies it is important to analyze the organization and its environment in the present moment and how it may develop in the future. The analysis has to be executed both at internal as well as at external levels to identify all opportunities and threats of the new strategy. Analysis of the external environment normally focuses on the customer (the needs of patients).

Goals, objectives and targets: Vision, mission and values. These are essential components of strategic planning. They are specific, time-bound statements of intended future results, and general and continuing statements of intended future results.

A **Vision statement** outlines what a clinical unit wants to be in the future; it is a source of inspiration and provides clear decision-making criteria. It reflects the optimistic, perhaps utopic view of the organization's future.

A **Mission statement** describes what the unit or service is at present. It defines the customers (kinds of patients), critical processes and it informs about the desired level of performance.

A vision statement is different from a mission statement. The Vision describes a future identity and the Mission describes why it will be achieved.

A **Values statement** describes the main values protected by the organization during the progression, reflecting the organization's culture and priorities.

Management by objectives

Management by Objectives (MBO)[18] is a method of agreeing on objectives within an organization. The management and the head of clinical units reach a consensus on the objectives to attain in a certain time period (typically one year). The objectives have to comply with some criteria, usually described by the acronym SMART (Specific, Measurable, Achievable, Realistic, and Time-Specific). They usually represent an increased level of performance than the one attained during the previous period of time. The objectives of each clinical unit are part of the global objectives of the hospital or the higher level organization (i.e., the health care area) which are the result of the consensus between the organization and the health care authorities. Some objectives are collective, while others can be individualized. Ideally, the responsible of a clinical unit shares and divides the objectives among the responsible staff under his authority.[19]

The achievement (or non-achievement) of the objectives should lead to incentives (or penalties). In health care, significant pay incentives (bonuses) are not common, and this is an advantage because high bonuses trigger unethical behaviour such as distorting financial figures to achieve short term individual targets. This is the main criticism of management by objectives.

Quality management

Quality management is an exceptionally useful tool for the running of clinical services. It is based on scientific excellence which includes research on efficacy and efficiency. In can include systems to control costs but accompanied by methods to sustain and increase quality. The basic assumption is that cost control and quality of care can run in parallel. Of the several approaches, we have chosen for this chapter, the European Foundation for Quality Management (EFQM)[20] model, because it is increasingly been used by health care administration in many countries and not only by Europeans, and although developed for the management of very large companies it soon became evidence of their advantages for the public sector.

The EFQM model relies on the strive over Excellence, which is defined as the *outstanding practice in managing the organization and achieving results*. Truly Excellent organizations are those that strive to satisfy their stakeholders by what they achieve, how they achieve it, what they are likely to achieve, and the confidence they have that the results will be sustained in the future. Being excellent requires total leadership commitment and acceptance of the fundamental concepts, a set of principles on which the organization bases its behaviours, activities, and initiatives.

Excellence relies on a few fundamental concepts (Fig. 1.6.1):

1 Results Orientation: Excellence is achieving results that delight all the organization's stakeholders.

2 Customer Focus: Excellence is creating sustainable customer (patient) value.

3 Leadership and Constancy of Purpose: Excellence is visionary and inspirational leadership, coupled with constancy of purpose.

4 Management by Processes and Facts: Excellence is managing the organization through a set of interdependant and interrelated systems, processes and facts.

5 People Development and Involvement: Excellence is maximizing the contribution of employees through their development and involvement.

6 Continuous Learning, Innovation and Improvement: Excellence is challenging the status quo and effecting change by utilizing learning to create innovation and improvement opportunities.

7 Partnership Development: Excellence is developing and maintaining value-adding partnerships.

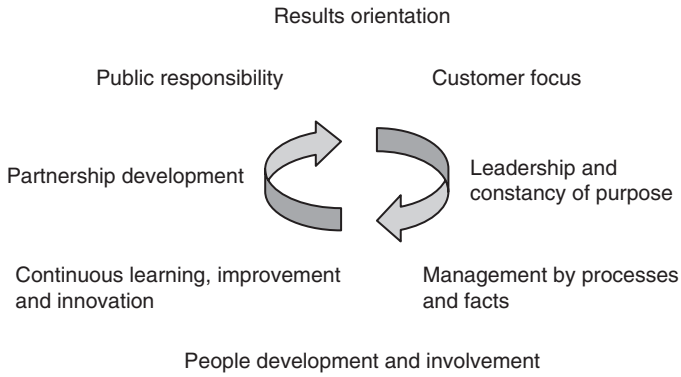


Fig. 1.6.1 Concepts of excellence.

8 Corporate Social Responsibility: Excellence is exceeding the minimum regulatory framework in which the organization operates and strives to understand and respond to the expectations of their stakeholders in society.

The EFQM Excellence Model is a practical tool that can be used in a number of different ways: as a tool for self-assessment; as a way to benchmark with other organizations; as a guide to identify areas for improvement; as the basis for a common vocabulary and a way of thinking; as a structure for the organization's management system.

The EFQM Excellence Model is a non-prescriptive framework based on nine criteria (Fig. 1.6.2). Five of these are 'Enablers' and four are 'Results'. The 'Enabler' criteria covers what an organization does. The 'Results' criteria covers what an organization achieves. 'Results' are caused by 'Enablers' and 'Enablers' are improved using feedback from 'Results'.

Risk management

Risk management is a discipline for living with the possibility that future events may cause adverse effects. The term is increasingly been used in the health sector.

Risk analysis includes risk assessment (identifying sources of potential harm, assessing the likelihood that harm will occur and the consequences if harm does occur), risk management (evaluation of the risks identified that require action and selection and implementation of the procedures to control those risks), and risk communication (interactive dialogue between all parties involved in the risk).

There are seven principles in the management of risk:

1 Global perspective (recognizing both the potential value of opportunities and the potential impact of adverse effects);

2 Forward-looking view (identifying uncertainties, anticipating potential outcomes);

3 Open communication (encouraging free-flowing information at and between all levels);

4 Integrated management (risk management is an integral part of management);

5 Continuity (maintaining constant vigilance);

6 Shared vision (common purpose, collective communication, and focusing on results); and

7 Teamwork (pooling talents, skills, and knowledge).

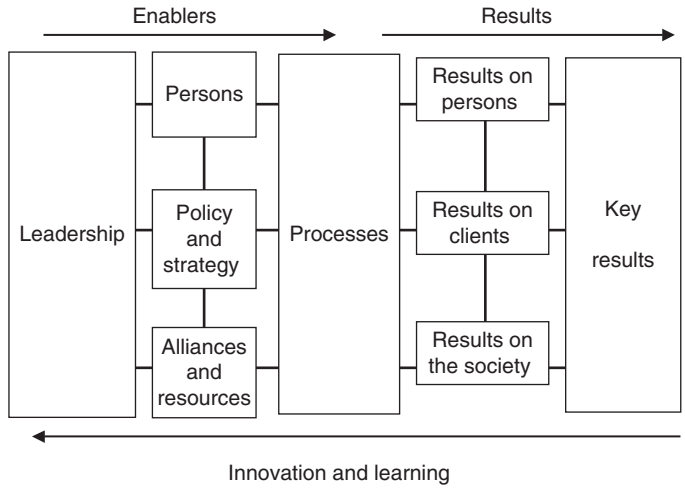


Fig. 1.6.2 European Foundation for Quality Management.

The experience of psychiatrists in disaster and traumatic events has made them highly valuable experts in risk management.

Types of health care and management systems

Health Systems throughout the world share the same, although sometime distant, goals: accessibility, equity, and extensive range of coverage. But there is also a wide array of organizational and financing types (taxation, employer-employee based private insurance, mixed), which have a strong impact on their efficiency, effectiveness, equity, and productivity. On the basis of financial policy, the following three main health care systems can be distinguished:

1 The social insurance system: It is also known as Bismarck model. It is based on obligatory insurance funded by both the employer and the employees. It guarantees collectivity, bilateral consent, and social solidarity. There is only marginal interference by the state, which however, is the main provider of health care facilities. Patients are free to choose their primary care physician, who is either a free professional or salaried to an insurance agency. In some countries, this system has evolved into the next one.

2 Government-state controlled through taxation system: Introduced by Beveridge in the U.K., it guarantees free of charge access of all citizens to health care services. Planning, programming, financing, and administration of services, as well as prevention and public health policy are centrally controlled. The system is financed through taxes as health care is considered a right of the citizens. Interestingly enough, it often co-exists with a non-negligible private insurance sector, which offers reduced waiting lists, better accommodation facilities, a more free election of the doctors and other privileges. In both cases, there are some limitations to access such as, having the primary care physician as the first contact and gate-keeper.

3 The private health system: The system is financed either through private insurance or by direct (out of the pocket) payment by the patient. Population insurance is not obligatory. Both, the patients and the physicians preserve the right of choice and payment is based on fee per service. The private system flourishes in the USA where about 2/3 of Health Services belong

to private or to managed care for profit or non-profit organizations. Governmental contribution to health care free of charge for patients is limited to the uninsured, very poor, and elderly as well as to patients in emergency through the Medicare and Medicaid Federal Government programs.

The more recent trend is towards adopting a mixed system of financing of health services, with an increasing collaboration of the private sector both in financing the services and in providing medical care.

Managed care

Managed care is the use of business managerial principles, strategies and techniques in health care. As it started in the USA during the presidency of Ronald Reagan, in a restricted sense is only applied to this country as a way to control Medicare and Health Maintenance Organizations' (HMO's) payouts. However, nowadays every system including those more controlled by governments use the same approach to control costs. Governmental run systems are just big nationwide HMO's and confront clinicians with the same restrictions to their work, patients with limitations of access, be it only waiting lists, conflicts of interests and confrontation of cultures. Essentially, it is a reform of health care from its longstanding not-for-profit business principles into a for-profit model that would be driven by the insurance industry or governmental bodies ruled by the same principles.

The reason for the beginning of managed care is the need to control cost and to reduce the so-called medical inflation which in the 1980s and 90s, was running at twice or thrice the general inflation rate. Nevertheless, managed care has not been so successful at this role, but has brought rationality in the use of public resources in health care, and often into attracting private resources to a successful for-profit business model.

There are several forms of managed care. Plans range from more restrictive to less restrictive, and include:

Health Maintenance Organization (HMO)

An HMO is an insurance plan under which an insurance company gears most aspects of the health care of the insured person. Each insured person is assigned a 'gatekeeper', usually a primary care physician who is responsible for the overall care of members assigned to him/her. Specialty services require a specific referral from the primary care physician to the specialist. Non-emergency hospital admissions also require specific pre-authorization by the PCP. Typically, services are not covered if performed by a provider not an employee of or specifically approved by the HMO, unless it is an emergency as defined by the HMO. The HMO concept was introduced in 1960 by Dr. Paul Elwood[21] and was adopted by the Nixon Administration.

Preferred Provider Organization (PPO)

PPO is a coinsurance system, which provides patients a co payment (generally around 80 per cent) of the costs of care, for an insurance fee. The deductible is the first part of the coverage and is paid by the patient. After the deductible is met, the coinsurance portion applies. Because the patient is picking up a substantial portion of the 'first dollars' of coverage, PPO is the least expensive type of coverage.

Point of Service (POS)

A POS plan utilizes some of the features of each of the above plans. Members of a POS plan do not make a choice about which system to use until the point at which the service is being used. For example, if the patient stays in a network of providers and seeks a referral to use a specialist, they may have a co payment only. However, if they use a network provider, but do not seek a referral, they will pay more.

Conclusions

It is clear that the modern role of both the physicians, in general, and psychiatrists, in particular, requires intensive decision-making which is helped by management principles.

Psychiatrists, in addition to their clinical qualifications and skills are asked to occupy positions and undertake responsibilities as clinical executives, directors of health care facilities, administrators of Academic units and even Mental Health Commissioners, all of them requiring managerial knowledge and leadership qualities.

Increased pressure by patients for improved quality of services and access to new and innovative treatments needs to be balanced against the expectation of the health care system of the physician to act 'economically', following cost containment guidelines and staying within expenditure ceilings.

Such decisions require specialized knowledge and a deep understanding of the principles and the functions of management and health economics. Such knowledge is only gained through specialized training by introducing management teaching, either at the undergraduate level or preferably at the residency level, as part of the core curriculum or as an elective which may include items such as administration principles, quality assurance, budgeting, resource allocation, accreditation procedures and what is close to the psychiatrist's clinical background the personnel management. This may be extended to ongoing professional education programmes for psychiatrists who are already active in the field.

Further information

Harvard Business Review http://harvardbusinessonline.hbsp.harvard.edu

World Psychiatric Association Section on Mental Health Economics Prof. Katarzyna Prot, Chair Institute of Psychiatry & Neurology. IV Dept. of Psychiatry Ceglowska 80. Warsaw, Poland 01-809 T: (40-21) 230-0653; F: (40-21) 230-7832 prot@bukareszt.ro; http://www.wpanet.org

References

1. Porter, M.E. and Teisberg, E.O. (2006). *Redefining Health Care: Creating Value-Based Competition on Results*. Harvard Business School Press, Boston.
2. World Psychiatric Association (1998). The Madrid Declaration. *Current Opinion in Psychiatry*, **11**, 1.
3. Follett, M. P. (1951). *Creative Experience*. Peter Smith, New York.
4. López-Ibor, J.J. and Crespo Hervás, M.D. (2000). A West Mediterranean Perspective. In *Ethics, Culture and Psychiatry. International Perspectives*. (eds. A. Okasha, J. Arboleda-Florez, and N. Sartorius). American Psychiatric Press, Inc. Washington, DC.
5. López-Ibor, J.J. and Crespo Hervás, M.D. (1996). Informed Consent in Spain. In *Informed Consent in Psychiatry. European Perspectives of Ethics, Law and Clinical Practice*. (eds. H.G. Koch, S. Reiter-Theil, and H. Helmchen), pp. 233–7. Medizin in Recht und Ethik, Band 33, Nomos, Baden-Baden.

6. López-Ibor, J.J. (1997). Psychiatric Care under the Present Economic Era. An International Perspective. *European Psychiatry*, **12** (suppl 2), 88–91.
7. Andlin-Sobocki, P., Jonsson, B., Wittchen, H.U., *et al*. (2005). Cost of disorders of the Brain in Europe. *European Journal of Neurology*, *Jun.* **12** (suppl 1), 1–27.
8. World Psychiatric Association (1990). *The WPA Bulletin* VI N.2–3.
9. Naisbitt, J. (1982). *Megatrends: Ten New Directions Transforming our Lives*. Macdonald.
10. Callahan, P. (1991). *What kind of Life*. Simon & Schuster, New York.
11. Morgenstern, O. and von Neumann, J. (1947). *The Theory of Games and Economic Behavior*. Princeton University Press
12. Berne, E. (1964). *Games People Play – The Basic Hand Book of Transactional Analysis*. New York: Ballantine Books.
13. Angell, I.O. and Smithson, S. (1991). *Information Systems Management: Opportunities and Risks*. Macmillan, London.
14. Langefors B. (1973). *Theoretical Analysis of Information Systems*. Philadelphia, Auerbach Publishers.
15. Tannenbaum, R.and Schmidt, W.H. (1985). How to Choose a Leadership Pattern. *Harvard Bussines Review*, March–April, 95–101.
16. Bradford, R.W., Duncan, P. and Tarcy, B. (2000). *Simplified Strategic Planning*. Chandler House Press, Worcester, MA.
17. Fahe, L. and Narayanan, V.K. (1986). *Macroenvironmental Analysis for Strategic Management*. West Publishing Company, New York.
18. Drucker, P. (1964). *Managemet for Results*. Harper & Row, New York.
19. Goldratt, E.M. and Cox, J. (2004). *The Goal: A Process of Ongoing Improvement*. North River Press, Great Barrington, MA.
20. http:// www.efqm.org
21. Ellwood, P.M. Jr., Anderson, N.N., Billings, J.E., *et al*. (1971). Health maintenance strategy. *Medical Care*, **9**(3), 291–8.

1.7

Descriptive phenomenology

Andrew Sims

Principles of descriptive phenomenology

Definitions and explanations

Psychopathology is the systematic study of abnormal experience, cognition, and behaviour. It includes the **explanatory psychopathologies**, where there are assumed causative factors according to theoretical constructs, and **descriptive psychopathology**, which precisely describes and categorizes abnormal experiences as recounted by the patient and observed in his behaviour.[1] Therefore the two components of descriptive psychopathology are the observation of behaviour and the empathic assessment of subjective experience. The latter is referred to by Jaspers as **phenomenology**,[2] and implies that the patient is able to introspect and describe what these internal experiences are, and the doctor responds by recognizing and understanding this description. Descriptive phenomenology, as described here, is synonymous with phenomenological psychopathology, and involves the observation and categorization of abnormal psychological events, the internal experiences of the patient, and consequent behaviour. The attempt is made to observe and understand this psychological event or **phenomenon** so that the observer can, as far as possible, know what the patient's experience must feel like.

Mental phenomena in health and cultural variation

It is not surprising that the identification and classification of the phenomena of mental illness is a difficult task as there is no consensus concerning what would be acceptable as normal healthy experiences. Mental illness has variously been considered as the products of a diseased brain, the symptoms that doctors treat, or a statistical variation from the norm carrying biological disadvantage, and mental illness often has legal implications. It is best to retain the use of the word 'normal' in a statistical sense; thus a phenomenon, such as hypnagogic hallucination, may be statistically abnormal but not an indicator of ill health or mental disease. Similarly, it is unwise to extrapolate from a population of mentally ill people and make assertions about the origins of behaviour in those who are not mentally ill.

It is important to recognize the effect of **culture** on subjective experience, the expression of psychological symptoms, and their manifestation in behaviour. In some cultures the very expression of subjective experience and emotion is discouraged and censored, in others feelings tend to be somatized, and in yet others the subjective experience of the individual tends to be subjugated to the sense of well being of the immediate social group. There are specific culture-bound expressions of subjective distress concerning body image in those who suffer from anxiety disorders. For delusions of passivity, although the psychopathological form remains relatively constant, the description of content will vary according to culture; for example, 'the djinn made me do it', 'my thoughts are controlled by the television'. Similarly, for possession state, although the psychopathological description remains similar, the actual cultural expression is very different between a member of a fundamentalist sect in the American Appalachian Mountains and a Buddhist girl in Sri Lanka.

Understanding the patient's symptoms

Although in internal medicine a clear distinction is made between **symptom** (the complaint which the patient makes) and **sign** (the indicator of specific disease observed or elicited on examination), in psychiatry both are contained within the speech of the patient. He complains about his unpleasant mood state, therefore identifying the *symptom*; he ascribes the cause of the pain in his knee to alien forces outside himself, thus revealing a *sign* of psychotic illness. Because both symptoms and signs emanate from the patient's conversation, in psychiatric practice the term symptom is often used to include both. For a symptom to be used diagnostically, its occurrence must be typical of that condition and it must occur relatively frequently.

Fundamental to psychiatric examination is the use of **empathic understanding** to explore and clarify the patient's subjective experiences. The method of empathy implies using the ability to 'feel oneself into' the situation of the other by proceeding through an organized series of questions, rephrasing, and reiterating where necessary until one is quite sure of what is being described by the patient. The final stage is recounting back to the patient what you, the psychiatrist, believe the patient's experience to be, and the patient recognizing that as indeed an accurate representation of their own internal state. Empathy uses the psychiatrist's capacity, as a fellow human being, to experience what the patient's subjective state must feel like as it arises from a combination of external environmental and internal personal circumstances.

Identifying phenomena as specific indicators of defined psychopathology may be difficult. It may require hearing much conversation from the patient for significant words and sentences to be revealed. The psychiatrist, when in the role of psychopathologist, has to assume that all speech of the patient, all behaviour of the patient, and every nuance has meaning, at least to the patient at the time the speech or behaviour takes place; it is not just an epiphenomenon of brain functioning.

Jaspers has contrasted **understanding** with **explaining**; descriptive phenomenology is concerned with the former. Understanding is the perception of personal meaning of the patient's subjective experience and involves the human capacity for empathy. That is, I understand because I am able to put myself into my patient's situation and know for myself how he is feeling, I feel those feelings of misery myself. Explanation is concerned with observation from outside and working out causal connections as in scientific method. In psychopathology, the terms *primary* and *secondary* are based upon this important distinction between meaningful and causal connections. That which is primary can be reduced no further by understanding, i.e. by empathy. What is secondary emerges from the primary in a way which can be understood by putting oneself into the patient's situation at the time; that is, if I were as profoundly depressed as my patient, I could have such a bleak feeling that I believed the world had come to an end—a nihilistic secondary delusion.

Subjective experience and its categorization

Within certain limits subjective experience is both predictive and quantifiable. When an individual loses a close relative it can be predicted that he or she will experience misery and loss. It is possible to quantify depressive symptoms and compare the degree of depression at different times in the same individual or differences between individuals at the same time. An important distinction for psychopathology is that between form and content. The **form** of psychological experience is the description of its structure in phenomenological terms (e.g. a delusion). Its **content** is the psychosocial environmental context within which the patient describes this abnormal form: 'Nurses are coming into the house and stealing my money'. The form is dependent upon the nature of the mental illness, and ultimately upon whatever are the aetiological factors of that condition. Content is dependent upon the life situation, culture, and society within which the patient exists. The distinction is important for diagnosis and treatment; determining the psychopathological form is necessary for accurate diagnosis, whereas demonstrating the patient's current significant concerns from the content of symptoms will be helpful in constructing a well-directed treatment regime.

Whereas most science is concerned with objectivity and with trying to eliminate the observer as far as possible from being a variable within the experiment, descriptive phenomenology tries to make evaluation of the subjective both quantifiable and scientific. It is a mistake to discredit subjectivity in our clinical practice. Inevitably we use it all the time and we should learn to use it skilfully and reliably. When I make an assessment that my patient is depressed, I am, at least to some extent, making a subjective judgement based upon the experienced and disciplined use of empathy: 'If I felt as my patient looks and describes himself to be, I would be feeling sad'. In psychopathology the distinction is also made between **development**, where a change of thinking or behaviour can be seen as emerging from previous patterns by understanding what the individual's subjective experience is, and **process**, where an event is imposed from outside and this cannot be understood in terms of a natural progression from the previous state. Anxiety symptoms could be seen as a development in a person with anankastic personality confronted with entirely new external circumstances; epilepsy and its psychiatric symptoms would be a process imposed upon the individual and not understandable in terms of previous life history.

Theoretical bases of descriptive phenomenology

There are important theoretical differences from dynamic psychopathology. Descriptive psychopathology does not propose explanations accounting for subjective experience or behaviour, but simply observes and describes them. Psychoanalytic psychopathology studies the roots of current behaviour and conscious experience through postulated unconscious conflicts and understands abnormalities in terms of previously described theoretical processes. The distinction between form and content and between process and development is not seen as important in psychoanalysis, but symptoms are considered to have an unconscious psychological basis. Descriptive phenomenology makes no comment upon the unconscious mind. It depends upon the subject being able to describe internal experiences, i.e. conscious material. Descriptive psychopathology is not dependent upon brain localization but on clarifying the nature of the subjective phenomena in discussion with the patient; if links can then be shown between certain phenomena and specific brain lesions, that is, of course, highly advantageous in furthering psychiatric knowledge. Descriptive phenomenology can be a unifying factor between concepts of brain and mind; it does not depend on philosophical stance on the nature of mind or brain.

Disorders of perception

Perception is not restricted to the screening of physical signals by sense organs but implies the processing of these data to represent reality. Ideas from the philosophy of mind have influenced psychiatric concepts of perception and the constitution of reality. Recently the distinction between sensory screening and interpretation has been confirmed by neurocognitive research.

Hundert[3] used the philosophical idea contained in the Kantian distinction between *a priori* categories and *a posteriori* experiences as a framework for differentiating perception by the sense organs from the secondary evaluation process. Kant's emphasis on the interplay between 'distal' perception and 'proximal' conceptualization can be exemplified by the perception and recognition of faces, disturbed in the Capgras syndrome and to a lesser degree in schizophrenia. The processing of visual perception is organized on at least four levels of complexity: the retina, the lateral geniculate body, the occipital visual cortex, and the hippocampus. The occipital cortex, where we actually 'see', does not contain an image any more than do the preceding levels; rather, it holds a database composed of signals from specific neurones for edges, angles, curves, sudden movements, and so on. Compared with the perceptual screen of the retina, these signals are 'scrambled' but even so they form a notion of what we perceive as reality. Recognition of faces needs further processing, probably in the hippocampal area where associations from other cortical fields are integrated with the

visual information (e.g. the voice belonging to the face). In psychiatry we deal with heterogeneous aetiologies, and perceptual disturbances may originate from different levels of processing, usually from a more integrated level than in neurological disease, and further from the immediate screening of physical stimuli by the sense organs. Thus, psychiatric disorders of perception affect different stages of information processing—from disturbances in the sense organs to complex phenomena involving feelings and ideas.

Here we shall mainly focus on hallucinations and some related phenomena, which are relevant for psychiatric illnesses.

Definitions of perceptual disturbances

Cutting[4] defines **hallucination** as 'perception without an object or as the appearance of an individual thing in the world without any corresponding material event'. The problem with this definition is that although some hallucinating patients mistake a hallucinatory perception for a real one, others can differentiate them: as demonstrated experimentally by Zucker,[5] there is an 'as if' quality even when patients assert that they perceive real objects or events. Voices described in detail by hallucinating patients were imitated and presented to the patients without warning. They had no difficulty in discriminating these external voices from their hallucinations. For this reason Janzarik[6] defined hallucinations, without associating them with perception at all, as 'free running psychic contents' (using a concept similar to Jackson's disinhibition). In keeping with this idea, lack of perception may facilitate hallucinations as in sensory deprivation or in the oneiroid states of paraplegic patients.[7]

The perceptual quality of hallucinations differs from similarity to sensory experiences, as in delirium, to the bizarre apprehensions of some with schizophrenia. Also, the extent to which the person is affected varies from descriptions of hallucinations as film-like in amphetamine psychoses to the affectively overwhelming experiences of hallucinations associated with delusional mood.

The term **pseudohallucination**, sometimes, is used to describe a perception recognized as *unreal*. Jaspers[8] defined hallucination as corporeal and tangible; pseudohallucination lacks this quality. According to Jaspers, pseudohallucinations are not tangible and real as hallucinatory perceptions; they appear spontaneously; they are discernible from real perception; and, they are difficult, but not impossible, to overcome voluntarily. Kandinsky illustrated Jaspers' definition of pseudohallucination: spontaneously arising images of acquaintances arose when a patient kept his eyes closed. He was fully aware of the unrealistic character of this experience and could abandon it by opening his eyes. Thus, to Jaspers, pseudohallucinations are close to imagined images except that they arise spontaneously and are more vivid. Jaspers' definition is not used consistently; in some Anglo–American literature it has been sufficient for the definition of pseudohallucination that there be subjective awareness that the percept lacks a real external equivalent and arises from the subject.

Imagery describes vivid visual experiences, which can be produced and manipulated voluntarily. It occurs in trance states when the perceptions are produced voluntarily, but are more real and last longer than in a normal state.

Illusions differ from hallucinations in being based on a misinterpretation of a real object or event, often associated with a mood. Illusions have to be distinguished from delusional perceptions which are percepts based on real objects to which an incorrect meaning has been attached. In delusional perceptions this 'error' cannot be corrected by the patient; in illusions the patient can recognize the true meaning.

Kurt Schneider described *Gedankenlautwerden* (*écho de la pensée*, or thoughts heard aloud) as a transitional phenomenon between vivid imagination and auditory hallucination. The patient recognizes that the words he hears are his own thoughts, but he cannot voluntarily control them. *Gedankenlautwerden* can disturb concentration when talking to other people. It can be differentiated from thought insertion and from auditory hallucinations in that there is a lesser degree of alienation.

Klosterkötter[9] has described transitions from elementary unformed hallucinatory sensations, like a crack, bump, or hiss, through more meaningful perceptions which still can be localized 'inside' the head, to complex hallucinations which become part of a delusional cognitive structure. These transitions were related to increasingly affective involvement in the themes of the hallucination. Klosterkötter's observations support Janzarik's interpretation of hallucinations as 'free running psychic contents', as do experimental studies of model psychoses which show a regular sequence of three psychopathological states: vegetative arousal, affective change, and 'productive' phenomena-like hallucinations and delusions.

Some misperceptions, found mainly in schizophrenic patients, are less complex than hallucinations, appear to be more closely related to neuropsychological disturbances, and include less systematization. They include **optical distortions** of size, colour, distance, and perspective, which can resemble experiences reported by people taking cannabis or other psychoactive drugs. These fluctuating, circumscribed misperceptions exemplify the way in which a more complex phenomenon of psychopathology can be built upon something more basic. Krause *et al.*[10] videotaped the non-verbal behaviour of schizophrenic patients and their healthy partners in a conversation. Brief non-verbal cues play an important part in dialogue. Schizophrenic patients miss these non-verbal brief cues and are poor at judging the intentions of others; their own non-verbal communication is poorly co-ordinated. The ensuing dysfunction diminishes social competence. Schizophrenic painters, trained before the onset of their illness, have been shown to misperceive perspective.[11]

Sensory modalities

Hallucinations can affect every sense modality. The most common, in the idiopathic psychoses, are **auditory hallucinations**, usually in the form of voices, although other kinds of sound may be associated with delusional contents. Voices talking to each other about the patient, and voices commenting about the patient's ongoing acting or thinking, are considered to be typical of, but not specific to, schizophrenia.[12] Voices calling the patient's name or talking without comments to the patient are diagnostically non-specific.

Visual hallucinations are most frequently found in organic psychosis, particularly delirium, in which they may occur for only a couple of hours during the night if the syndrome is not full blown. Visual hallucinations, more often than those in other sensory modalities, depict animals and scenes with several persons. In alcoholic delirium in particular, optical hallucinations of fine structures (such as hairs, threads, or spider webs) occur, and are especially likely to appear if the patient stares at a white wall. A typical, although not specific, combination of hallucination and delusion in organic psychosis is the 'siege experience', in which

patients believe they are besieged by enemies and have to bar their doors and windows.

Bodily, tactile, or coenaesthetic hallucinations are associated more often with schizophrenia than with affective or organic psychoses. The phenomenology includes simple tactile sensations of the skin, sexual sensations, sensations of the contraction, expansion, or rotation of inner organs, or atypical pain. Usually these sensations are associated with delusional explanations. Tactile hallucinations localized in the skin can underlie the **delusion of parasitosis**. Elderly patients in the early stages of organic cerebral alterations are at highest risk.

Coenaesthesia is a bodily misperception, which may last for minutes to days. It fluctuates (sometimes in relation to stress), and is usually not attributed to external agents or explained by delusional ideas. Patients seldom report them spontaneously. Klosterkötter[9] suggests that when coenaesthesia is attributed strongly to external influences, it is likely to be followed by schizophrenia.

Hallucinations may be **gustatory** or **olfactory**, for example, a smell of gas (perhaps thought to emanate from neighbours trying to kill the patient). Blunting of gustatory sensations or misperception of food as oversalted or overspiced is occasionally reported by melancholic patients.

Aetiological theories of hallucination

Aetiological theories are of three kinds:

1 overstimulation affecting different levels of information processing;

2 failure of inhibition of mental functions;

3 distortion of the processing of sensory information at the interpretive level.

The work of Penfield and Perot[13] has suggested that **overstimulation** may be a pathogenic mechanism. They stimulated the temporal regions of 500 patients, of whom 8 per cent reported scenic hallucinations, some in several modalities. Stimulation of the visual occipital cortex led to simple hallucinations-like flashes, circles, stars, or lines. This phenomenon has been observed in drug-induced experimental psychosis. It is interesting that schizophrenic patients can usually distinguish drug-induced hallucinations from those arising from their disorder. Using neural network theories, Emrich[14] simulated hallucinations by using Hopfield networks; overloading the storage capacity of the network generated what appeared to be the equivalent of hallucinations.

Disinhibition theory originated with Hughlings Jackson, who considered that productive symptoms were caused by the disinhibition of controlling neural activities, while negative symptoms resulted from damage to the systems, which generate the productive symptoms. More recently, sensory deprivation research has yielded inconsistent results; hallucinations, narrowly defined, seldom occur after deprivation, which may be of greater relevance to vivid, usually visual, imaginative experiences. Disinhibition may also underlie the 'hypnagogic hallucinations' which can occur in healthy subjects shortly before they fall asleep.

The role in the production of hallucinations of post-sensory interpretation and evaluation of stimuli is uncertain. In these terms, hallucinations are a sort of deception, but this is not a sufficient description of their nature. Recent neurophysiological hypotheses and findings from neuroimaging studies have suggested that there is an 'inner censorship' involved in clarifying ambiguities of perception.[14]

Disorders of thinking

Types of thinking

Three types of thinking can be distinguished which represent a continuum, without sharp boundaries, and intertwined in everyday life, from low to high regard for external reality and goal-directness: fantasy thinking, imaginative thinking, and rational thinking.[15] Since each of these types can predominate under some conditions, this distinction is useful to understand certain abnormal phenomena.

Fantasy thinking (also called dereistic or autistic thinking) produces ideas, which have no external reality. This process can be completely non-goal-directed, even if the subject is to some extent aware of the mood, affect, or drive, which motivates it. In other cases fantasy serves to exclude reality, which may require material with which the subject does not want to engage. This type of fantasy thinking is directed. Its goal is not to solve a problem but to avoid it via neglect, denial, or distortion of reality. Normal subjects use fantasy thinking deliberately and sporadically. However, if its content becomes subjectively accepted as fact, it becomes abnormal. This pathological exclusion of reality can remain limited in extent (e.g. in hysterical conversion and dissociation, pseudologia phantastica, and some delusions) or it may be manifested as withdrawal from the real world.

Rational (conceptual) thinking attempts to resolve a problem through the use of logic, excluding fantasy. The accuracy of this endeavour depends on the person's intelligence, which can be affected by various disturbances of the different components involved in understanding and reasoning.

Imaginative thinking comes between fantasy thinking and rational thinking. It is a process of forming a representation of an object or a situation using fantasy but without going beyond the rational and possible. This thinking is goal-directed but frequently leads to more general plans than the solution of immediate problems. Imaginative thinking becomes pathological if the person attaches more weight to his representation of events than to other objectively equally possible interpretations. In overvalued ideas, the imagined interpretation surpasses other interpretations in strength; in delusions, all other possibilities are excluded.

Delusions

The term 'delusion' signifies a complex edifice of thinking in which 'delusional ideas' are linked with other ('normal') thoughts. Delusions are communicated to others in the form of judgement. In this context, the term 'delusional idea' customarily refers to pathologically false judgement for which three criteria have been proposed: the unrivalled conviction with which they are held, their lack of amenability to experiences or compelling counter-arguments, and the impossibility of their content.[16] The last criterion must be discarded for two reasons. Firstly, collective beliefs derived from the socio-cultural setting of a person can be considered, in other surroundings, as false or impossible. Taking this into account, delusion is often defined as a 'false unshakable belief, which is out of keeping with the patient's social and cultural background'.[15] Secondly, in certain delusions (e.g. delusional jealousy) the content

does not go beyond the possible. Thus delusions are best defined as overriding, rigid, convictions which create a self-evident, private, and isolating reality requiring no proof.[17]

(a) The genesis of delusions

Jaspers[18] introduced a distinction between primary and secondary delusions. He supposed that the first, called true *delusional ideas*, are characterized by their 'psychological irreducibility', whereas the second, called *delusion-like ideas*, emerge understandably from disturbing life experiences or from other morbid phenomena, such as pathological mood state or misperception. This led to the assumption that primary delusions are the direct expression of the underlying condition considered to be the basis of schizophrenia. Four types of primary delusion have been distinguished in this perspective.

1 **Delusional intuition** (autochthonous delusion), occurring spontaneously, 'out of the blue'.

2 **Delusional percept**, in which a normal perception acquires a delusional significance. Schneider[19] assumed that 'psychological irreducibility' was clearly evident in this process, and included *delusional percept* among his 'first-rank symptoms' of schizophrenia.

3 **Delusional memory** can be distorted or false memory coming spontaneously into the mind, like delusional intuition. In other cases they occur, like delusional percept, in two stages, which means that normal memories are interpreted with delusional meaning.

4 **Delusional atmosphere** refers to an ensemble of minuscule and almost unnoticed experiences, which impart a new and bewildering aspect to a situation. The world seems to have been subtly altered; something uncanny seems to be going on in which the subject feels personally involved, but without knowing how. From this uncertainty evolves first certainty of self-reference, and then the formation of fully structured and specific delusional meaning. The apparent change in the surrounding situation is accompanied by tension, depression, or suspicion, and by anxious or even exciting expectations, so that it is often called 'delusional mood'.

The primary–secondary distinction assumes that the delusional atmosphere is part of the process underlying all primary delusional phenomena. If this preliminary disturbance is not perceived clearly or is not communicated by the patient as a general change in the situation, delusion may be manifested only as delusional percept, intuition, or memory. When the initial change in atmosphere is experienced clearly, a subsequent alteration in the environment, or a fully formed delusional idea, can lead to release from the preceding perplexity. The origin of primary delusions is commonly attributed to a basic cognitive anomaly disturbing information-processing, which reduces the influence of past experience on current perception. This is considered to entail a heightened awareness of irrelevant stimuli and an ambiguous unstructured sensory input allowing the intrusion of unexpected and unintended material from long-term memory.[20]

(b) The content of delusions

The content of delusions is determined by the mood in which they emerge and evolve, by the patient's personality and socio-cultural background, and by previous life experiences. In principle, the content can embrace all kinds of presumptions in separate categories. The following six delusional themes are usually distinguished:

- **delusion of persecution** based on the assumption that the patient is pursued, spied upon, or harassed
- **delusional jealousy**
- **delusion of love** characterized by the patient's conviction that another person is in love with him or her
- **delusion of guilt**, unworthiness, and poverty which may sometimes reach the degree of 'nihilistic delusion', in which the patient believes the real world has disappeared completely
- **grandiose delusion** in which patients are convinced that they have great talents, are prominent in society, or possess supernatural powers
- **hypochondriacal delusion** founded on the conviction of having a serious disease.

The mood state when delusional ideas emerge favours certain themes. Delusion of guilt, or unworthiness, and hypochondriacal delusion are strongly linked with depression. Grandiose and erotic delusion generally occurs in excited or manic states. Delusions of persecution and jealousy emerge most frequently from suspicious mood states or a delusional atmosphere, but may occur in depressed subjects.

Some further specific contents of delusions are:

- **religious delusion**, which may occur with grandiose delusion or delusion of guilt
- **delusion of infestation**, a subtype of hypochondriacal delusion, and characterized by the conviction of infestation by small organisms
- **delusional misidentification** in which the patient believes, on the basis of a delusional percept, that a perceived person has been replaced by an imposter, or in which he is convinced that another person has been physically transformed into his own self
- **delusion of control** in which the patient experiences sensations, feelings, drives, volition, or thoughts as *made* or influenced by others (this schizophrenic delusion is believed to result from cognitive dysfunction consisting of a failure of the system which monitors willed intentions).[21]

(c) The structure of delusions

1 The alternatives, 'logical' or 'paralogical', indicate whether or not the connection of ideas is consistent with logical thinking.

2 The notions, 'organized' or 'unorganized', indicate whether or not the delusional idea is integrated into a formed concept. Highly organized, logical delusions are described as *systematized*.

3 The relationship between delusion and reality varies:
   - in **polarized delusion**, delusional reality is inextricably intermingled with actual fact
   - if the delusional belief and reality exist side by side without influencing each other, we speak of **juxtaposition**
   - in **autistic delusion** the patient takes no account of reality and lives in a delusional world.

Overvalued idea

An overvalued idea is an acceptable, comprehensible idea pursued beyond the bounds of reason.[22] Overvalued idea causes disturbed functioning or suffering to the person himself or others.

Overvalued ideas of prejudice (overvalued paranoid ideas) are characterized by an underlying self-referent interpretation of the behaviour or sayings of others; patients assume themselves to be overlooked, slighted, unfairly treated, provoked, or loved. Overvalued apprehension may become apparent as morbid jealousy, hypochondriacal phobia (e.g. parasitophobia), or dysmorphophobia, in which patients assume that they attract attention because of a real or presumed bodily defect. In anorexia nervosa subjects are preoccupied by the endeavour to remain thin, and in transsexualism by the desire to change gender because they feel that they belong to the opposite sex.

Overvalued ideas generally occur with abnormal personality under stressful situations. Those with paranoid personality traits may develop, on the basis of a presumed injustice, querulous, or litiginous overvalued idea. Sometimes ideas become overvalued only during abnormal mood states (of various origins) which set aside counterbalancing influences.

Thinking in mood disorders

The content of thought in mood disorders is coloured by affect. Negative thinking about self, the future, and the world prevails.[23] Mishaps and failures are attributed to personal faults; success is attributed to the action of other people. This depressive thinking spreads from the starting point of negative life events to more general events, and it tends to become long lasting. The fixed viewpoint that emerges is called 'cognitive schema'. After recovery from an acute episode this schema may become latent, but it can be reactivated by distressing life events. It can also prolong symptoms. Negative thinking started by minor misfortunes can become autonomous, driving down mood—which in turn intensifies negative thinking. The negative schema can prolong a depressive episode or precipitate a new one. It is probable that such schemas are activated by both cognitions and emotions. Guilty thoughts are closely connected with this type of thinking, and may reach the intensity of a delusion. To a degree, guilty thinking in depression is dependent on culture. In mania, the content of thought is related to the mood of elation, with diminished self-criticism and excessive self-importance. In phobic and other anxiety states, thinking centres on situations leading to anxiety. Typical contents of delusional thinking in depression concern guilt, religious failure, condemnation, personal insufficiency, impoverishment, hypochondriasis, and nihilistic ideas. In mania, delusional ideas may be feelings of spiritual or economic power. In contrast with schizophrenic delusions, affective delusions grow out of the underlying, excessive mood and do not appear as something new and alien to the personality.

Phobic and anankastic phenomena

Phobic and anankastic (obsessional) phenomena have in common that the patient experiences them as unwanted, but cannot suppress them. They often occur together.

(a) Phobia

Phobias are inappropriate, exaggerated fears which are not under voluntary control, cannot be reasoned away, and entail avoidance behaviour.[24] The fears are kindled by particular stimuli. These may either be perceived objects, such as animals (animal phobia) or pustules (in some illness phobias), or situations such as open places (agoraphobia) or confined rooms (claustrophobia).

Phobias initially triggered by a very specific stimulus can eventually generalize. Thus, an elevator phobia may become extended to all kinds of closed rooms. Some phobias are linked with broader circumstances from the beginning. In social phobia, for instance, patients avoid meeting people because they fear that they will be noticed. Identical types of fears can be triggered by different stimuli in different subjects. Thus, illness phobia is activated in some patients by observed body changes, but in others by situations involving the risk of infection.

Phobias are characterized by avoidance behaviour: patients avoid anxiety-provoking objects or situations. Because of stimulus generalization, this can lead to severe impairment; for instance, they cannot leave home.

(b) Anankastic symptoms

Anankastic phenomena occur as obsessions or compulsions:

1 Obsessions occur as repeated thoughts, memories, images, ruminations, or impulses that patients know to be their own but are unable to prevent. The content of these ideas is often unpleasant, terrifying, obscure, or aggressive.

2 Compulsions are actions, rituals, or behaviours that the patient recognizes as part of his own behaviour, but cannot resist.

(c) Combined syndromes

In phobic–anankastic syndromes patients attempt to reduce their phobic fears by certain actions, such as hand washing in the case of an infection phobia. If obsessional thoughts or impulses induce anxiety (e.g. obscene ideas during worship, or the impulse to lean too far over a balustrade) and entail the avoidance of the situations that provoke them, the term anankastic–phobic syndrome is used.

Phobias, obsessions, and compulsions result most frequently from neurotic conflicts, but they also occur with functional or organic mental disorders. Anankastic personalities, characterized by perfectionism, rigidity, sensitivity, and indecisiveness, are especially prone to develop obsessions and compulsions.

Disorder of the thinking process

Disturbance of thinking may be recognized by the patient himself or deduced by an observer from the subject's speech.

Impairments of thought production are conventionally named 'formal thought disorder' and contrast with abnormalities of the 'content of thought' observed in delusions. In the deviant reality-testing of deluded patients there is always a disturbance of the form of thinking.

(a) Disorders of the flow of thinking

Each remembered idea is linked with a number of other notions, related closely as well as distantly. In rational thinking, a 'determining tendency'[25] guides the flow of ideas in the chosen direction and excludes associations which do not conform with this goal. This procedure can be disturbed in various ways which are commonly grouped together under the heading of 'formal thought disorder'.

(i) Disturbances of the speed of thinking

In **acceleration** of thinking, associations are still formed normally but at grossly accelerated speed. The goal is not maintained for long and the intervention of new thoughts produces 'flight of ideas'.

Retardation refers to a slowing down of the thinking process, which hampers formation of associations and may prevent reaching the original goal of thoughts. This results in difficulties in concentration and decision-making.

Acceleration and retardation of thinking are due to a change of affect, and are characteristic of mood disorders.

(ii) Circumstantiality

In circumstantiality the determining tendency is maintained but the patient can reach the goal only after having exhaustively explored unnecessary associations arising in his mind. When answering a question, he relates many irrelevant details before returning to the point. This inability to exclude unimportant associations occurs in organic mental disorders and in mental retardation.

(iii) Perseveration

Perseveration is found in organic mental disorders and is defined as an inability to shift from one theme to another; a thought is retained long after it has become inappropriate in the given context. For example, a patient may give a correct answer to the first question, but repeats the same response to a subsequent, completely different inquiry.

(iv) Interruptions in the flow of thinking

Thought blocking is a sudden unintended cessation in the train of thought, experienced by the patient as 'snapping off'. After this break, which may occur in the middle of a sentence, the previous thought may be taken up again or replaced by another. Thought blocking occurs in organic states, in depression, and frequently in schizophrenia where it is described as part of negative thought disorder.

In **loosening of association** the flow of thinking is interrupted by deviation towards distant or unrelated thought, in contrast with flight of ideas in which there is only a speeding up of access to nearby associations. Loosening of association is a type of formal thought disorder. In **tangentiality** the ideas deviate towards an obliquely related theme. In **fusion**, different kinds of association evoked by an original thought are blended to produce a word or sentence. **Derailment** is characterized by the interpolation of ideas which neither the patient nor the observer can link with the previous stream of thought. **Muddling** designates an extreme degree of derailment and fusion.

In organic states, incoherent thinking, which is clinically similar to derailment, may be attributable to a primary intellectual impairment and not to an increased spread of associations.

(b) Overinclusive thinking

This kind of thought disorder is not based on an interruption of the flow of thought but on an inability to preserve conceptual boundaries; ideas only distantly related to the concept under consideration become incorporated within it,[26] for example, when asked to indicate the essential components of a *room*, *table* might be included as well as *ceiling*, *wall*, and *floor*.

(c) Concrete and abstract thinking

In organic mental disorders and mental retardation, inability to think abstractly may be attributed to a diminished capacity to structure a concept. There have been various theories used to explain the **concrete thinking** of schizophrenia, involving memory, conceptualization, and intrusion of delusions. The process may be enhanced by loosening of associations. The fact that schizophrenia sometimes manifests excessively **abstract thinking** may also be explained by a disturbance of working memory such that the concrete meaning of the initial thought is not retained.

(d) Disorder of control of thinking

In **obsessions** and **compulsions** the subject recognizes his thoughts as being produced by himself but is unable to control them.

In **passivity of thought**, the patient experiences his thoughts as manipulated by outside influences. The interpretation resulting from this feeling is described as 'thought withdrawal', 'thought insertion', or 'thought broadcasting' (which denotes the patient's conviction that his thoughts are diffused to other people). These 'delusions of the control of thought' were included by Schneider[27] among his 'first-rank symptoms' of schizophrenia.

A particular variation of thought insertion occurring in schizophrenia is **crowding of thoughts**. In this condition, the patient experiences an excessive increase in the amount of thoughts imposed from the outside and compressed in his mind.

Language and speech disorder

'Speech disorder' refers to defects in the ability to generate and articulate verbal statements, whereas 'language disorder' designates deficits in the use of language. The terms 'aphasia' and 'dysphasia' are often used interchangeably for speech disorders.

(a) Disturbed generation and articulation of words

Aphonia designates the inability to vocalize. Thus, whispering occurs in somatic illnesses (paralysis of cranial nerve IX or disease of the vocal cords) and hysteria. **Dysphonia** is a somatic impairment with hoarseness.

Dysarthria refers to disorders of articulation occurring in various malformations or diseases, which impair the mechanisms of phonation, in lesions of the brain stem, in schizophrenia, and in psychogenic disorders.

The causes of **stuttering and stammering** are unclear, but are sometimes considered to be of neurotic origin. **Logoclonia** (the spastic repetition of syllables) occurs in Parkinsonism.

(b) Disturbance in talking

'Disturbances in talking' was proposed by Scharfetter[28] as a generic term for disorders of speech or language not belonging to the preceding group of disturbances.

Changes in volume of sound and in intonation occur in affective and schizophrenic states, and refer to loud excited and quiet monotonous speech.

Bradyphasia (decelerated talking) and **tachyphasia** (accelerated talking) occur in mood disorders, schizophrenia, and organic dysphasias.

Logorrhea (verbosity) is observed in various disorders, especially in manic states.

Alogia (poverty of speech) is a decrease in spontaneous talking; it occurs in depression and schizophrenia.

In **poverty of content of speech**, the amount of speech is adequate but conveys little information. This is often related to schizophrenic disorganization of thinking.

Verbigeration is the monotonous repetition of syllables and words observed in organic language disorders, schizophrenia, and agitated depression.

Echolalia is the repetition of words or parts of sentences that are spoken by others. It can be observed in schizophrenia, organic states, and subnormality.

Sometimes patients give **approximate answers**: i.e. they avoid giving the correct answer to a question that they have understood, just missing being correct. This occurs in organic disorders, schizophrenia, and hysteria.

Paraphasia denotes the enunciation of an inappropriate sound instead of a word or phrase. This happens in organic speech disorders but may also have psychogenic causes.

Speech may be unintelligible for various reasons. **Paragrammatism** and **parasyntax** (loss of grammatical and syntactical coherence) occur in organic mental disorders and excited manic states, and in schizophrenia, when severe thought derailments become manifest as 'word salad'. **Private symbolism** can be observed in schizophrenia in three forms: use of existing words with a particular symbolic meaning, creation of 'neologisms' (new words with an idiosyncratic meaning), and production of a private incomprehensible language, which may be spoken (cryptolalia) or written (cryptographia).

Mutism (refraining from speech) may be found in various kinds of psychiatric disorder. It is a cardinal feature of stupor and also occurs as an 'hysterical' reaction to stress.

Pseudologia fantastica is characterized by fluent lying, which is developed into a fantastic construct. This 'mythomania' occurs in histrionic and asocial personality disorders.

(c) Organic language disorders

This refers to impairments of spontaneous language, naming, writing, and reading, occurring as a result of brain dysfunction. These disorders can be divided into 'sensory' (receptive), 'motor' (expressive) defects, or both combined, containing the following principal subcategories:

(i) Sensory language disorders

In **primary sensory dysphasia** the patient cannot understand the speech of others. His own speech remains fluent, but contains errors in the use of words, syntax, and grammar. Writing and reading are also impaired. If, in this condition, the patient's speech becomes unintelligible, the disturbance is called 'jargon aphasia'. If only the repetition of a message is disturbed, the disorder is named 'conduction dysphasia'.

In **pure word-deafness** speech, reading, and writing are fluent and correct. The patient hears words as sounds, but cannot recognize their meaning. In **pure word-blindness** (alexia) speech and writing are normal but the patient cannot read with understanding.

(ii) Motor language disorders

In **primary motor dysphasia** the verbal or written expression of words and the construction of sentences is disturbed, but the understanding of speech and writing are preserved.

In **pure word-dumbness** the disturbance is limited to an inability to produce and repeat words at will. **Pure agraphia** is an isolated inability to write. **Nominal dysphasia** is an inability to produce names and nouns.

Disorders of intellectual performance

(a) Conceptualization of intelligence

'Intelligence' refers to the capacity to solve problems, to cope with new situations, to acquire skills through learning and experience, to establish logical deductions, and to form abstract concepts. There has been a classical debate amongst psychologists as to whether intelligence represents different and specific abilities or a unitary, general factor of intelligence.

(b) Measurements of intelligence

Individual intellectual capacity is graded by reference to the intelligence quotient (**IQ**), which is defined as the ratio of a subject's intelligence to the average intelligence for his or her age. The assessment of intelligence is considered in Chapter 1.8.3.

In addition to the global assessment of intelligence, numerous tests have been developed to assess organic impairment, scholastic achievement, and aptitude.

(c) Mental retardation (learning disability)

If the development of intellectual performance does not reach an IQ level of 70, the condition is designated 'mental retardation'. This is subdivided according to severity, with four levels recognized in ICD-10:

- mild (IQ 50–69)
- moderate (IQ 35–49)
- severe (IQ 20–34)
- profound (IQ below 20).

The causes of mental retardation are considered in Section 10.

Disorders of later onset

In these disorders normally developed intellectual performance declines. This can occur as a result of organic brain disorders, and in psychotic and affective disorders.

Organic disorders may have toxic, traumatic, inflammatory, or hypoxic causes. If these conditions are treated successfully, the disturbance can be arrested or even reversed.

In dementia there is a progressive disintegration of intellectual function, which usually begins insidiously and is often first recognized through an impairment of memory.

In psychotic states the distorted testing and evaluation of reality can impair intellectual performance. In schizophrenia, formal thought disorder may contribute to this effect.

Severe affective disorder can impair perception, attention, and motivation, leading to poor intellectual performance. These disturbances are observed more often in depression, but can occur in manic mood.

Disorders of mood

This section outlines the psychopathological elements comprising mood disorders, in particular the different varieties of depression, mania, anxiety state, and depersonalization.

Mood is a state of mind, which is longer lasting than affect or feeling. Mood encompasses all mental processes; it is not influenced

by will, and is strongly related to values. Heidegger[29] has considered mood as the fundamental expression of an individual's being. Kierkegaard[30] emphasized the role of existential orientation in determining mood, especially general anxiety.

The extent and type of deviation of mood is important in affective disorders. Although there are no sharp boundaries between the normal variations and pathological states of mood, severe states are clearly abnormal and difficult to understand. Mood can be abnormal in several ways: sad or anxious in depressive disorders; euphoric in mania; irritated in mania or agitated depression; dysphoric in depression or in mixed manic—depressive disorders; morose in chronic-depressed states, often with a component of resentment; blunted (the feeling of 'having no feelings' or 'petrified' feelings) in prolonged, severe depressive disorder. Stanghellini[31] analysed depressed patients and described how morose affect may emerge when the patient struggles against declining abilities and experiences resistance. In such cases feelings of timidity and despair may contrast with an outward appearance of hostility.

Two types of **euphoria** should be differentiated: one shows elation and feelings of increased spiritual, intellectual, or physical power, and the other results from disinhibition in organic states and dementia. This second type, rather than elation, may show lack of interest and an attitude of negligence towards the patient's actual situation.

These abnormal moods are related to altered **bodily feelings** and thinking.

Abnormal **somatic** symptoms can be divided into physical symptoms, such as cardiovascular dysregulation, increased sweating, and feeling cold, and hypochondriacal symptoms, such as headache and feeling of tightness in the chest, heavy limbs, being choked, or difficulty in swallowing. These latter symptoms are related to feeling of loss of energy.

Lopez-Ibor[32] suggested the term 'depression-equivalent' for conditions in which somatic symptoms (e.g. headaches which vary on a diurnal pattern) dominate the clinical picture. Cross-cultural research has found higher rates of such somatic symptoms in depression in Africa[33] and South America,[34] and a lower rate of guilt compared with Western industrialized countries. However, the results are not wholly consistent and variation may reflect cultural differences or differing patterns of consultation with doctors, and what patients expect doctors to treat.

A feedback loop may develop between anxiety and physical **arousal**, e.g. palpitations, which accompanies it.[35,36] The prevalence of mitral valve prolapse is higher in anxiety disorder (37 per cent) than in the general population (5 per cent).[37] This finding is consistent with the idea that palpitation may lead to a conditioned anxiety response. The behaviour therapy technique of exposure aims to decondition this reflex. In social phobia and panic disorder anxiety is often complicated by anxiety-provoking situations which may lead to severe social disablement. Somatic symptoms of anxiety may be so prominent in some depressive states that patients are misdiagnosed as medically ill, with loss of weight, atypical pain, or sensory or motor disturbances. This type of depression has been called 'depressio sine depressione', or 'somatoform depression'.

Disturbances of diurnal rhythm can influence all the other symptoms of mood disorder.[38] There are changes associated with sleep in the electroencephalogram, with shorter REM latency (phase advance), and also changes in endocrine and cardiovascular circadian rhythms. In depression, sleep disturbance is characterized by early awakening, whereas falling asleep in the evening is often undisturbed. About 70 per cent of melancholic patients show diurnal distribution of mood, psychomotor activity, somatic symptoms, and slowed and impoverished thinking.

Psychomotor retardation or acceleration is one of the most prominent symptoms of mood disorder. Often the patient's appearance and expressive movements reveal more than words. The retarded patient's movements are slow, the limbs are rigid, the body is bent, and the expression is sad or anxious, and does not respond to the situation. The subjective feeling may be of emptiness, weakness, and tension. If the condition is severe, it can be difficult to discriminate between depressive and catatonic stupor; patients with depressive stupor seldom have increased muscular tension or rigidity. Increased psychomotor activity can appear in depression as agitation, i.e. restlessness without the ability to attain goals or organize behaviour. In mania, increased psychomotor activity is also seen in sexual excesses and extravagant spending.

Psychomotor retardation, and probably also acceleration, may be accompanied by a changed experience of time.[39] Depressed patients overemphasize the past, remembering guilt-connected events; manic patients feel that the future is immanent. Inability to distinguish wishes from reality results in poor decision-making in both depression and mania. Some depressives are unable even to decide how to dress in the morning. A manic patient's workroom can reflect the dissolution of his ability to give priority to important things, for example tools for immediate and frequent use and those seldom used. Extreme retardation is seen in depressive stupor when patients do not move, speak, eat, or drink. Extreme acceleration occurs in mania ('boiling over') and may be accompanied by a sense of confusion.

Retardation and acceleration are closely related to depressive and manic **thought disorder**. In depression the flow of associations is reduced and slowed, and short-term memory can appear impaired (pseudodementia). Depressed patients often ruminate about negative topics and have difficulty in terminating these thoughts. In mania, acceleration of thinking leads to a plethora of associations, 'flight of ideas', and pressure of speech. Unlike patients with schizophrenic thought disorder, depressed patients retain logical connections.

Depersonalization (see later) can occur alone or as part of a depressive state. In the latter, part of the body, the self, the mind, actions, or thinking are sensed as being alienated—not belonging to the self. In mood disorders, depersonalization does not usually reach the intensity of delusion, as it can in schizophrenia.

Although anxiety disorders and major depression have been defined by operational criteria in the diagnostic manuals, the clinical symptoms of mood states vary considerably. Attempts have been made to define a core syndrome by using factor analysis to identify latent trait symptom profiles derived from several assessment scales and from different samples of depressed patients. Cross-cultural comparisons of symptom profiles can also help to identify core symptoms. Among the latent traits, retardation was found most often, together with loss of interest and alteration of diurnal rhythm. Guilt, death wish, and affective reactivity occurred inconsistently.[40]

Disorders of self and body image

Disorders of self

These describe the abnormal inner experiences of I-ness and my-ness which occur in psychiatric disorders. Scharfetter has added the characteristic of awareness of being or ego vitality to the four formal characteristics previously described by Jaspers: feeling of awareness of activity, awareness of unity, awareness of identity, and awareness of the boundaries of self.[41,42]

(a) Disorder of the awareness of being

This is demonstrated by **nihilistic delusions**, which frequently occur in severe depressive illness and are a feature of the eponymous Cotard's syndrome.[43] Non-psychotic abnormality is exemplified by **depersonalization** in which the sufferer experiences his mental activity, body, or surroundings as changed in quality to become unreal, remote, or automatized.

(b) Disorder of awareness of activity

Disorder of the awareness of activity occurs with neurological lesions, such as some dyspraxias, and also in psychotic conditions in which the individual believes that no action has occurred when it has, or vice versa. This does not include action that the patient knows he has executed but with a belief it was under the influence of another. Non-psychotic disorder of activity occurs when an individual believes that he has no freedom of action and that his range of choice is limited by external circumstances, for instance a person with depressive symptoms who believes that he is inevitably incompetent.

(c) Disorder of awareness of singleness

In health, one assumes that 'I am one person'. Disorder occurs in the rare visual perceptual experience of **autoscopy**.[44] Non-psychotic examples of disorder of singleness include the double phenomenon, described by Jaspers,[45] and **multiple personality disorder,** which is the apparent existence of two or more distinct personalities within an individual, only one of them being evident at any time. The **double phenomenon** is much more frequent, and describes the self-experience of those who feel that there are two different parts of themselves in conflict with each other; they are fully aware of both at the same time.

(d) Disorder of awareness of identity

Disorder of identity occurs in **delusion of control** or **passivity experience**, in which the sufferer believes that he has been taken over by an alien, with the belief that there is a break in continuity from 'myself' who was there before. Non-psychotic disorder of awareness of identity is exemplified by **possession disorder**, in which there is a temporary loss of the sense of personal identity and the individual may act *as if* they have been taken over by another personality, spirit, or force.

(e) Disorder of the awareness of boundaries of self

Disorder of boundaries of self occurs in first-rank symptoms of schizophrenia such as thought withdrawal, control, and diffusion.[46] The patient believes that thoughts 'are being taken out of me, influenced by an outside source'. Non-psychotic disorder of the boundaries of self occurs in ecstasy states, characteristically described as an 'as if' experience. There is disturbance of boundaries of self in that the individual may feel that there is no limit between self and the outside world.

Depersonalization and derealization

Depersonalization is the experience of one's own feelings and experiences being detached, distant, not one's own, lost or altered. Derealization is the same range of subjectivity describing awareness of the outside world. The sufferer recognizes that this is a subjective change and is not imposed by outside forces. Because the sufferer finds it difficult to describe, this experience tends to be underdiagnosed, but the misery it causes and the disturbance in functioning is considerable; it is experienced as being so subjectively unpleasant that not uncommonly deliberate self-harm results.

Insight

The clinical assessment of a patient's capacity to understand the nature, significance, and severity of his or her own illness has been called insight. There is current interest in describing its characteristics and establishing how it correlates with other measures of illness.[47] The attitude of patients towards their illness has clear clinical implications, and the assessment of insight tries to investigate the patient's awareness concerning the impact their illness has, and their capacity to adapt to the changes brought about by illness. The patient's awareness of illness and the extent to which it is interfering with function affects compliance for prescribed treatment. David has proposed that insight implies the ability to relabel unusual mental events as *pathological*, the recognition that one has mental illness, and compliance with treatment. Some parallels have been drawn between the loss of insight in psychiatric patients and the denial of disease or loss of function that occurs in certain neurological conditions.

Because of its importance for clinical management, there have been many attempts over recent years to measure insight, all of which depend upon a precise operational definition of the concept. McEvoy *et al.*[48] developed a questionnaire to measure the patients' awareness of the pathological nature of their experiences and also their acceptance of the need for treatment. The measure constructed by David *et al.*[49] added the ability to relabel unusual mental events as 'pathological' to the recognition of mental illness and compliance with treatment.

The relationship between impairment of insight and the presence of other aspects of psychopathology is complicated; there is no clear association between impairment of insight and intellectual or neuropsychological deficit.[50] Not surprisingly, patients with unimpaired insight are found to be significantly less likely to require readmission to hospital, tend to be more compliant with treatment, and show an improved prognosis.[51] Surprisingly, and this shows how little is known about this subject, many patients are prepared to comply with treatment, even though they do not believe themselves to be ill, if the social milieu is conducive to receiving treatment.[48]

Insight is a multifaceted phenomenon with considerable clinical significance as it predicts the likelihood of patients complying with treatment. Most studies of insight have been concerned with patients suffering from schizophrenia, and it is important to extend work to other serious mental illnesses.

Disorders of awareness of the body

(a) Bodily complaint without organic cause

Such conditions create difficulties for psychopathological understanding.

1 Aetiology is often obscure, sometimes with doubt that there may be an unrevealed physical cause.
2 The descriptive terms used come from different theoretical backgrounds and have changed their meaning over the years.
3 There is often discrepancy between the meanings attached to the symptoms by the patient and by the doctor.

'Somatoform disorders', which include both somatization and hypochondriacal disorders[52] are, characteristically, repeated presentation of physical symptoms with persistent requests for medical investigation, despite negative findings, and reassurance by doctors that the symptoms have no physical basis. The patient with **somatization** as the prominent disorder complains of multiple recurrent, and often changing, physical symptoms in different bodily systems over a prolonged time. The patient with **hypochondriasis** has a persistent preoccupation with bodily function, the possibility of illness, and the seriousness with which symptoms should be treated. Not infrequently these two groups of symptoms overlap. Co-morbid anxiety and depression is quite frequent with both somatization and hypochondriasis. The content of hypochondriasis may take the form of delusion, overvalued ideas, hallucination, anxious or depressive rumination, or anxious preoccupation.

The term 'Dissociative (conversion) disorder' has replaced the confusing, but graphic, hysteria. **Conversion symptoms** can be categorized as motor, sensory (including pain), or psychological. Motor symptoms include weakness or paralysis of limbs or part of a limb and abnormality of gait; sensory symptoms include glove, and stocking anaesthesia. Amongst the psychological symptoms is a narrowing of the field of consciousness with selective amnesia such as may occur in fugue states. For conversion disorder, or hysteria, to be diagnosed, symptoms should appear to be psychogenic in nature, causation should be considered unconscious, symptoms may carry some sort of advantage to the patient, and they occur by the mediation of the processes of conversion or dissociation.

Artefactual illness includes two categories: *elaboration* of physical symptoms for psychological reasons, and intentional production or *feigning* of symptoms or disabilities, either physical or psychological. Conversion symptoms are believed to arise without the patient's conscious involvement, but artefactual illness implies that the illness, lesion, or complaint is ultimately the individual's own conscious production. **Malingering** implies feigning or producing symptoms expressly for the social advantages of being regarded as ill, while the broader category of artefactual illness includes other motivations and simply describes the behaviour.

Narcissism is an exaggerated concern with one's self-image, especially with personal appearance. This absorption with self is usually associated with feelings of insecurity and ambivalence concerning the self and feelings of threat to one's integrity.

Dislike of the body and distortion of body image are subjectively different experiences but often occur together, for example in anorexia nervosa or with gross obesity. In **dysmorphophobia** the primary symptom is the patient's belief that he or she is unattractive. Sufferers believe themselves to have a physical defect, such as the size of their nose or breasts, that is noticeable to other people, but objectively their appearance may lie within normal limits. The content disorder of dysmorphophobia is an overvalued idea in which the degree of concern and consequent distress is clearly out of proportion and comes to dominate the whole of life. The overvalued idea of dysmorphophobia may be associated with an underlying personality disorder of anankastic or dependent type, or with other psychiatric disorders.

Awareness of body size and **disturbance of eating** frequently occur together; alteration of body image is associated with eating disorder. Obesity in adolescence, in diet-conscious Western societies, frequently results in self-loathing, more frequently in girls than boys, with overestimation of body fatness. Disturbance of body image occurs in sufferers from anorexia nervosa, characteristically an overestimate of width with an accurate estimation of height or the width of inanimate objects. The more 'over-fat' an individual considers herself to be, the more dissatisfaction with herself she will experience.[53] Such disorders of self-image, with significant overestimation of size and discrepancy between perceived and desired size, also occur in bulimia nervosa, and may be associated with depression of mood and feelings of guilt and unworthiness.

(b) Organic changes in body image

Organic change may result from either damage to the conceptualized object (e.g. following amputation, with a phantom limb) or damage to the process of conceptualization (e.g. section of the corpus callosum). Hyperschemazia, pathological accentuation of body image, occurs when physical illness or neurological lesion causes enhancement of perception of an organ. Diminished or absent body image (hyposchemazia, aschemazia) may occur with loss of innervation, or with parietal lobe lesions. The diminution of body image may be simple (e.g. loss or neglect of a limb) or complex. There may also be distortions of the body image (paraschemazia), in which enhancement or diminution of parts of the body may occur.

(c) Disorder of gender and sexuality

Core gender identity is established very early in life and then retained; it is biologically influenced and socially reinforced. **Transsexualism** is a disorder of gender identity, much more common in biological males, in which there is discrepancy between anatomical sex and the gender that the person assigns to himself. The subjective belief is an overvalued idea. Other disorders of sexuality are considered elsewhere. (See Chapter 4.11.3)

(d) Pain as a psychopathological entity

Pain is a subjective experience, which is hard to describe and categorize; it is not well-charted phenomenologically. It appears to have more in common with mood than perception. Pain associated with psychiatric illness tends to be more diffuse and less well localized and to spread with non-anatomical distribution. It also tends to be complained of constantly, becoming even more severe at times but persisting without remission. It may clearly be seen to be associated with underlying disturbance of mood, which appears to be primary in time and causation. Psychogenic pain tends to progress in severity and extent over time. Persistent, severe, and distressing pain, which cannot be explained fully by a physiological process or physical disorder has been designated **persistent somatoform pain disorder**. (See Chapter 5.2.6.)

Motor symptoms and signs

Motor symptoms and signs may be due to a neurological disorder causing organic brain syndrome, such as rigidity in Parkinson's disease, or may be related to emotional states such as restlessness or tremor in anxiety. However, there is a further group of symptoms, which affect voluntary movements and often occur in functional psychoses. These symptoms are neither unequivocally neurological nor clearly psychogenic in origin and are termed **motility disorder** by some authors. Table 1.7.1 gives a glossary of disordered motility. Whether patients are unable or unwilling to move normally is still a matter of debate. The origin of motility symptoms may well be a functional (rather than a morphological) abnormality of basal ganglia.

A further classification of motility disorder distinguishes psychomotor hyperphenomena (e.g. tic disorder), hypophenomena (e.g. stupor), and paraphenomena (e.g. mannerism). **Tics** are rapid, irregular movements involving groups of facial or limb muscles. **Stupor** is a state in which a patient does not communicate, i.e. does not speak (mutism) or move (akinesia), although he or she is alert. **Mannerisms** are uncommon; they are conspicuous expressions by gesture, speech, or objects (e.g. dress) that seem to have a particular meaning, often delusional.

Catatonia is a psychopathological syndrome of disturbed motor behaviour. It is generally reversible, and it occurs with mood disorders, general medical conditions, toxic and psychotic states, and neurological disorders. Brain tumour, encephalitis, endocrine, and metabolic disorders may elicit catatonic symptoms. In Western countries catatonia is considered, nowadays, to be uncommon in general psychiatric practice, however, some catatonic symptoms have been found to occur in 5–10 per cent of acute psychiatric in-patients.[54]

Table 1.7.1 Symptoms and signs of motility disorder

Catalepsy (synonym; waxy flexibility, flexibilitas cerea)	Maintaining uncomfortable positions against resistance
Posturing	Maintaining uncomfortable positions that may have a delusional meaning
Stupor	Inability to communicate despite being awake
Akinesia	Inability to move
Mutism	Inability to speak
Echolatia	Repetition of another person's speech
Echopraxia	Repetition of another person's acts
Mannerism	Uncommon conspicuous expression by gestures, speech, or objects
Grimacing	Uncommon conspicuous facial expression
Stereotypy	Repetition of actions
Verbigeration	Repetition of speech
Tic	Rapid movements of facial or limb muscles
Akathisia	Inability to remain seated or standing
Psychomotor retardation	Slowing of mental and motor activity
Psychomotor agitation	Arousal of mental and motor activity (typically by anxiety)

Stupor, *mutism*, and *negativism* are the classical triad of symptoms demarcating the catatonia syndrome, and automatic obedience and stimulus-bound behaviour, stereotypy, and catalepsy contribute to the syndrome in manic patients.[55] Lesser symptoms may occur, and catatonia may take the form of hypomobility, with stupor only in extreme cases. Patients in stupor remain persistently unresponsive for hours, days, or even longer. They appear to be unaware of events around them and are mute.

Catatonic excitement presents as excessive motor activity. Such patients may talk incessantly, especially when in an 'exalted stage'.[56] There may be outbursts of talking, singing, dancing, and removing their clothes. Such states carry the risk of exhaustion, dehydration, and injury; it may be harmful and dangerous to the patient and to others.

In **negativism** the patient resists the examiner's manipulations with force equal to that applied by the examiner. In **catalepsy** (posturing), the patient maintains posture for long periods. These include facial postures, such as grimacing or *schnauzkrampf* (lips in a pucker); body postures, such as *psychological pillow* (lying on his back with head elevated as if on a pillow); lying in a 'jack-knife' position; and many other uncomfortable and bizarre postures maintained against gravity or attempts to rectify them. An examiner trying to move a cataleptic limb passively will notice **waxy flexibility**, in which initial resistance to an induced movement changes to gradually allowing the imposition of a posture, like bending a candle.

Stereotypy is non-goal directed, repetitive behaviour, the verbal form of which is called *verbigeration*, endless repetition of phrases and sentences. **Automatic obedience** occurs when, despite instructions to the contrary, the patient permits the examiner to move his limbs into a new posture. This may then be maintained against instructions to the contrary. In **ambitendency** the patient appears stuck in indecisiveness, resisting the examiner's non-verbal signals, but showing hesitancy in doing so.

Echo phenomena may occur when the patient is interacting with another person and present as *echolalia* (imitation of the speech of others) or *echopraxia* (imitation of the actions of others). **Mannerisms** are strange but purposeful movements characteristic of that person. They may be exaggerated caricatures of ordinary movements.

Disordered speech may also be regarded as a sign of disordered motility, as in mutism or verbigeration.

In **delirium**, tremor often occurs. Anxiety is accompanied by restlessness. In a particular motor pattern in delirium tremens (alcohol withdrawal delirium) the patient appears to be collecting objects or brushing away dust. Typically, the movements never seem to achieve what they are meant to and, of necessity therefore, are repetitive. Suggestibility in delirium may lead to movements, which are based on erroneous assumptions, such as trying to take hold of a proffered, but non-existent, thread. Patients may develop panic and try to flee. Speech may be hurried and indistinct. In some cases of delirium, such as that due to hepatic failure, patients may be hypoactive before becoming drowsy and comatose. Hepatic failure may also result in catatonic disorder.

Many conditions, such as brain tumour, encephalitis, and endocrine and metabolic disorders, may elicit catatonic symptoms. Patients with a variety of mental disorders may show abnormal movements that are of histrionic nature. They may throw themselves to the ground, seek and maintain bodily contact, or show psychomotor agitation. Alternatively, there may be psychogenic paresis.

In dementia there may be general disturbance of psychomotor functions leading to disturbed co-ordination and clumsiness. During the further progress of dementia, lethargy, and akinesia may occur.

Sequelae of encephalitis are known to include a number of motor symptoms apart from parkinsonism, as seen in the epidemic of encephalitis lethargica that occurred around 1920. **Tardive dyskinesia** is a side-effect of neuroleptic therapy. However, since signs of tardive dyskinesia such as perioral hyperkinesia and dystonias were described before the introduction of neuroleptics,[57] it is also a motor symptom of mental disorder in its own right.

Disorders of memory

The psychology of memory is discussed in Chapter 2.5.3.

Memory may be differentiated into short-term or recent memory and long-term or remote memory. Furthermore, ultra-short-term memory may be distinguished from short-term memory. Ultra-short-term memory encompasses immediate registration within the span of attention. Short-term memory reflects new learning. Long-term memory is usually associated with earlier data or other information that has been stored for months or years.

Additional terms describing memory functions are *declarative* and *procedural* memory. Declarative memory contains facts, which may be consciously recalled, whereas procedural memory contains skills and automatic activities. In dementia—both degenerative (Alzheimer type) and vascular (multi-infarct dementia)—recent memory is usually impaired earlier than remote memory.

Biographical memory is the recall of events in a person's past, which have an emotional loading, and therefore an impact on understanding depression.

Amnesia is a period of time, which cannot be recalled, and it may be global or partial. With regard to time it may be retrograde—an expression derived from the idea that one is looking backwards from an event (such as brain trauma or electroconvulsive therapy) to find the period that is deleted before the event. Correspondingly, anterograde amnesia means a period of deleted memory after an event. Although it is difficult to distinguish between types of amnesia, focal lesions in the hippocampus seem to affect remote memory less than recent memory, whereas diffuse brain disease often affects both. In psychogenic amnesia it is sometimes possible to recognize specific personal meaning in the events which cannot be recalled.[58] Amnestic disorders should strongly alert the examiner to the possibility of cerebral pathology.

Disorders of memory are closely connected with other disorders, such as disorders of consciousness; there is often amnesia for episodes of disturbed consciousness.

Some patients are aware of memory disorder and complain about it; others tend to neglect their memory deficits and manifest secondary signs such as confabulation. Confabulations are inventions, which substitute for missing contents in gaps of memory; the patient is not aware that they are not true memories.

A disorder of short-term memory, as in Korsakoff's syndrome or transient global amnesia, is often neglected by the patient. Behaviour appears normal, and it often seems that the personality is intact. Such a patient may be engaged in lively conversation or seemingly purposeful actions, and only after further investigation does it become obvious that these activities are not based on facts. This memory disorder can be assessed directly by examining the patient. Other forms become apparent retrospectively on taking the patient's history. In these cases the patient complains about periods of global or partial amnesia. Memory for certain events may have faded or become obscured by layers of other events (palimpsest); this is typical of repeated amnestic periods following bouts of drinking. In mood disorder there may be complaints about impaired memory, although no memory deficit is found in objective tests. Examples of false memory (paramnesia) are *déjà vu*, an erroneous feeling of familiarity with, for example, a person or a room, and jamais vu, a feeling of unfamiliarity for a well-known object. *Déjà vu* may occur in temporal lobe epilepsy, although it is not specific for that disorder. Delusional memories are also examples of paramnesia.

Disorders of consciousness

Consciousness is the sum of various mental functions—in the words of Jaspers[59] 'the whole of present mental life'. Lipowski,[60] who regards the concept of consciousness to be 'completely redundant', describes what is commonly meant by clouding of consciousness on the basis of a number of behavioural features (Table 1.7.2). In contrast with Lipowski, the concept of consciousness has recently elicited fresh interest in philosophy and clinical neurology. (See Chapter 2.1)

Consciousness is a mode of relatedness between mind and world. Disordered consciousness may occur on a dimension of severity, which ranges from lucidity via clouding and then towards unconsciousness. The latter represents a state of coma. In addition, consciousness may be assessed on a dimension of vigilance.[61] Ey[62] regards consciousness as an attribute of wakefulness. Indeed, sleepiness implies a reduction in consciousness; but consciousness may also be reduced despite normal vigilance. Likewise, consciousness is impaired by a disorder of memory, orientation, or coherence, as in the clouded consciousness of delirium.

When consciousness is impaired there is clouding of perception, ideas, and images. The intensity of perception is diminished and there is a disintegration of order in the perceptive field. Accordingly, patients become disoriented.

The term 'confusional state' is a synonym for delirium that emphasizes thought disorder and disorientation. **Disorientation** may concern time, place, or person. Temporal and geographical disorientation are common. Remote contents are better remembered than recent ones; name or date of birth is usually more available than age, or name of the hospital. It is useful, after a polite excuse, to ask direct questions concerning orientation, even if they sound trivial, since some patients are skilful in avoiding topics that show the degree of their disorientation.

Table 1.7.2 Behavioural features indicating clouding of consciousness

The person is awake but may be drowsy
Awareness of the self and the environment is reduced
Both immediate and recent memory are impaired
Thinking is disorganized, and may be dreamlike; for instance perception is faulty and misperceptions may occur
The ability to (learn new material is reduced) learn
The person is unable to overcome this state by deliberate effort

Another abnormality is described by the term **narrowing of consciousness**, which means that awareness of a person's environment is restricted, for example, owing to an abnormal affective or delusional state.

In epileptic aura or after taking certain drugs, consciousness may be experienced as heightened with increased intensity of awareness.

Twilight state is a well-defined interruption of the continuity of consciousness. Consciousness is clouded and sometimes narrowed. Despite the disorder of consciousness, the patient is able to perform certain actions, such as dressing, driving, or walking around. Subsequently, there is amnesia for this state. Twilight states may occur in epilepsy, alcoholism (*mania à potu* is a twilight state), brain trauma, general paresis, and dissociative disorder. *Mania à potu* describes the situation where a person reacts excessively by developing twilight state with small amounts of alcohol. Often these patients have an increased vulnerability due to pre-existing organic brain pathology. Twilight state occasionally leads to violent behaviour.

In an **oneiroid state** the patient experiences narrowing of consciousness together with multiple scenic hallucinations. Oneiroid states may occur in schizophrenia, but are also observed in patients who have to be totally passive and dependent on others. The atmosphere is perceived as strange and dreamlike. Accordingly patients may be aloof and behave like dreamers.[63] Unlike twilight states, the contents of oneiroid states are often remembered.

Disorders of attention and concentration

Attention and concentration imply the directing of mental activities towards a particular object. Attention is associated with present alertness, and concentration with longer lasting achievement and performance; there is a distinction between *selective* and *shared* attention. Attention and concentration may be impaired by clouded consciousness or individual aspects, such as sleepiness, incoherence, or memory deficit. However, there may be other reasons for inattention such as hallucination or mood disturbances. Attention deficit is a permanent feature in the childhood disorder, attention-deficit hyperactivity disorder.

Assessment of attention and concentration may consist of simple arithmetical tasks and include psychometric performance test in addition to the clinical examination. Psychometric performance tests are also valuable tools in assessing disorder of memory and consciousness.

Disorders of sleep are described in Chapter 4.14.

Disorder of personality

It is the expression of disordered personality which is the consideration of descriptive phenomenology, the observation of characteristic behaviour and the subject's self description. Schneider has defined personality as 'the unique quality of the individual, his feelings and personal goals'[64] *Abnormality of personality* is present when a characteristic or trait of clinical relevance is developed in the patient to a statistically abnormal extent, that is either deficient or excessive. *Personality disorder* is present when that abnormality causes suffering to the patient or to other people. A person with antisocial personality disorder not uncommonly causes discomfort to other people; with obsessive-compulsive personality disorder the abnormal characteristics may frequently cause distress to the individual himself.

Both ICD 10 and DSM-IV are derived from Schneider's description of personality types. The advantage of such a typological approach is that it does not imply any specific theory of causation. The accurate description of personality characteristics and type is valuable in clinical practice for diagnosis, prognosis and the rational planning of treatment. The skills of psychopathology are ideally suited to the observation of consistent personality traits, and forming an opinion unprejudiced by preconceived theoretical considerations. Descriptions of different personality types and disorders are developed elsewhere.

Acknowledgements

This revised chapter, prepared by Andrew Sims, is based on that in the first edition by Andrew Sims, Christoph Mundt, Peter Berner, and Arnd Barocka. The contributions of these authors to the original text is gratefully acknowledged.

Further information

Sims, A. (2003). *Symptoms in the mind* (3rd edn). W.B. Saunders, London.
Jaspers, K. (1963). *General psychopathology* (7th edn) (trans. J. Hoenig and M.W. Hamilton). Manchester University Press.
Cutting, J. (1997). *Principles of psychopathology*. Oxford University Press.

References

1. Sims, A. (2003). *Symptoms in the mind* (3rd edn). W.B. Saunders, London.
2. Jaspers, K. (1963). *General psychopathology* (7th edn) (trans. J. Hoenig and M.W. Hamilton). Manchester University Press.
3. Hundert, E.M. (1995). *Lessons from an optical illusion*. Harvard University Press, Cambridge, MA.
4. Cutting, J. (1997). *Principles of psychopathology*. Oxford University Press.
5. Zucker, K. (1928). Experimentelles über Sinnestäuschungen. *Archiv für Psychiatrie und Nervenkrankheiten*, **83**, 706–54.
6. Janzarik, W. (1988). *Strukturdynamische Grundlagen der Psychiatrie*. Enke, Stuttgart.
7. Schmidt-Degenhard, M. (1992). *Die oneiroide Erlebnisform*. Springer, Berlin.
8. Jaspers, K. (1965). *Allgemeine psychopathologie*. Springer, Berlin.
9. Klosterkötter, J. (1992). The meaning of basic symptoms for the development of schizophrenic psychoses. *Neurology, Psychiatry, and Brain Research*, **1**, 30–41.
10. Krause, R., Steimer, E., Sänger-Alt, C., *et al.* (1989). Facial expression of schizophrenic patients and their interactionpartners. *Psychiatry: Interpersonal and Biological Processes*, **52**, 1–12.
11. Allderidge, P. (1974). *The late Richard Dadd*, pp. 1817–86. Tate Gallery, London.
12. Huber, G. and Gross, G. (1989). The concept of basic symptoms in schizophrenia and schizoaffective psychoses. *Recenti Progressi in Medicina*, **80**, 646–52.
13. Penfield, W. and Perot, P. (1963). The brain's record of auditory and visual experience. *Brain*, **86**, 595–696.
14. Emrich, H.M. (1989). A three-component-system hypothesis of psychosis. Impairment of binocular depth inversion as an indicator of a functional dysequilibrium. *British Journal of Psychiatry*, **155** (Suppl. 5), 37–8.

15. Hamilton, M. (1974). *Fish's clinical psychopathology. Signs and symptoms in psychiatry* (revised reprint). John Wright, Bristol.
16 Jaspers, K. (1975). *Allgemeine psychopathologie* (8th edn). Springer, Berlin.
17. Scharfetter, C. (1980). *General psychopathology*. Cambridge University Press.
18. Jaspers, K. (1963). *General psychopathology* (7th edn) (trans. J. Hoenig and M.W. Hamilton), p. 98. Manchester University Press.
19. Schneider, K. (1962). *Klinische psychopathologie* (6th edn). Thieme, Stuttgart.
20. Hemsley, D.R. (1994). An experimental psychological model for schizophrenia. In *Search for the causes of schizophrenia* (ed. H. Häfner, W.F. Gattaz, and W. Janzarik), pp. 179–88. Springer, Berlin.
21. Frith, C.D. and Done, D.J. (1989). Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychological Medicine*, **19**, 359–63.
22. McKenna, P.J. (1984). Disorders with overvalued ideas. *British Journal of Psychiatry*, **145**, 579–85.
23. Beck, A.T., Rush, A.J., Shaw, B.F., *et al.* (1979). *Cognitive therapy of depression*. Guilford Press, New York.
24. Marks, J.M. (1969). *Fears and phobias*. Heinemann, London.
25. Jaspers, K. (1963). *General psychopathology* (7th edn) (trans. J. Hoenig and M.W. Hamilton), p. 162. Manchester University Press.
26. Payne, R.W. (1996). The measurement and significance of overinclusive thinking and retardation in schizophrenic patients. In *Psychopathology of schizophrenia* (ed. P.H. Hoch and J. Zubin), pp. 77–97. Grune and Stratton, New York.
27. Schneider, K. (1962). *Klinische psychopathologie* (6th edn). Thieme, Stuttgart.
28. Scharfetter, C. (1980). *General psychopathology*, p. 114. Cambridge University Press.
29. Heidegger, M. (1963). *Sein und Zeit*. Niemeyer, Tübingen.
30. Kierkegaard, S. (1849). *Die Krankheit zum Tode*, Vol. 8. Diederichs, Jena.
31. Stanghellini, G. (1995). The interplay between personality and affective vulnerability: the case of dysphoria. Presented at World Psychiatry Association Regional Symposium, Prague.
32. Lopez-Ibor, J.J. (1969). Depressive Äquivalente. In *Das depressive syndrom* (ed. H. Hippius and H. Selbach), pp. 403–7. Urban und Schwarzbenberg, Munich.
33. Mundt, C. (1991). Endogenität von Psychosen—Anachronismus oder aktueller Wegweiser für die Pathogeneseforschung? *Nervenarzt*, **62**, 3–15.
34. Escobar, J.I., Gomez, J., and Tuason, V.B. (1983). Depressive phenomenology in North and South American patients. *American Journal of Psychiatry*, **140**, 47–51.
35. Lader, M. (1983). The psychophysiology of anxiety. *Encephale*, **9**, 205B–10B.
36. Kellner, R. (1988). Anxiety, somatic sensations, and bodily complaints. In *Handbook of anxiety* (ed. R. Noyes Jr, M. Roth, and G.D. Burrows), Vol. 2, pp. 213–37. Elsevier, Amsterdam.
37. Albus, M. (1990). Psychophysiologie der Angsterkrankungen. *Nervenarzt*, **61**, 639–46.
38. Stallone, F., Huba, G.J., Lawlor, W.G., *et al.* (1973). Longitudical studies of diurnal variations in depression: a sample of 643 patient days. *British Journal of Psychiatry*, **123**, 311–8.
39. Mundt, C., Richter, P., van Hees, H., *et al.* (1998). Zeiterleben und Zeitschätzung depressiver Patienten. *Nervenarzt*, **69**, 38–45.
40. Philipp, M., Maier, W., and Benkert, O. (1985). Operational diagnosis of endogenous depression. II. Comparison of 8 different operational diagnoses. *Psychopathology*, **18**, 218–25.
41. Jaspers, K. (1963). *General psychopathology* (7th edn) (trans. J. Hoenig and M.W. Hamilton), pp. 121–7. Manchester University Press.
42. Scharfetter, C. (1981). Ego-psychopathology: the concept and its empirical evolution. *Psychological Medicine*, **11**, 273–90.
43. Cotard, M. (1882). Nihilistic delusions. Reprinted in *Themes and variations in European psychiatry* (ed. S.R. Hirsch and M. Shepherd). John Wright, Bristol.
44. Lukianowicz, N. (1958). Autonoscopic phenomena. *Archives of Neurology and Psychiatry*, **80**, 199–220.
45. Jaspers, K. (1963). *General psychopathology* (7th edn) (trans. J. Hoenig and M.W. Hamilton), pp. 92–3. Manchester University Press.
46. Sims, A.C.P. (1993). Schizophrenia and permeability of self. *Neurology, Psychiatry and Brain Research*, **1**, 133–5.
47. David, A.S. (1990). Insight and psychosis. *British Journal of Psychiatry*, **156**, 798–808.
48. McEvoy, J.P., Apperson, L.J., and Appelbaum, P.S. (1989). Insight in schizophrenia: its relationship to acute psychopathology. *Journal of Nervous and Mental Disease*, **177**, 43–7.
49. David, A.S., Buchanan, A., and Reed, A. (1992). The assessment of insight in psychosis. *British Journal of Psychiatry*, **161**, 599–602.
50. Kemp, R. and David, A. (1996). Insight and psychosis: a social perspective. *Psychological Medicine*, **25**, 515–20.
51. McEvoy, J.P., Freter, S., and Everett, C. (1989). Insight and the clinical outcome in schizophrenia. *Journal of Nervous and Mental Diseases*, **177**, 48–51.
52. World Health Organization (1992). International statistical classification of diseases and related health problems, 10th revision. WHO, Geneva.
53. Slade, P.D. (1988). Body image in anorexia nervosa. *British Journal of Psychiatry*, **153** (Suppl. 2), 20–2.
54. Rosebush, P.I., Hildebrand, A.M., Furlong, B.G., *et al.* (1990). Catatonic syndrome in a general psychiatric population: frequency, clinical presentation and response to lorazepam. *Journal of Clinical Psychiatry*, **51**, 357–62.
55. Fink, M. and Taylor, M.A. (2003). *Catatonia: a clinician's guide to diagnosis and treatment*. Cambridge University Press.
56. Kahlbaum, K.L. (1874). *Die Katatonie oder das Spannungsirresein*. (trans. Y. Levis and T. Pridon 1973) Verlag August Hirshwald, Berlin. Johns Hopkins University Press, Baltimore.
57. Berrios, G. (1996). *The history of mental symptoms*. Cambridge University Press.
58. Lishman, W.A. (1997). *Organic psychiatry*. Blackwell Science, Oxford.
59. Jaspers, K. (1963). *General psychopathology* (7th edn) (trans. J. Hoenig and M.W. Hamilton), p. 138. Manchester University Press.
60. Lipowski, Z. (1990). *Delirium*. Oxford University Press, New York.
61. Sims, A. (2003). *Symptoms in the mind* (3rd edn), p. 41. W.B. Saunders, Edinburgh.
62. Ey, H. (1963). *La conscience*. Presse Universitaire de France, Paris.
63. Mayer-Gross, W., Slater, E., and Roth, M. (1955). *Clinical psychiatry*. Cassell, London.
64. Schneider, K. (1950). *Psychopathic personalities* (9th edn) (trans. M.W. Hamilton, 1958). Cassell, London.

1.8

Assessment

Contents

1.8.1. The principles of clinical assessment in general psychiatry

John E. Cooper and Margaret Oates

Introduction

This chapter is focused on the needs of the clinician in a service for general adult psychiatry, who has to carry-out the initial assessment of the patient and family, working either in the context of a multi-disciplinary team or independently. Within this quite wide remit, the discussion is limited to general principles that guide the practice of all types of psychiatry. The chapter does not include the special procedures and techniques also needed for assessment of children and adolescents, the elderly, persons with mental retardation, persons with forensic problems, and persons requiring assessment for suitability for special types of psychotherapy.

It is assumed that the reader has already had significant experience of clinical psychiatry and has completed the first stages of a postgraduate psychiatric training programme. Therefore details of the basic methods recommended in commonly used textbooks or manuals of instruction for obtaining and recording information on essentials such as the history, personal development, mental state, and behaviour of the patient are not included in this chapter.[1]

Three topics have been given special attention. These are assessment by means of a multi-disciplinary **team**, the trio of concepts **diseases, illness, and sickness**, and the development of **structured interviewing and rating schedules**. The first two have a special connection that justifies emphasis in view of the recent increase in multi-disciplinary styles of assessment. For instance, when different members of the team appear to be in disagreement about what should be done, it is usually a good idea to ask the question: 'What is being discussed—is it the patient's possible physical disease, the patient's personal experience of symptoms and distress, or the interference of these with social activities?' It will then often become apparent that the issues in question are legitimate differences in emphasis and priority of interest, rather than disagreements. The third topic is given prominence in order to illustrate some aspects of the background of the large number of such schedules (or 'instruments') that are now available. They are usually given the shortest possible mention in research reports, but since most advances in clinical methods and service developments come from studies in which an assessment instrument has been used, clinicians should know something about them.

The aim of the initial clinical assessment is to allow the clinician and team to arrive at a comprehensive plan for treatment and management that has both short-term and longer-term components. The achievement of this will be discussed under the following headings.

- Concepts underlying the procedures of assessment
- Contextual influences on assessment procedures
- Assessment as a multi-disciplinary activity
- Instruments for assessment
- The condensation and recording of information
- Making a prognosis
- Reviews
- Writing reports

Concepts underlying the procedures of assessment

The separation of form from content, and from effects on activities

In psychiatric practice more than in other medical disciplines, the key items of information that allow the identification of signs and symptoms of psychiatric disorders are often embedded in a mixture of complaints about disturbed personal and social relationships, together with descriptions of problems to do with work, housing, and money. These complaints and problems may be a contributing cause or a result of the symptoms of psychiatric disorders, or they may simply exist in parallel with the symptoms. A preliminary sorting out into overall categories of information is therefore essential.

The distinction between the form and the content of the symptoms is particularly important, together with the differentiation of both of these from their effects upon the functioning of the patient (function is used here in a general sense as applying to all activities, in contrast to the specific meaning given to it in the classification of disablements). This differentiation is discussed in Chapter 1.7, so only a brief mention is needed here.

The presenting complaint of the patient is often the interference with functions, but enquiry about the reasons for this should then reveal the contents of the patient's thoughts and feelings. The form of the symptoms (i.e. the technical term, such as phobia or delusion used to identify a recurring pattern of experience or behaviour known to be important) allows the identification of the psychiatric disorder. Knowledge of the effects on functions is essential for decisions about the management of patient and family, and is an important aspect of the severity of the disorder.

This sorting into different types of information often implies a conflict of priorities during the interview. The clinician must be seen to acknowledge the concerns and distress of the patient, but also must ask questions that will allow the identification of symptoms. Learning to balance this conflict of interest is an essential part of clinical training, and has been well recognized by previous generations of descriptive psychiatrists, including Jaspers. The separation of the social effects of a symptom from the symptom itself is also a necessary part of the assessment process. Further comments on this and related issues have been made by Post[2] and by McHugh and Slavney.[3]

Categories of information: subjective, objective, and scientific

Is there such a thing as a truly objective account of events? If 'objective' is intended to mean absolutely true and independent of all observers, the answer must be negative. Students and trainee psychiatrists often come to psychiatric clinical work from medical and surgical disciplines where they have been encouraged to 'search for the facts' with the implication that 'true' facts exist. They may need to be reminded that the supposed facts of all medical histories, even those of clearly physical illnesses, depend upon the perceptions, opinions, and memories of individuals who may give different versions of the same events at different times.

'Objective' has several shades of meaning in ordinary usage, but in clinical assessment it's most useful meaning is that an account of an event or behaviour is based on agreement between two or more persons or sources. In contrast, 'subjective' can be used to indicate that the account comes from only one person. Objective information is likely to be safer to act upon than subjective, so efforts should always be put into raising as much as possible of the information about a patient into the objective category. Nevertheless, many of the most important symptoms in psychiatry can only be subjective, since they refer to the inner experience of the one person who can describe them.

When assessing the reliability and usefulness of other types of information, such as the results of treatment or possible explanations of causes, a further useful distinction can be made between objective defined as above and 'scientific', taking this to mean that systematic efforts have been made to obtain evidence based upon comparisons (or 'controls') which demonstrate that one explanation can be preferred out of several possibilities that have been considered.

Simple definitions such as these are useful in clinical discussions, but it must be remembered that in the background are many complicated and unsolved problems of philosophy and semantics. Some of these suggestions on the status of information in clinical work are based upon the writings and clinical teaching of Kraupl Taylor.[4]

Disease, illness, and sickness

These concepts have existed in the medical and sociological literature for many years, and are best regarded as useful but inexact concepts that refer to different but related aspects of the person affected, namely pathology (disease), personal experience (illness), and social consequences (sickness), respectively.[5] They are useful as a trio because they serve as a reminder that all three levels should be considered in a clinical assessment, even though for different patients they will vary greatly in relative importance. There are no simple answers to questions about how they are best defined and how exactly they are related to each other, but time spent on these issues is not wasted because they reflect quite naturally some of the different interests and priorities of the different health professions (and are therefore often the basis of different viewpoints put forward by various members of a multi-disciplinary team).

Another reason for being familiar with these concepts is that in legal and administrative settings, simple and categorical pronouncements about the presence of mental illness or mental disease and their causes and effects may be required whatever the medical viewpoint might be about the complexity of these concepts.

Clinicians of any medical discipline know from everyday experience that the complete sequence of disease—illness—sickness does not apply to many patients. Although disease usually causes the patient to feel ill and the state of illness then usually interferes with many personal and social activities, in practice there are many exceptions. Potentially serious physical, biochemical, or physiological abnormalities (disease) may be discovered in surveys of apparently healthy persons before any symptoms, distress, or interference with personal activities (illness) have developed, and some patients may have either or both of illness and sickness (interference with social activities) without any detectable disease.

A number of sociologists, anthropologists, and philosophers have joined psychiatrists in trying to define mental illness and mental health, but without achieving much clarification. Aubrey

Lewis[6] and Barbara Wootton,[7] although writing from the different contexts of clinical psychiatry and sociology, both arrived at the conclusion that neither mental illness nor mental health could be given precise definitions, although they are useful terms in everyday language (and the same applies equally to physical health and physical illness).

More positive conclusions have resulted from attempts to define disease, in that Scadding (a general physician) has suggested that it should be defined as an abnormality of structure or function that results in 'a biological disadvantage'.[8,9] This seems reasonable if one is dealing only with conditions that have a clear physical basis, but if applied in psychiatry it implies that, for instance, behaviours such as homosexuality that reduce the likelihood of reproduction would have to be regarded as diseases alongside infections, carcinoma, and suchlike. This seems to be stretching a traditional concept too far, and different approaches clearly need to be explored.

One way forward is to accept that simple definitions and concepts encompassed by one word cannot cope with complicated ideas such as disease or health, and to take care to differentiate between definitions of these as concepts in their own right, and attempts to develop **models of medical practice**. The debate noted above refers to concepts of health, disease, and disorder, and it has been continued more recently with respect to psychiatry in two quite extensive reviews, in terms of the types of concepts,[10] and of their possible contents.[11] What follows below is better regarded as about *models of medical practice*, and two points are suggested as a basis for the discussion. First, more than one dimension or aspect of the person affected always needs to be included in descriptions of health status. Second, models of medical practice and thinking do not necessarily have to start with the assumption that physical abnormalities (diseases) are the basic concept from which all others are derived.

Regarding the first point (of more than one aspect or dimension), soon after the contribution of Susser and Watson[5] noted above, Eisenberg, a psychiatrist with social and anthropological interests,[12] made a plea for all doctors, but particularly psychiatrists, to recognize the importance of appropriate illness behaviours in addition to giving the necessary attention to the diagnoses and treatment of serious and dangerous disorders.[13] He gave special emphasis to the need to minimize problems that may arise from discrepancies between disease as it is conceptualized by the physician and illness as it is experienced by the patient: 'when physicians dismiss illness because ascertainable disease is absent, they fail to meet their socially assigned responsibilities'. A similar model with a more overtly three-dimensional structure usually referred to as 'bio-psychosocial' has also been described by Engel.[14] and also by Susser.[15] Historically, all these can be regarded as variations on and explicit developments of a theme that has been accepted implicitly by generations of psychiatrists influenced by the 'psychobiology' of Adolf Meyer and his many distinguished pupils, manifest in the importance given to the construction of the traditional clinical formulation.

The second point, to do with the disease level not being the best starting point for conceptual models of medical practice, is of more recent and specifically psychiatric origin. Both Kraupl Taylor[16] and, more recently, Fulford[17] give detailed arguments for the conclusion that the illness experience of the patient is the most satisfactory starting point from which to develop a model of medical practice. Taylor presents his case as a matter of logic, and Fulford works through lengthy philosophical and ethical justifications. This new viewpoint has the virtue of starting with the encounter between patient and doctor, which has the strength of being one of the few things that is common to all types of clinical practice. In Taylor's terms, by describing symptoms and distress the patient arouses 'therapeutic concern' in the doctor and so first establishes 'patienthood'. Whether or not a diagnosis is reached or a disease is later found to be present, and whether or not the social activities of the patient are also interfered with, are other issues of great importance, but they do not diminish the primary importance of the first interaction; in this, both patient and doctor play their appropriate roles according to their personal, social, cultural, and scientific backgrounds.

If medical training and practice are guided by this model, there is no interference with the essential obligation of the doctor to identify and treat any serious disease that may be present. However, a parallel obligation to satisfy the patient and family that the illness (comprising complaints and distress) and the sickness (interference with activities) have also been recognized and will be given attention, is equally clear.

How to answer questions by the patient and family about whether the patient has a mental illness or not, and what this implies, needs careful discussion. Within a multi-disciplinary team it is usually best for the team to reach early agreement on a particular way of describing the patient's illness so that conflicting statements will not be made inadvertently by different members if asked about it. This is because the patient or family may expect this type of statement, and not because distinctions between, for instance, mental illness and physical illness, or between nervous illness and emotional upset, are regarded as fundamental from a psychiatric viewpoint. This difficult issue will be made easier if something about the patient's ideas about the nature and implications of terms such as 'mental illness' and 'nervous breakdown' is always included as part of the initial assessment information. Similarly, all members of the team need to be familiar with the concept of illness behaviour and the way this is determined by cultural influences[18] (see Chapter 2.6.2).

The diagnostic process: disorders and diagnoses

Psychiatrists learn during their general medical training that the search for a diagnosis underlying the presenting symptoms is one of the central purposes of medical assessment. This is because if an underlying cause can be found, powerful and logically based treatments may be available. But even in general medicine, as Scadding pointed out 'the diagnostic process and the meaning of the diagnosis which emerges are subject to great variation . . . the diagnosis which is the end-point of the process may state no more than the resemblance of the symptoms and signs to a previously recognized pattern'.[8,9] In psychiatry, 'may' becomes 'usually', and this has been recognized by the compilers of both ICD-10 and DSM-IV, in that these are presented not as classifications of diagnoses, but of disorders. These classifications use similar definitions of a disorder; the key phrases in ICD-10 are 'the existence of a clinically recognizable set of symptoms or behaviour associated in most cases with distress and with interference with personal functions', and in DSM-IV 'a clinically significant behavioural or psychological syndrome or pattern that occurs in an individual and that is associated with present distress or disability . . . '.

The use of such broad definitions is necessary because of the present limited knowledge of the causes of most psychiatric disorders, and a similarly limited understanding of processes that underlie their constituent symptoms. To avoid overoptimistic assumptions, there is much to be said for psychiatrists avoiding the use of the term 'diagnosis' except for the comparatively small minority of instances in which it can be used in the strict sense of indicating knowledge of something underlying the symptoms. A consequence of this viewpoint is that the currently used 'diagnostic criteria' in both these classifications should be relabelled as 'criteria for the identification of disorders'.

In spite of this, it must be accepted that the patient and family are likely to expect statements to be made about the cause of their distress and symptoms. The members of all human groups expect their healers to discover the causes of their misfortunes (i.e. to make a diagnosis), and to provide remedies. This is so whether the group is a sophisticated and scientifically oriented modern society, or a non-industrialized society that relies on ethnic healers and folk remedies. The obvious relief of a patient or family on the pronouncement of an 'official' diagnosis is often evident in any type of healing activity, even though the diagnostic terms themselves mean very little. The pronouncement of an official diagnosis is taken to show that the doctor knows what is wrong, and therefore will be able to provide successful treatment or advice. If the diagnosis is expressed in terms that the patient can understand, it will have additional power as an explanatory force.

The readiness of ethnic healers and practitioners of complementary (or alternative) medicine to provide a diagnosis and treatment in terms that have a meaning and therefore a powerful appeal to their customers is probably one of the main reasons for their continued survival and popularity alongside scientifically based medicine. This is a separate issue from the question of whether or not the treatments of complementary practitioners are successful in the sense of having effects that could be demonstrated by means of a controlled clinical trial.

Within psychiatry and clinical psychology, the medical habit of searching for a diagnosis has at times been misunderstood as an unjustified preoccupation with the presence of physical disease as a cause of mental disorders. This was most marked in the United States during the 1950s and 1960s, expressed particularly in the writings of Menninger in which the diagnostic process and attempts to classify patients were dismissed as a waste of time.[19] This viewpoint ignores two points made here and by many others; first, the choice of a diagnostic term is only one part of the overall process of assessment that leads also to a personal formulation. Second, any assessment of a person, whether made as statements about psychodynamic processes, as statements about structural and biochemical abnormalities, or as statements about interference with activities, is unavoidably an act of classification of some sort.

More detailed discussions about the importance of diagnosis have been provided by Scadding[8] as a general physician, and by Kendell[20] and Cooper.[21] as psychiatrists.

Concepts of disablement

Disablement will be used here as an overall term to cover any type of interference with activities by illness. This is often of more concern to the patient than the symptoms of the illness itself, since the fear of long-term dependence upon others is usually present, even though not voiced in the early stages. The question arises whether to leave the description and assessment of disablement to different members of the team as it arises in various forms, or whether in addition to encourage reference to one of the systematic descriptive schemes that are now available. Even if not used as fully as their authors intend, these have the merit of serving as checklists or reminders for the whole team, to ensure that the many different effects of the illness have been considered.

Two widely used descriptive schema are the *International Classification of Functioning, Disability and Health* (*ICF*),[22] and a broadly similar framework described by Nagi[23] that is often used in the United States, particularly by neurologists. These are best regarded as descriptive conceptual frameworks rather than classifications, sharing a basic structure of several levels of concepts. For the ICIDH, these are **functioning**, **disability**, and the **contextual factors (both environmental and personal)** relevant to what is being assessed. These terms and concepts are defined in the manuals published by WHO Geneva. The ICF is published as both a short and a long version, and it is probably wise for interested users to start by examining the short version. As noted in Fig. 1.8.1.1, these three concepts can be put alongside the sequence of ideas that leads from complaints, through symptoms to the identification of disorders or diagnoses. This may represent a causal sequence in some individuals, and this is clearest in acute physically based illnesses. But for many patients encountered in psychiatric practice, whose illnesses often have prominent social components, causal relationships may be absent or even in the opposite direction. For instance, sudden bereavement, i.e. loss of a social relationship, may be the clear cause of interference with the ability to perform daily activities (disability), and also of uncontrolled weeping (an impairment of the normal control of emotions). Social handicaps can also be imposed unjustifiably by other persons, as when a patient who is partly or fully recovered from long-standing psychiatric illness and quite able to work is refused employment due to the prejudice of a potential employer.

Many mental health workers find that to use a scheme such as Fig. 1.8.1.1 or the ICF helps to clarify how different aspects of a patient's problems fit together. Similarly, the different members of the team may be able to see more clearly how their activities with the patient and family complement one another, since the different concepts in the framework correspond approximately with the interests of different health disciplines. Social workers will focus on assessment of work and social relationships, occupational therapists will have a special expertise in the assessment of daily activities, and clinical psychologists are skilled in the assessment of cognitive and other psychological functions. Researchers in the various health disciplines have naturally devised rating scales that reflect their own interests and ideas, independently of the ICF or other overall schema, but it is usually found that such scales correspond quite closely with one or other of the concepts just discussed. The reluctance of both researchers and clinicians to adopt a standard set of terms to cover the various levels or concepts continues to be a problem; the reader needs to be aware that the terms impairment, disability, and handicap are often used synonymously by different authors.

The description of social and interpersonal relationships is in principle included in comprehensive schemas such as the ICF and that of Nagi, but many separate instruments that cover relationships in great detail have been devised over the years by psychotherapists, family therapists, and others.[24–26]

The sequence of assessment: collection, analysis, synthesis, and review

As information accumulates and is discussed, several different but related aspects of the patient and the illness have to be kept in mind. Good psychiatric practice is a part of what is sometimes referred to as 'whole-person medicine' in which at different times the contrasting but complementary processes of both analysis and synthesis of the information available will be needed. The patient must be seen both as an individual with a variety of attributes, abilities, problems, and experiences, and as a member of a group that is subject to family, social, and cultural influences; at different stages in the process of assessment each of these aspects will need separate consideration.

Analysis is needed to identify those attributes, experiences, and problems of the patient and the family that might require specific interventions by different members of the team. This must then be followed by several types of synthesis (or bringing together of information) to enable attempts to understand both subjective and objective relationships between the patient and the illness. First, possible interventions must be placed in order of priority for action. Second, the whole programme needs to be reviewed at intervals so as to assess progress and decide about any additional interventions that are required. At these times of review, and particularly towards the end of the whole episode of illness, global statements about 'overall improvement', or changes in 'quality of life' may be additional useful ways of summarizing and evaluating what has been happening from the viewpoint of the patient.

From complaints to formulation

Figure 1.8.1.1 demonstrates how the information contained in the complaints presented by the patient needs to be sorted out into different conceptual categories so that it can form the basis of actions by the various members of the multi-disciplinary team.

The top box represents the complaints. Unpleasant symptoms are likely to head the list, but inability to do everyday activities or a description of problems with relationships may well come first. Symptoms that give a clue to disorders, diagnoses, and possible treatments may not be identified without close questioning by someone who knows what to ask about.

The second box indicates that the complaints need to be sorted out into symptoms and impairments (an impairment in this sense is interference with a normal physiological or psychological function, as explained below). Some complaints are both symptoms and impairments: symptoms because it is known that they can contribute towards the recognition of an underlying diagnosis or towards the identification of a disorder, and impairments because they indicate measurable interference with the function of a part of the body or of a particular organ. For instance, inability to remember the time of the day is a symptom (disorientation in time) that may contribute towards a diagnosis of some kind of dementia. It is also an impairment of cognitive functioning that is likely to interfere with the performance of everyday activities such as getting up and going to bed at the correct time, and organizing housework.

The left-hand side of Fig. 1.8.1.1 represents the progress towards the identification of a disorder and perhaps even an underlying diagnosis. These are important concepts because they may indicate useful treatments and likely eventual outcomes. The right-hand side shows the progression from impairment of functions of parts of the body or organs, through interference with personal and daily activities to interference with participation in social activities.

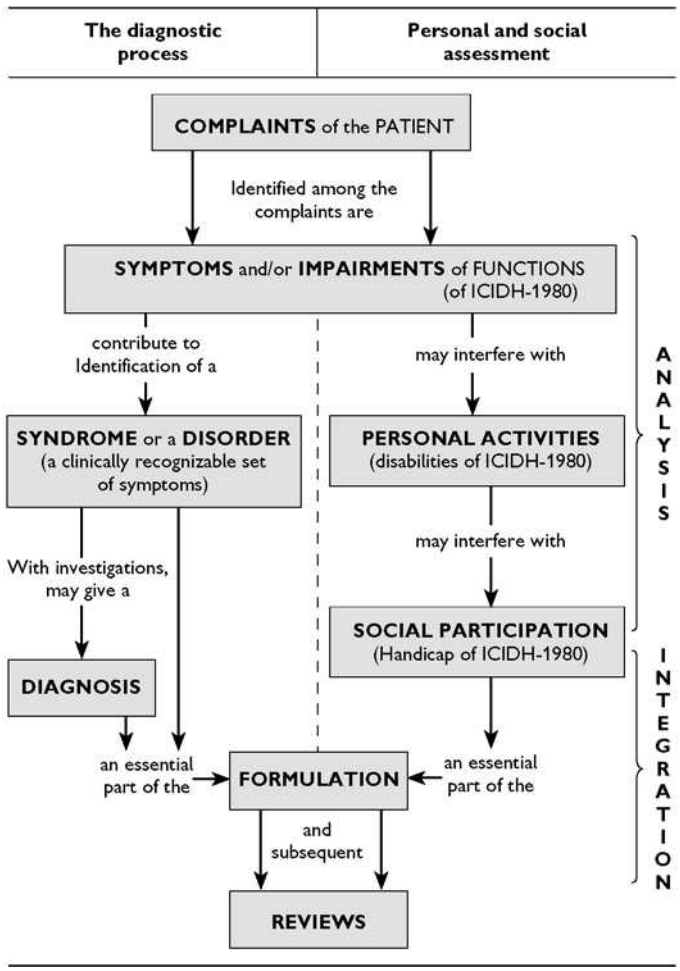


Fig. 1.8.1.1 Analysis and integration of information.

A clinical assessment is not complete until all the components of both sides of Fig. 1.8.1.1 have been considered. In doing this, the different components and the two pathways of concepts will need to be given widely varying emphasis for different patients, and also for the same patient at different times. For instance, if there is a physical cause for a disturbance of behaviour, an accurate diagnosis of this will lead to the best possible chance of rapid and successful treatment. In contrast, if a disturbance of social behaviour has its origin in personal relationships or has been imposed upon the patient by the social prejudices of others, the correct diagnostic category is unlikely to add much. An assessment of social networks and supportive relationships will be more relevant to deciding upon useful actions.

Life events and illness

For clarity, the right-hand side of Fig. 1.8.1.1 is given in a very compressed form, but in practice it is likely to need dividing into several components. The possibility of discovering relationships in time between life events and the onset of symptoms or interference with activities, particularly if repeated, should always be kept in mind, since this may be relevant to management plans and the

assessment of prognosis. The best guide to this will be a lifechart. The opinions of patients and families about the causes of illness must be listened to with respect, while bearing in mind that the attribution of illness to the effects of unpleasant experiences is a more or less universal human assumption that often has no logical justification. Clinicians have to arrive at their own conclusions about such relationships by means of experience, common sense, and some acquaintance with research findings. Researchers seeking robust evidence on this topic are faced with a very difficult task, since the assessment of vulnerability to life events is a surprisingly complicated and controversial issue. The leading method in this field is the Life Events and Difficulties Schedule developed by Brown and Harris; a bulky training manual has to be mastered during a special course, and this then serves as a guide to an interview which may last for several hours. The length and detail of these procedures illustrate well the technical and conceptual problems that have to be faced.[27–29]

Psychodynamics and the life story

'Psychodynamics' refers in a general sense to the interactions between discrete life events, personal relationships, and personality attributes, in addition to its use to cover internal psychological processes (such as defence mechanisms and coping strategies). All of these need to be examined when trying to understand which of several possible causes of an illness at a particular time is the most likely.

A mixture of knowledge about local social and cultural influences and more technical psychological issues is needed for this appraisal of the patient's life story, and suggestions about different components of the overall pattern may well come from different members of the team.

The internal psychodynamics of the patient often need to be considered in detail, and one way to do this would be to construct a subdivision of the right-hand side of Fig. 1.8.1.1 to show interpersonal relationships and psychodynamic processes. In some patients a major conclusion of the initial assessment will be that these aspects are paramount, indicating the need for referral to a specialist psychotherapy service. The assessment of suitability for specific forms of psychotherapy and cognitive behavioural approaches are dealt with in section 6.

Contextual influences on assessment procedures

The place of assessment should not be regarded as automatically fixed in the outpatient or other clinical premises. One or more assessment interviews at home should be considered,[30] since the patient and family may feel much more at ease and therefore likely to express themselves more freely in familiar surroundings, but with the proviso that privacy may be more difficult to achieve. The assessor will often be surprised how much useful information about the home and family circumstances is gained from an interview at home, even when there appeared to be no special reason for this at first. In addition, the behaviour of both the patient and family members in the clinic or hospital is often different from that observed in familiar home surroundings. There are also obvious advantages to both assessment and care at home for mothers who have psychiatric disorders in the puerperium.[31]

Interviews on primary care premises are also often appreciated by patients who dislike going to hospitals of any sort, and the ease of consultation with the general practitioner is an additional advantage. The adoption of regular visits by a consultant psychiatrist to primary care premises as a major element in cooperation between psychiatrists and general practitioners is a style of work that seems to be spreading, with advantages to all concerned.[32]

Privacy of interviewing and **confidentiality** of what is discussed needs careful consideration; there are few absolute rules, but the following points of procedure should be explained clearly to both patient and relatives from the start. First, the patient and any member of the family should know that if they wish they are entitled to speak to the doctor in private, and they must be able to feel that what they say will not be conveyed to any other member of the family unless they request this. Second, in addition to the usual rules of professional secrecy, the patient must agree not to question other family members about what they said to the doctor, and vice versa. These may seem to be elementary points to trained professionals, but they are often not appreciated by patients or relatives who may be in fear of each other, or at least apprehensive about the reaction of the other on learning that statements they might construe as critical have been made about them. These are all points by which trust is established and maintained between patient and doctor, and for the same reason any attempts by relatives to seek interviews on condition that the occasion is kept secret from the patient should be firmly resisted.

An interpreter should always be sought if the patient cannot speak fluently in the language of the interviewer. Mental health professionals who can also act as interpreters are increasingly available nowadays due to the presence of almost all communities of sizeable ethnic minorities. Because of the issues of confidentiality noted above, a professional of the same sex as the patient should always be preferred to family members when interpretation is needed.

Language barriers are usually, but not always, accompanied by a cultural difference. The interviewer must remember that the concept of a private interview between two strangers in which personal and often unpleasant events and experiences are discussed freely comes from 'middle-class western' culture, and is not necessarily shared by persons from other cultures. A discussion of this point before the interview with a mental health professional familiar with the patient's background will help the interviewer to determine what to aim at in terms of intimate or possibly distressing information.

Multiple sources of information are always an advantage for those topics (mainly events) for which objective accounts are possible. Clinical experience is the best guide as to which account to use when conflicts of information arise. Serious conflicts of information arising during the initial assessment that involve the patient's account of events are best resolved by trying to obtain more information. Confrontation of the patient with important conflicts of information should be avoided since it easily leads to misunderstandings. If done at all, confrontation should be reserved for later stages in the overall management when it forms part of a planned intervention with a special purpose.

Assessment as a multi-disciplinary activity

Multi-disciplinary teams can take many forms, varying from the tightly organized and necessarily hierarchical surgical theatre team in which the role of each member is clearly defined and unchanging, to loosely knit groups in other types of health service in which

only some of those attending meetings about patients regard what is taking place as a team event. For the purposes of discussion of the types of multi-disciplinary teamwork increasingly to be found in the mental health services, it is useful to differentiate between multi-disciplinary practice, familiar to many generations of mental health workers, and the more recently evolved multi-disciplinary teamwork. Both of these styles of work have many variations, but they both have some key features that need to be recognized by those involved.

In **multi-disciplinary practice** the consultant or most senior doctor present at clinical meetings or 'ward rounds' is accepted by all as the leader of the group, and listens to (and usually depends upon) the views of the senior nurses and other health professionals who may or may not be present. But the decisions about treatment and management are clearly acknowledged to be the responsibility of the doctors present. In most settings the only essential attendees at these meetings are the doctors and nurses; attendance of other health professionals is usually welcomed and valued, but they are not regarded as necessary members of the group.

Multi-disciplinary teamwork has probably developed in response to a marked increase in the number of social workers, occupational therapists, clinical psychologists, and others, in those medical services in which patients and families with multiple needs are the rule rather than the exception. Clinical skills and techniques that were not previously available are now available, and the health professionals offering these expect quite naturally to be given increased personal and professional recognition; this can usually be found as a member of a multi-disciplinary team of the sort described here. The most fully developed style of multi-disciplinary teamwork involves a very significant commitment of professional time by each member so that all the team meetings can be attended, in addition to the time spent directly with the patient and family.

Some sharing of responsibilities and blurring of roles is needed, but each member also must be seen as retaining the professional skills of their parent discipline. Role blurring is most obvious in the information-gathering and information-sharing phases of assessment, and in the team discussions that lead to agreement about the content of a programme of activities.

Leadership

The concept of a team implies that a team leader is recognized, but a leader does not have to be an obviously dominant speaker and decision maker. Many successful team leaders 'lead from behind' to great effect and to everyone's satisfaction. The main reasons for having an agreed leader are, first, to keep discussions acceptably brief and to a practical timetable; second, to facilitate decisions between reasonable alternatives; and third, to arbitrate when insoluble disagreements arise between team members.

There are a number of settings within the mental health services in which there may be no need for the leader for everyday purposes to be a doctor. This occurs most frequently in special crisis intervention and emergency units, in rehabilitation units, and in services for those with mental retardation. However, members of such teams have to acknowledge that decisions about the presence of physical illness, and the need for medication or laboratory investigations, can only be made by a medically qualified person. The team then has to accept the authority of the doctor on these occasions because of the unique ethical and legal responsibilities that accompany a medical qualification.

In teams running an inpatient unit such as an acute admission ward, there is a clear need for the whole team to accept that medical and nursing members have free access to all the patients for purposes of physical examination, laboratory investigations, the administration of medication, and a variety of nursing procedures.

Key workers and the planning of care

The allocation of a team member as key worker (or case manager) for each patient being assessed is the usual method of work in teams of this sort. Which member becomes key worker for which patient depends upon the ability of the team to match the needs of each patient with the skills available amongst the team members, according to their training. Although all patients are discussed in detail at team meetings, and any team member can contribute suggestions and viewpoints, it is usually accepted that once a programme of activities is identified and agreed upon, most if not all of the contacts with the patient and family will be made by the key worker. The key worker also has the responsibility of reporting back to the team about progress, and about problems encountered which might require new major decisions or changes in the programme.

In the United Kingdom, due to yet another recent reorganization of primary care and hospital services, the situation with respect to urgent assessments and psychiatric emergencies has recently become complicated. Fundamental structural changes are taking place as a result of administrative and financial pressures rather than because of evidence from studies of the previous pattern of services. There is, however, now an element common to all areas in that the provision of a written plan for the care proposed is now a statutory responsibility for all patients. Further changes seem likely, and it is beyond the remit of the general principles of assessment described here to try to describe the present situation in the United Kingdom in detail. Reviews by Burns[33] and others[34] are very helpful guides through the complexities and terms used to cover some of these developments in the United Kingdom.

In addition to the specific medical responsibilities noted above, psychiatrists as members of a multi-disciplinary team have other important areas of expertise that should be recognized by the other members. Experienced psychiatrists are likely to have special skills in the assessment of dangerousness and risks of various sorts, and psychiatrists at any stage in their training should be able to show that they are specially trained to summarize information by the production of an overall *formulation* that reflects the agreed policies of the team.

To be an efficient and accepted long-term member of a multi-disciplinary team of this type requires personal characteristics not necessarily possessed by all mental health professionals. Tolerance of the different viewpoints of other team members is essential, in addition to the professional skills needed to carry-out the work required.

The frequency of team meetings is determined by the size and nature of the workload. Special meetings to discuss topics not directly related to the patients are also usually found to be necessary, so as to deal with issues such as team policies, recruitment and appointments, relationships with outside agencies (for instance about too few, too many, or inappropriate referrals), interpersonal problems between team members, and work-related stress in the team members. This last problem is particularly important in teams dealing with crisis intervention and psychiatric emergencies

because of the need to maintain a rapid turnover of patients and families who are seen over only a limited period of time.

A different type of problem that may need sensitive handling by the team leader and others in authority outside the team itself is the relationship between the team members and their immediate superiors (or 'line managers') in the hierarchy of their own discipline. Each team member has to strike a balance between personal needs for professional supervision and training, and the ability to make decisions within the team because of special skills not possessed by other team members. This type of problem will be minimized if team members are comparatively senior and experienced within their parent discipline. Student health workers are not appropriate as team members, but they can benefit greatly if attached to the team as observers. They will have the opportunity to learn something about how other disciplines operate, which is an aspect of training usually absent from the rest of their training.

Disagreements often arise within a team about the best time for patients to be discharged from care, or about the precise time for referral when it is in the patient's interests to be assessed by another service. In countries where outpatient services and inpatient services are staffed by different teams under different organizations, there will be many such breaks in care, and multi-disciplinary teamwork can become frustrating. But where continuity of care between different parts of the general psychiatric services is the norm, the most frequent changes of care result from the need for the patient to be assessed for more specialized treatment such as rehabilitation, cognitive behavioural therapy, or intensive psychotherapy. The team needs to develop agreed policies for these occasions, and these will depend largely upon the structure of the local services available.

Although no systematic information is available, there is little doubt that the style of multi-disciplinary teamwork just described has become accepted in the mental health services in many countries. Its popularity and success are probably due to the recognition of multiple rather than single needs in a large proportion of psychiatric patients and families, plus an increased job satisfaction experienced by the non-medical team members. Multi-disciplinary styles of working are especially important in emergency psychiatric services and crisis intervention units.[35,36]

Instruments for assessment

Reasons for the development of structured interviewing and rating instruments

The training of all mental health professionals includes instruction in some system of information-gathering and recording based upon a conceptual structure that helps them to organize the large amount of information they usually need to collect. With training, the list of headings under which this information is collected becomes incorporated in the professional's mind as an automatically available guide to the conduct of assessment interviews. For research purposes, however, it is necessary to demonstrate overtly that the essential topics have been covered in a comprehensive and systematic manner. In many types of research not only the headings covered but the detailed items also need to be recorded so that others studying the results of the research can be confident that nothing was missed, and that the information obtained was not a biased selection of the total that might have been available. It has also been generally recognized since the 1950s that for purposes of communication between researchers in different centres, conclusions must be based upon information that has been shown to have a satisfactory inter-rater reliability.

With these aims in mind, detailed and comprehensive structured interviewing and rating schedules for recording many varieties of information have been developed (nowadays these are usually called 'instruments'; for brevity this term will be used to cover any sort of published interviewing and rating schedule). The most common types cover the present mental state and behaviour. Most of these instruments are not appropriate for use in everyday clinical work because they have been designed for research studies, but nevertheless it is useful for clinicians to know something about how they originated.[37]

Since the first appearance of partly or fully structured psychiatric rating instruments in the 1950s, there has been a steady increase in their number, type, and complexity. In the discussion that follows, some of the most widely used instruments are commented upon as examples but many others are available that are not mentioned. Comprehensive lists of such instruments can be found in catalogues of instruments and reviews by the WHO and others.[38–42]

A word of warning is needed about the use of these instruments in ordinary clinical work. Any of them can be used as useful checklists by clinicians to improve the range of information collected. But this does not mean that the quality of information recorded will necessarily be high. Most instruments were originally designed for use in research studies in which they were used by researchers specially trained in their use, and occasional use by untrained staff will not produce information of the same quality and usefulness.

Instruments for the assessment of mental state and behaviour

The instruments now available can be grouped according to the main purposes for which they were designed.

Screening instruments such as the General Health Questionnaire[43] are needed for the identification of likely cases or high-risk individuals amongst large populations. These tend to be short and economical in use, since they have to be administered to large numbers of subjects. They are designed to generate a simple score that indicates the status of the subject in relation to the populations upon which the instrument was developed and validated. This is essential for screening and for epidemiological studies, but this single score does not convey much about the details of the subject's feelings or behaviour, and so is of limited interest to the clinician.

Screening instruments are often questionnaires, defining this to mean that they are simply a means of recording the answers to a set of questions, without any further questions or enquiry about the extent to which the subject understands the question or wishes to qualify answers given. Questionnaires are usually filled in by the subject as a 'paper-and-pencil' exercise, as in the General Health Questionnaire, but one widely used questionnaire that has a very detailed content (the *Composite International Diagnostic Interview*[44,45]) is completed by an interviewer.

Detailed instruments may contain the following:

1 Symptoms of only one type, as in Hamilton's rating scale for depression,[46] or the Scale for Assessment of Negative Symptoms.[47]

2 A selection of symptoms for the study of the relationships between two closely related types, such as depressive and schizophrenic symptoms in the Schedule for Affective Disorders and Schizophrenia.[48]

3 A limited number of items covering different symptoms and behaviour selected as being of special importance, as in the recently developed Health of the Nation Scales of the United Kingdom.[49]

4 A more or less comprehensive array of symptoms that allows the study of the relative distribution of symptoms of many different types, such as Schedules for Clinical Assessment in Neuropsychiatry[50] and the Composite International Diagnostic Interview.[44,45] Other widely used but less tightly structured instruments with a comprehensive content are the Brief Psychiatric Rating Scale[51] and the Comprehensive Psychopathological Rating Scale,[52] aimed at measuring change.

The source and method for collection of information is usually specified by the designers of an instrument. These can include interviews with patients, relatives, and carers, observation of the patient, extracts from other documents, and any combination of these.

The more detailed instruments usually depend upon an interview, and the style of interviewing recommended and the training needed to achieve this depend upon both the quality of the information required and the type of research interviewer for whom the instrument is designed. These vary widely; for instance, Schedules for Clinical Assessment in Neuropsychiatry require a clinical professional training plus a special course for the instrument itself; the Comprehensive Psychopathological Rating scale, the Brief Psychiatric Rating scale, and the *Structured Clinical Interview for DSM*[53] assume a clinical professional training only. The Composite International Diagnostic Interview requires experience in interviewing such as market research plus a special course for the instrument itself, but no clinical professional training.

The time period covered varies from a cross-sectional picture of the present mental state and behaviour ('present' usually being taken to mean the immediately recent period of 2 or 4 weeks), to longer periods of follow-up, personal history, and development, and lifetime histories of psychiatric disorders. The more complicated and lengthy instruments that cover these longer periods are usually designed for particular studies, so are rarely suitable for general use.

Developments since the 1950s

A historical approach is helpful in trying to understand how and why the many instruments now available have developed. Hamilton's Rating Scale for Depression, published in 1959, is a good example of the first generation of instruments, most of which are comparatively short and simple.[46] Its contents can easily be printed on one page, and comprise the following:

1 the names of the symptoms to be rated

2 a rating scale, the same for all the symptoms, by which the presence or absence and the severity of each symptom is recorded

3 a box in which the rating of each symptom is placed

No special recommendations about length and style of interview are given, and no explanations or definitions of the symptoms are given other than what is provided on the rating sheet itself. In other words, the interpretation of the ratings is based on the assumption that the raters have sufficient experience and training to know what most of their contemporaries also mean by the named symptoms. Data analysis is left to the user, other than recommendations about the likely meaning of the sum of the ratings with respect to severity of illness and 'caseness'. This and other early instruments were not tied to the use of any particular set of diagnostic categories, probably because the diagnostic classifications that were available in the 1950s and 1960s were not widely used.

The first generation of instruments made it much easier for researchers to communicate the detailed results of their clinical studies to others, mainly by facilitating the study of changes in symptoms over comparatively short periods of time. The need for this was no doubt connected with the increasing numbers of psychotropic medicines that became available around that time. Measurement of change in symptoms is more immediately useful for the study of response to treatment than reliance upon statements about overall improvement or waiting for a change in diagnosis. But in the absence of guidance about how the symptoms are defined, problems still remain in the interpretation of the results.

Improvements in more recent instruments leading to better quality and meaning of the data they collect have been of two main types, in that the structure and the associated procedures of the instruments have become more elaborate as time has passed. First, the input has been improved by the provision of written descriptions and definitions of symptoms, and by recommending particular styles of interviewing. This implies that researchers using the instrument should carry-out preliminary training work so that satisfactory levels of inter-rater reliability are achieved before starting the main study. Second, the output has been improved by the use of computers to organize and summarize the symptom ratings, allied with the development of widely used psychiatric classifications.

Computer programs based upon decision trees (algorithms) first appeared in the 1970s, and are now commonplace. They allow the specification of sets of symptoms that identify disorders or indicate diagnoses, so that the resulting statements about symptom profiles or the presence of disorders or diagnoses are free from errors of human judgement such as carelessness, simple forgetting, and personal variations from one occasion to the next. But the biases and assumptions built into the programs by their authors still remain, and these may be a problem to others with different opinions.

Programs can also be written to assign disorders and diagnoses according to a selected classification, such as ICD-10 or DSM-IV, and some of the most recently developed instruments such as Schedules for Clinical Assessment in Neuropsychiatry and the Comprehensive International Diagnostic Interview are of this type. When used as intended, the data output from these more recent structured instruments is versatile and of high reliability, but to obtain these benefits the researcher has to pay the penalty of working hard to achieve and to maintain inter-rater reliability.

There are, of course, still plenty of uses for the simpler types of instruments; it is up to those designing and carrying out a study to decide what type of information they need and why, and to select their instruments accordingly. For the sake of those who will be interested in trying to interpret the results, a justification of the

quality of the information obtained should always be included in the description of the findings.

Once an instrument (or often a related group of instruments) has demonstrated its usefulness it is likely to stay in use for many years, while at the same time being subject to extensions and improvements. Families of instruments and traditions of interviewing style therefore develop and persist in the major research centres and groups, and it is possible to identify some of these and follow them over the years.

Three such traditions of instrument development are selected for mention so as to illustrate the continuity and close relationships that sometimes exist between different instruments; these relationships may not be apparent from reports of studies in which they have been used. Three research centres that have produced particularly prominent sets of instruments are the Medical Research Council Social Psychiatry Unit at the Institute of Psychiatry in London, Biometrics Research at the New York State Psychiatric Institute at Columbia University, New York, and the Department of Psychiatry at Washington University, St Louis, Missouri. The instruments mentioned below are only a small proportion of the many in the literature, but they are well known because of their association with some large collaborative international research studies and with widely used classifications of psychiatric disorders such as ICD-8, ICD-9, and ICD-10, and DSM-III, DSM-IIIR, and DSM-IV.

At approximately the same time in the early 1960s, but independently, research groups headed by John Wing, at the Institute of Psychiatry of the University of London (at the Maudsley Hospital), and by Robert Spitzer, at the Biometrics Research Unit at the New York State Psychiatric Institute at Columbia University, began to produce structured interviewing and rating schedules that provided extensive coverage of symptoms and were accompanied by recommendations for training procedures.

The present state examination

The present state examination (**PSE**)[54] is a semi-structured procedure, based upon an interview schedule containing items that are rated as the interview proceeds. The content of the PSE has always been more or less comprehensive and it contains a number of symptoms, such as worry, muscular tension, restlessness, etc. that are not associated with particular diagnoses. These symptoms are included because they are often clinically obvious and also important to the patient (see comments below on 'bottom-up' and 'top-down' organization of interview schedules).

The ratings made by the interviewer do not depend entirely upon the immediate reply of the subject, but represent the interviewer's clinical judgement as to whether or not the subject has the symptoms as described in the glossary of definitions learned during the interview training. Questions are provided for all the symptoms and items and are used whenever possible in the order provided, but the order may be varied if the interviewer thinks fit. The interviewer is also encouraged to ask any other questions that seem relevant to determine the timing, frequency, and severity of the symptoms, as in an ordinary clinical interview. In other words, the interviewer aims to conduct a clinical interview that has been structured as much as possible so as to allow symptoms to be rated with high inter-rater reliability, but without seeming to be unpleasantly rigid to either the subject or the interviewer. Much practice and training are required before these aims can be achieved, but there is no doubt that it is possible.

The PSE was not developed with any particular diagnostic classification in mind. It was intended from the start simply to be a means of arriving at a comprehensive and defined set of symptoms described in a reliable manner, with the user being left to decide whether and how to condense the symptoms into groups and what to do with the results. This is sometimes referred to as a 'bottom-up' style of instrument organization. Versions 7 and 8 of the PSE were first used on a large scale in two studies that involved international collaboration and comparisons, namely the United States–United Kingdom Diagnostic Project between London and New York,[55,56] and the International Pilot Study of Schizophrenia coordinated by the WHO, Geneva.[57] Since then its content has been revised and extended as versions 9 and 10, but the techniques of interviewing and rating remain the same. PSE-10 is one of the main components of Schedules for Clinical Assessment in Neuropsychiatry.

Schedule for affective disorders and schizophrenia, and the structured clinical interview for DSM

The series of instruments developed by Spitzer and his colleagues at Biometrics of the New York State Psychiatric Institute have been of several different kinds and, in the early years at least, had a much more rigid structure than the PSE. Users of the Mental Status Schedule and the longer Psychiatric Status Schedule were instructed to follow the order of the questions as printed in the schedule, the only deviation from this being a repetition of the same questions if thought necessary by the interviewer. However, later instruments such as the Schedule for Affective Disorders and Schizophrenia[48] and, more recently, the Structured Clinical Interview for DSM-III and DSM-IV[53] allow more flexibility for the interviewer in both interview style and the choice of a little or a lot of training (despite its length, the Structured Clinical Interview for DSM is recommended for clinical use as well as for research). There has also been an increasing tendency for instruments from the New York group to be dedicated to a particular purpose. For instance, the content of the Schedule for Affective Disorders and Schizophrenia is keyed towards the study of relationships between schizophrenia and affective disorders, and the Structured Clinical Interview for DSM contains only those items that are necessary for identifying disorders present in the corresponding DSM. Like the Diagnostic Interview Schedule mentioned in the next section, the Schedule for Affective Disorders and Schizophrenia and the Structured Clinical Interview for DSM have a 'top-down' structure, meaning that their content is determined from the start by an already existing set of criteria or symptoms.

The instruments produced by these two centres in the 1960s and 1970s have been used widely in many countries, and their success led to the production of many similar instruments by other researchers. The adoption of the PSE for use by the WHO in a number of international collaborative studies also led to its being translated into more than 25 languages, with varying but never extensive degrees of adaptation to fit the different cultures and social settings involved.

The diagnostic interview schedule

The third major research group is based at Washington University, St Louis, Missouri, and is well known as the originator of the first widely used sets of *Diagnostic Criteria for Research*.[58] Following the publication of DSM-III in 1980, there was considerable interest

in discovering how the disorders it contained were distributed in the American population. Supported by the National Institute of Mental Health, Lee Robins and her colleagues designed the Diagnostic Interview Schedule (**DIS**)[59] for this purpose. This is composed of questions covering the symptoms required to identify what were considered to be the 15 most important disorders in DSM-III. The Epidemiological Catchment Area study of the National Institute of Mental Health, the very large study in which the DIS was first used, included a population sample of more than 18 000 subjects in five largely urban areas.[60]

So as to avoid the costs and other problems involved in employing trained psychiatrists or psychologists as interviewers, the DIS was designed as a highly structured questionnaire administered as an interview by lay interviewers. The interviewers, usually already experienced in interviewing for market research, had to undergo a week-long intensive training course on the DIS. The DIS questions must be given in the order printed in the schedule. Possible symptoms are not rated as present if in the opinion of the subject they may be due to physical disorders, but there is no free questioning about timing, severity, and other details of the symptoms. Questions may be repeated, but only questions provided in the schedule may be asked of the subject. This is a very different concept from that of the PSE technique, and it is based upon the assumption made by the designers of the DIS that by controlling the interviewer in this way, the DIS would 'enable the interviewer to obtain psychiatric diagnoses comparable to those a psychiatrist would obtain'.[60] Put in another way, this is an assumption that expressed complaints can be used as near equivalents of inferred symptoms for the purposes of identifying psychiatric disorders.

Schedules of clinical assessment for neuropsychiatry and the composite international diagnostic interview

Although originating from different groups with different traditions and purposes, the PSE and the DIS have now given rise to direct descendants, namely the Schedules of Clinical Assessment for Neuropsychiatry (**SCAN**) and the Composite International Diagnostic Interview (**CIDI**), that are closely connected. During the early 1980s, a collaborative programme of work between WHO and the National Institute of Mental Health of the United States (known as the Joint Project) resulted in the transformation of DIS into CIDI[61] by increasing its contents by adding large parts of first DSM-IIIR and then of the drafts of ICD-10 and DSM-IV. This was matched by the evolution of PSE-9 into PSE-10, the centrepiece of SCAN,[50] whose content similarly covers almost all of both ICD-10 and DSM-IV. The only sections of ICD-10 and DSM-IV not now covered by SCAN and CIDI are those dealing with disorders of adult personality, disorders of childhood and adolescence, and mental retardation.

The coordination by WHO of the development of the final stages of SCAN and CIDI has been aimed at the production of two instruments with different but complementary uses in epidemiological studies. CIDI can be administered to comparatively large numbers of subjects in the community since the use of lay interviewers keeps costs to a minimum. SCAN is more suitable for the professional (and therefore more expensive) assessment of subjects with obvious or severe disorders, whether these have been selected from a larger population by means of CIDI or other screening instruments, or whether they are being studied clinically for other reasons. The latest development in this long-term programme has been the establishment of WHO-sponsored training centres in a number of countries. Psychiatrists and other mental health professionals can now obtain the necessary training for both SCAN and CIDI in English, French, German, Spanish, Chinese, Japanese, and Arabic.[61]

These and other instruments will no doubt be developed further, but every new instrument and every change to an existing one carries with it problems of data interpretation. Even though the content of changed or new instruments may seem to be the same as their predecessors, quite small changes in the method or the sequence of questions may have important effects, particularly for highly structured instruments in which the ratings are not filtered through the clinical judgement of a trained mental health professional. For instance, a recent report from the United States[62] discusses the possibility that the differences in prevalence rates for some disorders found between the Epidemiological Catchment Area study[60] and the more recent Co-morbidity Study[63] are due at least in part to changes in the 'stem questions' that introduce other specific questions rather than being due to real differences in the community subjects.

There are also unsolved problems in the study of individuals in the community, who have not sought professional help, by means of instruments originally designed for the study of psychiatric patients already in contact with services. To fulfil the criteria for a psychiatric disorder does not necessarily indicate a need for treatment, since the assessment of 'caseness' requires more than a simple count of symptoms. The debate about this problem has now stretched over 20 years, but needs to continue,[64,65] together with further examination of the closely related topic of clinical validity.[66]

Other selected issues

The importance of **negative symptoms** in the assessment of individuals with schizophrenic syndromes has led to the development of instruments devoted to these symptoms; the Scale for the Assessment of Negative Symptoms[47] is one of the most widely used, particularly in the United States. The Psychological Impairments Rating Scale (WHO/PIRS)[67] has been found to be acceptable in a variety of cultural settings, and has been used in several large international collaborative studies coordinated by the WHO. Both these instruments and a variety of others are useful as checklists for ordinary clinical purposes.[68] However, because of the nature of the symptoms being assessed, most of them are still beset with significant problems about inter-rater reliability and the exact meaning of their constituent items.

The **assessment of personality** poses special problems because to obtain a satisfactory account of a individual's personality, however the concept is defined, requires much more than the views of that individual; additional accounts of personal development and relationships from relatives or close friends are needed for comparison. Current concepts of personality disorders as listed in ICD-10 and DSM-IV also have serious limitations; the problems are well illustrated by a recent large international collaborative study coordinated by the WHO, using the International Personality Disorder Examination (IPDE).[61] Several hours of skilled interviewing are required, with at least two informants, to cover the content of the items that are needed to identify the disorders of adult personality contained in both the above classifications.

This study and others with similar aims have found that if an individual fulfils the criteria for one disorder of personality, they are quite likely to fulfil the criteria for at least one more. This implies that the present categories reflect only parts of the overall personality; this may be quite useful, but a fairly drastic overhaul of the currently used categories is clearly needed.

Clinically, it is useful to assess three aspects of personality, according to the salience of personal characteristics and problems arising from them. First, one or more of the personality disorders described in ICD-10 or DSM-IV should be used only if there are quite clear accounts of repeated problems and behaviours as specified, and they are not due to symptoms of any other disorders that may be present. Second, for less severe but repeated problems and behaviours, the concept of 'accentuated personality type' is often useful, described simply by a short list of ordinary adjectives. These indicate recurring behaviours and attitudes likely to cause a variety of mild interpersonal or social problems, again not attributable to symptoms of other disorders. Finally, even though neither of these first two types of personality disturbance is present, it is always worthwhile describing the usual characteristics of the patient by means of a few adjectives (such as 'a worrier', somewhat shy and socially inhibited, definitely gregarious, etc.). Vaguely optimistic terms commonly offered by friends and relatives, such as 'happy-go-lucky', should be avoided.

Multiaxial descriptive systems (often optimistically called classifications) have been available for many years,[69] and now apply to both ICD-10 and DSM-IV. Multiaxial systems describe several aspects of the person in addition to the disorder, and can be regarded as providing a systematized formulation that facilitates the coded recording of several aspects of the person concerned. Most of them have been designed more for research than for everyday clinical use, but they can all serve as very useful checklists when preparing for clinical reviews. DSM-IV, like DSM-III, is presented as a multiaxial scheme covering five aspects of the subject (Axis I, clinical disorders; Axis II, personality disorders and mental retardation; Axis III, General Medical Conditions; Axis IV, Psychosocial and Environmental Problems; Axis V, Global Assessment of Functioning). Similar instruments are now available for ICD-10,[70] covering general adult psychiatry and the psychiatry of childhood and adolescence (see Chapter 1.9 and 9.1.1).

Quality of life has come to the fore in recent years, but in the same way as for multiaxial assessment, to use this term does no more than make explicit something that has always been implicit in a good clinical assessment. Examination of the content of the many assessment instruments that are now available with this title shows that they contain various mixtures of almost every possible attribute of the person, the illness, and the environment. There is no point in using a new term when the information collected refers only to already familiar problems such as symptoms, disablements, how the patient's time is occupied, and contacts with medical services. There is even a considerable literature on the 'quality of life' of whole communities and countries, in which indices are calculated from national or regional statistics about, for instance, standards of housing, education, transport, and consumption of material resources. Such indices are of value to economists and demographers, but are far removed from clinical assessments. There is much to be said for using the term in clinical work only when it indicates 'higher-level' value-judgements and concepts such as personal satisfaction, self-fulfilment, and freedom from distress. Most of these are subjective and difficult to measure, but in many ways they reflect the ultimate aims of all medical care. An excellent recent review of this topic from the viewpoint of psychiatry is available,[71] which illustrates well the wide range of subjects now covered by the term.

Service research into the closely related **needs assessment** has resulted in the production of some detailed instruments that, again, are of interest to clinicians largely as potential checklists. A good example from the United Kingdom is the MRC Needs for Care Assessment.[72]

Administrative pressure to provide some sort of **quantification of clinical outcome** has resulted in several comparatively brief instruments designed for clinical use. Two widely used examples are the Global Assessment of Functioning (**GAF**) scale (Axis V of DSM-IV) and, in the United Kingdom, the recently developed Health of the Nation Outcome Scale (**HONOS**).[49]

In both of these, the assessor uses whatever information is available about the patient to make judgements about the presence and severity of symptoms and troublesome behaviour, and the extent to which these interfere with activities, relationships, and social performance. In the GAF scale this is expressed as a single overall score. In HONOS, 12 separate ratings are made which can be used independently, or added together to give an overall score if required. This type of instrument is likely to become increasingly important as the demand for 'evidence-based medicine' spreads, since they are designed for use by virtually any health professional in almost any setting. So long as precautions are always taken to ensure that the ratings made are as reliable and as valid as the setting permits, and likely sources of bias and error are kept in mind, their use can be a valuable aid to many forms of clinical assessment.

One further example of a **comprehensive assessment instrument** should be mentioned because it was designed for both research and clinical purposes, and it has been at times widely used in a number of European countries. The ADMP (an acronym in both German and English for the Association of Methodology and Documentation in Psychiatry) exists in English, German, French, Spanish, and Japanese versions, covering virtually all the information needed for a comprehensive assessment by means of lists of items to be coded as present or absent. It is up to the user to decide the meaning of each item, and how much to train with fellow raters (or not) so as to improve inter-rater reliability.[73]

The condensation and recording of information

Summary and formulation

The skills required to produce summaries and formulations should be acquired early on in professional training, since they are central to the process of getting the information about the patient into a form which facilitates the making of decisions and the allocation of priorities for actions. Useful preliminaries to the writing of both summary and formulation are the preparation of a problem list and a lifechart; how to prepare these should also be covered in the early stages of training. The summary for an individual patient should be more or less the same whoever prepares it, since it should be a simple record of what is known, arranged under conventional headings. A 'telegram' style of writing is acceptable for the sake of brevity. In contrast, a formulation should be written as a grammatically correct narrative, and there is no necessary expectation

that two different clinicians using the same summary about a patient would arrive at exactly the same conclusions in their formulations. This is because a formulation is an attempt by the writer to understand, and therefore to some extent to interpret, what has been influencing (and perhaps causing) the feelings and behaviour of the patient, and what relationships might exist between life events, illness, and contact with medical services.

Like the rest of the written medical records, the summary, formulation, and problem lists should be regarded as being as much for future readers as for the present carers. A clearly written summary and a well-argued formulation recorded in the case records will ensure that the reasons for treatments and decisions to do with the present illness are clear, and will be of great help to others if the patient has to be assessed in subsequent episodes of illness.

Summaries and formulations written by psychiatric members of a multi-disciplinary team should be freely available to all the team, so that they can be discussed before the meetings at which a diagnosis is agreed and care programmes are set-up. But it is not usually appropriate to send summaries and formulations made for hospital and team purposes to general practitioners or to consultants in other specialties. Specially written and shorter letters are best for this, taking into account the possibility that the patient or family may gain sight of, or even be shown, documents about them sent to other medical professionals.

Differential, main, subsidiary, and alternative diagnoses

A differential diagnosis should be placed in the case records in a prominent place, with a clear indication of who made it ('diagnosis' will be used in this section because of current conventions, but the difference between identifying a disorder and inferring an underlying diagnosis already noted must be kept in mind). When the patient suffers from more than one disorder it is usually possible to select one as the main diagnosis and specify the other(s) as additional or subsidiary diagnoses. The main diagnosis will usually be the one that is leading to immediate action, but the choice may depend upon the purposes for which the diagnoses are being recorded. Usually it reflects the reason for the current contact with services or admission but there are patients and occasions when, for instance, it makes more sense to record a lifetime diagnosis (such as schizophrenia or bipolar disorder) as the main diagnosis, even though something else such as anxiety or a phobic disorder is the reason for the current episode of care.

When one main diagnosis clearly applies yet does not account for some symptoms which, although a significant part of the clinical picture, still fall short of fulfilling the criteria for another disorder, it is useful to record these simply as 'additional symptoms' (for instance, depressive disorder with some obsessional symptoms; agoraphobia with some depressive symptoms, etc.). Neither ICD-10 nor DSM-IV mention this way of recording symptoms 'leftover' after the main disorder has been accounted for, even though it is a useful clinical custom familiar to many generations of clinicians in a variety of countries. However, omission from formal classifications should not be allowed to inhibit clinicians from following clinical habits they find useful.

When there is reasonable debate about what is the best diagnosis out of two or more possibilities, one must be chosen provisionally as the main diagnosis as a basis for action but the other should be recorded as an alternative diagnosis. It is also good practice in quite early stages of the assessment process to record provisional diagnoses, which can then be changed as more information becomes available. About a third of psychiatric patients fulfil the criteria for more than one disorder as defined in current classifications, but as already noted, this does not carry the same implications about underlying morbid physiological, psychological, or anatomical processes as would a statement about the presence of the same number of medical or surgical diagnoses.

Making a prognosis

The final statement in the formulation should be the prognosis. This attempt to predict what will happen to the patient in the future should be expressed as clear statements about likely outcomes, avoiding vague comments such as 'the prognosis is guarded' (found all too often in case records). The patient and family usually hope to be told about the prospects for recovery and the likelihood of relapse. Efforts should be made to do this, but with due care to emphasize that a prognosis is only an estimate that may be proved wrong by events. A prognostic statement should contain predictions of such things as:

1 immediate response to treatment, assuming compliance
2 duration of this episode of illness and/or stay in hospital
3 degree of recovery from this illness (i.e. partial or complete) in terms of both symptoms and return to previous activities
4 risk of recurrence, stated as the likely position at specific points in time, depending upon the circumstances of the case (6 months, 1 year, and 2 years from the present are often appropriate)

However difficult it may seem, attempts should be made to record a prognosis in these terms, and to sign it. To do this will fulfil the legitimate expectations of the patient and family, and the clinician will make possible a uniquely valuable learning experience when faced in the future with such statements about those patients seen in further episodes of care.

Reviews

The initial assessment should produce a list of agreed actions to be carried out in a stated order of priority by the various members of the team. Division into immediate and medium-term actions will help the whole process, and also indicate the timescale of reviews to assess progress. One of the main functions of the acknowledged team leader is to keep an eye on the progress of all the patients in the care of the team, discussing with each key worker both outside and within team meetings the best timing of the next review. Review meetings should be recorded as such in the case documents, with conclusions about progress made and any changes in plans or objectives. It is particularly important, again for future readers of the records, to write down clearly whether there was any response treatment (it is very frustrating to read in a case record of treatments given in the past, and then to find no indication of the result.)

Writing reports

Consultant psychiatrists are often asked by external agencies to provide written reports on patients for whom they have a current

clinical responsibility, and requests may also be received for a report on a patient they have not seen previously. The purpose of these external reports is usually different to that of the usual clinical communications undertaken in the ordinary clinical care of the patient, in that the request is usually for an opinion about one specific issue. These requests frequently involve an opinion on the risk posed to others by the patient's inability to perform certain skills, or by the positive adverse effect of the patient's problems on others. An opinion on the capacity of the patient to understand and competently agree about important issues is also frequently requested.

The purposes of reports

Reports requested by individuals or agencies (both judicial and non-judicial) will usually fall into one of four broad groups:

1 Protection of the public or an institution:
   (a) life assurance and mortgage companies who are interested in the risk of suicide, or loss of earnings due to future illnesses;
   (b) licensing authorities and transport companies who are concerned with fitness to drive or risks due to the public, due to impairment of skills and judgement consequent upon psychiatric illness;
   (c) employment and benefit agencies who are concerned with fitness to work;
   (d) employers or occupational health physicians who are concerned about the risk posed by the patient's psychiatric disorder to an institution's clients, or about the likelihood of periods of absence because of sick leave.
2 Protection of the patient:
   (a) solicitors or courts may require reports on the competence of individual patients to conduct their financial affairs, to protect themselves from exploitation, and to engage in civil contracts;
   (b) bodies concerned with the Mental Health Act (in the United Kingdom the Mental Health Act Commission and Mental Health Review Tribunals) may require reports on the competence of individuals to give informed consent to non-voluntary psychiatric treatment or inpatient care, and the risk to the safety and well-being of the individual patient posed by such treatment.
3 Child protection:
   (a) Social Services Child Care Departments and others involved in the welfare of children may request reports of the supervising psychiatrist for Child Protection Case Conferences on the contribution of a psychiatric disorder to the childcare problem, and on the likely impact of the psychiatric disorder on the future parenting of the patient;
   (b) solicitors acting for all parties in childcare proceedings (the child, the Social Services Department, and the patient) may ask for a psychiatric report on a mentally ill parent about the likely risk to the child of suffering significant physical, developmental, or emotional harm from the patient in question.
4 Medico-legal and compensation proceedings: lawyers acting for either the patient or an agency being sued may ask for a report on the impact of the event on the mental health of the patient, together with the nature of the psychiatric disorder and its prognosis. Such reports may be requested of the supervising consultant, or of another psychiatrist as an independent expert.

Reports for forensic or criminal proceedings are dealt with in Chapter 11.15.

Guiding principles

The general principles noted below apply to all reports, whether the psychiatrist knows the patient because of current or previous clinical responsibility, or whether the report is on a patient whom the psychiatrist has not seen before (this latter is known as providing an 'expert opinion').

(a) Confidentiality

In most situations written consent must be obtained from the patient before personal information can be given to an outside agency. In almost all situations involving the writing of a report on an individual to an outside agency, that individual will gain sight of the report or will be entitled to do so. It is therefore good clinical practice for the patient to be aware of the content of the report, and particularly of any opinions or recommendations it may contain. Nevertheless, there are certain situations where the duty of care to the public or to a child overrides the duty of confidentiality, and in such circumstances the psychiatrist may write the report even without the patient's consent.

(b) Partiality

The opinion of the psychiatrist is being sought as an expert professional. The report should not be biased in favour of one side or another and should not be influenced unduly by the interests of the commissioning agency or the psychiatrist's view of the best interests of their patient. This may cause difficulties if the patient is in the personal care of the psychiatrist because in most clinical situations psychiatrists try to be non-critical, non-judgemental, and supportive, tending to encourage rather than to prevent. But the best interests of the patient will not be served by being put in a situation where the likely outcomes are failure to do what is expected or to function at a suboptimal level.

Structure of reports

All reports should have three main sections. First, the report should begin with the patient's personal details and the reasons for which the report has been requested, together with the identity of the commissioning agency. It should also specify the relationship between the writer and the patient. If the patient is or was in the clinical care of the writer the duration of the care should be noted, and the date of the last occasion the patient was seen should be given. If a special interview had to be arranged with a patient not previously known, the duration and date of the interview should be stated. The sources of information other than the patient used to prepare the report should then be detailed, plus any other documents that have been read. Reports for civil, judicial, and child protection proceedings will also require a short paragraph on the current employment and status of the author of the report, and a note of any special experience of relevance.

The second section should describe in appropriate detail the patient's personal, social, medical, and psychiatric history, the mental

state and behaviour at examination, the diagnosis and differential diagnosis, and comments upon aetiology, management, and prognosis. In almost all reports, the prognosis is the primary concern, so this should be given special attention. It is important to remember that one of the most reliable predictors of the recurrence of behaviours or episodes of illness in the future is the frequency of their occurrence in the past. Similarly, the vulnerability of the patient in the past (that is, any enduring predisposing factors and patterns of past precipitants) will tend to predict future vulnerability and the likelihood of further episodes of illness. Some mention of the past will therefore always be necessary, but in many instances this can be brief and reduced to a commentary of a few lines. But in other situations, particularly those involving civil court actions or childcare proceedings, a more detailed account of the past will be necessary.

Certain aspects of the patient's past history and previous levels of functioning will need to be highlighted depending upon the purposes of the report and the nature of the questions asked of the psychiatrist. For example, if the report has been requested by an occupational physician about the fitness of a patient to return to work, then attention will need to be paid in the report to the duration of illnesses in the past and the amount of sick leave that has been taken. Detail will need to be given about the impact, if any, that the patient's ill health has had on the past to his or her capacity to work. If the report has been requested in relation to the safety of the patient to care for a child, then information will need to be given in the past history of the patient about the previous impact of the patient's illness on his or her capacity to care for children or any risks that the patient posed to a child in the past. Life assurance and mortgage companies are likely to be particularly interested in suicidal behaviour.

The last section should contain the opinion of the psychiatrist about the specific questions posed by the commissioning agent. These questions may be unrealistically simple or there may be requests for categorical assertions of outcome that are simply not possible. The writer must avoid falling into the trap of complying with unreasonable requests about certainties. One way of avoiding this is to give opinions about risks or outcomes by stating criteria that would indicate different outcomes with different likelihoods, expressed by words such as possible, probable, and definite.

In situations where one of the variables involved in the patient's prognosis is the response of helping agencies and the availability of resources, great care must be exercised on the part of the report writer to ensure that this contingency is made clear. If possible, suggestions should be made as to how the availability of the required resources can be assured. When considering the likely impact of a future breakdown in the mental health of a patient on some other person, such as a child, consideration should be given not only to the direct impact of the illness but also to the indirect consequences and the presence or absence of other protective factors. For example, if a woman with schizophrenia lives with her parents who can safely take over the care of her child, then the impact of a further episode on that child may be much less than if she is living alone and the child needs to be removed into the care of the local authority.

An opinion is often requested on whether an accident or an act of omission such as medical negligence caused the current psychiatric disorder or disabilities of the patient. If the psychiatrist concludes that the accident or omission was definitely a contributing cause but not in itself sufficient to cause all aspects of the existing disorder and disability, then further comments will be expected on other possible contributing influences, such as predisposing personal traits, or special vulnerability to current adversities. In such circumstances, there should be an attempt to weigh the contributing factors in order of their aetiological importance.

The last section of the report is usually the best place to list the sources of information used for the report, making clear distinctions between personal observations and information obtained by the writer, opinions and observations made by other team members, and written information obtained from other documents. There should always be a clear distinction between opinions based upon objective information and direct examination, and suppositions based upon interpretations, speculation, and past clinical experience. If opinions based upon research conducted by others are given, then the sources of this information should be acknowledged and referenced in the usual manner.

The language of the report should be appropriate to the commissioning agency. If the report has been requested by an occupational physician or medical officer working for a company, then it is appropriate to use accepted medical and psychiatric terminology. If the report has been requested by a civil or judicial authority, non-technical language should be used wherever possible and any medical or psychiatric terms used should be defined. At all times when writing psychiatric reports it is important to use psychiatric terms in an appropriate fashion according to a stated international classification, and to avoid idiosyncrasies.

Further information

Moller, H.J., Engel, R.R., and Helmsley, D.R. (2001). Standardised measurement instruments in psychiatry. In *Contemporary psychiatry. Foundations of psychiatry*, Vol. 1 (eds. H. Henn, N. Sartorius, H. Helmchen, and H. Lauter), pp. 113–35. Springer-Verlag, Berlin.

Grounds, A. (1993). Psychiatric reports for legal purposes in forensic psychiatry. In *Clinical, legal and ethical issues* (eds. J. Gunn and P.J. Taylor), Chap. 22, p. 826–55. Butterworth—Heineman Ltd, Oxford.

References

1. Gelder, M., Gath, D., Mayou, R., *et al.* (1996). *Oxford textbook of psychiatry* (3rd edn). Oxford University Press, Oxford.
2. Post, F. (1983). The clinical assessment of mental disorders. In *Handbook of psychiatry*, Vol. 1 (eds. M. Shepherd and O.L. Zangwill), pp. 210–20. Cambridge University Press, Cambridge.
3. McHugh, P.R. and Slavney, P.R. (1986). *The perspectives of psychiatry*. Johns Hopkins University Press, Baltimore, MD.
4. Taylor, F.K. (1971). A logical analysis of the medico-psychological concept of disease: Part 1. *Psychological Medicine*, **1**, 356–64.
5. Susser, M.W. and Watson, W.B. (1971). *Sociology in medicine* (2nd edn), pp. 16–17, 216–18, 295ff. Oxford University Press Oxford.
6. Lewis, A.L. (1953). Health as a social concept. *The British Journal of Sociology*, **4**, 109–204.
7. Wootton, B. (1959). Social pathology and the concepts of mental health and mental illness. *Social science and social pathology*. Allen and Unwin, London.
8. Scadding, J.G. (1967). Diagnosis, the clinician and the computer. *Lancet*, **2**, 877–82.
9. Scadding, J.G. (1996). Essentialism and nominalism in medicine: the logic of diagnosis in disease terminology. *Lancet*, **348**, 594–6.
10. Kerr, A. and McClelland, H. (1991). *Concepts of mental disorder: a continuing debate*. Gaskell Press, London.

11. Tyrer, P. and Steinberg, D. (1993). *Models for mental disorders: conceptual models in psychiatry*. Wiley, Chichester.
12. Eisenberg, L. (1998). The social construction of mental illness (editorial). *Psychological Medicine*, **18**, 1–9.
13. Eisenberg, L. (1977). Disease and illness. Distinctions between professional and popular ideas of sickness. *Culture, Medicine and Psychiatry*, **1**, 9–24.
14. Engel, G.L. (1978). The biopsychosocial model and the education of health professionals. *Annals of the New York Academy of Sciences*, **310**, 169–81.
15. Susser, M. (1990). Disease, illness, sickness; impairment, disability and handicap (editorial). *Psychological Medicine*, **20**, 471–4.
16. Taylor, F.K. (1979). *Psychopathology: its causes and symptoms*. Quartermaine House, Sunbury on Thames.
17. Fulford, K.W.M. (1989). *Moral theory and medical practice*. Cambridge University Press, Cambridge.
18. Mechanic, D. (1986). The concept of illness behaviour; culture, situation and personal predisposition. *Psychological Medicine*, **16**, 1–8.
19. Menninger, K. (1948). Changing concepts of disease. *Annals of Internal Medicine*, **29**, 318–25.
20. Kendell, R.E. (1975). *The role of diagnosis in psychiatry*. Blackwell Science, Oxford.
21. Cooper, J.E. (1983). Diagnosis and the diagnostic process. In *Handbook of psychiatry. General psychopathology*, Vol. 1, Part III (eds. M. Shepherd and O.L. Zangwill), pp. 199–209. Cambridge University Press, Cambridge.
22. World Health Organization. (2001). *International Classification of Functioning, disability and health (ICF) short version*. WHO, Geneva.
23. Nagi, S.Z. (1965). Some conceptual issues in disability and rehabilitation. In *Sociology and rehabilitation* (ed. M.B. Sussman), pp. 110–13. American Sociological Association, Washington, DC.
24. Weissman, M.M. (1975). The assessment of social adjustment: a review of techniques. *Archives of General Psychiatry*, **32**, 357–65.
25. Gurland, B.J., Yorkston, N.J., Stone, A.R., *et al.* (1972). The Structured and Scaled Interview to Assess Maladjustment (SSIAM). I. Description, rationale and development. *Archives of General Psychiatry*, **27**, 259–64.
26. Claire, A.W. and Cairns, V.E. (1978). Design, development and use of a standardised interview to assess social maladjustment and dysfunction in community studies. *Psychological Medicine*, **8**, 589–604.
27. Brown, G.W. and Harris, T.O. (1986). *Social origins of depression; a study of psychiatric disorder in women*. Tavistock Publications, London.
28. Brown, G.W. and Harris, T.O. (1986). Stressor, vulnerability and depression; a question of replication (editorial). *Psychological Medicine*, **16**, 739–44.
29. Andrews, G. and Tennant, C. (1978). Life events and psychiatric illness (editorial). *Psychological Medicine*, **8**, 545–9.
30. Jones, S.J., Turner, R.J., and Grant, J. (1987). Assessing patients in their homes. *Bulletin of the Royal College of Psychiatrists*, **11**, 117–19.
31. Oates, M.R. (1988). The development of an integrated community-oriented service for severe post-natal mental illness. In *Motherhood and mental illness*, Vol. 3 (eds. R. Kumar and I. Brockington), pp. 133–58. John Wright, London.
32. Tyrer, P. (1984). Psychiatric clinics in general practice: an extension of community care. *The British Journal of Psychiatry*, **145**, 9–14.
33. Burns, T. (1997). Case management, care management and care programming (editorial). *The British Journal of Psychiatry*, **170**, 393–5.
34. Holloway, F., Oliver, N., Collins, E., *et al.* (1995). Case management: a critical review of the literature. *European Psychiatry*, **10**, 113–28.
35. Katschnig, H. and Cooper, J.E. (1991). Psychiatric emergency and crisis intervention services. In *Community psychiatry: the principle* (eds. D.H. Bennet and H.L. Freeman), pp. 517–42. Churchill Livingstone, Edinburgh.
36. Cooper, J.E. (1990). Professional obstacles to implementation and diffusion of innovative approaches to mental health care. In *Mental health care delivery: innovations, impediments and implementation* (eds. I. Marks and R. Scott), pp. 233–53. Cambridge University Press, Cambridge.
37. Henderson, A.S. (1988). The tools of social psychiatry. In *An introduction to social psychiatry*, pp. 6–29. Oxford University Press, Oxford.
38. Thompson, C. (ed.) (1989). *The instruments of psychiatric research*. Wiley, Chichester.
39. World Health Organization. (1993). *Catalogue of assessment instruments used in the studies coordinated by the WHO mental health programme*. WHO/MNH/92.5. Division of Mental Health, World Health Organization, Geneva.
40. Sartorius, N. and Janca, A. (1996). Psychiatric assessment instruments developed by the World Health Organization. *Social Psychiatry and Epidemiology*, **31**, 55–69.
41. Janca, A. (1997). Current trends in the development of psychiatric assessment instruments. *Current Opinion in Psychiatry*, **10**, 457–61.
42. Thornicroft, G. and Tansella, M. (eds.) (1997). *Mental health outcome measures*. Springer-Verlag, Berlin.
43. Goldberg, D.P. and Hillier, V.F. (1979). A scaled version of the general health questionnaire. *Psychological Medicine*, **9**, 139–45.
44. Wittchen, H.-U., Robins, L.N., Cottler, L.B., *et al.* (1991). Cross-cultural feasibility, reliability and sources of variance of the Composite International Diagnostic Interview (CIDI)—results of the multi-centre WHO/ADAMHA field trials (wave 1). *The British Journal of Psychiatry*, **159**, 645–53.
45. World Health Organization. (1993). *The Composite Diagnostic Interview, core version 1.1*. American Psychiatric Press, Washington, DC.
46. Hamilton, M. (1960). Rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, **23**, 56–62.
47. Andreason, N.C. (1989). The Scale for Assessment of Negative Symptoms (SANS): conceptual and theoretical foundations. *The British Journal of Psychiatry*, **155**(Suppl. 7), 49–52.
48. Endicott, J. and Spitzer, R.L. (1978). A diagnostic interview: the schedule for affective disorders and schizophrenia. *Archives of General Psychiatry*, **35**, 837–44.
49. Wing, J.K., Curtis, R., and Beevor, A. (1994). 'Health of the Nation': measuring mental health outcomes. *Psychiatric Bulletin: Journal of Trends in Psychiatric Practice*, **18**, 690–1.
50. Wing, J.K., Babor, T., Brugha, T., *et al.* (1990). SCAN: Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry*, **47**, 589–93.
51. Overall, J.E. and Gorham, D.R. (1962). The Brief Psychiatric Rating Scale (BPRS). *Psychological Reports*, **10**, 799–812.
52. Asberg, M., Montgomery, S.A., Perris, C., *et al.* (1978). A Comprehensive Psychopathological Rating Scale (CPRS). *Acta Psychiatrica Scandinavica. Supplementum*, **271**, 5–28.
53. Spitzer, R.L., Williams, J.B.W., and Gibbon, M. (1987). *Structured clinical interview for DSM-IV* (*SCID*). Biometrics Research, New York State Psychiatric Institute, New York.
54. Wing, J.K., Cooper, J.E., and Sartorius, N. (1974). Measurement and classification of psychiatric symptoms: an instruction manual for the PSE and the CATEGO program. Cambridge University Press, London.
55. Cooper, J.E., Kendell, R.E., Gurland, B.J., *et al.* (1969). Cross-national study of diagnosis of mental disorders; some results from the first comparative investigation. *The American Journal of Psychiatry*, **125**(Suppl.), 21–9.
56. Cooper, J.E., Kendell, R.E., Gurland, B.J., *et al.*(1972). *Psychiatric diagnosis in New York and London: a comparative study of mental hospital admissions*. Maudsley Monograph No. 20. Oxford University Press, Oxford.
57. Sartorius, N., Jablensky, A., and Shapiro, R. (1977). Two year follow–up of the patients included in the WHO international pilot study of schizophrenia: preliminary communication. *Psychological Medicine*, **7**, 529–41.

58. Feighner, J.P., Robins, E., Guze, S.B., *et al.* (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, **26**, 57–63.
59. Robins, L.N., Helzer, J.E., Croughan, J., *et al.* (1981). National Institute of Mental Health Diagnostic Interview Schedule. Its history, characteristics and validity. *Archives of General Psychiatry*, **38**, 381–8.
60. Regier, D.A., Boyd, J.H., Burke, J.D., *et al.* (1988). One month prevalence of mental disorders in the United States: based on the five epidemiological catchment area sites. *Archives of General Psychiatry*, **45**, 977–86.
61. Pull, C. and Wittchen, H.-U. (1991). CIDI, SCAN and IPDE: structured diagnostic interviews for ICD–10 and DSMIII–R. *European Journal of Psychiatry*, **6**, 277–85.
62. Regier, D.A., Kaelber, C.T., Rae, D.S., *et al.* (1998). Limitations of diagnostic criteria and assessment instruments for mental disorders. *Archives of General Psychiatry*, **55**, 9–115.
63. Kessler, R.C., McGonagle, K.A., Zhao, S., *et al.* (1994). Lifetime and 12 month prevalence of DSM–III–R psychiatric disorders in the United States; results from the National Co-morbidity Study. *Archives of General Psychiatry*, **51**, 8–19.
64. Wing, J.K., Bebbington, P., and Robins, L.N. (1981). *What is a case? The problem of definition in psychiatric community surveys*. Grant McIntyre, London.
65. Spitzer, R.L. (1998). Diagnosis and need for treatment are not the same (commentary). *Archives of General Psychiatry*, **55**, 120.
66. Kendell, R.E. (1989). Clinical validity. *Psychological Medicine*, **19**, 45–56.
67. Biehl, H., Maurer, K., Jablensky, A., *et al.* (1989). The WHO/PIRS: introducing a new instrument for rating observed behaviour, and the rationale of the psychological impairment concept. *The British Journal of Psychiatry*, **155**(Suppl. 7), 68–81.
68. Barnes, T.R.E. (ed.) (1989). Negative symptoms in schizophrenia. *The British Journal of Psychiatry*, **155**(Suppl. 7).
69. Mezzich, J.E. (1979). Patterns and issues in multiaxial psychiatric diagnosis. *Psychological Medicine*, **9**, 125–37.
70. World Health Organization. (1997). *Multiaxial presentation of ICD–10 for use in adult psychiatry*. Cambridge University Press, Cambridge.
71. Katschnig, H., Freeman, H., and Sartorius, N. (1997). *Quality of life in mental disorders*. Wiley, Chichester.
72. Bebbington, P., Brewin, C.R., Marsden, L., *et al.* (1996). Measuring the need for psychiatric treatment in the general population: the community version of the MRC Needs for Care Assessment. *Psychological Medicine*, **26**, 229–36.
73. Woggon, B. (1986). The AMDP-III: a comprehensive instrument for recording psychiatric information. In *Assessment in depression* (eds. N. Sartorius and T. Baan), pp. 112–20. Springer–Verlag, Berlin.

1.8.2 Assessment of personality

C. Robert Cloninger

Introduction

The assessment of personality provides the context needed to understand someone as a whole person with particular goals and values that they pursue with a unique emotional style. A person's way of adapting to life experience can tell an experienced clinician much about his level of well-being and his vulnerability to various forms of psychopathology. Knowing a person's personality well can allow a psychiatrist to predict what other mental and physical disorders are likely to occur in the same person or in the same family. For example, individuals with antisocial personality are more likely to have substance abuse and less likely to have Parkinson's disease than others.[1] On the other hand, if you learn someone has substance abuse, then you can reasonably suspect that they may be impulsive or novelty seeking. Recognition of the many associations between personality and psychopathology can greatly enhance clinical assessment and differential diagnosis in general.

Assessment of personality also helps to establish a therapeutic alliance and mutual respect, because it involves the sharing of unique personal and social information that distinguishes one person's style of life from others. Patients feel understood and appreciated when their psychiatrist understands their motivation and can predict their reactions to different situations and people. On the other hand, no one likes to be reduced to a 'case' or a 'label'. Everyone is unique, and yet it is possible to explore the mystery of each person's uniqueness in a systematic way. Consequently, effective clinical assessment of personality is designed to understand a person's emotions, goals and values, strengths and weaknesses in the context of the narrative of his life.

Understanding personality also helps in treatment planning because people differ markedly in the types of treatments to which they respond and with which they will comply. For example, personality traits predict much of the variability in response to antidepressants, whereas the symptoms of depression or other psychopathology do not.[2,3]

Fortunately, personality can be well assessed clinically without psychometric testing in ways that are simple and brief as part of routine history taking and mental status examination. The clinical assessment of personality requires little extra time if the clinician is alert to non-verbal cues in a person's general appearance, expressions, and behavior, as well as to the significance of what is said and how it is said. Only brief questions to clarify complaints and their context may be needed.

Personality develops over time in response to a changing internal and external environment. As a result, the longitudinal course of a person's development of personality and psychopathology is a key element in the clinical assessment of personality. Specifically, it is highly informative to know what a patient's personality was like as a child when assessing him in the presence of additional psychopathology, like a depression or anxiety state that modifies his emotions, thoughts, and behaviour. However, personality traits are not fixed and completely stable. Rather, each of us has a range of thoughts, feelings, and behaviours at any given point in time. As a result, our personality traits frequently vary within that range and occasionally change by moving beyond the previous range in response to particular internal and external events. Understanding the course of a person's development during his life is what allows the psychiatrist to understand him as a unique person.

In this chapter, I will try to explain the basic constructs and methods of personality assessment, so that a clinician can apply this knowledge in a flexible and practical manner. If you have to ask standardized questions that are not tailored to opportunities that arise in the course of an interview, then you don't understand the basic constructs adequately. On the other hand, some clinical features about personality traits are sufficiently high in yield and diagnostic value, that they should be assessed in a final review if they haven't been come up more spontaneously during the interview.

What is personality?

In order to assess something it is crucial to have a good understanding of what it is and what it is not. People differ markedly from one another in their outlook on life, in the way they interpret their experiences, and in their emotional and behavioral responses to those experiences. These differences in outlook, thoughts, emotions, and are what actions characterize an individual's personality. More generally, personality can be defined as the dynamic organization within the individual of the psychobiological systems that modulate his or her unique adaptations to a changing internal and external environment.[4] Each part of this definition is important for a clinician to appreciate. Personality is 'dynamic', meaning, it is constantly changing and adapting in response to experience, rather than being a set of fixed traits. Inflexibility of personality is actually an indicator of personality disorder. Personality is regulated by 'psychobiological' systems, meaning, personality is influenced by both biological and psychological variables. Consequently, treatment of personality disorders requires growth in psychological self-understanding and not just treatment with medications, although these can be helpful adjuncts to therapy.[5,6] These systems involve interactions among many internal processes, so that each person's pattern of adjustment is 'unique' to them, even though they follow general rules and principles of development as complex adaptive systems.[7] Finally, to understand personality and its development, we must pay attention to both the 'internal' and 'external' processes by which an individual interacts with and adapts to his own internal milieu and external situation. For example, when a person is under stress, he is likely to think and feel differently about himself and other people. On the other hand, when he is calm and encouraged, he may act more maturely and happily. Everyone has personal sensitivities or 'rough spots' that surface when they are under stress. Everyone has 'good days' and 'bad days', and this pattern of variability over time is what characterizes a person's unique personality configuration.

An individual's personality can only be adequately characterized in terms of interactions among different internal and external forces that influence a person's emotions, thoughts, and behaviour. A person may feel and act differently on a date, at work, with trusted friends, at school, or in church. His personality doesn't change, rather his personality can only be adequately assessed when the psychosocial context is specified. Some traits are strong and pervasive regardless of the situation, but other aspects of personality may be markedly affected by the situation. Furthermore, the internal processes may modify a person's outlook, as when his outlook is influenced by prior or anticipated events, or when his goals and values allow him to change his outlook in ways that are not predictable by what he has previously done. Human beings have an amazing ability to change their outlook for the better or the worse in ways that are unpredicted by their past or present circumstances. Personality traits can be described in ways that are moderately stable over time and situations, but a prudent clinician must never mistake average probabilities with predictive certainty.

Five major types of situations are useful to distinguish for human beings: Sexual situations involving reproduction and sexuality; Material situations involving the quest for material possessions and power; Emotional situations involving emotions and social attachments; Intellectual situations involving communication and culture; Spiritual situations involving the quest for what is beyond individual human existence. The average person is concerned with material situations most of the time—obtaining food, clothing, shelter, transportation, and striving for power and wealth. However, to understand a person fully it is essential to recognize his feelings, thoughts, and intuitions in other types of situations ranging from the sexual to the spiritual. The way a person adapts to these five different types of situations correspond to layers of an individual's personality. The treatment of psychopathology can be viewed as a working-through of problems and blind spots in these five layers of everyone's personality, enabling the development of self-awareness in the full range of life situations.[8]

Personality involves much more than the description of a fixed set of traits that allow the prediction of a person's behavior. Personality involves the interaction of internal and external forces that influence the development of a person's behavior, but nevertheless allow for the potential of a person to grow in self-awareness and thereby change in ways that cannot be predicted from his past behavior.[9]

How can personality be described quantitatively?

Personality refers to the motivational systems *within* a person, not between individuals. In other words, to understand what motivates a person we need to recognize empathically what he is thinking and feeling within his own being. We need a model of the dynamic psychobiological processes within a human being. Unfortunately, the people who have developed most personality tests often treat each person as a black box that emits self-reports. As a result, most personality psychologists have failed to understand the internal dynamics underlying the thoughts and feelings of the people they assess. However, it is possible to describe a person's internal processes, which interact with his or her external situations. In order to account for both the internal and the external influences on personality, it is essential to distinguish the dimensions of a person's temperament and those of his character.[4]

The temperament traits are biases in emotional responses that are fully developed early in life and relatively stable thereafter. On the other hand, character involves higher cognitive processes that develop in a stepwise manner over the life course to enable a person to regulate his emotions, achieve certain goals, and maintain particular values and virtues. Initially, it was thought that character was less heritable than temperament, but empirical studies have shown that both are moderately heritable. The key difference is the difference in the pattern of learning and memory: the procedural learning of habits and skills influences the conditioning of temperament, whereas propositional learning of goals and values influences the development of character. Both procedural and propositional learning interact with one another in self-aware consciousness so that a person can maintain a personal sense of continuity throughout many episodes of experience as the story of his life unfolds.

Temperament can be assessed in terms of four quantifiable dimensions, as measured by the Temperament and Character Inventory.[4] These are described in Table 1.8.2.1, which shows that each trait is manifested in slightly different ways depending on the situation. A situation necessarily depends on both the person's outlook and the external circumstances themselves. For example, a person is described as high in Harm Avoidance if he is easily fatigued, fearful, shy, pessimistic, and inhibited. On the other hand, a person is described as low in Harm Avoidance if he is vigorous, risk-taking, beguiling, optimistic, and uninhibited.

Table 1.8.2.1 Descriptions of temperaments according to emotional responses elicited by particular external situations and internal outlooks

Temperament	Sexual situations	Material situations	Emotional situations	Intellectual situations	Spiritual situations
Harm Avoidance	Fatigable vs Vigorous	Fearful vs Risk-taking	Shy vs Beguiling	Pessimistic vs Optimistic	Inhibited vs Uninhibited
Novelty Seeking	Craving vs Reserved	Extravagant vs Frugal	Irritable vs Stoical	Impulsive vs Rigid	Exploratory vs Immobile
Reward Dependence	Insecure vs Independent	Sympathetic vs Aloof	Sociable vs Distant	Sentimental vs Indifferent	Attached vs Detached
Persistence	Ambitious vs Apathetic	Overachieving vs Underachieving	Loyal vs Fickle	Determined vs Ambivalent	Perfectionistic vs Pragmatic

However, the level of Harm Avoidance varies moderately between situations. For example, some people who are shy are not easily fatigued, and some people who are shy meeting strangers are risk-takers when driving an automobile. The components of Harm Avoidance that are manifested in different situations are moderately correlated, and so it is useful to consider all these as part of a higher order trait that is moderately heritable and moderately stable across time and situations. Likewise, Novelty Seeking, Reward Dependence, Persistence are also moderately heritable and stable dimensions of temperament.

Likewise, there are three dimensions of character, which quantify the nature of a person's goals and values (Table 1.8.2.2). Each of these character traits is comprised of components that are expressed in different situations. The character dimensions also correspond to key functions of a person's mental self-government. As a result, character traits provide a rich description of key features of the mental status examination, including insight and judgment.

Insight refers to the depth of a person's ability to recognize and understand the inner nature of things, rather than basing opinions on superficial appearances. Insight is quantifiable as the character trait of Self-transcendence. A person with deep insight is respectful, mindful, and holistic in perspective, whereas one with little insight is unrealistic, shallow, and fragmented in perspective.

Judgment refers to a person's legislative ability to cooperate and get along with others in ways that are appropriate and flexible, and can be quantified as the character trait of Cooperativeness. A person with good judgment is cooperative and principled, whereas a person who has poor judgment is uncooperative and opportunistic.

Foresight refers to a person's executive ability to anticipate what will be satisfying in the long-term or in the future. This executive function allows a person to follow a life path that maintains well-being. A person who is far-sighted is responsible, purposeful, resilient, and resourceful, whereas one who learns from hindsight only is irresponsible, aimless, fragile, and inadequate. In addition, foresight leads to cheerfulness and spontaneity, whereas reliance on hindsight is associated with moodiness and conventionality. Accordingly, the degree of a person's foresight provides important clinical information about a person's ability to appreciate what is real, meaningful, and satisfying. As a result, a person's self-directedness is an important indicator of reality testing, maturity, and vulnerability to mood disturbance. Self-directedness is high

Table 1.8.2.2 Descriptions of the three dimensions of character according to the five layers of everyone's personality, which are defined by the predominant focus of the person's internal perspective on the external situation. Within each layer of personality, maturation and integration involves increasing each of the three character dimensions, which describe the functions of insight, judgment, and foresight. Integration of the whole person requires working through these functions in each of the layers of personality

Cognitive function (Character dimension)	Characteristics of the sexual layer	Characteristics of the material layer	Characteristics of the emotional layer	Characteristics of the intellectual layer	Characteristics of the spiritual layer
Insight (Self-Transcendence)	Trustful vs Alienated (prelogical categorizing)	Free-Flowing vs Compulsive (concrete-vivid logic)	Identifying vs Avoiding (emotive imagery)	Creative vs Imitative (abstract symbols)	Intuitive vs Conventional (preverbal schemas)
Judgment (Cooperativeness)	Tolerant vs Prejudiced	Forgiving vs Revengeful	Empathic vs Inconsiderate	Helpful vs Unhelpful	Principled vs Opportunistic
Foresight (Self-Directedness)	Responsible vs Irresponsible	Purposeful vs Aimless	Resilient vs Moody	Resourceful vs Inadequate	Spontaneous vs Predetermined

in people who are mature and happy, whereas it is low in people with personality disorders and in those vulnerable to psychoses and mood disorders.

Psychometric testing of personality traits

A wide variety of psychometric tests can be used to describe personality traits, so it is useful for a clinician to understand the relationships among alternative measures. The number and content of traits describing personality vary but there is actually extensive overlap among the traits measured. Hans Eysenck popularized tests that measured three factors called Neuroticism, Extraversion, and Psychoticism.[10] The Eysenck Personality Questionnaire also includes validity measure called 'Lie'. Nearly all tests subsequently developed include factors corresponding closely to Neuroticism and Extraversion at least. Later, Jeffrey Gray showed that individual differences in rates of learning corresponded to weighted combinations of Neuroticism and Extraversion.[11] In other words, people who are most prone to anxiety and respond most sensitively to punishment are neurotic introverts (that is, they are high in Neuroticism and low in Extraversion). On the other hand, people who are most impulsive and respond most sensitively to rewards are stable extraverts (that is, they are low in Neuroticism and high in Extraversion). As a result, both Zuckerman and Cloninger developed tests that correspond to these individual differences in learning and vulnerability to psychopathology, as summarized in Table 1.8.2.3. Essentially, people who are most prone to anxiety are those who are described as neurotic introverts by Eysenck, neurotic or anxiety-prone by Zuckerman and Gray, and harm-avoidant by Cloninger. On the other hand, people who are most prone to impulsivity, anger, and substance abuse are called stable extraverts by Eysenck, impulsive sensation-seekers by Zuckerman and Gray, and Novelty seekers by Cloninger.

Later, Cloninger and others showed that all seven of the dimensions of his Temperament and Character Inventory had unique genetic determinants and unique brain processes, suggesting that a seven dimensional model is needed to account for the dynamic processes within each individual that regulates his personality. Nevertheless, five factor models like Zuckerman and Kuhlman's Personality Qeustionnaire or Costa and McCrae's NEO personality inventory can capture most of the information about personality in a statistical sense, even though they ignore the non-linear structure of personality resulting from its complex evolutionary history. Tables 1.8.2.3 and 1.8.2.4 and 4 show the correlations between measures of Cloninger's seven factor model and alternative five factor models (Zuckerman's ZKPQ in Table 1.8.2.3, Costa's NEO-PI in Table 1.8.2.4). As in Eysenck's Neuroticism factor, Neuroticism in five factor models is a composite of anxiety-proneness (as measured by high Harm Avoidance) and personality disorder (as measured by low Self-directedness). Extraversion is a composite of intrapsychic processes involving risk-taking (as measured by low Harm Avoidance), impulsivity (as measured by high Novelty Seeking), and sociability (as measured by high Reward Dependence), and personality maturity (as measured by high Self-directedness).[12] Essentially, Neuroticism and Extraversion are composites of traits leading to maladaptive and adaptive emotional styles. Five factor models now also distinguish traits related to agreeability and sociability (as measured by TCI Reward Dependence and Cooperativeness, low ZKPQ hostility, and high NEO agreeability). There is also consistent recognition of a trait variously identified as conscientiousness, persistence, and vigorous activity, which has been identified by a specific resistance to extinction of intermittently reinforced behaviour regulated by specific brain circuitry in rodents and humans.[13]

Beyond these four personality traits (anxiety-proneness, impulsive anger-proneness, social attachment, and persistence), alternative models of personality vary according to how the remaining features of personality are measured. Five factor models like the ZKPQ and NEO do not measure the personality trait underlying self-awareness, which leads to insight, creativity, and spirituality; however, this trait is measured as Self-transcendence in the TCI. Individual differences in serotonergic receptor function has been found to be strongly related to Self-transcendence.[14]

No consensus is possible to choose among alternative structures derived from factor analysis because an infinity of alternative rotations are statistically equivalent. Information beyond statistics is needed to choose among alternative models, as has been done by Gray, Zuckerman, and Cloninger. Such information includes brain imaging, genetics, development, or utility for developing insight

Table 1.8.2.3 Correlations (r x 100) between the Temperament and Character Inventory (TCI) scales and those of the Eysenck Personality Questionnaire (EPQ-revised) and the Zuckerman-Kuhlman Personality Questionnaire (correlations over 0.4 in bold, significant correlations only shown, n = 207. Reprinted from Personality and Individual Differences, 21, Zuckerman, M. and Cloninger, C.R. Relationship between Coninger's, Zuckerman's, and Eysenck's dimensions of personality, 283–5. Copyright 1996, with permission from Elsevier.

	Harm avoidance	Novelty seeking	Reward dependence	Persistence	Self-directed	Cooperative	Self-transcendent
EPQ Neuroticism	**59**				**– 45**		
EPQ Extraversion	**–53**	**44**	23		18		
EPQ Psychoticism		**41**	**–45**	–29	–31	**–42**	
EPQ Lie		–21			25	34	
ZKPQ Neuroticism	**66**				**–49**		
ZKPQ Impulsive Sensation	–39	**68**	–20				28
ZKPQ Hostility			–27		–32	**–60**	
ZKPQ Sociability	–38	37	31				
ZKPQ Activity	–29			**46**	36		

Table 1.8.2.4 Correlations between the scales of Temperament and Character Inventory-Revised (TCI-R) and the NEO-PI-Revised (correlations over 0.4 in bold, significant correlations only shown, multiple correlation also shown, n = 662, adults in the USA)

	Harm avoidance	Novelty seeking	Reward dependence	Persistence	Self-directed	Cooperative	Self-transcendent	mR
NEO Neuroticism	**63**			–20	**–62**	–28		75
NEO Extraversion	**–55**	40	**52**	40	25		22	77
NEO Openness	–25	**43**	25				37	54
NEO Conscience	–26	–34		**51**	**41**			70
NEO Agreeability		–23	40		31	**61**	20	66
mR	76	65	68	60	67	65	45	

into intrapsychic processes, as described in more detail elsewhere.[7] Fortunately, familiarity with the strong relationships among alternative measures will allow the clinician to interpret flexibly whatever information is available.

The assessment of personality can also be based upon abnormal traits indicative of personality disorder, as has been done by Livesley and others.[15] Whether the starting point is normal or abnormal personality traits, the same structure of personality is observed.[11,16] This shows that personality disorders are particular configurations of traits that vary quantitatively in the general population, not qualitatively in discrete disorders.

Clinical assessment of personality

Personality can be well assessed by allowing the patient to tell his life story and conducting a standard mental status examination. A checklist of signs and symptoms is not adequate for the assessment of personality because narratives only provide an account of a person's continuity of self-awareness over his lifespan. Within the life story, the key elements on which temperament ratings are made are the narrative account of emotional style, particularly in childhood, and general appearance and behaviour on mental status examination. The key elements on which character ratings are based are the range of a person's thoughts, the nature of his interpersonal relationships, and his insight and judgment. The clinician must consider not only the words of the patient, which may involve little or no cognitive insight or self-awareness, but also recognize the significance of non-verbal signs from body posture, facial expression, and gestures to understand his way of perceiving and relating to others.

The level of a person's foresight reflects all of these other sources of information. Lack of foresight is the cardinal feature of personality disorder. Other consistent features of personality disorders are summarized in Table 1.8.2.5.

Temperament involves emotional biases that can be directly observed and felt by an experienced clinician. The tendency of a person to elicit strong emotions from others or 'to get under the skin' of another is a sign of extreme temperament traits or personality disorder. For example, the person with extreme temperament may elicit an urge to be rescued or hostility in the examiner. His general appearance and behavior may be ingratiating or negativistic. Specific features of temperament that distinguish subtypes of personality disorders are summarized in Table 1.8.2.6.

Character traits are assessed partly on intuitive recognition and partly on history. A person who frequently blames others or elicits strong emotional responses in the examiner should be suspected of having a personality disorder. The ratings of character are more precisely based on observations of key functions of self-awareness obtained in the life narrative and the mental status examination. The most informative finding concerns the level of a person's self-awareness, as described in Table 1.8.2.7. The presence of personality disorder means essentially that a person is usually not self-aware (stage 0 in Table 1.8.2.7). Most adults are in the first stage of self-awareness most of the time: they are responsible, have initiative, and are able to delay gratification if they want, but are egocentric. As previously mentioned, they are preoccupied with material concerns. Such individuals may have problems with jealousy or pride, but are sufficiently self-aware so that they are not considered personality disordered. Elsewhere, a simple exercise to evaluate level of self-awareness is described as the Silence of the Mind meditation.[7] It can also be used to help a person improve his level of self-awareness, so it is useful for both assessment and treatment. The ability to reach the second or third stage of self-awareness is the key to improvement in psychotherapy, as described in detail elsewhere.[3] Such growth in self-awareness or

Table 1.8.2.5 Qualitative description of personality disorders

Discriminating features
A maladaptive pattern of responses to personal and social stress that is
stable and enduring since teens
inflexible and pervasive
causing subjective distress
and/or impaired work and/or social relations
Consistent features
lack of foresight (that is, the ability to anticipate what will be satisfying in the long run)
strong emotional reactions elicited from others
(like anger or urge to rescue)
efforts to blame and change others, rather than oneself
Variable features
odd, eccentric
erratic, impulsive
anxious, fearful

Table 1.8.2.6 Qualitative clusters and subtypes of personality disorders according to the American Psychiatric Association (DSM-IV, 1994)

Cluster	Subtype	Discriminating features
Odd/Eccentric	(Low Reward Dependence)	
	Schizoid	socially indifferent
	Paranoid	suspicious
	Schizotypal	eccentric
Erratic/Impulsive	(High Novelty Seeking)	
	Antisocial	disagreeable
	Borderline	unstable
	Histrionic	attention-seeking
	Narcissistic	self-centered
Anxious/Fearful	(High Harm Avoidance)	
	Avoidant	inhibited
	Dependent	submissive
	Obsessive	perfectionistic
Not otherwise specified		
	Passive-Aggressive	negativistic
	Depressive	pessimistic

character traits corresponds closely to the stages of cognitive and character development as described by Piaget, Freud, and Erikson (Table 1.8.2.8). For example, the first stage of self-awareness corresponds to the presence of initiative in Erikson's terms. The second stage involves the presence of generativity. More fine-grained ways to quantify the range of a person's thoughts and human relationships are also described elsewhere.[3] Such refined ratings are important for treatment but not for initial diagnosis.

Table 1.8.2.7 Three stages of self-awareness on the path to well-being (Reproduced from Cloniger, C.R. (2004). *Feeling Good: The Science of Well Being*, with permission of Oxford University Press, New York)

Stage	Description	Psychological characteristics
0	unaware	irresponsible, seeking immediate gratification ('child-like' ego-state)
1	average adult cognition	purposeful but egocentric able to delay gratification, but has frequent negative emotions (anxiety, anger, disgust) ('adult' ego-state)
2	meta-cognition	resourceful and allocentric aware of own subconscious thinking calm and patient, so able to supervise conflicts and relationships ('parental' ego-state,'mindfulness')
3	contemplation	creative and holistic perspective wise, spontaneous, and loving able to access what was previously unconscious as needed without effort or distress able to anticipate what will be satisfying in future ('state of well-being','foresight')

Insight and judgment are also important for assessing character because they are really simply alternative terms for describing the character traits of Self-transcendence and Cooperativeness, as previously discussed. The person's history about his family of rearing, education, marriage, and work history provide the key information for evaluating character. It is important to inquire about a person's goals and his hobbies and recreational activities, whether someone has secure friends, particularly anyone they fully trust and can confide in now or in the past, is important to know as a measure of capacity for intimacy and as a predictor of capacity for forming a therapeutic alliance. Relationships with prior counselors, as well as history of disability claims and law-suits provide important information about personality.

Remember that it is often important with psychiatric patients to assess their personality when they were children or adolescents. In other words, it is as important to evaluate their personality retrospectively as well as in practical, particularly at an age before the onset of other psychopathology, like substance abuse or depression. Current anxiety or depression is expected to inflate Harm Avoidance ratings. Stress or intoxication tends to release temperaments from higher cortical control by character. Likewise, chronic substance abuse, depression, or psychosis arrest character development while active, so early onset of mental disorders is often associated with character deficits. It is usually easy for a patient to provide meaningful information about his childhood personality if the clinician simply asks about the child's early relationships with parents, siblings, schoolmates, and other childhood friends.

Remember also that the single most important dimension of personality to assess in rating a person's level of maturity is his degree of foresight, measured as his Self-directedness. Is the person responsible, or does he tend to blame his problems on other people on unfortunate circumstances? Is the person purposeful, or does he lack clear goals in his life? Is he resourceful, or does he feel inadequate himself and depends on others to solve his problems for him? Assessment of Self-directedness alone is sufficient to determine if a person has a personality disorder of at least moderate severity.[2,4] In contrast, the finding of high Neuroticism is not the same as finding of low Self-directedness, even though they are strongly correlated: a person with anxiety or mood disorder and no personality disorder may be high in Neuroticism but not low in Self-directedness.

Some mild personality disorders also require consideration of the person's capacity to get along with others, as measured by his Cooperativeness. In addition, high functioning individuals who do not merit a diagnosis of personality disorder may nevertheless have specific blind spots in their insight and judgment that leads to severe problems. For example, a competent physician may usually be self-aware but lacks a capacity for intimacy or a sense of fairness in business. Such specific deficits may have severe impairment, even if a person is self-aware in other situations. As a result, it is important to consider the overall profile of a person's life in all five types of situations mentioned earlier. Simply deciding whether or not a person has a personality disorder is insufficient for an assessment of his personality and risk for psychopathology. An adequate initial assessment of a person's personality should allow ratings of all four dimensions of temperament and three dimensions of character, which in turn provide a basis for understanding a person's capacity for well-being and vulnerability to psychopathology.

Table 1.8.2.8 Comparison of different descriptions of character development

Stage of character development	Stage of Piaget	Stage of Freud	Stage of Erikson	Judgment (Cooperative)	Insight (Self-Transcendent)	Foresight (Self-directed)
0	Reflexive					
1	Enactive			Tolerance		
2		Oral	Trust		Trust	
3						Responsible
4	Intuitive	Anal	Autonomy	Forgiving		
5		Phallic	Initiative		Free-Flowing Productivity	
6	Concrete Operations	Latency	Industry			Purposeful
7		Early Genital		Empathic		
8					Transpersonal Identification	
9	Abstract Operations		Identity			Resilient
10		Later Genital	Intimacy	Helpful		
11			Keeper of Meaning		Creativity	
12						Resourceful
13			Integrity	Principled		
14					Holistic Intuition	
15						Spontaneous

Clinical value of psychometric testing

Most experienced clinicians should be able to make valid personality assessments without psychometric testing. However, psychometric testing may still be useful for at least three reasons. First, it helps the clinician to refine his or her clinical assessments by asking more questions with comparisons to normative data than is usually practical during a clinical session. Second, it provides the patient written feedback that can be studied and reflected upon, which does not depend on the clinicians' subjective biases—it reflects back to the patient what was said and provides a language that can be used for accurate communication between the patient and the doctor. Third, it provides a standard for comparison to later assessments as a means of measuring growth. As a result, it is often useful to supplement clinical impression with documentation that allows the patient to describe himself or herself without reliance on the judgment of anyone else. The patient's effort to describe himself or herself often has the therapeutic value of stimulating the patient to begin to understand the motives underlying the pattern of his behaviour. In addition, comparison of psychometric test scores with clinical impression is a helpful way for the clinicians to train themselves in the art of personality assessment.

Further information

The Washington University Center for Well-Being: psychobiology.wustl.edu
The Anthropaideia Foundation: aidwellbeing.org

References

1. Kaasinen, V., and Aalto, S. *et al.* (2004). Insular dopamine D2 receptors and novelty seeking personality in Parkinson's disease. *Movement Disorders*, **19**(11), 1348–51.
2. Cloninger, C. R. (2000). A practical way to diagnose personality disorder: A proposal. *Journal of Personality Disorders*, **14**(2), 99–108.
3. Joyce, P. R., Mulder, R. T. *et al.* (1994). Temperament predicts clomipramine and desipramine response in major depression. *Journal of Affective Disorders*, **30**, 35–46.
4. Cloninger, C. R., Svrakic, D. M. *et al.* (1993). A psychobiological model of temperament and character. *Archives of General Psychiatry*, **50**, 975–90.
5. Joyce, P. R., Mulder, R. T. *et al.* (2003). Borderline personality disorder in major depression: differential drug response, and 6-month outcome. *Comprehensive Psychiatry*, **44**(1), 35–43.
6. Oldham, J. M., Gabbard, G. O. *et al.* (2001). *Practice guidelines for the treatment of patients with borderline personality disorder*. Washington, DC., American Psychiatric Press.
7. Cloninger, C. R. (2004). *Feeling Good: The Science of Well Being*. New York, Oxford University Press.
8. Perls, F. S. (1969). *Gestalt Therapy Verbatim*. Lafayette, California, Real People Press.
9. Lowen, A. (2003). *The language of the body: Physical dynamics of character structure*. Alchua, Florida, Bioenergetics Press.
10. Eysenck, H. J., (ed.) (1981). *A Model of Personality*. New York, Springer-Verlag.
11. Gray, J. A. (1981). A critique of Eysenck's Theory of Personality. *A Model of Personality*. H. J. Eysenck. New York, Springer-Verlag: 246–76.
12. Zuckerman, M. and Cloninger, C. R. (1996). Relationships between Cloninger's, Zuckerman's, and Eysenck's dimensions of personality. *Personality & Individual Differences*, **21**, 283–5.
13. Gusnard, D. A., Ollinger, J. M. *et al.* (2003). Persistence and brain circuitry. *Proceedings of the National Academy of Sciences USA*, **100**(6), 3479–84.
14. Borg, J., Andree, B. *et al.* (2003). The serotonin system and spiritual experiences. *American Journal of Psychiatry*, **160**(11), 1965–1969.

15. Livesley, W. J., Jang, K. L. *et al.* (1998). Phenotypic and genetic structure of traits delineating personality disorder. *Archives of General Psychiatry* **55**, 941–48.
16. Strack, S., (ed.) (2006). *Differentiating normal and abnormal personality*. Springer Publishing, New York.

1.8.3 Cognitive assessment

Graham E. Powell

Principles of assessment

Assessment, testing, or measurement is the evaluation of the individual in numerical or categorical terms, adhering to a range of statistical and psychometric principles. Examples of measurement are, assigning people or behaviour to categories, using scales to obtain self-ratings or self-reports, using tests of ability and performance, or collecting psychophysiological readings. Even diagnosis is a form of measurement and should have various psychometric properties such as satisfactory reliability and validity. In this chapter we concentrate on cognitive or neuropsychological assessment, which typically employs standardized psychometric tests, but it is axiomatic that the basic principles are applicable to all forms of measurement without exception. For example, stating that a patient does or does not have a symptom is potentially just as much of a measurement as stating his or her IQ. It should be noted that this account is of English language tests, and readers elsewhere should note the principles but ask local psychologists what tests they use.

Psychometric tests aim to measure a real quantity—the degree to which an individual possesses or does not possess some feature or trait, such as social anxiety or spelling ability or spatial memory. This real quantity is known in classical test theory as the **true score** *t*, and the score that is actually obtained on the given test is the **observed score** *x*. It is assumed that the observed score is a function of two values, the true score plus a certain amount of **error** *e*, because no test is perfect. Therefore we have the most basic equation in psychometrics: $x = t + e$. The statistical aim of psychometric measurement is to keep the error term to an absolute minimum so that the observed score is equal to the true score, which happens when the error term is reduced to zero. Of course, this is never achieved, but the error term can be reduced to the minimum by making the test as reliable as possible, where reliability is simply the notion that the test gives the same answer twice.

In practice, of course, if a test were repeated many times, each occasion would give a slightly different result, depending on how the person felt, the precise way questions were asked, the details of how answers were scored, or whether there has been any lucky guessing. In other words, observed scores would cluster around the true score. Like the distribution of any variable, the distribution of observed scores would have a mean and a standard deviation. The mean is obviously the true score, and this standard deviation is called the **standard error of measurement (SEM)**. The aim of a good test is to keep the SEM as near as possible to zero, and test manuals should state the actual SEM.

There is a relationship between SEM and the reliability of the test:

$$SEM = SD\sqrt{(1 - r_{11})}$$

where SD is the standard deviation of the test and r_{11} is the test–retest reliability of the test (expressed as a correlation coefficient ranging from −1 to +1). If the reliability of the test is perfect (+1), as can be seen the SEM will be zero:

$$SEM = SD\sqrt{(1-1)} = SD\sqrt{0} = 0.$$

Thus a test should be as reliable as possible because then the observed score will be the true score and the standard error of measurement will be zero.

An unreliable test is always useless, but if reliability can be achieved then it is worth considering the test score and, more specifically, what it measures. The degree to which a test measures what it is supposed to measure is known as **validity**. There may be various threats to validity. For example, a test of numeracy may be so stressful that scores are highly dependent upon the patient's anxiety level rather than on his or her ability, or a test of social comprehension may have questions which are culturally biased and so scores may depend in part upon the person's ethnic background.

In practice, there are various types of reliability and validity, and these are summarized in Tables 1.8.3.1 and 1.8.3.2. Further discussion can be found in Kline.[1]

Having used a reliable and valid test, the next issue is how the numbers are analysed and expressed. It has to be noted first that there are three types of scale of measurement. A **nominal** scale is when numbers are assigned to various categories simply to label the categories in a manner suitable for entry onto a computer database—the categories actually bear no logical numerical relationship to each other. Examples would be marital test status or ethnic background or whether one's parents were divorced or not. Nominal scales are used to split people into groups and all statistics

Table 1.8.3.1 Types of reliability

Scorer or rater reliability	The probability that two judges will (i) give the same score to a given answer, (ii) rate a given behaviour in the same way, or (iii) add up the score properly. Scorer reliability should be near perfect.
Test–retest reliability	The degree to which a test will give the same result on two different occasions separated in time, normally expressed as a correlation coefficient. A reliability of less than 0.8 is dubious.
Parallel-form reliability	The degree to which two equivalent versions of a test give the same result (usually used when a test cannot be exactly repeated because, say, of large practice effects).
Split-half reliability	If a test cannot be repeated and there are no parallel forms, a test can be notionally split in two and the two halves correlated with each other (e.g. odd items versus even items). There is also a mathematical formula for computing the mean of all possible split halves (the Kuder–Richardson method).
Internal consistency	The degree to which one test item correlates with all other test items, i.e. an 'intraclass correlation' such as the a coefficient, which should not drop below 0.7.

Table 1.8.3.2 Types of validity

Face validity	Whether a test seems sensible to the person completing it; i.e. does it appear to measure what it is meant to be measuring? This is in fact not a statistical concept, but without reasonable face validity, a patient may see little point in co-operating with a test that seems stupid.
Content validity	The degree to which the test measures all the aspects of the quality that is being assessed. Again, this is not a statistical concept but more a question of expert judgement.
Concurrent validity	Whether scores on a test discriminate between people who are differentiated on some criterion (e.g. are scores on a test of neuroticism higher in those people with a neurotic disorder than in those without such a disorder!). Also, whether scores on a test correlate with scores on a test known to measure the same or similar quality.
Predictive validity	The degree to which a test predicts whether some criterion is achieved in the future (e.g. whether a child's IQ test predicts adult occupational success; whether a test of psychological coping predicts later psychiatric breakdown). For obvious reasons, these last two types of validity are often jointly referred to as *criterion-reiated validity.*
Construct validity	Whether a test measures some specified hypothetical construct, i.e. the 'meaning' of test scores. For example, if a test is measuring one construct, there should not be clusters of items that seem to measure different things; the test should correlate with other measures of the construct (*convergent validity*); it should not correlate with measures that are irrelevant to the construct (*divergent validity*).
Factorial validity	If a test breaks down into various subfactors, then the number and nature of these factors should remain stable across time and different subject populations.
Incremental validity	Whether the test result improves decision-making (e.g. whether knowledge of neuropsychological test results improves the detection of brain injury).

are based on the frequency of people in each group. The relationship or association between groups can be examined using χ^2 statistics, for example to test whether there is a relationship between being divorced and having parents who divorced. Next there is an **ordinal** scale, in which larger numbers indicate greater possession of the property in question. Rather like the order of winning a race, no assumptions are made about the magnitude of the difference between any two scale points; it does not matter whether the race is won by an inch or a mile. Ordinal scales allow people to be rank ordered and numerical scales can be subjected to non-parametric statistical analysis (which is that branch of statistics which makes minimal assumptions about intervals and distributions), including the comparison of means and distributions and the computation of certain correlation coefficients. Finally comes the **interval** scale in which each scale point is a fixed interval from the previous one, like height or speed. The types of test described in this chapter for the most part aspire to be interval scales, allowing use of the full range of parametric statistics (which assume equal intervals and normally distributed variables).

Having obtained a test score for someone, that score then has to be interpreted in the light of how the general population or various patient groups generally perform on that test. There are two general characteristics of a scale that have to be remembered. The first is the measure of central tendency. Typically one would consider the mean (the arithmetic average), but it is also sometimes useful to consider the median (the middle score) and the mode (the most frequently obtained score). This will be the first hint as to whether the score is normal or whether it is more typical of one group than another. However, in order to gauge precisely how typical a given score is, it is necessary to take into account the **standard deviation (SD)** of the test (other measures relating to the dispersion of test scores, such as the range or skew, can be considered but are not of such immediate relevance).

As long as the mean and SD of the test are known, it is possible to work out exactly what percentage of people obtain up to the observed score x. This is done by converting the observed score into a **standard score** z and converting the z-score to a **percentile**. A standard score is simply the number of SDs away from the mean m, and it will have both negative and positive values (because an observed score can be either below or above the mean, respectively). In other words, $z = (x - m)/\text{SD}$. For reference, Table 1.8.3.3 gives some of the main values of z and what percentage of people score up to those values. It is this percentage that is known as the percentile and it is obtained from statistical tables. For example, a score at the 25th percentile means that 25 per cent of people score lower than that specific score. Obviously, the 50th percentile is the mean of the test. For illustration, the equivalent IQ scores (IQ scores have a mean of 100 and SD of 15) and broad verbal descriptors are also given in Table 1.8.3.3.

A knowledge of percentile scores can help to decide to which category a patient may belong. For example, if a patient completes a token test of dysphasia and scores at the 5th percentile for normal controls and the 63rd percentile for a group of dysphasics, the score is clearly more typical of the dysphasic group.

However, in clinical practice it is often not just a comparison with others that is needed, but a comparison between two of the patient's own scores. For example, verbal IQ might seem depressed

Table 1.8.3.3 z-scores, percentiles, IQ scores, and descriptions

z-score	Percentile	IQ	Description
−2.00	2.5th	70	Scores below the 2.5th percentile are *deficient* or in the *mentally retarded* range
−1.67	5th	75	
−1.33	10th	80	Scores between the 2.5th and 10th percentile are *borderline*
−1.00	16th	85	
−0.67	25th	90	Scores between the 10th and 25th percentile are low average
−0.33	37th	95	
0.00	50th	100	The mean score
+0.33	63rd	105	
+0.67	75th	110	Scores between the 25th and 75th percentile are in the *average* range
+1.00	84th	115	
1.33	90th	120	Scores between the 75th and 90th percentile are high average
		120+	Scores over the 90th percentile are *superior*

in comparison with spatial IQ, or the patient's memory quotient might seem too low for his or her IQ. These are known as **difference scores**, and their analysis is a crucial part of the statistical analysis of a patient's profile. There are two key concepts: the **reliability of difference scores** and the **abnormality of difference scores**. Failure to distinguish between these two leads to all manner of erroneous conclusions. In brief, a reliable difference is one that is unlikely to be due to chance factors, so that if the person were to be retested then the difference would again be found. If the test is very reliable (see the previous discussion of reliability), even a small difference score, may be reliable. As a concrete example, the manual of the Wechsler Adult Intelligence Scale—Third Edition (**WAIS-III**)[1] indicates that a difference of about nine points between verbal IQ and performance IQ is statistically reliable at the 95 per cent level of certainty.

However, although a difference of this size would be reliable, this does not necessarily mean that it is abnormal and therefore indicative of pathology. The abnormality of a difference score is the percentage of the general population that has a difference score of this size or greater. Published tables,[2] show that 18 per cent of adults have a discrepancy of at least 10 points between verbal and performance IQ, so a difference of 10 points is not at all unusual. In fact, to obtain an abnormal difference between verbal and performance IQ the discrepancy has to be of the order of 22 points for adults and 26 points for children (i.e. less than 5 per cent of adults or children have discrepancy scores of this size).

Having introduced the basic concepts of psychometric assessment, this is an appropriate point, prior to the description of specific tests, at which to summarize the information that can (or should) be found in a typical test manual, and this is set out in Table 1.8.3.4.

Table 1.8.3.4 What to expect in a good test manual

Theory	The history of the development of the concept and earlier versions of the test The nature of the construct and the purpose of measuring it
Standardization	Characteristics of the standardization sample, how the sampling was carried out. and how well these characteristics match those of the general population Similar data on any criterion groups Similar data for each age range if the test is for children
Administration	How to administer the test in a standard fashion so as to minimize variability of administration as a factor in the error term
Scoring	How to score the test, and criteria for awarding different scores, so as to minimize scorer error
Statistical properties	Means and standard deviations of all groups Reliability coefficients and how they were obtained Validity measures and how they were derived Standard error of measurement Reliability of difference scores Abnormality of difference scores Other data on the scatter of subtest scores Scores of criterion groups
Special considerations	Groups for whom the test is not suitable or less suitable, i.e. the range of convenience of the tests Ceiling effects: at what point does the test begin to fail to discriminate between high scorers? Floor effects: at what point does the test begin to fail to discriminate between low scorers?

Tests of cognitive and neuropsychological functioning

General ability and intelligence

A very useful broad screening test, especially when it is suspected that mental functions are severely compromised, is the Mini-Mental State Examination.[3,4] It is brief, to the point, and can be repeated over time to gauge change. It measures general orientation in time and place, basic naming, language and memory functions, and basic non-verbal skills, and has good norms for a middle age range, especially the elderly, with appropriate adjustment for age. The maximum score is 30, and a score of 24 or less raises the possibility of dementia in older persons, especially if they have had nine or more years of education (a score of 24 is at about the 10th percentile for people aged 65 and older).

However, the Mini-Mental State Examination is only a screening test and the presence or nature of cognitive impairment cannot be diagnosed on the basis of this test alone. A detailed cognitive assessment is provided by the Wechsler scales, i.e. the Wechsler Adult Intelligence Scale—Third Edition UK Version (**WAIS-III^{UK}**),[1] the Wechsler Intelligence Scale for Children—IV UK Version (**WISC-IV^{UK}**),[5] or the Wechsler Preschool and Primary Scale of Intelligence—Revised (**WPPSI-III**).[6] Outlines of the WAIS-III^{UK} and WISC-IV^{UK} are given in Table 1.8.3.5.

IQ scores themselves are very broad measures, drawing upon a wide range of functions. This does not only mean that the scores are very stable (reliable), but also that the IQ score is relatively insensitive to anything except quite gross brain damage. Rather, a careful analysis of subtest scores is needed, always bearing in mind the concepts of reliability and abnormality of difference scores. For example, it takes a subtest range of 11 to 12 points to be considered abnormal (i.e. found in less than 5 per cent of people) on the WAIS-III^{UK} and the WISC-IV^{UK}.

Sometimes the patient may have a language disorder or English may not be his or her first language. In such circumstances Raven's Progressive Matrices Test,[7] which is a non-verbal test of inductive reasoning (non-verbal in the sense that it requires no verbal instructions and no verbal or written answers), can be used. The present author avoids the new norms because they were not collected in the normal fashion (i.e. not in a formal test session under the direct supervision of a psychologist), but the old norms are good. The Matrices Test has the additional advantage of having an advanced version for people in the highest range of ability.[8] No non-English versions of the WAIS-III^{UK} or the WISC-IV^{UK} are available, but the non-verbal scores can be used with caution as there may be unexpected cross-cultural effects.

Speed of processing

Reasoning is not just about solving difficult problems, but also about solving them quickly; the difference between power and speed. IQ tests as above do have timed subtests sensitive to speed, but it can be useful to administer specific tests that are not quite so confounded with intellectual ability.

One example, particularly sensitive to even quite mild concussion, is the Paced Auditory Serial Addition Test (**PASAT**).[9,10] Here, the

Table 1.8.3.5 Outline of the WAIS-III^{UK} and WISC-IV^{UK}

	WAIS-III^{UK}	WISC-IV^{UK}
Age range	16–89 years	6.0–16.11 years
Verbal subtests	Vocabulary Similarities Arithmetic Digit span Information Comprehension Letter-number sequencing	Similarities Digit span Vocabulary Letter-number sequencing Comprehension Information Arithmetic Word reasoning
Non-verbal or spatial subtests	Picture completion Digit symbol Block design Matrix reasoning Picture arrangement Symbol search Object assembly	Block design Picture concepts Coding Matrix reasoning Symbol search Picture completion Cancellation
IQ score	Verbal IQ (VIQ) Performance IQ (PIQ) Full scale IQ (FSIQ)	Full scale IQ (FSIQ)
Index scores	Verbal comprehension Perceptual organization Working memory Processing speed	Verbal comprehension Perceptual reasoning Freedom from distractibility Processing speed
Mean IQ or index scores	100 (SD of 15)	100 (SD of 15)
Mean subtest scores	10 (SD of 3)	10 (SD of 3)
Test–retest reliability of IQ	0.98 for Full scale IQ	0.97 for Full scale IQ
Standard error of measurement of FSIQ	About 2.5, so all scores are about ±5 points[a]	About 2.68, so all scores are about ±5 points
Reliable differences (p <.05)	About 9 points between VIQ and PIQ	About 11 points between VCI and PRI
Abnormal differences (p <.05)	About 22 points between VIQ and PIQ	About 26 points between VCI and PRI
Validity	Highly related to other tests of ability and to criteria related to ability	Highly related to other tests of ability and to criterion groups

[a]95% of the time, true scores are the observed score ±1.96 SEM. In other words, the likely true score is within the range defined by about 2 SEMs either side of the score obtained.

client is read a list of numbers, and as each one is read out so it has to be added to the previous number and the answer spoken aloud (Table 1.8.3.6). This has to be done quickly or the next number will come along. There are several trials in which the numbers are delivered at a faster and faster pace, from one number every 2.4 s down to every 1.2 s. It sounds easy but in actuality is very demanding; even at the slowest speed the average score is only about 70 per cent correct, and this falls away to only about 40 per cent at the fastest speed. Indeed, if a patient has any significant mental slowing, they often cannot do the test at all. Obviously the test cannot be used if the patient has a stammer, or is dysarthric or innumerate.

A less stressful test of mental speed is the Speed of Comprehension Test,[11] in which the person indicates as fast as possible whether simple sentences are true or false (e.g. tomato soup is a liquid, grapes are people). The test can be given orally for patients who cannot read.

Two visual tests of mental speed are Map Search (looking for target symbols on a map as fast as possible) and Telephone Search (looking for various symbols on a page from a telephone directory).[12]

One test that tries to disentangle the relative contribution of slowed motor speed versus slowed mental speed, often a crucial issue in patients with motor deficits, is the Adult Memory and Information Processing Battery,[13] which has two useful timed tests of cancelling target digits.

Attention and concentration

There are various aspects of attention and concentration: the ability to focus resources, the ability to focus on the right aspect, the ability to sustain this attention, the ability to ignore extraneous

Table 1.8.3.6 Sample from PASAT

Number on tape	(Mental process)	Patient says
7		
5	(5+7)	12
1	(1+5)	6
4	(4+1)	5
9	(9+4)	13

information or distracting events, and the ability to divide attention between different tasks. The tests on speed listed above are of course also measures of attention, because highly focused and selective attention has to be sustained for the duration of a pressured task. Digit span on the WAIS-III^{UK} is also a test of attention, as any lapse in attending to the incoming digits will necessarily result in a wrong answer.

However, in addition to these tests, a battery may be used, such as the Test of Everyday Attention[12] which has eight different subtests. Test–retest reliability is quite good, over 0.83 for Map Search and Telephone Search, for example. In terms of validity, the tests are very sensitive to the effects of head injury and stroke.

Memory

Memory is a complex set of processes whereby the person registers, stores, and retrieves information within different modalities (e.g. verbal memory versus spatial memory) and across different time periods (e.g. primary or shorter-term memory versus secondary or longer-term memory or learning). Therefore, as with intelligence, various batteries have evolved with subtests that tap these various aspects. Two examples of batteries are the Wechsler Memory Scale—Third Edition[14] for adults and the Children's Memory Scale,[15] which are both summarized in Table 1.8.3.7. Another battery, which makes a special effort to reflect real-life tasks, is the Rivermead Behavioural Memory Test,[16] which also has a child's version.[17]

Sometimes time constraints make it difficult to give complete memory batteries. Often, just a few key subtests are selected, or other individual tests may be given. For example, to gauge verbal learning the Rey Auditory-Verbal Learning Test is well researched,[18] a test of visual memory is the Rey–Osterrieth Complex Figure Test[19]; and a forced choice recognition tests for words and faces is the Recognition Memory Test.[20] If the ability of the patient to recall details of his or her past life is an issue, the Autobiographical Memory Interview can be used.[21]

Language

Commonly used batteries for the assessment of language deficits are the Boston Diagnostic Aphasia Examination[22] and the closely related Western Aphasia Battery.[23] The Boston Examination covers auditory comprehension, oral expression, understanding written language, and writing. These tests can take a long time to give and so brief screening tests are often used, such as the Boston Naming Test[24] or the Graded Naming Test,[25] which both assess word finding, or the Token Test,[26] which assesses verbal comprehension. Finally, a good test to gauge reading and spelling ability is the Wechsler Objective Reading Dimensions Test[27] which will produce reading and spelling ages, and give the abnormality of difference scores between IQ and reading or spelling scores.

Frontal and executive functions

The term executive function derives from the theory that there is a supervisory system exerting executive control of attention. Deficits of this system cause broad patterns of cognitive and behavioural change called the dysexecutive syndrome,[28] which includes changes in volition, poor planning, a disruption of purposive action, and reduced efficacy of performance. One of the most frequent causes of this syndrome is damage to the frontal lobes (frontal-lobe syndrome is a dysexecutive syndrome) but it may also be caused by other patterns of lesion. Table 1.8.3.8 lists some of the features of the dysexecutive syndrome, and examples of tests that are sensitive to them.

Table 1.8.3.7 Summary of two memory batteries

	Wechsler memory Scale-III	Children's memory scale
Age range	16–89 years	5–16 years
Subtests	Information and orientation Logical memory[a] Faces[a] Verbal paired associates[a] Family pictures[a] Word lists[a] Visual reproduction[a] Letter—number sequencing Spatial span Mental control Digit span (Means typically 10, SDs typically 3)	Dot locations[a] Stories[a] Faces[a] Word pairs[a] Family pictures[a] Word lists[a] Numbers Sequences Picture locations (Means of 10, SDs of 3)
Index Scores	Auditory immediate Visual immediate Immediate memory Auditory delayed Visual delayed Auditory recognition delayed General memory Working memory (Means of 100, SDs of 15)	Verbal immediate Verbal delayed Verbal delayed recognition Learning Visual immediate Visual delayed Attention/concentration General memory (Means of 100, SDs of 15)
Reliability	0.60–0.87 for the index scores (i.e. rather low, and note a practice effect of up to 15 points across 5 weeks)	0.76–0.91 for the index scores (note a large practice effect of about 10–15 points across a 2-month interval)
Standard error of measurement	3.88–7.40 (so true scores are at best ±8 points from the observed score)	4.5–7.4 (so true scores are at best ±9 points from the observed score)
Validity	See manual for content, criterion-related, construct, and other types of validity	See manual for content, construct, and criterion-related validity

[a]Also delayed trial.

Some clinical issues

Sources of tests and test data

A good summary of the principles of test theory is given by Halligan *et al.*[36] Information about tests and where to order them from can be found in Lezak's *Neuropsychological Assessment*,[37] Strauss's *Compendium of Neuropsychological Tests*[2] and Mitrushina's[38] *Handbook of Normative Data*.

Understanding tests

The onus is upon the test user to be sufficiently knowledgeable about test theory to gauge the strengths and limitations of tests. Common problems with tests are small standardization sample sizes, unknown or unstable factor structure, poor or no theoretical adequacy, poor or no use of criterion groups, vague scoring

Table 1.8.3.8 Features and tests of the dysexecutive syndrome

Features	Tests
Behavioural change	The Dysexecutive Questionnaire (DEX), both self-report and other report[29]
Planning and impulsivity	Mazes subtest of the WISC-IV[UK][7]
Fluency	Of generating words and designs, the DKEFS[30]
Concept formations and ability to shift mental set	Modified Card Sorting Test[31] Rule Shift Cards Test[33]
Estimation	Of various amounts, the Cognitive Estimates Test[32,33] Temporal Judgement Test of time estimation[29]
Alternating plans	Switching between plans based on numbers or letters, the Trail Making Test, see DKEFS[30]
Screening out distracting information	The Stroop Test,[34] e.g. reading the word 'BLUE' when it is printed in red ink
Suppression of competing responses	The Hayling Test[35] which requires patients to choose connected or unconnected words to finish a sentence
Rule attainment	Brixton Test, requiring the patient to learn the rule whereby a pattern changes[35]
Planning	Action Program Test, to use given materials to achieve a given end[29] Zoo Map Test of organizing a route[29] Modified Six Elements Test of planning the order of tasks according to rules[29]

criteria, and poor or no information on difference scores within or between tests.

Even a well-normed and proven test may not actually be applicable to the particular patient at hand; tests only have a certain range of convenience. Tests have to be very carefully chosen when confounding factors are present, such as when English is not the patient's first language or when there is a sensory or motor deficit.

Indeed, there are a range of potentially confounding variables even in those patients who are within the range of convenience of the test. These include effects of medication, fatigue as the testing progresses, motivation, mental state, disturbed behaviour, cultural background and beliefs, and educational background. Test manuals may provide information on such potentially confounding variables, but often it is necessary to know the primary research on the test and its sensitivity to such factors.

Assessing children

Special issues arise in the assessment of children because the neuropsychology theory and conceptual framework are different, the effects of specified lesions change with age, the pattern of recovery of function varies with age, extensive developmental norms, covering the age range, are needed, and children may be more stressed by tests or may find it harder to cooperate with test procedures. These and other issues are fully discussed in texts on developmental neuropsychology,[39] paediatric neuropsychology,[40] and head injury in children.[41] Several children's tests have already been cited by name in preceding sections, and many of the adult tests cited have also been standardized on children, often in subsequent research studies not necessarily carried out by the original test author. Examples of tests with children's norms are word fluency, design fluency, auditory verbal learning, the original Wechsler Memory Scale, the Stroop Test, the Token Test, the Trail-making Test, Wisconsin Card Sorting, the Paced Serial Addition Test, and the Progressive Matrices Test.

Assessing premorbid ability

In order to understand the effects of a brain injury, to gauge intellectual loss, to plan rehabilitation, and to advise on issues relating to personal injury compensation, it is necessary to estimate premorbid intelligence. A summary of strategies for estimating premorbid IQ is given in Table 1.8.3.9.

Capacity

Within the context of intellectual and cognitive functioning, a person is incapable of managing his or her own affairs if two

Table 1.8.3.9 Strategies for estimating premorbid IQ

Strategy	Test and/or comment
Assume highest subtest score on the WAIS-III represents original level	Normal individuals have a profile of abilities and show quite a wide range of subtest scores. It makes no sense at all to say that a person's best score is his or her 'real' potential. Anyone using this method will grossly over-estimate IQ loss
Consider scores on subtests thought to be relatively insensitive to the effects of brain injury	Vocabulary is highly correlated with IQ and is also such a deeply ingrained ability that it is relatively insensitive to the affects of brain injury. Therefore, scores on the Vocabulary subtest can indeed be a guide to premorbid IQ
Gauge IQ from educational record	This is reasonable as long as the person had (i) full access to education, and (ii) the motivation to take and pass exams. Gauging ability band is made easier now that national statistics on examination pass rates are published annually in the UK
Gauge IQ from occupational record	Again this is reasonable as a broad approximation, but cultural and sociological constraints on choice of work or progress in work have to be taken into account
Tests of overlearned skills such as reading	Reading ability is highly correlated with IQ, and is a very overlearned skill, not easily affected by brain injury. This is the best and safest method, as long as the patient had no history of dyslexia and is not currently dysphasic. Tests include the Wechsler Test of Adult Reading[42] and the Spot the Word Test[11]
Genetic endowment	If the patient was damaged at birth, or if the damage caused gross physical deficits adversely affecting educational and occupational potential, then the ability and educational and occupational record of natural parents and siblings may be considered

criteria are met. First, they must have an objective deficit likely to impair problem-solving and decision-making. Second, they must be incapable of sensibly delegating or of appropriately seeking advice. Cognitive assessment obviously bears upon both of these issues. In the first instance assessment can help gauge whether there is a deficit at all, and if so its severity and precise nature. For example, most people would be able to manage their own lives despite some mild reduction in intellectual efficiency or some mild memory problem—after all, this is in any case the course of natural ageing but it is much more difficult to cope with a severe memory deficit. In the second instance, cognitive assessment can point to deficits which make it unlikely that the person can appropriately delegate certain responsibilities. For example, those with dysexecutive syndrome may be gullible or impulsive over whom to approach for advice, or may be reluctant to accept advice, may delegate only inconsistently, or may say they accept certain advice but then do the opposite. In short, they cannot plan to delegate, or if they do make such plans, the plans are poorly monitored and inconsistently implemented.

Malingering

There is no single test of malingering (i.e. consciously motivated deliberate underperformance on tests). Rather, a pattern builds up which gradually raises the suspicion of malingering.[2] Features of test performance which raise the issue are as follows:

1 a degree of deficit that is disproportionate to the severity of the injury;

2 bizarre errors not typically seen in patients with genuine deficits;

3 patterns of test performance that do not make sense;

4 not showing expected patterns;

5 inconsistencies between test performance and behaviour in real life;

6 inexplicable claims of remote memory loss even for important life events like weddings;

7 random responding on forced-choice tests;

8 below random responding on forced-driven tests;

9 poor performance on effort tests that look hard but are in fact easy;

10 the absence of severe anxiety or profoundly low mood such as might cause a collapse in performance;

11 after head injury, the absence of any improvement or indeed a worsening of performance over time;

12 failure to report deficits following a brain injury when in retrospect those deficits are claimed to have been severe;

13 relative absence of a history of somatization or related disorders.

Typical clinical neuropsychological assessment

Having set out the theory, tests, and issues, we can build up a picture of a typical clinical neuropsychological assessment, as given in Table 1.8.3.10.

Table 1.8.3.10 Typical protocol for a clinical neuropsychological assessment and report

Aims	The purpose of the assessment, e.g. to describe deficits. monitor improvement, inform rehabilitation planning, address certain specific issues
Background of patient	Information relevant to the interpretation of test findings, e.g. language, handedness, age, educational history, occupational history, medical and psychiatric history
Nature of the brain injury	For example, time since injury, age at injury, mechanisms of injury, retrograde amnesia, loss of consciousness, post-traumatic amnesia, results of neurological examination, results of scans
Behaviour and mental state	Motivation, co-operation with procedures, anxiety, mood, any aspect that threatens reliability or validity, a clear statement as to whether or not reliability has been compromised
Intelligence	Verbal, non-verbal, skills profile
Speed	Verbal, non-verbal, motor
Attention and concentration	Verbal/spatial tasks General behaviour and lapses on tests
Memory	Verbal/non-verbal Immediate/delayed recall Learning
Language	Reading Screening tests for dysphasia Aphasia battery if needed
Construction skills	Refer to performance subtests Copying a complex figure
Sensory deficits	Note gross deficits and subjective account Record problems noted on tests Refer to neurological examination Tests of spatial neglect Tests of visual agnosia
Motor deficits	Note gross problems and subjective accounts Record problems noted on tests Refer to neurological examination Tests of apraxia
Executive functions	Fluency Planning Estimation Personality/behaviour Record dysexecutive problems noted on other tests
Life situation	Way of life, typical day, leisure activities Nature and amount of any support Impact of deficits upon everyday living
Interview with other informant	An observer's account of deficits, changes, coping, etc.
Formulation	A coherent account of the injury and its repercussions, taking all information into account, focusing on the aims of the assessment

Generalizability theory and ecological validity

There is a broad issue how to generalize from an observation made in one context to what might be observed in other contexts. Traditional divisions between reliability and validity become blurred because they are both expressions of the degree to which a score can be generalized. This sweeps away the notion of a true score, to be replaced by the notion of a universe score, which is the mean of all possible observations under all possible conditions. Classical test theory is replaced by generalizability theory, in which variance in a test score is apportioned to various factors. However, in practice generalizability theory informs test construction rather than replacing classical test theory.[15]

There is one very important aspect of generalizability, and this is ecological validity or the degree to which a test score predicts real-life functioning. Some of the various threats to ecological validity are listed in Table 1.8.3.11 (see Long[43] for a fuller discussion), but the recent trend is to make neuropsychological tests increasingly a distillation of real-life tasks so as to lessen this generalizability problem.[12,16,29]

Table 1.8.3.11 Threats to the ecological validity of test results

The assessment session	Data are collected in a quiet sterile focused environment, whereas real life is noisy and full of distractions
Type of test	Cognitive tests are often constructed to measure a single pure aspect of processing, whereas real-life tasks are multidimensional
Type of interaction	The behaviour of the patient is constrained by the nature of the examiner-patient relationship, and is unlike spontaneous behaviour
Content of tests	The limited number and content of tests that can be given may not tap the real-life problems that are complained of (e.g. reduced sense of humour)
Confounding factors	Test anxiety Motivation to co-operate with the assessment or to perform in a certain way Short test sessions to avoid fatigue whereas most problems are reported when the patient *is* fatigued
Over-reliance on test data	Blinkered adherence to numbers to the exclusion of background information, general observation, information from others, and common sense
Failure to consider ecological validity	Lack of understanding of the issue Failure to follow up patients in such a way as to obtain feedback on the ecological validity of the original assessment, which is necessary to shape ecologically valid assessment procedures
Solutions	Use tests of concentration and distraction Develop new tests or new versions of tests reflecting real-life tasks Find sources of information about real-life behaviour Continue to widen the range of tests available, focusing test development on clinical need Estimate effects of confounding factors Treat numbers as only one form of data Specifically address ecological issues in the final report

Cognitive assessment of psychiatric disorders

It is possible to give only a summary of findings from the cognitive assessment of psychiatric and neuropsychiatric disorders. A fuller account is given by Grant and Adams[44] and McCaffrey.[45]

Epilepsy

There is no single cognitive profile of people with epilepsy. The relationship between epilepsy and the presence of mental retardation (learning disability) will mainly be mediated by the original brain damage causing both the epilepsy and the mental retardation. However, some patients do deteriorate intellectually if seizures are frequent or uncontrolled or if there are lapses into status epilepticus. In terms of partial seizures, the most common pattern is disturbance of verbal memory if there is a left temporal (dominant) focus and of non-verbal memory if there is a right focus. Anticonvulsants themselves may mildly impair performance on a wide variety of intellectual, cognitive, and speeded tasks.

Parkinsonism

The pattern of deficits in patients without overt dementia is memory disturbance and dysexecutive syndrome (e.g. reduced fluency, concept formation, ability to shift set). If there is an overt (subcortical) dementia, aphasia, agnosia, and severe amnesia are relatively uncommon, but mood change is frequent.

Dementia

The most common early sign of Alzheimer's disease is poor performance on delayed verbal memory, possibly with dysexecutive signs, eventually joined by a deterioration in the meaningfulness of speech with a breakdown in semantic relationships and understanding; speech becomes empty of content and frontal dysexecutive deficits emerge.

Depression

In younger neurologically intact persons, depression affects attention and memory. After head injury, the presence of anxiety or depression can make a significant to test scores, including IQ, mental speed, and verbal and spatial memory.

Alcohol

There is a typical neurocognitive profile found in chronic detoxified alcoholics after 2 to 4 weeks abstinence: intact IQ and verbal skills, but impairment of novel problem-solving, abstract reasoning, learning and memory, visual spatial analysis, and complex perceptual–motor integration. If severe thiamine deficiency arises, Wernicke–Korsakoff syndrome may ensue, with profound anterograde amnesia.

Other drugs[37]

Findings regarding the long-term neuropsychological effects of marijuana are equivocal, but if there are long-term changes they probably involve attention. Long-term cocaine use may also affect attention and memory. There are conflicting reports about the long-term use of opiates, but there may be a diffuse effect upon visuospatial and visuomotor activities. Chronic solvent abuse

leads to cerebellar ataxia and also some impairment of IQ and memory.

Schizophrenia

There is a growing awareness of dysexecutive deficits in the aetiology of schizophrenic symptoms, relating to disorders of willed action for example. Patients with schizophrenia score poorly on the Behavioural Assessment of the Dysexecutive Syndrome and show other dysexecutive features.

Summary, conclusion, and future directions

Psychometric methods based on classical test theory have permitted the development of reliable and valid tests assessing a wide range of intellectual and cognitive functions. Test results assist in formulation and diagnosis, guide rehabilitation and management, provide baseline measures to detect change, and generally assist clinical decision-making regarding such issues as capacity. Tests and assessment procedures are being further developed so as to improve their ecological validity, enabling better prediction of real-life behaviour and functioning.

Further information

Appleton, R. and Baldwin, T. (ed.) (2006). *Management of brain-injured children*. Oxford University Press, New York. AssessmentPsychology.com

Halligan, P.W. and Wade, D.T. (ed.) (2005). *Effectiveness of rehabilitation for cognitive deficits*. Oxford University Press, London.

Stern, Y. (ed.) (2007). *Cognitive reserve: theory and applications*. Taylor & Francis, New York.

References

1. Wechsler, D. (1997). *Wechsler Adult Intelligence Scale* (3rd edn). Psychological Corporation, Cleveland, OH.
2. Strauss, E., Sherman, E.M.S., and Spreen, O. (2006). *A compendium of neuropsychological tests: administration, norms, and commentary* (3rd edn). Oxford University Press, New York.
3. Folstein, M.F., Folstein, S.E., and McHugh, P.R. (1975). 'Mini-Mental State'. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **12**, 189–98.
4. Tombaugh, T.N., McDowell, I., Krisjansson, B., *et al.* (1966). Mini-Mental State Examination (MMSE) and the modified MMSE (3MS): a psychometric comparison and normative data. *Psychological Assessment*, **8**, 48–59.
5. Wechsler, D. (2004). *Wechsler Intelligence Scale for Children* (4th edn) *UK: manual*. Harcourt Assessment, London.
6. Wechsler, D. (1989). *Manual for the Wechsler Preschool and Primary Scale of Intelligence* (revised). Psychological Corporation, San Antonio, CA.
7. Raven, J.C. (1996). *Progressive matrices*. Oxford Psychologists Press, Oxford.
8. Raven, J.C. (1994). *Advanced progressive matrices, sets I and II*. Oxford Psychologists Press, Oxford.
9. Gronwall, D.M.A. (1977). Paced auditory serial addition task: a measure of recovery from concussion. *Perceptual and Motor Skills*, **44**, 367–73.
10. Stuss, D.T., Stethem, L.L., and Pelchat, G. (1988). Three tests of attention and rapid information processing: an extension. *Clinical Neuropsychologist*, **2**, 246–50.
11. Baddeley, A.D., Emslie, H., and Nimmo-Smith, I.N. (1992). *The speed and capacity of language-processing tests*. Thames Valley Test Company, Bury St Edmunds.
12. Robertson, I.H., Ward, T., Ridgeway, V., *et al.* (1994). *The test of everyday attention*. Thames Valley Test Company, Bury St Edmunds.
13. Coughlan, A.K. and Hollows, S.E. (1985). *The Adult Memory and Information Processing Battery (AMIPB): test manual*. Psychology Department, St James's University Hospital, Leeds.
14. Wechsler, D. (1997). *Wechsler Memory Scale* (3rd edn). Psychological Corporation, San Antonio, CA.
15. Cohen, M.J. (1997). *Children's Memory Scale*. Psychological Corporation, San Antonio, CA.
16. Wilson, B., Cockburn, J., and Baddeley, A. (1985). *The Rivermead Behavioural Memory Test*. Thames Valley Test Company, Bury St Edmunds.
17. Wilson, B.A., Ivani-Chalian, R., and Aldrich, F. (1991). *The Rivermead Behavioural Memory Test for Children Aged 5–10 Years*. Thames Valley Test Company, Bury St Edmunds.
18. Schmidt, M. (1996). *Rey Auditory–Verbal Learning Test*. Western Psychological Services, Los Angeles, CA.
19. Meyers, J. and Meyers, K. (1995). *The Meyers scoring system for the Rey Complex Figure and the Recognition Trial: professional manual*. Psychological Assessment Resources, Odessa, TX.
20. Warrington, E.K. (1984). *Recognition Memory Test manual*. NFER–Nelson, Windsor.
21. Kopelman, M., Wilson, B., and Baddeley, A. (1990). *The autobiographical memory interview*. Thames Valley Test Company, Bury St Edmunds.
22. Goodglass, H. and Kaplan, E. (1983). *The assessment of aphasia and related disorders* (2nd edn). Lea and Febiger, Philadelphia, PA.
23. Kertesz, A. (1982). *Western aphasia battery*. Psychological Corporation, San Antonio, CA.
24. Kaplan, E.F., Goodglass, H., and Weintraub, S. (1983). *The Boston Naming Test* (2nd edn). Lea and Febiger, Philadelphia, PA.
25. McKenna, P. and Warrington, E.K. (1983). *Graded Naming Test*. NFER–Nelson, Windsor. (New norms in 1997.)
26. McNeil, M.M. and Prescott, T.E. (1978). *Revised Token Test*. Pro-ed, Austin, TX.
27. Wechsler, D. (1993). *Wechsler objective reading dimensions: manual*. Psychological Corporation, London.
28. Baddeley, A.D. and Wilson, B.A. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition*, **7**, 212–30.
29. Wilson, B.A., Alderman, N., Burgess, P.W., *et al. Behavioural assessment of the dysexecutive syndrome*. Thames Valley Test Company, Bury St Edmunds.
30. Delis, D.C., Kaplan, E., and Kramer, J.H. (2001). *Delis Kaplan Executive Function System (DKEFS): manual*. The Psychological Corporation, San Antonio, TX.
31. Nelson, H.E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, **12**, 313–24.
32. Shallice, T. and Evans, M.E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex*, **14**, 292–303.
33. O'Carroll, R., Egan, V., and Mackenzie, D.M. (1994). Assessing cognitive estimation. *British Journal of Clinical Psychology*, **33**, 193–7.
34. Sachs, T.L., Clark, C.R., Pols, R.G., *et al.* (1991). Comparability and stability of performance of six alternate forms of the Dodrill–Stroop Color–Word Tests. *Clinical Neuropsychologist*, **5**, 220–5.
35. Burgess, P.W. and Shallice, T. (1997). *The Hayling and Brixton tests*. Thames Valley Test Company, Bury St Edmunds.
36. Halligan, P.W., Kischka, U., and Marshall, J.C. (2003). *Handbook of clinical neuropsychology*. Oxford University Press, New York.

37. Lezak, M.D., Howieson, D.B., Loring, D.W., *et al.* (2004). *Neuropsychological assessment* (4th edn). Oxford University Press, New York.
38. Mitrushina, M., Boone, K.B., Razani, J., *et al.* (2005). *Handbook of normative data for neuropsychological assessment* (3rd edn). Oxford University Press, New York.
39. Spreen, O., Risser, A.H., and Edgell, D. (1995). *Developmental neuropsychology*. Oxford University Press, New York.
40. Baron, I.S, Fennell, E.B., and Voeller, K.K.S. (1995). *Paediatric neuropsychology in the medical setting*. Oxford University Press, New York.
41. Broman, S.H. and Michel, M.E. (ed.) (1995). *Traumatic head injury in children*. Oxford University Press, New York.
42. The Psychological Corporation. (2001). *Wechsler Test of Adult Reading (WTAR): manual*. The Psychological Corporation, San Antonio, TX.
43. Long, C.J. (1996). Neuropsychological tests: a look at our past and the impact that ecological issues may have on our future. In *Ecological validity of neuropsychological testing* (ed. R.J. Sbordone and C.J. Long), pp. 1–41. GR Press–St Lucie Press, Delray Beach, FL.
44. Grant, I. and Adams, K.M. (ed.) (1996). *Neuropsychological assessment of neuropsychiatric disorders* (2nd edn). Oxford University Press, New York.
45. McCaffry, R.J., Palav, A.A., O'Bryant, S.E., *et al.* (ed) (2003). *Practitioner's guide to symptom base rates in clinical neuropsychology*. Kluwer Academic/Plenum Publishers, New York.

1.8.4 Questionnaire, rating, and behavioural methods of assessment

John N. Hall

The earliest forms of psychiatric assessment were based on direct interviews with patients, on reported observations by those who knew the patient, and on direct observations by attendants—later nurses—in the care setting. Attempts to codify these forms of assessment had begun over 90 years ago, as illustrated by the 'Behavior Chart' of Kempf.[1] The present range of structured psychiatric assessment methods grew from the 1950s in association with the introduction of neuroleptic medication and the development of psychiatric rehabilitation programmes. The two most frequently used types of systematic and structured assessment used in both clinical practice and research continue to be questionnaires and ratings. Their value lies in the systematic coverage of relevant content, and the potential for comparing scores across individuals and groups and over time.

This section covers assessment methods that are appropriate for both self-report by patients and others—questionnaires—and observations and judgements made by others about the patient and their immediate circumstances—rating methods. This section will also briefly describe behavioural approaches to assessment of clinical relevance.

Questionnaires offer the respondent a preset range of written questions covering the area of clinical interest, such as depression. The questions are usually completed by marking one of a set of provided response categories (**forced-choice** questions), but may be completed by the patient writing their own response in free text.

Self-report and 'self-monitoring' methods are similar to the latter form of questionnaire, in that the patient completes a diary or premarked sheets. These are more open-ended, and any associated thoughts of the patient may be included. Self-report measures are used widely in cognitive behavioural interventions.

Ratings are judgements about the quality or characteristics of a defined attribute or behaviour, completed subjectively, or on the basis of direct observation of the behaviour in question. While questionnaires are usually self-completed, ratings may be completed by one person with respect to another person. In psychiatric practice, ratings include those made by professional staff, often a nurse or care worker, or by a family member or informal carer, about a patient.

Ratings and behavioural measures have a special use in the assessment of disturbed or bizarre behaviour, where the patient may have little insight or knowledge of the nature or degree of their disturbance, which may pose a major ongoing management problem, or a barrier to their placement in the community. An example of such a measure is the Aberrant Behavior Checklist.[2] This is a 58-item behavioural rating scale completed by an informant, with the content covering five subscales: irritability, agitation, and crying; social withdrawal and lethargy; stereotyped behaviour; hyperactivity and non-compliance; and inappropriate speech.

The purpose of questionnaire, rating, and behavioural assessments

Scales may be used for a number of purposes:

- for the initial assessment of a patient as part of a clinical formulation
- for ongoing monitoring during the course of treatment
- as outcome measures
- for assigning patients from a larger population to a particular therapeutic regime
- for service planning

Normally an assessment will focus on the presented or referred patient. However, it may be helpful to either focus on a family member or on a formal or informal direct carer of the patient. Another potential focus is the patient's environment. The range of behaviour a patient can display is limited by the physical nature of their environment, by the range of equipment or materials available to the patient, and by the social rules of the setting (such as rules against smoking). A rating of environmental restrictiveness would then survey both environmental constraints and the range of formal institutional regulations and informal rules followed by care staff.

Most measures simply describe the current functioning of the patient, without offering a framework for translating the obtained scores into clinical priorities for treatment. An important development in rating methodology has been the 'needs assessment' approach that incorporates the views of patients and carers when taking into account the extent to which their needs have been met, or remain unmet. The Camberwell Assessment of Need (CAN) family of measures[3] has adopted a consistent set of content domains, which has been applied to separate need assessment schedules which can now be applied to adults, older adults, people with learning disabilities, and in forensic settings.

Scale content

A questionnaire or rating is defined by both **overall content** and **item format**. The content of a measure should logically be determined by its purpose. One model of assessment[4] suggests that there are four main content areas for assessment, including cognition, affect (including verbal-subjective components of behaviour), physiological activity, and overt behaviour. The content should cover all the domains of clinical relevance, including current and past behaviour, and psychopathology. While most rating scales cover a relatively limited number of functional areas, and are often designed for use with a specific client group or clinical population, some measures are designed for wider use.

The format of each item typically consists of an **item stem**, or question, followed by a set of **response options**. The item stem and responses should be grammatically complementary, and the total set of response options for each item should together cover all logically possible response options. Responses should use exact frequencies (such as 'twice a day' or 'at least every hour') rather than vague terms such as 'often' or 'frequently'. Response options may be set out verbally, or may have a numerical value attached. Usually the responses for each item will be laid out in sequence to form a graded series of increasing or decreasing severity or quality of response. These items may be set out in **unipolar** (where one end is 'zero' or nil occurrence) or **bipolar** (with the mid-point being neutral or 'normal') form. In general, if an item has more than five response options there is a risk of poor reliability. Figure 1.8.4.1 illustrates the most common individual item formats.

Most items in questionnaires and rating scales are designed to produce ordinal scores—that is, the score simply gives the relative order of items, without implying any mathematical equality of the differences between scores. This limits the statistical methods that can be used with the scores arising from these measures.

Criteria for evaluating questionnaires and rating scales

There are a number of technical and practical factors to bear in mind in appraising and selecting a measure. Anyone using a published questionnaire or rating scale should examine the technical qualities of the measure, which should be included in the original publication or on a scale manual. This is both to critically assess whether or not a specific measure is suitable for the intended purpose, and also to understand how best to use the scale in practice—including any training requirements.

Direct frequency count

How many cigarettes a day do you smoke? ....

Dichotomous or binary item (used in checklists)

Have you ever smoked a cigarette? Yes ☐ No ☐

Three or more response options

a. Nominal scaling: please describe your marital status

Single ☐ Married ☐ Separated/divorced ☐ Widowed ☐

b. Ordinal Unipolar: how often do you feel constipated?

Never ☐ Between 1 and 3 times a week ☐ 4 or more times a week ☐

c. Ordinal Bipolar: how good are you as a car driver?

Above average ☐ average ☐ below average ☐

Linear or **Visual analogue** scale (the line is usually 10 cm long)

What quality of care have you received?

Very poor ——— Very good

Fig. 1.8.4.1 Examples of item formats.

Psychometric adequacy

Psychometric criteria are the most important technical ways to evaluate the quality of a measure. Chapter 1.8.3 outlines those minimum psychometric principles and properties that are applicable to all psychological measures, including the classical approaches to scaling, reliability, validity, and sensitivity to change, as well as generalizability approaches. The most widely known psychometric requirements of any scale are validity and reliability: neither of these are inherent properties of a scale. Scales are valid for specific purposes, which should be clearly described in the original published article about a scale.

The form of reliability most characteristic of rating scales, and of other behavioural measures, is inter-rater or inter-observer reliability, which examines the similarity of scores when two or more different raters administer a scale. Ideally the manual for a scale should describe a rater training procedure. If not, it is always sensible to carry out some basic rater training, carrying out some assessments prior to the main study, analyzing score discrepancies, and ensuring that raters have discussed these differences and why they arose.

Typically, ratings may be completed after an observation period varying from a few minutes or hours, to a few days. For those ratings based on observation periods of more than a day, there is then a variable delay between the relevant observations and the completion of the rating, and also any one observer will only have been present or on duty for a proportion of the observation period. Unless the observer against whom reliability is being assessed is observing for exactly the same period, the periods of observation will not be coincident. Under these circumstances a double rating may not be strictly a reliability check, but more of a check of the stability of the behaviour. Patel *et al.*[5] discuss how the use of a simple checklist of the occurrence of key events can substantially improve the reliability of this type of scale.

A further technical issue for observer-based scales is that of reactivity, which is the change in behaviour due to the patient's awareness of the presence of an observer. *Any* live observer will induce some reactivity effects, and this issue is often ignored. Reactivity factors may be reduced by making the observation procedure as minimally intrusive as possible by, for example, careful siting of the observer.

Practical factors and an example

Because of their apparent simplicity, the limitations of questionnaires are not always considered. Bowling[6] has reviewed the main sources of variation in data quality between four modes of administering questionnaires, comparing face-to-face interviews, telephone interviews, self-administration, and electronic procedures. The quality of data was defined in terms of overall response rates, item response rates, response accuracy, and social desirability response bias. These different ways of administering questionnaires have differing cognitive demands on respondents,

so that face-to-face interviews tend to yield higher quality data. Regier *et al.*[7] point out the need to standardize measures for both clinical and epidemiological work, given the 'drift' or 'mutations' that can occur with repeated use of even the most carefully designed measures, with major consequences, for example, for public health and policy if prevalence estimates cannot be made reliably.

Practically, scales should be written at the level of vocabulary simple enough for the lowest level of educational attainment likely to be found among scale users. Usually one type of item format is used throughout the scale: sometimes the wish to have one format means that some items are then not easy to understand because that one format is not suited to the content of every item. Shorter scales are usually preferred to longer ones. Some scales are in the public domain. Others are copyright and payments should then be made each time a copy of the scale is used.

There are several steps involved in creating a questionnaire or rating scale. The creation of the original item pool is the first step in designing a new measure, and often existing measures provide some of the items, as well as those identified from any other surveys or studies. Individual items must be selected from this pool, implying, making judgements about the most important topics to cover, and about the bandwidth of the possible range of response options from 'normal' to the most extreme likely to be found.

McGuire *et al.*[8] give a very clear description of the construction of a new rating scale to assess the quality of the clinician–patient therapeutic relationship in community care. They describe the four stages in the creation of the scale:

- generating an item pool
  - conducting semi-structured interviews with clinicians and patients
  - reviewing the content of existing scales covering the same phenomena
- identifying factors and items in the new scale
  - original pool of items rated by clinicians and patients
  - rating scores subjected to principal components factor analysis
- conducting a test–retest reliability study of the new scale
- testing the factorial structure of the revised new scale in a new sample of clinicians and patients

Multiple measures

In both clinical practice and research, the concurrent use of several different measures may be helpful, addressing different categories of functioning and behaviour, with care taken to consider the overall assessment load on any one staff member in the light of their other clinical commitments. Rutter,[9] in reviewing changes in child psychiatry, pays particular attention to the importance of sound measurement by contrasting standardized interviews and checklists, and points out that multiple measures involving different informants, which are repeated over time, are necessary to reduce error and minimize rater bias. Self-completed questionnaires may supplement observer-completed ratings or checklists in the assessment of specific behavioural problems. Deale *et al.*[10] in evaluating the outcome of a treatment trial for chronic fatigue, used 10 outcome measures, namely: three functional impairment measures; two fatigue measures; two psychological distress questionnaires or inventories; and three other variables, including a global self-rating and a self-written statement of illness attributions.

Behavioural and observational assessment methods

A number of existing simple observational methods were refined alongside the clinical introduction of behaviour therapy procedures in the 1960s, and these continue to be associated with contemporary cognitive behavioural interventions (see Fig. 1.8.4.2).

These methods focus on current overt behaviour, and are used for the immediate recording of events, for example counting the number of times an event occurs (leading to frequency counts), or coding observed event by using a set of prescribed behavioural categories that exhaustively and mutually exclusively cover all anticipated possibilities. Direct observation methods can be used in a standardized manner for a group of patients. But they also lend themselves to flexible modification, so that the frequency and the duration of observation periods, and coding systems, can be chosen to suit the requirements of the behavioural difficulties of an individual.

Functional analysis

The term 'functional analysis' generally refers to attempts to discern the variables controlling or maintaining a phenomenon. It usually describes the observation of an individual's behaviour of clinical significance, linked to the observation of those events in the immediate environment that directly preceded, were concurrently associated with, and followed, the target behaviour. This sequence of **antecedent** environmental events, target **behaviour** and concurrent events, and **consequent** environmental events, is often called an **ABC** analysis. For example, incidents of aggression by a particular client in a day-setting may be a function of who is near them or talking to them, so that a record of their presence or absence would be important.

Sampling

A key issue with both rating and direct observational methods is the spread of observation periods over the waking day. Since many behaviours vary in their natural frequency during the waking day, many events should, theoretically, be observed continuously throughout the day. However, since the time needed for continual observation is usually unavailable, a representative sample of the whole day should be observed. Ideally, a random sample of time

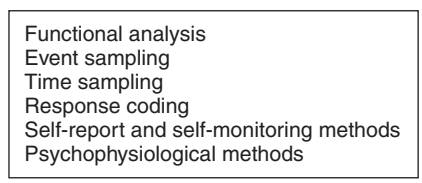


Fig. 1.8.4.2. Categories of behavioural and observational assessment methods.

periods throughout the day should be observed. When events of clinical interest are very complex, it may take too long to code each event fully, so then only, say, every fifth or tenth event is coded in detail—this is termed as **event sampling**. When events are happening very rapidly, or if they tend to happen at about the same rate during the day, it is time-wasting to observe all the time, so observations may then be made only every 15 or 30 min—this is termed as **time sampling**.

Response coding

A set of qualitative coding categories should cover all the most likely events, using clear and unambiguous language, coding categories should be simple enough to be entered quickly. When continuous observation is used, especially for high rates of behaviour, the observer must be allowed regular rest periods.

Psychophysiological methods

These methods typically involve the use of surface sensors to measure changes in, for example, skin electrical conductivity (as in measuring changes in sweating), in light transmission (as in measuring changes in finger blood flow), and in volume (as in plethysmography). These sensors are now always electronic, so the electrical signals from them are calibrated with the associated physical change, which is taken as a proxy measure for the assumed underlying change in physiological arousal. The small size, low weight, and the sophistication and reliability of electronic recording equipment, including hand-held electronic event recorders, now allows the immediate and unobtrusive recording of events, with concurrent data analysis.

Standard or individualized measures?

There are many standard ratings and questionnaires already in existence covering most areas of clinical interest. However, the fact that a scale is well used does not necessarily mean that it is psychometrically sound, or that the content covered is appropriate for a given clinical or research purpose. Rating scales may continue to be used widely when they are no longer the best scales technically, but because the volume of published research in which they have been used permits comparisons with other studies.

Solutions to this dilemma are either to create a totally new measure, making sure it *is* better than its predecessor, or to improve systematically the properties of an existing measure. Parker *et al.*[11] describe a modification of the well-established parental bonding instrument to include abusive parenting, which was omitted in the original version. Their article demonstrates both how to modify an original measure to improve item wording, and at the same time incorporate additional material to increase the value of the measure. An unsound solution is to change the measure in an *ad hoc* way—scale vandalization!—with no awareness of the principles of sound scale and item construction. This only results in an instrument that will then be of poor or unknown reliability or validity.

Conclusions

The value of rating and questionnaire methods is that they can be used by a variety of assessors who do not need to be qualified mental health professionals—although the need for at least some training in rating methods should not be overlooked. They can be presented in very short versions therefore making minimal demands on both assessors and patients, and so can lead to high levels of compliance. Since they can be used by patients, they can themselves be tools to increase engagement and to give patients direct feedback about their own current state. Behavioural and observational assessment measures constitute a clinically useful subgroup of methods, which potentially have high validity with respect to day-to-day functioning. However, the apparent simplicity of all of these methods masks the need for care in their construction, the importance of training in their use, and caution in over-sophisticated interpretation of data arising from their use.

Further information

There are a number of helpful core texts, describing general technical issues in questionnaire and rating scale design (see Streiner & Norman[a] and McDowell[b]). Similarly Haynes & O'Brien[c] and Hersen[d] cover the general principles of behavioural assessment. Andrews *et al.*[e] give an excellent review of outcome measures in mental health, covering many of the most commonly used rating scales. McDowell[b] and Bowling[f] include extensive lists of the most commonly used scales in health and social care, including those covering mental health issues.

(a) Streiner, D.L. and Norman, G.R. (2003). *Health measurement scales*. Oxford University Press, Oxford.

(b) McDowell, I. (2006). *Measuring health: a guide to rating scales and questionnaires*. Oxford University Press, Oxford.

(c) Haynes, S.N. & O'Brien, W.H. (2000). *Principles and practice of behavioural assessment*. Springer, New York.

(d) Hersen, M. (2005). *Clinician's handbook of adult behavioral assessment*. Academic Press, New York.

(e) Andrews, G., Peters, L., and Teesson, M. (1994). *The measurement of consumer outcome in mental health: a report to the National Mental Health Information Strategy Committee*. Clinical Research Unit for Anxiety Disorder, Sydney, Australia.

(f) Bowling, A. (2004). *Measuring health*. Open University Press, Maidenhead.

References

1. Kempf, E.J. (1915). The behavior chart in mental diseases. *American Journal of Insanity*, **71**, 761–72.
2. Aman, M.G., Singh, N.N., and Stewart, A.W. (1985). The aberrant behavior checklist: a behavior rating scale for the assessment of treatment effects. *American Journal of Mental Deficiency*, **89**, 485–91.
3. See www.iop.kcl.ac.uk/hsr/prism/can
4. Eifert, G.H. and Wilson, P.H. (1991). The triple response approach to assessment: a conceptual and methodological reappraisal. *Behaviour Research and Therapy*, **29**, 283–92.
5. Patel, V., Hope, T., Hall, J.N., *et al.* (1995). Three methodological issues in the development of observer-rated behaviour rating scales: experience from the development of the rating scale for aggressive behaviour in the elderly. *International Journal of Methods in Psychiatric Research*, **5**, 21–7.
6. Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health*, **27**, 281–91.

7. Regier, D.A., Kaelber, C.T., Rae, D.S., *et al.* (1998). Limitations of diagnostic criteria and assessment instruments for mental disorders. *Archives of General Psychiatry*, **55**, 109–15.
8. McGuire-Snieckus, R., McCabe, R., Catty, R., *et al.* (2007). A new scale to assess the therapeutic relationship in community mental health care: STAR. *Psychological Medicine*, **37**, 85–95.
9. Rutter, M. (1997). Child psychiatric disorder: measures, causal mechanisms, and interventions. *Archives of General Psychiatry*, **54**, 785–9.
10. Deale, A., Chalder, T., Marks, I., *et al.* (1997). Cognitive behaviour therapy for chronic fatigue syndrome: a randomized controlled trial. *American Journal of Psychiatry*, **154**, 408–14.
11. Parker, G., Roussos, J., Hadzi-Pavlovic, D., *et al.* (1997). The development of a refined measure of dysfunctional parenting and assessment of its relevance in patients with affective disorders. *Psychological Medicine,* **27**, 1193–203.

1.9

Diagnosis and classification

Michael B. First, and Harold Alan Pincus

In Psychiatry, as in all of medicine, diagnosis is a key function and central to developing a plan of treatment for patients. Psychiatry, however, faces special challenges. The etiopathogenesis of most psychiatric disorders is not known. For the most part, a clinician must rely on reports from, and direct observation of patients to gather the necessary information to determine a diagnosis. Until very recently, laboratory tests had little relevance. Even diagnostic information found in medical records may not be useful, since the clinician cannot ascertain whether the historically recorded diagnoses of previous clinicians were based on reliable observations, the application of similar diagnostic approaches, or even the same system of classification. These special challenges faced by the field have ensured that diagnosis and classification in psychiatry has a long and rich history.

Definitions

The term 'diagnosis' can mean both the name of a particular disease as well as the process of determining or 'making' a diagnosis. In medicine, generally, various terms are used to describe a pathological entity. When there is the presence of objective pathology or the presumed understanding of aetiology, the term 'disease' is generally used, e.g., pancreatic cancer, strep throat, Alzheimer's disease. In instances of unknown aetiology or when the disease process is not apparent, the term 'disorder' is usually applied with a syndromic characterization, i.e., definition based on symptom presentation, history, and sometimes, associated laboratory findings. Other terms are also used in common parlance, such as 'illness' for an individual's subjective awareness of distress and 'sickness' for the inability to perform usual social roles. For the most part, in psychiatry, the term 'disorder' is used.

Classification represents the process of placing diagnostic entities into various groupings in a systematic way, based on a set of principles with regard to the similarities and differences among these categories. Depending upon the principles and conceptual framework underlying the categorization process, classifications can be very different.

Goals of the classification

In some ways, the most important question may be 'whose needs is the classification primarily intended to address?' Clinicians want a classification that can categorize as many people that come in for help as possible. They want the classification to facilitate the identification and treatment of patients and provide guidance on prognosis and cause. Researchers want to have groupings that are highly homogeneous in order to test the efficacy of specific treatments and to better understand the aetiology of specific disorders. Educators want a classification system to offer a structure for teaching about psychopathology and differential diagnosis. Public health administrators want to track epidemiology, health utilization, and costs over time. Some argue that psychiatric diagnosis is a reductionistic labelling of individual differences or social deviance and exposes individuals to potential stigma. At a minimum, they would like a psychiatric diagnostic system to be less prone to misuse. Ultimately, most classifications attempt to balance among those competing priorities, not always successfully. In some cases, e.g., the ICD-10, different products are developed for different target groups, i.e., research diagnostic criteria for investigators, a simpler, more aggregated classification for primary care providers, etc.

Conceptual issues

A range of conceptual issues and their resolution determines the principles and rules governing a system of classification. It is important to note, however, that classification systems do not necessarily apply these rules in a consistent manner. Some of the issues noted below may not be resolved in an absolute manner, but in a way that employs compromises among multiple priorities, e.g. balancing the needs of clinicians, researchers, educators, and public health administrators or having some diagnostic groupings based on a descriptive approach and others on a theory-based approach.

Descriptive v. theory-based: Do the classification principles emanate from a theory regarding the aetiology or mechanisms of psychopathology (e.g. psychodynamics, behavioural, neurobiological) or does the classification attempt to provide a theoretical heuristic framework for describing syndromic entities?

Pathology v. normalcy: What assumptions underlie distinguishing what constitutes a 'mental disorder' or 'caseness' from normative behavior?

Categorical v. dimensional: Does the classification assume discrete categories with sharp boundaries or does it assume that psychopathology lies on a continuum across a range of dimensions (and if so, what dimensions and how were they chosen?)

Lumping v. splitting: Does the classifications system establish a smaller number of broad, relatively heterogeneous categories or numerous homogeneous categories?

Multiple v. single diagnosis: Is there a hierarchy where certain diagnoses have priority and 'trump' other diagnoses if an individual fits into more than one category or are multiple simultaneous diagnoses (i.e. comorbidity) encouraged to communicate more complete diagnostic information?

The development of modern classifications

The first international classification of diseases in 1855 was concerned with a nomenclature of causes of death.[1] After many revisions this list was adopted by the World Health Organization (WHO) in 1948 and the so-called *Sixth Revision of the International Statistical Classification of Diseases, Injuries and Causes of Death* (ICD-6) was produced.[2] The sixth edition of the ICD included for the first time a classification of mental disorders, containing 10 categories of psychoses, nine categories of psychoneurosis, and seven categories of character, behaviour, and intelligence. A number of problems with this classification (e.g., many important categories such as the dementias, many personality disorders, and adjustment disorders were not included) rendered it unsatisfactory for use in most countries; only five countries, including the United Kingdom, adopted it officially. The first edition of the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM-I)[3] was published in the United States as an alternative to ICD-6. For the first time in an official classification, glossary definitions of the various disorders were included in addition to the names of the disorders.

Work on ICD-8 began in 1959 with the goal of developing a classification system that would be acceptable to all of its member nations. The resulting system, ICD-8, went into effect in 1968, and in 1974 added a glossary which was largely based on British views about diagnostic concepts.[4] Coincident with the development of ICD-8, the American Psychiatric Association prepared a second edition of its DSM based on ICD-8, defining each disorder for use in the United States.[5]

The early 1970s saw the introduction of explicit operationalized diagnostic criteria that were developed for research purposes. Although the glossary definitions of disorders in DSM-II and ICD-8 were an improvement over just having a list of diagnostic categories, these brief descriptions were too vague to be useful in identifying diagnostically homogeneous populations for study. Researchers responded to this need by developing their own operationalized criteria. The first set of diagnostic criteria that covered a wide range of disorders was developed by Robins and Guze at Washington University in St. Louis[6] with the stated purpose of 'provid[ing] common ground for different research groups so that diagnostic definitions can be emended constructively as further studies are completed'. (p. 57). They were known as the 'Feighner criteria' after the first author of the paper that presented them. Criteria sets for 16 disorders were presented and listed those features required for each diagnosis (known as 'inclusion criteria') as well as features whose presence would rule out the disorder (known as 'exclusion criteria'). The Feighner criteria proved to be enormously useful to the research community as illustrated by the large number of times they were cited in other papers (i.e. 1650 citations from 1972 to 1982 as compared to the typical average of 2.1 citations per paper). Several years later, an expanded set of research criteria based on the Feighner criteria was developed to meet the needs of a National Institute of Mental Health-sponsored collaborative project on the psychobiology of depression.[7] These criteria, known as the Research Diagnostic Criteria (RDC) subsequently became very popular among researchers and were heavily used, especially in research on mood and psychotic disorders.

To develop the mental disorders section for ICD-9, WHO initiated an intensive program to identify problems encountered by psychiatrists in different countries in the use of the mental disorders section of ICD-8 and to formulate recommendations for their solution. A series of eight international seminars were held annually from 1965 to 1972, each of which focussed on a recognized problem in psychiatric diagnosis. The outcome of the seminars formed the basis for the recommendations made for ICD-9,[8] which was ultimately published in 1978.

As work progressed on the development of ICD-9, the American Psychiatric Association decided to develop a third edition of its diagnostic manual, DSM-III.[9] This decision was made both because of identified inadequacies of the ICD-9 for research and clinical use, and because the ICD-9 did not include important innovations that had already been demonstrated by researchers to be both technically feasible and useful, like operationalized diagnostic criteria. Under the leadership of Robert L. Spitzer, successive drafts of DSM-III were prepared by 14 advisory committees, with the drafts being distributed among both American and international psychiatrists for comments and review. Many of the DSM-III criteria sets were based on the RDC criteria, with the rest developed based on expert consensus.

The improvement in reliability over DSM-II (which provided only glossary definitions) was demonstrated by a large NIMH-supported field trial in which clinicians were asked to independently evaluate patients using drafts of the DSM-III criteria.[10] The explicit diagnostic criteria provided for each of the disorders in the classification were based on the symptomatic presentation of the disorder rather than on theories about the underlying cause. Even though the DSM-III was a product of the American Psychiatric Association, its adoption of this 'descriptive approach' resulted in its widespread acceptance by all mental health professionals in the United States, regardless of their theoretical orientation. For example, clinicians from different orientations might have very different understandings of what causes panic attacks; a cognitively-oriented clinician might attribute a panic attack to the person's tendency to catastrophize in response to normal physical sensations like increased heart rate; a neurobiologically-oriented clinician might consider panic attacks to be due to overactivity of brain circuitry involved in fight-or-flight responses, and a psychodynamically-oriented clinician might see panic as a consequence of the breakdown of the defense organization at various levels. Despite these divergent hypotheses, however, all of these clinicians can agree on how a panic attack presents (i.e. a discrete period of apprehension or fear with at least four symptoms such as shortness of breath, palpitations, chest pain, choking, dizziness, etc), thus facilitating communication among them.

DSM-III also introduced the use of a multiaxial system for recording the diagnostic evaluation. The multiaxial system facilitated the use of a biopsychosocial model of evaluation by separating (and thereby calling attention to) developmental and personality disorders (Axis II), physical conditions (Axis III), stressors (Axis IV),

and degree of adaptive functioning (Axis V) from the usually more florid presenting diagnoses (Axis I).

Despite initial opposition among some psychiatrists (most especially those with a psychoanalytic orientation), DSM-III proved to be a great success, becoming the common language of mental health clinicians and researchers for communicating about mental disorders. Although it was intended primarily for use in the United States, it was translated into 13 languages and was widely used by the international research community.

Experience with DSM-III in the few years after its publication in 1980, revealed a number of inconsistencies and lack of clarity in the diagnostic criteria sets. Furthermore, research conducted in the early 1980's demonstrated errors in some of the assumptions that went into the construction of the DSM-III criteria sets. For example, the DSM-III prohibition against giving an additional diagnosis of Panic Disorder to individuals with both Major Depressive Disorder and panic attacks was shown to be incorrect based on data demonstrating that relatives of individuals with both Major Depressive Disorder and Panic Attacks can have either Major Depressive or Panic Disorder.[11] For these reasons, work began on a revision of the DSM-III, which was published as DSM-III-R in 1987.[12]

Initial work began on the development of the psychiatric section of the *10th Revision of the International Classification of Diseases* (ICD-10), in 1982 under the chairmanship of Norman Sartorius. After a meeting of WHO representatives and consultants together with representatives of the American Drug and Mental Health Administration in Copenhagen in 1982, several further meetings took place (e.g. in Djakarta and in Geneva in 1984) in which a provisional psychiatric classification was designed. It was decided that the ICD-10 classification of mental disorders would be produced in several versions. The first of these is to be used, as are other parts of the International Classification of Diseases, mainly for statistical purposes, and included a short glossary definition for each category.[13] This is the version that was officially approved by the World Health Assembly and thus, is the version for which international compatibility is mandated by treaty agreements. The second version, Clinical Descriptions and Diagnostic Guidelines is for the use of the practicing clinician.[14] Each category in this version has a detailed definition specifying the main features of the disorder followed by diagnostic guidelines. The third version, the Diagnostic Criteria for Research is primarily intended for research and contains diagnostic criteria which are stricter in form than those in the clinical diagnostic guidelines from which they were derived.[15] For example, while the guidelines may indicate that a particular disorder 'usually starts in early childhood', the diagnostic criteria for research would specify that the diagnosis 'should not be made if the onset is after the age of 30', The decision to separate the criteria for research from the clinical guidelines was made because clinicians in their daily work do not observe overly strict rules when making diagnoses, which are of cardinal importance for research.[16] Finally, a version of the mental disorders section was produced for use in primary care settings.[17] It contains a much smaller number of categories (i.e. those that are frequently encountered in every day general practice) as well as treatment guidelines corresponding to these categories.

By 1986, a first draft of the psychiatric chapter, including details of the categories, code numbers, diagnostic guidelines, and precise diagnostic criteria for research had been written, and by June 1987, the clinical diagnostic guidelines were being circulated by WHO's division of mental health for field trials in 194 different centers in 55 different countries.[18] In 1989, the International Revision Conference, attended by representatives of the health ministries of a majority of WHO member states gave formal approval to the basic categories and text. A draft of the diagnostic criteria for research was produced in 1990 and field trials to evaluate inter-rater agreement, confidence in diagnosis, and ease of use began later in the year.[19] Finally, in 1990, the World Health Assembly formally approved its introduction in member states starting in January 1, 1993.

The American Psychiatric Association started work on the development of DSM-IV in 1988, shortly after the publication of DSM-III-R, spurred on by the need to coordinate its development with the already ongoing development of ICD-10. DSM-IV continued the descriptive atheoretical approach advanced by both DSM-III and DSM-III-R, but this time also incorporated a meta-analytic data-based revision process to guide changes.[20,21] This was in contrast to both DSM-III and DSM-III-R which by necessity, given the paucity of available empirical data, relied almost exclusively on expert consensus. The DSM-IV workgroups began their deliberations by identifying a series of diagnostic questions to be considered and problems to be addressed and employed a three-stage empirical review process to address, these questions. The first stage involved a systematic comprehensive review of the published literature guided by literature searches using rules established at a DSM-IV methods conference. The second stage involved supplementing the literature reviews with a data reanalysis project funded by the MacArthur foundation in which existing data sets collected for other studies were combined and analyzed using meta-analytic methods. These data reanalyses were useful in answering a number of diagnostic questions (e.g. determining the minimum number of panic attacks required in order to justify a diagnosis of panic disorder) but were unfortunately limited by incompatibilities in the data sets and the fact that the data needed to answer specific diagnostic questions often had not been collected. Proposed criteria sets formulated based on the literature reviews and data reanalyses were then tested in 15 NIMH-funded multi-site field trials. The entire empirical review process and the reasons underlying the decisions made by the DSM-IV workgroups have been documented in the four volume DSM-IV Sourcebook.[22–25]

In order to increase compatibility between ICD-10 and DSM-IV, a collaborative relationship was established between the DSM-IV workgroups and the developers of ICD-10. Two meetings were convened in which the respective workgroups joined forces with the goal of minimizing the differences between diagnostic definitions in the two systems. Unfortunately, the potential to make the two systems identical was seriously constrained by differences in the timelines between the two revision processes. By the time the DSM-IV workgroups were first convened in 1989, the categories and basic text of the ICD-10 had already been settled by the International Revision Conference.[26] Thus, although final DSM-IV and ICD-10 systems were much more similar than were DSM-III and ICD-9, a number of mostly small differences in criteria sets persist. While some of the discrepancies are the result of genuine differences in diagnostic outlook (e.g., the one month duration of ICD-10 schizophrenia vs. 6 month duration of DSM-IV schizophrenia), the overwhelming majority appear not to have any justification.[27]

One of the most important uses of the DSM-IV has been as an educational tool. This is especially true of the descriptive text that accompanies the criteria sets for the DSM-IV disorders. Given that the interval between DSM-IV and DSM-V was being extended from seven years between DSM-III and DSM-III-R, and between DSM-III-R and DSM-IV to at least 12 years, concerns were raised that the information in the text would become increasingly out-of-date over time. Therefore, in order to bridge the span between DSM-IV and DSM-V, a revision of the DSM-IV text was undertaken.[28] The primary goal of the DSM-IV-TR was to maintain the currency of the DSM-IV text, which reflected the empirical literature up to 1992. Thus, most of the major changes in DSM-IV-TR were confined to the descriptive text. Changes were made to a handful of criteria sets in order to correct errors identified in DSM-IV. In addition, some of the diagnostic codes were changed to reflect updates to the ICD-9-CM coding system adopted by the U.S. Government.

Differences between DSM-IV and ICD-10

A fundamental difference between the ICD-10 and the DSM-IV reflects the different purposes of the two systems, i.e., that ICD-10 is set up as a classification system whereas DSM-IV is a diagnostic nomenclature. The primary goal of the ICD is to facilitate the collection of statistics about those individuals who present themselves to a health care professional. Thus, the ICD has been designed to provide the coder with an unambiguous choice of diagnostic category given a particular case. The main rule for deciding whether to include a diagnostic category in the ICD is its common international usage. Inclusion of a category in the ICD carries with it no implication of diagnostic validity—in fact, a number of categories included in ICD-10 were considered for inclusion in DSM-IV—but were not added because of concerns about their validity (e.g., mixed anxiety depression). In contrast, inclusion of a category in the DSM implies that the category has been officially sanctioned by the American Psychiatric Association as appropriate for clinical and research usage, i.e., the category has both clinical utility and is backed up by an empirical data base. It should be noted, however, that the empirical data base is not equivalent for all of the categories—to minimize disruption, diagnostic categories, that were included in earlier editions of the DSM have been 'grandfathered' in. Starting with DSM-IV, new categories were only added if they met these higher standards.

Another important difference between the DSM and ICD approach is the role of impairment in the definition of a disorder. With only a few exceptions (e.g., dementia, phobias), mental disorders in ICD-10 are defined exclusively by the symptomatic presentations—there is no requirement that the symptoms cause any impairment in the individual's level of functioning. Impairment in functioning caused by the symptoms is indicated in ICD-10 by using an orthogonal scale, the International Classification of Functioning.[29] In contrast, most of the DSM-IV criteria sets include a criterion (known as the 'clinical significance criterion') requiring that the disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning. According to the introduction of the DSM-IV, this criterion has been included to help establish 'the threshold for the diagnosis of a disorder in those situations in which the symptomatic presentation by itself (particularly in its milder forms) is not inherently pathological and may be encountered in individuals for whom a diagnosis of 'mental disorder' would be inappropriate'. (p. 8). Accordingly, the only diagnoses that do not include this criterion are those whose symptomatic presentations are considered to be inherently indicative of psychopathology (e.g., the psychotic disorders).

The diagnostic implications of this difference can be illustrated in the different ways that specific phobia is defined in DSM-IV and ICD-10. In DSM-IV, a phobia is diagnosed only if 'the avoidance, anxious anticipation, or distress in the feared situation interferes significantly with the person's normal routine, occupational (or academic) functioning or social activities or relationships, or there is marked distress about having the phobia'. (p. 449, DSM-IV-TR). ICD-10 has no such requirement; the phobia is diagnosed so long as there is a marked fear or avoidance of a specific object or situation. Thus, an individual residing in New York City who has a snake phobia but who never has any occasion to encounter a snake would not be diagnosed as having a mental disorder in DSM-IV-TR because the phobia does not have any impact on the person's functioning whereas in ICD-10 such an individual would be diagnosed as having a snake phobia because the person would react with fear or avoidance if he or she had the occasion to be confronted with a snake.

Separating symptoms from the functional impairment that results from them certainly makes conceptual sense. In other areas of medicine, the diagnosis of a disorder is based solely on the presence of pathology and not on the effect that the pathology exerts on a person's life (e.g., a patient is diagnosed with pneumococcal pneumonia if the patient's lungs are infected with the pneumococcus bacillus regardless of the impact of the pneumonia on the patient's level of functioning). The problem with diagnosing mental disorders in this way is that it is not currently possible to define the presence of a mental disorder based on the identification of its underlying pathology. The descriptive symptoms that make up the definitions of mental disorders in the DSM-IV and ICD-10 are not specific to mental disorders but can and do occur in individuals without any mental disorder. Thus, defining disorders exclusively in terms of presenting symptomatology, much of which can occur in normal individuals, can lead to false positive diagnoses. For this reason, in order to avoid false positive diagnoses in the absence of objective evidence of disease, DSM relies on functional impairment or distress to help set the diagnostic threshold between normality and disorder.

The structure of ICD-10, (Chapter V)

The psychiatric classification is part of the general medical classification. There are 21 chapters, each designated by a Roman numeral. The psychiatric disorders are included in Chapter V which is also identified by the letter F. The letter F is followed by Arabic numbers, the so-called second digit for the larger diagnostic groups and the third digit for more special groups. Thus the use of three digits allows a choice of 100 diagnostic possibilities. Proceeding further with a fourth digit, 1000 possible diagnoses are available, of which about one-third are used at present. This system is thus designed to allow the addition of new diagnoses in future without having to change substantial parts of the classification.

Furthermore, it is possible to code the course over time or characteristic features of a disorder by using a fifth or sixth digit. By using codes from other chapters of ICD-10, such as X, Y, and Z, additional

circumstances (e.g. suicide) or special symptoms (e.g. nausea) as well as psychosocial factors can be coded. Somatic comorbidity is coded from the related chapters, for example diseases of ear, nose, and throat from Chapter VIII headed by the letter H (e.g. tinnitus H93.1) or diseases of the gastrointestinal system from Chapter XI headed by the letter K (e.g. alcohol gastritis K29.2). The specific challenges encountered in diagnosing psychiatric disorders reliably over the years has led WHO to include short definitions plus inclusion and exclusion terms for all psychiatric disorders in Chapter V (F). In all other chapters, diagnoses are named without further explanation.

As described earlier in this chapter, Chapter V of ICD-10 is not just a catalogue of disorders for statistical purposes, but is also a clinical manual, a textbook of diagnoses, and an instrument for research for different users. Therefore, a group of texts had to be produced to serve the various purposes—the so-called 'ICD-10 family of documents'

The *Short Glossary of ICD-10, Chapter V (F)* is part of the basic work known as the *International Statistical Classification of Diseases and Related Health Problems* The Short Glossary is part of the first of three volumes, the general systematic classification, and gives short definitions which are useful not only for medical personnel but also for statisticians, health insurance clerks, and others who are not in medical or related professions.

The *Clinical Descriptions and Diagnostic Guidelines (CDDG Version)*, known as the Blue Book because of the colour of its cover, was developed first and can be regarded as the central part of the psychiatric classification[14] intended for use by psychiatric clinicians in their daily practice. The *Diagnostic Criteria for Research* (DCR), known as the Green Book, has been developed for scientific use[15] and is intended to be used together with the diagnostic guidelines. Compared with the Blue Book, the symptom criteria are more clearly defined, the time criteria are stricter, and the inclusion and exclusion criteria are more precise in the Green Book. Thus, many unclear cases which are unsuitable for research are excluded. However, despite its title, this book is also useful for diagnosticians in clinical practice.

The *multiaxial version* of the ICD-10 classification of mental disorders allows different aspects of the patient's health and social situation to be assessed. Introduced by Rutter and colleagues,[30] multiaxial diagnosis has been employed for many years in child and adolescent psychiatry. It contains clinical syndromes, problems of development, intelligence, somatic disorders, and psychosocial problems. To a considerable degree, the multiaxial version of ICD-10 is comparable with that of DSM-IV. However, in DSM-IV, axis I is for psychiatric clinical disorders, axis II is for personality disorders and intellectual disability, and axis III is for general medical conditions. In ICD-10, axis I includes all disorders. Thus, psychiatric disorders (F1–F5), personality disorders (F6) and intellectual disability (F7), and the chapters on somatic comorbidity all use one axis.

Axis II of ICD-10 is for disability. To facilitate its use, WHO developed an instrument, the short disability assessment schedule (WHO DAS-S), which helps to describe and assess the consequences of axis I disorders.[31] Axis II corresponds to the widely used DSM-IV axis V, Global Assessment of Functioning (GAF). In connection with the disability axis, the International Classification of Functioning created by WHO for the whole of rehabilitative medicine, of which psychiatry is only a part, should be mentioned.[29] Axis III of ICD-10 covers psychosocial and other problems, and corresponds to DSM-IV axis IV (psychosocial and environmental problems).

The *primary health care* (PHC) version of the ICD-10 classification of mental disorders was developed because of the great importance of psychiatric disorders in general practice, for example the high prevalence of depressions, anxiety disorders, and dependence on alcohol and psychotropic drugs.[17] There are 24 syndromes, including dementia, delirium, depression, etc. Each disorder is understood in a rather broad sense, and not subdivided, and the descriptions are simpler than those in the main classification. A flipcard containing symptoms, diagnostic criteria, differential diagnoses, and counselling and treatment of the patient and the family is provided for every syndrome.

At first glance, the structure of ICD-10, Chapter V (F), follows that of ICD-8 or ICD-9 (See Appendix 1). The classification begins with the 'organic disorders', followed by disorders due to the abuse of psychoactive substances. The next section of the classification contains schizophrenia and other psychotic disorders. This is followed by affective disorders and then neurotic and personality disorders. The chapter ends with intellectual disability and disorders of childhood and adolescence. Closer examination of the classification reveals that the traditional dualistic principle—psychoses on the one hand (in ICD-9: codes 290–299) and neuroses on the other (in ICD-9: codes 300–310)—has been abandoned. The diagnostic terms now used take a more phenomenological descriptive approach. According to the authors of ICD-10, the same psychiatric disorder may show both psychotic and non-psychotic symptoms. 'Psychotic' is defined as the manifestation of productive symptoms. The term 'neurosis' did not appear in the first drafts of ICD-10 because it is used in different and contradictory ways and is supposedly based on theories of intrapsychic processes which many of the WHO experts regarded as not generally accepted. However, after protests and objections by many clinicians worldwide, it was concluded that 'psychotic' and 'neurotic' should be used, although only as descriptive terms and not as diagnostic rubrics. Thus the term 'neurotic disorders' follows the traditional use of the word but does not imply an etiological theory.

(a) Organic, including symptomatic, mental disorders

Disorders of organic aetiology are grouped in this subchapter, independent of whether they contain psychotic or non-psychotic symptoms. However, the use of the term 'organic' does not imply that conditions elsewhere in the classification are non-organic in the sense of having no cerebral substrate.

(b) Mental and behavioural disorders due to psychoactive substance use

An improvement over ICD-9 is the compilation of all mental and behavioural disorders due to psychoactive substances within a single subchapter. The third digit indicates which substance or class of substances (e.g. F10 Alcohol) is responsible for the disorder, which is coded as a fourth digit (e.g. F10.3 Alcohol withdrawal state) or a fifth digit (e.g. F10.31 Alcohol withdrawal state with convulsions). It is possible to differentiate acute intoxication, harmful use, dependence syndrome, withdrawal state with or without delirium, different psychotic disorders, amnesic syndrome, and a number of other disorders. Thus, the psychopathological syndrome can be described and related to the dominant substance class.

(c) Schizophrenia, schizotypal, and delusional disorders

This subchapter covers schizophrenia, acute psychotic disorders, schizoaffective disorders, delusional disorders, and schizotypal

disorders. Before schizophrenia can be diagnosed the symptoms have to be observed for at least one month, unlike DSM-IV where symptoms should be observed for six months before using this diagnosis. Special care is taken with the description of short-lasting psychoses, since acute and transient psychotic disorders are of particular interest to psychiatrists from developing countries where short-lasting acute psychoses with a good prognosis are observed quite frequently.

(d) Mood (affective) disorders

All mood disorders are combined in this subchapter, which represents a considerable change compared with ICD-9. The disorders previously known as endogenous and neurotic depressions are coded in this subchapter; the differentiation between these categories has been abandoned. The ICD-9 category of neurotic depression (300.4) is no longer found in ICD-10; most of these cases are now coded as dysthymia (F34.1). Single manic episodes are coded as F30, while recurrent manic episodes are now coded as bipolar affective disorder (F31), regardless of whether or not there has been a previous depressive episode.

(e) Neurotic, stress-related, and somatoform disorders

The disorders in this subchapter are divided into a large number of categories. For instance, dissociative disorders are divided into seven subcategories, some of which represent rather rare disorders. The term hysteria is no longer used. In this subchapter, reactions to severe stress and adjustment disorders are enumerated according to time criteria and severity. Here, aetiology is generally accepted to mean exceptional mental stress or special life events. A new group of disorders in this classification are the somatoform disorders, which are of particular importance in developing countries. The traditional term neurasthenia is still maintained for a special category, in contrast with DSM-IV.

(f) Behavioural syndromes associated with physiological disturbances and physical factors

This subchapter brings together eating disorders, non-organic sleep disorders, sexual dysfunction, mental and behavioural disorders associated with the puerperium, and abuse of non-dependence-producing substances. In ICD-9, all sexual disorders were contained in one subchapter. In ICD-10, only disorders of sexual dysfunction are in F5; disorders of gender identity and sexual preference have been assigned to two different sections in subchapter F6 on personality disorders. The special code F54, psychological and behavioural factors associated with disorders or diseases classified elsewhere, allows classification of psychosomatic disorders by coding an additional somatic diagnosis.

(g) Disorders of adult personality and behaviour

Specific personality disorders are coded in this subchapter. Cyclothymic personality is not included, but an equivalent appears in F3 as cyclothymia. Also, schizotypal disorders could have been assigned to this subchapter but appear instead in F2 (as F21). The emotionally unstable personality disorder is found in this subchapter, where it is subdivided into an impulsive type (F60.30) and a borderline type (F60.31). A new entity is the factitious disorder, i.e. the intentional production or feigning of symptoms or disabilities, either physical or psychological (F68.1). If desired, narcissistic personality disorder and passive–aggressive personality disorder may be coded by using the criteria in Annex 1 of the Diagnostic Criteria for Research.

An important aspect of this subchapter is the inclusion of enduring personality changes after catastrophic experience (F62.0) or after psychiatric illness (F62.1). Personality changes after surviving a concentration camp or torture are coded under the first of these.

(h) Remaining subchapters

F7 intellectual disability, F8 Disorders of psychological development, and F9 Behavioural and emotional disorders with onset during childhood and adolescence are

The structure of DSM-IV-TR

The 'DSM-IV-TR Classification of Mental Disorders' refers to the comprehensive listing of the official diagnostic codes, categories, subtypes, and specifiers (see Appendix 2). It is divided into various 'diagnostic classes' which group disorders together based on common presenting symptoms (e.g., mood disorders, anxiety disorders), typical age-at-onset (e.g., disorders usually first diagnosed in infancy, childhood, and adolescence), and aetiology (e.g., substance-related disorders, mental disorders due to a general medical condition).

Disorders usually first diagnosed in infancy, childhood, or adolescence

The DSM-IV-TR classification begins with disorders usually first diagnosed in infancy, childhood, or adolescence. The inclusion of a separate 'childhood disorders' section in DSM-IV-TR is only for convenience—some of these conditions are sometimes diagnosed for the first time in adulthood (e.g., attention-deficit/hyperactivity disorder) and many disorders included in the rest of DSM-IV-TR can start in childhood (e.g., major depressive disorder, schizophrenia). Thus, a psychiatrist doing a diagnostic assessment of a child or adolescent should not only focus on those disorders listed in this section but also consider disorders from throughout the DSM-IV-TR. Similarly, when evaluating an adult, the psychiatrist should also consider the disorders in this section since many of them persist into adulthood (e.g., stuttering, learning disorders, tic disorders).

While the first set of disorders included in this section (i.e., intellectual disability learning and motor skills disorders, and communication disorders) are not, strictly speaking, regarded as mental disorders they are included in the DSM-IV-TR to facilitate differential diagnosis. Autism and other pervasive developmental disorders are characterized by gross qualitative impairment in social relatedness, in language, and in repertoire of interests and activities and include autistic disorder, Asperger's disorder, Rett's disorder, and childhood disintegrative disorder. The Disruptive Behaviour Disorders (i.e. Attention-deficit/hyperactivity disorder, conduct disorder, and oppositional-defiant disorder) are grouped together because they are all characterized (at least in their childhood presentations) by disruptive behavior. The Feeding Disorders of Infancy and Early Childhood include the DSM-IV-TR categories of pica, rumination disorder, and feeding disorder of infancy and early childhood (also known as failure to thrive). Tic disorders, elimination disorders, and other disorders of infancy and early childhood (which include separation anxiety disorder, selective mutism, reactive attachment disorder, and stereotypic movement disorder) round out the childhood section.

Delirium, dementia, amnestic disorder, and other cognitive disorders

In DSM-III-R, delirium, dementia, amnestic disorder, and other cognitive disorders, along with substance-induced mental disorders and mental disorder due to a general medical condition, were included in a section called 'organic mental disorders', which contained all disorders that were due to either a general medical condition or substance use. In DSM-IV, the term 'organic' was completely eliminated from the classification because of the misleading implication that disorders not included in that section (e.g., schizophrenia, bipolar disorder) did not have an organic component[32]. In fact, virtually all mental disorders have both psychological and biological components, and to designate some disorders as 'organic' and the remaining disorders in the DSM-IV as 'non-organic' reflected a reductionistic mind-body dualism.

As a result of the elimination of the Organic Mental Disorder diagnostic grouping, those disorders originally included in that section had to be redistributed throughout DSM-IV into other diagnostic classes. Delirium, dementia, and amnestic disorder were thus grouped together into a major diagnostic class because of their central roles in the differential diagnosis of cognitive impairment. Although both delirium and dementia are characterized by multiple cognitive impairments, delirium is distinguished by the presence of clouding of consciousness which is manifested by an inability to appropriately maintain or shift attention. Three types of delirium are included in DSM-IV based on causative factors: delirium due to a general medical condition, substance-induced delirium, and delirium due to multiple etiologies.

Dementia is defined by clinically significant memory impairment accompanied by impairment in one or more other areas of cognitive functioning (e.g. language, executive functioning). DSM-IV-TR includes several types of dementia based on aetiology, including dementia of the Alzheimer's type, vascular dementia, a variety of dementia due to general medical and neurological conditions (e.g., HIV, Parkinson's disease), substance-induced persisting dementia, and dementia due to multiple etiologies. In contrast to dementia, amnestic disorder is characterized by memory impairment occurring in the absence of other cognitive impairments. Two types are included in DSM-IV: amnestic disorder due to a general medical condition and substance-induced persisting amnestic disorder.

Mental disorders due to a general medical condition not elsewhere classified

In DSM-IV-TR, most of the mental disorders due to a general medical condition have been distributed alongside their 'non-organic' counterparts in the classification (e.g. mood disorder due to a general medical condition and substance-induced mood disorder was included in the mood disorders section). Two specific types of mental disorders due to a general medical condition (i.e. catatonic disorder due to a general medical condition and personality change due to a general medical condition) do not fit into any of the other diagnostic classes and therefore, are included here in this diagnostic class.

Substance-related disorders

In DSM-IV, substance-related disorders include psychiatric disturbances that result from medication side effects and the consequences of toxin exposure, in addition to those that arise due to drug and alcohol abuse. Two types of substance-related disorders are included in DSM-IV: substance use disorders (dependence and abuse), which focus on the maladaptive nature of the pattern of substance use; and substance-induced disorders, which cover psychopathological processes caused by the direct effects of substances (including toxins and medications) on the central nervous system.

Schizophrenia and other psychotic disorders

Included in this grouping are those disorders in which psychosis is the primary characteristic symptom (i.e. schizophrenia, schizophreniform disorder, schizoaffective disorder, delusional disorder, shared psychotic disorder and brief psychotic disorder). It should be noted that other disorders that may have psychotic features are not included in this grouping (e.g. mood disorders with psychotic features, delirium).

Mood disorders

This diagnostic class includes disorders in which the predominant disturbance is in the individual's mood. Although the term 'mood' is generally considered to include emotions such as depression, euphoria, anger, and anxiety, DSM-IV includes in this section only disorders characterized by depressed, elevated, or irritable mood. This diagnostic class is further divided into depressive and bipolar disorders. The term 'bipolar' is misleading because the name implies the presence of both 'down' and 'up' moods. In fact, bipolar disorder is defined by the presence of one or more manic or hypomanic episodes. Thus, patients with multiple manic episodes (i.e. unipolar mania) are considered to be bipolar despite the lack of the second 'pole'.

Anxiety disorders

The common thread tying together disorders in this section is the fact that the clinical presentation of these disorders is typically characterized by significant anxiety. The rationale for this grouping has been criticized because of evidence suggesting that at least some of the disorders are likely to be etiologically distinct from the others. For example, it has been argued that obsessive-compulsive disorder is most likely part of an obsessive-compulsive spectrum that might include tic disorders, hypochondriasis, body dysmorphic disorder, and perhaps trichotillomania.[33]

Somatoform disorders

Somatoform disorders are characterized by their presentation in general medical settings by individuals who do not consider themselves to be suffering from a mental disorder. Individuals with somatoform disorders present with somatic complaints or bodily concerns that are not adequately explained by an underlying general medical condition. Conceptually, the somatoform disorders can be divided into three general types: 1) those in which the individual's focus is on the physical symptoms themselves (somatization disorder, undifferentiated somatoform disorder, pain disorder, and conversion disorder); 2) those who are preoccupied by the belief that one has a serious physical illness despite medical reassurance (hypochondriasis); and 3) those who are preoccupied by the belief that a part or parts of their body are physically defective (body dysmorphic disorder).

Factitious disorders

Individuals with a factitious disorder intentionally produce or feign a physical or psychological symptom, motivated by the psychological need to assume the sick role and be taken care of. This is in contrast to malingering (which is not considered to be a mental disorder) in which the person is motivated by secondary gain (e.g. to evade criminal responsibility, to receive disability benefits).

Dissociative disorders

Dissociation is the core element of this group of disorders, which is defined as a disruption in the usually integrated functions of consciousness, memory, identity, and perception. Four specific disorders are included (dissociative amnesia, dissociative fugue, dissociative identity disorder, and depersonalization disorder).

Sexual and gender identity disorders

This diagnostic class groups together disorders involving three relatively distinct aspects of human sexuality: sexual dysfunctions, which involve disturbances in sexual desire or functioning, paraphilias which involve unusual sexual preferences that interfere with functioning (or in the case of preferences that involve harm to others like paedophilia, merely acting on those preferences), and gender identity disorder that entail one's internal identity of maleness and femaleness (gender identity) being at odds with one's anatomical sexual characteristics.

Eating disorders

Disorders in this section involve abnormal eating behavior; either the refusal to maintain adequate body weight (anorexia nervosa) or discrete episodes of uncontrolled eating accompanied by excessive effects to counteract the effects of these binges (bulimia nervosa).

Sleep disorders

Sleep disorders are subdivided into four groups based on presumed aetiology (primary, related to another mental disorder, due to a general medical condition, and substance induced). Two types of primary sleep disorders have been included: dyssomnias (problems in regulation of amount and quality of sleep) and parasomnias (events that occur during sleep). The dyssomnias include primary insomnia, primary hypersomnia, circadian rhythm sleep disorder, narcolepsy, and breathing-related sleep disorder, whereas the parasomnias include nightmare disorder, sleep terror disorder, and sleepwalking disorder.

Impulse control disorders not elsewhere classified

Many disorders in the DSM-IV-TR are characterized by problems with impulse control (e.g. borderline personality disorder, substance dependence, attention-deficit/hyperactivity disorder). This diagnostic grouping is for those impulse-control disorders not included in other sections of the DSM-IV-TR. Included are problems controlling angry impulses (intermittent explosive disorder), problems controlling impulses to steal (kleptomania) or set fires (pyromania), problems controlling impulses to pull out one's hair (trichotillomania) and problems controlling impulses to gamble (pathological gambling).

Adjustment disorders

This diagnostic class is for presentations that do not meet criteria for specific disorders (i.e. subthreshold presentations) that represent a maladaptive response to a stressor. For example, if depression occurring after a job loss is severe enough to meet full symptomatic criteria for a major depressive episode, then major depressive disorder would be diagnosed. If the job-loss-related depression falls symptomatically short of this diagnostic threshold, then adjustment disorder with depressed mood is diagnosed.

Personality disorders

Each of us has a personality, that is our characteristic way of experiencing and processing the world and ourselves. When an individual's characteristic patterns of relating, feeling, and thinking are so inflexible and maladaptive that they interfere with his or her functioning, then that person is considered to have a personality disorder. 10 specific personality disorders are included in DSM-IV-TR: paranoid personality disorder (pervasive distrust and suspiciousness of others), schizoid personality disorder (lack of desire for social relationships and a restricted expression of emotions), schizotypal personality disorder (acute discomfort with close relationships, odd beliefs, perceptual distortions, and eccentricities of behaviour), antisocial personality disorder (disregard for the rights of others), borderline personality disorder (instability of personal relationships and self-image, fears of abandonment, and marked impulsivity), histrionic personality disorder (extensive emotionality and attention seeking), narcissistic personality disorder (grandiosity, need for admiration, and lack of empathy), avoidant personality disorder (social inhibition, and hypersensitivity to negative evaluation), dependent personality disorder (excessive need to be taken care of), and obsessive-compulsive personality disorder (preoccupation with orderliness, perfectionism, stubbornness).

Research planning for DSM-V and ICD-11

It is currently anticipated that the DSM-V will be published in 2012 and that the ICD-11, although likely to be in a final draft form around the same time, will be officially published a few years later after approval by the WHO Assembly. When the last major revision of the DSM, DSM-IV, was published in 1994, the American Psychiatric Association decided to hold off on starting work on the next revision of the DSM until at least 2010, at least partly in response to the criticism that the seven year interval between prior versions of the DSM was too short.[34] Similarly, resistance to the implementation of ICD-10 by many countries (including the United States) ensured that the next revision of the ICD would also be put off for a number of years. The American Psychiatric Association decided to take advantage of this delay in the diagnostic revision process by partnering with the National Institute of Mental Health and the World Health Organization in order to initiate a research planning process with the aim of stimulating potentially informative research prior to the formal beginning of the DSM-V and ICD-11 revision processes.

Part of the impetus for encouraging research in advance of the next diagnostic revision is the general frustration felt by both researchers and clinicians with the superficially descriptive approach taken by DSM-IV and ICD-10. Although the operationalized criteria in DSM-III were developed based largely on expert consensus, there was a general understanding that the categories would continually be revised and improved in future editions of the DSM, ultimately culminating in the identification of the underlying disease processes.

Unfortunately, in the more than 25 years since the publication of DSM-III, the goal of validating these syndromes and discovering the underlying pathophysiology has remained elusive. Despite many proposed candidates, not one laboratory marker has been found to be diagnostically useful for any DSM category.[35] Epidemiological and clinical studies have shown extremely high rates of comorbidities among the disorders, undermining the hypotheses that these syndromes represent distinct etiologies. Regarding treatment, lack of treatment specificity is the rule: SSRI's effective for treating disorders across the diagnostic spectrum (e.g. depression, panic, generalized anxiety disorder, posttraumatic stress disorder, social anxiety, body dysmorphic disorder, obsessive-compulsive disorder, pathological gambling, trichotillomania, borderline personality disorder, etc.). Results of twin studies have also contradicted DSM-IV's assumptions that separate syndromes have a distinct underlying genetic basis (e.g. major depressive disorder and generalized anxiety disorder have the same genetic risk factors).

The considerable limitations of the DSM paradigm have fueled the desire that DSM-V and ICD-11 would be etiologically-based rather just descriptive. The main barrier to making DSM-V and ICD-11 more etiological is, of course, the enormous gaps in our understanding of the pathophysiology of mental disorders. Therefore, in order to help move the field forward towards the goal of a primarily etiological classification, a series of 'white papers' was commissioned under joint sponsorship of the American Psychiatric Association, the National Institute of Mental Health (NIMH), National Institute for Alcoholism and Alcohol Abuse (NIAAA), and the National Institute for Drug Abuse (NIDA). Research planning workgroups responsible for the development of these white papers were constituted for two primary reasons: 1) to stimulate research that will enrich the empirical data base prior to the start of the DSM-V revision process; and 2) to devise a research and analytic agenda that would facilitate the integration of findings from animal studies, genetics, neuroscience, epidemiology, clinical research, cross-cultural research, and clinical services research, which will lead to the eventual development of an etiologically-based, scientifically-sound classification system. In order to encourage thinking beyond the current DSM-IV framework, most of the workgroup members had not been closely involved in the DSM-IV development process. Furthermore, rather than organizing the white paper workgroups around the traditional diagnostic categories, the workgroups focused instead on cross-cutting issues, which included 1) a basic nomenclature workgroup, focusing on a variety of issues that had to do with the way disorders are classified in the DSM; 2) a neuroscience and genetics workgroup whose focus was to develop a basic and clinical neuroscience and genetics research agenda to guide the development of a future pathophysiologically-based classification; 3) a developmental science workgroup which outlined a research agenda to inform developmental aspects of the diagnostic classification; 4) a workgroup focusing on two major gaps in the DSM-IV, namely inadequacies in the classification of personality disorders and of relational disorders; 5) a mental disorders and disability workgroup which focused on disentangling the concepts of symptom severity and disability; and 6) a culture and psychiatric diagnosis workgroup which considered cross-cultural issues in diagnosis and classification. It should be noted that given the breakthrough nature of the suggested research and the relatively short time frame leading up to the anticipated publication of DSM-V and ICD-11, it was understood that most of the proposed research agenda was unlikely to bear fruit until DSM-VI/ICD-12 or later.

The six white papers were published by American Psychiatric Publishing, Inc. in 2002 in a monograph entitled 'A Research Agenda for DSM-V'.[36] Three additional white papers, one focusing on gender issues, one focusing on diagnostic issues in the geriatric population, and one focusing on mental disorders in infants and young children were commissioned subsequently and appear in a second volume of the research agenda.[37]

The second phase of the DSM-V Research Planning Process consisted of 11 research planning conferences (plus a methods conference) that occurred from 2004 to 2007. These conferences were being organized with the assistance and support of the World Health Organization and are co-funded by APA, NIMH, NIAAA, and NIDA. Unlike the white papers in the first phase which focused on general cross-cutting issues, these conferences for the most part focussed on specific diagnostic topics. The primary goals of these conferences were to stimulate the empirical research necessary to allow informed decision-making regarding crucial diagnostic deficiencies identified in DSM-IV and ICD-10, and to promote international collaboration in order to increase the likelihood of developing a future unified DSM/ICD (i.e. each conference had two co-chairs, one from the United States and the other outside the US, each conference included an equal number of the US and international participants, and half the conferences took place outside the US). Conference topics were selected after consultation with the US and international experts. Finite resources necessitated that the number of conferences be limited to a total of 11 (and the number of participants to 25); thus a number of potentially important topics could not be included. The 11 diagnostic-topic focused conferences covered Dimensional Approaches Personality Disorders (December 2004, Arlington, VA);[38] Substance-Related Disorders (February 2005, Rockville, MD);[39] Stress-Induced and Fear Circuitry Disorders (June 2005, Arlington, VA); Dementia (September 2005, Geneva, Switzerland),[40] Deconstructing Psychosis (February 2006, Arlington, VA);[41] Obsessive-Compulsive Spectrum Disorders (June 2006, Arlington, VA);[33] Dimensional Approaches to Diagnosis (July 2006, Bethesda, MD);[42] Somatic Presentations (September 2006, Beijing, China); Externalizing Disorders of Childhood (February 2007, Mexico City); Comorbidity of Anxiety and Depression (June 2007, London, UK), and Public Health Implications (September 2007, Geneva, Switzerland). An additional conference on Autism spectrum disorders was also convened in Sacramento, CA in February 2008. Summaries of the conferences are available on the DSM-V website: www.dsm5.org.

The future

Despite the ubiquitous desire to move from a descriptive classification system to an etiologically-based classification system defined

by objective laboratory findings, results of the research planning process indicate that disorders in DSM-V and ICD-11 will continue to be defined based on descriptive symptomatology. Despite the advances in neuroimaging, genetics, and biological markers over the past 10 years, it is unlikely that any objective laboratory findings will be part of the definition of any DSM-V or ICD-11 disorder, with the one exception being polysomnography findings to define sleep disorders given that such findings are already part of the diagnostic definitions of sleep disorders in the International Classification of Sleep Disorders – 2nd Edition (ICSD-2).[43] Although research studies have reliably demonstrated differences in a wide variety of objective measures between groups of affected individuals and controls (for example, brain ventricular size in individuals with schizophrenia as compared to unaffected controls),[44] when it comes down to applying the findings to a particular individual for the purpose of making a diagnosis, none of these findings have been shown to be sufficiently sensitive or specific.

A central question being raised as part of the DSM-V/ICD-11 revision process is whether psychiatric diagnosis would be better served by a dimensional approach rather than the current categorical approach. This topic was the exclusive focus of one of the 11 research planning conferences being held in advance of the DSM-V/ICD-11 revision (i.e. 'Dimensional Approaches in Diagnostic Classification: A Critical Appraisal', held in Arlington VA, July 27–28, 2006)[42], was the main focus of the Personality Disorders Research Planning conference which proposed a research agenda for adopting a dimensional approach to personality disorders[45] and was a important component of the other diagnostic-related conferences (e.g. 'Should there be both categorical and dimensional criteria for the substance use disorders in DSM-V'.[46]) Much of the impetus for moving toward a dimensional approach comes from dissatisfaction with categorical diagnoses expressed by the research community.[47–58] There are a number of persuasive arguments for the superiority of a dimensional approach over a categorical one, including, 1) the lack of evidence for discrete breaks or demarcations in distributions of symptoms; 2) evidence of a superior fit of empirical data to latent structuring models that correspond to dimensional vs. categorical approaches; 3) higher levels of diagnostic reliability and stability over time; and 4) elimination of the problematic artifacts of the categorical system, like excessive diagnostic comorbidity and arbitrary diagnostic thresholds.[59] On the other hand, a categorical approach to diagnosis remains critically important because of its clinical and administrative utility. Clinicians typically must make dichotomous decisions in everyday practice (i.e. whether to treat or not treat, to hospitalize or not hospitalize, to refer or not refer, etc.) and need to assign diagnostic categories to their patients for the purposes of reimbursement.[60] Categorical labels also facilitate clinical communication by providing a convenient shorthand when discussing a patient's diagnosis (e.g. it is more efficient for a clinician to use a single diagnostic term when talking about a patient who has a borderline personality disorder rather than having to describe all the dimensions that went into that summary judgment). Given the relative advantages of each approach, DSM-V and ICD-11 will most likely adopt some sort of hybrid approach, retaining the categorical diagnoses for communication and decision-making purposes but also providing accompanying dimensions.

Finally, there is a strong push for DSM-V and ICD-11 to become more developmentally focused. ICD-10 and DSM-IV provide definitions of mental disorders that, as far as possible, are applicable across age groups. Given that most disorders can occur at any time during an individual's lifespan, for the most part, the definitions of disorder ignore developmental variations in presentation. It should be noted that DSM-IV does include some age-appropriate modifications in the diagnostic definitions (e.g. in PTSD, criterion B(1), 'recurrent and intrusive distressing recollections of the event, including images, thoughts, or perceptions' is supplemented by 'in young children, repetitive play may occur in which themes or aspects of the trauma are expressed'), these modifications are the exceptions. It is anticipated that DSM-IV and ICD-11 will pay considerably more attention to these issues, reviewing longitudinal studies that track the evolution of disorders across the lifespan and also consider whether definitions should be modified to take into account developmental context.

Appendix 1

International classification of diseases, 10th revision

F00–F09 Organic, including symptomatic, mental disorders

F00 Dementia in Alzheimer's disease

F00.0 Dementia in Alzheimer's disease with early onset
F00.1 Dementia in Alzheimer's disease with late onset
F00.2 Dementia in Alzheimer's disease, atypical or mixed type
F00.9 Dementia in Alzheimer's disease, unspecified

F01 Vascular dementia

F01.0 Vascular dementia of acute onset
F01.1 Multi-infarct dementia
F01.2 Subcortical vascular dementia
F01.3 Mixed cortical and subcortical vascular dementia
F01.8 Other vascular dementia
F01.9 Vascular dementia, unspecified

F02 Dementia in other diseases classified elsewhere

F02.0 Dementia in Pick's disease
F02.1 Dementia in Creutzfeldt–Jakob disease
F02.2 Dementia in Huntington's disease
F02.3 Dementia in Parkinson's disease
F02.4 Dementia in human immunodeficiency virus (HIV) disease
F02.8 Dementia in other specified diseases classified elsewhere

F03 Unspecified dementia

A fifth character may be added to specify dementia in F00–F03, as follows:

.*x*0 Without additional symptoms
.*x*l Other symptoms, predominantly delusional
.*x*2 Other symptoms, predominantly hallucinatory
.*x*3 Other symptoms, predominantly depressive
.*x*4 Other mixed symptoms

F04 Organic amnesic syndrome, not induced by alcohol and other psychoactive substances

F05 Delirium, not induced by alcohol and other psychoactive substances

F05.0 Delirium, not superimposed on dementia, so described
F05.1 Delirium, superimposed on dementia
F05.8 Other delirium
F05.9 Delirium, unspecified

F06 Other mental disorders due to brain damage and dysfunction and to physical disease

F06.0 Organic hallucinosis
F06.1 Organic catatonic disorder
F06.2 Organic delusional (schizophrenia-like) disorder
F06.3 Organic mood (affective) disorders
.30 Organic manic disorder
.31 Organic bipolar disorder
.32 Organic depressive disorder
.33 Organic mixed affective disorder
F06.4 Organic anxiety disorder
F06.5 Organic dissociative disorder
F06.6 Organic emotionally labile (asthenic) disorder
F06.7 Mild cognitive disorder
F06.8 Other specified mental disorders due to brain damage and dysfunction and to physical disease
F06.9 Unspecified mental disorder due to brain damage and dysfunction and to physical disease

F09 Unspecified organic or symptomatic mental disorder

F07 Personality and behavioural disorders due to brain disease, damage and dysfunction

F07.0 Organic personality disorder
F07.1 Postencephalitic syndrome
F07.2 Postconcussional syndrome
F07.8 Other organic personality and behavioural disorders due to brain disease, damage, and dysfunction
F07.9 Unspecified organic personality and behavioural disorder due to brain disease, damage, and dysfunction

F10–F19 Mental and behavioural disorders due to psychoactive substance use

F10 Mental and behavioural disorders due to use of alcohol
F11 Mental and behavioural disorders due to use of opioids
F12 Mental and behavioural disorders due to use of cannabinoids
F13 Mental and behavioural disorders due to use of sedatives or hypnotics
F14 Mental and behavioural disorders due to use of cocaine
F15 Mental and behavioural disorders due to use of other stimulants, including caffeine
F16 Mental and behavioural disorders due to use of hallucinogens
F17 Mental and behavioural disorders due to use of tobacco
F18 Mental and behavioural disorders due to use of volatile solvents
F19 Mental and behavioural disorders due to multiple drug use and use of other psychoactive substances

Four- and five-character categories may be used to specify the clinical conditions, as follows

F1*x*.0 Acute intoxication
.00 Uncomplicated
.01 With trauma or other bodily injury
.02 With other medical complications
.03 With delirium
.04 With perceptual distortions
.05 With coma
.06 With convulsions
.07 Pathological intoxication
F1*x*.1 Harmful use
F1*x*.2 Dependence syndrome
.20 Currently abstinent
.21 Currently abstinent, but in a protected environment
.22 Currently on a clinically supervised maintenance or replacement regime (controlled dependence)
.23 Currently abstinent, but receiving treatment with aversive or blocking drugs
.24 Currently using the substance (active dependence)
.25 Continuous use
.26 Episodic use (dipsomania)
F1*x*.3 Withdrawal state
.30 Uncomplicated
.31 Convulsions
F1*x*.4 Withdrawal state with delirium
.40 Without convulsions
.41 With convulsions
F1*x*.5 Psychotic disorder
.50 Schizophrenia-like
.51 Predominantly delusional
.52 Predominantly hallucinatory
.53 Predominantly polymorphic
.54 Predominantly depressive symptoms
.55 Predominantly manic symptoms
.56 Mixed
F1*x*.6 Amnesic syndrome
F1*x*.7 Residual and late-onset psychotic disorder
.70 Flashbacks
.71 Personality or behaviour disorder
.72 Residual affective disorder
.73 Dementia
.74 Other persisting cognitive impairment
.75 Late-onset psychotic disorder
F1*x*.8 Other mental and behavioural disorders
F1*x*.9 Unspecified mental and behavioural disorder

F20–F29 Schizophrenia, schizotypal, and delusional disorders

F20 Schizophrenia

F20.0 Paranoid schizophrenia
F20.1 Hebephrenic schizophrenia
F20.2 Catatonic schizophrenia
F20.3 Undifferentiated schizophrenia
F20.4 Post-schizophrenic depression
F20.5 Residual schizophrenia
F20.6 Simple schizophrenia
F20.8 Other schizophrenia
F20.9 Schizophrenia, unspecified

A fifth character may be used to classify course
*x*0 Continuous
.*x*1 Episodic with progressive deficit
.*x*2 Episodic with stable deficit
.*x*3 Episodic remittent
.*x*4 Incomplete remission
.*x*5 Complete remission
.*x*8 Other
.*x*9 Period of observation less than one year

F21 Schizotypal disorder

F22 Persistent delusional disorders

F22.0 Delusional disorder
F22.8 Other persistent delusional disorders
F22.9 Persistent delusional disorder, unspecified

F23 Acute and transient psychotic disorders

F23.0 Acute polymorphic psychotic disorder without symptoms of schizophrenia
F23.1 Acute polymorphic psychotic disorder with symptoms of schizophrenia
F23.2 Acute schizophrenia-like psychotic disorder
F23.3 Other acute predominantly delusional psychotic disorders
F23.8 Other acute and transient psychotic disorders
F23.9 Acute and transient psychotic disorders unspecified

A fifth character may be used to identify the presence or absence of associated acute stress

.*x*0 Without associated acute stress
.*x*1 With associated acute stress

F24 Induced delusional disorder

F25 Schizoaffective disorders

F25.0 Schizoaffective disorder, manic type
F25.1 Schizoaffective disorder, depressive type
F25.2 Schizoaffective disorder, mixed type
F25.8 Other schizoaffective disorders
F25.9 Schizoaffective disorder, unspecified

F28 Other non-organic psychotic disorders

F29 Unspecified non-organic psychosis

F30–F39 Mood (affective) disorders

F30 Manic episode
F30.0 Hypomania
F30.1 Mania without psychotic symptoms
F30.2 Mania with psychotic symptoms
F30.8 Other manic episodes
F30.9 Manic episode, unspecified

F31 Bipolar affective disorder

F31.0 Bipolar affective disorder, current episode hypomanic
F31.1 Bipolar affective disorder, current episode manic without psychotic symptoms
F31.2 Bipolar affective disorder, current episode manic with psychotic symptoms
F31.3 Bipolar affective disorder, current episode mild or moderate depression
.30 Without somatic symptoms
.31 With somatic symptoms
F31.4 Bipolar affective disorder, current episode severe depression without psychotic symptoms
F31.5 Bipolar affective disorder, current episode severe depression with psychotic symptoms
F31.6 Bipolar affective disorder, current episode mixed
F31.7 Bipolar affective disorder, currently in remission
F31.8 Other bipolar affective disorders
F31.9 Bipolar affective disorder, unspecified

F32 Depressive episode

F32.0 Mild depressive episode
.00 Without somatic symptoms
.01 With somatic symptoms
F32.1 Moderate depressive episode
.10 Without somatic symptoms
.11 With somatic symptoms
F32.2 Severe depressive episode without psychotic symptoms
F32.3 Severe depressive episode with psychotic symptoms
F32.8 Other depressive episodes
F32.9 Depressive episode, unspecified

F33 Recurrent depressive disorder

F33.0 Recurrent depressive disorder, current episode mild
.00 Without somatic symptoms
.01 With somatic symptoms
F33.1 Recurrent depressive disorder, current episode moderate
.10 Without somatic symptoms
.11 With somatic symptoms
F33.2 Recurrent depressive disorder, current episode severe without psychotic symptoms
F33.3 Recurrent depressive disorder, current episode severe with psychotic symptoms
F33.4 Recurrent depressive disorder, currently in remission
F33.8 Other recurrent depressive disorders
F33.9 Recurrent depressive disorder, unspecified

F34 Persistent mood (affective) disorders

F34.0 Cyclothymia
F34.1 Dysthymia
F34.8 Other persistent mood (affective) disorders
F34.9 Persistent mood (affective) disorder, unspecified

F38 Other mood (affective) disorders

F38.0 Other single mood (affective) disorders
.00 Mixed affective episode
F38.1 Other recurrent mood (affective) disorders
.10 Recurrent brief depressive disorder
F38.8 Other specified mood (affective) disorders

F39 Unspecified mood (affective) disorder

F40–F48 Neurotic, stress-related, and somatoform disorders

F40 Phobic anxiety disorders
F40.0 Agoraphobia
.00 Without panic disorder
.01 With panic disorder
F40.1 Social phobias
F40.2 Specific (isolated) phobias
F40.8 Other phobic anxiety disorders
F40.9 Phobic anxiety disorder, unspecified

F41 Other anxiety disorders

F41.0 Panic disorder (episodic paroxysmal anxiety)
F41.1 Generalized anxiety disorder
F41.2 Mixed anxiety and depressive disorder
F41.3 Other mixed anxiety disorders
F41.8 Other specified anxiety disorders
F41.9 Anxiety disorder, unspecified

F42 Obsessive-compulsive disorder

F42.0 Predominantly obsessional thoughts or ruminations
F42.1 Predominantly compulsive acts (obsessional rituals)
F42.2 Mixed obsessional thoughts and acts
F42.8 Other obsessive-compulsive disorders
F42.9 Obsessive-compulsive disorder, unspecified

F43 Reaction to severe stress, and adjustment disorders

F43.0 Acute stress reaction
F43.1 Post-traumatic stress disorder
F43.2 Adjustment disorders
.20 Brief depressive reaction

.21 Prolonged depressive reaction
.22 Mixed anxiety and depressive reaction
.23 With predominant disturbance of other emotions
.24 With predominant disturbance of conduct
.25 With mixed disturbance of emotions and conduct
.28 With other specified predominant symptoms
F43.8 Other reactions to severe stress
F43.9 Reaction to severe stress, unspecified

F44 Dissociative (conversion) disorders

F44.0 Dissociative amnesia
F44.1 Dissociative fugue
F44.2 Dissociative stupor
F44.3 Trance and possession disorders
F44.4 Dissociative motor disorders
F44.5 Dissociative convulsions
F44.6 Dissociative anaesthesia and sensory loss
F44.7 Mixed dissociative (conversion) disorders
F44.8 Other dissociative (conversion) disorders
.80 Ganser's syndrome
.81 Multiple personality disorder
.82 Transient dissociate (conversion) disorders occurring in childhood and adolescence
.88 Other specified dissociative (conversion) disorders
F44.9 Dissociative (conversion) disorder, unspecified

F45 Somatoform disorders

F45.0 Somatization disorder
F45.1 Undifferentiated somatoform disorder
F45.2 Hypochondriacal disorder
F45.3 Somatoform autonomic dysfunction
.30 Heart and cardiovascular system
.31 Upper gastrointestinal tract
.32 Lower gastrointestinal tract
.33 Respiratory system
.34 Genitourinary system
.38 Other organ or system
F45.4 Persistent somatoform pain disorder
F45.8 Other somatoform disorders
F45.9 Somatoform disorder, unspecified

F48 Other neurotic disorders

F48.0 Neurasthenia
F48.1 Depersonalization-derealization syndrome
F48.8 Other specified neurotic disorders
F48.9 Neurotic disorder, unspecified

F50–F59 Behavioural syndromes associated with physiological disturbances and physical factors

F50 Eating disorders

F50.0 Anorexia nervosa
F50.1 Atypical anorexia nervosa
F50.2 Bulimia nervosa
F50.3 Atypical bulimia nervosa
F50.4 Overeating associated with other psychological disturbances
F50.5 Vomiting associated with other psychological disturbances
F50.8 Other eating disorders
F50.9 Eating disorder, unspecified

F51 Non-organic sleep disorders

F51.0 Non-organic insomnia
F51.1 Non-organic hypersomnia
F51.2 Non-organic disorder of the sleep-wake schedule
F51.3 Sleepwalking (somnambulism)
F51.4 Sleep terrors (night terrors)
F51.5 Nightmares
F51.8 Other non-organic sleep disorders
F51.9 Non-organic sleep disorder, unspecified

F52 Sexual dysfunction, not caused by organic disorder or disease

F52.0 Lack or loss of sexual desire
F52.1 Sexual aversion and lack of sexual enjoyment
.10 Sexual aversion
.11 Lack of sexual enjoyment
F52.2 Failure of genital response
F52.3 Orgasmic dysfunction
F52.4 Premature ejaculation
F52.5 Non-organic vaginismus
F52.6 Non-organic dyspareunia
F52.7 Excessive sexual drive
F52.8 Other sexual dysfunction, not caused by organic disorders or disease
F52.9 Unspecified sexual dysfunction, not caused by organic disorder or disease

F53 Mental and behavioural disorders associated with the puerperium, not elsewhere classified

F53.0 Mild mental and behavioural disorders associated with the puerperium, not elsewhere classified
F53.1 Severe mental and behavioural disorders associated with the puerperium, not elsewhere classified
F53.8 Other mental and behavioural disorders associated with the puerperium, not elsewhere classified
F53.9 Puerperal mental disorder, unspecified

F54 Psychological and behavioural factors associated with disorders or diseases classified elsewhere

F55 Abuse of non-dependence-producing substances

F55.0 Antidepressants
F55.1 Laxatives
F55.2 Analgesics
F55.3 Antacids
F55.4 Vitamins
F55.5 Steroids or hormones
F55.6 Specific herbal or folk remedies
F55.8 Other substances that do not produce dependence
F55.9 Unspecified

F59 Unspecified behavioural syndromes associated with physiological disturbances and physical factors

F60–F69 Disorders of adult personality and behaviour

F60 Specific personality disorders
F60.0 Paranoid personality disorder
F60.1 Schizoid personality disorder
F60.2 Dissocial personality disorder
F60.3 Emotionally unstable personality disorder
.30 Impulsive type
.31 Borderline type
F60.4 Histrionic personality disorder
F60.5 Anankastic personality disorder

F60.6 Anxious (avoidant) personality disorder
F60.7 Dependent personality disorder
F60.8 Other specific personality disorders
F60.9 Personality disorder, unspecified

F61 Mixed and other personality disorders

F61.0 Mixed personality disorders
F61.1 Troublesome personality changes

F62 Enduring personality changes, not attributable to brain damage and disease

F62.0 Enduring personality change after catastrophic experience
F62.1 Enduring personality change after psychiatric illness
F62.8 Other enduring personality changes
F62.9 Enduring personality change, unspecified

F63 Habit and impulse disorders

F63.0 Pathological gambling
F63.1 Pathological fire-setting (pyromania)
F63.2 Pathological stealing (kleptomania)
F63.3 Trichotillomania
F63.8 Other habit and impulse disorders
F63.9 Habit and impulse disorder, unspecified

F64 Gender identity disorders

F64.0 Transsexualism
F64.1 Dual-role transvestism
F64.2 Gender identity disorder of childhood
F64.8 Other gender identity disorders
F64.9 Gender identity disorder, unspecified

F65 Disorders of sexual preference

F65.0 Fetishism
F65.1 Fetishistic transvestism
F65.2 Exhibitionism
F65.3 Voyeurism
F65.4 Paedophilia
F65.5 Sadomasochism
F65.6 Multiple disorders of sexual preference
F65.8 Other disorders of sexual preference
F65.9 Disorder of sexual preference, unspecified

F66 Psychological and behavioural disorders associated with sexual development and orientation

F66.0 Sexual maturation disorder
F66.1 Egodystonic sexual orientation
F66.2 Sexual relationship disorder
F66.8 Other psychosexual development disorders
F66.9 Psychosexual development disorder, unspecified

A fifth character may be used to indicate association with:

.*x*0 Heterosexuality
.*x*l Homosexuality
.*x*2 Bisexuality
.*x*8 Other, including prepubertal

F68 Other disorders of adult personality and behaviour

F68.0 Elaboration of physical symptoms for psychological reasons
F68.1 Intentional production or feigning of symptoms or disabilities, either physical or psychological (factitious disorder)
F68.8 Other specified disorders of adult personality and behaviour
F69 Unspecified disorder of adult personality and behaviour

F70–F79 Mental retardation (intellectual disability)

F70 Mild mental retardation (intellectual disability)

F71 Moderate mental retardation (intellectual disability)

F72 Severe mental retardation (intellectual disability)

F73 Profound mental retardation (intellectual disability)

F78 Other mental retardation (intellectual disability)

F79 Unspecified mental retardation (intellectual disability)

A fourth character may be used to specify the extent of associated behavioural impairment:

F7x.0 No, or minimal, impairment of behaviour
F7x.1 Significant impairment of behaviour requiring attention or treatment
F7x.8 Other impairments of behaviour
F7x.9 Without mention of impairment of behaviour

F80–F89 Disorders of psychological development

F80 specific developmental disorders of speech and language

F80.0 Specific speech articulation disorder
F80.1 Expressive language disorder
F80.2 Receptive language disorder
F80.3 Acquired aphasia with epilepsy (Landau-Kleffner syndrome)
F80.8 Other developmental disorders of speech and language
F80.9 Developmental disorder of speech and language, unspecified

F81 Specific developmental disorders of scholastic skills

F81.0 Specific reading disorder
F81.1 Specific spelling disorder
F81.2 Specific disorder of arithmetical skills
F81.3 Mixed disorder of scholastic skills
F81.8 Other developmental disorders of scholastic skills
F81.9 Developmental disorder of scholastic skills, unspecified

F82 Specific developmental disorder of motor function

F83 Mixed specific developmental disorders

F84 Pervasive developmental disorders

F84.0 Childhood autism
F84.1 Atypical autism
F84.2 Rett's syndrome
F84.3 Other childhood disintegrative disorder
F84.4 Overactive disorder associated with mental retardation and stereotyped movements
F84.5 Asperger's syndrome
F84.8 Other pervasive developmental disorders
F84.9 Pervasive developmental disorder, unspecified

F88 Other disorders of psychological development

F89 Unspecified disorder of psychological development

F90–F98 Behavioural and emotional disorders with onset usually occurring in childhood and adolescence

F90 Hyperkinetic disorders

F90.0 Disturbance of activity and attention
F90.1 Hyperkinetic conduct disorder
F90.8 Other hyperkinetic disorders
F90.9 Hyperkinetic disorder, unspecified

F91 Conduct disorders

F91.0 Conduct disorder confined to the family context
F91.1 Unsocialized conduct disorder

F91.2 Socialized conduct disorder
F91.3 Oppositional defiant disorder
F91.8 Other conduct disorders
F91.9 Conduct disorder, unspecified

F92 Mixed disorders of conduct and emotions

F92.0 Depressive conduct disorder
F92.8 Other mixed disorders of conduct and emotions
F92.9 Mixed disorder of conduct and emotions, unspecified

F93 Emotional disorders with onset specific to childhood

F93.0 Separation anxiety disorder of childhood
F93.1 Phobic anxiety disorder of childhood
F93.2 Social anxiety disorder of childhood
F93.3 Sibling rivalry disorder
F93.8 Other childhood emotional disorders
F93.9 Childhood emotional disorder, unspecified

F94 Disorders of social functioning with onset specific to childhood and adolescence

F94.0 Elective mutism
F94.1 Reactive attachment disorder of childhood
F94.2 Disinhibited attachment disorder of childhood
F94.8 Other childhood disorders of social functioning
F94.9 Childhood disorders of social functioning, unspecified

F95 Tic disorders

F95.0 Transient tic disorder
F95.1 Chronic motor or vocal tic disorder
F95.2 Combined vocal and multiple motor tic disorder (de la Tourette's syndrome)
F95.8 Other tic disorders
F95.9 Tic disorder, unspecified

F98 Other behavioural and emotional disorders with onset usually occurring in childhood and adolescence

F98.0 Non-organic enuresis
F98.1 Non-organic encopresis
F98.2 Feeding disorder of infancy and childhood
F98.3 Pica of infancy and childhood
F98.4 Stereotyped movement disorders
F98.5 Stuttering (stammering)
F98.6 Cluttering
F98.8 Other specified behavioural and emotional disorders with onset usually occurring in childhood and adolescence
F98.9 Unspecified behavioural and emotional disorders with onset usually occurring in childhood and adolescence

F99 Unspecified mental disorder

F99 Mental disorder, not otherwise specified

Appendix 2

Diagnostic and Statistical Manual of Mental Disorders (4th edition)

NOS = Not Otherwise Specified.

An x appearing in a diagnostic code indicates that a specific code number is required.

An ellipsis (. . .) is used in the names of certain disorders to indicate that the name of a specific mental disorder or general medical condition should be inserted when recording the name (e.g., 293.0 Delirium Due to Hypothyroidism).

If criteria are currently met, one of the following severity specifiers may be noted after the diagnosis

Mild
Moderate
Severe

If criteria are no longer met, one of the following specifiers may be noted

In Partial Remission
In Full Remission
Prior History

Disorders usually first diagnosed in infancy, childhood, or adolescence

Mental retardation

Note: *These are coded on Axis II.*

317 Mild Mental Retardation
318.0 Moderate Mental Retardation
318.1 Severe Mental Retardation
318.2 Profound Mental Retardation
319 Mental Retardation, Severity Unspecified

Learning disorders

315.00 Reading Disorder
315.1 Mathematics Disorder
315.2 Disorder of Written Expression
315.9 Learning Disorder NOS

Motor skills disorder

315.4 Developmental Coordination Disorder

Communication disorders

315.31 Expressive Language Disorder
315.32 Mixed Receptive–Expressive Language Disorder
315.39 Phonological Disorder
307.0 Stuttering
307.9 Communication Disorder NOS

Pervasive Developmental disorders

299.00 Autistic Disorder
299.80 Rett's Disorder
299.10 Childhood Disintegrative Disorder
299.80 Asperger's Disorder
299.80 Pervasive Developmental Disorder NOS

Attention-deficit and disruptive behaviour disorders

314.xx Attention-Deficit/Hyperactivity Disorder
.01 Combined Type
.00 Predominantly Inattentive Type
.01 Predominantly Hyperactive-Impulsive Type
314.9 Attention-Deficit/Hyperactivity Disorder NOS
312.xx Conduct Disorder
.81 Childhood-Onset Type
.82 Adolescent-Onset Type
.89 Unspecified Onset

313.81 Oppositional-Defiant Disorder
312.9 Disruptive Behaviour Disorder NOS

Feeding and eating disorders of infancy or early childhood

307.52 Pica
307.53 Rumination Disorder
307.59 Feeding Disorder of Infancy or Early Childhood

Tic disorders

307.23 Tourette's Disorder
307.22 Chronic Motor or Vocal Tic Disorder
307.21 Transient Tic Disorder
Specify if: Single Episode/Recurrent
307.20 Tic Disorder NOS

Elimination disorders

—.— Encopresis
787.6 With Constipation and Overflow Incontinence
307.7 Without Constipation and Overflow Incontinence
307.6 Enuresis (Not Due to a General Medical Condition)
Specify type: Nocturnal Only/Diurnal Only/Nocturnal and Diurnal

Other disorders of infancy, childhood, or adolescence

309.21 Separation Anxiety Disorder
Specify if: Early Onset
313.23 Selective Mutism
313.89 Reactive Attachment Disorder of Infancy or Early Childhood
Specify type: Inhibited Type/Disinhibited Type
307.3 Stereotypic Movement Disorder
Specify if: With Self-Injurious Behaviuor
313.9 Disorder of Infancy, Childhood, or Adolescence NOS

Delirium, dementia and amnestic and other cognitive disorders

Delirium

293.0 Delirium Due to . . . [Indicate the General Medical Condition]
—.— Substance Intoxication Delirium (refer to Substance-Related Disorders for substance-specific codes)
—.— Substance Withdrawal Delirium (refer to Substance-Related Disorders for substance-specific codes)
—.— Delirium Due to Multiple Etiologies (code each of the specific etiologies)
780.09 Delirium NOS

Dementia

294.xx Dementia of the Alzheimer's Type, With Early Onset (also code 331.0 Alzheimer's disease on Axis III)
.10 Without Behavioural Disturbance
.11 With Behavioural Disturbance
294.xx Dementia of the Alzheimer's Type, With Late Onset (also code 331.0 Alzheimer's disease on Axis III)
.10 Without Behavioural Disturbance
.11 With Behavioural Disturbance
290.xx Vascular Dementia
.40 Uncomplicated
.41 With Delirium
.42 With Delusions
.43 With Depressed Mood
Specify if: *With Behavioural Disturbance*

Code presence or absence of a behavioural disturbance in the fifth digit for Dementia Due to a General Medical Condition:
294.10 = Without Behavioural Disturbance
294.11 = With Behavioural Disturbance
294.1x Dementia Due to HIV Disease (*also code 042 HIV on Axis III*)
294.1x Dementia Due to Head Trauma (*also code 854.00 head injury on Axis III*)
294.1x Dementia Due to Parkinson's Disease (*also code 331.82 Dementia with Lewy Bodies on Axis III*)
294.1x Dementia Due to Huntington's Disease (*also code 333.4 Huntington's disease on Axis III*)
294.1x Dementia Due to Pick's Disease (*also code 331.11 Pick's disease on Axis III*)
294.1x Dementia Due to Creutzfeldt–Jakob Disease (*also code 046.1 Creutzfeldt–Jakob disease on Axis III*)
294.1x Dementia Due to . . . [*Indicate the General Medical Condition not listed above*] (*also code the general medical condition on Axis III*)
—.— Substance-Induced Persisting Dementia (*refer to Substance-Related Disorders for substance-specific codes*)
—.— Dementia Due to Multiple Etiologies (*code each of the specific etiologies*)
294.8 Dementia NOS

Amnestic disorders

294.0 Amnestic Disorder Due to . . . [*Indicate the General Medical Condition*]
Specify if: Transient/Chronic
—.— Substance-Induced Persisting Amnestic Disorder (*refer to Substance-Related Disorders for substance-specific codes*)
294.8 Amnestic Disorder NOS

Other cognitive disorder

294.9 Cognitive Disorder NOS

Mental disorders due to a general medical condition not elsewhere classified

293.89 Catatonic Disorder Due to . . . [*Indicate the General Medical Condition*]
310.1 Personality Change Due to . . . [*Indicate the General Medical Condition*]
Specify type: Labile Type/Disinhibited Type/Aggressive Type/Apathetic Type/Paranoid Type/Other Type/Combined Type/Unspecified Type
293.9 Mental Disorder NOS Due to . . . [*Indicate the General Medical Condition*]

Substance-related disorders

The following specifiers apply to Substance Dependence as noted:
[a]With Physiological Dependence/Without Physiological Dependence

[b]Early Full Remission/Early Partial Remission Sustained Full Remission/Sustained Partial Remission
[c]In a Controlled Environment
[d]On Agonist Therapy/
The following specifiers apply to Substance-Induced Disorders as noted:
[I]With Onset During Intoxication/[W]With Onset During Withdrawal

Alcohol-related disorders

Alcohol use disorders

303.90 Alcohol Dependencea,b,c
305.00 Alcohol Abuse

Alcohol-induced disorders

303.00 Alcohol Intoxication
291.81 Alcohol Withdrawal
Specify if: With Perceptual Disturbances
291.0 Alcohol Intoxication Delirium
291.0 Alcohol Withdrawal Delirium
291.2 Alcohol-Induced Persisting Dementia
291.1 Alcohol-Induced Persisting Amnestic Disorder
291.x Alcohol-Induced Psychotic Disorder
.5 With Delusions[I,W]
.3 With Hallucinations[I,W]
291.89 Alcohol-Induced Mood Disorder[I,W]
291.89 Alcohol-Induced Anxiety Disorder[I,W]
291.89 Alcohol-Induced Sexual Dysfunction[I]
291.82 Alcohol-Induced Sleep Disorder[I,W]
291.9 Alcohol-Related Disorder NOS

Amphetamine (or Amphetamine-like)-related disorders

Amphetamine use disorders

304.40 Amphetamine Dependence[a,b,c]
305.70 Amphetamine Abuse

Amphetamine-Induced Disorders

292.89 Amphetamine Intoxication
Specify if: With Perceptual Disturbances
292.0 Amphetamine Withdrawal
292.81 Amphetamine Intoxication Delirium
292.xx Amphetamine-Induced Psychotic Disorder
.11 With Delusions[I]
.12 With Hallucinations[I]
292.84 Amphetamine-Induced Mood Disorder[I,W]
292.89 Amphetamine-Induced Anxiety Disorder[I]
292.89 Amphetamine-Induced Sexual Dysfunction[I]
292.85 Amphetamine-Induced Sleep Disorder[I,W]
292.9 Amphetamine-Related Disorder NOS

Caffeine-related disorders

Caffeine-induced disorders

305.90 Caffeine Intoxication
292.89 Caffeine-Induced Anxiety Disorder[I]
292.85 Caffeine-Induced Sleep Disorder[I]
292.9 Caffeine-Related Disorder NOS

Cannabis-related disorders

Cannabis use disorders

304.30 Cannabis Dependence[a,b,c]
305.20 Cannabis Abuse

Cannabis-induced disorders

292.89 Cannabis Intoxication
Specify if: With Perceptual Disturbances
292.81 Cannabis Intoxication Delirium
292.xx Cannabis-Induced Psychotic Disorder
.11 With Delusions[I]
.12 With Hallucinations[I]
292.89 Cannabis-Induced Anxiety Disorder[I]
292.9 Cannabis-Related Disorder NOS

Cocaine-related disorders

Cocaine use disorders

304.20 Cocaine Dependencea,b,c
305.60 Cocaine Abuse

Cocaine-induced disorders

292.89 Cocaine Intoxication
Specify if: With Perceptual Disturbances
292.0 Cocaine Withdrawal
292.81 Cocaine Intoxication Delirium
292.xx Cocaine-Induced Psychotic Disorder
.11 With Delusions[I]
.12 With Hallucinations[I]
292.84 Cocaine-Induced Mood Disorder[I,W]
292.89 Cocaine-Induced Anxiety Disorder[I,W]
292.89 Cocaine-Induced Sexual Dysfunction[I]
292.85 Cocaine-Induced Sleep Disorder[I,W]
292.9 Cocaine-Related Disorder NOS

Hallucinogen-related disorders

Hallucinogen use disorders

304.50 Hallucinogen Dependence[b,c]
305.30 Hallucinogen Abuse

Hallucinogen-induced disorders

292.89 Hallucinogen Intoxication
292.89 Hallucinogen Persisting Perception Disorder (Flashbacks)
292.81 Hallucinogen Intoxication Delirium
292.xx Hallucinogen-Induced Psychotic Disorder
.11 With Delusions[I]
.12 With Hallucinations[I]
292.84 Hallucinogen-Induced Mood Disorder[I]
292.89 Hallucinogen-Induced Anxiety Disorder[I]
292.9 Hallucinogen-Related Disorder NOS

Inhalant-related disorders

Inhalant use disorders

304.60 Inhalant Dependence[b,c]
305.90 Inhalant Abuse

Inhalant-induced disorders

292.89 Inhalant Intoxication
292.81 Inhalant Intoxication Delirium
292.82 Inhalant-Induced Persisting Dementia

292.xx Inhalant-Induced Psychotic Disorder
.11 With Delusions[I]
.12 With Hallucinations[I]
292.84 Inhalant-Induced Mood Disorder[I]
292.89 Inhalant-Induced Anxiety Disorder[I]
292.9 Inhalant-Related Disorder NOS

Nicotine-related disorders

Nicotine use disorder

305.1 Nicotine Dependence[a,b]

Nicotine-induced disorders

292.0 Nicotine Withdrawal
292.9 Nicotine-Related Disorder NOS

Opioid-related disorders

Opioid use disorders

304.00 Opioid Dependence[a,b,c,d]
305.50 Opioid Abuse

Opioid-induced disorders

292.89 Opioid Intoxication
Specify if: With Perceptual Disturbances
292.0 Opioid Withdrawal
292.81 Opioid Intoxication Delirium
292.xx Opioid-Induced Psychotic Disorder
.11 With Delusions[I]
.12 With Hallucinations[I]
292.84 Opioid-Induced Mood Disorder[I]
292.89 Opioid-Induced Sexual Dysfunction[I]
292.85 Opioid-Induced Sleep Disorder[I,W]
292.9 Opioid-Related Disorder NOS

Phencyclidine (or Phencyclidine-like)-related disorders

Phencyclidine use disorders

304.60 Phencyclidine Dependence[b,c]
305.90 Phencyclidine Abuse

Phencyclidine-induced disorders

292.89 Phencyclidine Intoxication
Specify if: With Perceptual Disturbances
292.81 Phencyclidine Intoxication Delirium
292.xx Phencyclidine-Induced Psychotic Disorder
.11 With Delusions[I]
.12 With Hallucinations[I]
292.84 Phencyclidine-Induced Mood Disorder[I]
292.89 Phencyclidine-Induced Anxiety Disorder[I]
292.9 Phencyclidine-Related Disorder NOS

Sedative-, Hypnotic-, Oranxiolytic-related disorders

Sedative, Hypnotic, or Anxiolytic use disorders

304.10 Sedative, Hypnotic, or Anxiolytic Dependence[a,b,c]
305.40 Sedative, Hypnotic, or Anxiolytic Abuse

Sedative-, Hypnotic-, or Anxiolytic-induced disorders

292.89 Sedative, Hypnotic, or Anxiolytic Intoxication
292.0 Sedative, Hypnotic, or Anxiolytic Withdrawal
Specify it: With Perceptual Disturbances
292.81 Sedative, Hypnotic, or Anxiolytic Intoxication Delirium
292.81 Sedative, Hypnotic, or Anxiolytic Withdrawal Delirium
292.82 Sedative-, Hypnotic-, or Anxiolytic-Induced Persisting Dementia
292.83 Sedative-, Hypnotic-, or Anxiolytic-Induced Persisting Amnestic Disorder
292.xx Sedative-, Hypnotic-, or Anxiolytic-Induced Psychotic Disorder
.11 With Delusions[I,W]
.12 With Hallucinations[I,W]
292.84 Sedative-, Hypnotic-, or Anxiolytic-Induced Mood Disorder[I,W]
292.89 Sedative-, Hypnotic-, or Anxiolytic-Induced Anxiety Disorder[W]
292.89 Sedative-, Hypnotic-, or Anxiolytic-Induced Sexual Dysfunction[1]
292.85 Sedative-, Hypnotic-, or Anxiolytic-Induced Sleep Disorder[I,W]
292.9 Sedative-, Hypnotic-, or Anxiolytic-Related Disorder NOS

Polysubstance-related disorder

304.80 Polysubstance Dependence[a,b,c,d]

Other (or Unknown) substance-related disorders

Other (or Unknown) substance use disorders

304.90 Other (or Unknown) Substance Dependence[a,b,c,d]
305.90 Other (or Unknown) Substance Abuse

Other (or Unknown) substance-induced disorders

292.89 Other (or Unknown) Substance Intoxication
Specify it: With Perceptual Disturbances
292.0 Other (or Unknown) Substance Withdrawal
Specify it: With Perceptual Disturbances
292.81 Other (or Unknown) Substance-Induced Delirium
292.82 Other (or Unknown) Substance-Induced Persisting Dementia
292.83 Other (or Unknown) Substance-Induced Persisting Amnestic Disorder
292.xx Other (or Unknown) Substance-Induced Psychotic Disorder
.11 With Delusions[I,W]
.12 With Hallucinations[I,W]
292.84 Other (or Unknown) Substance-Induced Mood Disorder[I,W]
292.89 Other (or Unknown) Substance-Induced Anxiety Disorder[I,W]
292.89 Other (or Unknown) Substance-Induced Sexual Dysfunction[1]
292.85 Other (or Unknown) Substance-Induced Sleep Disorder[I,W]
292.9 Other (or Unknown) Substance-Related Disorder NOS

Schizophrenia and other psychotic disorders

295.xx Schizophrenia
The following Classification of Longitudinal Course applies to all subtypes of Schizophrenia.

Episodic With Interepisode Residual Symptoms (*specify if*: With Prominent Negative Symptoms)/Episodic With No Interepisode Residual Symptoms/Continuous (*specify if*: With Prominent Negative Symptoms)

Single Episode In Partial Remission (*specify if*: With Prominent Negative Symptoms)/Single Episode In Full Remission
Other or Unspecified Pattern
.30 Paranoid Type
.10 Disorganized Type
.20 Catatonic Type
.90 Undifferentiated Type
.60 Residual Type
295.40 Schizophreniform Disorder
Specify it: Without Good Prognostic Features/With Good Prognostic Features
295.70 Schizoaffective Disorder
Specify type: Bipolar Type/Depressive Type
297.1 Delusional Disorder
Specify type: Erotomanic Type/Grandiose Type/Jealous Type/Persecutory Type/Somatic Type/Mixed Type/Unspecified Type
298.8 Brief Psychotic Disorder
Specify it: With Marked Stressor(s)/Without Marked Stressor(s)/With Postpartum Onset
297.3 Shared Psychotic Disorder
293.xx Psychotic Disorder Due to . . . [Indicate the General Medical Condition]
.81 With Delusions
.82 With Hallucinations
—.— Substance-Induced Psychotic Disorder (refer to Substance-Related Disorders for substance-specific codes)
Specify it: With Onset During Intoxication/With Onset During Withdrawal
298.9 Psychotic Disorder NOS

Mood disorders

Code current state of Major Depressive Disorder or Bipolar I Disorder in fifth digit:
1 = Mild
2 = Moderate
3 = Severe Without Psychotic Features
4 = Severe With Psychotic Features
Specify: Mood-Congruent Psychotic Features/Mood-Incongruent Psychotic Features
5 = In Partial Remission
6 = In Full Remission
0 = Unspecified
The following specifiers apply (for current or most recent episode) to Mood Disorders as noted
[a]Severity/Psychotic/Remission Specifiers/[b]Chronic/[c]With Catatonic Features/[d]With Melancholic Features/[e]With Atypical Features/[f]With Postpartum Onset
The following specifiers apply to Mood Disorders as noted:
[g]With or Without Full Interepisode Recovery/[h]With Seasonal Pattern/[i]With Rapid Cycling

Depressive disorders

296.xx Major Depressive Disorder,
.2x Single Episode[a,b,c,d,e,f]
.3x Recurrent[a,b,c,d,e,f,g,h]
300.4 Dysthymic Disorder
Specify it: Early Onset/Late Onset
Specify: With Atypical Features
311 Depressive Disorder NOS

Bipolar disorders

296.xx Bipolar I Disorder,
.0x Single Manic Episode[a,c,f]
Specify if: Mixed
.40 Most Recent Episode Hypomanic[g,h,i]
.4x Most Recent Episode Manic[a,c,f,g,h,i]
.6x Most Recent Episode Mixed[a,c,f,g,h,i]
.5x Most Recent Episode Depressed[a,b,c,d,e,f,g,h,i]
.7 Most Recent Episode Unspecified[g,h,i]
296.89 Bipolar II Disorder[a,b,c,d,e,f,g,h,i]
Specify (current or most recent episode): Hypomanic/Depressed
301.13 Cyclothymic Disorder
296.80 Bipolar Disorder NOS
293.83 Mood Disorder Due to . . . [*Indicate the General Medical Condition*]
Specify type: With Depressive Features/With Major Depressive-like Episode/With Manic Features/With Mixed Features
—.— Substance-Induced Mood Disorder (*refer to Substance-Related Disorders for substance-specific codes*)
Specify type: With Depressive Features/With Manic Features/With Mixed Features
Specify if: With Onset During Intoxication/With Onset During Withdrawal
296.90 Mood Disorder NOS

Anxiety Disorders

300.01 Panic Disorder Without Agoraphobia
300.21 Panic Disorder With Agoraphobia
300.22 Agoraphobia Without History of Panic Disorder
300.29 Specific Phobia
Specify type: Animal Type/Natural Environment Type/Blood-Injection-Injury Type/Situational Type/Other Type
300.23 Social Phobia
Specify if: Generalized
300.3 Obsessive–Compulsive Disorder
Specify if: With Poor Insight
309.81 Posttraumatic Stress Disorder
Specify if: Acute/Chronic
Specify if: With Delayed Onset
308.3 Acute Stress Disorder
300.02 Generalized Anxiety Disorder
293.89 Anxiety Disorder Due to . . . [*Indicate the General Medical Condition*]
Specify if: With Generalized Anxiety/ With Panic Attacks/With Obsessive–Compulsive Symptoms
—.— Substance-Induced Anxiety Disorder (*refer to Substance-Related Disorders for substance-specific codes*)
Specify if: With Generalized Anxiety/ With Panic Attacks/With Obsessive–Compulsive Symptoms/With Phobic Symptoms
Specify if: With Onset During Intoxication/With Onset During Withdrawal
300.00 Anxiety Disorder NOS

Somatoform disorders

300.81 Somatization Disorder

300.82 Undifferentiated Somatoform Disorder
300.11 Conversion Disorder
Specify type: With Motor Symptom or Deficit/With Sensory Symptom or Deficit/With Seizures or Convulsions/With Mixed Presentation
307.xx Pain Disorder
.80 Associated With Psychological Factors
.89 Associated With Both Psychological Factors and a General Medical Condition
Specify if: Acute/Chronic
300.7 Hypochondriasis
Specify if: With Poor Insight
300.7 Body Dysmorphic Disorder
300.82 Somatoform Disorder NOS

Factitious disorders

300.xx Factitious Disorder
.16 With Predominantly Psychological signs and Symptoms
.19 With Predominantly Physical Signs and Symptoms
.19 With Combined Psychological and Physical Signs and Symptoms
300.19 Factitious Disorder NOS

Dissociative disorders

300.12 Dissociative Amnesia
300.13 Dissociative Fugue
300.14 Dissociative Identity Disorder
300.6 Depersonalization Disorder
300.15 Dissociative Disorder NOS

Sexual and gender identity disorders

Sexual dysfunctions

The following specifiers apply to all primary Sexual Dysfunctions:
Lifelong Type/Acquired Type Generalized Type/Situational Type Due to Psychological Factors/Due to Combined Factors

Sexual desire disorders

302.71 Hypoactive Sexual Desire Disorder
302.79 Sexual Aversion Disorder

Sexual arousal disorders

302.72 Female Sexual Arousal Disorder
302.72 Male Erectile Disorder

Orgasmic disorders

302.73 Female Orgasmic Disorder
302.74 Male Orgasmic Disorder
302.75 Premature Ejaculation

Sexual pain disorders

302.76 Dyspareunia (Not Due to a General Medical Condition)
306.51 Vaginismus (Not Due to a General Medical Condition)

Sexual dysfunction due to a general medical condition

625.8 Female Hypoactive Sexual Desire Disorder Due to . . . [*Indicate the General Medical Condition*]
608.89 Male Hypoactive Sexual Desire Disorder Due to . . . [*Indicate the General Medical Condition*]
607.84 Male Erectile Disorder Due to . . . [*Indicate the General Medical Condition*]
625.0 Female Dyspareunia Due to . . . [*Indicate the General Medical Condition*]
608.89 Male Dyspareunia Due to . . . [*Indicate the General Medical Condition*]
625.8 Other Female Sexual Dysfunction Due to . . . [*Indicate the General Medical Condition*]
608.89 Other Male Sexual Dysfunction Due to . . . *Indicate the General Medical Condition*]
—.— Substance-Induced Sexual Dysfunction (*refer to Substance-Related Disorders for substance-specific codes*)
Specify if: With Impaired Desire/With Impaired Arousal/With Impaired Orgasm/With Sexual Pain
Specify if: With Onset During Intoxication
302.70 Sexual Dysfunction NOS

Paraphilias

302.4 Exhibitionism
302.81 Fetishism
302.89 Frotteurism
302.2 Pedophilia
Specify if: Sexually Attracted to Males/Sexually Attracted to Females/Sexually Attracted to Both
Specify if: Limited to Incest
Specify type: Exclusive Type/Nonexclusive Type
302.83 Sexual Masochism
302.84 Sexual Sadism
302.3 Transvestic Fetishism
Specify if: With Gender Dysphoria
302.82 Voyeurism
302.9 Paraphilia NOS

Gender identity disorders

302.xx Gender Identity Disorder
.6 in Children
.85 in Adolescents or Adults
Specify if: Sexually Attracted to Males/Sexually Attracted to Females/Sexually Attracted to Both/Sexually Attracted to Neither
302.6 Gender Identity Disorder NOS
302.9 Sexual Disorder NOS

Eating disorders

307.1 Anorexia Nervosa
Specify type: Restricting Type; Binge-Eating/Purging Type
307.51 Bulimia Nervosa
Specify type: Purging Type/Nonpurging Type
307.50 Eating Disorder NOS

Sleep disorders

Primary sleep disorders

Dyssomnias

307.42 Primary Insomnia
307.44 Primary Hypersomnia
Specify if: Recurrent

347.00 Narcolepsy
780.57 Breathing-Related Sleep Disorder
327.3X Circadian Rhythm Sleep Disorders
.31 Delayed Sleep Phase Type
.35 Jet Lag Type
.36 Shift Work Type
.30 Unspecified Type
Type/Unspecified Type
307.47 Dyssomnia NOS

Parasomnias

307.47 Nightmare Disorder
307.46 Sleep Terror Disorder
307.46 Sleepwalking Disorder
307.47 Parasomnia NOS

Sleep disorders related to another mental disorder

327.02 Insomnia Related to . . . [*Indicate the Axis I or Axis II Disorder*]
327.15 Hypersomnia Related to . . . [*Indicate the Axis I or Axis II Disorder*]

Other sleep disorders

780.xx Sleep Disorder Due to . . . [*Indicate the General Medical Condition*]
.52 Insomnia Type
.54 Hypersomnia Type
.59 Parasomnia Type
.59 Mixed Type
—.— Substance-Induced Sleep Disorder (*refer to Substance-Related Disorders for substance-specific codes*)
Specify type: Insomnia Type/Hypersomnia Type/Parasomnia Type/ Mixed Type
Specify if: With Onset During Intoxication/With Onset During Withdrawal

Impulse control disorders not elsewhere classified

312.34 Intermittent Explosive Disorder
312.32 Kleptomania
312.33 Pyromania
312.31 Pathological Gambling
312.39 Trichotillomania
312.30 Impulse-Control Disorder NOS

Adjustment disorders

309.xx Adjustment Disorder
.0 With Depressed Mood
.24 With Anxiety
.28 With Mixed Anxiety and Depressed Mood
.3 With Disturbance of Conduct
.4 With Mixed Disturbance of Emotions and Conduct
.9 Unspecified
Specify if: Acute/Chronic

Personality disorders

Note: *These are coded on Axis II*

301.0 Paranoid Personality Disorder
301.20 Schizoid Personality Disorder
301.22 Schizotypal Personality Disorder
301.7 Antisocial Personality Disorder
301.83 Borderline Personality Disorder
301.50 Histrionic Personality Disorder
301.81 Narcissistic Personality Disorder
301.82 Avoidant Personality Disorder
301.6 Dependent Personality Disorder
301.4 Obsessive–Compulsive Personality Disorder
301.9 Personality Disorder NOS

Other conditions that may be a focus of clinical attention

Psychological factors affecting medical condition

316 . . . [*Specified Psychological Factor*]
Affecting . . . [*Indicate the General Medical Condition*]
Choose name based on nature of factors:
Mental Disorder Affecting Medical Condition
Psychological Symptoms Affecting Medical Condition
Personality Traits or Coping Style Affecting Medical Condition
Maladaptive Health Behaviours Affecting Medical Condition
Stress-Related Physiological Response Affecting Medical Condition
Other or Unspecified Psychological Factors Affecting Medical Condition

Medication-induced movement disorders

332.1 Neuroleptic-Induced Parkinsonism
333.92 Neuroleptic Malignant Syndrome
333.7 Neuroleptic-Induced Acute Dystonia
333.99 Neuroleptic-Induced Acute Akathisia
333.82 Neuroleptic-Induced Tardive Dyskinesia
333.1 Medication-Induced Postural Tremor
333.90 Medication-Induced Movement Disorder NOS

Other medication-induced disorder

995.2 Adverse Effects of Medication NOS

Relational problems

V61.9 Relational Problem Related to a Mental Disorder or General Medical Condition
V61.20 Parent–Child Relational Problem
V61.10 Partner Relational Problem
V61.8 Sibling Relational Problem
V62.81 Relational Problem NOS

Problems related to abuse or neglect

V61.21 Physical Abuse of Child (*code 995.54 if focus of attention is on victim*)
V61.21 Sexual Abuse of Child (*code 995.53 if focus of attention is on victim*)
V61.21 Neglect of Child (*code 995.52 if focus of attention is on victim*)
—.— Physical Abuse of Adult
V61.12 (if by partner)
V62.83 (if by person other than partner) (*code 995.83 if focus of attention is on victim*)

—.— Sexual Abuse of Adult
V61.12 (if by partner)
V62.83 (if by person other than partner) (*code 995.83 if focus of attention is on victim*)

Additional conditions that may be a focus of clinical attention

V15.81 Noncompliance With Treatment
V65.2 Malingering
V71.01 Adult Antisocial Behaviour
V71.02 Child or Adolescent Antisocial Behaviour
V62.89 Borderline Intellectual Functioning

Note: *This is coded on Axis II*

780.93 Age-Related Cognitive Decline
V62.82 Bereavement
V62.3 Academic Problem
V62.2 Occupational Problem
313.82 Identity Problem
V62.89 Religious or Spiritual Problem
V62.4 Acculturation Problem
V62.89 Phase of Life Problem

Additional codes

300.9 Unspecified Mental Disorder (nonpsychotic)
V71.09 No Diagnosis or Condition on Axis I
799.9 Diagnosis or Condition Deferred on Axis I
V71.09 No Diagnosis on Axis II
799.9 Diagnosis Deferred on Axis II

Acknowledgement

ICD-10 section of the chapter prepared by Michael B. First and Harold Alan Pincus is adapted from Delling's original text. The contribution of the author to the original text is gratefully acknowledged.

References

1. Kramer, M. (1988). Historical roots and structural basis of the International Classification of Diseases. In *International classif cation in psychiatry* (eds. Mezzich, J. and von Cranach, M.), pp. 3–29. Cambridge University Press, Cambridge.
2. World Health Organization. (1949). *Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death (ICD-6).* World Health Organization, Geneva, Switzerland.
3. American Psychiatric Association. (1952). *Diagnostic and Statistical Manual of Mental Disorders.* American Psychiatric Association, Washington, DC.
4. World Health Organization. (1967). *Glossary of mental disorders and guide to their classification, for use in conjunction with the International Classification of Diseases* (8th revision). World Health Organization, Geneva, Switzerland.
5. American Psychiatric Association. (1968). *Diagnostic and Statistical Manual of Mental Disorders,* (2nd edn.). Psychiatric Association, Washington, DC.
6. Feighner, J., Robins, E., Guze, S.B., *et al.* (1972). Diagnostic criteria for use in psychiatric research. *Archive of General Psychiatry,* **26**(1), 57–63.
7. Spitzer, R.L., Endicott, J. and Robins, E. (1978). Research diagnostic criteria: rationale and reliability. *Archives of General Psychiatry,* **35**(6), 773–82.
8. World Health Organization. *(1978) Mental disorders: glossary and guide to their classification, for use in conjunction with the Ninth Revision of the International Classification of Diseases.* World Health Organization, Geneva, Switzerland
9. American Psychiatric Association. (1980). *Diagnostic and Statistical Manual of Mental Disorders,* (3rd edn.) Psychiatric Association, Washington, DC.
10. Spitzer, R., Forman, J. and Nee, J. (1979) DSM-III field trials: I. initial interrator diagnostic reliabiltiy. *American Journal of Psychiatry,* **136,** 815–7.
11. Leckman, J., Weissman, M., Merikangas, K., *et al.* (1983). Panic disorder and major depression. Increased risk of depression, alcoholism, panic, and phobic disorders in families of depressed probands with panic disorder. *Archives of General Psychiatry,* **40**(10), 1055–60.
12. American Psychiatric Association. (1987). *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn. – Revised) *(DSM-III-R).* American Psychiatric Association, Washington, DC.
13. World Health Organization. (1992). *International Statistical Classification of Diseases and Related Health Problems,* (10th Revision) *(ICD-10).* World Health Organization, Geneva
14. World Health Organization. (1992). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines.* World Health Organization, Geneva
15. World Health Organization. (1993). *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research.* World Health Organization, Geneva.
16. Sartorius, N. (1990). Classifications in the field of mental health. World Health. *Statistics Quarterly,* **43**(4), 269–72.
17. World Health Organization. (1996). *Diagnostic and management guidelines for mental disorders in primary care: ICD-10* Chapter V Primary Care Version. Göttingen.: WHO-Hogrefe and Huber.
18. Sartorius, N., Kaelber, C., Cooper, J., *et al.* (1993). Progress toward achieving a common language in psychiatry. Results from the field trial of the clinical guidelines accompanying the WHO classification of mental and behavioural disorders in ICD-10. *Archives of General Psychiatry,* **50**(2), 115–24.
19. Sartorius, N., Ustun, T., Korten, A., *et al.* (1995). Progress toward achieving a common language in psychiatry, II: Results from the international field trials of the ICD-10 diagnostic criteria for research for mental and behavioural disorders. *American Journal of Psychiatry,* **152**(10), 1427–37.
20. Frances, A., Widiger, T. and Pincus, H. (1989). The development of DSM-IV. *Archive of General Psychiatry,* **46**, 373–5.
21. Widiger, T., Frances, A., Pincus, H., *et al.* (1991). Toward an empirical classification for the DSM-IV. *Journal of Abnormal Psychology,* **100**(3), 280–8.
22. Widiger, T., Frances, A., Pincus, H. *et al.* (eds.) (1994). *DSM-IV Sourcebook Vol. 1.* American Psychiatric Association, Washington, DC.
23. Widiger, T., Frances, A., Pincus, H., *et al.* (eds.) (1996). *DSM-IV Sourcebook Vol. 2.* American Psychiatric Association, Washington DC.
24. Widiger, T., Frances, A., Pincus, H., *et al.* (eds.) (1997). *DSM-IV Sourcebook Vol. 3.* American Psychiatric Association, Washington, DC.
25. Widiger, T., Frances, A., Pincus, H., *et al.* (eds.) (1998). *DSM-IV Sourcebook Vol. 4.* American Psychiatric Association, Washington, DC.
26. Kendell, R. (1991). The relationship between DSM-IV and ICD-10. *Journal of Abnormal Psychology,* **100**(3), 297–301.
27. First, M. and Pincus, H. (1999). Classification in psychiatry: ICD-10 vs. DSM-IV. A response. *British Journal of Psychiatry,* **175**, 205–9.
28. First, M. and Pincus, H. (2002) The DSM-IV Text Revision: rationale and potential impact on clinical practice. *Psychiatric Services*, **53**(3), 288–92.
29. World Health Organization. (2001). *International Classification of Functioning, Disability, and Health (ICF).* World Health Organization, Geneva.
30. Rutter, M., Shaffer, D. and Sturge, C. (1975). *A multi-axial classification of child psychiatric disorders.* World Health Organization, Geneva.

31. Janca, A., Kastrup, M., Katschnig, H., *et al.* (1996). The World Health Organization Short Disability Assessment Schedule (WHO DAS-S): a tool for the assessment of difficulties in selected areas of functioning of patients with mental disorders. *Social Psychiatry and Psychiatric Epidemiology,* **31**(6), 349–54.
32. Spitzer, R., First, M., Williams, J., *et al.* (1992). Now is the time to retire the term 'organic mental disorders'. *American Journal of Psychiatry,* **149**(2), 240–4.
33. Hollander, E., Kim, S. and Zohar, J. (2007). OCSDs in the forthcoming DSM-V. *CNS Spectrums,* **12**(5), 320–3.
34. Zimmerman, M. (1988). Why are we rushing to publish DSM-IV? *Archive of General Psychiatry,* **45**(12), 1135–8.
35. Charney, D., Barlow, D., Botteron, K., *et al.* (2002). Neuroscience research agenda to guide development of a pathophysiologically based classification system. In *A research agenda for DSM-V* (ed. Regier D.), pp. 31–84. American Psychiatric Association, Washington, DC.
36. Kupfer, D., First, M. and Regier, D. (eds.) (2002). *A Research Agenda for DSM-V.* American Psychiatric Publishing, Washington, DC.
37. Narrow, W., First, M., Sirovatka, P., *et al.* (2007). *Age and Gender Considerations in Psychiatric Diagnosis: A Research Agenda for DSM-V.* Arlington, American Psychiatric Association, VA
38. Widiger, T. and Simonsen, E. (2005). Alternative Dimensional Models of Personality Disorder: finding a common ground. *Journal of Personality Disorders*, **19**(2), 110–30.
39. Saunders, J., Schuckit, M., Sirovatka, P., *et al.* (eds.) (2007). *Diagnostic Issues in Substance Use Disorders Refining the Research Agenda for DSM-V.* Amerincan Psychiatric Association, Arlington, VA.
40. Sunderland, T., Jeste, D., Baiyewu, O., *et al.* (eds.) (2007). *Diagnostic Issues in Dementia Advancing the Research Agenda for DSM-V.* American Psychiatric Association, Arlington, VA
41. Allardycce, J., Gaebel, W., Zielasek, J., *et al.* (2007). Deconstructing Psychosis Conference February 2006: the validity of schizophrenia and alternatiuve approaches to the classification of psychosis. *Schizophrenia Bulletin,* **33**(4), 863–7.
42. Helzer, J., Kraemer, H., Krueger, R., *et al.* (eds.) (2008). *Dimensional Approaches in Diagnostic Classification: Refining the Research Agenda for DSM-V.* American Psychiatric Association, Arlington, VA.
43. American Sleep Disorders Association. (2005). *Diagnostic Classification Steering Committee. International Classification of Sleep Disorders: Diagnostic and Coding Manual.* American Academy of Sleep Medicine, Westchester, IL.
44. Andreasen, N., Olsen, S., Dennert, J., *et al.* (1982). Ventricular enlargement in schizophrenia: relationship to positive and negative symptoms. *American Journal of Psychiatry,* **139,** 297–302.
45. Widiger, T., Simonsen, E., Krueger, R., *et al.* (2005). Personality disorder research agenda for the DSM-V. *Journal of Personality Disorder,* **19**(3), 315–38.
46. Helzer, J., Van den Brink, W. and Guth, S. (2006). Should there be both categorical and dimensional criteria for the substance use disorders in DSM-V. *Addiction,* **101**(suppl. 1), 17–22.
47. Krueger, R., Watson, D. and Barlow, D. (2005). Introduction to the Special Section: toward a dimensionally-based taxonomy of psychopathology. *Journal of Abnormal Psychology,* **114**, 491–3.
48. Cloninger, C. (1998). A new conceptual paradigm from genetics and psychobiology for the science of mental health. *Australian and New Zealand Journal of Psychiatry,* **33,** 174–86.
49. Lenzenweger, M. (1999). Schizophrenia: refining the phenotype, resolving the endophenotype. *Behaviour and Research Therapy,* **37**(3), 281–95.
50. Livesley, W. and Jang, K. (2000). Toward an empirically-based classification of personality disorder. *Journal of Personality Disorder,* **14,** 137–51.
51. Meyer, R. (2001). Finding paradigms for the future of alcoholism research: an interdisciplinary perspective. *Alcohol, Clinical and Experimental Research,* **25**(9), 1393–406.
52. Peralta, V., Cuesta, M., Giraldo, C., *et al.* (2002). Classifying psychotic disorder: issues regarding categorical vs. dimensional approaches and time frame to assess symptoms. *European Archives of Psychiatry and Clinical Neuroscience,* **252**, 12–18.
53. Serreti, A., Macciardi, F. and Smeraldi, E. (1996). Identification of symptomatologic patterns concern in major psychoses: proposal for a phenotype definition. *American Journal of Medical Genetics,* **67**(4), 393–400.
54. Van Os, J., Gilvarry, C., Bale, E., *et al.* (1999). A comparison of the utility of dimensional and categorical representations of psychosis. *Psychological Medicine*, **29**, 595–606.
55. Clark, L. (2005). Temperament as a Unifying Concept in the Study of Personality and Psychopathology. *Journal of Abnormal Psychology,* **114**(4), 505–21.
56. Goldberg, D. (1996). A dimensional model for common mental disorders. *British Journal of Psychiatry,* **168**(suppl. 30), 44–9.
57. Krueger, R. and Piasecki, T. (2002). Toward a dimensional and psychometricallyinformed approach to conceptualizing psychopathology. *Behaviour and Research Therapy,* **40**(5), 485–99.
58. Widiger, T. and Samuel, D. (2005). Diagnostic categories or dimensions: a question for the diagnostic and statistical manual of mental disorders (5th edn.). *Journal of Abnormal Psychiatry,* **114**(4), 494–504.
59. First, M. (2005). Clinical Utility: A Prerequisite for the Adoption of a Dimensional Approach in DSM. *Journal of Abnormal Psychiatry,* **114**(4), 560–4.
60. Kraemer, H., Noda, A. and O'Hara, R. (2004). Categorical versus dimensional approaches to diagnosis: methodological challenges. *Journal of Psychiatric Research,* **38,** 17–25.

1.10

From science to practice

John R. Geddes

The difficulties in keeping up to date

Clinicians need accurate and up-to-date information about emerging knowledge on assessment and treatment as well as other developments in practice. The presentation of this knowledge needs to be timely, accurate, and unbiased. In an ideal world, every psychiatrist would have instantaneuous access to the original scientific articles. As this is not feasible because clinicians are busy and the skills needed for an adequate systematic search, critical appraisal, and interpretation of research articles are not routinely available. Further, the volume of research articles is staggering: about 2 million papers are published in 20 000 biomedical journals every year,[1] and even if a psychiatrist restricted her reading to those clinical psychiatry journals it would be necessary to read about 5500 papers each year—equivalent to 15 papers every day.[2] Clearly, a strategy is required for efficient and timely identification of research that is both methodologically sound and clinically relevant.

Traditionally, clinicians have used a number of methods of keeping up to date with research, including consulting colleagues and reading textbooks and journals. Smith[3] reviewed the research on the information needs of doctors and rated sources of information on several dimensions: their relevance to clinical practice, their scientific validity, how easy they were to use, and an overall estimate of their usefulness. Most of the sources that scored highly on all dimensions (such as regularly updated evidence-based textbooks) were of limited availability. Traditional methods of obtaining information (such as conventional textbooks and lecture-based continuing medical education) were more widely available, but of limited validity.

The difficulty in accessing reliable information means that many clinical decisions are made with a greater degree of uncertainty than is necessary. The gap between research and clinical practice is often filled by an unsystematic combination of beliefs, opinions, and clinical experience, which inevitably leads to unnecessary variations in clinical practice. These have been widely documented in psychiatry and include variations in the use of electroconvulsive therapy,[4,5] the use of antipsychotics,[6,7] and the treatment of depression.[8–10] The existence of these variations can only mean that some patients are not receiving the optimum treatment.

Methods of improving use of best available evidence

A coherent set of strategies designed as a clinical tool to link the best available evidence directly to the care of individual patients was first formulated at McMaster University in Canada—an approach called **evidence-based medicine**.[11] Evidence-based medicine is problem-based and splits the process of linking research to practice into five stages (Fig. 1.10.1) plus the identification of clinical questions in need of more research.

To make evidence-based practice feasible in real-life clinical practice, a number of problems need to be solved at each stage of the process.

Formulating a structured clinical question

When uncertainty arises in clinical practice, the clinician needs to formulate a structured clinical question. This step is fundamental to the process of evidence-based medicine because it allows the

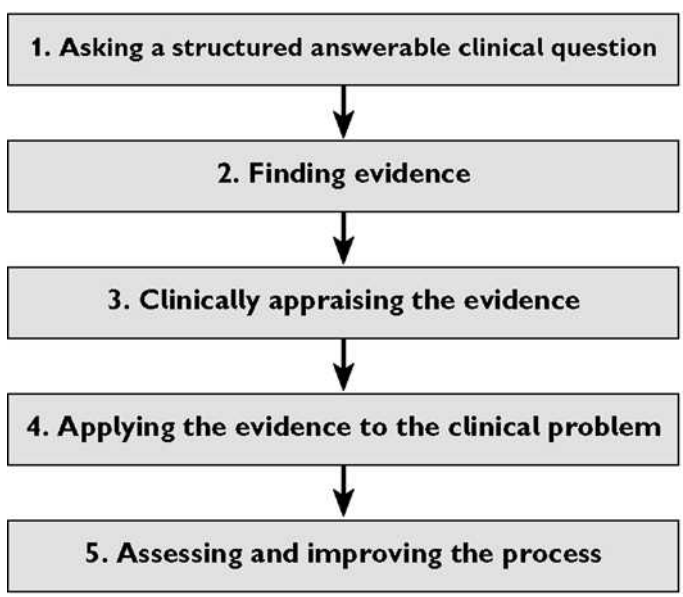


Fig. 1.10.1 The five stages of evidence-based medicine.

clinician first to classify the question, second to identify the research architecture that is most likely to yield a reliable result, and finally to determine the most efficient way of looking for the most reliable research.

(a) Example

Consider a patient who has suffered from two episodes of major depressive disorder both of which have caused substantial functional impairment. On each occasion his symptoms have responded to treatment with a selective serotonin reuptake inhibitor. Following remission of symptoms, the patient has been advised to continue treatment for 6 months before gradually discontinuing the drugs. His psychiatrist is now considering whether or not to advise long-term treatment with antidepressant medication to reduce the risk of relapse. The patient wants to know the risk of relapse without treatment and how much this would be reduced by continuing the drugs. The process of rapidly finding the best answer begins by formulating a clinical question:

1 in patients with major depressive disorder who have responded to drug treatment (the **problem**)

2 how effective are antidepressants (the **intervention**)

3 compared with alternative treatments (including none) (the **comparison intervention**)

4 in preventing relapse (the **outcome**)?

The next step is to classify the question. This example clearly concerns a question about therapy. Most of the questions that arise in clinical practice concern therapy, diagnosis, prognosis, or aetiology. Once the question has been formulated and classified, this suggests the most reliable research architecture (Table 1.10.1)

Finding evidence and advances in the organization of clinical knowledge

Identification of the nature of the clinical question and the most reliable study design enables the clinician to do a focused and efficient literature search. One of the main advances of evidence-based medicine has been the development of methods of research synthesis, or the process of identifying, appraising, and summarizing primary research studies into clinically usable knowledge. There are two main approaches to research synthesis—systematic reviews and clinical practice guidelines. Both these approaches are based on an explicit methodology that begins with the construction of a hierarchy of evidence in which certain forms of research architecture are considered to be reliable than others. The methodology is most clearly developed for questions about therapy and these will be the focus here.

(b) Levels of evidence

A commonly used hierarchy of evidence for studies of treatments is as follows:

Ia Evidence from a systematic review of randomized controlled trials,

Ib Evidence from at least one randomized controlled trial,

IIa Evidence from at least one controlled study without randomization,

IIb Evidence from at least one other type of quasi-experimental study,

Table 1.10.1 Types of clinical question and most reliable study architecture

Type of question	Form of the question	Most reliable study architecture
Diagnosis	How likely is a patient who has a particular symptom, sign, or diagnostic test result to have a specific disorder?	A cross-sectional study of patients suspected of having the disorder comparing the proportion of the patients who really have the disorder who have a positive test result with the proportion of patients who do not have the disorder who have a positive test result.
Treatment	Is the treatment of interest more effective in producing a desired outcome than an alternative treatment (including no treatment)?	Randomized evidence in which the patients are randomly allocated to receive either the treatment of interest or the alternative: this is usually a systematic review of RCTs or a single high-quality RCT.
Prognosis	What is the probability of a specific outcome in this patient?	A study in which an inception cohort (patients at a common stage in the development of the illness—especially first onset) are followed up for an adequate length of time.
Aetiology	What has caused the disorder?	A study comparing the frequency of an exposure in a group of persons with the disease (cases) of interest with a group of persons without the disease (controls)—this may be an RCT, a case-control study. or a cohort study.

RCT, randomized controlled trial.

III Evidence from non-experimental descriptive studies, such as comparative studies, correlation studies, and case-control studies,

IV Evidence from expert committee reports or opinions and/or clinical experience of respected authorities.

In this hierarchy, randomized evidence is considered, on average, to be more reliable thsan non-randomized evidence, and a systematic review of randomized evidence is considered to be the best defence against systematic bias.

Hierarchies of evidence have also been formulated for non-therapeutic studies, such as studies of aetiology, diagnosis, and prognosis. Again, the fundamental feature of these hierarchies is that the study architectures with the least susceptibility to bias are considered most reliable. The study design considered most reliable for each type of clinical question is shown in Table 1.10.1.

(c) Systematic reviews

The need for systematic reviews and the methodology used are described in Chapter 6.1.1.2. The recognition of the need for systematic reviews of randomized controlled trials, and the development of the scientific methodology of reviews, has been one of the most striking advances in health services research over the last decade. One key development was the founding of the Cochrane Collaboration, an international organization with the objective of

producing regularly updated systematic reviews of the effectiveness of all health care interventions.[12]

(d) Clinical practice guidelines

In some areas of health care there is sufficient evidence, coexisting with substantial clinical uncertainty, that it is worth developing clinical practice guidelines. Clinical practice guidelines have been defined as 'systematically developed statements to assist practitioner decisions about appropriate health care for specific clinical circumstances'.[13] Guidelines have been developed for several years, but there have been recent advances in the methodology of producing explicitly evidence-based guidelines. Evidence-based clinical practice guidelines are developed by a guideline development group consisting of key stakeholders who decide on the precise clinical questions to be answered. The evidence is then systematically reviewed and classified according to a hierarchy of evidence (see above) and presented to the guideline development group. The group then makes recommendations as appropriate. The degree to which the recommendations are directly based on the evidence is described using a second level of classification[14]:

1 directly based on category I evidence;

2 directly based on category II evidence or extrapolated recommendation from category I evidence;

3 directly based on category III evidence or extrapolated recommendation from category I or II evidence;

4 directly based on category IV evidence or extrapolated recommendation from category I, II, or III evidence.

Clinical practice guidelines are usually developed at a national level and need tailoring to suit local circumstances. Professional and scientific bodies such as the American Psychiatric Association and the British Association for Psychopharmacology often take the lead in developing national guidelines. Increasingly, health care providing organizations are developing clinical practice guidelines as a way of assuring quality, increasing standardization of care and controlling costs. For example, in the United Kingdom, the National Institute for Health and Clinical Excellence (NICE) is now the main body producing guidelines across all disease areas. These guidelines are extremely rigorous in terms of methodology and also routinely include economic analyses of the cost-effectiveness of health care technologies. At their best, these guidelines produce the most accurate syntheses of current knowledge available. NICE also produces Health Technology Appraisals (HTAs) of single interventions to determine the cost-effectiveness of new technologies (mainly medicines) prior to their introduction into the taxpayer-funded National Health Service. Cost-effectiveness analysis requires the translation of disease-specific estimates of clinical effectiveness into the common metric of Quality Adjusted Life Years (QALYs) using sophisticated modelling techniques. Decisions about whether to allow reimbursement of the treatment depends on the cost per QALY (incremental cost-effectiveness ratio, ICER): a treatment with an ICER of more than £30 000 is unlikely to be approved. NICEs HTA decisions are particularly likely to be controversial when there is some evidence that the treatment works, but that the ICER is found to be too high—for example, in the case of acetylcholinesterase inhibitors in Alzheimer's disease.

There are several limitations to clinical practice guidelines. Firstly, evidence-based clinical practice guidelines are expensive and time consuming to produce and rapidly become out of date. Secondly, to influence practice, evidence-based clinical practice guidelines need to be actively disseminated and implemented. Guidelines that are developed nationally and passively sent out to doctors are often not used.[15] A number of active approaches are effective in helping change clinicians' behaviour:[16]

- outreach visits (also known as academic detailing)
- local opinion leaders
- patient-mediated interventions (including patient education)
- multifaceted interventions involving a range of techniques.

There is some evidence that guidelines can improve patient outcomes by their effect on clinical practice, especially when they are made relevant to local circumstances.[15] In one study in the United States, 217 patients with depressive disorders were randomly assigned to usual care or a multifaceted intervention designed to achieve the Agency for Health Care Policy and Research guidelines on management of depression.[17] Patients in the intervention group were much more likely to be treated in accordance with the guidelines, and this led to improved outcomes in patients with major depressive disorder (more than 50 per cent reduction on the Symptom Checklist-90 Depressive Symptom Scale at 4 months in 74 per cent of experimental patients compared with 44 per cent of control patients).

Understandably, clinicians also seek guidance in important clinical questions that are poorly served by high-quality, especially randomized, evidence. To assist in these clinical decisions, Frances and his colleagues have developed an innovative method of guideline development based on a systematic survey of the views of clinical experts.[18,19]

(e) Use of electronic communication and the Internet

The development of the Internet—or World Wide Web—during the 1990s facilitated the development of evidence-based practice. The Internet has now become a vast information resource for doctors and patients. Improved access to information afforded to patients means that they are often very well informed about their condition. This is one of the factors contributing to the need for doctors to improve their own access to information. The Internet has several drawbacks including the disorganization of the information and the lack of quality control.[20,21] Web portals have been developed that provide organized and indexed access to critically appraised websites (e.g. www.nelh.nhs.uk). The Web has now become the main medium for transmitting and storing knowledge.

(f) Improving current awareness

Another area of improvement has been the development of tools to assist doctors to maintain their current awareness of advances in research. The idea of review and abstracting journals is not new, but there has been a recognition that such journals also need a methodology to allow them to identify the most reliable and clinically important research studies.

A number of new journals have been produced with the aim of improving the availability of high-quality evidence to clinicians. The first of these was *ACP Journal Club* (targeted primarily at general physicians), followed by *Evidence-Based Medicine* (targeted primarily at family doctors) and, more recently, Evidence-Based Mental Health (aimed at mental health clinicians of all disciplines).

Evidence-Based Mental Health scans over 200 journals regularly and selects only those articles that both meet explicit methodological criteria (see Box 1.10.1) and are clinically important. The articles are then summarized in structured abstracts and published on one page with an accompanying commentary by a clinical expert.

A systematic approach to searching for the best available evidence

The developments in the organization of clinical knowledge make it possible for a clinician to search rapidly and efficiently for current best evidence using a standard approach (Fig. 1.10.2). This approach will change as new methods of organizing knowledge are developed.

(a) Example (continued)

Although a recent clinical guideline has been produced by NICE, it does not include sufficient quantitative information to answer the patient's question.

Box 1.10.1 Examples of the criteria for selection and review of articles for abstracting in Evidence-Based Mental Health

Articles are considered for abstracting if they meet the following criteria

Basic criteria

- Original or review articles
- In English
- About humans
- About topics that are important to the practice of clinicians in the broad field of mental health

Studies of prevention or treatment must meet these additional criteria

- Random allocation of participants to comparison groups
- Follow-up (endpoint assessment) of at least 80% of those entering the investigation
- Outcome measure of known or probable clinical importance
- Analysis consistent with study design

Studies of diagnosis must meet these additional criteria

- Clearly identified comparison groups, at least one of which is free of the disorder or derangement of interest
- Interpretation of diagnostic standard without knowledge of test results
- Interpretation of test without knowledge of diagnostic standard result
- Diagnostic(gold) standard(e.g. diagnosis according to DSM-IV or ICD-10 criteria after assessment by clinically qualified interviewer) preferably with documentation of reproducible criteria for subjectively interpreted diagnostic standard(e.g. report of statistically significant measure of agreement among observers)
- Analysis consistent with study design.

Studies of prognosis must meet additional criteria

- Inception cohort (first onset or assembled at a uniform point in the development of the disease) of individuals, all initially free of the outcome of the interest
- Follow-up of at least 80% of patients until the occurrence of a major study endpoint
- Or to the end of the study
- Analysis consistent with study design.

Studies of causation must meet these additional criteria

- Clearly identified comparison group for those at risk of, or having, the outcome of interest (i.e. randomized controlled trial quasi-randomized controlled trial, non-randomized controlled trial, cohort analytical study with case by-case matching, or statistical adjustment to create comparable groups, case-control study)
- Masking of observers of outcomes to exposures (this criterion is assumed to be met if the outcome is objective), observers of exposures masked to outcomes for case-control studies, or masking of subjects to exposure for all other study designs
- Analysis consistent with study design.

Geddes, J.R., Carney, S.M., Davies, C., Furukawa, T.A., Kupfer, D.J., Frank, E., Goodwin. G.M. Relapse prevention with antidepressant drug treatment in depressive disorders. *Lancet* 2003, **361:** 653–61.[22]

This is a systematic review of randomized controlled trials comparing a number of antidepressants (including tricyclics, selective serotonin reuptake inhibitors, monoamine oxidase inhibitors, and low-dose antipsychotics) with placebo in the prevention of relapse in depressive disorder.

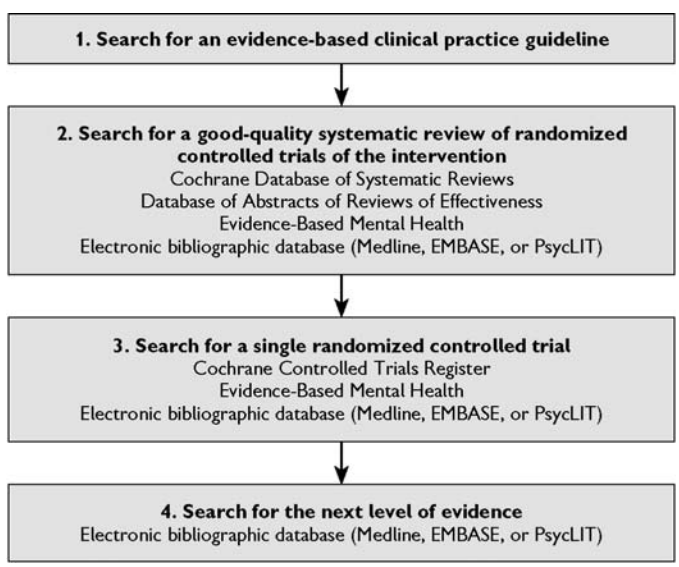


Fig. 1.10.2 A systematic approach to identifying the best evidence about a therapy.

Critical appraisal of research articles

Once the evidence has been found, it needs to be critically appraised for its reliability and usefulness. Psychiatrists need to be able to assess the scientific value and clinical importance of a study. This requires a range of epidemiological and biostatistical skills that have not traditionally been considered to be key skills for psychiatrists. In the United Kingdom, the Royal College of Psychiatrists introduced in 1999 a new part of the main professional examination that is designed to test these skills, recognizing their fundamental importance for clinical psychiatrists.[23]

Structured critical appraisal is an active process that involves a systematic assessment of the key methodological aspects of the paper. In particular, critical appraisal focuses systematically on those aspects of the study methodology that are most likely to lead to unreliability of results. A number of checklists, designed to make the appraisal quicker and more systematic, have been produced for different research study designs.[24] For example, the critical appraisal of a systematic review involves an assessment of those aspects of methodology described in Chapter 6.1.1.2. A commonly used checklist for systematic reviews is shown in Table 1.10.2.

(a) Example (continued)

Using the checklist, the review can be quickly critically appraised. It is a review of the efficacy of antidepressant drugs in preventing recurrence of depression in patients who responded to acute phase therapy with antidepressants and so appears relevant to the clinical question. The authors have only included randomized controlled trials, and this will make systematic error less likely and improve the reliability of the review. The literature search strategy is clearly documented and included electronic databases (Medline, Psyclit, Embase, Lilacs, and the Cochrane Library), hand searching of journals, and correspondence with researchers active in the field and drug companies. The quality of the randomized controlled trials was rated both from the description of the allocation of treatment and by assessing other methodological issues such as whether the primary analysis was done as an intention-to-treat analysis and the degree of blinding of the clinician and patient. It can be concluded that the reviewers have made a reasonable effort to identify the primary studies, although it is possible that other studies, perhaps with negative results, have not been published (publication bias, see Chapter 6.1.1.2).

The results of the primary studies are shown graphically as odds ratios (see Glossary) in Fig. 1.10.3. Odds ratios falling to the left of the vertical line indicate that the outcome (relapse of depressive symptoms) occurred less frequently in patients who continued treatment with an antidepressant. The smaller the odd ratio, the larger the treatment effect found in that particular study. For each study, the central diamond represents the most likely value of the relative risk and the box around the relative risk shows the 95 per cent confidence interval (**CI**). The larger studies (e.g. **Rouillon *et al.*[25]) have narrower confidence intervals because they are larger and therefore have less random error and greater precision.

From the figure, it can be seen that, the odds ratio of the all the studies included fall on the left-hand side of the line and therefore found the same direction of treatment effect. There is considerable overlap in the confidence intervals from study to study and, although there is some statistically significant heterogeneity between them (see Chapter 6.1.1.2), this appears to be in the size—rather than the nature—of the positive treatment effect. At the bottom of the figure is the estimate of the combined treatment effect of the trials. This is expressed as a pooled odds ratio. Combining the study results produces a more precise estimate of the drug's relative effectiveness, with tighter confidence intervals. The overall pooled odds ratio of relapse for patients taking antidepressants compared with placebo is 0.30 (95 per cent CI (see Glossary), 0.22–0.38).

Table 1.10.2 Checklist to assist the critical appraisal of a systematic review

Validity
1. Did the review address a clearly focused clinical question? Did the review describe: the population studied? the intervention given? the outcomes considered?
2. Did the authors select the right sort of studies for the review? The right studies would: address the review's question have an adequate study design (e.g. for a question re therapy, an RCT)
3. Were the important relevant studies included in the review? Which bibliographic databases were used? Personal contact with experts Search for unpublished as well as published studies Search for non-English language studies
4. Did the review's authors assess the quality of the included studies? Did they use: description of randomization? a rating scale?
Results
5. Were the results similar from study to study? Are the results of all the included studies clearly displayed? Are the results from different studies similar? If not, are the reasons for variations between studies discussed?
6. What is the overall result of the review? Is there a clear clinical conclusion (a clinical bottom-line)? What is it? What is the numerical result?
7. How precise are the results? Is there a confidence interval?
Clinical relevance of the results
8. Can I apply the results to my patient? Is this patient so different from those in the trial that the results do not apply?
9. Should I apply the results to my patient? How great would the benefit of therapy be for this particular patient? Is the intervention consistent with my patient's values and preferences? Were all the clinically important outcomes considered? Are the benefits worth the harms and costs?

RCT, randomized controlled trial.

Methods of using research findings at the level of individual patients

After the study has been critically appraised for its validity, the clinician needs to determine what the results are, and their importance for the patient. Patients in research studies are always different from those in real-life clinical practice in ways that may be difficult to determine even only if because they–or at least the

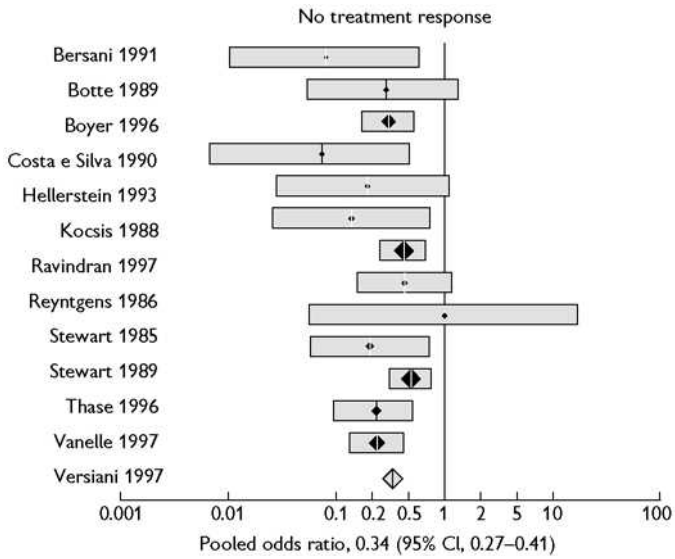


Fig. 1.10.3 Antidepressant versus placebo for the prevention of relapse in depressive disorder. (Reprinted from the *Lancet*, **361**, Geddes, J.R., Carney, S.M., Davies, C., *et al.* (2003), Relapse prevention with antidepressant drug treatment in depressive disorders, 653–61 Copyright 2003, with permission from Elsevier).[22]

episode of illness–was in the past rather than the present. Therefore, the use of results from research studies in clinical practice should be cautious and always requires a degree of extrapolation. The contribution of clinical epidemiology is in developing methods of applying research results to individual patients that are biologically and statistically robust and are explicit about any assumptions made. One of the most useful questions to ask when applying the results of a research study is: 'Is my patient so different from those in the study that the results cannot be used'? The next step is to try to interpret the study results for a particular patient, in terms of his or her clinical characteristics and treatment preferences.

An alternative measure of treatment effect which many find more clinically useful than the relative risk or odds ratio is the number needed to treat (**NNT**).[26,27] The NNT is an estimate of the number of patients that would need to be treated with the intervention of interest, compared with the alternative, in order to achieve one good outcome or to avoid one harmful outcome. The NNT is calculated by taking the reciprocal of the difference between the rates of the outcome of interest in the experimental and control groups.

(a) Example (continued)

Although the review supports the general conclusion that continued therapy with the same antidepressant that was effective in acute phase therapy is very much likely to decrease the risk of relapse, a clinician may wish to check the evidence for a specific drug. For example, how strong is the evidence for reboxetine? The systematic review identifies one trial that directly answers this question[28]. This was a multicentre trial performed in centres in Europe and South America. The trial recruited 358 outpatients who met DSM-IIIR criteria for an acute recurrence of major depressive disorder to open-label reboxetine and randomized the 285 patients who achieved >50 per cent reduction in baseline depressive symptoms on the Hamilton Depression Rating Scale to either continued reboxetine or placebo (mean age reboxetine 43.4 years, placebo 42.3 years, women 79.3 per cent reboxetine: 67.4 per cent men). One group of 145 patients were allocated to reboxetine, 4 mg b.d. and 141 patients were allocated to placebo. Treatment was continued for a further 46 weeks and the primary outcome was relapse (increase in the Hamilton Depression Rating Scale of 50 per cent or more and/or a total score of ≥18 points).

(i) Appraising the validity of a randomized controlled trial

The appraisal of an individual randomized controlled trial also addresses issues of internal and external validity. The main difference compared with the appraisal of a systematic review is that the reader can appraise the trial directly, rather than relying on the author of the review to have made an adequate assessment of the quality of the studies included. The most important questions to answer during the appraisal of the article are those relating to methodological issues that have been randomly shown to affect the reliability of the results (Table 1.10.3).

This study[28] was a **randomized controlled trial**. The method of randomization is not stated and no details are provided about the **concealment of treatment allocation**. Concealment of allocation is one of the most important features to appraise in a randomized controlled trial, and refers to how well the treatment allocation of the next patient was concealed from the participating clinician.[29] If a clinician has definite knowledge of, or can reasonably predict, the next allocation, he may decide not to enter a patient if he favours either treatment in that specific case. This would obviously lead to a subversion of randomization and biased results. This is why methods of quasi-randomization, such as alternate allocation, are often unsatisfactory. The most satisfactory methods of random allocation are when allocation is performed by a third party following entry of the patient into the study, for example using a centralized telephone service. Concealment of allocation should be distinguished from **blinding** of treatment allocation. Blinding refers to whether the patient (**single blind**), or both the patient and the clinician (**double blind**), are kept unaware of which arm of the study the patient has been allocated to *following* randomization. Blinding protects against bias in the treatment during the trial and in the assessment of subjective outcomes, but can be difficult to maintain when the experimental treatment has characteristic side-effects (e.g. antidepressants). This study was reported to be double blind; however, it is difficult to tell how effective the blinding was and whether this affected the treatment that the patients received or the ratings of outcome.

Of the patients who were randomized, 79/145 (54 per cent) of those allocated to reboxetine and 65/141 (46 per cent) of those allocated to placebo completed the trial. The reasons for drop-out are clearly stated and all the patients who entered the study are accounted for. Patients who drop out of a study may be different from those who complete it—this is a particular problem if there

Table 1.10.3 Questions that must be answered during the appraisal of an article

1. Was the assignment of the patients randomized?
2. Was the randomization list concealed?
3. Were all the subjects who entered the trial accounted for at its conclusion?
4. Were they analysed in the groups to which they were randomized?
5. Were subjects and clinicians 'blind' to which treatment was being received?
6. Other than the experimental treatment, were the groups treated equally?
7. Were the groups similar at the start of the trial?

is **differential drop-out** between the arms of the study. For this reason, the most statistically reliable and clinically useful method of analysis is to include all patients who were randomized in the analysis. This is called an **intention-to-treat analysis (ITT)**. In the reboxetine trial[28] the primary analysis was done on an ITT basis (although three patients who dropped out before the first follow-up assessment were excluded from the primary analysis—these were included in the systematic review), and the patients were analysed in the groups to which they were randomized. Lastly, although randomization will avoid bias in treatment allocation, it is possible that, by chance, the groups will be unbalanced on some key prognostic factors such as age, sex, duration, and severity of illness. Therefore it is important to assess the baseline characteristics of the patients to identify any obvious differences. In Versiani *et al.*[28] the patients were reasonably similar on baseline characteristics at entry into the trial, although the proportion of females was a little higher in the reboxetine group.

(ii) Are the results relevant for your patient?

To determine the relevance of the study to real-life patients, it is important to examine the inclusion and exclusion criteria of the trial. The main inclusion criteria are discussed above. Patients excluded from the trial were those with coexisting psychotic features or chronic depression, those with a first episode at the time of screening and patients with a history of seizures, serious brain injury, clinically significant haemopoietic or cardiovascular disease, urinary retention, or glaucoma. The user of the study results will have to take these inclusion and exclusion criteria into account, and the clinician needs to judge the relevance of the results for the individual patient.

(iii) What are the results?

Of the 145 patients treated with reboxetine, 29 relapsed, giving an experimental event rate (**EER**), of 20 per cent or a probability of relapse of 0.2. In the placebo-treated group 73 of 141 patients relapsed giving a control event rate (**CER**) of 52 per cent. Therefore the absolute difference between the rates was

$$\text{EER} - \text{CER} = 32 \text{ per cent}$$

and the difference between the probabilities is 0.32. This means that for every 100 patients treated with reboxetine, compared with placebo, 32 fewer relapsed. Therefore, to prevent one relapse over 46 weeks, 100/32, or about three patients, would need to be treated (NNT) with reboxetine.

(b) Interpretation of numbers needed to treat

The clinical interpretation of NNT depends on the seriousness of the outcome and the nature (and cost) of the intervention.

For example, if the number needed to treat with aspirin following acute stroke to avoid one death in the short-term is 100, this seems a very useful intervention because death is such a serious outcome that it is worth treating a lot of patients to save a few from dying—especially as aspirin is very cheap. Some examples of NNTs are given in Table 1.10.4.

Table 1.10.4 Examples of numbers needed to treat for interventions in psychiatry assessment and report

Intervention	Outcome	NNT	95% CI	Reference
Cognitive therapy in bulimia nervosa	Long-term remission	2		30
Antidepressants in dysthymic disorder	Clinical response	5	3–10	31
Family therapy in schizophrenia	Relapse at 1 year	**7**	4–14	32
SSRIs compared with TCAs in depressive disorder	Remain in treatment at 6 weeks	39	20–426	28

SSRI, selective serotonin reuptake inhibitor, TCA tricyclic antidepressant.

Critical appraisal of other research designs

The approach taken to the appraisal of research designs applied to other clinical questions is similar to that outlined above. Rather than passively reading the article or abstract, the clinician actively searches out the most important methodological features to determine the reliability of the study. By applying the methods developed by clinical epidemiologists, the results presented in the paper can be used to calculate more clinically meaningful measures. These can, in turn, be tailored to suit the characteristics of the individual patient. The clinician needs to develop a practical knowledge of the tools to allow him to use them routinely and quickly in clinical practice and also to develop sufficient familiarity to be able describe the results of studies to colleagues and patients.

Glossary

- **Absolute risk reduction (risk difference)** is the absolute arithmetic difference in the risk of the adverse outcome between control group (CER) and experimental group (EER). When an intervention increases the probability of a beneficial outcome it is known as the absolute benefit increase (ABI).
- **Confidence interval (CI)** is the range within which the true value of a statistical measure can be expected to lie. The CI is usually accompanied by a percentage value (usually 95 per cent) which shows the level of confidence that the true value lies within this range.
- **Event rate** is the proportion of patients in a group in whom the event is observed. In control patients, this is called the control event rate (CER) and in experimental patients it is called the experimental event rate (EER). The patient expected event rate (PEER) refers to the rate of events that would be expected in a patient who received no treatment or conventional treatment.
- **Number needed to treat (NNT)** is the reciprocal of the absolute risk reduction and is the number of patients that need to be treated to prevent one bad outcome or to achieve one beneficial outcome.
- **Odds ratio** is the odds of the outcome in the experimental group divided by the odds of the outcome in the control group. The odds ratio is often reported in meta-analyses because of its useful statistical properties.
- **Relative risk** is the risk of the outcome in the experimental group divide by the risk of the outcome in the control group. The risk ratio is increasingly reported in meta-analyses because it is easier to interpret clinically than the odds ratio.

Further information

Straus S.E., Richardson, S., Rosenberg, W., *et al.* (2001). *Evidence-based medicine: how to practise and teach EBM* (2nd edn). Churchill Livingstone, London.

Geddes, J.R. and Harrison, P.J. (1997). Evidence-based psychiatry: closing the gap between research and practice. *British Journal of Psychiatry*, **176**, 220–5.

References

1. Mulrow, C.D. (1994). Rationale for systematic reviews. *British Medical Journal*, **309**, 597–9.
2. Geddes, J.R., Wilczynski, N., Reynolds, S., *et al.* (1999). Evidence-based mental health—the first year. *Evidence-Based Mental Health*, **2**, 3–5.
3. Smith, R. (1996). What clinical information do doctors need? *British Medical Journal*, **313**, 1062–8.
4. Pippard, J. (1992). Audit of electroconvulsive treatment in two national health service regions. *The British Journal of Psychiatry*, **160**, 621–37.
5. Hermann, R.C., Dorwart, R.A., Hoover, C.W., *et al.* (1995). Variation in ECT use in the United States. *The American Journal of Psychiatry*, **152**, 869–75.
6. Meise, U., Kurz, M., and Fleischhacker, W.W. (1994). Antipsychotic maintenance treatment of schizophrenia patients: is there a consensus? *Schizophrenia Bulletin*, **20**, 215–25.
7. Lehman, A.F. and Steinwachs, D.M. (1998). Patterns of usual care for schizophrenia: initial results from the Schizophrenia Patient Outcomes Research Team (PORT) Client Survey. *Schizophrenia Bulletin*, **24**, 11–20.
8. Wells, K.B., Katon, W., Rogers, B., *et al.* (1994). Use of minor tranquilizers and antidepressant medications by depressed outpatients: results from the medical outcomes study. *The American Journal of Psychiatry*, **151**, 694–700.
9. Hirschfeld, R.M., Keller, M.B., Panico, S., *et al.* (1997). The National Depressive and Manic-Depressive Association consensus statement on the undertreatment of depression. *Journal of the American Medical Association*, **277**, 333–40.
10. Munizza, C., Tibaldi, G., Bollini, P., *et al.* (1995). Prescription pattern of antidepressants in out-patient psychiatric practice. *Psychological Medicine*, **25**, 771–8.
11. Evidence-Based Medicine Working Group (1992). Evidence-based medicine. A new approach to teaching the practice of medicine. *Journal of the American Medical Association*, **268**, 2420–5.
12. Chalmers, I., Dickersin, K., and Chalmers, T.C. (1992). Getting to grips with Archie Cochrane's agenda. *British Medical Journal*, **305**, 786–8.
13. Field, M.J. and Lohr, K.N. (1990). *Clinical practice guidelines: direction of a new agency*. Institute of Medicine, Washington, DC.
14. Eccles, M., Freemantle, N., and Mason, J. (1988). North of England evidence based guidelines development project: methods of developing guidelines for efficient drug use in primary care. *British Medical Journal*, **316**, 1232–5.
15. Grimshaw, J.M. and Russell, I.T. (1993). Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet*, **342**, 1317–22.
16. Oxman, A.D., Thomson, M.A., Davis, D.A., *et al.* (1995). No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *Canadian Medical Association Journal*, **153**, 1423–31.
17. Katon, W., Von Korff, M., Lin, E., *et al.* (1995). Collaborative management to achieve treatment guidelines. Impact on depression in primary care. *Journal of the American Medical Association*, **273**, 1026–31.
18. Kahn, D.A., Docherty, J.P., Carpenter, D., *et al.* (1997). Consensus methods in practice guideline development: a review and description of a new method. *Psychopharmacology Bulletin*, **33**, 631–9.
19. Frances, A.J., Docherty, J.P., and Kahn, D.A. (1998). The Expert Consensus Guideline Series: treatment of bipolar disorder. *The Journal of Clinical Psychiatry*, **57**, 1–88.
20. Jadad, A.R. and Gagliardi, A. (1998). Rating health information on the Internet: navigating to knowledge or to Babel? *Journal of the American Medical Association*, **279**, 611–14.
21. Silberg, W.M., Lundberg, G.D., and Musacchio, R.A. (1997). Assessing, controlling, and assuring the quality of medical information on the Internet: caveant lector et viewor–let the reader and viewer beware. *Journal of the American Medical Association*, **277**, 1244–5.
22. Geddes, J.R., Carney, S.M., Davies, C., *et al.* (2003). Relapse prevention with antidepressant drug treatment in depressive disorders. *Lancet*, **361**, 653–61.
23. Critical Review Paper Working Party (1997). MRCPsych Part II examination: proposed critical review paper. *Psychiatric Bulletin*, **21**, 381–2.
24. Sackett, D.L., Richardson, S., Rosenberg, W., *et al. Evidence-based medicine: how to practise and teach EBM*. Churchill Livingstone, London.
25. Rouillon, F., Berdeaux, G., Bisserbe, J.C., *et al.* (2000). Prevention of recurrent depressive episodes with milnacipran: consequences on quality of life. *Journal of Affective Disorders*, **58**, 171–80.
26. Sackett, D.L., Deeks, J.J., and Altman, D.G. (1996). Down with odds ratios! *Evidence-Based Medicine*, **1**, 164–6.
27. Cook, R.J. and Sackett, D.L. (1995). The number needed to treat: a clinically useful measure of treatment effect. *British Medical Journal*, **310**, 452–4.
28. Versiani, M., Mehilane, P., Gaszner, P., *et al.* (1999) Reboxetine, a unique selective NRI, prevents relapse and recurrence in long-term treatment of major depressive disorder. *The Journal of Clinical Psychiatry*, **60**, 400–406.
29. Shulz, K. (2000). Assessing allocation concealment and blinding in randomised controlled trials: why bother? *Evidence-Based Mental Health*, **3**, in press.
30. Fairburn, C.G., Norman, P.A., Welch, S.L., *et al.* A prospective study of outcome in bulimia nervosa and the long-term effects of three psychological treatments. *Archives of General Psychiatry*, **52**, 304–12.
31. Vanelle, J.M., Attar, L.D., Poirier, M.F., Bouhassira, M., Blin, P., and Olie, J.P. (1997). Controlled efficacy study of fluoxetine in dysthymia. *The British Journal of Psychiatry*, **170**, 345–50.
32. Mari, J.J. and Streiner, D. (1998). Family intervention for schizophrenia. In *Cochrane database of systematic reviews* (ed. M. Oakley-Brown), Update Software.
33. Anderson, I.M. and Tomenson, B.M. (1995). Treatment discontinuation with selective serotonin reuptake inhibitors compared with tricyclic antidepressants: a meta-analysis. *British Medical Journal*, **310**, 1433–8.

This page intentionally left blank

SECTION 2

The Scientific Basis of Psychiatric Aetiology

This page intentionally left blank

2.1

Brain and mind

Martin Davies

History of the mind–brain relation

The thesis that the brain, rather than the heart, is the seat of the mind was already widely accepted by the ancient Greeks; but it was not universally accepted—Aristotle was an exception. Many issues in psychiatry resonate with the ancient debates over the roles of the heart and the brain. But a brief review of modern thinking about the mind–brain relation can begin two millennia later with René Descartes, who held that minds are real things of a fundamentally different kind from material bodies.(1)

Dualism: Descartes

Descartes's world-view included bodies or material things, whose essence is to be extended in space, and minds, which are immaterial things whose essence is thinking. According to **Cartesian dualism**, the mind is not literally housed within the body, because spatial properties belong to matter and not to mind. But, when he talked about the way we experience the states of our own body, Descartes sometimes spoke of the mind being 'mixed up with' the body.

Early theories of the brain as the seat of the mind assigned an important role to the ventricles. On Descartes's view, mechanical operations involving the release of animal spirits in the ventricles were adequate to explain animal behaviour but intelligent human action required something more. He postulated that the immaterial mind could modulate processes in the material brain by way of a causal interaction operating through the pineal gland.

The motion of bodies and the completeness of physics

Dualist interactionism is challenged by theories about the motion of bodies. According to Descartes's own theory, quantity of motion (defined as mass times speed) is conserved. Because motion is not a directional notion, this conservation law allowed that the immaterial mind could bend the trajectory of a physical particle in the pineal gland. But Gottfried Leibniz's superior theory, with conservation laws for momentum (a directional notion) and kinetic energy, had the consequence that only impacts with other bodies could cause changes in the direction or speed of physical particles. This left no room for immaterial causes of material changes and, while Leibniz was a dualist, he was not an interactionist dualist but believed, instead, in a **pre-established harmony** between the material and immaterial worlds.

By departing from the idea that impact was the only force on bodies, and allowing action at a distance, Isaac Newton reopened the possibility of distinctively mental forces affecting the trajectory of bodies. These forces were not even ruled out by the law of conservation of energy, which was widely accepted by the middle of the nineteenth century, but advances in biochemistry and neurophysiology during the first half of the twentieth century made appeal to vital and mental forces seem increasingly unmotivated. Since around 1950, the dominant theories of the mind–brain relation have been compatible with a broadly physicalist world-view and with the completeness of physics: physical effects have wholly physical causes.(2)

Behaviourism: Ryle

From the 1920s to the 1950s, particularly in the United States, behaviourism was a dominant approach within psychology. This was not just methodological behaviourism, which is a restriction on the kinds of evidence that can be used, but a radical reconception of psychology as the science of behaviour rather than the science of the mind. In philosophy, **analytical behaviourism** was a doctrine about the meaning of our mental discourse. The idea was to analyse or translate our mental talk into talk about patterns of behaviour.

Gilbert Ryle promoted behaviourism as a response to what he called 'Descartes's myth' of 'the ghost in the machine'.(3) A dualist would regard talk about being in love, or wanting to visit Paris, as talk about an immaterial mind whose states lie hidden behind observable bodily behaviour. Ryle proposed to analyse this mental talk as being about the observable behaviour itself. He did not, however, aim to replace all mental terms by terms appropriate to the science of material bodies moving through space. He analysed believing that the ice is thin as, in part, being 'prone to skate warily' and it was enough, for his purposes, that skating warily is an observable and recognizable kind of behaviour, even if it is not readily defined in terms of the trajectories of body parts.

Because action is explained in terms of what the agent believes and what the agent wants, analytical behaviourism faces a major objection of principle. There is no pattern of behaviour associated with a belief, by itself. Someone who believes that the ice is thin but has an unusual desire to be immersed in ice-cold water may not skate warily. So there is no prospect of analysing any belief in terms

of behaviour. We might elevate this point into a general requirement on the description of any creature as having beliefs. Attributions of beliefs are not warranted if they merely summarize the creature's dispositions to exhibit patterns of behaviour. A belief is a mental state that can figure in the explanation of indefinitely many different actions in pursuit of different goals.

The identity theory: Place and Smart

Ryle's behaviourism involved a clear rejection of Descartes's duality of material and immaterial substances, but **central state materialism** (also known as the **identity theory**) encapsulated a more thoroughgoing commitment to the physicalist world-view. If the physical effects of our experiences, thoughts, and volitions have wholly physical causes then there is no causal work left for distinct mental items to do. To avoid epiphenomenalism, mental states, processes and events were to be identified with physical states, processes and events, and mental properties with physical properties. Place advanced a precursor of the identity theory, restricted to the case of conscious experiences,[4] and this was generalized by Smart, who identified beliefs and desires, intentions and hopes, as well as sensations and experiences, with brain states or processes.[5]

The identity theory defends the idea of **mental causation** by identifying each mental state with a physical state that is a locus of causal powers. But, taken literally, the identity theory is bound to seem chauvinistic. No being with a physical constitution radically different from ours could be described as feeling anything, or thinking anything, or wanting anything.

Functionalism: Putnam and Lewis

The functionalist response to the identity theory is that what a system does is more important than what it is made from. Physically different computing machines can run the same software and one version of the functionalist theory of the mind–brain relation is that the mind is the software of the brain.[6]

In an early version of **functionalism**, Hilary Putnam proposed that mental states are functional states like the states of an abstractly defined Turing machine rather than physical states like the states of a human brain.[7] This **machine functionalism** had the advantage of not tying mental states to a particular physical substrate but also a disadvantage. Since a Turing machine is in only one state at a time, machine states are not analogous to mental states like being in love or wanting to visit Paris.

The dominant version of contemporary functionalism, attributable to David Lewis, is **analytical functionalism**.[8] The leading idea is that commonsense specifications of the interconnected causal roles of mental states can be taken as interlocking analyses of mental state terms. For any physical being with a mind, there will be physical states playing each of the mental state causal roles but different physical states may play the same causal role in physically different minded beings—in human beings and Martians, for example.

Functionalism thus avoids the apparent chauvinism of the identity theory by allowing that a human being may be in the same mental state as a being with a very different physical constitution. But functionalism faces the opposite problem of apparently being too liberal. It seems to be possible to make up examples in which physical states play the causal roles that are supposed to define mental states, yet where, intuitively, there is no intelligence and no mental life.[9]

Challenges to functionalism

The dominant contemporary theories of the mind–brain relation are compatible with a broadly physicalist world-view and analytical functionalism, in particular, is consistent with a version of physicalism, **a priori physicalism**, that is both ontologically and conceptually **reductionist**.[10]

Ontologically, analytical functionalism is like the identity theory in its commitment to types of physical state that *realize* mental states. Functionalism does not quite say that being in pain is to be identified with having C-fibres firing (the standard example for the identity theory); but it does say that the causal role of the mental state of being in pain is played, in human beings, by the physical state of having C-fibres firing.

Despite this ontological similarity to the identity theory, analytical functionalism is *conceptually* more like analytical behaviourism in being a thesis about the meanings of our mental terms. According to behaviourism, it is a matter of meaning, or **conceptual analysis**, that being in a mental state is being disposed to produce particular patterns of behaviour. According to functionalism, it is equally a matter of meaning that being in a mental state is being in a state that plays a particular causal role. Consequently, analytical functionalism is *conceptually reductionist*. The mental facts, as conceived by the functionalist, are *entailed a priori* by the physical facts.

As we shall now see, both the ontological and the conceptual commitments of analytical functionalism face challenges.

Rylean behaviourism revisited

A theorist of the mind–brain relation who was sympathetic to Rylean behaviourism might challenge the *ontological* commitments that are shared by functionalism and the identity theory. The neo-behaviourist might accept the idea that if a system has a disposition to exhibit a particular pattern of behaviour then there must be a basis for this disposition in the system's inner constitution. But he might argue that identifying individual mental states with physical states, or insisting that mental states are individually realized by physical states, goes beyond what is required by this idea.

Dispositions do not float free of inner constitution and the behavioural dispositions of human beings are, presumably, underpinned by states and processes of the brain. But it is not obviously required that there must be a single brain state that underpins precisely the dispositions that are associated with the attribution of a single mental state. This neo-behaviourism may draw support from remarks made by Wittgenstein.[11]

Neo-behaviourism will be open to objection so long as it retains the unattainable commitment to an analysis of belief attributions in behavioural terms. But there is an alternative view that abandons those analytical ambitions. The **interpretationist** says that mentalistic interpretation is answerable to a creature's behaviour in various actual and hypothetical circumstances, but that this answerability is a matter of 'making sense' of the creature and cannot be codified mental state by mental state.[12] Rather, the interpreter casts a net of psychological description—'X is in pain; X is in love; X wants to visit Paris; X believes that the ice is thin; …'—over a writhing mass of behaviour. Tracts of human behaviour normally support this interpretive project and, presumably, the behaviour is susceptible of causal explanation. But we should not assume that the physical causes of behaviour must have an articulation that matches the structure of the interpreter's description.[13]

Interpretationism is compatible with a broadly physicalist worldview but it involves some departure from apparently plausible claims about mental reality and mental causation. The interpretationist is not committed to the claim that there are individual mental states—for example, individual beliefs such as my belief that there is a bottle of white wine in the refrigerator, or that I have an appointment at 9 a.m.—that are bearers of causal powers.

Consciousness and the explanatory gap

The *conceptual* commitments of analytical functionalism are challenged by our conception of **conscious mental states**.

According to functionalism, all mental states are realized by physical, specifically neural, states and the phenomenal properties of conscious mental states are physical properties of those neural states. We can ask what makes the difference between conscious mental states and unconscious mental states. Is there, for example, something distinctive about the neural underpinnings of conscious mental states? If we had a plausible answer to that question, there would be the further question *why* mental states with that distinctive neural nature are *conscious* mental states. This question is apt to seem puzzling and even unanswerable. But, according to analytical functionalism, there would be no puzzling 'why?' question about consciousness. All the mental facts, including the facts about consciousness and phenomenology, are entailed a priori by the physical facts.

A powerful intuition thus speaks against the conceptual commitments of analytical functionalism. For it seems that even the full physical story about the world would not settle a priori the question whether a creature was in a conscious mental state. It seems to be conceivable (not ruled out a priori) that there could be a creature physically just like one of us yet lacking consciousness—a *zombie*—or even a complete physical duplicate of our world from which consciousness was totally absent—a *zombie world*.[14] Between the physical sciences and the facts of consciousness there seems to be an **explanatory gap**.[15]

Thomas Nagel has drawn attention to a difference between two kinds of conception. Conceptions of conscious mental states are **subjective**; they are accessible from some, but not all, points of view. The conscious mental states that we can *conceive* are limited to relatively modest imaginative extensions from the conscious mental states that we ourselves *undergo*. In contrast, the conceptions deployed in grasping theories in the physical sciences are **objective**; they are accessible from many different points of view. The physical theories that we can grasp are limited, not by our sensory experience, but by our intellectual powers.[16] Many contemporary philosophers of consciousness argue that the explanatory gap is a product of this duality of conceptions. There is no a priori entailment from the physical and functional facts *objectively conceived* to the phenomenal facts *subjectively conceived*.

The majority of these philosophers maintain that a duality of conceptions does not require an ontological dualism of substances, states, or properties and that the explanatory gap is consistent with physicalism as an ontological doctrine. But there is an important minority view that acceptance of an explanatory gap must lead to a rejection of physicalism. David Chalmers, beginning from the intuition of an explanatory gap, recommends a return to some form of dualism.[17] Others argue in the opposite direction, embracing physicalism, denying that there is an explanatory gap, and accepting the counterintuitive conceptual reductionism of analytical functionalism.

Personal and subpersonal levels of description and explanation

The mind–brain relation is an aspect of a more encompassing relationship between persons and the physical systems of which they are constituted, including systems of neural information processing.

Our conception of persons as such is a conception of subjects and agents. At the personal level of description and explanation, we describe what people feel, think, want and do, and we explain what people do in terms of their sensations, beliefs, and desires. As the case of conscious mental states illustrates, our personal-level *descriptions* are not always entailed a priori by physical and functional descriptions of the systems that constitute us. Personal-level descriptions involve subjective and normative concepts that are different from the objective and descriptive concepts that figure in the physical sciences.

Our personal-level *explanatory practices* seem to be different in kind from our scientific practices of explaining the operation of mechanical systems. McDowell describes personal-level explanations as 'explanations in which things are made intelligible by being revealed to be, or to approximate to being, as they rationally ought to be'.[18] In a similar spirit, Dennett describes them as 'non-mechanistic'. A mechanistic account of what happens when a person feels, thinks, wants, and acts would belong at a quite different level of description and explanation, not the 'level of people and their sensations and activities', but 'the *subpersonal* level of brains and events in the nervous system'.[19]

One extreme view of the relationship between the personal and subpersonal levels highlights what is distinctive about the personal level and regards it as substantially independent from the subpersonal level. This view might encourage the interpretationist account of personal-level psychological descriptions, minimizing the ontological and causal commitments of personal-level discourse to avoid constraints on that discourse from the subpersonal level of neuroscience.

The opposite extreme view is the conceptually reductionist view of analytical functionalism. The personal level is the level of mental states whose causal roles are revealed by conceptual analysis while the subpersonal level is the level of neural states that play those roles. There are no explanatory gaps. All that is true at the personal level is entailed a priori by physical truths at the subpersonal level.

According to an attractive view that is intermediate between these two extremes, the relationship between the personal and subpersonal levels is one of *interaction without reduction*.[20] As against the first extreme view, the personal level is not independent of the subpersonal level but constrained by it, because our personal-level descriptions—cast in terms of experience, thought, planning, and agency—carry commitments about causal structure in the brain. But, as against the second extreme view, there are also explanatory gaps that reveal themselves when we try to construct illuminating accounts of those personal-level notions using only the subpersonal-level resources of neuroscience.

Conclusion

Descartes's ontological dualism of mind and body made it difficult for him to describe the phenomenology of embodiment, the way we experience our own body. Contemporary theories of the mind–brain relation are predominantly physicalist, rather than dualist, in their ontology. But the duality of objective and subjective conceptions

still presents a challenge for the sciences of the mind. Persons understood as such, partly from the first-person perspective—persons conceived as subjects and agents, with their experiences, thoughts, plans and actions—will not be visible in a purely objective, scientific story of the physical world.

Further information

Anthology

Chalmers, D.J. (ed.) (2002). *Philosophy of mind: classical and contemporary readings*. Oxford University Press, Oxford.

Textbook

Braddon-Mitchell, D. and Jackson, F.C. (2007). *Philosophy of mind and cognition: an introduction* (2nd edn). Blackwell Publishing, Oxford.

References

1. Descartes, R. (1641). *Meditations on first philosophy* In Descartes, R. (1988). *Selected philosophical writings* (trans. eds. J. Cottingham, R. Stoothoff, and D. Murdoch). Cambridge University Press, Cambridge, UK.
2. Papineau, D. (2002). *Thinking about consciousness*, Appendix. Oxford University Press, Oxford.
3. Ryle, G. (1949). Descartes' myth. In *The concept of mind*, Chap. 1. Penguin Books, Harmondsworth.
4. Place, U.T. (1956). Is consciousness a brain process? *The British Journal of Psychology*, **47**, 44–50.
5. Smart, J.C.C. (1959). Sensations and brain processes. *Philosophical Review*, **68**, 141–56.
6. Block, N. (1995). The mind as the software of the brain. In *An invitation to cognitive science* (2nd edn), Vol. 3 (eds. E.E. Smith and D.N. Osherson), pp. 377–425. MIT Press, Cambridge, MA.
7. Putnam, H. (1967). The nature of mental states. Originally titled 'Psychological predicates'. In *Art, mind, and religion* (eds. W.H. Capitan and D.D. Merrill), pp. 37–48. University of Pittsburgh Press, Pittsburgh, PA.
8. Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, **50**, 249–58.
9. Block, N. (1978). Troubles with functionalism. In *Minnesota studies in the philosophy of science*, Vol. 9 (ed. C. Wade Savage), pp. 261–325. University of Minnesota Press, Minneapolis, MN.
10. Jackson, F.C. (2005). The case for a priori physicalism. In *Philosophy—science—scientific philosophy: main lectures and colloquia of GAP.5* (eds. C. Nimtz and A. Beckermann), pp. 251–65. Fifth International Congress of the Society for Analytical Philosophy, Bielefeld, 22–26 September 2003. Mentis, Paderborn.
11. Wittgenstein, L. (1981). *Zettel*, (eds. G.E.M. Anscombe and G.H. von Wright, trans. G.E.M. Anscombe), Sections 608–9. Basil Blackwell, Oxford.
12. Davidson, D. (1973). Radical interpretation. *Dialectica*, **27**, 313–28. Reprinted in Davidson, D. (1984). *Inquiries into truth and interpretation*, pp. 125–39. Oxford University Press, Oxford.
13. Dennett, D.C. (1987). True believers: the intentional strategy and why it works. In *The intentional stance*, pp. 13–35. MIT Press, Cambridge, MA.
14. Chalmers, D.J. (1996). *The conscious mind: in search of a fundamental theory.* Oxford University Press, Oxford.
15. Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, **64**, 354–61.
16. Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, **83**, 435–50. Reprinted in Nagel, T. (1979). *Mortal questions*, pp. 165–80. Cambridge University Press, Cambridge, UK.
17. Chalmers, D.J. (2007). Naturalistic dualism. In *The Blackwell companion to consciousness* (eds. M. Velmans and S. Schneider), pp. 359–68. Blackwell Publishing, Oxford.
18. McDowell, J. (1985). Functionalism and anomalous monism. In *Actions and events: perspectives on the philosophy of Donald Davidson* (eds. E. LePore and B.P. McLaughlin), pp. 387–98. Basil Blackwell, Oxford.
19. Dennett, D.C. (1969). *Content and consciousness*, pp. 91–4. Routledge and Kegan Paul, London.
20. Davies, M. (2000). Interaction without reduction: the relationship between personal and sub-personal levels of description. *Mind and Society*, **1**, 87–105.

2.2

Statistics and the design of experiments and surveys

Graham Dunn

Introduction

Research into mental illness uses a much wider variety of statistical methods than those familiar to a typical medical statistician. In many ways there is more similarity to the statistical toolbox of the sociologist or educationalist. It would be a pointless exercise to try to describe this variety here but, instead, we shall cover a few areas that are especially characteristic of psychiatry. The first and perhaps the most obvious is the problem of measurement. Measurement reliability and its estimation are discussed in the next section. Misclassification errors are a concern of the third section, a major part of which is concerned with the estimation of prevalence through the use of fallible screening questionnaires. This is followed by a discussion of both measurement error and misclassification error in the context of modelling patterns of risk.

Another major concern is the presence of missing data. Although this is common to all areas of medical research, it is of particular interest to the psychiatric epidemiologist because there is a long tradition (since the early 1970s) of introducing missing data by design. Here we are thinking of two-phase or double sampling (often confusingly called two-stage sampling by psychiatrists and other clinical research workers). In this design a first-phase sample are all given a screen questionnaire. They are then stratified on the basis of the results of the screen (usually, but not necessarily, using two strata—likely cases and likely non-cases) and subsampled for a second-phase diagnostic interview. This is the major topic of the third section.

If we are interested in modelling patterns of risk, however, we are not usually merely interested in describing patterns of association. Typically we want to know if genetic or environmental exposures have a causal effect on the development of illness. Similarly, a clinician is concerned with answers to the question 'What is the causal effect of treatment on outcome?' How do we define a causal effect? How do we measure or estimate it? How do we design studies in order that we can get a valid estimate of a causal effect of treatment? Here we are concerned with the design and analysis of randomized controlled trials (RCTs). This is the focus of the fourth section of the present chapter.

Finally, at the end of this chapter pointers are given to where the interested reader might find other relevant and useful material on psychiatric statistics.

Reliability of instruments

In this section we consider two questions:

- What is meant by 'reliability'?
- How do we estimate reliabilities?

Models and definitions

Most clinicians have an intuitive idea of what the concept of reliability means, and that being able to demonstrate that one's measuring instruments have high reliability is a good thing. Reliability concerns the consistency of repeated measurements, where the repetitions might be repeated interviews by the same interviewer, alternative ratings of the same interview (as a video recording) by different raters, alternative forms or repeated administration of a questionnaire, or even different subscales of a single questionnaire, and so on. One learns from elementary texts that reliability is estimated by a correlation coefficient (in the case of a quantitative rating) or a kappa (κ) or weighted κ statistic (in the case of a qualitative judgement such as a diagnosis). Rarely are clinicians aware of either the formal definition of reliability or of its estimation through the use of various forms of intraclass correlation coefficient rho (ρ).

First consider a quantitative measurement X. We start with the assumption that it is fallible and that it is the sum of two components: the 'truth' T and 'error' E. If T and E are statistically independent (uncorrelated), then it can be shown that

$$\mathrm{Var}(X) = \mathrm{Var}(T) + \mathrm{Var}(E) \tag{2.2.1}$$

where $\mathrm{Var}(X)$ is the variance of X (i.e. the square of its standard deviation), and so on. The reliability ρ_x of X is defined as the proportion of the total variability of X (i.e. $\mathrm{Var}(X)$) that is explained by the variability of the true scores (i.e. $\mathrm{Var}(T)$):

$$\rho_x = \frac{\mathrm{Var}(T)}{\mathrm{Var}(X)} = \frac{\mathrm{Var}(T)}{\mathrm{Var}(T) + \mathrm{Var}(E)} \tag{2.2.2}$$

This ratio will approach zero as the variability of the measurement errors increases compared with that of the truth. Alternatively, it will approach one as the variability of the errors decreases. The standard deviation of the measurement errors (i.e. the square root of $\mathrm{Var}(E)$) is usually known as the instrument's standard error of measurement. Note that reliability is not a fixed characteristic of

an instrument, even when its standard error of measurement (i.e. its precision) is fixed. When the instrument is used on a population that is relatively homogeneous (low values of Var(T)), it will have a relatively low reliability. However, as Var(T) increases then so does the instrument's reliability. In many ways the standard error of measurement is a much more useful summary of an instrument's performance, but one should always bear in mind that it too might vary from one population to another—a possibility that must be carefully checked by both the developers and users of the instrument.

Now let us complicate matters slightly. Suppose that a rating depends not only on the subject's so-called true score T and random measurement error E, but also on the identity R, say, of the interviewer or rater R. That is, each rater has his or her own characteristic bias (constant from assessment to another) and the biases can be thought of as varying randomly from one rater to another. Again, assuming statistical independence, we can show that, if $X = T + R + E$, then

$$\text{Var}(X) = \text{Var}(T) + \text{Var}(R) + \text{Var}(E) \tag{2.2.3}$$

But what is the instrument's reliability? It depends. If subjects in a survey or experiment, for example, are each going to be assessed by a rater randomly selected from a large pool of possible raters, then

$$\rho_{xa} = \frac{\text{Var}(T)}{\text{Var}(X)} = \frac{\text{Var}(T)}{\text{Var}(T) + \text{Var}(R) + \text{Var}(E)} \tag{2.2.4}$$

However, if only a single rater is to be used for all subjects in the proposed study, there will be no variation due to the rater and the reliability now becomes

$$\rho_{xb} = \frac{\text{Var}(T)}{\text{Var}(T) + \text{Var}(E)} \tag{2.2.5}$$

Of course, $\rho_{xb} > \rho_{xa}$. Again, the value of the instrument's reliability depends on the context of its use. This is the essence of generalizability theory.[1] The three versions of ρ given above are all intraclass correlation coefficients and are also examples of what generalizability theorists refer to as generalizability coefficients.

Designs

Now consider two simple designs for reliability (generalizability) studies. The first involves each subject of the study being independently assessed by two (or more) raters but that the raters for any given subject have been randomly selected from a very large pool of potential raters. The second design again involves each subject of the study being independently assessed by two (or more) raters, but in this case the raters are the same for all subjects. Equations (2.2.1) and (2.2.2) are relevant to the analysis of data arising from the first design, whilst eqns (2.2.3–2.2.5) are relevant to the analysis of data from the second design.

Estimation of ρ and κ from ANOVA tables

When we come to analyse the data it is usually appropriate to carry out an analysis of variance (ANOVA). For the first design we carry out a one-way ANOVA (X by subject) and for the second we perform a two-way ANOVA (X by rater and subject). In the latter case we assume that there is no subject by rater interaction and accordingly constrain the corresponding sum of squares to be zero. We assume that readers are reasonably familiar with an analysis of variance table. Each subject has been assessed by, say, k raters. The one-way ANOVA yields a mean square for between-subjects variation (**BMS**) and a mean square for within-subjects variation (**WMS**). WMS is an estimate of Var(E) in eqn. (2.2.1). Therefore, the square root of WMS provides an estimate of the instrument's standard error of measurement. The corresponding estimate of ρ_x is given by

$$r_x = \frac{\text{BMS} - \text{WMS}}{\text{BMS} + (k-1)\text{WMS}} \tag{2.2.6}$$

where r_x is used to represent the estimate of ρ_x rather than the true, but unknown, value. In the case of $k = 2$, r_x becomes

$$r_x = \frac{\text{BMS} - \text{WMS}}{\text{BMS} + \text{WMS}} \tag{2.2.7}$$

In the slightly more complex two-way ANOVA, the ANOVA table provides values of mean squares for subjects or patients (**PMS**), raters (**RMS**), and error (**EMS**). We shall not concentrate on the details of estimation of the components of eqn. (2.2.3) (see Fleiss[2] or Streiner and Norman[3]) but simply note that ρ_{xa} is estimated by

$$r_{xa} = \frac{n(\text{PMS} - \text{EMS})}{n \times \text{PMS} + k \times \text{RMS} + (nk - n - k)\text{EMS}} \tag{2.2.8}$$

where n is the number of subjects (patients) in the study.

In reporting the results of a reliability study, it is important that investigators give some idea of the precision of their estimates of reliability, for example by giving an appropriate standard error or, even better, an appropriate confidence interval. The subject is beyond the scope of this chapter, however, and the interested reader is referred to Fleiss[2] or Dunn[4] for further illumination.

Finally, what about qualitative measures? We shall not discuss the estimation and interpretation of κ in any detail here but simply point out that for a binary (yes/no) measure one can also carry out a two-way analysis of variance (but ignore any significance tests since they are not valid for binary data) and estimate r_{xa} as above. In large samples r_{xa} is equivalent to κ.[3] A corollary of this is that κ is another form of reliability coefficient and, like any of the reliability coefficients described above, will vary from one population to another (i.e. it is dependent on the prevalence of the symptom or characteristic being assessed).

Prevalence estimation

Following Dunn and Everitt,[5] we ask the following questions of a survey report.

- Do the authors clearly define the sampled population?
- Do the authors discuss similarities and possible differences between their sampled population and the stated target population?
- Do the authors report what sampling mechanism has been used?
- Is the sampling mechanism random? If not, why not?
- Exactly what sort of random sampling mechanism has been used?
- Do the methods of data analysis make allowances for the sampling mechanism used?

Of course, it is vital that what counts as a case should be explained in absolute detail, including the method of eliciting symptoms (e.g. structured interview schedule), screening items, additional impairment criteria, and so on, as well as operational criteria or algorithms used in making a diagnosis. In the following we concentrate on the statistical issues. First, we consider survey design (and the associated sampling mechanisms), and then we move on to discuss the implications of design for the subsequent analysis of the results.

Survey design

Here we are concerned with the estimation of a simple proportion (or percentage). We calculate this proportion using data from the sample and use it to infer the corresponding proportion in the underlying population. One vital component of this process is to ensure that the sampled population from which we have drawn our subjects is as close as possible to that of the target population about which we want to draw conclusions. We also require the sample to be drawn from the sample population in an objective and unbiased way. The best way of achieving this is through some sort of random sampling mechanism. Random sampling implies that whether or not a subject finishes up in the sample is determined by chance. Shuffling and dealing a hand of playing cards is an example of a random selection process called simple random sampling. Here every possible hand of, say, five cards has the same probability of occurring as any other. If we can list all possible samples of a fixed size, then simple random sampling implies that they all have the same probability of finishing up in our survey. It also implies that each possible subject has the same probability of being selected. But note that the latter condition is not sufficient to define a simple random sample. In a systematic random sample, for example, we have a list of possible people to select (the sampling frame) and we simply select one of the first 10 (say) subjects at random and then systematically select every tenth subject from then on. All subjects have the same probability of selection, but there are many samples which are impossible to draw using this mechanism. For example, we can select either subject 2 or subject 3 with the same probability (1/10), but it is impossible to draw a sample which contains both.

What other forms of random sampling mechanisms might be used? Perhaps the most common is a stratified random sample. Here we divide our sampled population into mutually exclusive groups or strata (e.g. men and women, or five separate age groups). Having chosen the strata, we proceed, for example, to take a separate simple random sample from each. The proportion of subjects sampled from each of the strata (i.e. the sampling fraction) might be constant across all strata (ensuring that the overall sample has the same composition as the original population), or we might decide that one or more strata (e.g. the elderly) might have a higher representation. Another commonly used sampling mechanism is multistage cluster sampling. For example, in a national prevalence survey we might chose first to sample health regions or districts, then to sample post codes within the districts, and finally to select patients randomly from each selected post code. (See Kessler[6] and Jenkins *et al.*[7] for discussions of complex multistage surveys of psychiatric morbidity.)

One particular design that has been used quite often in surveys designed to estimate the prevalence of psychiatric disorders is called two-phase or double sampling. Psychiatrists frequently refer to this as two-stage sampling. This is unfortunate, since it confuses the two-phase design with simple forms of cluster sampling in which the first-stage involves drawing a random sample of clusters and the second-stage a random sample of subjects from within each of the clusters. In two-phase sampling, however, we first draw a preliminary sample (which may be simple, stratified, and/or clustered) and then administer a first-phase screening questionnaire such as the General Health Questionnaire (see Chapter 1.8.1). On the basis of the screen results we then stratify the first-phase sample. Note that we are not restricted to two strata (likely cases versus the rest), although this is perhaps the most common form of the design. We then draw a second-phase sample from each of the first-phase strata and proceed to give these subjects a definitive psychiatric assessment. The point of this design is that we do not waste expensive resources interviewing large numbers of subjects who not appear (on the basis of the first-phase screen) to have any problems. Accordingly, the sampling fractions usually differ across the first-phase strata. However, it is vital that each of the first-phase strata have a reasonable representation in the second-phase, and it is particularly important that all of the first-phase strata provide some second-phase subjects. The reader is referred to Pickles and Dunn[8] for further discussion of design issues in two-phase sampling (including discussion of whether it is worth the bother).

Analysis of the results

Here we are particularly concerned with the last of the questions posed at the beginning of the section. In fact, it is a question that should be asked not only of prevalence surveys but of all investigations whether they are epidemiological surveys, intervention studies, or laboratory experiments. How was the design incorporated in the analysis? Frequently the required information is missing. Either the authors are ignorant of the implications of the design, or the journal editor has insisted that technical details are stripped from the report, or both.

Consider a hypothetical sample of 100 participants who have contributed to an estimate of prevalence of, say, depression using a definitive psychiatric interview.

Seventy of the participants have been given a diagnosis of depression. What is a valid estimate of prevalence? What is the standard error of the estimate? Assuming that the data have arisen through simple random sampling, the prevalence p is estimated by 0.70 and its variance is given by

$$\mathrm{Var}(p) = \frac{1}{np(1-p)} \tag{2.2.9}$$

where n is the sample size. The standard error is then given by the square root of this expression.

Suppose that we are now told that the results were obtained from a two-phase survey. The size of the first-phase sample was 300. Of these, 100 were screen positive and 200 were screen negative. The second-phase sample consisted of 70 screen positives, of whom 65 were found to be depressed on interview, together with 30 screen negatives, of whom 5 were found to be depressed on interview. The estimate of prevalence is given by

$$\begin{aligned} p &= P(\text{screen +ve}) \times P(\text{interview +ve}|\text{screen +ve}) + P(\text{screen} - \text{ve}) \\ &\quad \times P(\text{interview +ve}|\text{screen}-\text{ve}) \\ &= (100/300) \times (65/70) + (200/300) \times (5/30) \\ &= 0.42 \end{aligned} \tag{2.2.10}$$

where $P(A)$ should be read as 'probability of A' and $P(A|B)$ should be read as 'probability of A given B' or 'probability of A conditional on B having occurred'. The vertical '|' should not be confused with division, represented by '/'.

The prevalence estimate from the two-phase survey is considerably lower than if simple random sampling had been assumed. How has this arisen? Obviously the second-phase sample has been enriched for people who are likely to be depressed. The sampling fraction for the screen positives is 70/100, that is each second-phase participant can be thought of as representing 100/70 of screen positives from the original sample. Similarly, the sampling fraction for the screen negatives is 30/200, and each second-phase participant represents 200/30 screen-negative participants from the first-phase sample. The reciprocal of the sampling fraction is called the sampling weight. The total weighted second-phase sample size is $70 \times (100/70) + 30 \times (200/30) = 300$, the first-phase sample. Similarly, the total weighted number of cases of depression is $65 \times (100/70) + 5 \times (200/30) \approx 126$. The latter is the estimate of the number of cases in the first-phase sample. Hence the estimate of prevalence is 126/300 = 0.42, as before. To recapitulate in a slightly more technical way, if the ith individual in the second-phase sample is assigned a sampling weight w_i, and if the interview outcome y_i has a value of 1 if the ith subject is a case and is 0 otherwise, then the weighted prevalence estimate is given by

$$p = \Sigma w_i y_i / \Sigma w_i \tag{2.2.11}$$

where Σ means 'sum over all observations in the second-phase sample' and x_i is simply an indicator that the observation is, indeed, a second-phase observation ($x_i = 1$ for everyone). This estimator is an example of the well-known Horwitz–Thompson estimator from the sampling survey literature[9] but it is not particular familiar to psychiatrists or medical statisticians. We shall discuss the use of weighting adjustments again below.

Returning to our original two-phase calculations, let $A = P$(screen +ve) and $B = 1–A = P$(screen –ve). Also, let $p = P$(interview +ve|screen +ve) and $q = P$(interview +ve|screen –ve), so that eqn. (2.2.10) becomes

$$p = Ap + Bq \tag{2.2.12}$$

The variance of the estimate of prevalence from the two-phase design is given by[10]

$$\mathrm{Var}(p) = \frac{A^2 p(1-p)}{n_1} + \frac{B^2 q(1-q)}{n_2} + \frac{(p-q)^2 AB}{n_3} \tag{2.2.13}$$

where n_1 is the number of first-phase screen positives and n_2 is the number of first-phase screen negatives (and $n = n_1+n_2$ is the total (first phase) sample size).

Validation of screening questionnaires

It is frequently the case that data from a two-phase survey which has been designed to estimate prevalence are also used to examine the characteristics of the screen questionnaire (in particular, sensitivity and specificity). Readers who are unfamiliar with these concepts are referred to Chapter 2.7 or to Goldberg and Williams.[11] Sensitivity is the proportion of true cases who are screen positive. Specificity is the proportion of true non-cases who are screen negative. The trouble is caused because we used the screen first and then differentially subsampled to carry out the definitive diagnostic interview. Readers familiar with the use of Bayes' theorem will realize how to solve the problem, but here we use another version of the Horwitz–Thompson estimator:

$$\text{Sensitivity} = \Sigma w_i y_i z_i / w_i y_i \tag{2.2.14}$$

and

$$1\text{–specificity} = \Sigma w_i (1-y_i) z_i / \Sigma w_i (1-y_i) \tag{2.2.15}$$

where, as before, y_i indicates whether the ith subject was a true case of depression (1 = yes, 0 = no). This ensures that the calculations in eqn (2.2.14) are only being carried out on the true cases and, similarly, that the calculations in (2.2.15) are only being carried out on the non-cases. Again, w_i is the second-phase sampling weight. The new variable z_i indicates whether the screen result was positive (1 = yes, 0 = no). An alternative, and perhaps easier, approach is to split the second-phase sample into two: cases and non-cases. Estimation of sensitivity and specificity in these two subfiles (assuming that they are being stored on a computer) is then computationally exactly the same as the weighted estimation of prevalence discussed in the previous section. In the first file, sensitivity is simply the weighted sum of the screen positives divided by the weighted sum of the cases. Similarly, in the second file, specificity is the weighted sum of the screen negatives divided by the weighted sum of the non-cases.

Many readers will be familiar with the idea of choosing a range of cut-points for the screen questionnaire and then estimating sensitivity and specificity at each of the choices. A plot of sensitivity against 1—specificity is called a receiver operating characteristic (**ROC**) curve. If the screen is of no use, then the plot will be a straight line through the origin with unit slope. A good screen will produce a convex curve (the greater the area between the observed curve and that indicated by a straight line with unit slope, the better the screen is at discriminating between cases and non-cases). It is sometimes said that one cannot investigate ROC curves using two-phase data. This view is, in fact, mistaken. One can think of the two-phase sampling design as a mechanism by which one can deliberately introduce the analogues of verification bias.[12] Note that there is no necessity to restrict the first-phase stratification to just two strata (potential cases versus non-cases) to define the sampling fractions for the second-phase of the survey. We start by calculating observed sampling fractions for each discrete outcome of the screening questionnaire. These define the corresponding sampling weights. We then consider all the possibilities for defining z_i in eqns (2.2.14) and (2.2.15)—there is no need for the z_i to correspond to the way that the second-phase sampling fractions were determined. We then repeatedly use eqns (2.2.14) and (2.2.15), keeping the weights constant as we change the definition of the z_i. One important point to bear in mind is that if the characteristics of the screen are not fairly well known beforehand and if one of the major aims of the survey is to carry out an ROC analysis, then this is not a particularly efficient design to use. It would be better to go back to the simple random sample—all subjects assessed by both screen and interview.

If one needs, say, confidence intervals for estimates of sensitivity and specificity, it is relatively straightforward to do this via a weighted logistic regression (see next section). The file can be split into cases and non-cases and then, using appropriate software (see below), one fits a logistic model containing a predictor variable

which has the value of 1 for all subjects (i.e. just fitting a constant). One then obtains the confidence interval for the intercept term in the output. Finally, the inverse of the logistic transformation of the lower and upper confidence limits will yield the corresponding limits for the sensitivity (specificity) itself. Note that the interval will be asymmetric and will be within the permitted bounds of zero and unity.

Modelling patterns of risk

Henderson (see Chapter 2.7) has introduced the idea of an odds ratio to measure the association between a suspected risk factor and disease. In a linear logistic model the response variable is the natural logarithm of the odds (of disease). Therefore the difference between two groups on this logistic scale (i.e. $\log(a/b)$–$\log(c/d)$ using Henderson's notation) is equivalent to the logarithm of the odds ratio (i.e. $\log(ad/bc)$). This provides us with an easy way to calculate confidence intervals for odds ratio: log(odds ratio) or $\log(ad/bc)$ is normally distributed with variance $1/a + 1/b + 1/c + 1/d$ (the corresponding standard error is the square root of this variance). The 95 per cent confidence interval for the log(odds ratio), for example, is then the point estimate plus or minus 1.96 standard errors. Taking exponents (antilogarithms) of these limits provides the corresponding limits for the odds ratio itself. The exponent of the parameter estimate in the output of the logistic regression run provides the point estimate for the corresponding odds ratio.

The great advantage of logistic regression is that it enables us to model the potential effects of several risk factors simultaneously. It allows us to adjust for the effects of suspected confounder(s) in assessing the effects of a risk factor of interest. Logistic regression can also be generalized to cope with the use of sampling weights, either to cope with data missing by design (as in a two-phase survey) or to allow for non-response and/or attrition.[13] However, one must be very wary of using weights in software packages that do not explicitly deal with sampling weights. Many packages have weighting functions but these are interpreted as frequency weights—the number of times the observation has been made instead of the number of times it might have been made (as in the case of a sampling weight). The use of frequency weights, as opposed to sampling weights, produces standard errors, and confidence intervals that are far too small. This is not a subtle effect; it can make an enormous difference to a *P* value, giving the impression of a highly significant effect when, in reality, there is little or nothing there.[14] We illustrate this point by reference to a two-phase survey of psychiatric morbidity in Cantabria in Northern Spain.[15,16] The weighted prevalence estimate from these data is 31 per cent. The appropriate 95 per cent confidence interval (**CI**), obtained using sampling weights is (26 per cent, 40 per cent). The naive use of frequency weights produces a 95 per cent CI of (28 per cent, 35 per cent), which is much too narrow. The odds ratio indicating how much higher the prevalence of disorder is in women compared with men is 2.02 with a 95 per cent CI of (0.86, 4.74). The naive use of frequency weights gives the same point estimate (2.02), but here the 95 per cent CI is (1.45, 2.77); again, this is much too narrow. In the analysis of a similar study from Verona in Northern Italy, Dunn *et al.*[14] found a corresponding odds ratio of 2.85 with a 95 per cent CI of (1.31, 6.19). The *P* value for this odds ratio is about 0.008. The incorrect use of frequency weights gives us a 95 per cent CI of (2.31, 3.53) and a corresponding *P* value of less than 0.00001—at least an 800-fold difference!

Evaluating treatment effects

Readers will be familiar with the challenges posed by confounding in trying to validly infer an effect of an exposure on the development of illness from data arising from an observational study (or epidemiological survey). We can model patterns of risk (as in the above section) but we can never be sure that we have allowed for all possible sources of confounding (i.e. the effects of unmeasured variables that are associated with the exposure of interest and also influence outcome) in assessing the effect of a particular risk factor. The same challenges apply to the use of observational data to the evaluation of the effect of a treatment or other intervention. Although it may be possible to obtain valid treatment effect estimates from observational data,[17,18] the ideal is to use a randomized experiment. Allocation of treatments by randomization ensures that there are no systematic selection effects (i.e. no biases arising from hidden confounding) and enables the investigator to obtain valid measurements of uncertainty (i.e. valid p-values, standard errors, and confidence intervals).

How do we define a causal effect of treatment? Following Rubin[19] we define it as the comparison between potential outcomes or counterfactuals. For a patient who has received treatment we define the treatment effect as a comparison between the outcome that we have observed after the receipt of treatment (Y_t, say) with that we would have observed if, contrary to fact, the same patient had not received the treatment (Y_c, say the subscript c indicating a control condition). For a patient who has not received treatment we compared the observed outcome (Y_c, in this case) with the unobserved counterfactual (Y_t). Typically, we might be interested in the difference (Y_t–Y_c) but ratios might also be of interest. We define the Average Causal Effect (ACE) of treatment as the average (*Ave*—over the whole of the population of interest) of the individual treatment effects, that is

$$\text{ACE} = \text{Ave}(Y_t - Y_c) \tag{2.2.16}$$

But, of course we can only ever observe either Y_t or Y_c, not both. We can never observe an individual treatment effect and therefore cannot obtain a direct estimate of the average. But we note that

$$\text{ACE} = \text{Ave}(Y_t) - \text{Ave}(Y_c) \tag{2.2.17}$$

and if we could obtain valid estimates of $\text{Ave}(Y_t)$ and $\text{Ave}(Y_c)$ then the problem would be solved. The only sure way of knowing that we have a valid estimate of these two averages (i.e. no selection biases) is to make sure that the allocation of treatment is completely random. Hence the Randomized Controlled Trial (RCT). Of course, we have to ensure that we have eliminated or reduced all other potential sources of bias (e.g. by masking treatment allocation from the people assessing outcomes) and one must not automatically assume that randomization has led to a perfect trial. In the critical appraisal of a clinical trial report, one might ask:

- What is the target population for the evaluation of the treatment, and were the trial participants representative of this population?
- Were the treatment and control conditions clearly defined and operationalized? In particular, was the control condition convincing?

- Were outcomes clearly defined and reliably measured?
- Was the trial adequately powered (i.e. was the number of randomized participants adequate for the job at hand)?
- Was randomization used to allocate treatment and how was the randomization actually implemented?
- Was it possible to blind mask treatment allocation from all those concerned with the assessment of outcomes? What evidence has been provided to indicate that masking has been effective?
- What methods were used to estimate the treatment effects and their precision? Were they fully compatible with the trial design (the randomization procedure)? Did the analysis take account of everyone randomized? In particular, did the analysis deal adequately with missing outcome data?
- If the trial used a non-standard design (randomization of groups of participants rather than of individuals; use of group-based therapies; and so on) were the power calculations and methods of statistical analysis appropriate for this design?

Estimation of treatment efficacy in the presence of non-compliance

What if the trial participants, despite giving their consent to be randomized, do not receive the treatment to which they were allocated? In a drug trial the participants may not take the tablets, or take less than the prescribed amount. They may even receive the active medication despite being allocated to the placebo (control) condition. In a psychotherapy trial they may not turn up to all or even any of the planned sessions of therapy. The standard method of analysis (and justifiably so) is based on the so-called Intention-to-Treat (ITT) principle: we analyse as randomized. We evaluate the effect of the offer of treatment as opposed to its receipt. But what do we do if we wish to evaluate (estimate) the effect of the treatment actually received? Two commonly used methods of analysis are called per Protocol and As Treated estimators. In the first method the effect of randomization is evaluated after first discarding the people who did not adhere to (or comply with) their allocated treatment. In the second, we compare the outcomes of all those who receive the treatment with all of those who received the control condition, regardless of their random allocation. Both methods are potentially flawed—they take no account of possible selection effects (confounding). A much better approach is to estimate the effect of randomization in that subgroup of participants who would always have complied with their treatment allocation whatever the outcome of the randomization. We call this subgroup the Compliers (the rest being non-compliers), and we aim to estimate the Complier-Average Causal Effect of Treatment (CACE).[20,21] In effect we are obtaining an ITT effect for the Compliers. Randomization ensures that, on average, the proportion (P) of Compliers is the same in each arm of the trial. The overall ITT effect is a weighted average of the ITT effect in the Compliers and the ITT effect in the non-compliers (the weights being P and 1-P, respectively). If we are prepared to assume that there is no direct effect of randomization on outcome (in particular, there is no effect of randomization on outcome in the non-compliers) then

$$\begin{aligned} \mathrm{ITT}_{\mathrm{All}} &= P.\mathrm{ITT}_{\mathrm{Compliers}} + (1-P).\mathrm{ITT}_{\mathrm{non\text{-}compliers}} \\ &= P.\mathrm{ITT}_{\mathrm{Compliers}} \end{aligned} \qquad (2.2.18)$$

It follows that

$$\mathrm{CACE} = \mathrm{ITT}_{\mathrm{compliers}} = \frac{\mathrm{ITT}_{\mathrm{All}}}{P} \qquad (2.2.19)$$

It is easy to show that in a trial in which no-one allocated to the control condition gets access to the (potentially) active treatment the proportion of Compliers (P) is estimated by the proportion of Compliers in the treatment group (this follows directly from randomization). In the situation in which participants allocated to the control condition can get access to treatment P is estimated by the difference between the two randomized groups with respect to the proportion of participants receiving treatment. Considering everyone randomized, eqn (2.2.19) is the ratio of the ITT effect on outcome to the ITT effect on receipt of treatment. Note that $\mathrm{ITT}_{\mathrm{All}}$ and CACE ($\mathrm{ITT}_{\mathrm{Compliers}}$) share the same null hypothesis (when one is zero, so is the other) and a significance test for one of them is equivalent to the corresponding significance test for the other. The estimation of the CACE is simply a way of adjusting the overall ITT estimate to allow to attenuation in the presence of non-compliance. It will not reveal a significant treatment effect when we already have a non-significant ITT estimate.

We do not advocate the replacement of ITT estimates by CACE estimation, but we do recommend that investigators supplement their primary ITT analyses with more detailed explanatory methods, particularly when there are situations in which non-compliance is strongly associated with subsequent loss to follow-up (i.e. missing outcome data). Valid ITT (and any other) estimates are particularly difficult to obtain in the presence of missing data. This is an active area of theoretical development in medical statistics, but the more easily understood methods are now beginning to make the transition to the more easily accessible clinical journals. For example, a relatively non-technical discussion of the potential application to mental health trials is provided by Dunn *et al.*[22,23]

We refer readers to Everitt and Wessely[24] and Dunn[25] for a much more detailed discussion of the methodological pitfalls for RCTs in psychiatry. One important point needs stressing again, however. It is vital that RCTs involve the randomization of sufficient numbers of subjects to be confident that the trial has sufficient power to detect the *minimum* treatment effect that still has *clinical* (as opposed to statistical) significance. This minimum effect should be defined in terms of its importance and *not* by naively observing the results of previous trials or pilot investigations.

Conclusions

The recently published *Encyclopedia of Biostatistics*[26] comprises of six large volumes of chapters such as this one, covering every area of conceivable interest to the statistically interested clinical research worker. Therefore it is inevitable that this chapter should be very restricted. Inevitably, the choice of topics might be thought to be rather idiosyncratic. Areas which might have been covered, but have been ignored, include survival modelling (particularly recent developments in so-called frailty modelling), longitudinal data analysis (with special reference to modelling patterns of attrition), genetics, and a whole range of classical multivariate methods such as principal components and factor analysis, discriminant analysis (although logistic modelling is one of the better methods of discriminant analysis), multidimensional scaling, and cluster analysis.

Henderson (see Chapter 2.7) has also mentioned exciting possibilities for the development of latent trait (item response theory) and latent class models. Several years ago I was often asked to teach trainee psychiatrists all they needed to know about statistics in two 2-h sessions. The first was to cover univariate methods, and the second, multivariate analysis. It cannot be done!

Further information

So, where should the reader go from here? Which are the most useful textbooks? In terms of general medical statistics, the obvious choice is Armitage, Berry, and Matthews. Everitt and Dunn and Everitt provide general introductions to multivariate methodology. Measurement error problems, including structural equation modelling, are covered by Dunn. The role of statistics in genetics is well covered by Sham. Although there are many texts on the use of statistics in psychology and education (e.g. Plewis which includes an introduction to multilevel modelling), the only specialist reference for psychiatrists appears to be that by Dunn.

Armitage, P., Berry, G., and Matthews, J.N.S. (2002). *Statistical methods in medical research* (4th edn). Blackwell Science, Oxford.

Everitt, B.S. and Dunn, G. (2001). *Applied multivariate data analysis* (2nd edn). Arnold, London.

Everitt, B.S. (1996). *Making sense of statistics in psychology*. Oxford University Press, Oxford.

Sham, P. (1998). *Statistics in human genetics*. Arnold, London.

Dunn, G. (1999). *Statistics in psychiatry*. Arnold, London.

Plewis, I. (1997). *Statistics in education*. Arnold, London.

References

1. Shavelson, R.J. and Webb, N.M. (1991). *Generalizability theory: a primer*. Sage, Thousand Oaks, CA.
2. Fleiss, J.L. (1987). *The design and analysis of clinical experiments*. Wiley, New York.
3. Streiner, D.L. and Norman, G.R. (2003). *Health measurement scales: a practical guide to their development and use* (3rd edn). Oxford University Press, Oxford.
4. Dunn, G. (2004). *Statistical evaluation of measurement errors: design and analysis of reliability studies* (2nd edn). Arnold, London.
5. Dunn, G. and Everitt, B.S. (1995). *Clinical biostatistics: an introduction to evidence based medicine*, Chap. 4. Arnold, London.
6. Kessler, R.C. (1994). The National Comorbidity Survey of the United States. *International Review of Psychiatry*, **6**, 365–76.
7. Jenkins, R., Bebbington, P., Brugha, T., *et al.* (1997). The National Psychiatric Morbidity Surveys of Great Britain—strategy and methods. *Psychological Medicine*, **27**, 765–74.
8. Pickles, A. and Dunn, G. (1998). Prevalence of disease, estimation from screening data. In *Encyclopedia of biostatistics*, Vol. 5 (eds. P. Armitage and T. Colton), pp. 3484–90. Wiley, Chichester.
9. Lehtonen, R. and Pahkinen, E.J. (1995). *Practical methods for design and analysis of complex surveys*. Wiley, Chichester.
10. Cochran, W.G. (1977). *Sampling techniques* (3rd edn). Wiley, New York.
11. Goldberg, D.P. and Williams, P. (1988). *A user's guide to the General Health Questionnaire*. NFER-Nelson, Windsor.
12. Begg, C.B. and Greenes, R.A. (1983). Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics*, **39**, 207–15.
13. Dunn, G. (1998). Compensating for missing data in psychiatric surveys. *Epidemiologia e Psichiatria Sociale*, **6**, 159–62.
14. Dunn, G., Pickles, A., Tansella, M., *et al.* (1999). The role of two-phase epidemiological surveys in psychiatric research. *The British Journal of Psychiatry*, **174**, 95–100.
15. Pickles, A., Dunn, G., and Vázquez-Barquero, J.L. (1995). Screening for stratification in two-phase ('two-stage') epidemiological surveys. *Statistical Methods in Medical Research*, **4**, 75–91.
16. Vázquez-Barquero, J.L., Garcia, J., Artal Simón, J., *et al.* (1997). Mental health in primary care: an epidemiological study of morbidity and use of health resources. *The British Journal of Psychiatry*, **170**, 529–35.
17. Rubin, D.B. (2007). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in Medicine*, **26**, 20–36.
18.. Luellin, J.K., Shadish, W.R., and Clark, M.H. (2005). Propensity scores: an introduction and experimental test. *Evaluation Review*, **29**, 530–58.
19. Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, **66**, 688–701.
20. Bloom, H.S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review*, **8**, 225–46.
21. Angrist, J.D., Imbens, G.W., and Rubin, D.B. (1996). Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association*, **91**, 444–72.
22. Dunn, G. Maracy, M., Dowrick, C., *et al*. Estimating psychological treatment effects from an RCT with both non-compliance and loss to follow-up. *The British Journal of Psychiatry*, **183**, 323–31.
23. Dunn, G., Maracy, M., and Tomenson, B. (2005). Estimating treatment effects from randomized clinical trials with non-compliance and loss to follow-up: the role of instrumental variable methods. *Statistical Methods in Medical Research*, **14**, 369–95.
24. Everitt, B.S. and Wessely, S. (2004). *Clinical trials in psychiatry*. Oxford University Press, Oxford.
25. Dunn, G. (2006). Psychotherapy for depression. In *Textbook of clinical trials* (2nd edn) (eds. D. Machin, S. Day, and S. Green), pp. 377–93. John Wiley & Sons, Chichester.
26. Armitage, P. and Colton, T. (eds.) (1998, 2nd edn 2005). *Encyclopedia of biostatistics*. Wiley, Chichester.

2.3

The contribution of neurosciences

Contents

2.3.1 Neuroanatomy

†R. C. A. Pearson

Introduction

The symptoms, signs, and syndromes of psychiatry, whether organic or biological psychiatric disease or not, in the main reflect alterations in functions which reside in the cerebral cortex, including the limbic lobe, and those structures and pathways closely related to the cortex. These cortical manifestations of psychiatric disease include alterations in thought, language, perception, mood, memory, motivation, personality, behaviour, and intellect. Therefore, this brief account of brain structures and pathways that are important in psychiatry will concentrate on the cerebral cortex and related structures and pathways. Readers who require a fuller account of central nervous system anatomy are referred to the many standard texts, which give a more complete coverage of the subject.

Broadly speaking, neuroanatomy can be subdivided into two parts—the topographical organization of the brain and spinal cord, and the anatomical connections forming functional pathways in the central nervous system. The former is of vital importance clinically, since pathologies rarely respect the boundaries of functional systems, and knowledge of the spatial relationships of different brain structures is increasingly useful as modern imaging methods more accurately visualize detailed brain structure *in vivo*. However, it is the second subdivision of the subject which makes the greater contribution to understanding the biological basis of psychiatric disease, and it is this that will be at the centre of the present account.

The structure and organization of the cerebral cortex

The lobes of the cerebral cortex

A variable pattern of fissures (sulci) and folds (gyri), many of which have specific names, extensively groove the surface of the cerebral hemisphere. A few are relatively constant and are used to subdivide the cerebral hemisphere into lobes, named for the bones of the skull which they underlie (Fig. 2.3.1.1).

The deep lateral sulcus, also called the Sylvian fissure, extends from the uncus, anteriorly and medially, to the parietal lobe, posteriorly and medially. It has a short stem, and anterior, ascending, and posterior rami. The anterior and ascending rami embrace the pars triangularis of the frontal lobe, which houses Broca's motor speech area. The much longer posterior ramus is used in defining the lobes of the hemisphere. The central sulcus is prominent approximately midway along the anteroposterior extent of the

† Dr. Pearson died while this new edition was being prepared. The editors pay tribute to his scientific achievements and to his contributions to this book.

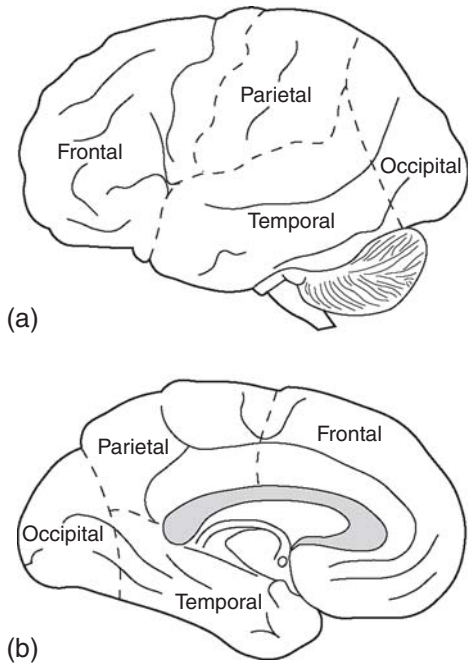


Fig. 2.3.1.1 The lobes of the cerebral cortex.

lateral surface of the hemisphere and, most commonly, extends over the medial margin, where its inferomedial tip is embraced by the U-shaped paracentral lobule. On the lateral surface, it passes from the medial margin, forwards and laterally, to reach the lateral sulcus. The line of the central sulcus closely approximates the line of the coronal suture of the adult skull, i.e. the junction between the frontal and parietal bones; consequently, the sulcus separates the frontal and parietal lobes. The demarcation of the occipital lobe is the parieto-occipital sulcus dorsally and medially, and the pre-occipital sulcus ventrally and laterally, with an imaginary line connecting the two and intersecting the posterior tip of the lateral sulcus. The temporal lobe lies anterior to this line and inferior (ventral) to the lateral sulcus. The deep lateral sulcus broadens out at its fundus, with an area of cortex forming the extensive floor of the sulcus, particularly in its anterior two-thirds. This cortex is the insula, which does not form part of any of the lobes mentioned above. The insula is surrounded by the circular sulcus, and is overhung by the frontal and parietal opercula superiorly, and the temporal operculum inferiorly (ventrally). The anatomical borders of the lobes of the cerebral cortex, and other sulcal and gyral landmarks, are only loosely paralleled by functional boundaries. However, lobar terminology is so firmly embedded in clinical and non-clinical neuroscience that consideration of their anatomical features is essential.

(a) The frontal lobe (Fig. 2.3.1.2)

The precentral gyrus, immediately in front of the central sulcus and continuing onto the medial surface, contains the primary motor cortex. The precentral sulcus usually defines the anterior boundary, and in front of this lies the premotor cortex. The inferior

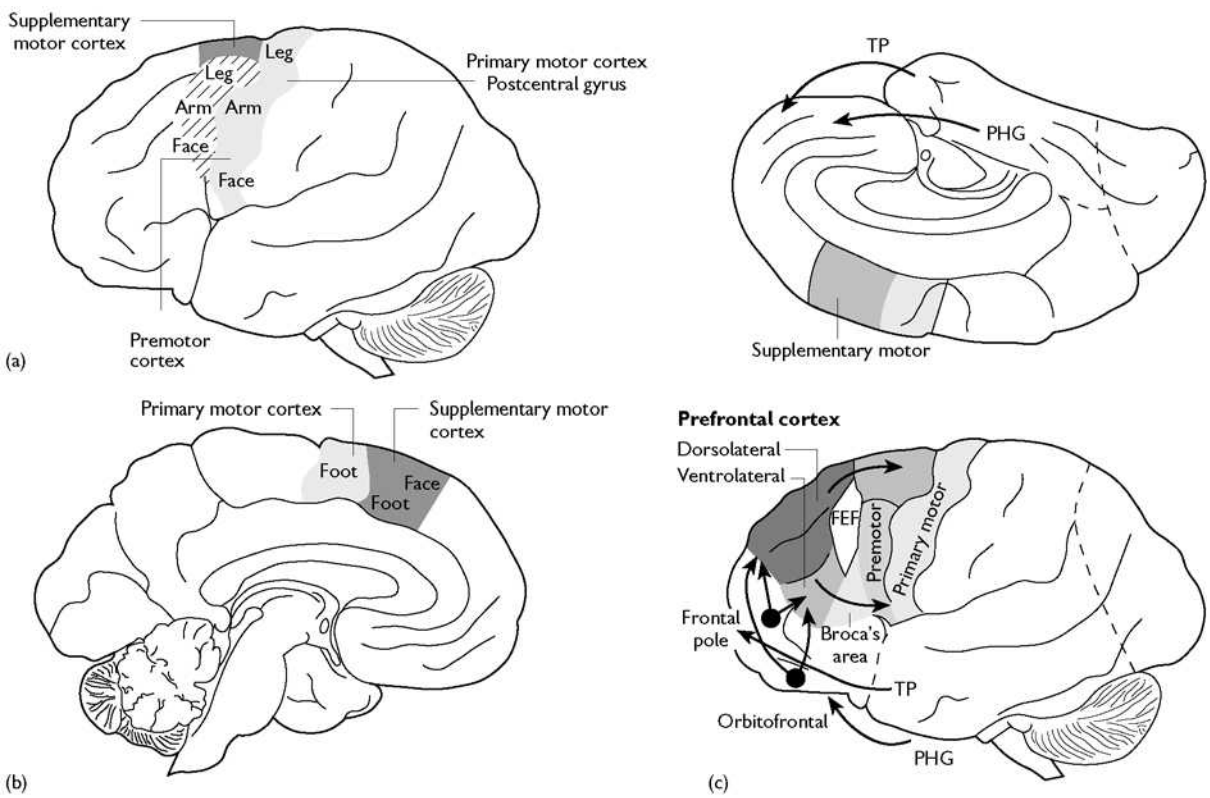


Fig. 2.3.1.2 Motor areas of the frontal lobe: TP, temporal pole; PHG, parahippocampal gyrus; FEF, frontal eyefields.

margin of the sulcus runs into the pars triangularis, which includes Broca's motor speech area. On the medial margin and surface, the cortex includes the supplementary motor area. The lateral prefrontal cortex, in front of these motor and associated areas, is usually grooved by two major horizontal sulci, defining the superior, middle, and inferior frontal gyri. The cortex on the medial surface of the frontal lobe anterior to the prefrontal gyrus forms the medial prefrontal cortex. The concave inferior surface of the frontal lobe, overlying the bony orbit *in vivo*, is the orbitofrontal cortex.

(b) The parietal lobe (Fig. 2.3.1.3)

Behind the central sulcus, the primary somatic sensory cortex (SI), extending onto the medial surface to occupy the posterior part of the paracentral lobule, where the sacral spinal segments are represented, occupies the postcentral gyrus. The postcentral sulcus limits the postcentral gyrus posteriorly. Behind this, the sulcal pattern is variable, but one or more horizontal intraparietal sulci divide the lobe into superior and inferior parts. The second somatic sensory cortex (SII) is located in the parietal operculum, close behind the inferolateral tip of the central sulcus. Specific sulcal patterns in the transition region between the parietal lobe and the occipital and temporal lobes (the supramarginal and angular gyri) are important landmarks for the detailed localization of language functions.

(c) The occipital lobe (Fig. 2.3.1.4)

The occipital lobe is predominantly involved in vision and visual perception. The medial surface is grooved by the deep horizontal calcarine sulcus, which typically reaches the posterior pole of the hemisphere. Within its walls is the primary visual cortex. This area (area 17 of Brodmann) is often called the striate cortex; in the freshly sliced brain, a thin band of white matter, the stria of Gennari, is clearly visible running in the centre of the cortical grey ribbon. The extent of this stria precisely demarcates the primary visual cortex. Surrounding areas of the medial and lateral surface, the prestriate and peristriate cortex, contain some of the numerous separate visual association areas.

(d) The temporal lobe (Figs 2.3.1.5 and 2.3.1.6)

Two horizontal sulci, the superior and inferior temporal sulci, divide the lateral surface of the temporal lobe into the superior, middle, and inferior gyri. The latter, extending onto the inferior surface, is also known as the inferotemporal cortex. On the medial surface, the collateral sulcus runs from close to the temporal pole to the calcarine sulcus posteriorly. Medial to this is the parahippocampal gyrus. Anteriorly, this curves dorsally and caudally to form the hook-shaped uncus. The entorhinal cortex occupies approximately the anterior third of the parahippocampal gyrus. Lateral to this, in the walls of the rhinal sulcus, lies the perirhinal cortex. The uncus closely overlies the amygdala; the primary olfactory cortex, the piriform cortex, lies immediately in front. The choroid fissure limits the parahippocampal gyrus medially. Passing into the floor of the lateral ventricle, the subicular areas of cortex lead to the hippocampus proper. A detailed consideration of the anatomy of the hippocampus is given below.

The cortex of the temporal operculum contains Heschl's gyri, within which lies the primary auditory cortex. Diverse auditory

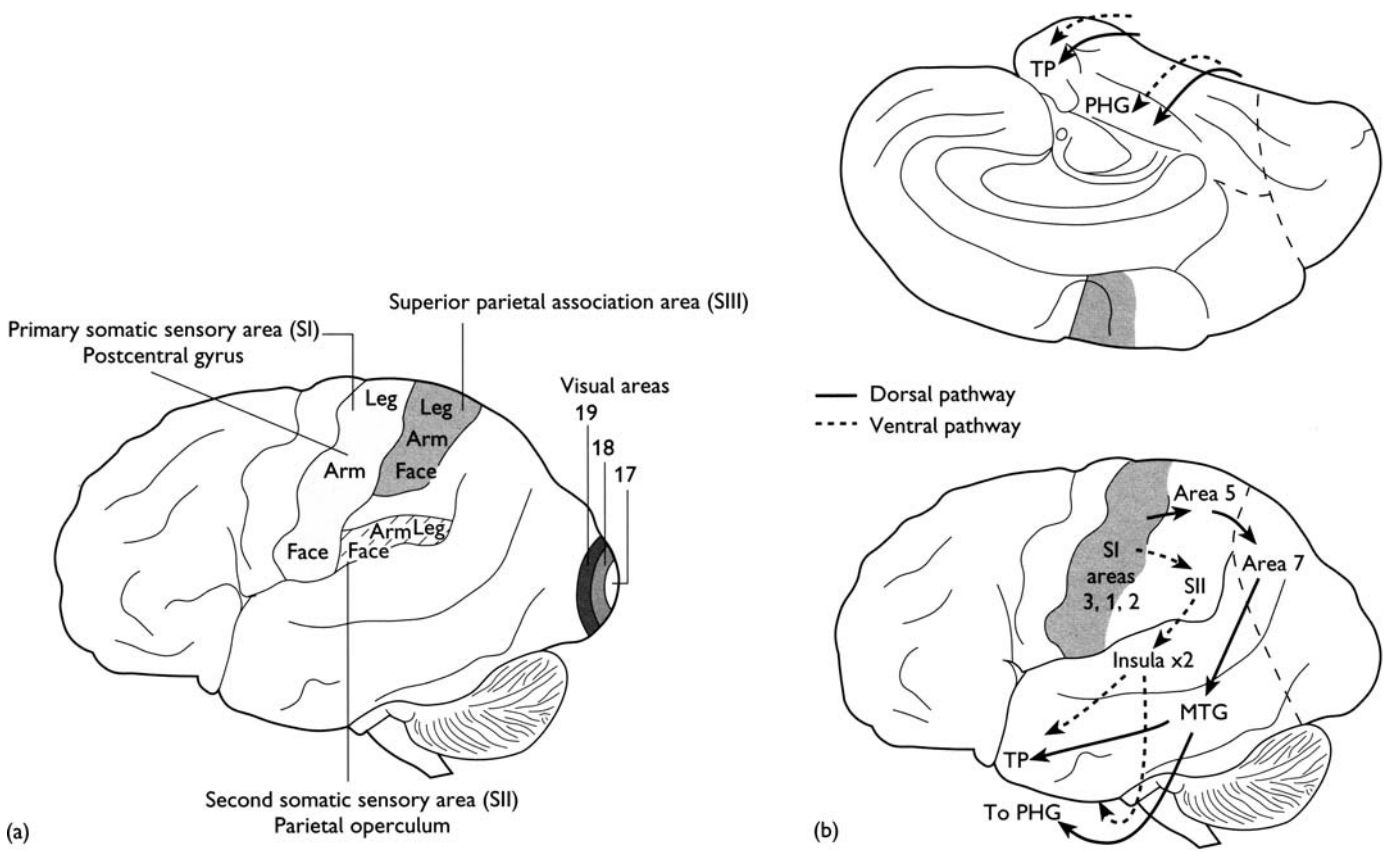


Fig. 2.3.1.3 (a) Areas of the parietal lobe; (b) somatic sensory association pathways. TP, temporal pole; PHG, parahippocampal gyrus; MTG, middle temporal gyrus.

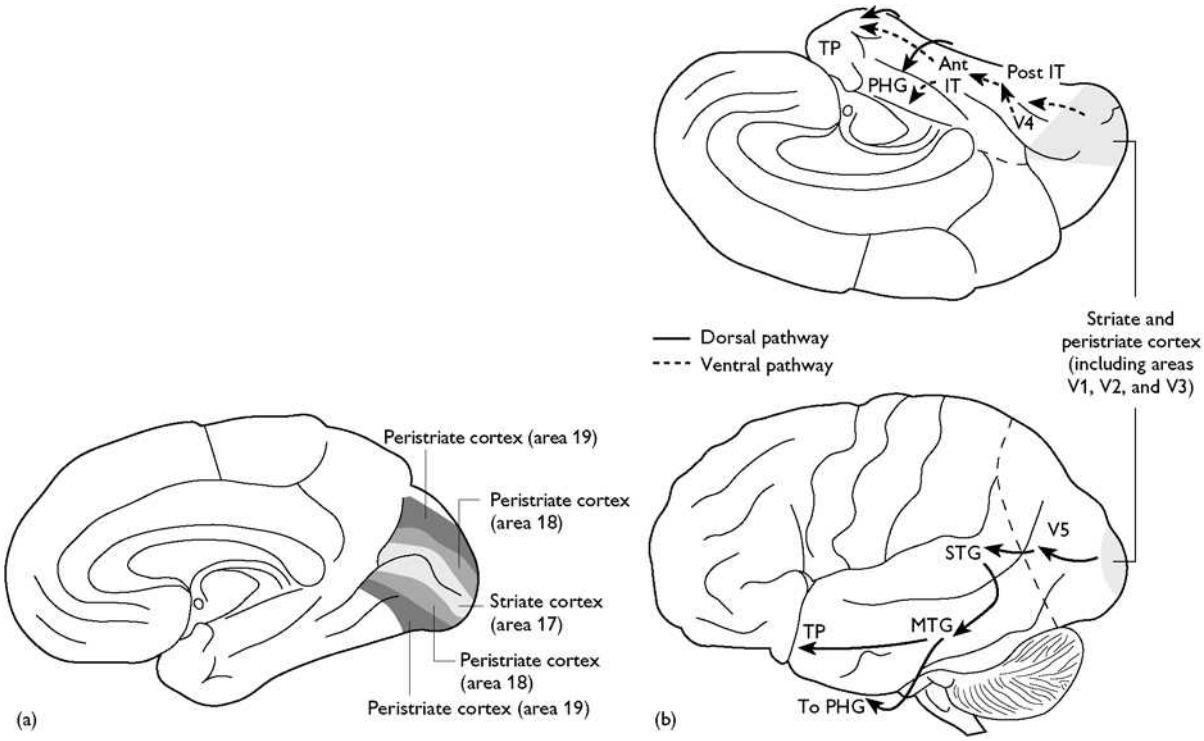
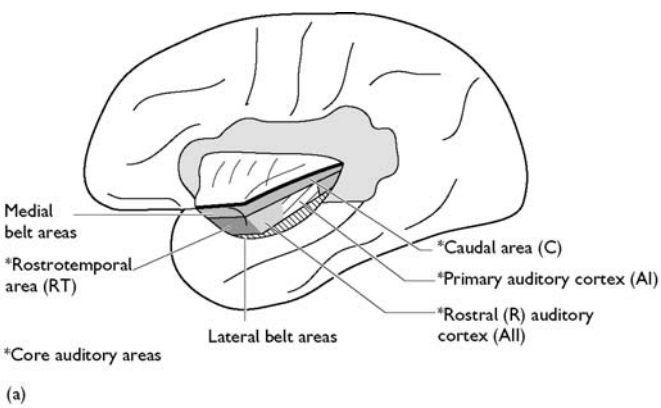


Fig. 2.3.1.4 The occipital lobe: (a) visual areas; (b) visual association pathways. TP, temporal pole; IT, inferotemporal cortex; PHG, parahippocampal gyrus; STG, superior temporal gyrus; MTG, middle temporal gyrus; Ant, anterior; Post, posterior.



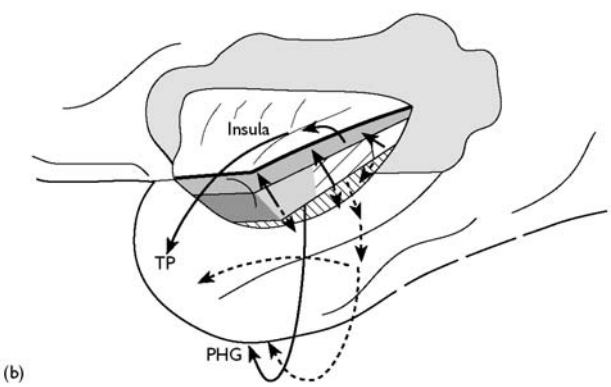
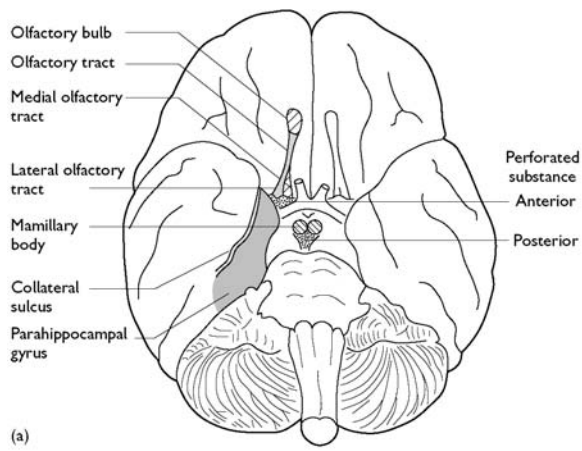


Fig. 2.3.1.5 Auditory association connections. TP, temporal pole; PHG, parahippocampal gyrus.



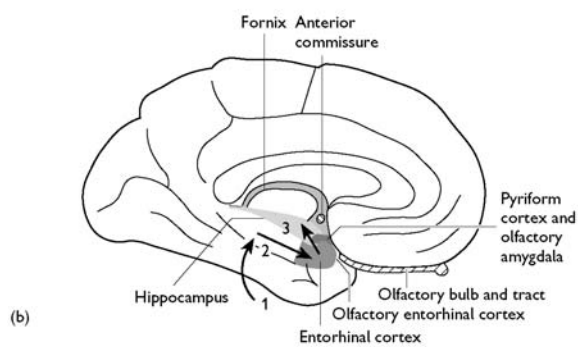


Fig. 2.3.1.6 Olfactory and limbic structures: (a) ventral surface of the brain; (b) medial temporal lobe. 1, Association cortex to parahippocampal gyrus; 2, parahippocampal gyrus to entorhinal cortex; 3, entorhinal cortex to hippocampus (perforant path).

association areas surround this, extending into the superior temporal gyrus. The functions of the cortex of the middle temporal gyrus are uncertain, but include complex visual, auditory, and somatic sensory association areas. The inferotemporal cortex is largely concerned with visual perception and cognition. The anatomical pathways that underlie these functions are considered below.

(e) The insula

The anterior margin of the insula, where the cortex becomes continuous with the anterior perforated substance, is known as the limen insulae. Above and below, where the insular cortex rolls round onto the opercula, lies the circular sulcus, the superior and inferior rami of which fuse posterosuperiorly to form the apex of the insula. Several variable sulci mark the insula, but little is known of the functional subdivisions of this cortex; gustatory, somatic sensory, and auditory areas have been described.

The structure of the neocortex

The neocortical grey matter is usually described as having six layers (Plate 1(a)). Wide variation in the nature of this microscopic lamination underlies the subdivision of neocortex into a multiplicity of (usually numbered) areas. At its simplest, two types of neurones make up the grey matter—pyramidal and non-pyramidal (or granular) cells. An apparent predominance of one or other type gives the extremes of granular and agranular cortex, equating with sensory areas (granular) and the motor cortex (agranular). In fact, the proportion of different cell types is constant in all areas. Indeed, with the single exception of the primary visual cortex, the numbers of neurones under a fixed surface area is also constant in all cortical areas. Variations in the size of pyramidal cells in particular lead to an apparent change in proportions. These variations probably reflect differences in the axonal volume of individual pyramidal cells, reflecting the distance and volume of projection fibres from a cortical area.

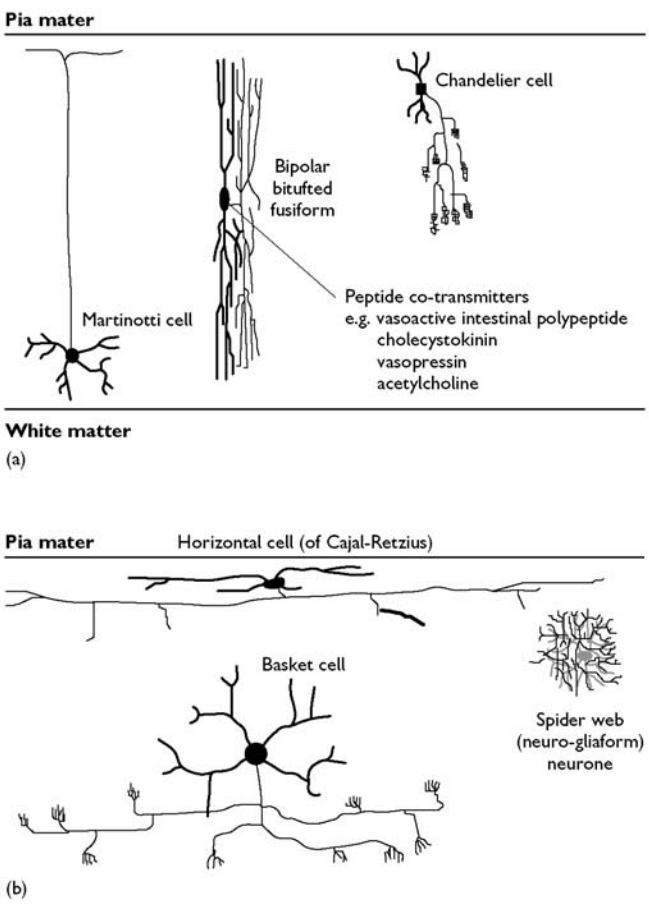


Fig. 2.3.1.7 Inhibitory (GABA-ergic) neurones of the neocortex: (a) vertical; (b) horizontal.

Pyramidal cells have a single main apical dendrite ascending towards the pial surface, and several horizontally spreading basal dendrites. All dendrites bear dendritic spines, which receive synapses (Plate 1(b)). All pyramidal cells use excitatory amino acids as neurotransmitters and have axons which enter the subcortical white matter; hence they are all projection neurones. They constitute approximately 60 per cent of all the neurones in the cortex. A second spiny neuronal type, the spiny stellate cells, is the next most numerous. These also use an excitatory amino acid, most probably glutamate, as their neurotransmitter. Unlike pyramidal cells, however, their axons remain confined to the cortical grey matter; they are interneurones, accounting for a further 25 per cent of cortical neurones. All the other neurones are inhibitory interneurones, using γ-aminobutyric acid (GABA) as their major neurotransmitter. Many also contain one or more neuropeptides, and their content of specific calcium-binding proteins varies. They have a wide range of axonal and dendritic forms, and have been multiply classified in the past. Broadly speaking, they can be grouped into those with horizontal axonal arborizations, those whose axons ramify at right angles to the pial surface, i.e. through the depth of the cortex, and those with radial axons (Fig. 2.3.1.7).

The structure of the allocortex

The allocortex comprises a number of different areas, all with very different structures. They are either limbic or olfactory (or both) and are found predominantly in the medial temporal lobe. The largest of these regions is the hippocampus. Essentially, this includes the three-layered cortex of the hippocampus, together with the transitional areas between it and neocortex, which are variably said to have three, four, five, or six laminae. The hippocampal formation comprises the dentate gyrus, Ammon's horn (CA fields), and the subiculum. Both the dentate gyrus and the CA fields have a prominent single layer of neurones, with an overlying molecular layer and a subjacent polymorphic layer. In the dentate gyrus, the cells are granule cells, whereas in the CA fields they are predominantly large pyramidal cells—the stratum pyramidale. Both cell types are excitatory and are projection neurones. Scattered populations of GABA-ergic inhibitory interneurones are found immediately subjacent to the main cellular laminae and in the molecular layers. The CA fields are numbered 1, 2, and 3, from the subiculum to the dentate gyrus. The subiculum is the zone of transition between the three-layered hippocampus proper, and the entorhinal cortex and cortex of the parahippocampal gyrus laterally. It is sometimes further subdivided into subzones including the presubiculum, between the lateral cortex and the subiculum, and the prosubiculum, between the subiculum proper and CA1 of the hippocampus. The histological appearance of the hippocampus and adjacent areas are shown in Fig. 2.3.1.8, and their connections are considered in detail below.

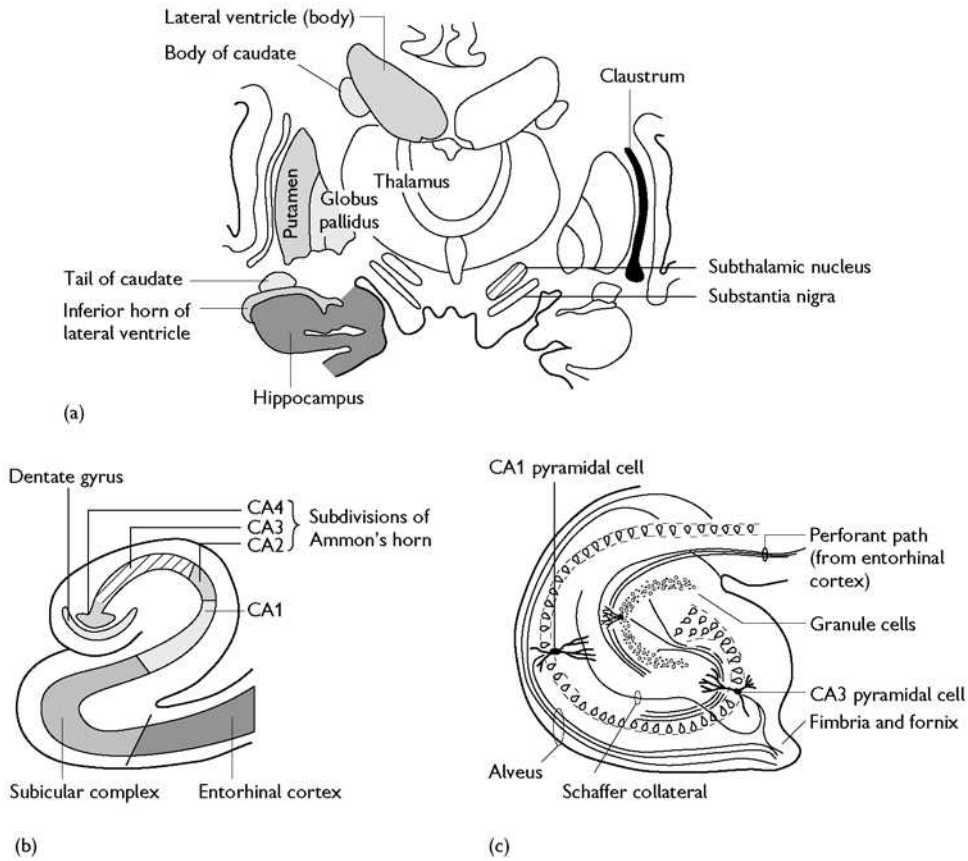


Fig. 2.3.1.8 (a) Hippocampal formation in the temporal lobe; (b) subdivisions of the hippocampal formation; (c) organization of the hippocampus.

Table 2.3.1.1 General connections of the cerebral cortex

Nucleus/area	Afferents	Efferents
Subcortical		
Thalamus		
Principal	+	+
Intralaminar	+	+
Locus coeruleus (NA)	+	–
Raphe (serotonin)	+	–
Basal nucleus (ACh)	+	–
Hypothalamus (histamine)	+	–
SN and VTA (dopamine)	+	–
Claustrum	+	+
Striatum	–	+
Pons	–	+
Superior colliculus and reticular formation	–	+
Specific (e.g. corticospinal)	–	(+/–)
Cortical		
Contra lateral		
Homotopic	+	+
Heterotopic	+	+
Ipsilateral		
Short	+	+
Long	+	+

NA, noradrenaline; ACh, acetylcholine; SN, substantia nigra; VTA, ventral tegmental area.

The general pattern of connectivity of the cortex

All cortical areas share a broadly similar pattern of connections. What differs is the relative quantity of connections in each category, as well as the precise detail of origin and termination. The broad categories of connections which all cortical areas share are most easily seen in a table (Table 2.3.1.1). These will be dealt in detail in subsequent sections.

A central feature of the connections of the neocortex is their organization relative to the cortical laminae (Plate 1(a)). For simplicity, it can generally be assumed that the forward-flowing stream of connections for perception and action use the central laminae, descending projections originate in deeper layers, and feedback and non-specific afferents terminate in both superficial and deep laminae. Thus, in primary sensory areas the thalamic input from the principal nuclei (e.g. the lateral geniculate nucleus in the visual cortex) terminates in layer IV. Projections to higher association areas originate from pyramidal cells immediately adjacent, in layer III. As the ipsilateral association pathway progresses from area to area, the association fibres terminate in this central region, and the thalamic afferents end in superficial and deep laminae. Projections to subcortical nuclei, such as the striatum, arise in layer V, and corticothalamic axons come from layer VI. It is as if the input of major importance, requiring detailed, focused, and faithful transmission, arrives in layer IV and is passed on from the deeper part of layer III. Inputs which moderate or modulate this pass to either side of the central layers, and outputs which are not directly part of this progression arise from deeper layers.

There is a general principle in the organization of connectivity with the cerebral cortex of overlapping connectivity of functionally related areas throughout the range of connections made. In other words, functionally related cortical areas, which are connected with each other, also tend to have overlapping or interdigitating connections with other structures. This is seen on the striatum, the thalamus, the claustrum, the cholinergic basal forebrain, and the pontine nuclei.

Subcortical afferents to the cerebral cortex

The thalamus (Plate 2)

Thalamic nuclei can be classified as specific (principal) or non-specific. In general, the specific nuclei degenerate completely when the cortex is removed, whereas the non-specific nuclei do not. This is because the sole major projection of the specific nuclei is to the cerebral cortex. The non-specific nuclei are the intralaminar and midline nuclei, which project to the basal ganglia as well as to cortex, and the reticular nucleus, which projects only to other thalamic nuclei.

All cortical areas receive afferents from at least one specific thalamic nucleus, and an additional input from the intralaminar/midline nuclei. Corticothalamic fibres reciprocate all thalamocortical projections, probably without exception.

(a) Specific nuclei

The cortical projections of the specific nuclei are shown in Plate 2. Subclassification of the specific thalamic nuclei depends upon their afferent subcortical connections. The primary relay nuclei receive the major sensory pathways—the lateral and medial geniculate nuclei for the optic tract and auditory pathway respectively, and the ventral posterior nucleus for the somatic sensory pathways. The secondary relay nuclei are those which receive a known major subcortical pathway which is not sensory. These are usually taken to include the caudal subdivision of the ventral lateral nucleus, which receives the cerebellar pathway, the anterior division of the ventral lateral nucleus and the ventral anterior nucleus, which both receive fibres from the internal segment of the globus pallidus, and the anterior nuclei, which receive the mammillothalamic tract. The association nuclei are those which were not previously understood to receive a major subcortical pathway, predominantly the medial nucleus and the pulvinar. This view is no longer really tenable, since the pulvinar is now known to receive a major input from the superior colliculus, and the various subdivisions of the medial nucleus receive fibres from the ventral pallidum and the olfactory pathway, amongst others. The subcortical afferents to the principal thalamic nuclei are shown in Plate 2(a).

(b) The non-specific nuclei

The reticular nucleus lies lateral to the main body of the thalamus, separated from it by the external medullary lamina. Cells in the nucleus are inhibitory, using GABA as their neurotransmitter. Excitatory thalamocortical and corticothalamic fibres, passing to and from the main nuclei, traverse the reticular nucleus and give axon collaterals to the reticular nucleus cells. There is a very tight arrangement, whereby the branches of axons from the thalamus to a particular cortical area terminate in the reticular nucleus in close proximity to the corticothalamic axons from the same cortical area. The reticular nucleus cells in turn send their inhibitory axons into the thalamus, to terminate precisely in the nuclei from which they receive a collateral input.

The intralaminar and midline nuclei receive afferents from the major pathways to their adjacent principal nuclei and project to the striatum (caudate, putamen, and ventral striatum) as well as to the cortex. There is a broad topographic relationship in their projection to both areas. The midline nuclei innervate the ventral striatum (nucleus accumbens) and the limbic cortex, including the hippocampal formation. The anterior intralaminar nuclei connect with the prefrontal, parietal, occipital, and temporal cortex, and project to all parts of the caudate. The posterior intralaminar nuclei, of which the centromedian nucleus is the largest, project to the putamen and are reciprocally related to the motor, premotor, and supplementary motor areas of cortex in the frontal lobe. Many axons to the cortex from the intralaminar nuclei are collateral branches of thalamostriate axons. The cortical and striatal areas to which they each project are precisely those parts of the cortex and striatum which are themselves directly connected by corticostriate fibres.

Non-thalamic subcortical afferents to the cerebral cortex

A variety of nuclei send afferents to the cerebral cortex. These fall into two categories: fibres from the so-called isodendritic core of the brainstem and basal forebrain, which are non-reciprocal and aminergic, and the two-way interconnections with the claustrum.

Acetylcholine and the basal forebrain

A system of cholinergic nuclei extend from the septum verum anteriorly, through the nuclei of the diagonal band of Broca to the basal nucleus of Meynert in the substantia innominata, ventral to the globus pallidus, most posteriorly. From these, cholinergic fibres pass to the entire cerebral cortex. Alternative nomenclature for these nuclei uses a numeric system, (Ch1 to Ch4). The anterior cell groups project to the hippocampus and entorhinal cortex, and to the olfactory bulb and cortex. There is an approximate topographic relationship in the projection of the basal nucleus to the neocortex, but with considerable overlap; adjacent or overlapping regions of the nucleus project to widely separated but functionally related and interconnected areas of neocortex. Degeneration of this system is associated with the dementias of Alzheimer's disease and Lewy body disease.

Cholinergic cells in the pedunculopontine nucleus of the midbrain project to the thalamus, particularly to the midline and intralaminar nuclei.

(a) Serotonin, noradrenaline (norepinephrine) and adrenaline (epinephrine): the raphe and associated nuclei

The median group of nuclei of the brainstem reticular formation is made up largely of the various raphe nuclei, all of which use serotonin (5-hydroxytryptamine) as their neurotransmitter. Their projections are very widespread throughout the central nervous system. Broadly speaking, there is a rostrocaudal topography to the efferent projections of these nuclei. The most rostral, notably the dorsal median raphe nucleus in the midbrain tegmentum, send projections to the entire cerebral cortex and striatum. There is a prominent projection to the thalamus, notably to the midline nuclei which in turn project in part to the hippocampal formation.

The serotoninergic raphe-spinal tract, which has important functions in the gating of pain, arises from the most rostral nuclei, the raphe obscurus and the raphe magnus.

The locus coeruleus lies in the dorsolateral pons, immediately deep to the ventricular ependyma. Together with the subceruleus immediately deep to it, this nucleus supplies noradrenergic fibres to most of the central nervous system. Ascending fibres pass to the thalamus and hypothalamus, the entire cerebral cortex (neo- and allocortex), the amygdala, the septal nuclei, and the olfactory bulb. Adrenergic cells in the brainstem do not appear to send ascending projections to the cortex.

(b) Dopamine: the substantia nigra and adjacent nuclei

The major dopaminergic cell group of the midbrain is the pars compacta of the substantia nigra, which projects to the striatum (caudate, putamen, nucleus accumbens, and olfactory tubercle). Two other adjacent nuclei, the ventral tegmental area of Tsai and the pigmented parabrachial nucleus, are also dopaminergic. Together, these three cell groups project rostrally in the medial forebrain bundle to innervate the thalamus and hypothalamus, the hippocampal formation, the entorhinal cortex and amygdala, and widespread areas of the neocortex, especially the prefrontal, orbitofrontal, and cingulate cortex.

(c) Histamine and the posterior hypothalamus

The entire cerebral cortex, including the limbic lobe, receives a histaminergic projection from the tuberomamillary nucleus of the posterior hypothalamus. Postsynaptic receptors appear to be of two types, H1 and H2, with broadly opposing effects.

(d) The cortex and the claustrum

The claustrum is a thin plate of grey matter lying immediately deep to the cortex of the insula, and separated from it by the white matter of the extreme capsule. Medially, the external capsule (Fig. 2.3.1.8(a)) separates it from the putamen. The nucleus receives fibres from and projects to the whole cortex, including the allocortex. The reciprocal connection is topographically organized, but with overlapping zones projecting to widely separated but functionally related and interconnected cortical areas. Many of the neurones of the claustrum have branching axons with collaterals going to two or more such interconnected areas. For example, the superior parietal cortex (area 5) is widely separated from the premotor cortex (area 6) but is connected to it by ipsilateral association fibres, and the two areas are functionally closely related. Both these areas project to and receive from an overlapping zone of the claustrum, and many axons of claustral cells in this zone may branch to both areas.

(e) Modulation of cortical activation and the anatomy of the reticular activating system

Specific information to the cerebral cortex, for example relating to sensory stimuli in the periphery, is relayed via the main thalamic nuclei. The other, diffusely projecting, systems are most likely to be involved in the regulation of cortical responsivity. Such a role has been demonstrated electrophysiologically for the claustrum, and the pharmacology of the antihistamines indicates a role for this transmitter system in the regulation of cortical arousal. The cholinergic input from the basal forebrain is necessary for the proper functioning of the cortex, and its degeneration is associated with cognitive decline and memory impairment. The possible relationship of mesolimbic dopamine pathways to schizophrenia is well known. Similarly, the psychopharmacology of serotonin also implies a major role for this transmitter system in the proper functioning of the cortex. There are two routes by which these 'non-specific' pathways affect the cortex: direct projections, and an indirect pathway through the thalamus. Brainstem nuclei send fibres to the intralaminar and midline nuclei, which in turn send fibres to the entire cortex including the hippocampus. The cholinergic input from the interpeduncular nucleus to the intralaminar nuclei is prominent. Serotonin is particularly concentrated in the midline nuclei. There are other reticular formation projections to these nuclei, but the transmitters remain uncertain. This latter indirect route by which the reticular formation of the brainstem impinges on the cerebral cortex via the thalamus constitutes the reticular activating system. The role of this system in cortical arousal is well documented.

Corticocortical connections

(a) The corpus callosum and the commissural connections of the cerebral cortex

All cortical areas, both send fibres to and receive fibres from the opposite hemisphere, although the connections are not necessarily throughout the whole area. In most of the cortex, commissural fibres pass to the contralateral side in the corpus callosum. The anterior commissure carries fibres that interconnect the anterior third or so of the temporal lobe with its partner, as well as fibres that interconnect the olfactory bulbs on each side. Some fibres in the fornix cross the midline, the so-called commissure of the fornix, to interconnect the two hippocampal formations. Commissural fibres are of two types, homotopic and heterotopic. Homotopic fibres pass from one area of cortex to the same area on the other side. Heterotopic fibres pass from one area to a different, although, often functionally related, area on the other side. As a generalization, the functions of the commissures can be subdivided into two categories. First, they serve to interconnect representations of the contralateral sensory surround across the midline, for example, the representations of the two halves of the body, the two visual hemifields, and so on. In areas containing a lateralized sensory representation, of either the body or the visual field, the callosal fibres are confined, in both origin and termination, to the parts of the area containing a representation of midline and adjacent regions. Thus the representation of the trunk in somatic sensory areas sends and receives commissural fibres, whereas the hand and foot representations are not connected across the midline. Similarly, the vertical meridians in visuotopic representations are interconnected by callosal fibres, whereas the periphery is not. In contrast, the second function of the commissures is to connect areas in one hemisphere with areas on the other side, where the functions of each are represented on only one side, i.e. they are lateralized. Of course, this is most apparent for language areas; for example, objects held in the non-dominant hand cannot be named following callosal section because the sensory cortex of the non-dominant hemisphere cannot communicate across the midline with the language and speech areas in the dominant hemisphere.

(b) Ipsilateral corticocortical association connections

All cortical areas interconnect with other areas in the same hemisphere. The primary sensory areas are in the parietal, occipital, and temporal lobes. From these, parallel pathways emanate in an

approximately hierarchical sequence through multiple areas in the adjacent association cortex, passing towards the medial temporal lobe where all pathways converge on the parahippocampal gyrus and the cortex of the temporal pole. In general, three 'tiers' of association areas can be recognized in this sequence; the first tier receives from the core sensory areas, and the third projects into the temporal pole and the parahippocampal gyrus. Connections passing in the direction of the medial temporal cortex from the primary sensory areas have a feedforward pattern of termination, whereas the reciprocal connections have a feedback character. Although the interconnections of the multiple areas along this sequence of connections are complex, a broad pattern common to all the sensory pathways can be discerned. Essentially, each sensory modality has a core zone in the cortex, comprising three areas. Each of these is linked to the relevant main thalamic nuclei and contains a complete representation of the sensory surround. They are linked together by short association connections passing forwards from the first thalamo-recipient zone in a stepwise fashion to the other two. Two streams of connections, dealing with different aspects of sensory perception, emanate from these into the surrounding association areas. One of these, the ventral stream (the 'stimulus-relevant' or 'what' system), is primarily concerned with a detailed perception and characterization of the stimulus. The second, the dorsal stream (the 'self-relevant' or 'where' system), is concerned mainly with spatial location, particularly in extrapersonal space. Each area in this sensory hierarchy is reciprocally connected to a part of the frontal lobe, where the dual-pathway streams are represented by a dorsal and a ventral prefrontal subdivision, feeding onto the supplementary and premotor cortex, and so to the motor cortex.

(c) Cortical association pathways for vision (Fig. 2.3.1.4)

The three core visual areas are V1 (the primary visual cortex, striate cortex), V2 which surrounds V1 and is contained within Brodmann's area 18, and outside this the V3 complex which is probably still within area 18. From these, feedforward projections pass to areas in the inferior parietal and superior temporal cortices, as the dorsal stream, and to V4, within Brodmann's area 19, at the junction of the occipital association cortex with the inferior temporal gyrus on the medial surface, forming the ventral pathway. Areas in the superior temporal sulcus and middle temporal gyrus form the next two tiers of association cortex for the dorsal pathway. The ventral route progresses via the posterior inferotemporal cortex on to the anterior inferotemporal cortex. Both the anterior inferotemporal and middle temporal cortical areas at the distal end of these two visual association pathways project to the cortex of the temporal pole, and to the parahippocampal gyrus. The parahippocampal gyrus projects to the entorhinal cortex more anteriorly, and the temporal pole connects with the amygdala. The entorhinal cortex provides a major input to the dentate gyrus of the hippocampus via the perforant path as seen in the figure. Similarly, the amygdala projects into the hippocampal formation.

(d) Cortical association pathways for somatic sensation (Fig. 2.3.1.3)

The three core areas for the somatic sensory cortex are Brodmann's areas 3, 1, and 2 on the postcentral gyrus, together constituting the classical primary somatic sensory cortex. Each receives a major input from the ventral posterior nucleus of the thalamus and contains a complete representation of the body. Area 3 projects forwards to areas 1 and 2, and area 1 projects similarly to area 2. All three areas project in a feedforward fashion to the second somatic sensory area in the parietal operculum. This represents the first step on the ventral association pathway. The dorsal pathway begins with the projection of all subdivisions of the primary somatic sensory cortex to Brodmann's area 5 in the superior parietal cortex. Further steps along the ventral pathway are areas in the insula. In the ventral pathway, information flows from the superior to the inferior parietal cortex, and so to the middle temporal gyrus. Areas at the ends of both these paths project onto the parahippocampal gyrus and the temporal pole.

(e) Cortical association pathways for hearing (Fig. 2.3.1.5)

The association pathways from the auditory cortex are less well understood. The three core areas lie along Heschl's gyrus in the temporal operculum, with the classical primary auditory area (AI) most posterior, the rostral auditory area (AII) more anterior, and the rostrotemporal auditory area in front of that. Each of these contains a complete representation of the auditory surround, organized tonotopically, and connects with the medial geniculate nucleus. On either side of these lie multiple auditory association areas, termed the medial and lateral belt areas, representing the first steps along the two association streams of connections, although it is unclear which is 'dorsal' and which is 'ventral' in a functional sense. The medial belt areas have connections with parts of the insula, whereas the lateral belt areas project into the association cortex of the superior temporal gyrus. The first association pathway probably continues through areas of the anterior part of the superior temporal gyrus, and so to the parahippocampal gyrus and the temporal pole. The second continues from the superior to the middle temporal gyrus, and thence to the parahippocampal cortex and the temporal pole.

(f) The olfactory pathway to the cerebral cortex (Fig. 2.3.1.6)

The olfactory pathway is unique among the sensory modalities in having direct access to the cerebral cortex without passing through the thalamus. The primary olfactory receptor neurones of the olfactory mucosa send their axons (the fila olfactaria) through the cribriform plate of the ethmoid bone directly into the overlying olfactory bulb, where they contact the mitral cells in synaptic glomeruli. Axons of the mitral cells pass caudally in the olfactory tract to the anterior perforated substance. Here the olfactory tract splits into medial and lateral olfactory striae. All the mitral cell axons pass in the lateral stria. The medial stria contains axons mainly from the anterior olfactory nucleus, which are destined for the contralateral olfactory bulb by way of the anterior commissure. The lateral olfactory stria passes to the medial temporal lobe, where the axons terminate in the anterior margin of the entorhinal cortex, the pyriform cortex, and the corticomedial subdivision of the amygdala. All three termination zones interconnect. The olfactory entorhinal cortex and the olfactory amygdala both have connections with their non-olfactory partners, i.e. with the more posterior entorhinal subdivisions and the basolateral part of the amygdala respectively.

(g) The limbic cortex and the amygdala (Figs 2.3.1.6 and 2.3.1.8)

The latter stages of the sensory association pathways all converge on the entorhinal cortex, the perirhinal cortex, and the amygdala, which are themselves interconnected. These in turn project to the hippocampus. The perirhinal cortex lies lateral to the entorhinal

cortex in the banks of the rhinal fissure. It appears to receive afferents from the later stages of the sensory association pathways, notably from the temporal pole, and is extensively interconnected with the amygdala, the entorhinal cortex, and the hippocampus. The amygdala projects to the CA pyramidal cells, notably to CA1. The entorhinal input to the hippocampus arises from cell clusters in layer II and forms the perforant pathway, with axons terminating on the dendrites of granule cells of the dentate gyrus. Additional entorhinal fibres pass to the CA pyramidal cells. Axons of the dentate gyrus granule cells pass out into the molecular layer of the CA fields, notably CA3, where they synapse on the apical dendrites of pyramidal neurones. CA3 pyramidal cells project out via the fimbria, but also send collateral axons to synapse with the pyramidal cells of CA1 in particular. The activity of pyramidal neurones in the CA fields is regulated by inhibitory GABA-ergic interneurones in the molecular layer and by the basket cells, which are also GABA-ergic, sited immediately subjacent to the pyramidal layer. The latter inhibitory neurones have axons that branch around the pyramidal cell bodies, forming the baskets of terminal fibres from which they are named. CA1 sends some fibres into the fornix, but projects heavily to the subicular complex. The major output of the hippocampal formation comes from the subicular complex, and passes out into the fornix via the alveus and fimbria. The projection of fibres in the fornix to the hypothalamus, including the mamillary nuclei, is considered below. However, some hippocampal efferent fibres bypass these nuclei and enter the mamillothalamic tract without synapsing. Rather they pass directly to the anterior thalamic nuclei, where they terminate. From the anterior thalamic nuclei, of which there are several subdivisions, axons project to the cortex of the cingulate gyrus along the whole of its length, extending into the parahippocampal gyrus inferiorly. Fibres forming the cingulum bundle interconnect these medial cortical areas, running predominantly from anterior to posterior and ending in the parahippocampal gyrus, so that they reach the entorhinal cortex. This completes Papez's circuit and defines the structures of the limbic lobe—hippocampal formation, mamillary nuclei and anterior thalamus, cingulate cortex, parahippocampal gyrus and cingulate cortex. The term 'limbic system' is often extended to include structures, such as the amygdala, which have strong connections with these components.

(h) Association connections of the frontal lobe (Fig. 2.3.1.2)

The hierarchical sequence of connections in the sensory association pathways outlined above is reflected in a similar sequence passing from association areas of the prefrontal cortex back towards the primary motor cortex of the precentral gyrus. The two streams of connections are tightly linked together by long association pathways, with each tier of connections in the sensory association pathways interconnected with an area of the frontal lobe. The medial temporal areas, including the entorhinal cortex, the perirhinal cortex and the parahippocampal gyrus are closely interconnected with areas in the orbitofrontal cortex. The temporal pole is reciprocally connected with the frontal pole (Brodmann's area 10). The tiers of sensory association areas in the parietal, occipital, and temporal cortex are interconnected with the dorsolateral and ventrolateral prefrontal association cortex, occupying Brodmann's areas 9, 46, and 45. Broadly speaking, the association areas on the dorsal stream of connections for each modality interconnect with the dorsolateral prefrontal areas 9 and 46. In contrast, the ventral stream areas interconnect with the ventrolateral prefrontal cortex in areas 45 and 46. Like the sensory association areas, these regions seem to separate functionally into a dorsal hierarchy, dealing with internally generated actions, and a ventral hierarchy related to externally guided behaviours. Connections from the dorsolateral prefrontal cortex pass preferentially to the supplementary motor cortex, whereas the more ventrally placed prefrontal cortex feeds into the premotor cortex. Both the supplementary and the premotor cortex feed into the primary motor cortex of the precentral gyrus. The frontal eye-field (Brodmann's area 8), which has major ipsilateral association connections with the visual areas in the occipital lobe, is strategically placed between the prefrontal and premotor areas. It also receives ipsilateral association connections from the dorsolateral and ventrolateral prefrontal association areas.

(i) Speech areas of the cerebral cortex

Because speech and language are not present in the subhuman primate species commonly used for neuroanatomical investigation of cortical connections, little is known about the connections of these in the human. The posterior speech area (Wernicke's area) occupies a large extent of the posterior temporal and inferior parietal cortex at the posterior limits of the superior temporal and lateral sulci, and is often taken as extending anteriorly along the superior temporal gyrus. It includes the angular and supramarginal gyri (Fig. 2.3.1.9). This would represent a region where the association pathways of all three modalities—somatic sensation, vision, and hearing—lie close together. It would include many auditory association areas along the superior temporal gyrus. Within the frontal lobe, Broca's area occupies the pars triangularis between the anterior and ascending rami of the lateral sulcus. It is closely adjacent to the face area in the lateral premotor cortex. Similarly, the medial speech area lies on the medial surface immediately in front of the face representation in the supplementary motor cortex. It is, perhaps, not unreasonable to suppose that these two areas function to some extent as the premotor and supplementary motor speech areas.

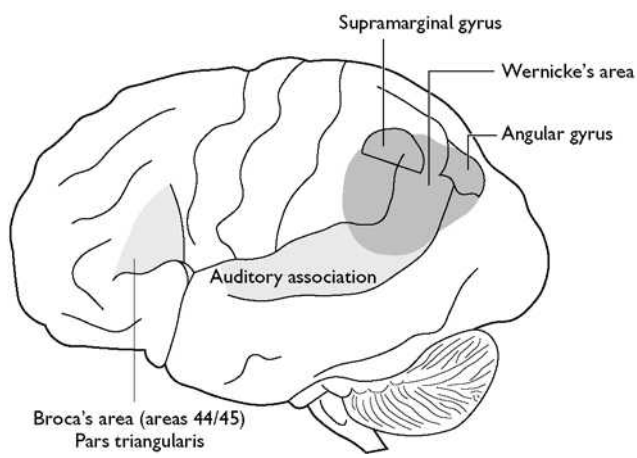


Fig. 2.3.1.9 Auditory association cortex and Wernicke's sensory speech area.

Subcortical efferent pathways of the cerebral cortex

The corticostriate pathway and the basal ganglia (Fig. 2.3.1.10)

The anatomical definition of the basal ganglia comprises those deep grey matter nuclei that develop from the telencephalic (cerebral) vesicle. Strictly, this would include the components of the striatum, the globus pallidus, the amygdala, and the claustrum. Because these latter two are better considered with the cortex and limbic system respectively, they are usually excluded from the term basal ganglia. Similarly, because of their close anatomical and functional relationship with the striatum and globus pallidus, the substantia nigra and subthalamic nucleus are usually included with the basal ganglia. The caudate and putamen are developmentally, anatomically, pharmacologically, and functionally a single structure, secondarily subdivided by the development of the internal capsule. The two fuse below the inferior margin of the anterior limb of the internal capsule to form the nucleus accumbens. This fusion comes to the surface of the hemisphere at the anterior perforated substance, an area sometimes called the olfactory tubercle. Together, the caudate, putamen, nucleus accumbens, and olfactory tubercle make up the striatum. The globus pallidus consists of two parts, an external segment (GPe) and an internal segment (GPi), separated by a thin white-matter lamina. Both segments extend ventrally in the region of the substantia innominata, below the anterior commissure, to form the ventral pallidum. The pars reticulata of the substantia nigra (SNpr) is developmentally, anatomically, pharmacologically and functionally a part of the internal segment of the globus pallidus, which has been separated off by the development of the fibres of the internal capsule passing into the crus cerebri of the midbrain. Therefore, in this account the basal ganglia and related nuclei will comprise the striatum, GPe, GPi-SNpr, and the subthalamic nucleus (see Fig. 2.3.1.10).

The entire cerebral cortex projects to the striatum. The putamen receives predominantly from the sensorimotor cortex around the central sulcus. The caudate receives the input from most of the parietal, occipital, temporal, and frontal lobes. The limbic areas, including the entorhinal and perirhinal cortex, the hippocampus, and the amygdala, project to the ventral striatum (nucleus accumbens and olfactory tubercle) and the adjacent ventral portion of the head of the caudate. The corticostriate pathway arises from pyramidal cells predominantly in layer V of the cortex and is excitatory, using glutamate as its neurotransmitter. The termination of the pathway is broadly topographically organized, but with considerable interdigitation or overlap of the projections from different cortical areas. Individual cortical areas project to a longitudinal strip of striatum orientated anteroposteriorly. The strips receiving from the frontal lobe extend more rostrally, whereas those receiving from the parietal, occipital, and temporal lobes extend more caudally. Anteriorly, zones receiving projections from interconnected areas in the frontal lobe interdigitate or overlap; centrally, zones receiving from frontal areas overlap with zones connected to areas within the parietal, temporal, or occipital cortex with which the frontal areas are connected by ipsilateral association fibres. More posteriorly, interconnected sensory association areas

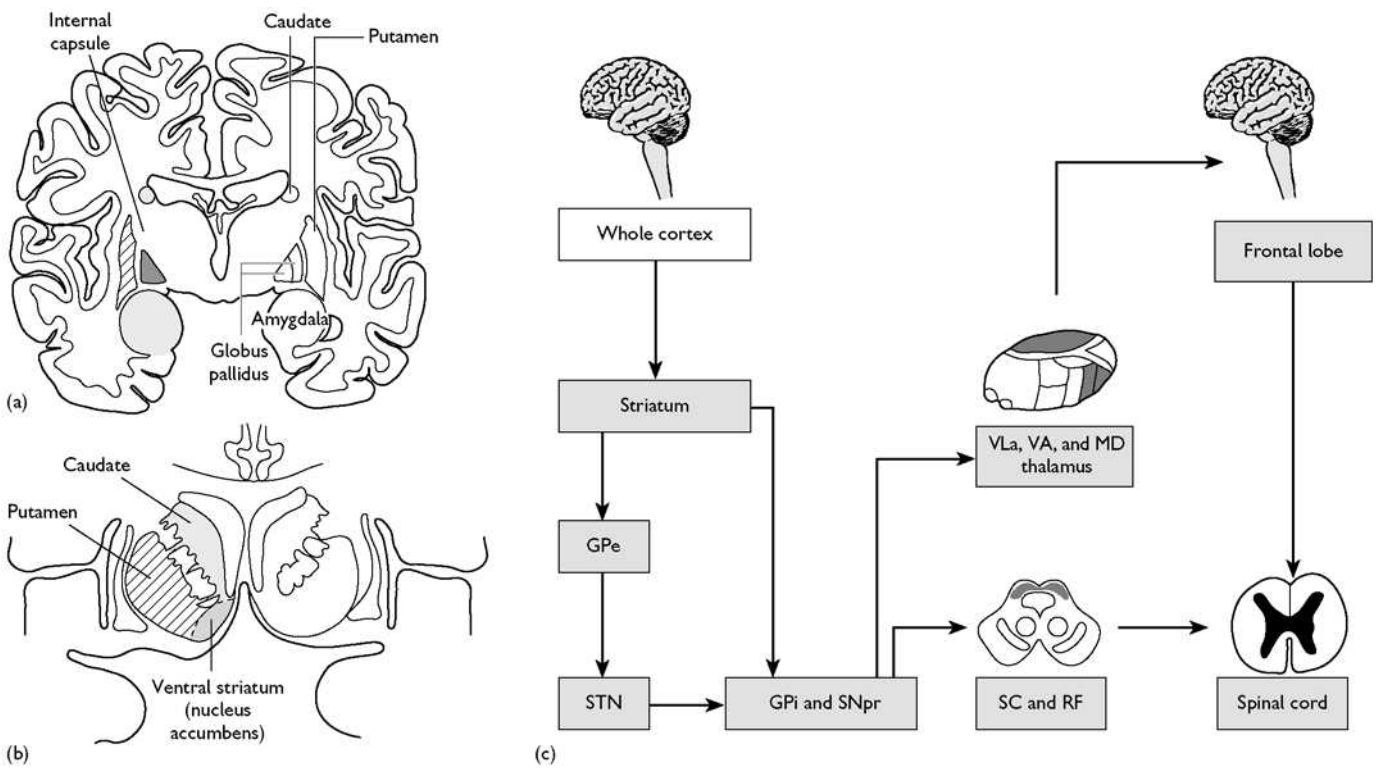


Fig. 2.3.1.10 The corticostriate pathway and the basal ganglia: (a) component nuclei of the basal ganglia; (b) striatum; (c) major pathways through the basal ganglia. GPe, external segment of the globus pallidus; STN, subthalamic nucleus; GPi, internal segment of the globus pallidus; SNpr, pars reticula of the substantia nigra.

project to overlapping zones in the striatum. The striatum itself is compartmentalized, on the basis of acetylcholinesterase (AChE) histochemistry, into AChE-poor patches (striosomes) and AChE-rich matrix. The cortex projects into both compartments, but with a slightly different laminar origin. Pyramidal neurones in the deeper part of lamina V project to the patches, whereas more superficial lamina V cells and some lamina III cells project to the matrix. It is possible that apparently overlapping corticostriate projections from different cortical areas are segregated into the different compartments. The projection from a single cortical area may switch between compartments along the anteroposterior length of its zone of termination. The intralaminar nuclei of the thalamus sends excitatory (glutamatergic) axons to the entire striatum. The nuclei project to those parts of the caudate or putamen which receive from the cortical areas and to which the particular intralaminar nucleus projects. Many, if not all, interconnected areas in the frontal, parietal, occipital, and temporal lobes also receive a shared projection from a single intralaminar nucleus. Thus, there is a tightly organized but complex topographical relationship between the connections of functionally related cortical areas with each other, with the striatum, and with the intralaminar nuclei of the thalamus. A closely similar relationship is seen in the cortical connections of the claustrum and the basal forebrain, and possibly even in the cortical projection to the pontine nuclei. Another major projection to the striatum is dopaminergic and comes from the pars compacta of the substantia nigra and adjacent nuclei. There is some topography in this projection, with the lateral and central parts of pars compacta of the substantia nigra projecting to the caudate and putamen. The medial part and the adjacent nuclei, such as the ventral tegmental area, project to the ventral striatum. The effect of dopamine on striatal neurones appears to be different in the two compartments. Additional striatal afferents come from the brainstem raphe nuclei (serotonin) and the amygdala (to the ventral striatum and the head of the caudate).

The output of the striatum passes to all parts of the globus pallidus (the ventral pallidum, GPe, and GPi-SNpr). These fibres are inhibitory, using GABA as their neurotransmitter. Here, the pathway through the basal ganglia separates into a direct and an indirect route, ultimately passing to the thalamus. Fibres from the striosomes (patches) of the striatum are rich in substance P (as well as being GABA-ergic) and project to GPi. Axons from GPi go directly to the anterior part of the ventral lateral nucleus and the adjacent ventral anterior nucleus of the thalamus, forming the direct route. Neurones in the striatal matrix compartment contain enkephalin as a cotransmitter with GABA and project to GPe. Axons from GPe go to the subthalamic nucleus, which in turn projects to GPi (the indirect pathway). The ventral pallidum has equivalent pathways, but the final destination of the pallidal output is the mediodorsal nucleus of the thalamus. All efferents from the globus pallidus are GABA-ergic and inhibitory. The neurones of the subthalamic nucleus are glutamatergic and excitatory. Activation of the striatal output through the direct pathway leads to a reduced tonic inhibition of the thalamus. In contrast, activation via the indirect pathway leads to increased activation of GPi, and hence increased inhibition of the thalamus. The balance between these opposite effects is crucial in the normal functioning of the basal ganglia. Disruption of this balance is used to explain much of the pathophysiology of the extrapyramidal disorders. The ventral anterior and rostral ventral lateral nuclei of the thalamus project mainly to the premotor and supplementary motor areas of the frontal lobe. The ventral pallidal pathway through the mediodorsal thalamus feeds onto the prefrontal association areas. The major pathways through the basal ganglia are summarized in Fig. 2.3.1.10(c)).

The corticopontine pathway and the cerebellum

There is a major projection from the cortex to the pontine nuclei. The extent to which individual cortical areas contribute to this pathway varies. The greatest projection comes from the regions around the central sulcus, which also contribute to the pyramidal tract (see below). However, there is a substantial projection from the prefrontal cortex and a significant number of fibres from the occipital lobe. Many fewer corticopontine axons arise from temporal neocortex, although some areas send some and it is possible that most areas send at least a few.

The pontine nuclei send their axons to the cerebellar cortex of the lateral parts of the posterior lobe. They terminate as mossy fibres, contacting granule cells. Axon collaterals pass, in addition, to the lateral dentate deep cerebellar nucleus. The anterior lobe and the midline and paramedian region of the posterior lobe are related to the spinocerebellar inputs. The flocculonodular lobe is connected to the vestibular pathway. The Purkinje cells of the cerebellar cortex send inhibitory fibres to the deep cerebellar nuclei, and it is from these that the output of the cerebellum arises. The intermediate (globose and emboliform) and medial (fastigial) deep nuclei project to the red nucleus and the vestibular nuclei and reticular formation. The dentate nucleus projects to the posterior part of the ventral lateral nucleus of the thalamus. This in turn provides the major thalamic input to the primary motor cortex of the precentral gyrus. In this way, the neocortical input to the cerebellum is transmitted via the pontine nuclei,and to the motor cortex via the thalamus.

The fornix and the cortical projection to the hypothalamus

Fibres leaving and entering the hippocampus form a thin white-matter covering on the ventricular surface, deep to the ependyma, called the alveus. These fibres pass into the fornix via the fimbria. The fornix passes initially posteriorly and superiorly, then anteriorly, curving around the outer curve of the lateral ventricle and angling towards the midline in its course. The fornices of the two sides come together at about the junction of the posterior third and anterior two-thirds of the corpus callosum. Many fibres pass across the midline, the commissure of the fornix, and turn caudally to enter the contralateral hippocampus. The two fornices, united in the midline, are suspended from the corpus callosum by the septum pellucidum as they arch over the roof of the third ventricle and the choroid fissure of the body of the lateral ventricle. They turn ventrally, immediately in front of the interventricular foramen of Monroe. The two fornices separate, and each divides into an anterior and a posterior column, passing in front of and behind the anterior commissure. The anterior column carries axons to and from the septal nuclei CA3, CA1, and the subiculum project to the lateral septal nucleus. This has diverse efferent projections to the hypothalamus, the epithalamus, and the midline thalamus, but also projects to the adjacent medial septal nucleus. The medial septal nucleus is the major source of cholinergic fibres to the hippocampus via the fornix. The posterior column of the fornix curves posteriorly through the hypothalamus, giving off many fibres to

medial and lateral hypothalamic nuclei. It ends in the mamillary nuclei, which in turn project via the mamillothalamic tract to the anterior thalamus. This projection is partially bilateral.

There is a major input to the hypothalamus from the amygdala, via the stria terminalis, a white-matter tract that follows the curve of the caudate nucleus around the lateral ventricle, lying between the caudate and the thalamus. The connections between the hypothalamus and amygdala are reciprocal.

Direct projections to the cortex from the hypothalamus have been discussed earlier. Direct projections from neocortex to the hypothalamus have been described, but their extent and distribution are disputed. If they exist in the human, they probably arise from the prefrontal/orbitofrontal and insular cortex.

The corticobulbar and corticospinal pathways

The direct projection of the cortex to the brainstem and spinal cord, the pyramidal tract, arises from the cortex in front of and behind the central sulcus. About 40 per cent of fibres arise from the primary somatic sensory cortex and the adjacent superior parietal lobe. In the frontal lobe, axons arise from the primary motor, premotor, and supplementary motor areas. Apart from the direct innervation of the spinal cord and motor nuclei of the cranial nerves, direct cortical fibres innervate the red nucleus, the vestibular nuclei, the reticular formation, and the superior colliculus. In the case of the first two of these, the origin of the fibres is probably very similar to the areas of origin of the pyramidal tract. In contrast, the cortical projections to the reticular formation and superior colliculus have a much wider origin, and may include most neocortical areas to a greater or lesser degree.

The contribution of neuroanatomy to psychiatry

The above is a necessarily abbreviated account of the anatomy of the central nervous system, centring on the organization and connections of the cortex including the limbic lobe. Topographical neuroanatomy has been ignored, despite its importance in the reading of modern images of the brain in living patients. For psychiatry, the importance of the connectionist view of the brain lies in the contribution it can make to the understanding of the normal and pathological functioning of the central nervous system. Present-day neuropsychology and cognitive neuroscience are aimed towards understanding the highest levels of central nervous system processing and function. It is these areas or systems that are commonly involved in the signs and symptoms of major psychiatric disease. It is to the pathways underlying cognition, perception, memory, mood, and attention that the psychiatrist interested in the pathophysiology of mental illness must turn his or her attention, and it is this area that the above account has attempted to review.

Further information

Gloor, P. (1977). *The temporal lobe and limbic system*. Oxford University Press.

Pearson, R.C.A. (1995). Dorsal thalamus. In *Gray's anatomy* (38th edn) (ed. P.L. Williams, L.H. Bannister, M.M. Berry, *et al.*), pp. 1080–91. Churchill Livingstone, Edinburgh.

Pearson, R.C.A. (1995). Cerebral cortex. In *Gray's anatomy* (38th edn) (ed. P.L. Williams, L.H. Bannister, M.M. Berry, *et al.*), pp. 1141–71. Churchill Livingstone, Edinburgh.

Passingham, R. (1993). *The frontal lobes and voluntary action*. Oxford University Press.

Paxinos, G. (ed.) (1990). *The human nervous system*. Academic Press, San Diego, CA.

Tovée, M.J. (1996). *An introduction to the visual system*. Cambridge University Press.

Zeki, S. (1993). *A vision of the brain*. Blackwell Science, Oxford.

2.3.2 Neurodevelopment

Karl Zilles

Neural induction

The central nervous system originates from the midline region of the embryo as a specialized area of ectoderm, the neuroectoderm, or neural plate. FGF (fibroblast growth factor) signalling and BMP (bone morphogenetic protein) as well as Wnt (wingless gene) inhibition are required as steps for neural induction.[1] However, the complete set of factors and their respective interactions are presently not sufficiently understood in the mammalian embryo. As the neuroectodermal cells proliferate the neural plate is transformed into an indentation, the neural groove. The lateral parts of this groove approach each other and join in the midline, forming the neural tube. The folds begin to fuse in the central part of the groove but the most rostral and caudal parts close only later, leaving initially rostral and caudal neuropores. A small transitional zone between the neural plate and the surrounding ectoderm provides the cells of the neural crest, which develop into the postganglionic cells of the sympathetic and parasympathetic nervous system, the sensory neurons of the spinal ganglia and ganglia of cranial nerves, Schwann cells, and chromaffine cells of the suprarenal glands.

Neural tube formation requires a controlled expression of cell adhesion molecules in the lateral folds of the neural groove. If the rostral neuropore fails to close, the development of the forebrain is impaired leading to anencephaly. If the caudal neuropore fails to close, the most severe result is rachischisis, a malformation with a dorsally exposed neural groove. The mildest result is spina bifida occulta which is a cleft of a vertebral arch covered by epidermis.

As development continues, the neural tube and crest move to a position between the ectoderm and the notochord. The rostral part of the neural tube differentiates into the brain; the caudal part (behind the fifth somite) differentiates into the spinal cord.

Organogenesis of the central nervous system

The embryonic brain has three vesicular enlargements: the forebrain, telencephalon and diencephalon, the midbrain, mesencephalon, and the hindbrain, rhombencephalon. Because the brain grows much faster than the rest of the embryo, it becomes deflected ventrally. A dorsally convex cephalic flexure marks the border between hindbrain and midbrain, and a cervical flexure marks the border between spinal cord and hindbrain (Fig. 2.3.2.1). The ventrally convex pontine flexure forms the hindbrain. During week 5, the forebrain differentiates further into the rostral telencephalon and the more caudal diencephalon. The telencephalon consists of two

hemispheric vesicles connected by a thin lamina terminalis, the most dorsal part of which develops into the commissural plate, the *Anlage* of the corpus callosum. The ventral part of the lamina terminalis differentiates into the anterior commissure. The central cavity of the diencephalon (the third ventricle) is connected with the cavities of the hemispheric vesicles (the lateral ventricles) by the interventricular foramen. The diencephalon develops bilateral evaginations, the eye vesicles, which differentiate into the retina and the optic nerve. Meanwhile the hindbrain becomes subdivided into a rostral metencephalon and a caudal myelencephalon.

The cerebellum starts to develop from the metencephalon during week 6 (Fig. 2.3.2.1). At first, the enlarged central cavity of the neural tube (the future fourth ventricle) has a thin roof plate bordered by two thickenings of the neural tube, the rhomboid lips, which merge in the midline. These thickenings develop into the cerebellar hemispheres, while the midline part develops into the vermis of the cerebellum. Fissures appear in the cerebellar hemispheres, forming the anterior and posterior lobes and the uvula.

The rhombencephalon is temporarily divided into eight rhombomeres,[2] whose borders are specified by specific combinations of transcription factors (e.g. Hox, Krox, Wnt genes) disappear during further development. Local expression of homeobox genes leads to the formation of the pallium and the ganglionic hill in the forebrain.

The hindbrain develops in close association with the visceral archs, which appear during week 4. It innervates these archs and the organs derived from them by a group of branchial nerves, which later become the trigeminal (V), facial (VII), glossopharyngeal (IX), vagal (X), and accessory (XI) cranial nerves. Other cranial nerves develop connections between the hindbrain and peripheral organs not derived from the visceral archs. They are the oculomotor (III), trochlear (IV), abducens (VI), vestibulocochlear (VIII), and hypoglossal (XII) nerves. The olfactory (I) and optic (II) nerves arise separately as evaginations of the forebrain.

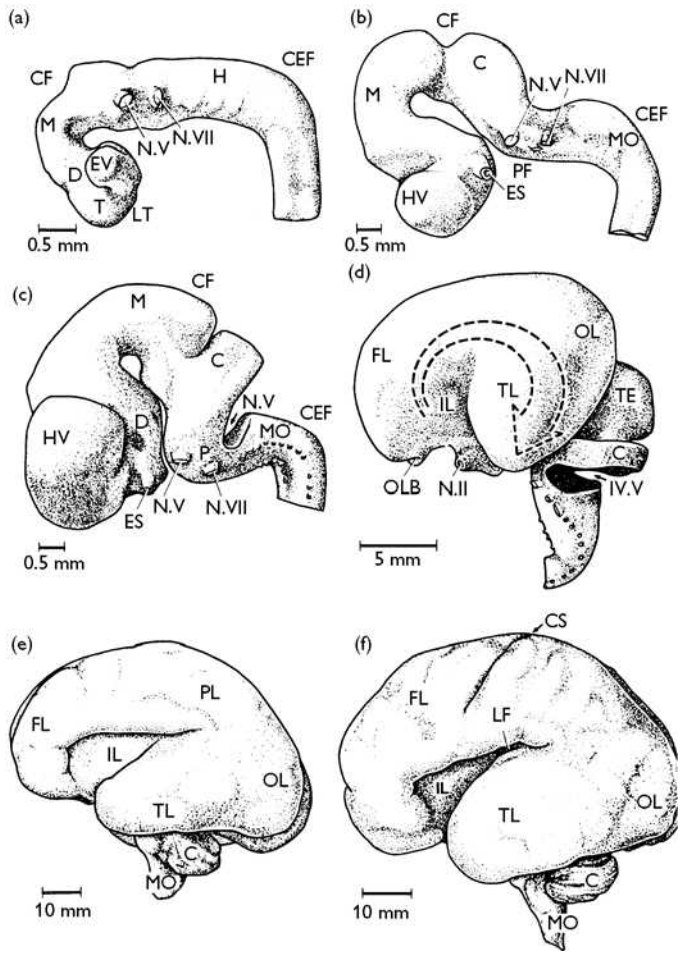


Fig. 2.3.2.1 Development of the human brain. Brains of (a) 4 mm, (b) 10.4 mm (c), 13.8 mm, and (d) 53 mm human embryos, and (e) 21 week and (f) 24 week fetuses. C, cerebellum or cerebellar *Anlage*; CEF, cervical flexure; CF, cephalic flexure; CS, central sulcus; D, diencephalon; ES, eye stalk; EV, eye vesicle; FL, frontal lobe; H, hindbrain; HV, hemispheric vesicle; IL, insular lobe; LF, lateral fissure; LT, lamina terminalis; M, mesencephalon; MO, medulla oblongata; N.II, optic nerve; N.V, trigeminal nerve; N. VII, facial nerve; OL, occipital lobe; OLB, olfactory bulb; P, pons; PF, pontine flexure; PL, parietal lobe; T, telencephalon; TE, tectum; TL, temporal lobe; IV.V, fourth ventricle.

Histogenesis of the spinal cord

The neural tube initially consists of a single layer of neuroepithelial cells surrounding a central canal filled with the cerebrospinal fluid. The outer surface of the future spinal cord has an external limiting membrane, and the inner surface bordering the central canal has an inner limiting membrane. The entire wall of the early neural tube is called the ventricular zone.[3]

The cells of the ventricular zone proliferate, and the surface of the spinal cord enlarges. The cord then thickens as cells divide further to produce a multilayered epithelium. The daughter cells have different potentialities: one type of cell (the neuroblast) retains the capability for mitosis, whereas another type (the proneurone) is postmitotic and represents an immature neurone. The proliferation of neurones is almost complete around birth.

Some neuroepithelial cells develop into precursors of glial cells, glioblasts, which differentiate into astroglial, oligodendroglial, and microglial cells. The first glioblasts differentiate into radially extended cells spanning the entire width of the wall of the spinal cord (the same occurs in the cerebral hemispheres and the cerebellar cortex as described below). During later development these cells are transformed into ependymal cells and astroglia.

The histogenesis of the spinal cord starts at the cervical level and progresses in a caudal direction. After week 3, a longitudinal sulcus limitans is recognizable on the inner surface of the neural tube, dividing the wall into dorsal (alar) and ventral (basal) plates. The dorsal plates of both sides are connected by a thin roof plate, and the ventral plates by a thin floor plate. The dorsal plate differentiates into a sensory zone, the dorsal horn of the adult spinal cord, and the ventral plate differentiates into a motor zone, the ventral horn. The sympathetic preganglionic neurones form the lateral horn, which is present only at thoracic levels. The subdivision into dorsal and ventral plates is functionally important not only in the spinal cord, but also in the brainstem (see below).

During week 5, three concentrically organized zones—a cell-dense ventricular zone, a less-dense intermediate or mantle zone, and a superficially located marginal zone free of neuronal cells—develop in the wall of the spinal cord. Proneurones leave the ventricular zone and migrate along radial glial cells into the intermediate zone where they become organized into cell groups, nuclei. The motor neurones develop the axons of the ventral root, and the processes from spinal ganglionic cells grow into the spinal cord to

form the dorsal roots. Synapses develop first in the motor zone, and later in the sensory zone, during the weeks 10 to 13.

During the third month, the ventricular zone is reduced to a small rim surrounding the central canal and is finally transformed into the ependymal cell layer. The intermediate zone becomes organized into dorsal, ventral, and (at thoracic levels) lateral horns. The ascending and descending fibre tracts of the spinal cord are increased in size in the marginal zone. During weeks 14 and 15, oligodendrocytes begin to myelinate these fibre tracts. The corticospinal, or pyramidal, tract is visible for the first time during week 14 and reaches its target neurones, mainly motor neurones of the ventral horn, between weeks 17 and 29. The myelination of the pyramidal tract is completed between the first and second postnatal years. This late myelination explains the presence of the Babinski reflex in newborns and its disappearance during the first 2 years.

Histogenesis of the brainstem and cerebellum

At the level of the fourth ventricle the various zones of the hindbrain are arranged in a lateral-to-medial sequence (somatosensory-viscerosensory-visceromotor-somatomotor). In the hindbrain, proneurones not only migrate radially, as in the spinal cord, but also tangentially and longitudinally. This complex migration and the growth of fibre tracts lead to changes of the lateral-to-medial sequence of cranial nerve nuclei (e.g. the facial nucleus) in the adult.

In the cerebellum, between weeks 10 and 11, neuroblasts migrate from the ventricular zone through the intermediate zone into an area (the external granular layer) at the surface of the marginal zone. During weeks 12 and 13, proneurones from the ventricular zone begin to migrate along radially extended glial cells, Bergmann glia, into a region below this external granular layer, where they form the Purkinje cells of the ganglionic layer of the cerebellar cortex. Other proneurones from the ventricular zone develop into the cerebellar nuclei. Proneurones from the external granular layer then migrate inwards to form the internal granular layer and the basket and star cells of the molecular layer. The migration of cerebellar proneurones is not completed until the first postnatal year. During weeks 16 and 26, synapses develop and afferent fibre systems begin to form. The external granular layer finally disappears during the first 2 years of life, leaving the three-layered organization (molecular, ganglionic, and internal granular layers) of the adult cerebellar cortex.

Histogenesis of the cerebral cortex

Initially, the entire wall of the hemispheric vesicle consists of very densely packed mitotic cells. These cells undergo more than 28 mitotic rounds in the human brain.[4] In week 5, this develops into an inner cell-dense periventricular zone and an outer cell-poor marginal zone. In week 6, postmitotic proneurones leave the inner periventricular zone and form an intermediate (mantle) zone between the marginal and periventricular zones. By the end of week 6, the periventricular zone is further subdivided into a cell-dense ventricular zone and a less cell-dense subventricular zone.

During week 8, the cortical plate between the marginal and intermediate zones is formed by proneurones which have migrated along radial glial cells from the ventricular and subventricular zones through the intermediate zone.[3–7] A single radial glial cell can span the entire distance between the ventricular and pial surfaces. As the proneurones 'climb' to the cortical plate along the processes of the radial glial cell, they produce a vertically oriented cortical cell column. This radially guided migration is responsible for the architectonic organization of cortical layers II–VI. It is a prototype of the cortical map of the adult brain.[8] A further feature of cortical migration is the inside-to-outside layering, with the earliest proneurones being found in the deepest layers of the cortical plate and the latest in the most superficial layers. Thus layers V and VI of the adult cortex are generated before layer IV, and layer IV is generated before layers III and II. When cortical proneurones are migrating radially, their 'stop signal' is Reelin, produced by specialized cells (Cajal–Retzius cells) in the marginal layer. In this way, Reelin organizes the inside-to-outside layering of the cortex.[4] In addition to Reelin, the microtubule-associated protein (LIS1), Doublecortin (DCX), and the tumour-suppressor p73 are crucial for the normal migration of the proneurones. These radially migrating cells develop into the glutamatergic projection neurones (pyramidal neurones) of the cortex. At the same time there is much tangential migration of cells born in the ganglionic eminence, migrating into the cortical *Anlage*, and developing into γ-aminobutyric acid producing (GABAergic) cortical interneurones.[4,9]

Regional differences in the development of the cortical plate subdivide the hemisphere into segments. The lateral segment, with a well-developed cortical plate and presubplate, develops into the neocortex. The mediodorsal segment, with a wide marginal zone and a thin-folded cortical plate, develops into the archicortex, including the hippocampus. The mediobasal segment, with its inconspiciously developed cortical plate, is the precursor of the palaeocortex. The basolateral segment, ganglionic eminence, generates cortical interneurones (see above) and develops into the corpus striatum, the amygdala, and the septum.

During weeks 10 and 12, the axons of the serotoninergic and noradrenergic neurones contribute to the first synapses in the marginal and presubplate zones, where neurotrophin receptors (see below) are expressed. During the following 3 weeks, the subplate zone develops as axons grow in from the basal forebrain and thalamus, dendrites enlarge, and synapses form. From weeks 16 to 24, the cortical *Anlage* has a small marginal zone, a wide cell-dense cortical plate, and a very wide and less cell-dense subplate.

The transformation into the adult neocortical pattern starts between weeks 25 and 34 as the migration and proliferation of proneurones diminishes. Dendrites begin to differentiate and synapses begin to develop in the deepest cortical layers, progressing to the most superficial layer. Before birth, six cortical layers can be recognized in all regions of the neocortex. In the postnatal period, layer IV (inner granular layer) disappears as part of the differentiation of the motor cortex, leaving the five-layered agranular neocortex of the motor region.[10] Shortly before birth, the subplate, the subventricular zone, and most of the ventricular zone disappear, neuronal proliferation ceases, and the intermediate zone is transformed into the white matter of the pallium. The remaining ventricular zone contributes to the ependymal layer of the ventricular surface.

Dendritic and axonal differentiation continues after birth and into adult life. Synaptogenesis reaches a maximum during the first

postnatal year, but continues at a lower rate during childhood. The myelination of the vestibular system is finished shortly before birth, that of the somatosensory, visual, auditory, pyramidal, and extrapyramidal fibre tracts is nearly complete by the end of the third postnatal year, and that of the associative fibre tracts in the cerebral hemispheres is continued until the second decade.[11] The key change in synapses after birth is pruning; the density of synapses in the adult brain is half that in neonates.

The development of the neocortex is summarized in Fig. 2.3.2.2.

Hemispheric shape and the formation of gyri

The spherical shape of the early foetal hemisphere is transformed into the adult shape by differing rates of growth in the various regions of the telencephalon (Fig. 2.3.2.1). The future insular lobe grows less than other telencephalic regions, so that by the eighth month it is covered by the frontal, parietal, and temporal lobes. In the adult brain the insula is completely buried in the depth of the lateral fissure.

The extensive growth of the parieto-occipito-temporal association cortex leads to a bend in the temporal lobe around the lateral fissure. At the same time the temporal pole is pushed rostrally. This direction of growth (Fig. 2.3.2.1(d)) also affects the structures situated dorsomedially, i.e. the archicortex with the hippocampus, the corpus striatum, and the lateral ventricles. The corpus striatum is split by the ingrowing fibres of the internal capsule into the caudate nucleus and the putamen. The head of the caudate is situated ventrolaterally to the corpus callosum in the frontal lobe, and the tail of the caudate is located in the temporal lobe dorsal to the inferior horn of the lateral ventricle. The hippocampus forms its largest extension (the retrocommisssural part) in the temporal lobe, bends around the posterior end (splenium) of the corpus callosum, and reaches a position on top of the corpus callosum (the supracommissural part). The precommissural part of the hippocampus ends in front of the genu of the corpus callosum.

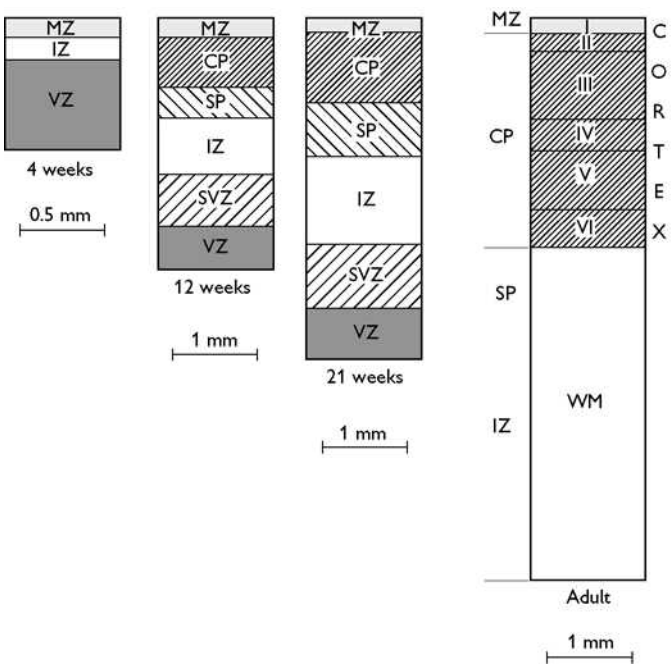


Fig. 2.3.2.2 Development of the neocortex between 4 weeks of gestation and adulthood. Roman numerals indicate the six cytoarchitectonically defined layers of the neocortex, which originate from the marginal zone (layer I) and the cortical plate. CP, cortical plate; IZ, intermediate zone; MZ, marginal zone; SP, subplate; SVZ, subventricular zone; VZ, ventricular zone; WM, white matter.

After the appearance of the lateral fissure, the neocortical surface develops many sulci and gyri. The central, collateral, cingulate, parieto-occipital, superior temporal, and calcarine sulci appear between weeks 16 and 21, followed by the pre- and postcentral, frontal, temporal, and intraparietal sulci. Highly variable secondary and tertiary sulci develop between week 29 and birth, when all sulci have been formed.[12,13]

The reasons for the formation of gyri in many mammalian brains, including the human brain, are not completely understood. Since the basic organization of the cerebral cortex is vertically oriented, with cell columns positioned side by side, growth of the cortex inevitably leads to a considerable enlargement of the cortical surface. A large unfolded cortical surface would have two disadvantages: the volume of the skull would increase to such a degree during foetal development that a normal delivery would be impossible; the distance between cortical regions interconnected by intrahemispheric projection fibres would increase and with it the information transmission time. Gyri allow the maximal cortical surface in the minimal volume, and they increase the speed of neural transmission between neighbouring cortical areas. Recent measurements show that gyrification is greatest in the association cortices. Although all gyri and sulci are present at birth, the depth of the sulci increases until two-thirds of the cortical surface is hidden in them.[12]

Genetic factors during development

The co-ordinated expression, in space and time, of many genes underlies neurodevelopment. Mutations in these 'neurodevelopmental genes' are increasingly being recognized as causes of developmental neurological disorders such as cortical dysplasia and epilepsy; they may also be relevant to learning disability and schizophrenia. Different gene families are involved in the major component processes of neurodevelopment, such as organogenesis, neurogenesis, neuronal migration, synaptogenesis, and programmed cell death (apoptosis).[14,15] The details are beyond the scope of this book, but a few examples are given here.

Neurotrophins (growth factors) are genes which, as their name suggests, are critical for neuronal growth and survival, especially via their influence on apoptosis which is promoted by insuffiency of neurotrophins such as nerve growth factor and inhibited by enhanced nerve growth factor functioning. The effects of neurotrophins are mediated by specific tyrosine kinase (Trk) receptors (Table 2.3.2.1). Classical neurotransmitters and their receptors are also involved in neurodevelopment, both directly in the formation of synaptic connections and indirectly through regulation of neurotrophins and Trk receptors; particular roles have been shown for glutamate, acting via *N*-methyl-D-aspartate receptors, as well as for γ-aminobutyric acid (GABA) and acetylcholine.

Table 2.3.2.1 Important neurotrophins, their sites of synthesis in the central nervous system, receptors, and target structures

Neurotrophin	Site of synthesis	Receptors	Target structures
NGF	Hippocampus Neocortex	TrkA, $p76^{NTR}$	Cholinergic neurones in the basal forebrain
BDNF	Hippocampus Neocortex	TrkB. $p75^{NTR}$	Dopaminergic neurones in the midbrain; retinal ganglionic cells; cholinergic neurones in the basal forebrain
NT-3	Hippocampus Cerebellum	TrkC, $p75^{NTR}$	Dopaminergic neurones in the midbrain; neurones of the nucleus mesencephalicus nervi trigemini; neurones of origin of the pyramidal tract

NGF, nerve growth factor; BDNF, brain-derived neurotrophic factor; NT-3, neurotrophin 3.

Further information

Rao, M.S. and Jacobson, M. (eds.) (2005). *Developmental neurobiology*. Kluwer Academic/Plenum Publishers, New York.

Sanes, D.H., Reh, T.H. and Harris, W.A. (2006). *Development of the nervous system* (2nd edn). Academic Press, Amsterdam.

Squire, L.R., Bloom, F.E., McConnell, S.K., *et al.* (eds.) (2005). *Fundamental neuroscience* (2nd edn), chapters 14–21. Academic Press, Amsterdam.

References

1. Ladher, R. and Schoenwolf, G.C. (2005). Making a neural tube: neural induction and neurulation. In *Developmental neurobiology* (eds. M.S. Rao and M. Jacobson), pp. 1–20. Kluwer Academic/Plenum Publishers, New York.
2. Lumsden, A. and Krumlauf, R. (1996). Patterning the vertebrate neuraxis. *Science*, **274**, 1109–15.
3. Kostovic, I. (1990). Zentralnervensystem. In *Humanembryologie* (ed. K.V. Hinrichsen), pp. 381–448. Springer-Verlag, Berlin.
4. Meyer, G. (2007). Genetic control of neuronal migrations in human cortical development. *Advances in Anatomy Embryology and Cell Biology*, **189**, 1–114.
5. Levitt, P. and Rakic, P. (1980). Immunoperoxidase localization of glial fibrillary acid protein in radial glial cells and astrocytes of the developing rhesus monkey brain. *Journal of Comparative Neurology*, **193**, 815–40.
6. Rakic, P. (1985). Limits of neurogenesis in primates. *Science*, **227**, 154–6.
7. Supèr, H., Soriano, E., and Uylings, H.B.M. (1998). The functions of the preplate in development and evolution of the neocortex and hippocampus. *Brain Research Reviews*, **26**, 40–64.
8. Rakic, P. (1988). Specification of cerebral cortical areas. *Science*, **241**, 170–6.
9. Rakic, P. (1995). Radial versus tangential migration of neuronal clones in the developing cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 11323–7.
10. Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth, Leipzig.
11. Yakovlev, P.I. and Lecours, A.-R. (1967). The myelogenetic cycles of regional maturation of the brain. In *Regional development of the brain in early life* (ed. A. Minkowski), pp. 3–70. Blackwell Science, Oxford.
12. Zilles, K., Armstrong, E., Schleicher, A., *et al.* (1988). The human pattern of gyrification in the cerebral cortex. *Anatomy and Embryology*, **179**, 173–9.
13. Armstrong, E., Zilles, K., Omran, H., *et al.* (1995). The ontogeny of human gyrification. *Cerebral Cortex*, **5**, 56–63.
14. Levitt, P., Barbe, M.F., and Eagleson, K.L. (1997). Patterning and specification of the cerebral cortex. *Annual Review of Neuroscience*, **20**, 1–24.
15. Lambert de Rouvroit, C. and Goffinet, A.M. (1998). A new view of early cortical development. *Biochemical Pharmacology*, **56**, 1403–9.

2.3.3 Neuroendocrinology

Charles B. Nemeroff and Gretchen N. Neigh

Definitions and principles

In the late 1950s, a controversy arose in endocrinology and neuroscience concerning whether neurones are capable of manufacturing and secreting hormones; that is to say, is it possible that certain neurones subserve endocrine functions? Much of the ensuing debate, which persisted for approximately two decades, was centred on two major findings. First, led by a husband and wife team, the Scharrers, a number of neurohistologists working with mammalian as well as with lower vertebrate and invertebrate species documented the presence—by both light and electron microscopy—of neurones that had all the characteristics of previously studied endocrine cells. They stained positive with the Gomori stain, which is claimed to be specific to endocrine tissues, and they incorporated granules or vesicles containing the endocrine substance purportedly released by these cells. The second major avenue of investigation centred around the brain's control of the secretion of the pituitary trophic hormones, which in turn were long known to control the secretion of the peripheral target endocrine hormones such as thyroid hormones, gonadal steroids, adrenal steroids, etc. A critical observation by several investigators had earlier demonstrated the existence of a vital neuroendocrine system, namely the magnocellular cells of the paraventricular nucleus of the hypothalamus that synthesized vasopressin and oxytocin, the nonapeptides that are transported down the axon to the nerve terminals of these neurones in the posterior pituitary (neurohypophysis) and released in response to physiological stimuli. For example, the release of vasopressin, or the antidiuretic hormone, as it is commonly named, acts as a critical regulator of fluid balance, and oxytocin is known to regulate the milk-letdown reflex during breast feeding.

It is now firmly established that neurones are indeed capable of functioning as true endocrine tissues, synthesizing and releasing substances, known as (neuro)hormones, which are released directly into the circulatory system and transported to distant sites of action. The release of vasopressin from the posterior pituitary gland and its action on the kidney is one often cited example; the action of the hypothalamic release and release-inhibiting factors on the anterior pituitary trophic hormone-producing cells is another.

Pleiotropic roles—endocrine factors and neurotransmitters

Although, it was clearly important to document the ability of neurones to function as neuroendocrine cells, particularly those in the central nervous system (CNS), the focus on an artificial classification system with clear demarcations of endocrine versus neuronal versus neuroendocrine, quickly lost its heuristic value. Indeed, we now recognize that the very same substance, often function at one site as an endocrine substance and at another as a neurotransmitter. Thus, adrenaline (epinephrine) functions as a hormone in the adrenal medulla and as a conventional neurotransmitter substance in the mammalian CNS. Similarly corticotrophin-releasing factor (CRF) functions as a true hormone in its role as a hypothalamic hypophysiotrophic factor in the hypothalamic anterior pituitary complex, yet it is apparently a 'conventional' neurotransmitter in cortical and limbic brain areas. It may act as a paracrine substance in the adrenal medulla (see Fig. 2.3.3.1). Thus, the field has progressed to the stage where we now strive to characterize the role of a particular chemical messenger in a particular region or endocrine axis. The traditional endocrine and neurotransmitter roles for several peptides alluded to above are firmly established, but the equally important paracrine roles for such substances, namely the secretion of a substance from one cell where it acts upon nearby cells, remain largely unexplored. This is, perhaps, best illustrated in the gastrointestinal tract where several peptides that function as hormones or neurotransmitter substances in other sites, including the CNS, act to influence local cellular function. Examples would include vasoactive intestinal peptide, cholecystokinin, and somatostatin.

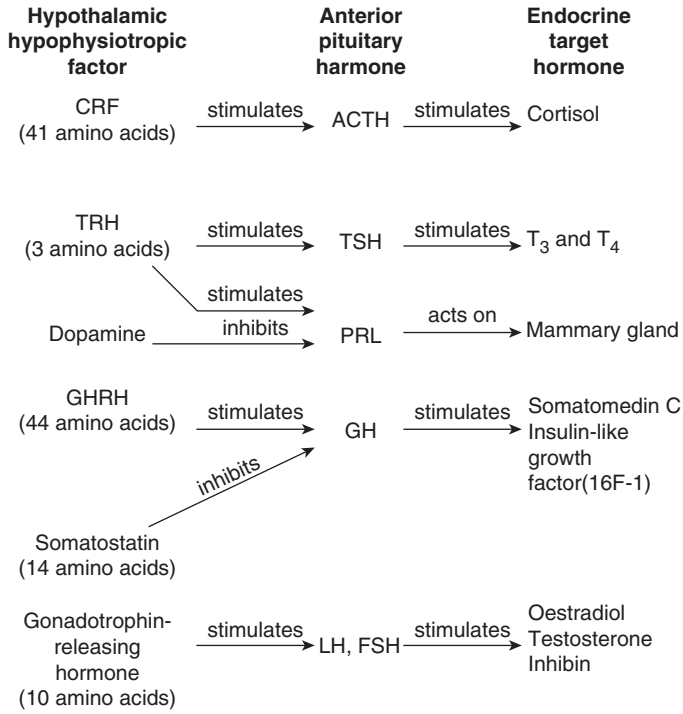


Fig. 2.3.3.1 Hormones of the hypothalamic-pituitary-end-organ axes: CRF, corticotrophin-releasing factor; TRH, thyrotrophin-releasing hormone; TSH, thyroid-stimulating hormone; T_3, tri-iodothyronine; T_4, thyroxine; PRL, prolactin; GHRH, growth-hormone releasing hormone; GH, growth hormone; LH, luteinizing hormone; FSH, follicle-stimulating hormone.

Neuroendocrinology

Neuroendocrinology thus, comprises the study of the endocrine role of neuronal or glial cells as well as the neural regulation of endocrine secretion, with a major portion of the latter consisting of the biology of the various hypothalamic-pituitary-end-organ axes and the major neurohypophyseal hormones, vasopressin and oxytocin. Because of the elegant and precise regulation of peripheral endocrine hormone secretion, afforded in part by the feedback of peripherally secreted hormones at pituitary and a variety of CNS sites, the actions of such hormones on the brain has become an integral part of this discipline. The related discipline of **psychoneuroendocrinology** arose with the realization that there are binding sites (receptors) for peripheral hormones within the CNS that have little to do with the feedback regulation of the hypothalamic-pituitary-end-organ axes, and the further recognition of the seminal role of the CNS in regulating endocrine function, for example in the effect of stress on several such measures. Often cited as beginning with the pioneering observations of Berthold, who reported that removing roosters' testosterone-secreting gonads abolished their sexual behaviour, psychoneuroendocrinology has been expanded to include the effects of hormones on behaviour, as well as the study of endocrine alterations in psychiatric disorders and, the converse, psychiatric symptomatology in endocrine disorders. Indeed, this stepchild of neuroendocrinology and psychosomatic medicine has been one of the most rapidly growing areas of research in psychiatry, now boasting an international society (International Society of Psychoneuroendocrinology), an annual meeting, and its own journal (Psychoneuroendocrinology).

Neuroendocrine window

One of the most commonly used strategies in the 1970s and 1980s, still in occasional use today, is the so-called neuroendocrine 'window' strategy. Until the relatively recent development and availability of functional brain-imaging techniques, the brain remained relatively inaccessible for study, with the exception of cerebrospinal fluid (CSF) and postmortem tissue studies. With the emergence of the monoamine theories of mood disorders and schizophrenia, many investigators attempted to draw conclusions about the activity of noradrenergic, serotonergic, and dopaminergic circuits in patients with various psychiatric disorders by measuring the basal and stimulated secretion of pituitary and end-organ hormones in plasma. There is little doubt that such an approach has severe limitations, but the results, now coupled with more modern approaches, have contributed to the substantial progress made in elucidating the pathophysiology of mood and anxiety disorders and, to a considerably lesser extent, schizophrenia. One assumption of the neuroendocrine window strategy is that the monoamine-containing neurones that regulate endocrine secretion are disordered (or not disordered) to the same extent as those monoamine circuits posited to be involved in the pathophysiology of the disorder under study. Such an assumption may well be true in neural circuits in which the cells of origin are found in a circumscribed area and project to widely diffused areas of the CNS (for example, the serotonergic and noradrenergic projections to the forebrain from the raphe and locus coeruleus cells in the brainstem, respectively). In contrast are the various dopaminergic circuits in the CNS, with their well-known topographic point-to-point

distribution. Thus, there is little reason to believe that the activity of the major dopamine-containing hypothalamic projection, the tuberoinfundibular system, with perikarya in the arcuate and periventricular hypothalamic nuclei and projections to the median eminence, is in any way related to the activity of the mesolimbicocortical dopamine pathway, with cell bodies in the ventral tegmentum of the midbrain and projections to the nucleus accumbens, amygdala, and cortical regions. This latter pathway has been implicated in the pathophysiology of schizophrenia. Thus, the study of the dopamine modulation of prolactin secretion in schizophrenia is unlikely to inform about the nature of limbic and cortical dopamine neuronal alterations in this devastating disorder.

Hypothalamic-pituitary-end-organ axes

Because a large portion of neuroendocrinology relevant to psychiatry is concerned with the hypothalamic-pituitary-end-organ axes, alterations in each of these systems in patients with a major psychiatric disorder are described later in this chapter. To avoid repetition, a generic description of the hierarchical organization of the various components of these systems is briefly outlined here. More comprehensive reviews of this subject are available.[1] In general, the hypothalamus contains neurones that synthesize release and release-inhibiting factors. These peptide hormones, summarized in Fig. 2.3.3.1, are synthesized by a process beginning with transcription of the DNA sequence for the peptide prohormone. After translation in the endoplasmic reticulum and processing during axonal transport and packaging in the vesicles in the nerve terminals, the now biologically active peptide is released from nerve terminals in the median eminence and secreted into the primary plexus of the hypothalamohypophyseal portal vessels. They are transported humorally to the sinusoids of the adenohypophysis where they act on specific membrane receptors on their specific target: the pituitary trophic hormone-producing cells. Activation of their receptors results in the release or inhibition of release of the pituitary trophic hormone. The increase or decrease in pituitary trophic hormone secretion produces a corresponding increase or decrease in end-organ hormone secretion. Thus **gonadotrophin-releasing hormone** (**GnRH**), a decapeptide, induces the release of the gonadotrophins, luteinizing hormone, and follicle-stimulating hormone from the anterior pituitary gland, which in turn stimulate the secretion of oestrogen and progesterone in women, and testosterone in men. The exogenous, intravenous administration of GnRH, comprises the GnRH stimulation test, a very sensitive test of hypothalamic-pituitary-gonadal (**HPG**) axis activity. Such stimulation tests are thought to be a sensitive measure of the activity of the axis because it is influenced by GnRH secretion, gonadotrophin secretion, and feedback at the pituitary and brain by gonadal steroids. The organization of, and major feedback mechanisms thus far demonstrated in the hypothalamic–anterior pituitary–end-organ axes are illustrated in Fig. 2.3.3.2.

In summary, neuroendocrinology (and the related discipline psychoneuroendocrinology) broadly encompasses the study of the following:

- the neural regulation of peripheral, target-organ hormone secretions, pituitary trophic hormone secretions, and secretions of the hypothalamic-hypophysiotrophic hormones;
- the role of neurotransmitter systems in the regulation of the above;

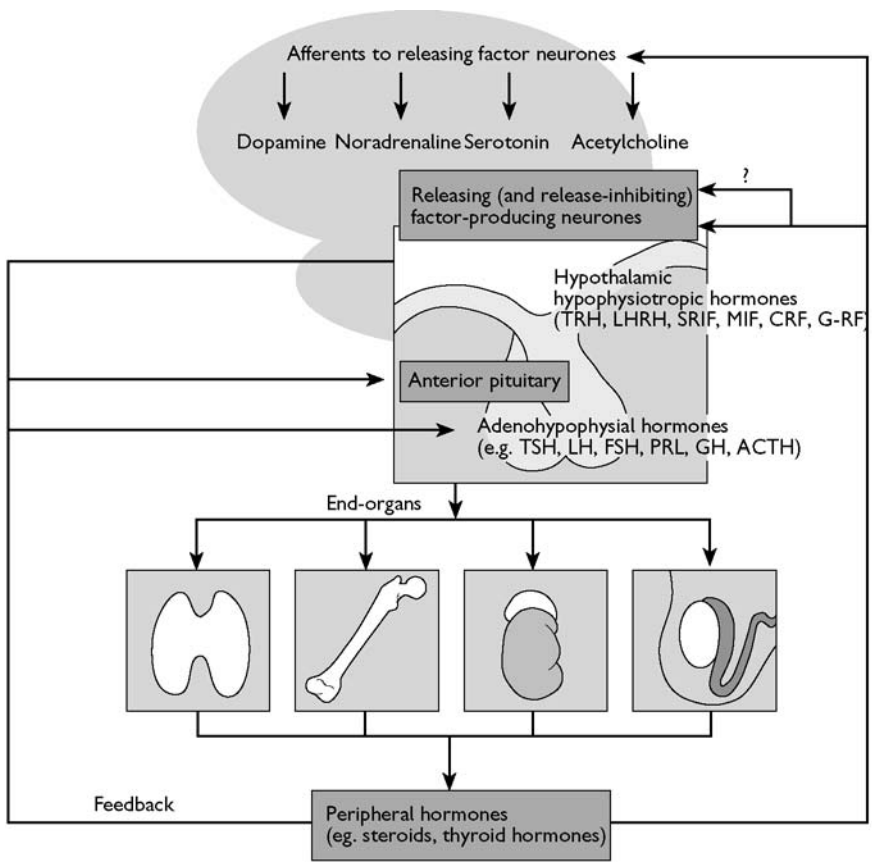


Fig. 2.3.3.2 Relationships between brain neurotransmitter systems, hypothalamic peptidergic (releasing-factor) neurones, anterior pituitary, and peripheral endocrine organs, illustrating established feedback loops: LHRH, luteinizing-hormone releasing hormone; SRIF, somatostatin; MIF, melanotrophin release-inhibiting factor; GHRF, growth-hormone releasing factor. Other abbreviations as in Fig. 2.3.3.1.

- the hormone effects of each of the endocrine axes on the CNS, alterations in the activity of the various endocrine axes in major psychiatric disorders, and conversely the behavioural consequences of endocrinopathies;
- the effects of target gland hormones on the CNS in normal individuals—for instance, the effects of synthetic glucocorticoids on memory processes.

The hypothalamic-pituitary-thyroid axis

It has been recognized for more than a century that adult patients with hypothyroidism exhibit profound disturbances in CNS function, including cognitive impairment and depression. In more recent years, attention has focused on more subtle alterations of the hypothalamic-pituitary-thyroid (**HPT**) axis in depressed patients. **Hypothyroidism** is most frequently subclassified as in four grades as follows:

- Grade 1 hypothyroidism is classic primary hypothyroidism (increased **thyroid-stimulating hormone** (**TSH**), decreased peripheral thyroid hormone concentrations, and an increased TSH response to **thyrotrophin-releasing hormone** (**TRH**)).
- Grade 2 hypothyroidism is characterized by normal, basal thyroid-hormone concentrations, but an increase in basal TSH concentrations and an exaggerated TSH response to TRH.
- Grade 3 hypothyroidism can only be detected by a TRH-stimulation test; patients have a normal basal thyroid hormone and TSH concentrations, but an exaggerated TSH response to TRH.
- Grade 4 hypothyroidism is defined as normal findings on the three thyroid axis function tests noted above, but the patients have antithyroid antibodies.

Left untreated, most, if not all, patients' progress from grade 4 to grade 1 hypothyroidism. Several studies have revealed an inordinately high rate of HPT axis dysfunction, largely hypothyroidism, in patients with major depression. In our pilot study, patients with comorbid depression and anxiety were especially likely to exhibit HPT axis abnormalities, especially the presence of grade 4 hypothyroidism, i.e. symptomless autoimmune thyroiditis. Patients with other major psychiatric diagnoses including schizophrenia and anxiety disorders appear to exhibit normal HPT axis function. For patients who require thyroid hormone replacement secondary to thyroid ablation, Bunevicius *et al.*[2] recently reported that treatment with a combination of **tri-iodothyronine** (**T3**) and **thyroxine** (**T4**) is optimal for mood and cognitive function, rather than the standard medication of T_4 alone.

In addition, a blunted TSH response to TRH is observed in approximately 25 per cent of patients with major depression. This observation, first reported by Prange *et al.*[3] and Kastin *et al.*,[4] more than 25 years ago, has been replicated in many studies. Unfortunately, the pathophysiological underpinnings of this observation remain obscure, though there is considerable evidence that it may be due, at least in part, to chronic hypersecretion of TRH and subsequent TRH-receptor downregulation in the anterior pituitary gland. Indeed, our group[5] and others have reported elevated TRH concentrations in the CSF of drug-free depressed patients.

TRH was the first of the hypothalamic-releasing factors to be chemically characterized. Immunohistochemical and radio-immunoassay methods revealed a heterogeneous brain distribution of TRH. This was the first in a series of experimental results that led to the inexorable conclusion that this peptide, and, as discussed below, other release and release-inhibiting hormones, function in extrahypothalamic brain regions as neurotransmitter substances. Thus, TRH has been shown to produce direct brain effects, independent of its action on the pituitary thyrotrophs. Antidepressant effects of intravenously and intrathecally applied TRH have been reported,[3] but the results have not been confirmed in large, controlled clinical trials.[6] In contrast, several reports over the last 30 years have documented the efficacy of T_3 (25–50 mg daily) in both accelerating the rate of onset of antidepressant response,[7] and in converting antidepressant non responders to responders,[8] the later confirmed in the recent NIMH-funded STAR-D trial.[9] Studies also suggest that T_4 (100–300 μg daily) supplementation is a viable strategy for increasing the response to antidepressants.[10] These studies were initiated when it became apparent that patients (and laboratory animals) with hypothyroidism do not respond to antidepressant agents. This led to the hypothesis that patients with subtle forms of hypothyroidism (grades 2–4) may not respond optimally to antidepressants unless they are adequately treated with exogenous thyroid hormone. With the cloning of the thyroid hormone receptor and its localization within the CNS, studies of its regulation in depression, as well as the regulation of TRH biosynthesis, can now be conducted in postmortem brain tissue.

The hypothalamic-pituitary-adrenal axis

Although, first identified in crude form in 1955 by Saffran *et al.*,[11] **corticotrophin-releasing factor** (**CRF**) was not chemically identified until 1981 by Vale and colleagues.[12] This discovery finally permitted the comprehensive assessment of hypothalamic-pituitary-adrenal (**HPA**) axis activity, and also led to scrutiny of the role of this peptide, which is now known to co-ordinate the endocrine, immune, autonomic, and behavioural effects of stress in a variety of psychiatric disorders.

CRF hypothesis of depression

Most investigators would agree that the most important finding in all of biological psychiatry is the hyperactivity of the HPA axis observed in a significant subgroup of patients with major depression. The magnitude of HPA axis hyperactivity has been reported to be correlated to the severity of the depression.[13] Literally thousands of reports on this subject have appeared since the original and independent observations of research groups led by Board, Bunney, and Hamburg, as well as by Carroll, Sachar, Stokes, and Besser. These studies, conducted from the late 1950s to the 1980s, applied the tests largely developed for the diagnosis of Cushing's syndrome to patients with major depression and other psychiatric disorders. A panoply of such tests, ranging from urinary free-cortisol, to CSF levels of cortisol, to the dexamethasone suppression test led to the inexorable conclusion that a sizeable percentage of depressed patients exhibited HPA axis hyperactivity. Our group and others, focusing on the mechanism(s) responsible for these findings, documented adrenocortical enlargement[14] and pituitary gland enlargement[15] in depressed patients, using CT and magnetic resonance imaging (**MRI**) respectively. The hypersecretion of cortisol is associated with hypersecretion of ACTH (and its co-secreted

product of the precursor pro-opiomelanocortin), which due to its trophic properties also causes adrenocortical gland enlargement. Both direct measurements of CRF in CSF,[16] and of CRF and gene expression for CRF (CRF mRNA expression) in postmortem tissue confirmed the hypothesized hypersecretion of CRF in depressed patients.[17] In addition, CRF hypersecretion results in CRF1 receptor downregulation observed in both receptor binding studies[18] and CRF mRNA expression studies conducted on post-mortem tissue from suicide victims.[19] Like the hypersecretion of cortisol, CRF hypersecretion normalizes upon recovery from depression. Indeed, there is now considerable evidence, derived from both preclinical and clinical studies, that CRF neuronal hyperactivity is reduced by treatment with several antidepressants including paroxetine and fluoxetine, the selective serotonin-reuptake inhibitors (**SSRIs**), reboxetine (the noradrenaline-reuptake inhibitor), venlafaxine (the dual noradrenaline/serotonin-reuptake inhibitor), desipramine (the tricyclic antidepressant), and tranylcypromine (the monoamine oxidase inhibitor), as well as by electroconvulsive therapy.[20]

When patients who are drug free but depressed are given CRF intravenously, they exhibit, as a group, a blunted ACTH response compared with normal control subjects. This is believed to be secondary to either the downregulation of CRF receptors on the corticotrophs after long-standing CRF hypersecretion and/or intact negative feedback of the hypersecreted cortisol at anterior pituitary and higher CNS centres. As with the other measures of HPA axis function, this endocrine abnormality normalizes upon recovery from depression. Indeed, persistent alterations in HPA axis function, whether due to dexamethasone non-suppression or CSF-CRF hypersecretion, is the harbinger of a poor response to antidepressant treatment. In recent years, Holsboer *et al.*[21] have pioneered the use of the combined **dexamethasone-CRF test**, in which patients are given the synthetic glucocorticoid and the following day receive a standardized CRF stimulation test. The results reveal that this test has a much greater sensitivity in detecting increases in HPA axis activity; it has now been used to detect axis alterations in the first-degree relatives of depressed patients who have never been symptomatic, raising for the first time the question of a trait (vulnerability) component to this measure.[22]

Space constraints preclude a more comprehensive discussion of this rich literature, but a few additional points are certainly worth interjecting. First, a robust preclinical literature has documented the depressogenic and anxiogenic effects of exogenously administered CRF in laboratory animals. When CRF was directly injected into the central nervous system it produced effects reminiscent of the cardinal symptoms of depression in patients, including decreased libido, reduced appetite and weight loss, sleep disturbances, and neophobia. Indeed, newly developed CRF1-receptor antagonists represent a novel putative class of antidepressants. Such compounds exhibit activity in virtually every preclinical screen for antidepressants and anxiolytics currently employed, and in an open study a CRF1 receptor antagonist was shown to possess antidepressant properties.[23] A second CRF receptor, the CRF2 receptor, exhibits genetic polymorphism (i.e. it occurs in more than one naturally occurring isoform or splice variant), and it is believed to utilize the urocortins as endogenous ligands. The long-term consequences of cortisol hypersecretion in depression are just now being scrutinized. One such sequela appears to be neuronal loss in the hippocampus, one of the major feedback sites for glucocorticoids in the CNS. This has now been documented by structural brain-imaging studies that utilized MRI techniques.[24] If the degeneration of neurones that represent glucocorticoid feedback sites indeed occur in depressed patients, then this should further increase HPA axis hyperactivity, which would explain the many reports of increasing adrenocortical hyperactivity of elderly depressed patients when compared with matched younger depressed patients.

Future studies will focus on the development of positron emission tomography (**PET**) ligands for both the glucocorticoid receptor and the CRF receptor, the role of the CRF binding protein—a unique protein which binds CRF in extracellular fluid and in plasma, preventing its availability to act on its receptor—and the role of the CRF peptidase, which degrades the peptide, in normal and pathological states. Finally, are the studies from our group and others, which have documented the long-term persistent increases in the HPA axis and extrahypothalamic CRF neuronal activity after exposure to early untoward life events—for example, child abuse and neglect in both laboratory animals (rats and non-human primates) as well as in both male and female patients.[20,25,26] This phenomenon has been posited to underlie the now well-documented association between early abuse and neglect and increased vulnerability to mood disorders.[27–29] Indeed, we have recently demonstrated single nucleotide polymorphisms (SNPs) of the CRF1 receptor that confer vulnerability or resistance to the development of depression after exposure to child abuse.[30] An early intervention strategy using CRF receptor antagonists may prevent such long-term alterations in the central nervous system.

HPA axis and other psychiatric disorders

HPA axis alterations have also been investigated in other psychiatric disorders. When depression is comorbid with a variety of other disorders such as multiple sclerosis, Alzheimer's disease, multi-infarct dementia, Huntington's disease, and others, both CRF hypersecretion and HPA axis hyperactivity are common. There is little evidence for HPA axis dysfunction in schizophrenia. In contrast, at least one anxiety disorder, post-traumatic stress disorder, is associated with extrahypothalamic CRF hypersecretion, as evidenced by elevated CSF concentrations of CRF,[31] but normal or reduced measures of adrenocortical activity. Finally, CRF neuronal degeneration is now well known to occur in the cerebral cortex of patients with Alzheimer's disease,[32, 33] an effect which temporally occurs prior to the better-studied cholinergic neuronal involvement. With the reduction in CRF concentrations in the cerebral cortex there is a reciprocal increase in CRF receptor density. Whether modification of the disease-associated effects on the CRF neuronal system in the cortex and hippocampus represents a novel strategy for the treatment of this common dementing disorder remains unclear at the present time.

The hypothalamic-growth hormone axis

Although the HPA and HPT axes have been more closely scrutinized in patients with psychiatric disorders, there is virtually universal agreement that the blunted growth-hormone response to a variety of provocative stimuli (particularly clonidine, an α2-adrenergic agonist) in depressed patients is the most consistent finding in affective disorders research.[34] The mechanism underlying this

phenomenon remains obscure, but it is of particular interest that, at least in some studies, it appears to persist upon recovery from depression, suggesting that it is a trait marker for depression vulnerability. There are reports of similar findings with other growth hormone-provocative stimuli, such as the use of apomorphine, desipramine, or levodopa. In addition, the blunted growth-hormone response to clonidine in depressed patients is particularly robust in those who have recently attempted suicide. Clearly, further work in this area is warranted, especially in the context of several reports of alterations in basal growth-hormone secretion in this disorder. The nature of this alteration is a reduction in the normal nocturnal rise in growth-hormone secretion, though this is not a universally agreed-upon finding. Alterations in growth-hormone secretion in other psychiatric disorders (particularly schizophrenia) have also been reported, though the results may have largely been due to long-term treatment of such patients with dopamine-receptor antagonists, antipsychotic drugs.

The secretion of growth hormone and the regulation of this axis are distinct from that of the other endocrine axes for several reasons. First, this is the one axis in which two hypothalamic-hypophysiotrophic hormones have unequivocally been shown to play a physiological role. The first discovered was **somatostatin** or **growth hormone-inhibiting hormone**. It is distributed in the CNS not only in cells of the periventricular nucleus of the hypothalamus, which projects to the median eminence, but in a variety of extrahypothalamic areas as well. Indeed, somatostatin is known to function as a CNS neurotransmitter and is of particular interest to psychiatrists because of its early involvement in the Alzheimer's disease process. Our group and others have documented the marked reduction in somatostatin concentrations in this dementing disorder.[33] In addition, somatostatin concentrations are markedly elevated in the basal ganglia of patients with Huntington's disease;[35] the pathological implications of this finding remain obscure. In contrast to the other peptide receptors described above in which one or at most two receptor subtypes have been identified, several distinct somatostatin receptor subtypes have now been structurally identified. Such diversity suggests the possibility of specific receptor-subtype agonists and antagonists as putative therapeutic agents.

Several years after the elucidation of the structure of somatostatin, the long-postulated **growth hormone-releasing factor** (**GHRF**) was found. This peptide has the most limited CNS distribution of all the hypothalamic-releasing hormones thus far studied. The vast majority of the peptide is found in the arcuate nucleus of the hypothalamus from where it projects nerve terminals to the median eminence. Unlike the other axes, the growth-hormone axis is also unique in not having a single target endocrine gland. Indeed, growth hormone does stimulate the release of **somatomedin C** from the liver and it also exerts direct effects on a variety of targets including bone and muscle. Most, but not all, investigators have reported a blunted growth-hormone response to GHRF in depressed patients, but the total number of patients studied pales in comparison to the TRH- and CRF-stimulation test data. There are no data published on somatomedin C responses to GHRF in depressed patients. No published studies are available in which GHRF concentrations or GHRF-mRNA expression have been studied in postmortem tissue of depressed patients and matched controls, an obvious study in view of the data reviewed above.

The hypothalamic-pituitary-gonadal axis

In view of the remarkable gender differences in the prevalence rate of depression, the relatively high rates of postpartum depression, as well as the reduction in libido that is so characteristic of depression, it is plausible to posit a reduction in HPG axis activity in depressed patients. Therefore, it is somewhat surprising that so little research has been conducted on HPG axis activity in depression and other psychiatric disorders. Indeed, a comprehensive database on this extraordinarily important area is simply not available, but the field has been reviewed.[36] A series of older studies documented no differences in basal gonadotrophin levels in depressed patients when compared to controls. The **gonadotrophin-releasing hormone** (**GnRH**) stimulation test has only been administered to a relatively small number of depressed patients; although the results revealed a blunted or normal response, no firm conclusions can be drawn from this limited data set. Indeed, such studies require control for menopausal status, menstrual-cycle phase, use of oral contraceptives, as well as the measurement of baseline progesterone, oestrogen, and gonadotrophin plasma concentrations. One remarkable finding relevant to these questions is the remarkable effectiveness of the GnRH agonist, leuprolide, in the treatment of the premenstrual syndrome. It is believed to act by producing a chemical ovariectomy through a marked downregulation of adenohypophyseal GnRH receptors and the expected resultant reduction in gonadotrophin and gonadal steroid secretion. Long-term treatment with this compound could theoretically result in bone-density reductions and a risk of cardiovascular disease. Therefore, supplementation with oestrogen and progesterone has been suggested in combination with leuprolide, though there are reports that such a strategy reduces the effectiveness of the treatment.

GnRH was the second of the hypothalamic-hypophysiotrophic hormones to be chemically characterized. It is relatively limited in distribution to hypothalamic regions and to the preoptic area and septum. It stimulates the secretion of both luteinizing hormone and follicle-stimulating hormone in both men and women. GnRH is known to act by stimulating GnRH receptors in the anterior pituitary gland, which results in the increased synthesis and release of the pituitary gonadotrophins, in turn causing the release of oestrogen and progesterone in women and testosterone in men. There is some evidence that oestrogens, which have receptors localized extensively throughout the CNS, may possess some antidepressant activity, in perimenopausal and postpartum depression, though the data is far from clear.[37] There are hints from the clinical trial literature that postmenopausal women on oestrogen replacement may respond better to treatment with fluoxetine than women who are not receiving such treatment,[38] though the database is small and fraught with many confounds. Further, emphasizing the potential role of sex steroids in affective disorders is the clinical study that demonstrates that in hypogonadal depressed men, testosterone treatment possessed antidepressant properties.[39]

The hypothalamic-prolactin axis

Prolactin, a pituitary hormone which acts on the mammary gland, plays a critical role in lactation. Unlike the other axes described, this axis is unique in having a non-peptide release-inhibiting factor, dopamine. In addition, although there is relatively strong

evidence for the existence of a prolactin-releasing factor, its isolation and characterization has not yet been realized. One of the difficulties in completing this task is the presence of TRH, which is a potent prolactin-releasing factor, and may in fact function physiologically in this regard. Interestingly, although the TSH response to TRH in depressed patients is often blunted, the prolactin response is not. Although the results are not unequivocal, most studies have not observed alterations of prolactin secretion in depressed patients.[40] In contrast to this small database is a remarkably large database on the use of provocative tests of prolactin secretion in patients with psychiatric disorders. To summarize briefly, the prolactin response to agents that increase serotonergic neurotransmission such as l-tryptophan, 5-hydroxytryptophan (5-HTP), l-(+)- and d-(+)-fenfluramine, clomipramine, and also to direct serotonin-receptor agonists, is blunted in depressed patients, as well as in patients with cluster-B personality disorder and borderline personality disorder. The available data would suggest that this blunted prolactin response is mediated by alterations in 5-HT1A-receptor responsiveness.

Discussion and summary

Although, the hypothesis that the neuroendocrine window strategy would ultimately provide the long searched for information concerning the nature of monoamine circuit alterations in patients with psychiatric disorders has never been realized, the approach has led to major advances in biological psychiatry. It has led to the **CRF hypothesis of depression**, which is supported by a considerable multidisciplinary database, and this in turn has directed the field towards the development of novel therapeutic approaches, namely CRF receptor antagonists. It also apparently explains the neurobiological mechanisms responsible for the increase in depression (first postulated by Freud in the early part of the 20th century) in patients exposed to trauma during their early life. If CRF is indeed the 'black bile' of depression, responsible for the endocrinopathy of depression, as well as several of the other cardinal features of this disorder, then CRF-receptor antagonists should represent a novel class of antidepressants that will be a welcome addition to the armamentarium. Indeed, a number of pharmaceutical companies are now testing CRF-receptor antagonists as novel anxiolytics and antidepressants in preclinical studies and clinical trials.

In addition to the now widely replicated HPA axis and CRF alterations in depression, are the HPT axis abnormalities. Most depressed patients, in fact, exhibit alterations in one of these two axes. Furthermore, there is the widely replicated blunting of the growth-hormone response to clonidine and other provocative stimuli and the blunted prolactin response to serotonergic stimuli in depressed patients. The vast majority of studies have focused on patients with mood disorders, particularly unipolar depression. Clearly other disorders, including eating disorders, anxiety disorders, schizophrenia, and axis II diagnoses should also be critically evaluated and compared to the literature on depression. The original neuroendocrine window strategy may well have failed in terms of gleaning information about monoamine-circuit activity, but the mechanistic studies that followed have been remarkably fruitful. As repeatedly noted above, the availability of ligands that can be utilized with positron-emission tomography to determine peptide-receptor alterations in the brain and pituitary of patients with psychiatric disorders will advance the field, as will the long-elusive ability to measure receptors for the endocrine target hormones (glucocorticoids, oestrogens, thyroid hormones, etc.) in the brains of patients with these severe mental illnesses.

Finally, it is important to note the increasing database suggesting that depression is a systemic disease with major implications for vulnerability to other disorders. Thus, depressed patients are much more likely to develop coronary artery disease and stroke, and perhaps cancer. They have been shown to have reduced bone density, rendering them more at risk for hip fracture and increasing a variety of measures of inflammation. Such findings may well be mediated by the described endocrine alterations in depression. This should provide a further impetus for investigating the neuroendocrinology of psychiatric disorders.

Further information

Psychoneuroendocrinology http://www.elsevier.com/wps/find/journaldescription.cws_home/473/description#description

Hormones and Behavior http://www.elsevier.com/wps/find/journaldescription.cws_home/622842/description#description

Neuroendocrinology http://content.karger.com/ProdukteDB/produkte.asp?Aktion=JournalHome&ProduktNr=223855

References

1. Campeau, S., Day, H.E.W., Helmreich, D.L., *et al.* (1998). Principles of psychoneuroendocrinology. *Psychiatric Clinics of North America*, **21**, 259–76.
2. Bunevicius, R., Kazanavicius, G., Zalinkevicius, R., *et al.* (1999). Effects of thyroxine as compared with thyroxine plus triiodothyronine in patients with hypothyroidism. *New England Journal of Medicine*, **340**, 424–9.
3. Prange, A.J., Wilson, I.C., Lara, P.P., *et al.* (1972). Effects of thyrotropin-releasing hormone in depression. *Lancet*, **2**, 999–1002.
4. Kastin, A.J., Ehrensing, R.H., Schlach, D.S., *et al.* (1972). Improvement in mental depression with decreased thyrotropin response after administration of thyrotropin-releasing hormone. *Lancet*, **ii**, 740–2.
5. Banki, C.M., Bissette, G., Arato, M., *et al.* (1988). Elevation of immunoreactive CSF TRH in depressed patients. *American Journal of Psychiatry*, **145**, 1526–31.
6. Nemeroff, C.B. and Evans, D.L. (1989). Thyrotropin-releasing hormone (TRH), the thyroid axis and affective disorder. *Annals of the New York Academy of Sciences*, **553**, 304–10.
7. Prange, A.J., Wilson, I.C., Rabon, A.M., *et al.* (1969). Enhancement of imipramine antidepressant activity by thyroid hormone. *American Journal of Psychiatry*, **126**, 457–69.
8. Joffe, R.T., Singer, W., Levitt, A.J., *et al.* (1993). A placebo-controlled comparison of lithium and triiodothyronine augmentation of tricyclic antidepressants in unipolar refractory depression. *Archives of General Psychiatry*, **50**, 387–94.
9. Nierenberg, A.A., Fava, M., Trivedi, M.H., *et al.* (2006). A comparison of lithium and T(3) augmentation following two failed medication treatments for depression: a STAR*D report. *American Journal of Psychiatry*, **163**, 1519–30.
10. Baumgartner, A. (2000). Thyroxine and the treatment of affective disorders: an overview of the results of basic and clinical research. *International Journal of Neuropsychopharmacology*, **3**, 149–65.
11. Saffran, M., Schally, A.V., and Benfey, B.G. (1955). Stimulation of the release of corticotropin from the adenohypophysis by a neurohypophysial factor. *Endocrinology*, **57**, 439–44.
12. Vale, W., Spiess, J., Riveir, C., *et al.* (1981). Characterization of a 41 residue ovine hypothalamic peptide that stimulates secretion of corticotropin and β-endorphin. *Science*, **213**, 1394–7.

13. Evans, D.L. and Nemeroff, C.B. (1987). The clinical use of the dexamethasone suppression test in DSM-III affective disorders: correlation with the severe depressive subtypes of melancholia and psychosis. *Journal of Psychiatric Research*, **21**, 185–94.
14. Nemeroff, C.B., Krishnan, K.R.R., Reed, D., *et al.* (1992). Adrenal gland enlargement in major depression: a computed tomography study. *Archives of General Psychiatry*, **49**, 384–7.
15. Krishnan, K.R.R., Doraiswamy, P.M., Lurie, S.N., *et al.* (1991). Pituitary size in depression. *Journal of Clinical Endocrinology and Metabolism*, **72**, 256–9.
16. Heit, S., Owens, M.J., Plotsky, P., *et al.* (1997). Corticotropin-releasing factor, stress and depression. *Neuroscientist*, **3**, 186–94.
17. Raadsheer, F.C., Hoogendijk, W.J.G., Stam, F.C., *et al.* (1994). Increased number of corticotropin-releasing hormone neurones in the hypothalamic paraventricular nuclei of depressed patients. *Neuroendocrinology*, **60**, 436–44.
18. Nemeroff, C.B., Owens, M.J., Bissette, G., *et al.* (1988). Reduced corticotropin releasing factor binding sites in the frontal cortex of suicide victims. *Archives of General Psychiatry*, **45**, 577–9.
19. Merali, Z., Du, L., Hrdina, P., *et al.* (2004). Dysregulation in the suicide brain: mRNA expression of corticotropin-releasing hormone receptors and GABA(A) receptor subunits in frontal cortical brain regions. *Journal of Neuroscience*, **24**, 1478–85.
20. Nemeroff, C.B. (1999). The preeminent role of early untoward experience on vulnerability to major psychiatric disorders: the nature-nurture controversy revisited and soon to be resolved. *Molecular Psychiatry*, **4**, 106–8.
21. Holsboer, F., Von Bardeleben, U., Weidemann, K., *et al.* (1987). Serial assessment of corticotropin-releasing hormone response after dexamethasone in depression-implications for pathophysiology of DST nonsuppression. *Biological Psychiatry*, **22**, 228–34.
22. Holsboer, F., Lauer, C.J., Schreiber, W., *et al.* (1995). Altered hypothalamic-pituitary-adrenocortical regulations in healthy subjects at high familial risk for effective disorders. *Neuroendocrinology*, **62**, 340–7.
23. Zobel, A.W., Nickel, T., Kunzel, H.E., *et al.* (2000). Effects of the high-affinity corticotropin-releasing hormone receptor 1 antagonist R121919 in major depression: the first 20 patients treated. *Journal of Psychiatric Research*, **34**, 171–81.
24. Sheline, Y.I., Wang, P.W., Gado, M.H., *et al.* (1996). Hippocampal atrophy in recurrent major depression. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 3908–13.
25. Coplan, J.D., Andrews, M.W., Rosenblum, L.A., *et al.* (1996). Persistent elevations of cerebrospinal fluid concentrations of corticotropin-releasing factor in adult non-human primates exposed to early life stressors: implications for the pathophysiology of mood and anxiety disorders. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 1619–23.
26. Heim, C., Newport, D.J., Bonsall, R., *et al.* (2001). Altered pituitary-adrenal axis responses to provocative challenge tests in adult survivors of childhood abuse. *American Journal of Psychiatry*, **158**, 575–81.
27. Kendler, K.S., Neale, M.C., Kessler, R.C., *et al.* (1992). Childhood parental loss and adult psychopathology in women. A twin-study perspective. *Archives of General Psychiatry*, **49**, 109–16.
28. McCauley, J., Kern, D., Kolodner, K., *et al.* (1997). Clinical characteristics of women with an history of childhood abuse. *Journal of the American Medical Association*, **277**, 1362–8.
29. Agid, O., Shapira, B., Zislin, J., *et al.* (1999). Environment and vulnerability to major psychiatric illness: a case control study of early parental loss in major depression, bipolar depression and schizophrenia. *Molecular Psychiatry*, **4**, 163–72.
30. Bradley, R.G., Binder, E.B., Epstein, M.P., *et al.* (2007). Influence of child abuse and trauma on adult depression is moderated by the corticotrophin releasing hormone receptor gene. *In review*.
31. Bremner, J.D., Licinio, J., Darnell, A., *et al.* (1997). Elevated CSF corticotropin-releasing factor concentrations in posttraumatic stress disorder. *American Journal of Psychiatry*, **154**, 624–9.
32. Bissette, G., Reynolds, G.P., Kilts, C.D., *et al.* (1985). Corticotropin-releasing factor-like immunoreactivity in senile dementia of the Alzheimer type: reduced cortical and striatal concentrations. *Journal of the American Medical Association*, **254**, 3067–9.
33. Bissette, G., Cook, L., Smith, W., *et al.* (1998). Regional neuropeptide pathology in Alzheimer's disease: corticotropin-releasing factor and somatostatin. *Journal of Alzheimer's Disease*, **1**, 1–15.
34. Checkley, S.A., Slade, A.P., and Shur, P. (1981). Growth hormone and other responses to clonidine in patients with endogenous depression. *British Journal of Psychiatry*, **138**, 51–5.
35. Nemeroff, C.B., Youngblood, W.W., Manberg, P.J., *et al.* (1983). Regional brain concentrations of neuropeptides in Huntington's chorea and schizophrenia. *Science*, **221**, 972–5.
36. Young, E. and Korszun, A. (2002). The hypothalamic-pituitary-gonadal axis in mood disorders. *Endocrinology and Metabolism Clinics of North America*, **31**, 63–78.
37. Schmidt, P., Rubinow, D., Neuman, L., *et al.* (2000). Estrogen replacement in perimenopause-related depression: a preliminary report. *American Journal of Obstetrics and Gynecology*, **183**, 414–20.
38. Schneider, L.S., Small, G.W., Hamilton, S.H., *et al.* and the fluoxetine collaborative study group (1997). Estrogen replacement and response to fluoxetine in a multicenter geriatric depression trial. *American Journal of Geriatric Psychiatry*, **5**, 97–106.
39. Pope, H.G., Cohane, G.H., Kanayama, G., *et al.* (2003). Testosterone gel supplementation for men with refractory depression: a randomized, placebo-controlled trial. *American Journal of Psychiatry*, **160**, 105–11.
40. Nicholas, L., Dawkins, K., and Golden, R. (1998). Psychoneuroendocrinology of depression-prolactin. *Psychiatric Clinics of North America*, **21**, 341–58.

Nemeroff, Charles B (3-27-08)

Research Grants
NIH

Scientific Advisory Board
AFSP; AstraZeneca; Forest Laboratories; NARSAD; Quintiles; Janssen/Ortho-McNeil, PharmaNeuroboost, Mt. Cook Pharma, Inc

Stockholder or Equity
Corcept; Revaax; NovaDel Pharma; CeNeRx, PharmaNeuroboost

Board of Directors
American Foundation for Suicide Prevention (AFSP); George West Mental Health Foundation; NovaDel Pharma, Mt. Cook Pharma, Inc

Patents
Method and devices for transdermal delivery of lithium (US 6,375,990 B1)
Method to estimate serotonin and norepinephrine transporter occupancy after drug treatment using patient or animal serum (provisional filing April, 2001)

Neigh, Gretchen N. (7-11-08)

Research Grants
NIH, NARSAD, AHA

2.3.4 Neurotransmitters and signalling

Trevor Sharp

By the end of the 19th century it was recognized that signalling from one neurone to the next occurs at specialized contacts – Sherrington coined the term 'synapse'. It took another 50 years for scientists to accept that information passes between neurones principally through the movement across synapses of chemicals and not electrical current. Today changes in chemical transmission at brain synapses are accepted as being key to the successful drug treatment, and cause, of many forms of psychiatric illness. This article focuses on general aspects of chemical transmission and describes some recent advances relevant to psychiatry that point the direction of future research.

Otto Loewi identified the first chemical neurotransmitter, acetylcholine, in 1921. Today evidence suggests that there are many tens if not hundreds of molecules in the brain that have neurotransmitter properties. These molecules include not only the three major classes of neurotransmitters—amines, amino acids and neuropeptides—but also specific purines, trophic factors, inflammatory mediators (chemokines and cytokines), lipids, and even gases. Examples of molecules that serve neurotransmitter functions in the brain are listed in Table 2.3.4.1. This list is not exhaustive and more are likely to be discovered.

Table 2.3.4.1 Examples of neurotransmitters in the brain

Chemical class	Example
Amines	Dopamine Noradrenaline 5-hydroxytryptamine Histamine Melatonin Acetylcholine
Amino acids	γ-aminobutyric acid (GABA) Glutamate Glycine
Neuropeptides	Substance P Leu- and Met-enkephalin Galanin
Purines	Adenosine Adenosine triphosphate (ATP)
Neurotrophic factors#	Neurotrophins (e.g. BDNF, NGF) Insulin-like growth factor (IGF) Vascular endothelial growth factor (VEGF)
Cytokines*	Interleukin-1 (IL-1) Tumor necrosis factor-α (TNF-α)
Chemokines*	CC Chemokines (e.g. Interleukin-8 [IL-8]) CXC Chemokines
Endocannabinoids#	Anandamide 2-Arachidonyl-glycerol (2-AG)
Gases#	Nitric oxide (NO) Carbon monoxide (CO)

*Putative class of neurotransmitters.

#Retrograde messengers.

Basic principles of chemical transmission

Typically a molecule is classified as a neurotransmitter if it is localized in neurones, released from nerve terminals (and often soma and dendrites) on membrane depolarization, and exerts physiological and molecular effects through acting on postsynaptic receptors. However, the degree to which a particular molecule satisfies these criteria may vary. For example, the term 'neurotransmitter' was once used to cover only those molecules that exert fast synaptic effects, whereas molecules that exerted slow synaptic effects were termed 'neuromodulators'. These distinctions are less useful today (and will not be used herein) as it is recognized that many molecules are capable of exerting both fast and slow synaptic effects, depending on the receptor activated. Moreover, it is now recognized that certain molecules transfer information at a synapse in a 'retrograde' direction. In this case the molecules are located in the postsynaptic neurone and, when their synthesis is activated, the molecule diffuses across the synapse to act presynaptically.

The general principles of the chemical transmission at central synapses are similar across all neurotransmitter molecules but there are often important differences between molecules especially across molecules of different size, in particular small molecule transmitters such as amines and amino acids, and peptides (Fig. 2.3.4.1).[1]

Small molecule neurotransmitters

Typically small neurotransmitter molecules such as amines and amino acids are synthesized at the nerve terminal by one or a few enzymatic steps and packaged in small vesicles via proton-coupled vesicular transporters, prior to release into the synapse on arrival of a depolarizing action potential. After release, the neurotransmitter diffuses across the synapse to interact with receptors on the postsynaptic neurone to trigger electrical and/or biochemical changes in the postsynaptic cell. Small molecule neurotransmitters are also released from the soma and dendrites of neurones, one purpose being to interact with presynaptic receptors that signal negative feedback to the neurone.

Once released the small neurotransmitters are selectively taken up by sodium-coupled transporters that are located in the plasma membrane of the nerve terminal or neighbouring cells (neurones or glial cells). This transport terminates transmission at the postsynaptic receptor, maintains low extracellular levels of transmitter, and allows reuse of the neurotransmitter by the neurone. Transport into the nerve terminal also presents the transmitter to catabolic enzymes to generate biologically inactive metabolites, for instance, monoamine oxidase in the case of the amine transmitters.

Neurotransmitter transporters

Advances in cloning technology has lead to new discoveries regarding the structural and pharmacological identity of transporters located on the plasma membrane, as well as vesicular transporters

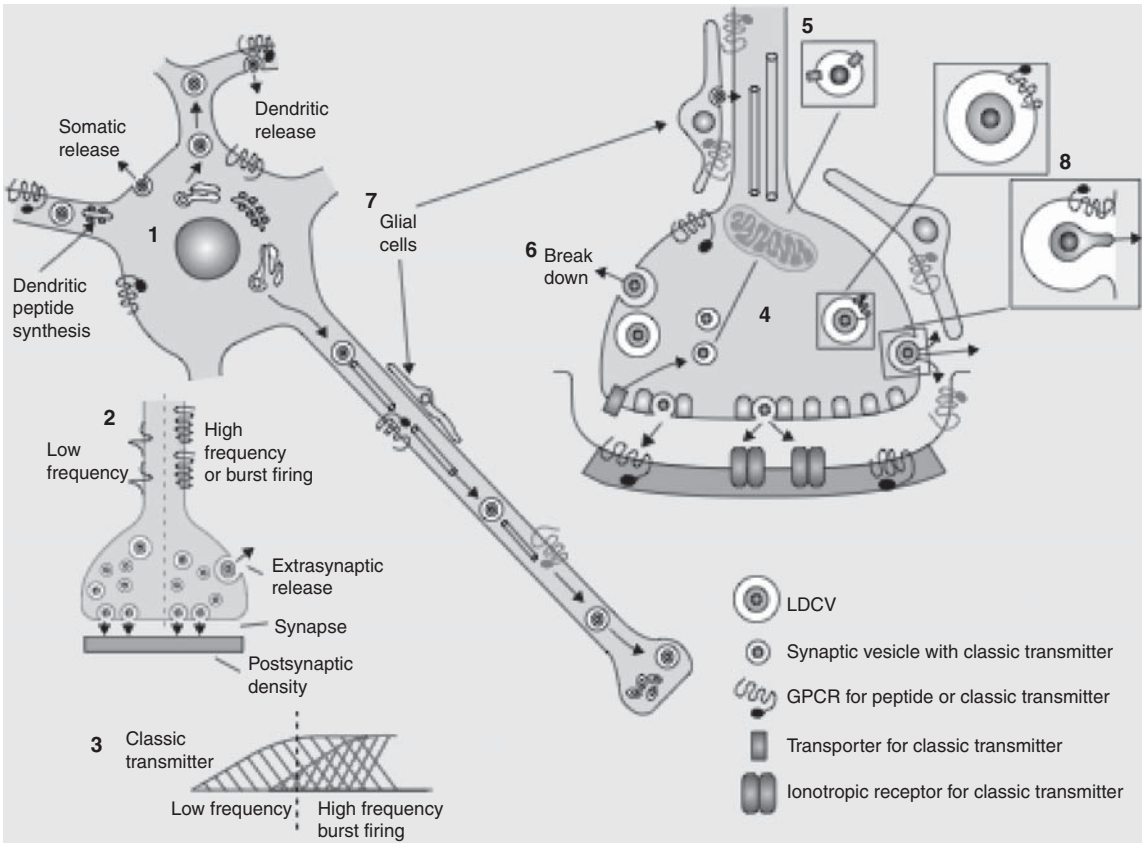


Fig. 2.3.4.1 Summary of the principal steps involved in chemical neurotransmission in the brain. Neuropeptides are synthesized in the cell body and then packaged in large dense core vesicles (LDCVs) that are transported into axons and dendrites (1). Small 'classic' neurotransmitters (e.g. amines and amino acids), are synthesized at the nerve terminal and stored in synaptic vesicles, and released into the synaptic cleft (2). LDCVs contain proteolytic enzymes (convertases) that generate the active neuropeptide from the precursor. Neurotransmitter receptors are either of the G-protein-coupled (metabotropic) or ligand-gated ion channel (ionotropic) type and are present on cell soma, dendrites, axons, and nerve endings (1,4). The small neurotransmitters are released during low and high frequency firing, whereas neuropeptides are preferentially released under burst or high frequency firing (2–4). Small transmitters have reuptake mechanisms (transporters) at both the plasma membrane and the vesicle membrane (5), which terminate neurotransmitter action and allow recycling (4). In contrast, neuropeptides are broken down by extracellular peptidases (6), and replacement occurs via axonal transport. Glial cells can express neurotransmitter receptors and transporters (7). Receptors are trafficked to and from the cell membrane by G-protein interacting proteins (8). Reprinted from The Lancet Neurology, 2, Hokfelt T, Bratfai T, Bloom F, Neuropeptides: opportunities for drug discovery, 465, copyright 2003, with permission from Elsevier.

located inside the nerve terminal.[2] The latter transporters concentrate transmitters in synaptic vesicles prior to release and play a key role in determining the neurotransmitter phenotype of a neurone.[3] A summary of plasma membrane and vesicular transporters is given in Table 2.3.4.2.

Plasma membrane transporters

Specific plasma membrane transporters for the amine neurotransmitters, dopamine (DAT), noradrenaline (NET), 5-HT (SERT), have been identified, sequenced and, investigated in detail at the molecular level because they are the site of action of stimulants such as amphetamines and cocaine, as well as many antidepressant drugs.[2] Indeed, the drug-binding site and the precise molecular mechanism of transporter inhibition by tricyclic antidepressant drugs have recently been revealed.[4]

Molecular cloning techniques have uncovered the genes for four transporters for the inhibitory neurotransmitter GABA; GAT-1, GAT-2, GAT-3 and BTG-1, the latter being localized primarily in the kidney. These transporters are highly homologous but pharmacologically distinct, with GAT-1 and GAT-3 being the most abundant subtypes and present on both neurones and glial cells. Although the significance of these transporters is yet to be fully understood, they display overlapping but different expression patterns in the CNS, suggesting distinct functional roles. The anticonvulsant effect of tiagabine is likely mediated by blockade of GAT-1, and there is much scope for new GABA uptake inhibitors of as yet unclear utility.

Glycine, another inhibitory amino acid transmitter, also has specific transporters located preferentially on the plasma membranes of glial cells in the forebrain (GLYT1) and neurones of the hindbrain and spinal cord (GLYT2). Interestingly, glycine is a positive allosteric co-modulator of glutamate NMDA receptors, and glycine

Table 2.3.4.2 Example of neurotransmitter transporters in the brain

Neurotransmitter	Transporter name
Plasma membrane transporters	
Dopamine	DAT
Noradrenaline	NET
5-Hydroxytryptamine	SERT
GABA	GAT-1
	GAT-2
	GAT-3
	BGT-1 (primarily in kidney)
Glutamate	EAAT-1 (GLAST1)
	EAAT-2 (GLT-1)
	EAAT-3 (EAAC1)
	EAAT-4
	EAAT-5
Glycine	GLYT-1
	GLYT-2
Acetylcholine (choline)	CHT
Vesicular transporters	
Monoamines (dopamine, noradrenaline, 5-HT, histamine)	VMAT1
	VMAT2
GABA	
Glutamate	VGAT
	VGLUT1
	VGLUT2
Acetylcholine	VGLUT3
	VAchT

transport blockade may offer a means to facilitate the functioning of this receptor without incurring excitotoxic effects. The antipsychotic potential of glycine transport blockers is currently under investigation because of consistent evidence of a link between the symptoms of schizophrenia and low glutamate function.[2]

Four transporters for the excitatory amino acid neurotransmitter, glutamate, have been cloned; EAAT1 (excitatory amino acid transporter1; synonym GLAST), EAAT2 (synonym GLT1), EAAT3, EAAT4, and EAAT5.[5] These transporters are located on both neurones (EAAT3) and glial cells (predominantly EAAT1/2) and serve to maintain low extracellular concentrations of glutamate, as well as providing a source of intracellular glutamate for metabolism. Whilst pharmacological blockade of glutamate transport increases excitotoxicity, pharmacologically enhanced EAAT expression appears to be neuroprotective.[6]

Vesicular transporters

Vesicular monoamine transporter 2 (VMAT2) has been identified and shown to be present in neurones of dopamine, noradrenaline and 5-HT (and histamine). In addition, vesicular monoamine transporter 1 (VMAT1) is an integral protein in the membrane of secretory vesicles of neuroendocrine and endocrine cells. Reserpine is a blocker of VMAT and causes depletion of amines, and the drug's tranquillizer effects are directly linked to this action. Acetylcholine is loaded into synaptic vesicles by a distinct transporter, VAchT.

Three homologous vesicular transporters for glutamate, VGLUT1, VGLUT2, and VGLUT3, have been cloned and characterized at the molecular level. Whilst all possess similar molecular properties, they are expressed on different neurone populations.[3] Interestingly, VGLUT3 is found in amine-containing neurones, raising the possibility of glutamate being a co-transmitter in these neurones. The vesicular transporter for GABA is VGAT. Loss of VGAT causes a drastic reduction in release of not only GABA but also glycine, indicating that glycinergic neurones do not express a separate vesicular glycine transporter.

Neuropeptides

Following the chemical identification of the first neuropeptide Substance P in 1971, evidence has accumulated that numerous peptides play neurotransmitter roles in the brain.[1] Some examples are shown in Table 2.3.4.3. The neuropeptides comprise 3–100 amino acids and together with other putative signalling peptides such as growth factors and cytokines, are synthesized in the nucleus by DNA transcription followed by translation from mRNA into precursor polypeptides (Fig. 2.3.4.1). These precursors typically undergo extensive post-translational processing that includes cleavage into smaller peptides by endopeptidases as well as other enzymic modifications. The precursor peptide usually contain an N-terminal signal sequence that directs the transport of newly synthesized protein to the lumen of the endoplasmic reticulum, and then the Golgi complex where the peptide is packaged into vesicles (so called large dense core vesicles) that are transported along the axon to the synapse. This obviates the need for neuropeptide vesicular transporters.

Proteolytic processing of a single precursor peptide often generates not one but a family of biologically active peptides, although the proteolytic steps may be tissue-specific. The opioid peptides provide one of the best worked out examples of this form of processing. Proopiomelanocortin (POMC) is a hypothalamic precursor opioid peptide whose structure contains sequences for adrenocorticotropic hormone (ACTH), α-melanocyte stimulating hormone (α-MSH) and β-endorphin. In the anterior lobe of the pituitary gland POMC is processed to form ACTH, whilst in the intermediate lobe POMC is processed to form α-MSH and β-endorphin. On the other hand, post-translational processing of the opioid precursor peptide, proenkephalin, gives rise to multiple copies of the pentapeptide met-enkephalin as well as a copy of leu-enkephalin, whilst a third opioid precursor, prodynorphin, gives rise to dynorphin. In total, the 3 separate opioid peptide genes give rise to at least 18 endogenous peptides with opiate-like activity.

Multiple proteolytic enzymes have been cloned and extensively characterized, including prohormone convertases that produce striking phenotypic effects when genetically manipulated in mutant mouse models.[7] The therapeutic utility of pharmacological manipulation of neuropeptide synthesis and degradation in the brain has yet to be realized. However, the success of inhibitors of the prohormone convertase that synthesizes angiotensin in the periphery (ACE inhibitors), for the treatment of hypertension, sets an important precedent.

Another mechanism to generate neuropeptide diversity is through alternative RNA splicing of a single gene. For example, in the case of the tachykinins alternative splicing of preprotachykinin gene A mRNA results in 3 splice variants which, after translation and post-translational processing, collectively generate the five biologically active peptides of the tachykinin family (including Substance P).

Table 2.3.4.3 Examples of families of neuropeptides

Opioid peptides	Leu-enkephalin Met-enkephalin Dynorphin β-endorphin Nociceptin
Tachykinins	Substance P Neurokinin A Neurokinin B
Hypothalamic releasing factors	Thyrotrophin releasing factor (TRH) Corticotrophin releasing factor (CRF) Growth hormone releasing hormone (GHRH) Somatostatin
Gut-brain peptides	Cholecystokinin (CCK) Galanin Insulin Neurotensin Neuropeptide Y (NPY) Vasointestinal polypeptide (VIP)
Other peptides	Bradykinin Calcitonin gene-related peptide Melanin concentrating hormone (MCH) Melanocortin Orexin Oxytocin Vasopressin

To date, there is little evidence that neuropeptides are cleared from the synapse by transporters in the plasma membrane, indicating that these particular transmitters are not recycled after release. Rather, evidence suggests that their action is terminated by peptidases that are thought to be located on extracellular membranes. Thus, replenishment of neuropeptides during high levels of synaptic activity is dependent on the proteolytic enzymes that generate the active peptides in the neurones.

A feature of most if not all neuropeptides is their co-localization with classic neurotransmitters. Some of the best examples include co-localization between GABA/dynorphin in movement control pathways (striatonigral neurones), CCK/dopamine in reward pathways (mesoaccumbens neurones), and glutamate/Substance P in pain pathways (dorsal root ganglion neurones). The functional significance of this co-localization is not fully clear but evidence suggests that peptide release requires higher frequencies of neuronal discharge than classic transmitters, and once released the neuropeptide either facilitates or opposes the function of the co-localized transmitter.[1] In a recent example, co-localization between 5-HT and galanin in midbrain raphe neurones was investigated to reveal an action of the peptide on 5-HT feedback mechanisms. This knowledge has been exploited to develop galanin ligands that are under development as novel antidepressant strategies.[8]

Neurotrophic factors

Neurotrophic factors are brain peptides that were originally recognized for their role in supporting growth, differentiation and survival of neurones but today are thought to possess many of the properties of neurotransmitters including neuronal localization and release, and an ability to modulate synaptic function. Moreover, there is evidence that trophic factors signal in a retrograde fashion (see later). Neurotrophic factors are currently named according to the action with which they were originally characterized (brain-derived neurotrophic factor — BDNF, nerve growth factor — NGF) and they comprise many families.[9]

Certain features distinguish neurotrophic factors from neuropeptides (see above). In particular, neurotrophic factors are larger molecules; for example, BDNF has a molecular size of 14 kDa whereas neuropeptides are typically much smaller peptides. Also whilst neuropeptides signal via G-protein coupled receptors (see below), neurotrophic factors signal via direct activation of a class of transmembrane spanning protein kinases called protein tyrosine kinases (Trk receptors) of which four types have been identified so far (TrkA, TrkB, TrkC and p75), and that phosphorylate proteins on tyrosine residues.[10] In some cases, the neurotrophic factor receptor and protein tyrosine kinase reside in the same protein, while in other cases, the receptor recruits an intracellular protein tyrosine kinase. Specific neurotrophic factors signal via specific protein kinases (e.g. NGF – via TrkA, BDNF – via TrkB). Activation of the protein tyrosine kinase triggers cascades of further protein phosphorylation that produce not only trophic effects but also changes in synaptic transmission.

Much recent interest in neurotrophic factors derives from findings that they regulate synaptic transmission in the adult brain, and that neurotrophic factor expression can be modulated through interactions with amine and amino acid neurotransmitters. For example, evidence suggests that repeated administration with amine-targeted antidepressants increases BDNF expression in animal models and depressed patients, whereas decreases in BDNF have been linked to depression. These data contribute to a currently popular hypothesis that changes in BDNF are important to the symptoms of depression as well as the relief of these symptoms by antidepressant drug treatment.[11, 12]

Chemokines and cytokines

Chemokines and cytokines comprise large families of homologous small proteins (6–10 kDa) and differ from neuropeptides and neurotrophic factors in that they are key signalling molecules of the immune system. However, recent findings suggest that these molecules and some of their receptors are also present in the brain in both glial cells and neurones, raising the possibility that they might also have neurotransmitter-like functions. Although the evidence is incomplete, data show that chemokine and cytokine molecules have several of the characteristics that define neurotransmitters including modulation of release of other neurotransmitters or neuropeptides.[13] The pharmacological development of agonists and antagonists that are selective for chemokine and cytokine receptors and can cross the blood-brain barrier, would open an intriguing new era of research in neuroscience.

Retrograde messengers

It is now recognized that in contrast to classical neurotransmitters and neuropeptides, a small number of brain molecules signal information at a synapse in a 'retrograde' direction that is released from the postsynaptic neurone to act on the presynaptic neurone. Molecules falling into this category include certain neurotrophic factors (see above), gaseous molecules and lipid messengers.

Nitric oxide

One putative retrograde gaseous messenger is nitric oxide (NO) that is produced in neurones from the amino acid L-arginine by a neurone-specific isoform of NO synthase (NOS), which has a widespread abundance in the brain. The first evidence of a role for NO as a neurotransmitter came from findings that stimulation of glutamate NMDA receptors by glutamate caused the release of a diffusible messenger, which was subsequently identified as NO.[14] The current thinking is that increased glutamatergic activity triggers in postsynaptic neurones an NMDA-mediated activation of NOS, and then increased synthesis of NO that diffuses across the synapse to enhance presynaptic transmission. The latter occurs at least in part through NO acting on guanylate cyclase to increase production of the second messenger cGMP. In postsynaptic neurones, NO regulates certain protein kinase pathways and gene transcription factors, and also changes cell signalling events by S-nitrosylation.

On the basis of studies on the effects of NO donors, and pharmacological and genetic modulation of NOS, increased NO production is associated with a range of CNS functions including cognition, induction and maintenance of synaptic plasticity, and NO may be neuroprotective under some conditions.[15] However, because excess NO has neurotoxic potential, and because of the difficulty of delivering NO to the CNS without inducing side effects through the many actions of NO on peripheral tissues, the development of NO-based therapies for the treatment of CNS disorders will be very challenging.

Endocannabinoids

Examples of lipid retrograde messengers are the endocannabinoids, which are a recently discovered family of naturally occurring lipids (including anandamide) that interact with cell surface receptors that are targeted by Δ^9-tetrahydrocannabinol (THC). The latter is the principle biologically active constituent of marijuana, the dried leaves of the cannabis plant.[16] In essence, endocannabinoids appear to be to THC what opiate peptides are to morphine.

It is now apparent that endocannabinoids are synthesized enzymically on demand within the postsynaptic neurone and once produced, diffuse across the synapse in a retrograde direction. Endocannabinoids then suppress neurotransmitter release through activation of presynaptic cannabinoid receptors, the main type in the brain being classified CB_1 (analogous in terms of structure and function to opiate receptors but quite distinct pharmacologically).

Many of the central effects of THC, including analgesia, increased appetite and euphoria, are thought to be mediated by CB_1 receptors. Since CB_1 receptors have a powerful influence on synaptic transmission in the brain and have limited distribution in the periphery (although CB_2 receptors are abundant in the immune system), drugs targeting these receptors and/or the enzymes involved in endocannabinoid synthesis and metabolism may have interesting therapeutic possibilities. Such drugs are currently under investigation as analgesic and anorectic agents, amongst many others.

Neurotransmitter receptors

Neurotransmitter receptors are located on the cell surface of both pre- and postsynaptic neurones, and as a general rule can be divided into two main types; one type activates an ion channel that is intrinsic to the receptor (ligand-gated ion channel, sometimes called ionotropic receptor), the other type activates a GTP binding protein which acts as a transducer between the receptor and effector system (G-protein coupled receptor [GPCR], sometimes called metabotropic receptor). An exceptions to this general rule are certain trophic factors and cytokines which signal via direct activation of a unique class of protein kinases, protein tyrosine kinases. In addition, steroid hormones signal in the brain by crossing the plasma membrane and activating receptors in the neuronal cytoplasm that translocate to the nucleus where they bind DNA and function as transcription factors.

Ligand-gated ion channels typically comprise a multimeric plasma membrane receptor complex (4–5 subunits each with 4 transmembrane spanning domains) that gate the influx of ions to evoke fast changes in synaptic signalling. GPCRs comprise a superfamily of single proteins (7 transmembrane spanning domains) that evoke slower changes in synaptic signalling through the generation of second messengers and interactions with intracellular signalling pathways.

One of the most remarkable advances in molecular neuropharmacology in the last 20 years, which has been made possible through advances in molecular cloning technology and the virtual completion of the human genome project, has been the discovery of huge diversity in neurotransmitter receptors. This complexity takes the form of not only several hundred GPCRs[17] but also considerable heterogeneity in ligand-gated ion channels produced through the assembly of multiple receptor subunits.[18, 19]

Most, and probably all, neurotransmitters have more than one receptor type, which were once classified solely according to their pharmacological properties, but today are grouped more precisely in terms of their pharmacological, functional, and structural properties. Many neurotransmitter receptor subtypes have been cloned and their distribution within the brain is known; the challenge is to turn these advances at the molecular level into a better understanding of synaptic function, and novel therapies. Neurotransmitter receptors are too numerous to describe individually but examples are shown in Table 2.3.4.4 and recent detailed reviews are available elsewhere.[10]

Ligand-gated ion channels

The amino acids glutamate and GABA are, respectively, the principal fast excitatory and inhibitory transmitters in the brain and exert their effects via ligand-gated ion channels. However, other transmitters are also capable of fast transmission via ligand-gated ion channels and these include acetylcholine (nicotinic receptors), 5-HT (5-HT_3) and ATP (P_{2X}). However, to date there is no known neuropeptide with a ligand-gated ion channel.

Ligand-gated ion channels for glutamate

Glutamate elicits fast excitatory effects by activating ligand-gated ion channels on postsynaptic membranes of which there are three types; α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors, N-methyl-d-aspartate (NMDA) receptors and less abundant kainate receptors, all of which were originally named according to their preferred synthetic agonist and gate cations (Na^+, K^+, and Ca^{2+}) with varying degrees of selectivity. There is likely to be additional heterogeneity because each receptor can be assembled as a tetramer of one or more of multiple subunits. For instance, AMPA receptors can be formed from tetramers of combinations of

Table 2.3.4.4 Examples of neurotransmitter receptors

Transmitter	Receptor	Signal transduction
Dopamine	D_1 family (Dopamine D_1, D_5)	Adenylate cyclase (G_s)
	D_2 family (Dopamine D_2, D_3, D_4)	Adenylate cyclase ($G_{i/o}$)
Noradrenaline	α_1 family ($\alpha_{1A,B,D}$)	Phospholipase C (G_q)
	α_2 family ($\alpha_{2A,B,C}$)	Adenylate cyclase ($G_{i/o}$)
	β family ($\beta_{1,2,3}$)	Adenylate cyclase (G_s)
5-hydroxytryptamine	5-HT_1 family (5-$HT_{1A, B, D,E, F}$)	Adenylate cyclase ($G_{i/o}$)
	5-HT_2 family (5-$HT_{2A, B, C}$)	Phospholipase C (G_q)
	5-HT_3	Cation channel
	5-HT_4	Adenylate cyclase (G_s)
	5-HT_5 family (5-$HT_{5A, B}$)	Not certain
	5-HT_6	Adenylate cyclase (G_s)
	5-HT_7	Adenylate cyclase (G_s)
Acetylcholine	M_1 (muscarinic)	Phospholipase C (G_q)
	M_2	Adenylate cyclase ($G_{i/o}$)
	M_3	Phospholipase C (G_q)
	M_4	Adenylate cyclase ($G_{i/o}$)
	M_5	Phospholipase C (G_q)
	Nicotinic (α1–10, β1–4, δ, ε, γ)	Cation channel
GABA	$GABA_A$ (α1–6, β1–3, γ1–3, σ1–3, δ, ε, π, o)	Chloride channel
	$GABA_B$	Adenylate cyclase ($G_{i/o}$)
Glutamate	AMPA ($GluR_{1–4}$)	Cation channel
	NMDA (NR_1, $NR_{2A–D}$, $NR_{3A–B}$)	Cation channel
	Kainate ($GluR_{5–7}$, KA_1, KA_2)	Cation channel
	Group I family ($mGluR_{1/5}$)	Phospholipase C (G_q)
	Group II family ($mGluR_{2–3}$)	Adenylate cyclase ($G_{i/o}$)
	Group III family ($mGluR_{4'6,7,8}$)	Adenylate cyclase ($G_{i/o}$)
Tachykinin (including substance P)	NK_1	Phospholipase C (G_q)
	NK_2	Phospholipase C (G_q)
	NK_3	Phospholipase C (G_q)
Opioid	δ	Adenylate cyclase ($G_{i/o}$)
	κ	Adenylate cyclase ($G_{i/o}$)
	μ	Adenylate cyclase ($G_{i/o}$)
Galanin	GAL1	Adenylate cyclase ($G_{i/o}$)
	GAL2	Adenylate cyclase ($G_{i/o}$)
	GAL3	Adenylate cyclase ($G_{i/o}$)
Adenosine	A_1	Adenylate cyclase ($G_{i/o}$)
	A_2 family ($A_{2A, B}$)	Adenylate cyclase (G_s)
	A_3	Adenylate cyclase ($G_{i/o}$)
ATP	P2X family ($P2X_{1-7}$)	Cation channel
	$P2Y_1$	Phospholipase C (G_q)
	$P2Y_2$	Phospholipase C (G_q)
	$P2Y_4$	Phospholipase C (G_q)
	$P2Y_6$	Phospholipase C (G_q)
	$P2Y_{11}$	Phospholipase C (G_q)
	$P2Y_{12}$	Adenylate cyclase ($G_{i/o}$)
	$P2Y_{13}$	Adenylate cyclase ($G_{i/o}$)
	$P2Y_{14}$	Phospholipase C (G_q)
Cannabinoid	CB1	Adenylate cyclase ($G_{i/o}$)
	CB2	Adenylate cyclase ($G_{i/o}$)

four subunits (GluR1–GluR4) and NMDA receptors from two subunits (NR1, NR2). There are a large number of naturally occurring variants of both AMPA and NMDA subunits generated through RNA editing and alternative splicing.

The pharmacological and functional significance of this complexity is not yet clear although evidence suggests that certain agents are able to distinguish between different receptor assemblies.[10] Interestingly, recent data suggest that changes in AMPA receptor subunit composition at the postsynaptic membrane cause differences in ion (Ca^{2+}) permeability that change synaptic efficacy, the best-characterized form of which is long-term potentiation (LTP), a widely accepted neurophysiological correlate of learning and memory.[19]

An interesting feature of many ligand-gated ion channels is the presence of multiple chemically-sensitive sites (allosteric modulatory sites), in addition to the site(s) which bind the natural transmitter ligand. This is the case for both the AMPA and NMDA receptors. Thus, in addition to the glutamate binding site AMPA receptors are sensitive to 'AMPAkines' which comprise a chemically diverse group of agents that potentiate the function of the receptor in *in vitro* models, and elicit associated procognitive effects *in vivo*.[20] Non-glutamate sites on the NMDA receptor include a site for Mg^{2+} that is the source of a voltage-dependent NMDA receptor block which requires membrane depolarization to open, as well as positive modulatory sites for glycine and polyamines. Whilst glutamate is released from presynaptic terminals in an activity-dependent fashion and acts as a neurotransmitter, glycine and the polyamines act as extracellular modulators that are present at more constant levels. The latter binding sites are under investigation as a possible source of NMDA receptor enhancing agents that do not suffer the potential excitotoxic effects of agonists at the glutamate site.[21]

Ligand-gated ion channels for GABA

GABA elicits fast inhibitory effects by activating the $GABA_A$ receptor ion channel complex. $GABA_A$ receptors are heteropentameric membrane proteins that form a GABA-gated chloride channel. There are at least 18 types of $GABA_A$ receptor subunits (α1–6, β1–3, γ1–3, Π, δ, θ and p1–3). Although studies co-expressing different $GABA_A$ receptor subunits indicate the potential for several hundreds if not thousands of $GABA_A$ receptor subunit combinations, studies on $GABA_A$ receptor subunit distribution and abundance in native brain tissue indicate that the number of naturally occurring types of $GABA_A$ receptors may be of the order of 10 or fewer.[22]

As with the glutamate ionotropic receptors, $GABA_A$ receptors have a number of allosteric modulatory sites including those sensitive to certain endogenous steroids, anaesthetic agents (for example alfaxalone) and alcohol (ethanol). However, the vast majority of $GABA_A$ receptors are characterized by their sensitivity to benzodiazepines. Both genetic and medicinal chemistry approaches have been used to identify the pharmacological significance of $GABA_A$ receptor subtypes, and specifically to determine whether the multiple behavioural effects of benzodiazepines such as diazepam (sedation, anxiolysis, amnesia, motor in-coordination etc.) can be attributed to specific $GABA_A$ receptor subunit combinations.[18]

In particular, studies with point-mutated mice have revealed that the sedative effect of diazepam is mediated by α1-containing $GABA_A$ receptors, whereas the anxiolytic action is mediated by α2/α3-containing $GABA_A$ receptors. Moreover, findings that ligands with selective actions at α2- and/or α3-containing $GABA_A$ receptors display anxiolytic activity, raise the possibility of future benzodiazepines with behaviourally selective actions. Interestingly, α5-containing $GABA_A$ receptors may be an important site of action of alcohol. The $GABA_A$ receptor subunit(s) targeted by steroids and anaesthetics to produce the CNS inhibitory effects of these agents are currently under investigation.

G-protein coupled receptors

Almost all neurotransmitters signal effects via GPCRs and most neurotransmitters signal via more than one type of GPCR. For example, the amine 5-HT possesses 14 receptor subtypes (comprising seven receptor families, 5-HT_{1-7}), 13 of which are GPCRs and one is a ligand-gated ion channel. Each 5-HT GPCR has high affinity and selectivity for 5-HT but individually the receptors demonstrate different selectivity for other ligands, arise from different (but homologous) genes, and are formed from different protein sequences with different distributions and signalling effects.[23] Since several 5-HT GPCRs can co-localize at a single synapse, the signal received by a postsynaptic neurone may be quite complicated. This complexity for 5-HT can be seen in many other transmitters including dopamine (D_{1-5}), glutamate ($mGluR_{1-8}$), noradrenaline ($\alpha_{1A,B,D}$, $\alpha_{2A,B,C}$, β_{1-3}), endocannabinoids (CB_{1-2}) and neuropeptides (Table 2.3.4.4).

Typically GPCRs comprise a single membrane protein with seven transmembrane spanning domains, an N-terminus facing the extracellular space, a C-terminus facing the cytoplasm, and several intracellular transmembrane domain linking loops. The N-terminus of some GPCRs ($mGluR_{1-8}$, $GABA_B$) contains the ligand binding site, while for most GPCRs the predicted ligand binding site lies within the transmembrane domains. Both the C-terminus and the third transmembrane intracellular loops are targets for phosphorylation by protein kinases; the third intracellular loop is the main site of G-protein interaction.

G-proteins

Each G-protein is a heterotrimer comprised of α, β and γ subunits that dissociate on binding of ligand to the GPCR. On dissociation, the α subunit binds GTP and through intrinsic GTPase activity directly regulates a number of specific downstream effector enzymes and ion channels. The β/γ subunits are also biologically active and regulate some of the same effector proteins.

There are four major types of G-proteins, Gs, Gi, Gq and G_0 that produce the following signalling effects; activation of adenylate cyclase, inhibition of adenylate cyclase, activation of phospholipase C and interaction with Ca^{2+} and K^+ channels, respectively. Changes in the activity of adenylate cyclase results in altered intracellular levels of the 'second messenger' cyclic adenosine monophosphate (cAMP). Similarly, phospholipase C (PLC) alters intracellular levels of inositol triphosphate (IP_3) and diacyl glycerol (DAG). Altered levels of these second messengers trigger changes in activity of specific signalling cascades and ultimately changes in physiological responses (see below).

The opening of ion channels in response to neurotransmitter-induced GPCR activation leads to direct effects (excitatory or inhibitory) on the electrical properties of neurones, albeit on a slightly slower timescale than effects produced by ligand-gated ion channels. Almost all neurotransmitter classes are able to evoke

changes in ion channel opening via GPCRs, and some may be clinically important. For example, the α_2-adrenoceptor-induced opening of K^+ channels on noradrenaline neurones that causes a fall in noradrenergic activity and release, may contribute to the anxiolytic and sedative properties of α_2-adrenoceptor agonists such as clonidine. On the other hand, the 5-HT_{2A} receptor-induced closing of K^+ channels on cortical neurones that causes an increase in cortical neurone activity, may underlie the psychotic effects of LSD and related hallucinogens.[23]

GPCR regulation

Recent discoveries of interactions between GPCRs and other intracellular proteins have lead to a new understanding of how the receptors are regulated and trafficked to and from the plasma membrane. Studies commencing on the β-adrenceptor, have identified two families of regulatory proteins called β-arrestins and GPCR kinases (GRKs). Within seconds of being activated by an agonist the GPCR is phosphorylated by a GRK on the C-terminal cytoplasmic tail and other intracellular domains. This phosphorylation promotes the interaction of β-arrestins with the GPCR, which limits the signal duration, and causes loss of sensitivity to agonist activation, and then receptor internalization from the cell surface.[24]

In addition to β-arrestins the C-termini of GPCRs associate with a large variety of transmembrane or soluble proteins, termed 'GPCR-interacting proteins' (GIPs). Some GIPs are themselves GPCRs that form homo- or heterodimers, while other GIPs are ionic channels, ionotropic receptors and proteins that control GPCR trafficking[25]. One interesting example of a GIP is the molecule p11, which reportedly functions to traffic a 5-HT GPCR (5-HT_{1B}) to the plasma membrane. Evidence suggests that p11 expression is reduced in postmortem brain of patients committing suicide, and that mice with genetic deletion of p11 have a depressive-like phenotype.[26]

Small G proteins

In addition to the G-proteins associated with GPCRs, there is a super family of 'small G-proteins' which also bind GTP and possess intrinsic GTPase activity but these are not modulated by agonist binding. Rather, small G-proteins function as molecular switches that control several cellular processes ranging from vesicle trafficking and exocytosis (e.g. Rab) to assembly of cytoskeletal structures (e.g. Rho). Among the best characterized small G-proteins are those that comprise the Ras family. Numerous types of cell signals including those of most neurotrophic factors converge on Ras and related proteins to regulate MAP-kinase pathways.

Second messengers

The generation of the second messengers cAMP and DAG by adenylate cyclase and phospholipase C, respectively, leads to activation of protein kinases that add phosphate groups to specific protein targets to change their activity and ultimately trigger diverse physiological responses (Fig. 2.3.4.2). Enzymes called phosphatases,

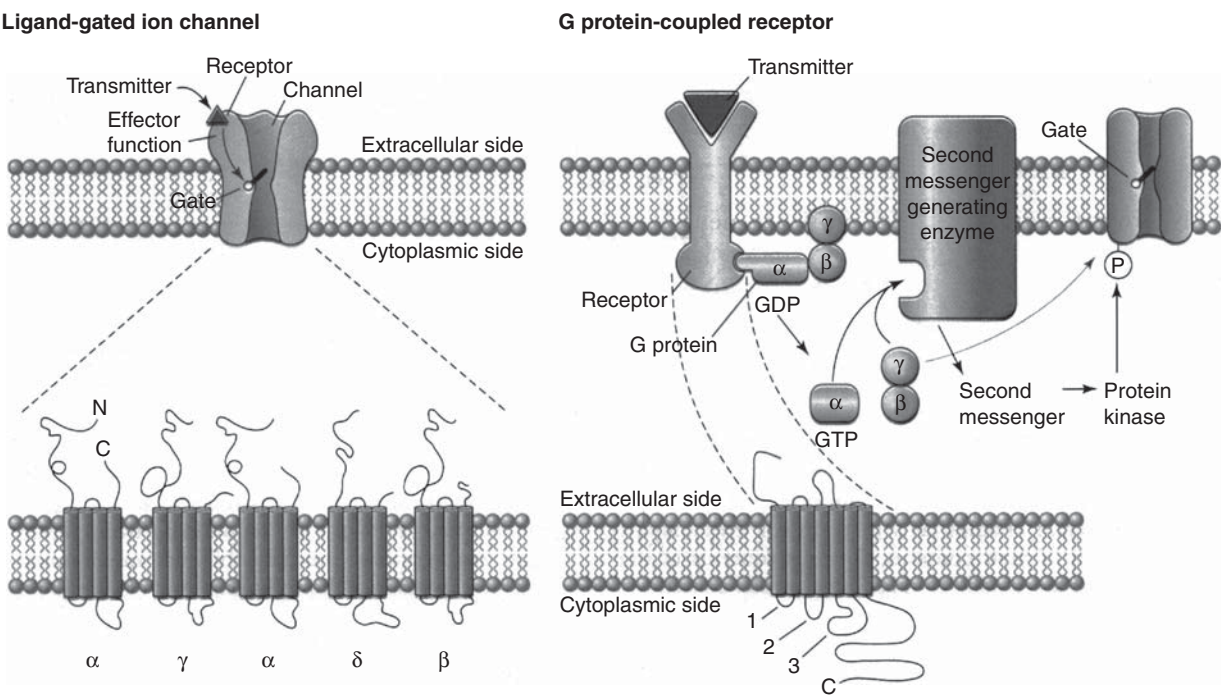


Fig. 2.3.4.2 Diagrammatic representation of ligand-gated ion channel and G-protein coupled receptors. Ligand-gated ion channels comprise multiple protein subunits that form a central pore in the plasma membrane. On binding of the neurotransmitter, this receptor mediates fast excitatory or inhibitory transmission depending on whether the pore gates cations or chloride ions, respectively. G-protein coupled receptors comprise a single membrane spanning protein. On binding of the neurotransmitter, this receptor mediates slow transmission by enabling the dissociation of the G-protein into α subunit monomer and β/γ subunit dimer, both of which may activate an effector enzyme to generate a second messenger. Also, the β/γ subunit dimer may directly interact with ion channels. Second messengers may also indirectly modulate ion channels through phosphorylation by activating protein kinases. Reproduced from Nestler EJ, Hyman SE and Malenka RC (2001) *Molecular Neuropharmacology*, p. 64, Copyright 2001, McGraw-Hill, New York.

which remove the phosphate groups, oppose these signalling effects. Guanylate cyclase is a cytosolic enzyme which also generates a second messenger, cGMP. As noted above, guanylate cyclase is activated by NO to produce effects on presynaptic function.

Based on molecular cloning studies, nine forms of adenylate cyclase have been identified (I–IX) and each exhibits a distinct distribution in brain and peripheral tissues.[27] The full implication of this complexity is not yet understood but it suggests that regulation of cAMP formation varies depending on the form of adenylyl cyclase expressed in neuronal cells.

Both cAMP and cGMP are degraded by phosphodiesterases (PDEs) which are expressed in numerous forms (types 1–11) in brain and periperal tissues.[10] At high concentrations, caffeine and related methylxanthines inhibit PDE and this action contributes to the pharmacological effects of these drugs. Much effort is being made to develop inhibitors that are selective for brain-specific forms of PDE. Rolipram inhibits all isoforms of PDE4; this drug showed promise as an antidepressant, but its clinical utility was limited by peripheral side effects. However, because the PDE4 enzymes comprise a number of isoforms, an inhibitor of one isoform may lead to the development of an effective antidepressant without the side effects of rolipram.

GPCR-induced activation of PLC causes the breakdown of phosphatidylinositol, resulting in the generation and recycling of the second messengers, IP_3 and DAG (phosphoinositide cycle). Both IP_3 and DAG produce downstream signalling effects, IP_3 through the mobilization of intracellular calcium stores and DAG through activating a protein kinase. There are two major isoforms of PLC in brain, β and γ, the β isoform being predominantly responsible for mediating the effects of GPCRs linked to Gq.

After its formation, IP_3 is recycled via a series of dephosphorylations to form inositol which is used in the regeneration of phosphatidylinositol. Interestingly, lithium, which is an important drug in the treatment of manic depressive illness, inhibits one of the enyzymes involved in the recycling of IP_3 (inositol-1-monophosphatase) and causes inositol depletion. Because inositol does not easily enter the blood-brain-barrier, brain inositol levels are thought to fall and production of the second messengers diminishes. It is a popular hypothesis that inositol depletion is responsible for lithium's clinical effects but this remains unproven. Indeed, lithium is known to interact with a range of signalling systems including various ion channels, adenlyate cyclases and protein kinases. Recent interest has focussed on the inhibition by lithium and mood stabilizing anticonvulsants such as valproate, of glycogen synthase kinase 3β that has a range of targets and effects ranging from formation of inositol to trophic mechanisms.[28]

Downstream signalling cascades

The activation or inhibition of second messenger signalling cascades by GPCRs can profoundly change the intracellular environment of a neurone by regulating the activity of protein kinases and other proteins, including gene transcription factors and even enzymes involved in regulation of chromatin structure. Consequently, these cascades may regulate gene transcription and protein synthesis and activate multiple downstream effectors, including those that form the cytoskeleton or contribute to mechanisms underlying synaptic plasticity. Such effects can have long-lasting effects on neuronal function. Increasing evidence suggests that the neuroadaptive responses to repeated psychotropic drug administration are underpinned by changes in gene expression that result in the remodelling of synaptic function and structure. This thinking has been applied to explain a range of processes including compulsive recreational drug use as well as the therapeutic action of antipsychotic drugs.

As an example, the past decade has seen the evolution of a fascinating theory to explain the delayed onset of antidepressant effect of drugs like fluoxetine and imipramine that act to inhibit plasma membrane amine transporters (see above). This theory posits that elevated amine levels trigger GPCR signalling cascades that activate gene programmes to enhance neuronal survival and connectivity, the latter having failed because of the adverse effects of stress and other environmental factors.[11, 29, 30] Some of the key genes involved in this process include trophic factors such as BDNF, which may be a trigger for the production of newly formed neurones (Fig. 2.3.4.3). Although this line of thought is driving promising pharmacological strategies for improved antidepressant therapies, our knowledge of the key molecules that are changed by antidepressants to bring about the relief of the symptoms of depression is far from complete.

Concluding remarks

Until recently, studies on the chemistry of synaptic neurotransmission have focused on a small number of neurotransmitters and a narrow group of proteins involved in neurotransmitter function,

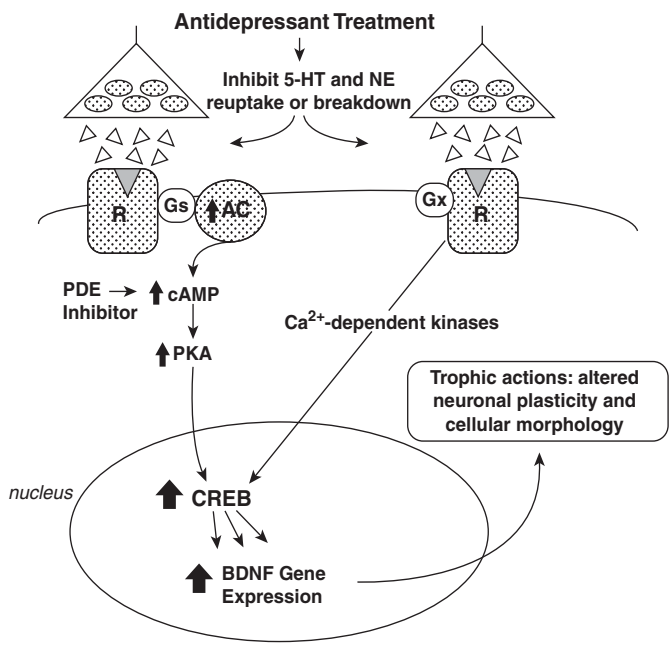


Fig. 2.3.4.3 Hypothetical signal transduction pathways regulated by antidepressant drugs leading to increased neurotrophic factor expression and neurogenesis. Antidepressant treatment increases synaptic amine levels that stimulate GPCRs (Gs) linked to adenylyl cyclase (AC). The subsequent increase in cAMP levels activates cAMP-dependent protein kinase (PKA) which, possibly together with Ca^{2+}-dependent protein kinases, increases the function and expression of the gene transcription factor, cAMP response element binding protein (CREB). CREB enhances the expression of brain derived neurotrophic factor (BDNF) that may underpin trophic effects of antidepressant treatment, including synaptic remodeling and increased neurogenesis. Reprinted by permission from Macmillan Publishers Ltd: Molecular Psychiatry, DOI: 10.1038/sj/mp/4001016.

specifically neurotransmitter receptors, transporters and enzymes which bring about neurotransmitter synthesis or degradation. Today, powerful molecular and genetic approaches are being used to identify and understand new proteins and mechanisms involved in neurotransmitter function and control. So far, just a few tens of perhaps thousands of neurotransmitter-related proteins, have been successfully targeted by pharmacological agents that have translated into important treatments of psychiatric disorder but there is promise of many more such treatments to come. Moreover, this huge diversity of neurotransmitter-related proteins is now emerging as a large resource for studies of genetic risk factors of psychiatric disorder and investigations of biological markers of illness diagnosis and progression, and treatment outcome.

Further information

International Union of Basic and Clinical Pharmacology (IUPHAR) committee official database on receptor nomenclature and drug classification, http://www.iuphar-db.org/index.jsp.

Kandel, E.R., Schwartz, J., Jessell, T. (2000) *Principles of Neural Science*, (4th edn.), McGraw-Hill, New York. Further details are available at http://en.wikipedia.org/wiki/Principles_of_Neural_Science.

Nestler, E.J., Hyman, S.E., and Malenka, R.C. (2008). *Molecular Neuropharmacology*, 2nd edn. McGraw-Hill, New York.

References

1. Hökfelt, T., Bartfai, T. and Bloom, F. (2003). Neuropeptides: opportunities fordrug discovery. *Lancet Neurology,* **2**, 463–72.
2. Iversen, L. (2006). Neurotransmitter transporters and their impacton the development of psychopharmacology. *British Journal of Pharmacology,* **147**, Suppl 1, S82-8.
3. Fremeau ,R.T., Voglmaier, S., Seal, R.P., *et al.* (2004). VGLUTs define subsets of excitatory neurones and suggest novel roles for glutamate. *Trends in Neurosciences,* **27**, 98–103.
4. Zhou, Z., Zhen, J., Karpowich, N.K., *et al.* (2007). LeuT-desipramine structure reveals how antidepressants block neurotransmitter reuptake. *Science,* **317**, 1390–3.
5. Shigeri, Y., Seal, R.P. and Shimamoto, K. (2004). Molecular pharmacology of glutamate transporters, EAATs and VGLUTs. *Brain Research,* **45**, 250–65.
6. Rothstein, J.D., Patel, S., Regan, M.R., *et al.* (2005). Beta-lactam antibiotics offer neuroprotection by increasing glutamate transporter expression. *Nature,* **433**, 73–7.
7. Scamuffa, N., Calvo, F., Chretien, M., *et al.* (2006). Proprotein convertases: lessons from knockouts. *Faseb Journal,* **20**, 1954–63.
8. Ögren, S.O., Kuteeva, E., Hökfelt, T, *et al.* (2006). Galanin receptor antagonists: a potential novel pharmacological treatment for mood disorders. *CNS Drugs,* **20**, 633–54.
9. Chao, M.V. (2003). Neurotrophins and their receptors: a convergence point for many signalling pathways. *Nature Reviews* **4**, 299–309.
10. Alexander, S.P., Mathie, A. and Peters, J.A. (2007). Guide to Receptors and Channels (GRAC), 2nd edition (2007 Revision). *British Journal of Pharmacology,* **150**, Suppl 1, S1–S168.
11. Castren, E. (2005). Is mood chemistry? *Nature Reviews,* **6**, 241–6.
12. Martinowich, K., Manji, H. and Lu, B. (2007). New insights into BDNF function in depression and anxiety. *Nature Neuroscience,* **10**, 1089–93.
13. Rostene, W., Kitabgi, P. and Parsadaniantz, S.M. (2007). Chemokines: a new class of neuromodulator? *Nature Reviews,* **8**, 895–903.
14. Garthwaite, J. and Boulton, C.L. (1995). Nitric oxide signaling in the central nervous system. *Annual Review of Physiology,* **57**, 683–706.
15. Calabrese, V., Mancuso, C., Calvani, M., *et al.* (2007). Nitric oxide in the central nervous system: neuroprotection versus neurotoxicity. *Nature Reviews,* **8**, 766–75.
16. Piomelli, D. (2003). The molecular logic of endocannabinoid signalling. *Nature Reviews,* **4**, 873–84.
17. Fredholm, B.B., Hokfelt, T. and Milligan, G. (2007). G-protein-coupled receptors: an update. *Acta physiologica,* **190**, 3–7.
18. Rudolph, U. and Mohler, H. (2006). GABA-based therapeutic approaches: GABAA receptor subtype functions. *Current Opinion in Pharmacology,* **6**, 18–23.
19. Schuman, E.M. and Seeburg, P.H. (2006). Signalling mechanisms. *Current Opinion in Neurobiology,* **16**, 247–50.
20. Lynch, G. and Gall, C.M. (2006). Ampakines and the threefold path to cognitive enhancement. *Trends in Neurosciences,* **29**, 554–62.
21. Kemp, J.A. and McKernan, R.M. (2002). NMDA receptor pathways as drug targets. *Nature Neuroscience,* **5** Suppl,1039–42.
22. McKernan, R.M. and Whiting, P.J. (1996). Which GABAA-receptor subtypes really occur in the brain? *Trends in Neurosciences,* **19**,139–43.
23. Barnes, N.M. and Sharp, T. (1999). A review of central 5-HT receptors and their function. *Neuropharmacology,* **38**,1083–152.
24. Lefkowitz, R.J. (2007). Seven transmembrane receptors: something old, something new. *Acta physiologica,* **190**, 9–19.
25. Bockaert, J., Roussignol, G., Becamel, C., *et al.* (2004). GPCR-interacting proteins (GIPs): nature and functions. *Biochemical Society Transactions,* **32**, 851–5.
26. Svenningsson, P., Chergui, K., Rachleff, I., *et al.* (2006). Alterations in 5-HT1B receptor function by p11 in depression-like states. *Science,* **311**, 77–80.
27. Cooper, D.M. (2003). Regulation and organization of adenylyl cyclases and cAMP. *Biochemical Journal,* **375**, 517–29.
28. Gould, T.D. and Manji, H.K. (2005). Glycogen synthase kinase-3: a putative molecular target for lithium mimetic drugs. *Neuropsychopharmacology,* **30**, 1223–37.
29. Pittenger, C. and Duman, R.S. (2008). Stress, depression, and neuroplasticity: a convergence of mechanisms. *Neuropsychopharmacology,* **33**, 88–109.
30. Duman, R.S. (2002). Synaptic plasticity and mood disorders. *Molecular Psychiatry*, **7**, Suppl 1, S29–34.

2.3.5 Neuropathology

Peter Falkai and Bernhard Bogerts

Introduction

The traditional domains of neuropathology are well-defined organic brain diseases with an obvious pathology, such as tumours, infections, vascular diseases, trauma, or toxic and hypoxemic changes, as well as degenerative brain diseases (e.g. Alzheimer's disease, Parkinson's disease, and Huntington's chorea). Neuropathological investigations of these brain disorders have been rewarding, because patients with any of these conditions can be expected to have gross morphological or more or less specific neurohistological anomalies related to the clinical symptoms of the disorders. Moreover, the type of brain pathology of these well-defined disease entities is quite homogenous. For example, it is highly unlikely that a patient with Parkinson's disease would not exhibit morphological changes and Lewy bodies

in the nigrostriatal system, just as much a person with Huntington's chorea would have a normal striatum, or a patient with Pick's or Alzheimer's disease would have no changes in the cerebral cortex.

In contrast, the history of the neuropathology of psychiatric disorders outside primary degenerative diseases is much more controversial, because no such obvious and homogenous types of brain pathology (as seen in neurological disorders) have yet been detected for the major psychiatric illnesses such as schizophrenia, affective disorders, substance-related disorders, or personality disorders.

The scope of this chapter is to summarize the neuropathological findings in schizophrenia, affective disorders, and alcoholism. Tables 2.3.5.1, 2.3.5.2, 2.3.5.3, and 2.3.5.4 highlight the significant findings. It goes beyond the scope of this chapter to review the large body of literature on the dementias, including specifically Alzheimer's disease. Concerning this matter, the reader is referred to several comprehensive reviews (e.g. Jellinger and Bancher 1998).[1]

Table 2.3.5.1 Gross morphometric findings in schizophrenia

Region/parameter	Finding
General	
Brain length	(↓)
Brain weight	↓
Ventricular area/volume	↑
Cortex thickness	↓
Temporal lobe	
Lobar area/volume	—
Hippocampal area/volume	↓
Parahippocampal area/volume	↓
Parahippocampal cortical thickness	↓
Amygdala area/volume	—
Sylvian fissure length, planum temporal volume	↓
Sulcogyral pattern	Abnormal
Frontal, parietal, and occipital lobes	
Cingulate cortical thickness	—
Insula area/volume	—
Corpus callosum thickness	(↑)
Internal capsula area/volume	—
Basal ganglia	
Globus pallidum area/volume	(↓)
Nucleus accumbens area/volume	↓
Cautdate-putamen area/volume	↑
Thalamus	
Mediodorsal nucleus area/volume	(↓)
Whole and various nuclei area/volume	—
Cerebellum	
Anterior vermis area	↓
Brainstem	
Substantia nigra volume	↓
Locus coeruleus volume	—
Periventricular grey volume	↓

In comparison with controls: ↓, reduced; ↑, increased; —, no difference; ( ), finding not or only partially replicated.

Adapted from Arnold and Trojanowski.[3]

Schizophrenia and other psychotic disorders

Studies between 1898 and 1975

In 1898, Alois Alzheimer (1898)[2] described subtle changes in the neocortex of patients with schizophrenia. Subsequently to Alzheimer, Southard reported cortical atrophy in schizophrenia and mentioned association areas of the cerebral cortex to be most affected in this disorder.[4] Vogt and Vogt and their coworkers reported cellular alterations in the cortex, thalamus, and basal ganglia of schizophrenics.[5] These considerable efforts on the part of many well-known neuroanatomists and neuropathologists to prove schizophrenia to be a primary brain disorder ended in inconsistent and unsubstantiated findings.[6] To a large extent, these inconsistencies can be attributed to a variety of methodological inadequacies including diagnostic uncertainties, inadequate control samples, flawed tissue-handling procedures, variable choice of brain regions for neuropathological studies, limitations in the sensitivity and specificity of classical histological stains, as well as lack of quantitative methods to delineate and analyse subtle brain abnormalities.[7]

Table 2.3.5.2 Neuronal morphometric findings in schizophrenia

Region/parameter	Finding
Temporal lobe	
Superior temporal gyrus (Tpt) neurone density	↓
Hippocampal neurone density	(↓)
Hippocampal neurone size	(↓)
Entorhinal cortex neurone density	(↓)
Entorhinal cortex neurone size	↓
Amygdala neurone density (basolateral n.)	—
Frontal lobe	
Prefrontal cortex pyramidal neurone density	↑
Prefrontal cortex interneurone density	(↓)
Prefrontal cortex neurone size	↓
Cingulate (anterior) pyramidal neurone density	↓
Cingulate interneurone density	↓
Cingulate neurone size	↓
Motor cortex neurone density	(↓)
Motor cortex neurone size	—
Basal ganglia	
Globus pallidus neurone counts	—
Nucleus accumbens neurone counts	↓
Nucleus basalis of Meynert neurone counts	—
Thalamus	
Mediodorsal nucleus neurone counts	(↓)
Cerebellum	
Purkinje cell density	↓
Brainstem	
Substantia nigra neurone density	↓
Substantia nigra neurone size	—
Locus coeruleus neurone density	
Locus coeruleus neurone size	
Pedunculopontine nucleus neurone density	↓

In comparison with controls: ↓, reduced; ↑, increased; —, no difference; ( ), finding not or only partially replicated.

Adapted from Arnold and Trojanowski.[3]

Neuropathological findings in schizophrenia since 1975

Advances in the last 30 years have produced more reliable psychiatric diagnostic criteria, improved structural and functional neuroimaging techniques, a large array of highly sensitive and specific molecular probes and labeling procedures, suitable for use in neuropathological studies, and computer-assisted image analysis methodologies. For these and other reasons, there has been a resurgence of interest in the neurobiological substrates of schizophrenia, and contemporary neuropathological studies have enumerated many findings in the brains of patients with schizophrenia (for reviews see[7, 8, 9]). Finally, the recent description of the first risk genes like Neuregulin-1 or Dysbindin has provided this field with reliable research targets.[8, 9] To identify the role of these genes for the pathophysiology of schizophrenia their expression pattern in human brain tissue has to be established in the near future.

(a) Diagnostic neuropathology

Stevens (1982)[10] surveyed the brains of 28 schizophrenic patients for gross and microscopic abnormalities using standard diagnostic stains. She discovered no abnormalities in temporal (including the amygdalohippocampal region), frontal, or parietal lobes or in the thalamus, but detected assorted abnormalities in other regions, including neuronal loss or infarction in the globus pallidus in five patients, increased cerebellar white matter gliosis in five patients, excessive Purkinje cell loss in 13 cases, and, most notably, increased fibrillary gliosis in periventricular, periaqueductal, and basal forebrain regions bilaterally.

In another prospectively accrued series,[10] she found that of 56 schizophrenics five were afflicted with other distinct neurological illnesses (multiple sclerosis, Friedreich's ataxia, epilepsy, stroke) and three had been treated with prefrontal leukotomies. The remaining 48 showed no differences to controls in the frequency of large- or small-vessel cerebrovascular disease, senile plaques, or neurofibrillary degeneration. However, there was an 'increased incidence of unexpected pathology in the schizophrenic group compared with the control group'. Of these 48 schizophrenics, 21 exhibited some degree of focal pathology compared to 12 of 56 controls, but these abnormalities were diverse in nature and location. Holzer staining suggested a significant increase in fibrous gliosis in the cortex, white matter, and periventricular structures, but generally for those brains showing other focal pathology. After removal of these cases, the 'adjusted' group showed no evidence of increased gliosis.

In a series of 101 elderly schizophrenics,[12] Golier *et al.* found only 10 with definite or probable Alzheimer's disease by modern neuropathological diagnostic criteria, 29 with senile plaques, 15 with vascular lesions, two with Parkinson's disease, three with unspecified tumour, and five with 'other' findings.

Another review concluded extensive neuropathological investigations due to lack of any evidence of neurodegeneration or neural injury beyond what typically is observed in brains of individuals without neuropsychiatric illness.[8]

(b) Morphometric studies

Macroscopic findings (Table 2.3.5.1)

Several planimetric postmortem studies of the entire cortex have been performed, some reporting significant reduction of cortical volume (12 per cent) and central grey matter (6 per cent), and others reporting no difference in volumes of cortex, white matter and whole hemispheres between schizophrenics and controls. Others that measured general brain parameters have shown reduced brain length, brain weight, and increased ventricular area/volume.

Since the publication of the first report of reduced tissue volume in temporolimbic structures of schizophrenics,[13, 14] numerous quantitative or qualitative anatomical postmortem studies on limbic structures of schizophrenics have been conducted. The majority of these studies substantiated subtle structural changes (15–20 per cent mean volume reduction) in at least one of the investigated areas, whereas only a few yielded entirely negative results. The findings comprise reduced volumes or cross-sectional areas of the hippocampus, amygdala, parahippocampal gyrus, which were later corroborated by morphometric magnetic resonance imaging (MRI) studies. Figure 2.3.5.1 demonstrates the subtle bilateral volume reduction of the hippocampus in schizophrenics and furthermore visualizes the kind of hypoplastic appearance of the anterior hippocampus, which can be seen in about one third of the patients

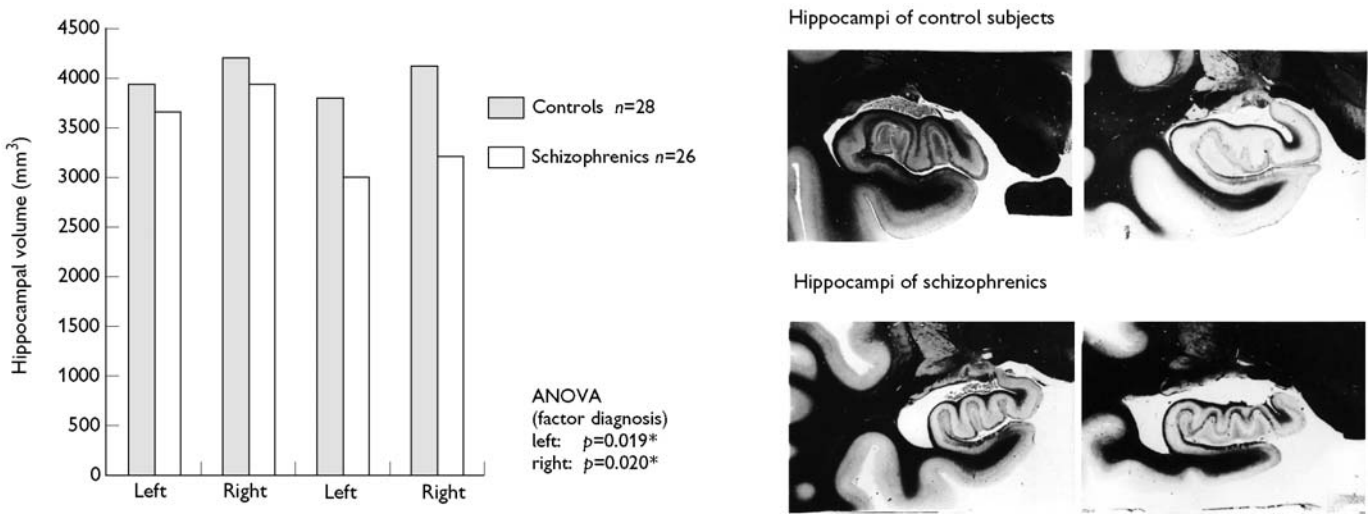


Fig. 2.3.5.1 Left side: hippocampal volumes in schizophrenic patients and controls; Right side: lippocampal atrophy macroscopically seen in about one-third of patients with schizophrenia (upper row) compared to control subjects (lower row) (from Bogerts 1990).

(lower row of the photographs). Other findings in limbic brain regions are left temporal horn enlargement, white-matter reductions in parahippocampal gyrus or hippocampus, and an increased incidence of a cavum septi pellucidi.

Unchanged volumes of the striatum and external pallidum but a subtle volume decrease in the internal pallidal segment were found in brains from the preneuroleptic era. Pallidal volume reduction was due to a reduction in the catatonic subgroup.[15] These initial findings have to be pursued, as longitudinal MRI studies suggest that enlargement of basal ganglia can be seen in schizophrenia as a consequence of treatment with classical neuroleptics, which can be reversed by the use of atypical substances.[16]

After initially finding no volumetric changes in the thalamic nuclei, subsequently the area/volume of the mediodorsal nucleus and anteroventral thalamic nucleus were found to be decreased.[17]

Changes in area measurements of the corpus callosum were described in some studies. The findings, however, are inconsistent; there are reports of increased as well as of decreased midline areas. More consistent are reports of shape abnormalities, in that the sex difference in anterior and posterior callosal thickness in normal controls seems to be reversed in schizophrenics and the mean curvature in the corpus callosum is bent upwards.[18]

Findings of decreased volume of the substantia nigra and the periventricular grey matter as well as no volumetric change in the locus coeruleus await replication.

Microscopic findings (Table 2.3.5.2)

There are a number of studies of neurone number, density, and size in schizophrenia. As summarized in Table 2.3.5.2, the majority of these have focused on the ventromedial temporal and frontal lobes.

In the lateral prefrontal cortex, an increase in neurone density has been reported inconsistently, which may relate to the observed decrease in neurone size (with decreased dendritic arborization and a decreased neuropil compartment).[20] In the anterior cingulate could be observed decreased pyramidal and local circuit neurone density accompanied by increased vertical axon density and altered dopaminergic innervation. These findings have been interpreted as representing disturbed connections in the anterior cingulate.

Within the ventromedial temporal lobe, reduced cell numbers or cell size and abnormal cell arrangements in the hippocampus or entorhinal cortex were described. However, some groups could not confirm cellular disarray in the hippocampus[21] just as little find significant volume and cell number reductions in the hippocampus and entorhinal cortex.[22]

Original studies demonstrating decreased neuronal counts in the mediodorsal and anteriorventral nucleus of the thalamus have been partially supported by subsequent investigations.[17]

The lateral (nigrostriatal) and medial (mesolimbic) parts of the mesencephalic dopaminergic systems have been evaluated and the size of the nerve cell bodies found to be significantly reduced in the medial part by 16 per cent, while the cell numbers were unchanged. The reduced cell size of the medial, mesolimbic neurones were taken to indicate dopaminergic underactivity. Two qualitative reports on degenerative changes in cholinergic cells in the basal nucleus of Meynert of schizophrenics have been published; more recent quantitative studies found normal cell numbers in the basal nucleus of schizophrenics. Volume measurements and cell counts in the noradrenergic locus coeruleus revealed a trend for decreased locus coeruleus volume without loss of neurones, indicating a reduction of neuropil in schizophrenics. These results appear comparable to those described in the substantia nigra, as mentioned above. Investigating the brainstem reticular formation revealed a twofold increased number of the cholinergic neurones of the pedunculopontine nucleus and the dorsal tegmental nucleus as well as a reduced cell size in the locus coeruleus.[23,24] However, these results are not undisputed as newer studies using state of the art stereology demonstrate opposite findings.[25]

Schizophrenia as a disorder of brain maturation

There is evidence from clinical research implicating aberrant neurodevelopmental processes in the pathophysiology of schizophrenia,[26] but there is also a growing literature suggestive of progressive deterioration in the disease for a substantial proportion of patients.[27] It should be noted that abnormal neurodevelopmental processes are not mutually exclusive of neurodegenerative mechanisms in the pathogenesis of complex neuropsychiatric disorders. Indeed, while some genetic disorders are mainly developmental (e.g. fragile X syndrome) and others mainly neurodegenerative (e.g. Huntington's disease), some have both developmental and degenerative pathologies (e.g. Down syndrome). Based on the neuropathological literature of the last 30 years some suggestions can be made concerning the pathophysiology of schizophrenia.

(c) Gliosis

Glial cells, mainly astrocytes (Figs. 2.3.5.2 and 2.3.5.3), show changes in response to almost every type of injury or disease in the central nervous system. Therefore, in typical degenerative brain disorders such as Alzheimer's disease or Huntington's chorea increased glial cell densities are found. Most studies using glial cell counts, neuron-to-glial ratios and glial cell nuclear volumes found no difference between schizophrenics and controls in temporolimbic structures, the thalamus, and cingulate cortex. In our own large-scale study we counted the number of astrocytes in several key regions such as the area surrounding the temporal horn and found no evidence for astrogliosis in schizophrenia (Fig. 2.3.5.3).[28] Although the question of fibrous gliosis (i.e. increase in glial cell fibres) remains more controversial, the well-controlled study by Bruton *et al.* (1990)[11] also rejects fibrous gliosis in schizophrenia.

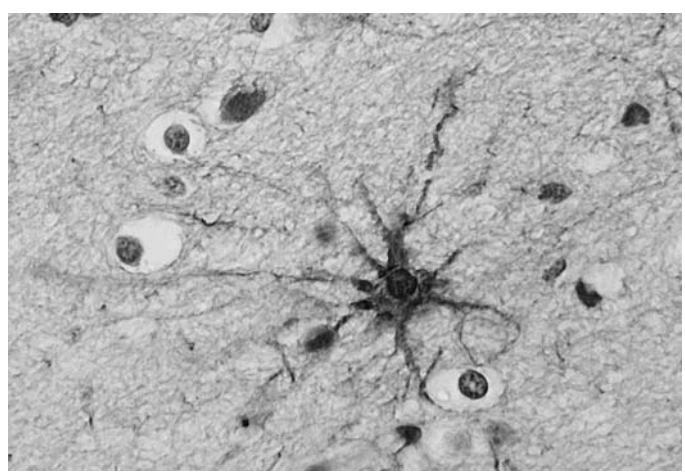


Fig. 2.3.5.2 Macroglia in a control subject: Glial fibrillary acid protein (GFAP) positive astrocyte in the human cortex.

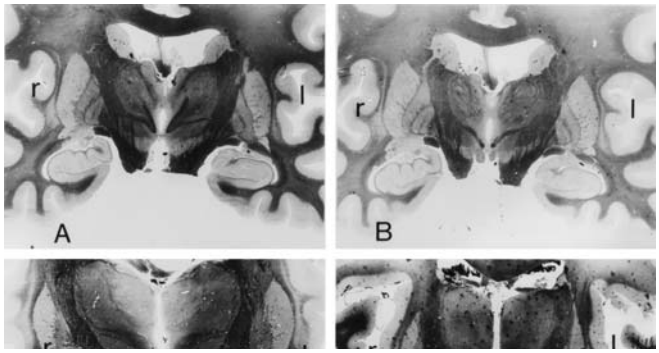


Fig. 2.3.5.3 No significant increase of GFAP positive cells around the left temporal horn in schizophrenia.

Therefore, it seems unlikely that the majority of schizophrenic patients show a considerable degree of astrogliosis. There is, on the contrary, some evidence demonstrating reduced macroglial densities in major depression and schizophrenia.[29] In this respect, specifically the oligodendroglia demonstrates qualitative and quantitative changes in schizophrenia[30] which is an interesting view of the riskgene Neuregulin-1 regulating myelin thickness via these cells. Some recent studies found evidence for the activation of microglia in the cortex of patients with schizophrenia.[31, 32, 33] As microglia respond to neuronal injury within 24 to 48 hours, studies are needed to link psychopathology with these markers. The examination of apoptic processes constitutes another interesting line of research supporting atypical degeneration with schizophrenia. Some recent studies demonstrate low-grade apoptotic processes in circumscribed brain regions in schizophrenia[34, 35] which is in line with other degenerative disorders showing similar features.[36]

(d) Neurohistological indications of disturbed brain development

Subtle cytoarchitectural anomalies were described in the hippocampal formation, frontal cortex, cingulate gyrus, and entorhinal cortex in patients suffering from schizophrenia compared with control subjects. For example, significant cellular disarray in the CA3-CA4 interface was described in the left and replicated in the right hippocampus.[37] This was interpreted as a bilateral migrational abnormality and broadly correlated with the degree of disease severity. One subsequent study was not able to fully replicate these findings, but did confirm a within-case correlation with severity; whereas another examination did not find any significant disarray distinguishing schizophrenics from controls. Another prominent finding was of an abnormal sulcogyral pattern or abnormal gross configuration of the temporal lobe and cytoarchitectonic abnormalities of the rostral entorhinal region as well as of the ventral insular cortex of schizophrenics.[38, 39] The cytoarchitectonic abnormalities of the rostral entorhinal region consisted of heterotopic pre-α-cells in the pre-β-layer (layer III), which would normally belong to the pre-α-layer (layer II). This observation stimulated considerable research, with some studies supporting these findings,[40, 41] while others did not.[42]

In conclusion, cytoarchitectonic abnormalities recently described in different limbic structures in schizophrenia are very subtle and can easily be missed using classical neuropathological methods. Quantifying them often needs sophisticated staining methods, for example, immunohistochemistry, serial sections and a matched control group and even then replicating original findings seems difficult, as outlined above. These findings can be interpreted as a sign for disturbed late neuronal migration or could mirror disturbed programmed cell death as heterotopias are frequently found in the temporal cortex of autopsied children, which seem to disappear in adults.

(e) Summary and pathophysiological conclusions

There is growing evidence for pathomorphological abnormalities in the postmortem brains of patients suffering from schizophrenia. The changes are focused on the frontal lobe and temporolimbic regions. They are subtle, lacking the typical signs of degeneration, and point to problems in prenatal (cell migration) and postnatal (connectivity) periods of brain development. Currently, underlying causes remain ambiguous, but the interaction between genetic and non-genetic factors (e.g. birth complications) is presently discussed on the basis of the recently found risk genes like Neuregulin-1 or Dysbindin.

Mood disorders (Table 2.3.5.3)

The number of published pathoanatomical studies in schizophrenia contrasts with the scant number of neuropathological examinations in affective disorders. In reviewing the world literature up to 1988, Jeste *et al.* (1988)[43] counted 15 neuropathological studies on affective disorders. Seven of them were published between 1949 and 1969, attended to less than four cases, and utilized qualitative tissue evaluation. Searching for 'unipolar depression and neuropathology' and 'bipolar depression', resulted in 56 and 77 hits in pubmed up to 2007, proving this field of research to speed up lately. Comprehensive reviews[44, 45, 46] highlight several aspects in more detail and are suggested for further reading. Table 2.3.5.3 summarizes the most relevant findings from morphometric post-mortem studies.

Macroscopic findings

Four studies examined macroscopic measures such as the gross brain morphology,[47] brain weight and ventricular volume,[48] and the area or volume of specific regions such as the hippocampus, parahippocampal gyrus (Altshuler *et al.* 1990), striatum, globus pallidus, and corpus callosum.[48, 49] In comparison to schizophrenic patients and/or non-psychiatric control subjects, patients with affective disorders revealed caudate lesions,[47] reduced area of the right parahippocampal gyrus,[50] increased brain and reduced ventricular volume.[48]

Microscopic findings

Several studies examined the cytoarchitecture and nerve cell, interneuronal or glial numbers of the pre- and orbitofrontal cortex,[51, 52, 53] entorhinal and insular cortex,[54, 55] anterior cingulated[56, 57] cerebellum,[58] brainstem[49, 59] and the peripheral nervous system.[60] Findings in patients with affective disorders included overall reduced neuronal numbers, together with disurbed cytoarchitecture of entorhinal cortex, reduced neuronal and glial numbers in the prefrontal cortex, the rostroventral insula and dorsal raphe, reduced Purkinje cells in anterior and posterior vermis and hemispheres of the cerebellum, reduced interneuronal numbers in layer II of the cingulate, but increased neuronal numbers of the locus coeruleus and peripheral motor neurone branching. In a series of studies focusing on the hypothalamus in affective disorders, the number of nitric oxide synthase (NOS) positive cells was

reduced in the nucleus suprachiasmaticus[61] and the paraventricular nucleus[62] in patients with affective disorder stressing the importance of this anatomical region for these illnesses.

Summary and pathophysiological conclusions

In summary, the number of postmortem studies on mood disorders is still limited but growing. There is some evidence for changes in key cortical regions, the basalganglia, the hypothalamus and brainstem. Structural brain imaging studies support the notion of mood disorders being associated with regional structural brain abnormalities, in particular regions involved in mood regulation. Because small numbers of subjects were studied, only some postmortem studies distinguished between unipolar and bipolar depression.[49] Nevertheless, recent structural imaging studies regarded this distinction worthwhile. The main abnormalities found in unipolar depression are smaller basal ganglia, cerebellum, frontal lobe and hippocampus, which may reflect disease-course-related atrophy. Bipolar disorder appears to be associated with larger third ventricle, smaller cerebellum, possibly smaller temporal lobe, and perhaps increased amygdala volume on the right side. In both groups, there seems to be an increased rate of subcortical white-matter lesions and periventricular hyperintensities. Whether the rate of subcortical white-matter lesions and periventricular hyperintensities are predictive of a later development of Alzheimer´s disease is yet unclear. Hippocampal plaques and tangles are increased in patients with Alzheimer´s disease with a lifetime history of major depression.[63] Further studies are needed, combining endocrine/biochemical parameters with structural parameters to identify the key regions involved in processes central to mood disorders such as changes in the regulation of the hypothalamic-pituitary-adrenal axis.

Alcoholism (Table 2.3.5.4)

The best known neuropathological feature of alcoholism is Wernicke's encephalopathy, which is characterized by degenerative changes including gliosis and small hemorrhages in structures surrounding the third ventricle and aqueduct (i.e. the mamillary bodies, hypothalamus, mediodorsal thalamic nucleus, colliculi, and midbrain tegmentum), as well as cerebellar atrophy. Most of the clinical features associated with the Wernicke-Korsakoff syndrome including ophthalmoplegia, nystagmus, ataxia, and mental symptoms such as confusion, disorientation, and even coma can be related to damaged functional systems in the hypothalamus, midbrain, and cerebellum.[64] Other important neuropathological manifestations of chronic alcoholism are central pontine myelinolysis, Marchiafava syndrome, and foetal alcohol syndrome (see Chapter 4.2.2.3).

Studies on alcohol-specific brain damage

Most of the changes mentioned above occur in association with thiamin deficiency, which is frequently, but not always, correlated with the long-term use of excessive amounts of alcohol. One major challenge is to identify those lesions caused by alcohol itself (uncomplicated alcoholism)[65] and those caused by other common

Table 2.3.5.3 Morphometric post-mortem studies in affective disorders

Study	Number of patients/controls	Region/parameter	Finding
General			
Jellinger (1977)[46]	4/15	Entire brain	Caudate lesions
Temporal lobe			
Altshuler *et al.* (1990)[49]	12/27	Hippocampal area	—
		Parahippocampal area	↓
Beckmann and Jakob (1991)[53]	4/0	Cytoarchitecture of entorhinal cortex	Disturbed
		Rostroventral insula nerve cell number	↓
Casanova *et al.* (1991)[54]	5/10	Entorhinal cortex neuronal numbers	↓
Other cortical and subcortical regions			
Brown *et al.* (1986)[47]	70/32	Lobar structures	—
		Callosal thickness	—
		Corpus striatum	—
		Brain weight	↑
		Lateral ventricles	↓
Diekmann *et al.* (1998)[55]	12/12	Cingulate cortex (interneurones in layer II)	↓
Baumann *et al.* (1999)[48]	8/8	Accumbens, putamen, caudate, external pallidal volumes	↓
Bernstein *et al.* (1998)[62]		NOS positive cells in the nucleus paraventricularis	↓
Bernstein *et al.* (2002)[61]		NOS positive cells in the nucleus suprachiasmaticus	↓
Cerebellum			
Lohr and Jeste (1986)[57]	12/23	Cerebellum: Purkinje cells in anterior and posterior vermis and hemispheres	—
Brainstem			
Hankoff and Peress (1981)[58]	4/26	Brainstem	—
Baumann *et al.* (1999)[48]	12/12	Locus coeruleus neuronal number	↑
Peripheral nervous system			
Ross-Stanton and Meltzer (1981)[60]		Motor neurone branching (peripheral)	↑

In comparison with controls: ↓, reduced; ↑, increased; —, no difference; ( ), finding not or only partially replicated.

Table 2.3.5.4 Morphometric post-mortem studies in alcoholism

Region/parameter	Finding
General	
Brain weight	↓
Intracranial volume	↓
Ventricular volume	↑
White > grey matter volume	↓
Temporal lobe	
Hippocampal neuronal numbers	(↓)
Amygdala neuronal numbers	↓
Frontal, parietal, and occipital lobes	
Superior frontal cortex, neuronal numbers (BA 8)	↓
Primary motor cortex, neuronal numbers (BA 4)	—
Frontal cingulated cortex, neuronal numbers (BA 32)	—
Inferior temporal cortex, neuronal numbers (BA 20 + 36)	—
Superior frontal cortex, GABAergic pyramidal neurones	—
Thalamus	
Medial dorsal and anterior nuclei of the thalamus, volumes	(↓)
Supraoptic and paraventricular nuclei of the hypothalamus, neuronal numbers	↓
Arginine-vasopressin immunoreactive neurones	↓
Basal ganglia	
Caudate, putamen, or globus pallidus volumes	—
Cerebellum	
Cerebellar volume in general	↓
Vernal, intermediate, and lateral zone volumes	↓
Purkinje cell densities	↓
Brainstem	
Locus coeruleus noradrenergic, neuronal numbers	(↓)
Median and dorsal raphe nuclei, neuronal numbers	—
Other brain structures	
Basal nucleus, neuronal structures	—

In comparison with controls: ↓, reduced; ↑, increased; —, no difference; ( ), finding not or only partially replicated.

alcohol-related factors, principally thiamin deficiency. The following paragraph summarizes recent results in this field, which has been reviewed in detail by others.[65, 66] Brain shrinkage can be found in uncomplicated alcoholism, which can largely be accounted for by loss of white matter. Some of this damage appears to be reversible. However, alcohol-related neuronal loss has been documented in specific regions of the cerebral cortex (superior frontal association cortex), hypothalamus (supraoptic and paraventricular nuclei), and cerebellum. The data are conflicting for the hippocampus, amygdala, and locus coeruleus. No changes are found in the basal ganglia, nucleus basalis, or serotonergic raphe nuclei. Concerning the prefrontal lobe it is interesting to note that although alcohol related pathology affects both neuronal and glial cells, the effects on glia are more dramatic than on neurones.[52] The cellular changes are more prominent and spread across cortical layers in alcohol dependent subjects compared to subjects with mood disorders.[52] As pointed out above, many of the regions being normal in uncomplicated alcoholics are damaged in those with Wernicke-Korsakoff syndrome. Dendritic and synaptic changes have been documented in uncomplicated alcoholics, and these, together with receptor and transmitter changes, may explain functional changes and cognitive deficits that precede the more severe structural neuronal changes.

Summary and pathophysiological considerations

In summary, there is neuropathological evidence showing that alcohol *per se* causes damage to both grey and white matter. White-matter damage is predominant and results in a reduction in brain volume. A component of the white-matter loss appears to be reversible in some cases, given a significant period of abstinence. The grey-matter damage appears to be regionally selective, but many areas of the brain appear to be resistant to damage.

Thiamin deficiency accounts for a major component of the brain damage in alcoholics. Animal models suggest the distribution and extent of neuronal loss to be dependent on the duration of alcohol exposure, the magnitude and mode of exposure (ingestion, inhalation, etc.), the genetic susceptibility of the species, and the strain of animals studied.[65] It has been suggested that alcohol withdrawal may play a role in brain damage, evidenced by the fact that a number of workers have shown loss of granule cells in the dentate gyrus of the hippocampus continuing even after alcohol exposure stops.[67] It was furthermore suggested that up-regulation of *N*-methyl-d-aspartate receptors may lead to withdrawal seizures and enhanced susceptibility to excitotoxicity, which may explain the continuing damage described.[68]

Further information

http://www.psychiatrie.med.uni-goettingen.de/falkai_publikationen.html
http://www.med.uni-magdeburg.de/fme/znh/kpsy/cv/bogerts%20de.htm

References

1. Jellinger, K.A. and Bancher, C. (1998). Neuropathology of Alzheimer's disease: a critical update. *Journal of Neural Transmission Supplementum*, **54**, 77–95.
2. Alzheimer, A. (1898). Beiträge zur pathologischen Anatomie der Hirnrinde und zur anatomischen Grundlage der Psychosen. *Monatsschrift Psychiatrie und Neurologie*, **2**, 82–120.
4. Southard, E.E. (1915). On the topographic distribution of cortex lesions and anomalies in dementia praecox with some account of their functional significance. *American Journal of Insanity*, **71**, 603–71.
5. Vogt, C. and Vogt, O. (1952). Resultats de l'etude anatomique de la schizophrenie et d'autres psychoses dites fontionelles faite a l'institut du cerveau de Neustadt, Schwarzwald. In *First International Congress of Neuropathology*, Vol. 1, pp. 515–32. Rosenberg and Sellier, Turin.
7. Harrison, P.J. (1999). The neuropathology of schizophrenia. A critical review of the data and their interpretation. *Brain*, **122**, 593–624.
8. Arnold, S.E., Talbot, K. and Hahn, C.G. (2005) Neurodevelopment, neuroplasticity, and new genes for schizophrenia. *Progress Brain Research*, **147**, 319–45.
9. Harrison, P.J. and Weinberger, D.R. (2005). Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Molecular Psychiatry*, **10**, 40–68.
10. Stevens, J.R. (1982). Neuropathology of schizophrenia. *Archives of General Psychiatry*, **39**, 1131–9.
11. Bruton, C.J., Crow, T.J., Frith, C.D., *et al.* (1990). Schizophrenia and the brain: a prospective clinico-neuropathological study. *Psychological Medicine*, **20**, 285–304.
12. Golier, J.A., Davidson, M., Haroutunian, V., *et al.* (1995). Neuropathological study of 101 elderly schizophrenics: preliminary findings. *Schizophrenia Research*, **15**, 120.
13. Bogerts, B., Meertz, E., and Schönfeldt-Bausch, R. (1983). Limbic system pathology in schizophrenia: a controlled post mortem study. *7th World Congress of Psychiatry*, Vienna, 1983, Vortrag, Abstract F6.

14. Bogerts, B., Meertz, E. and Schonfeldt-Bausch, R. (1985). Basal ganglia and limbic system pathology in schizophrenia. A morphometric study of brain volume and shrinkage. *Archives of General Psychiatry*, **42**(8), 784–91.
15. Stevens, J.R. (1997). Anatomy of schizophrenia revisited. *Schizophr Bull*, **23**(3), 373–83.
    Bogerts, B. and Lieberman, J. (1993). Neuropathology in the study of psychiatric disease. In *International review of psychiatry* (ed. N.C. Andreasen and M. Sato), Vol. 1 pp. 515–55. American Psychiatric Press, Washington, DC.
16. Chakos, M.H., Lieberman, J.A., Alvir, J., *et al.* (1995). Caudate nuclei volumes in schizophrenic patients treated with typical antipsychotics or clozapine. *Lancet*, **345**(8947): 456–57.
    Broca, P. (1865). Sur la faculté du langage articulé. *Bulletins de la Societé d'Anthropologie*, **6**, 377–93.
17. Danos, P., Schmidt, A., Baumann, B., *et al.* (2005). Volume and neuron number of the mediodorsal thalamic nucleus in schizophrenia: a replication study. *Psychiatry Research* 140(3), 281–9.
18. Innocenti, G.M., Ansermet, F., and Parnas, J. (2003). Schizophrenia, neurodevelopment and corpus callosum. *Molecular Psychiatry*, **8**(3), 261–74.
19. Bogerts, B. (1990). *Die Hirnstruktur Schizophrener und ihre Bedeutung für die Pathophysiologie und Psychopathologie der Erkrankung.* Thieme, Stuttgart
20. Lewis, D.A., Cruz, D., Eggan, S., *et al.* (2004). Postnatal development of prefrontal inhibitory circuits and the pathophysiology of cognitive dysfunction in schizophrenia. *Annuals of the New York Academy of Scienes*, **1021**, 64–76.
21. Christison, G.W., Casanova, M.F., Weinberger, D.R., *et al.* (1989). A quantitative investigation of hippocampal pyramidal cell size, shape and variability of orientation in schizophrenia. *Archives of General Psychiatry*, **46**, 1027–32.
22. Heckers, S., Heinsen, H., Heinsen, Y.C., *et al.* (1990). Limbic structures and lateral ventricle in schizophrenia. *Archives of General Psychiatry*, **47**, 1016–22.
23. Karson, C.N., García-Rill, E., Biedermann, J., *et al.* (1991). The brain stem reticular formation in schizophrenia. *Psychiatry Research*, **40**(**1**), 31–48.
24. Lohr, J.B. and Jeste, D.V. (1988) Locus ceruleus morphometry in aging and schizophrenia. *Acta Psychiatr Scand*, **77**(6), 689–97.
25. Marner, L., Soborg, C., and Pakkenberg, B. (2005). Increased volume of the pigmented neurons in the locus coeruleus of schizophrenic subjects: a stereological study. *Journal of Psychiatric Research*, **39**(4), 337–45.
26. Weinberger, D.R. (1996). On the plausibility of 'the neurodevelopmental hypothesis' of schizophrenia. *Neuropsychopharmacology*, **14**, 1S-11S.
27. Woods, B.T., Yurgelun-Todd, D., Goldstein, J.M., *et al.* (1996). MRI brain abnormalities in chronic schizophrenia: one process or more? *Biological Psychiatry*, **40**, 585–96.
28. Falkai, P., Honer, W.G., David, S., *et al.* (1999). No evidence for astrogliosis in brains of schizophrenic patients. A post-mortem study. *Neuropathology and Applied Neurobiology*, **25**, 48–53.
29. Cotter, D.R., Pariante, C.M. and Everall, I.P. (2001). Glial cell abnormalities in major psychiatric disorders: the evidence and implications. *Brain Research Bulletin*, **55**, 585–95.
30. Uranova, N.A., Vostrikov, V.M., Vikhreva, O.V., *et al.* the role of oligodendrocyte pathology in schizophrenia. *International Journal of Neuropsychopharmacology*, **Feb 21**, 1–9 (Epub ahead of print).
31. Bayer, T.A., Buslei, R., Havas, L., *et al.* (1999). Evidence for activation of microglia in patients with psychiatric illnesses. *Neuroscience Letters*, **271**, 126–8.
32. Radewicz, K., Garey, L.J., Gentleman, S.M. *et al.* (2000). Increase in HLA-DR immunoreactive microglia in frontal and temporal cortex of chronic schizophrenics. *Journal of Neuropathology and Experimental Neurology*, **59**, 137–50.
33. Foster, R., Kandanearatchi, A., Beasley, C., *et al.* (2006). Calprotectin in microglia from frontal cortex is up-regulated in schizophrenia: evidence for an inflammatory process? *European Journal of Neuroscience*, **24**, 3561–6.
34. Jarskog, L.F., Miyamoto, S. and Lieberman, J.A. (2007). Schizophrenia: new pathological insights and therapies. *Annu Rev Med*, **58**, 49–61.
35. Benes, F.M., Matzilevich, D., Burke, R.E., *et al.* (2006). The expression of proapoptosis genes is increased in bipolar disorder, but not in schizophrenia. *Molecular Psychiatry*, **11**, 241–51.
36. Glantz, L.A., Gilmore, J.H., Lieberman, J.A. *et al.* (2006). Apoptotic mechanisms and the synaptic pathology of schizophrenia. *Schizophrenia Research*, **18**, 47–63.
37. Kovelman, J.A. and Scheibel, A.B. (1984). A neurohistological correlate of schizophrenia. *Biological Psychiatry*, **19**, 1601–21.
38. Jakob, J. and Beckmann, H. (1986). Prenatal developmental disturbances in the limbic allocortex in schizophrenics. *Journal of Neural Transmission*, **65**, 303–26.
39. Jakob, H. and Beckmann, H. (1989). Gross and histological criteria for developmental disorders in brains of schizophrenics. *Journal of the Royal Society of Medicine*, **82**, 466–9.
40. Falkai, P. Schneider-Axmann, T. and Honer, W.G. (2000). Entorhinal cortex pre-alpha cell clusters in schizophrenia: quantitative evidence of a developmental abnormality. *Biological Psychiatry*, **47**, 937–43.
41. Kovalenko, S., Bergmann, A., Schneider-Axmann, T.,*et al.* (2003). Regio entorhinalis in schizophrenia: more evidence for migrational disturbances and suggestions for a new biological hypothesis. *Pharmacopsychiatry*, **36**, **Suppl 3**: S158–61.
42. Krimer, L.S., Herman, M.M., Saunders, R.C., *et al.* (1997). A qualitative and quantitative analysis of the entorhinal cortex in schizophrenia. *Cereb Cortex*, **7**, 732–9.
43. Jeste, D.V., Lohr, J.B., and Goodwin, F.K. (1988). Neuroanatomical studies of major affective disorders. A review and suggestions for further research. *British Journal of Psychiatry*, **153**, 444–59.
44. Harrison, P.J. (2002). The neuropathology of primary mood disorder. *Brain*, **125**, 1428–49.
45. Baumann, B. and Bogerts, B. (2001). Neuroanatomical studies on bipolar disorder. *British Journal of Psychiatry Supplemental*, **41**, S142–7.
46. Haldane, M. and Frangou, S. (2004) New insights help define the pathophysiology of bipolar affective disorder: neuroimaging and neuropathology findings. *Progress neuropsychopharmacol Biological Psychiatry*, **28**, 943–60.
47. Jellinger, K.A. (1977). Neuropathologic findings after neuroleptic long term therapy. In *Neurotoxicology* (eds. L. Roizin, H. Shiraki, and N. Grcevic), pp. 25–42. Raven Press, New York.
48. Brown, R., Colter, N., Corsellis, J.A.N., *et al.* (1986). Postmortem evidence of structural brain changes in schizophrenia. Differences in brain weight, temporal horn area and parahippocampal gyrus compared with affective disorder. *Archives of General Psychiatry*, **43**, 36–42.
49. Baumann, B., Danos, P., Krell, D., *et al.* (1999). Unipolar-bipolar dichotomy of mood disorders is supported by noradrenergic brainstem system morphology. *Journal of Affective Disorders*, **54**, 217–24.
50. Altshuler, L.L., Casanova, M.F., Goldberg, T.E., *et al.* (1990). The hippocampus and parahippocampus in schizophrenic, suicide, and control brains. *Archives of General Psychiatry*, **47**, 1029–34.
51. Rajkowska, G., Halaris, A. and Selemon, L.D. (2001). Reductions in neuronal and glial density characterize the dorsolateral prefrontal cortex in bipolar disorder. *Biological Psychiatry*, **49**, 741–52.
52. Miguel-Hidalgo, J.J. and Rajkowska, G. (2003) Comparison of prefrontal cell pathology between depression and alcohol dependence. *Journal of Psychiatric Research*, **37**, 411–20.

53. Bielau, H., Steiner, J., Mawrin, C., *et al.* (2007) Dysregulatin of GABAergic neurotransmission in mood disorders: a postmortem study. *Annals of New York Academy of Sciences*, **1096**, 157–69.
54. Beckmannn, H. and Jakob, H. (1991). Prenatal disturbances of nerve cell migration in the entorhinal region: a common vulnerability factor in functional psychoses? *Journal of Neural Transmission. General Section*, **84**, (1–2), 155–64.
55. Casanova, M.F., Saunder, R., Altshuler, L., *et al.* (1991). Entorhinal cortex pathology in schizophrenia and affective disorders. In *Biological Psychiatry* (ed. G. Racagni *et al.*), pp. 504–6. Elsevier, Amsterdam.
56. Diekmann, S., Baumann, B., Schmidt, U., *et al.* (1998). Significant decrease in calretinin immunoreactive neurons in layer II in the cingulate cortex in schizophrenics. *Pharmacopsychiatry*, **31**, 12.
57. Chana, G., Landau, S., Beasley, C., *et al.* (2003) Two-dimensional assessment of cytoarchitecture in the anterior cingulate cortex in major depressive disorder, bipolar disorder, and schizophrenia: evidence for decreased neuronal somal size and increased neuronal density. *Bioliological Psychiatry*, **53**, 1086–98.
58. Lohr, J.B. and Jeste, D.V. (1986). Studies of cerebellum and hippocampus in major psychiatric disorders. In *Biological psychiatry* (ed. C. Shagass *et al.*), pp. 1024–6. Elsevier, New York.
59. Hankoff, L.D. and Peress, N.S. (1981). Neuropathology of the brain stem in psychiatric disorders. *Biological Psychiatry*, **16**, 945–52.
60. Ross-Stanton, J. and Meltzer, F.A. (1981). Motor neuron branching patterns in psychotic patients. *Archives of General Psychiatry*, **38**, 1097–103.
61. Soares, J.C. and Mann, J.J. (1997). The anatomy of mood disorders-review of structural neuroimaging studies. *Biological Psychiatry*, **41**, 86–106.
61. Bernstein, H.G., Heinemann, A., Krell, D., *et al.* (2002) Further immunohistochemical evidence for impaired NO signaling in the hypothalamus of depressed patients. *Annals of New York Academy of Sciences*, **973**, 91–3.
62. Bernstein, H.G., Stanarius, A., Baumann, B., *et al.* (1998). Nitric oxide synthase-containing neurons in the human hypothalamus: reduced number of immunoreactive cells in the paraventricular nucleus of depressive patients and schizophrenics. *Neuroscience*, **83**, 867–75.
63. Rapp, M.A., Schnaider-Beeri, M., Grossman, H.T., *et al.* (2006). Increased hippocampal plaques and tangles in patients with Alzheimer disease with a lifetime history of major depression. *Archive of General Psychiatry*, **63**, 161–7.
64. Victor, M., Adams, R.D., and Collins, G. (1989). *The Wernicke-Korsakow syndrome and related neurologic disorders due to alcoholism and malnutrition*. Davis, Philadelphia, PA.
65. Harper, C.G. and Kril, J.J. (1990). Neuropathology of alcoholism. *Alcohol and Alcoholism*, **25**, 207–16.
66. Harper, C.G. (1998). The neuropathology of alcohol-specific brain damage, or does alcohol damage the brain? *Journal of Neuropathology and Experimental Neurology*, **57**, 101–10.
67. Vogt, C. and Vogt, O. (1948). Über anatomische Substrate. Bemerkungen zu pathoanatomischen Befunden bei Schizophrenie. *Ärztliche Forschritte*, **3**, 1–7.
67. Cavazos, J.E., Das, I., and Sutula, T.P. (1994). Neuronal loss induced in limbic pathways by kindling: evidence for induction of hippocampal sclerosis by repeated brief seizures. *Journal of Neuroscience*, **14**, 3106–21.
68. Hoffmann, P.L., Iorio, K.R., Snell, L.D., *et al.* (1995). Attenuation of glutamate-induced neurotoxicity in chronically ethanol-exposed cerebellar granule cells by NMDA receptor antagonists and ganglioside GM1. *Alcoholism, Clinical and Experimental Research*, **19**, 721–6.

2.3.6 Functional positron emission tomography in psychiatry

P. M. Grasby

Introduction

Positron emission tomography (**PET**) and single-photon emission tomography (**SPET**) are powerful tools for investigating the pathophysiology of psychiatric illnesses and the action of psychotropic drugs. With these techniques monoaminergic, cholinergic, opioid and benzodiazepine receptors, regional cerebral blood flow, glucose and oxygen metabolism can be measured in the living brain (Table 2.3.6.1). Thus, neural function of direct relevance to neurochemical and anatomical theories of psychiatric illnesses can be sampled.

Methodology of PET and SPET[(1)]

In brief, PET and SPET comprise the following:

- The production and incorporation of a positron or gamma-emitting radio-isotope into a molecule of biological interest to form a radiotracer administered to humans (Plate 3).
- The use of PET or SPET cameras to detect the emitted gamma radiation from the decaying radio-isotope and hence the 3D distribution of the radiotracer, over minutes to hours, in living human brain (Plate 4).
- Quantification of a physiological parameter of interest, such as number of available receptors or regional cerebral blood flow, from the mathematical modeling of the measured radio-activity in the brain over time (Plates 5 and 6).

Production of isotopes

Common PET radio-isotopes, produced by a cyclotron, are oxygen-15 ($^{(15)}$O), carbon-11 ($^{(11)}$C), and fluorine-18 ($^{(18)}$F) (with half-lives of 2.03 min, 20.4 min, and 109.8 min respectively) whilst SPET radio-isotopes include technetium ($^{(99m)}$Tc) and iodine-123 ($^{(123)}$I) (with half-lives of 6.02 h and 13.2 h respectively). With appropriate radiochemistry, isotopes can be incorporated into specific molecules to make radiotracers. Following quality control procedures, to estimate specific activity and radiochemical purity, the radiotracer is injected intravenously into subjects lying in the PET camera (Plate 3). Importantly, the total mass of radiotracer injected is very small (typically less than 5 μg) and therefore the radiotracer has no pharmacological effect itself.

PET versus SPET

SPET radiotracers are less diverse than PET tracers. However, SPET is cheaper than PET and less technically demanding, making it more readily available in hospitals and research centres. PET radiochemical procedures require in-house automated rapid synthetic chemistry facilities in dedicated hot cells, whereas SPET chemistry is more straight-forward and does not require such extensive facilities. For research and quantitation purposes, PET is far superior to SPET, although any widespread commercial/clinical application is

Table 2.3.6.1 Established and novel radiotracers for psychiatry

Radiotracer	Application	Comments
PET radiotracers		
$H_2{}^{15}O$	Blood flow	Used to map dysfunctional brain areas involved in psychiatric illnesses. Effectively replaced by functional MRI techniques such as BOLD.
^{18}F-FDG	Glucose metabolism	Used for many resting state studies and nowadays to define psychotropic drug effects.
^{11}C-SCH 23390	Dopamine D_1 receptor	Receptor occupancy studies with neuroleptics. Reports of altered cortical D_1 receptors in drug naive schizophrenics.
^{11}C-NNC 112		
^{11}C-Raclopride	Dopamine D_2 receptor	Robust demonstration of no elevation of striatal D2 receptors in drug naive schizophrenics. Striatal D_2 receptor occupancy studies with many neuroleptics. Frequently used to index dopamine release.
^{11}C-FLB-457	Dopamine D_2 receptor	High affinity ligand; enabling extrastriatal D_2 populations to be measured. Studies in schizophrenia in progress. Binding may be sensitive to endogenous dopamine release.
^{18}F-Fallypride	Dopamine D_2 receptor	High affinity ligand; enabling striatal and extrastriatal D_2 populations to be measured. Binding may be sensitive to endogenous dopamine release.
^{18}F-Fluorodopa	Dopamine synthesis capacity	Radiotracer predominantly imageable in basal ganglia, cortical signal weak. Consistent reports of raised ^{18}FDOPA in schizophrenia.
^{11}C-Flumazenil	Central benzodiazepine receptors	Labels all subtypes of central receptor.
^{11}C-MDL-100907	5-HT_{2A} receptors	Most suitable ligand for imaging 5-HT_2 receptors.
^{11}C-WAY 100635	5-HT_{1A} receptors	Reports of reduced 5-HT1A availability in depressive and anxiety disorders
^{11}C-desmethyl WAY		
^{11}C – FCWAY		
^{18}F - MPPF		
^{11}C-DASB	5-HT transporter	Studies in depressive illness. Occupancy studies of SSRIs.
^{11}C-McN 5652		Used to examine effects of ecstasy
SPECT radiotracers		
^{99m}TcHMPAO	Blood flow	Many resting state and two scan activation studies in psychosis.
^{123}I-Iodobenzamide	Dopamine D2 receptors	Occupancy and dopamine release studies in schizophrenia
^{123}I-Epidepride	Dopamine D2 receptors	Striatal and Extrastriatal D2 receptors. Used to show 'limbic selectivity' of certain neuroleptics
^{123}I-QNB	Muscarinic acetycholine receptors	
^{123}I-CIT	Dopamine and 5-HT reuptake sites	Studies in depressive illness

likely to be SPET based because of the technology restraints and costs associated with PET scanning.

Imaging of radiotracer, data collection, and analysis

PET utilizes the disintegration of positrons emitted from unstable nuclei such as$^{(11)}$C (Plate 4). Emitted positrons travel a short distance in tissue before annihilation by collision with an electron.[1] On annihilation, two high-energy gamma rays are generated with a separation angle of 180° (Plate 4). Radiation detectors (e.g. bismuth germanate), 180° apart and linked in electronic coincidence circuits, detect the resulting gamma radiation and therefore localize the source of radiation to a volume between any two detectors (Plate 3). By arranging rings of detectors around the subject's head and using computer-based back-projection techniques, the distribution of radiotracer within tomographic slices of the brain can be obtained.[1] SPET radioisotopes, in contrast, decay by emitting a single gamma ray and therefore the radiation detectors are not linked in coincidence circuits. State-of-the-art PET and SPET cameras have transaxial spatial resolutions of the order of 4 to 5 mm and can detect subnanomolar concentrations of receptors.[1]

Positron-emitting isotopes can be incorporated into molecules associated with diverse biochemical processes in the brain. For example, the positron emitter$^{(11)}$C can be incorporated into a molecule WAY 100635, which selectively binds to 5-HT_{1A} receptors, and injected intravenously in tracer amounts. Brain regions will show different profiles of radio-activity accumulation over time as the radiotracer binds in areas with a high density of 5-HT_{1A} receptors (medial temporal cortex) whilst in regions with no or sparse receptors (cerebellum), it will be washed out (Plate 5). By this means, specific and non-specific binding can be distinguished. With an appropriate model of the radiotracer's history in tissue over time, a quantitative measurement of 5-HT_{1A} receptor number in tomographic slices of the human brain can be obtained.[1] With some radiotracers (e.g. [11C]diprenorphine to label opiate receptors) it may be necessary to undertake radial artery cannulation to obtain an 'input function'[1] that describes the time course of presentation of radiotracer to the brain (Plate 6), whereas others tracers can be modeled with a 'pseudo' input function from a reference region.

Technical and practical limitations of PET and SPET compared with other imaging modalities

PET and SPET excel in the measurement of neurochemical parameters *in vivo* at very low (subnanomolar) concentration. Such sensitivity cannot be matched by other *in vivo* methods such as proton magnetic resonance spectroscopy (millimolar range). However, radiation dosimetry limits the number of scans that subjects may receive. Full quantitation can often be achieved with PET, unlike

SPET. However, for imaging blood flow change, or its correlates such as BOLD arterial spin labeling (ASL) and functional magnetic resonance imaging (fMRI), now offer the possibility of repeated measures (without radiation exposure) that far exceed that possible with PET- and SPET-based methods of flow mapping. Full quantitation of blood flow is not yet readily achievable with functional MRI without injection of contrast agents. In contrast, MRI based ASL can achieve full quantification. One disadvantage of functional MRI over PET, for some subjects, is the noisy claustrophobic environment of the scanner, but generally subjects and paradigms studied with PET flow mapping can be readily investigated with functional MRI (see Chapter 2.3.8), although all test materials in the vicinity of the scanner have to be non-magnetic.

Structural MRI scanning is often used in conjunction with PET activation and ligand binding techniques. The high-resolution anatomical information contained in MRI images can be used to precisely define areas of activation or radiotracer binding observed in PET studies from single subjects.

PET and SPET, and even functional MRI, have relatively poor temporal resolution (seconds) compared with electrophysiological methods such as EEG, event-related potentials, and magnetoencephalography (milliseconds), but these methods in turn suffer from poor spatial resolution. Attempts to integrate information from these different modalities are a major focus of methodological research in many imaging centres.

PET and SPET imaging strategies in psychiatry

These techniques (see Table 2.3.6.2) are used to either measure brain receptors and neurochemistry, or map functional brain activity via the indices of regional blood flow and glucose utilization.

Table 2.3.6.2 Summary of PET functional brain imaging approaches

Functional brain mapping: rCBF or metabolism is measured as an index of local neural activity

(a) Studies in normal volunteers in which 'activation' paradigms are used to identify functional anatomy that is relevant to psychiatric disorders
(b) Activation studies in patients who are compared with matched control subjects
(c) Studies in which the biological variable (e.g. rCBF) is correlated with a relevant clinical variable (e.g. hallucinations) within the patient group
(d) The longitudinal comparison of patients before and after various treatments and into clinical recovery
(e) Cross-sectional studies of resting-state brain activity in patient groups in comparison with appropriate controls

Radioligand imaging: the specific uptake and binding of radiolabelled tracer compounds is measured

(a) To estimate baseline radioligand uptake at rest in patient groups in comparison with controls
(b) Within-patient group correlations between radioligand uptake and particular symptoms/signs
(c) Longitudinal comparison of radioligand uptake in patients before and after various treatments and into clinical recovery
(d) 'Displacement' or radioligand activation studies designed to detect changes in the levels of intrasynapcic neurotransmitters in response to a pharmacological or cognitive challenge
(e) Investigation of the receptor binding and occupancy characteristics of psychotropic drugs

rCBF, regional cerebral blood flow.

Each approach attempts to define trait and state abnormalities of psychiatric illnesses or the effect of psychotropic drug action.

Because of the technical complexities, it is important to bear in mind the following questions when judging experimental results.

- What assumptions are made about the behaviour of the radiotracer *in vivo*?
- Has the radiotracer been well validated for the apparent physiological parameter measured ?
- Does the mathematical model of the radiotracer's behaviour give a good fit to the raw data?
- Is the spatial resolution of the PET camera sufficient for the regions measured?
- What is the test-test reliability for the PET radiotracer measure?
- How have the raw PET images been modified/treated in the data analysis?
- How have regions of interest been defined?
- Is there a possibility of observer bias in the measurements made?
- What statistical techniques have been used, and are the statistical thresholds appropriate?
- Do the statistics reflect fixed or random effects and the multiple comparisons made?

Measuring brain receptors and neurochemisty

Many neurochemical hypotheses, generated by post-mortem and animal data, can be rigorously tested with PET and SPET in the living brain whilst avoiding many of the confounding variables inherent in *in vitro* techniques. Many receptor systems can be studied (Table 2.3.6.1). For receptor mapping, particular successes in this area are the range of tracers available to image select components of dopaminergic and serotoninergic neurotransmission. For example, with presently available radiotracers it is possible to measure Dopamine D_1, striatal D_2, extrastriatal D_2, Dopamine reuptake sites, index Dopamine synthesis and endogenous dopamine release. For the serotonin system, 5-HT_{1A}, 5-HT_{2A} and 5-HT transporters can be readily measured. However, the rate of discovery of new radiotracers, suitable for use in humans, is relatively slow. Many conditions have to be satisfied to produce a suitable radiotracer for human use including blood-brain barrier permeability, high specific binding, receptor selectivity, absence of radioactive metabolites in brain, and adequate modelling of tracer kinetics (Plate 6). Practically, the increasing use of PET neuroreceptor mapping in the pharmaceutical sector should lead to a greater range of tracer availability due to the large chemical libraries within the industry.

Mapping brain activity by imaging blood flow and glucose metabolism

Regional cerebral blood flow and glucose metabolism are indicators of regional neuronal synaptic activity.[2] Radiotracers for these processes, such as $H_2{}^{(15)}O$ to index blood flow, are used to image brain activity in psychiatric illness. Glucose metabolic mapping using [$^{(18)}$F]deoxyglucose has some disadvantages over flow mapping. Radiation dosimetry limits for [$^{(18)}$F]deoxyglucose and the long half-life of the tracer restrict repeated measurement in a subject over a short time-scale. In contrast,$^{(15)}$ O-based methods allow, for example, 12 measurements of regional blood flow over a 3 h period in a single subject.

The rapid development of PET cameras and automated data analysis techniques in the last few decades has established flow-based functional imaging as a large and active research activity.[2]

As the techniques of regional cerebral blood flow mapping have advanced there has been an equivalent sophistication of experimental design with rest state studies being overshadowed by activation paradigms.[2] In an activation design, subjects are engaged in a specific cognitive task whilst being scanned, for instance generating words, and the blood flow pattern is compared with flow present in a baseline condition such as repeating words. PET activation experiments may involve categorical, correlational, and factorial designs.[2] Although, the more recent advent of fMRI-based methods is now surpassing PET-based methods for functional brain activation mapping, there are certain situations where PET-based methods may be preferred. These include functional measurements in brain areas prone to susceptabilty artifacts in fMRI such as the temporal poles/basal frontal areas. In addition, PET [$^{(18)}$F]deoxyglucose remains a more direct measure of neural activation than all flow-related methods including fMRI and ASL, a potentially important consideration when investigating the central functional effects of CNS drugs that might also alter blood flow directly.

Novel designs and data analysis for PET studies

Developments in this area have been very rapid.[2] Examples would include the use of principal components technique to analyze PET activation data sets and attempts to determine measures of functional and effective connectivity between brain regions activated by a given task. For PET receptor studies, the use of cluster analysis, parametric approaches and simplified reference region models are of considerable interest and are now in common use.[3] Furthermore, attempts are being made to relate PET neurochemical measures to other imaging modalities and genetic factors. Examples of studies in these areas would include explorations of the relationship between densities of 5-HT$_{1A}$ receptors and amygdala reactivity during emotional processing,[4] and influence of genetic polymorphisms on 5-HT$_{1A}$ receptor expression.[5]

Imaging pathophysiology: examples from schizophrenia research

Imaging dopamine receptors

Much research effort has focused on *in vivo* PET/SPET measurement of striatal dopamine D_2-receptor number in schizophrenia following the initial post-mortem reports of increased striatal dopamine receptor number. Initially, using [$^{(11)}$C]N-methylspiperone as a radiotracer, a two- to threefold raised striatal D_2-receptor number in drug-naive schizophrenics was reported.[6] However, subsequently other investigators using [$^{(11)}$C]raclopride, [$^{(11)}$C]N-methylspiperone, [$^{(123)}$I]iodobenzamide, [$^{(76)}$Br]bromolisuride failed to detect such elevations of striatal dopamine D_2-receptor number.[7,8] The different radiotracer methodologies used, the selectivity of radiotracers for dopamine D_2, D_3, and D_4 receptor subtypes, and the clinical characteristics of the patients studied have been advanced as possible explanations for the failure to replicate raised striatal dopamine D_2-receptor number. However, given these conflicting but essentially negative results, attention has shifted in recent years to reports of increased presynaptic dopaminergic function measured with [$^{(18)}$F]dopa[9] and cortical dopamine D_1 receptors measured with [$^{(11)}$C]SCH 23390 and [$^{(11)}$C]NNC 112[10] in schizophrenia. Most recently low density extrastriatal D_2 receptors are being imaged with [$^{(11)}$C]FLB-457 and [$^{(18)}$F]Fallypride. However, the newer patient studies reporting changes of cortical D_1 and extrastriatal D_2 receptors await further replication.

A novel extension to studies utilizing PET/SPET radiotracers for imaging dopamine D_2 receptors has been to index dopamine release during a pharmacological challenge in schizophrenia. Theoretically, PET/SPET has the potential to detect neurotransmitter release associated with behavioural and pharmacological challenges if sufficient endogenous neurotransmitter is released to cause appreciable change (via receptor occupancy) in the number of 'available' receptors that can be 'seen' by a radioligand. For example, pre-dosing animals and human subjects with d-amphetamine, which releases dopamine, results in decreased [$^{(11)}$C]raclopride and [$^{(123)}$I]iodobenzamide binding to dopamine D_2 receptors.[11] Enhanced release of striatal dopamine in acutely symptomatic patients with schizophrenia following pharmacological challenge has been reported in a large cohort.[10,12] In these studies, the displacement of radiotracer (presumably reflecting increased release of dopamine) correlated with worsening of positive symptoms. This important finding of increased dopaminergic responsivity, together with the consistent reports of raised [$^{(18)}$F] FDOPA uptake, provide some of the most convincing *in vivo* evidence to support the hypothesis of subcortical dopamine overactivity in schizophrenia.

Imaging blood flow change, hypofrontality and cortical inefficiency

From the outset of the functional neuroimaging of schizophrenia there has been discussion as to whether the frontal lobes of patients are 'less active' than those of normal subjects.[13] However, hypofrontality, whether at rest or during cognitive challenge, has not been a universal finding in all studies, making it a somewhat unreliable trait marker of schizophrenia. It has been found in about 50 per cent of resting-state studies but more often in activation paradigms.[14] Discrepant results might be attributable to the nature and demands of the task used, task performance, and the symptom profiles of patients studied. For example, Frith *et al.*[15] and Fletcher *et al.* used PET to study paced verbal fluency activations in chronic and acute schizophrenic patients, on and off neuroleptic medication, respectively, and failed to find hypofrontality. But pacing tasks could be criticized on the grounds that slowing the task so that patients and normal subjects perform equally, fails to address a dysfunction that may be expressed when patients are required to produce 'normal' levels of performance. This is a difficult issue to resolve. Pacing patients and controls means that performance levels may be matched, and therefore differences of brain activation are not confounded by the patients' failure to do the task. Yet it may be instructive to image patients attempting to perform a task that stresses (dysfunctional) cognitive processes, produces altered brain activity and hence impaired performance.

Many authors have suggested hypofrontality may be most pronounced in schizophrenic patients who have predominantly 'negative' symptoms. This view has received support from a large cross-sectional study of chronically symptomatic patients where resting blood flow was measured in 30 patients.[16] Within this

group of schizophrenic patients, greater hypofrontality was seen among those with the most pronounced negative symptoms as assessed by factor analysis. That symptoms and not the diagnosis of schizophrenia *per se* may be an important factor in hypofrontality is apparent in one study where poverty of speech (a sign of psychomotor retardation or poverty) was associated with reduced regional cerebral blood flow in the left dorsolateral prefrontal cortex, irrespective of diagnosis of depression or schizophrenia.[17] Some SPET and PET studies also suggest a relationship between hypofrontality and the presence of positive symptoms,[18,19] and hypofrontality (at rest or on activation) may resolve when symptoms improve.[20] Finally, contrary to the earlier notions of hypofrontality an alternative view of 'cortical inefficiency' has been advocated whereby patients with schizophrenia are suggested to have over activation (i.e. hyperfrontality) of frontal areas. In these studies it has been suggested that when patients and controls are performing at similar levels, patients show enhanced cortical activation reflecting the inefficient signal processing within the frontal cortex.[21]

Imaging pathophysiology: examples from depressive disorders

Imaging 5-hydroxytryptamine receptors

Impressive progress has seen the development of new radiotracers for the 5-hydroxytryptamine (**5-HT**) system; hypothesized to be dysfunctional in affective illness and to be the prime target for many antidepressant treatments. In particular, radioligands for 5-HT_{1A},5-HT_{2A} receptors, and the 5-HT transporter, are now established (Table 2.3.6.1).

One notable success is the radioligand [$^{(11)}$C]WAY 100635 for imaging 5-HT_{1A} receptors in the human brain.[22] As many antidepressant treatments alter 5-HT_{1A} receptor function in rodents, and 5-HT_{1A} knock out mice are anxious this ligand is proving useful investigating 5-HT_{1A} receptor populations in depressed or anxious patients before and after treatment. Studies in anxiety and depression are now being reported with suggestions of reductions of 5-HT_{1A} availability in these conditions.[23]

[$^{(11)}$C]N-methylspiperone, [$^{(18)}$F]altanserin, [$^{(18)}$F]ethylspiperone, [$^{(18)}$F]setoperone or [$^{(18)}$F]altanserin, and the SPET tracer [$^{(123)}$I]ketanserin have been used to measure 5-HT_2 receptor number; a receptor implicated in depressive illness, suicidal behaviour, and psychosis. Many of these 5-HT_2 ligands have been hampered by either the lack of selectivity, or the relatively low ratio of specific to non-specific signal obtained in the human brain,[24] although a few studies have appeared, reporting reduced 5-HT_2 receptor number in drug-free depressed patients. Further studies are needed using more selective ligands with higher signal-to-noise ratios, such as [$^{(11)}$C]MDL 100907, a promising selective ligand for 5-HT_{2A} receptors (Table 2.3.6.1). For the serotonin transporter, of the available tracers, [$^{(11)}$C]DASB gives a reasonable signal to noise ratio and studies have convincingly shown that standard efficacious doses of SSRI antidepressants are associated with substantial occupancy at this site.[25] Occupancy at the 5-HTT site occurs after first dosing, and responders and non responders do not differ in terms of SSRI occupancy levels, suggesting 5-HTT occupancy is perhaps a necessary but not sufficient explanation for the antidepressant effect.

Imaging blood-flow change in depressive disorder

Similarly to brain-mapping studies of patients with schizophrenia, regional deficits of neural activity (indexed by cerebral blood flow or glucose utilization) can be detected in the 'resting' brains of depressed patients.[26,27] Many resting-state studies have shown a reduction of regional brain functional activity, most frequently reported in the prefrontal cortex, compared with normal controls. However, the exact location of prefrontal change (dorsolateral, ventrolateral, orbitofrontal, and medial frontal areas) has been variably emphasized by different authors.[26–28]

As demonstrated in schizophrenia, significant associations between cortical activity and cognitive function, symptom clusters, including mood, and response to treatments are apparent.[27,28] Similar to schizophrenia, the resting-state functional brain abnormalities may represent the physiological correlates of aspects of the depressed state such as depressed mood, retardation, or cognitive impairment rather than trait markers of the illness itself.

Psychological challenge paradigms have been applied in studies in depressed cohorts to test whether specific brain regions, subserving select cognitive processes, are impaired in depressed patients. Currently, however, the majority of challenge paradigms in depressed subjects are undertaken with fMRI methods.

Imaging psychotropic drug action

Of direct clinical relevance are imaging studies of antipsychotic drug action where clinical efficacy and side-effects are related to receptor occupancy.[29,30]

Many studies have investigated the occupancy of striatal dopamine D_1 and D_2 and cortical serotonin 5-HT_2 receptors by neuroleptic drugs. Farde's group first demonstrated that clinically efficacious doses of a variety of classical antipsychotics cause between 65 per cent and 89 per cent occupancy of central dopamine D_2 receptors.[29] Higher receptor occupancy [(gt)85 per cent] is associated with an increased incidence of extrapyramidal side effects. Thus, there may be a therapeutic window for occupancy of between 65 per cent and 85 per cent, which is antipsychotic and yet less likely to cause extrapyramidal side effects. In contrast, treatment with classical antipsychotics produces variable levels of occupancy of striatal D_1 receptors. Interestingly, efficacious doses of the atypical antipsychotic clozapine are associated with a relatively low D_2 receptor occupancy (38–63 per cent) and a D_1 occupancy of 38 to 52 per cent.[29] This unexpected finding of low D_2 receptor occupancy, reproduced in different patient groups with both PET and SPET techniques, has challenged theories of a simple relationship between D_2 occupancy *per se* and clinical efficacy. Further evidence for this view comes from studies showing that schizophrenic antipsychotic non-responders have the same levels of dopamine D_2 occupancy as responders and that occupancy occurs as rapidly as 2 h after acute administration of the antipsychotic yet efficacy takes weeks to appear.

PET/SPET is also proving useful in the characterization of atypical antipsychotics. The binding of antipsychotic drugs to central 5-HT_2 receptors is a possible candidate for the mechanism of 'atypicality' and studies suggest high cortical 5-HT_2 occupancy with many atypicals including risperidone (80 per cent) and clozapine (84–90 per cent).[29,30] Other targets for some atypical neuroleptics include the 5-HT_{1A} receptor[31] and in addition, limbic

selectivity (i.e. neuroleptic limbic occupancy > striatal occupancy) is an actively researched explanation for atypicality.[32]

Conclusions

So far, the PET and SPET radiotracer techniques have been immediately valuable in assessing the receptor occupancy effects of antipsychotic drugs and of mapping the neural correlates of dysfunctional cognitive processes and psychiatric symptoms. Although the applications described have not yielded a diagnostic test, the techniques are undoubtedly providing unique information about the pathophysiology of psychiatric illnesses. Such information is likely to be key for the development of truly novel treatments.

Further information

www.crump.ucla.edu/lpp/lpphome.html A website that gives more detailed information on the methodology of PET scanning and clinical applications.

Journals publishing regular research articles on PET neuroscience include *Biological Psychiatry, Neuroimage, Synapse, Brain, Human Brain Mapping, Journal of Nuclear Medicine, Molecular Psychiatry and Neuropsychopharmacology* amongst others. Annual and biannual meetings of the Society of Nuclear Medicine, Organization for Human Brain Mapping and the Society of Cerebral Blood Flow and Metabolism include many 'cutting edge' PET neuroscience reports.

References

1. Myers, R., Cunningham, V., Bailey, D., *et al.* (1996). *Quantification of brain function using PET*. Academic Press, London.
2. Frackowiak, R.S.J., Friston, K.J., Frith, C.D., *et al.* (1997). *Human brain function.* Academic Press, San Diego, CA.
3. Gunn, R.N., Gunn S.R. and Cunningham, V.J. (2001). Positron emission tomography compartmental models. *Journal of Cerebral Blood Flow and Metabolism,* **6**, 635–52.
4. Fischer, P.M., Meltzer, C.C., Ziolko, S.K., *et al.* (2006). Capacity for 5-HT1A-mediated autoregulation predicts amygdala reactivity. *Nature Neuroscience,* **11**, 1362–3.
5. David, S.P., Murthy, V.M., Rabiner, E.N., *et al.* (2005). A functional genetic variation of the serotonin (5-HT) transporter affects 5-HT1A receptor binding in humans. *Journal of Neuroscience*, **10**, 2586–90.
6. Wong, D.F., Wagner, H.N., Tune, L.E., *et al.* (1986). Positron emission tomography reveals elevated D2-dopamine receptors in drug-naïve schizophrenia. *Science*, **234**, 1558–63.
7. Farde, L., Wiesel, F.-A., Stone-Elander, S., *et al.* (1990). D2 dopamine receptors in neuroleptic naive schizophrenic patients. *Archives of General Psychiatry*, **47**, 213–19.
8. Sedvall, G. and Farde, L. (1995). Chemical brain anatomy in schizophrenia. *Lancet*, **346**, 743–9.
9. Hietala, J., Syvalahti, E., Vuorio, K., *et al.* (1995). Presynaptic dopamine function in striatum of neuroleptic naive schizophrenic patients. *Lancet*, **346**, 1130–1.
10. Guillin, O., Abi-Dargham, A. and Laruelle, M. (2007). Neurobiology of Dopamine in Schizophrenia. *International Review of Neurobiology*, **78**, 1–39.
11. Volkow, N.D., Wang, G.-J., Fowler, J.S., *et al.* (1994). Imaging endogenous dopamine competition with 11C-raclopride in the human brain. *Synapse*, **16**, 255–62
12. Laruelle, M., Abi-Dargham, A., van Dyck, C.H., *et al.* (1996). Single photon emission computerized tomography imaging of amphetamineinduced dopamine release in drug-free schizophrenia subjects. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 9235–40.
13. Weinberger, D.R. and Berman, K.F. (1996). Prefrontal function in schizophrenia: confounds and controversies. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **351**, 1495–503.
14. Chuam, S.E. and McKenna, P.J. (1996). Schizophrenia—a brain disease? A critical review of structural and functional cerebral abnormality in the disorder. *British Journal of Psychiatry*, **166**, 563–82.
15. Frith, C.D., Friston, K.J., Herold, S., *et al.* (1995). Regional brain activity in chronic schizophrenic patients during the performance of a verbal fluency task. *British Journal of Psychiatry*, **167**, 343–9.
16. Liddle, P.F., Friston, K.J., Frith, C.D., *et al.* (1992). Patterns of cerebral blood flow in schizophrenia. *British Journal of Psychiatry*, **160**, 179–86.
17. Dolan, R.J., Bench, C.J., Liddle, P.F., *et al.* (1993). Dorsolateral prefrontal cortex dysfunction in the major psychoses; symptom or disease specificity? *Journal of Neurology, Neurosurgery and Psychiatry*, **56**, 1290–4.
18. Andreasen, N.C., O'Leary, D.S., Flaum, M., *et al.* (1997). Hypofrontality in schizophrenia: distributed dysfunctional circuits in neurolepticnaive patients. *Lancet*, **349**, 1730–4.
19. Sabri, O., Erkwoh, R., Schreckenberger, M., *et al.* (1997). Correlation of positive symptoms exclusively to hyperperfusion or hypoperfusion of cerebral cortex in never-treated schizophrenics. *Lancet*, **349**, 1735–9.
20. Spence, S.A., Hirsch, S.R., Brooks, D.J., *et al.* (1998). Prefrontal cortex activity in people with schizophrenia and control subjects. Evidence from positron emission tomography for remission of 'hypofrontality' with recovery from acute schizophrenia. *British Journal of Psychiatry*, **172**, 316–23.
21. Callicott, J.H., Mattay, V.S., Verchinchi, B.A., *et al.* (2003). Complexity of prefrontal cortical dysfunction in schizophrenia: more than up or down. *American Journal of Psychiatry*, **12**, 2209–15.
22. Pike, V., McCarron, J.A., Lammerstma, A.A., *et al.* (1995). First delineation of 5-HT1A receptors in the living human brain using (11)CWAY-100635. *European Journal of Pharmacology*, **283**, R1–R3.
23. Sargent, P.A., Kjaer, K.H., Bench, C.J., *et al.* (2000). Brain serotonin1A receptor binding measured by positron emission tomography with [11C]WAY-100635: effects of depression and antidepressant treatment. *Archives of General Psychiatry*, **2**, 174–80.
24. Pike, V. (1993). Positron emitting radioligands for studies *in vivo*—probes for human psychopharmacology. *Journal of Psychopharmacology*, **7**, 139–58.
25. Meyer, J. (2007). Imaging the serotonin transporter during major depressive disorder and antidepressant treatment. *Journal Psychiatry and Neuroscience*, **2**, 86–102.
26. Goodwin, G. (1996). Functional imaging, affective disorder and dementia. *British Medical Bulletin*, **52**.
27. Drevets, W.C., Price, J.L., Simpson, J.R., *et al.* (1997). Subgenual prefrontal cortex abnormalities in mood disorders. *Nature*, **386**, 824–7.
28. Grasby, P.M. and Bench, C.J. (1997). Neuroimaging of mood disorders. *Current Opinion in Psychiatry*, **10**, 73–8.
29. Farde, L., Nordstrom, A.-L., Wiesel, F.A., *et al.* (1992). Positron emission tomographic analysis of central D1 and D2 dopamine receptor occupancy in patients treated with classical neuroleptics and clozapine. Relation to extrapyramidal side effects. *Archives of General Psychiatry*, **49**, 538–54.
30. Nyberg, S., Farde, L., and Erikson, L. (1993). 5-HT2 and D2 dopamine receptor occupancy in the living human brain: a PET study with risperidone. *Psychopharmacology*, **110**, 265–72.

31. Bantick, R.A., Deacon, J.F., and Grasby, P.M. (2001) The 5-HT1A receptor in schizophrenia: a promising target for novel atypical neuroleptics? *Journal of Psychopharmacology*, **1**, 37–46.
32. Bressan, R.A., Erlandsson, K., Jones, H.M. *et al.* (2003) Optimizing limbic selective D2/D3 receptor occupancy by risperidone: a [123I]-epidepride SPET study. *Journal of Clinical Psychopharmacology*, **1**, 5–14.

2.3.7 Structural magnetic resonance imaging

J. Suckling and E. T. Bullmore

Introduction

Magnetic resonance imaging (**MRI**) is a versatile and evolving technology for visualizing the structure, function, and metabolism of the living human brain. All kinds of MRI data can be acquired without exposing subjects to ionizing radiation or radioactive isotopes. Installing the hardware for MRI represents a major capital investment, of approximately £1.5 million. For these three reasons of versatility, safety, and (relative) affordability, MRI continues to be the dominant brain-imaging technique in psychiatric practice and research.

In this chapter, we introduce the principles and practicalities of MRI and describe common methods of structural MRI data acquisition and analysis. Chapter 2.3.8 on functional MRI provides greater detail on statistical issues arising in image analysis.

Magnetization

If iron filings are scattered on a piece of paper they will be oriented at random. If a bar magnet is then placed under the paper, the iron filings will align themselves so that each filing lies parallel to the magnetic field produced by the magnet. More technically, we can say that iron filings have a susceptibility to be magnetized by a static magnetic field.

Susceptibility refers both to the effect of a magnetic field on an object, and the effect of that object on the field. Paramagnetic materials, like some metals, tend to be attracted by magnets and cause a local increase in the magnetic field strength. Diamagnetic materials, like carbon and many organic compounds, tend to be repulsed by magnets and cause a local decrease in field strength.

The brain also has a susceptibility to be magnetized. It is largely composed of water and each molecule of water comprises, of course, two hydrogen atoms and one oxygen atom. The hydrogen nucleus is a single positively charged proton, which has a dynamic property called spin. Like all moving charged particles, spinning protons generate a magnetic field. The axis of the magnetic dipole generated by a spinning proton is sometimes called its magnetic moment, and is drawn as a vector.

When the brain is placed in a strong magnetic field, the spinning protons align themselves with the external field, just as iron filings align themselves to the field of a bar magnet. The angle of alignment between each proton's moment and the (longitudinal) axis of the external magnetic field is α. Protons obey the laws of quantum mechanics, and so two modes of alignment or spin states are possible, one with the magnetic moment in the direction of the field ($\alpha = 0°$) and one with the moment in the opposite direction ($\alpha = 180°$). Depending on the strength of the applied field, the spin states have slightly different probabilities, with those protons aligned in the direction of the field in excess by about 5 ppm at an external field strength of 1.5 T. (Magnetic field strength is measured in units of gauss (**G**) or tesla (**T**): 1 T = 10 000 G. The earth's magnetic field is approximately 0.5 G; a child's toy magnet has a field of around 10 G.)

Thus, if the magnetic moments for all spinning protons are averaged, the net, or bulk, magnetization vector for the brain as a whole will have $\alpha = 0°$. The length of the net magnetization vector then represents the strength of longitudinal magnetization (Fig. 2.3.7.1).

Protons aligned with a static magnetic field are not static themselves, they rotate or precess at very high frequency around the axis of the external field. The precession frequency, or Larmor frequency, is constant for a given type of atomic nucleus and external field strength. For protons, the Larmor frequency at 1.5 T is 63.9 MHz. However, although all hydrogen nuclei in the brain precess at the same frequency in the same field, they will not all precess with the same phase. At any given time, different nuclei have reached a different point in their rotation around the external field axis.

Nuclear magnetic resonance

When a wineglass is tapped by a knife, it produces a high-pitched sound of characteristic frequency. If a singer can exactly match that frequency with her voice then the glass will resonate and may break. The basic idea is that if an object has a characteristic frequency of oscillation, exposing it to energy precisely at that frequency will cause a change in physical state.

Analogously, if we supply a pulse of radio-frequency energy at (and only at) the Larmor frequency to a brain located in a magnetic field, the protons within the brain will absorb the energy and resonate—this is nuclear magnetic resonance (**NMR**)—and their angle of alignment α with the external field will increase. If sufficient energy is supplied to cause $\alpha = 90°$, the radio-frequency pulse is called a '90° pulse'. If the net magnetization vector is flipped to an angle $\alpha = 180°$, the radio-frequency pulse is called a '180° pulse'. At the same time as the angle of alignment is increased by radio-frequency irradiation, the phase of precession becomes coherent over all protons. In other words, in place of the random variation in the phase of precession that existed before the radio-frequency pulse, protons are now 'marching in step' with each other around the axis of the external field.

After the radio-frequency pulse has ceased, the resonating nuclei gradually relax back to the equilibrium state of random precession in alignment with the external field. The two components of this relaxation process are characterized by relaxation times. The first relaxation time (T_1), also called the spin-lattice relaxation time, describes the time taken for the strength of longitudinal magnetization to return to 63 per cent of its value before radio-frequency irradiation. This is a measure of the time taken for α to return to zero having been flipped to 90° or 180°. T_1 is determined by interactions between protons and their long-range (molecular) environment

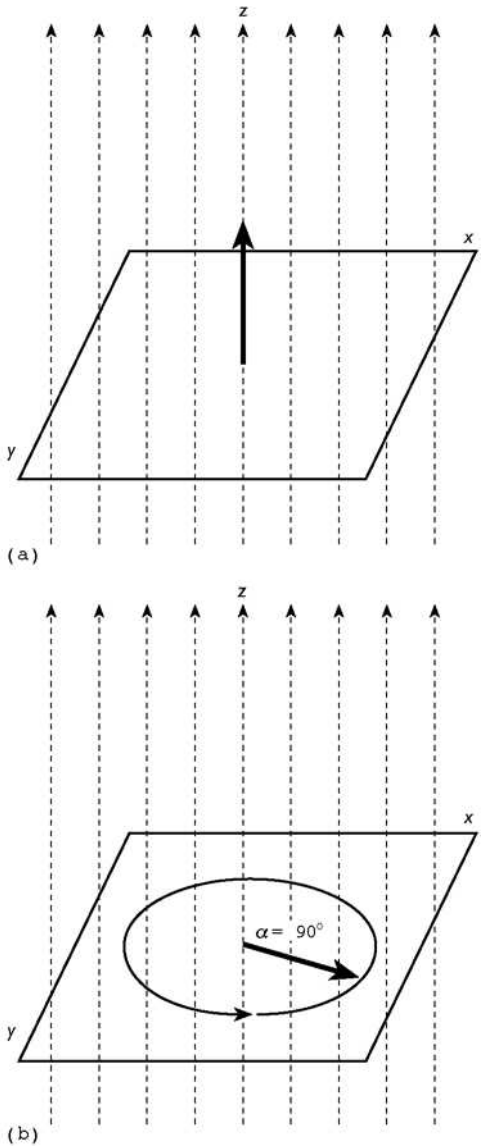


Fig. 2.3.7.1 Net magnetization vector.
(a) In a static magnetic field, the vector is aligned parallel to the longitudinal **z** axis of the field and **α** = 0.
(b) Immediately after a 90° excitation pulse of radio-frequency energy at Larmor frequency, the angle of alignment **α** is increased (transverse magnetization) and the phase of precession in the **x–y** plane is coherent over all protons in the brain. As protons relax following excitation, the angle of the net vector becomes smaller (return of longitudinal magnetization) and the phase of precession becomes more variable from one proton to another (dephasing).

or lattice. The second relaxation time (T_2), also called the spin–spin relaxation time, describes the time taken for the flipped nuclei to stop 'marching in step' around the axis of the field. This process of dephasing begins as soon as the radio-frequency pulse stops, but its rate is determined by the immediate (atomic) environment of the protons. Small variations in the applied magnetic field accentuate spin–spin relaxation, resulting in an observed relaxation time T_2^* which is somewhat faster than the 'true' relaxation time T_2 that would have been observed in an ideally homogeneous field.

As protons relax, they release the energy absorbed from the radio-frequency pulse in the form of a weak radio-frequency signal, which decays at a rate normally determined by T_2^*. This process is called free induction decay, and the emitted signal forms the data from which magnetic resonance images are ultimately constructed.

Magnetic resonance imaging

Spin echo sequence

A widely used MRI technique is the spin echo sequence. A 90° radio-frequency pulse is repetitively applied to the brain with a constant repetition time (TR ms) between consecutive pulses. Following each 90° pulse, protons are excited and then relax. The dephasing component of relaxation can be reversed by applying a second 180° pulse some time (TE/2 ms) after the 90° pulse. Following the first 90° pulse, protons immediately begin to precess idiosyncratically and the emitted signal decays. By reversing this process, the 180° pulse causes rephasing and an increase in emitted signal which has a maximum or echo at TE ms (time to echo) after the initial 90° pulse (Fig. 2.3.7.2).

The spatial location of the radio-frequency signal emitted by free induction decay in a given volume of the brain is encoded in three spatial dimensions by slice-selective radio-frequency irradiation combined with frequency- and phase-encoding gradients. To improve scan time, multiple slices can be excited in an interleaved fashion (multislice acquisition). This means that after the output signal is detected from one slice, and while the net magnetization vector is relaxing back to its equilibrium state, other slices can be excited. The in-plane resolution (voxel size) of the image is determined by the field of view and the number of voxels in the image. Typically, in-plane resolution at 1.5 T is in the order of 0.5 to 2 mm, and slice thickness is 2 mm or more.

Tissue contrast

The outstanding advantage of MRI for the anatomical examination of the brain is the easily visible contrast in the images between grey matter, white matter, and cerebrospinal fluid. In particular, contrast between parenchymal tissues (grey and white matter) has made MRI the imaging research tool of choice for identifying subtle cortical abnormalities in a wide variety of psychiatric disorders.

Tissue contrast in magnetic resonance images is determined by differences in the density of protons, and their physical and chemical environment. A tissue such as the cerebrospinal fluid, that is

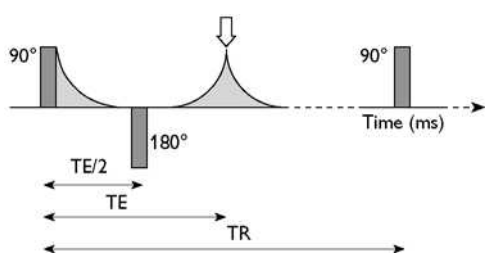


Fig. 2.3.7.2 Spin echo pulse sequence. A 90° excitation pulse of radio-frequency energy is immediately followed by exponential decay of the T_2-weighted signal. A 180° pulse TE/2 ms later causes rephasing of proton spins and an exponential increase in T_2-weighted signal with maximum (echo) TE ms after the 90° pulse. Images are acquired at TE (thick arrow). The protons are allowed to relax completely before the next 90° pulse, TR ms after the previous excitation.

composed largely of water, will have a lower proton density than parenchymal brain tissues. The physicochemical environment of protons has a marked effect on spin-lattice relaxation. If protons are mainly in freely diffusing water molecules, as they are in cerebrospinal fluid, T_1 will be prolonged, whereas if they are mainly bound to large macromolecules, as they are in fat, T_1 will be short (Table 2.3.7.1). Since grey matter contains proportionally less fat than myelinated white matter, T_1 is longer for grey matter. Spin–spin relaxation is likewise determined, in part, by the immediate physical environment of protons in the tissue; liquid tissues will have prolonged T_2 times compared with solid tissues. Other effects on apparent relaxation times (T_2^*) include minute fluctuations or inhomogeneities in the strength of the external magnetic field, which may be due to the local paramagnetic effects of iron-containing compounds such as haemoglobin.

The parameters of the spin echo pulse sequence, repetition time (TR), and time to echo (TE), can be judiciously adjusted to acquire images that are sensitive to or weighted by one or other of these possible sources of tissue contrast.

If TR is long (>1000 ms) and TE is short (<20 ms), contrast in the images will be weighted by differences between tissues in proton density. Proton-density-weighted images show good contrast between relatively hyperintense parenchymal tissue and hypointense cerebrospinal fluid (Plate 7).

If TR is short (<1000 ms) and TE is also short (<20 ms), contrast in the images will be weighted by tissue differences in spin-lattice relaxation. T_1-weighted images show excellent contrast between hyperintense white matter and relatively hypointense grey matter (Plate 7). For this reason, T_1-weighted images are widely used to measure quantitative abnormalities in size or shape of the cerebral cortex.

If TR is long (>1000 ms) and TE is also long (>20 ms), contrast in the images will be weighted by tissue differences in spin–spin relaxation. T_2-weighted images show strong contrast between hyperintense cerebrospinal fluid and parenchymal tissues (Plate 7), unless there is congestion or oedema of the parenchyma, in which case the T_2-weighted signal will be increased. For this reason, T_2-weighted images are widely used to identify acute, inflammatory, and ischaemic lesions.

Structural imaging sequences

An enormous range of sequences are available for brain imaging. For example, fast spin echo imaging provides a pair of dual-echo images with complementary tissue contrasts (proton density and T_2-weighted) for no increase in scan time over the spin-echo sequence. Spins can also be manipulated through changes in the applied magnetic field gradients. These sequences yield images of high-spatial resolution and a range of contrasts. Additionally, there are methods for suppressing signals due to blood flow or fat, improving signals from pathology.

Table 2.3.7.1 Relaxation times at 1.5 T for different tissue types

Tissue type	T_1 (ms)	T_2 (ms)
Grey matter	980–1040	64–71
White matter	740–770	64–70
CSF	>2000	>300
Fat (at 1 T)	180	90

CSF, cerebrospinal fluid.

Diffusion-weighted imaging

Diffusion-weighted imaging can provide information about the organization of white matter tracts in the brain that cannot be obtained by other MRI methods.

The basic idea is that protons move within and between cells by random motion. Typically, a proton may travel around 20 μm in 100 ms by this Brownian motion or diffusion. The rate of proton diffusion is related to how constrained they are by physical barriers such as myelinated cell membranes. The rate of diffusion affects the spin–spin relaxation time, with rapidly diffusing protons tending to relax more quickly. To acquire images that are weighted by differences in diffusion, two extra gradients are briefly applied during a spin echo sequence.[1]

White matter is generally hyperintense in diffusion-weighted imaging because closely packed axonal tracts provide the greatest barrier to the free diffusion of water in the brain. Furthermore, it is possible to deduce from diffusion-weighted imaging data how compactly organized the white matter is, and even to estimate in what direction the fibre tracts are oriented (Plate 8). This information is of considerable interest to psychiatry, since the pathology of many psychiatric disorders may involve the axonal connections between multiple cortical areas.[2]

Safety

MRI is absolutely contraindicated in patients who have any strongly magnetized metal object in their heads. This includes aneurysm clips, reconstructive metal plates, traumatically embedded metal fragments, or implanted electronic devices such as cardiac pacemakers. It is advisable to screen all subjects undergoing MRI by questionnaire for possible contraindications. A skull radiograph is a useful preliminary examination if there is any doubt about the presence of intracranial metal. All subjects need to provide informed consent in writing.

Static magnetic fields used in MRI cause no harmful effects to biological tissue. Rapid switching of field gradients can induce electrical currents in tissue, but at the switching speeds used in MRI these induced currents are several times less than needed for muscle contraction. Since radio-frequency energy can cause heating, limits to the amounts of energy absorbed are set by national standards.

Artefacts

Quality assurance protocols and diligent hardware servicing are necessary to maintain high standards in MRI. Image artefacts refer to loss of image quality (i.e. spatial resolution or tissue contrast) owing to a specific cause. It is often possible to effect a remedy and an awareness of their causes can often pre-empt the problem.

Movement

Subject motion is the most common artefact and has two components: voluntary and involuntary (physiological). The result of voluntary motion during acquisition may be an obvious blurring (Plate 7). The cooperation of subjects is vital and special regard is

required for children, the elderly, and those with neurological or psychiatric disorders who may find the MRI environment disconcerting. The subject's head may be physically restrained by additional padding and by Velcro straps placed across the forehead. Involuntary motion arises primarily from the cardiorespiratory cycle causing pulsation of the blood vessels and cerebrospinal fluid. This is most apparent in structural images showing major arteries.

Susceptibility

Where two materials with very different susceptibilities are closely adjacent, there may be severe distortion of the magnetic field, causing artefactual loss or exaggeration of the magnetic resonance signal. This is clearly seen in an image acquired with a metallic clip placed close to the scalp (Plate 7). Ferromagnetic (highly paramagnetic) materials that may cause artefacts include metallic dental fillings and plates, hairgrips, ear or nose rings, and even some cosmetics. A less obvious form of susceptibility artefact arises if more (diamagnetic) tissue is situated at one end of the image field of view than the other. A field gradient is set-up across the image resulting in signal loss. This is termed bulk susceptibility artefact and is often seen in images of both the head and neck (Plate 7).

Partial volume

Voxel sizes are larger than some scales of anatomical organization in the brain. A voxel may represent a heterogeneous mixture of tissue classes, or be only partially occupied by tissue of a single type. This partial volume artefact is particularly evident at the interface between cortical grey matter and sulcal cerebrospinal fluid, and at the interface between cortical grey matter and central white matter. It causes error in the estimation of tissue class volumes.

MRI hardware

Superconducting magnet

The superconducting magnets used for MRI require liquid helium cooling equipment to keep the temperature low enough (4 K) for superconduction. Cooling consumes the majority of the supplied power. Only a small current is initially required to generate the field, which is then self-sustaining.

Small variations in the homogeneity of the magnetic field give rise to distortions and artefacts. The field is minutely adjusted to improve homogeneity using additional magnets in an automated procedure known as shimming.

Magnetic field gradients are essential to MRI. Rapid switching of gradient coils produces the loud 'knocking' sound associated with magnetic resonance scanners. Three orthogonal gradients are available which are coupled to generate gradients in any direction.

Radio-frequency coil

The function of the radio-frequency coil is two-fold—to transmit the radio-frequency pulses of the imaging sequences and to receive the emitted signal. Head coils used for neuroimaging fit snugly over the subject's head, and are often of a three-dimensional design with good sensitivity throughout the volume they enclose. Surface coil designs are used to image small regions with high-spatial resolution, and phased-array coils combine several surface coils for more extensive coverage.

Computers

MRI produces large quantities of data, which require rapid processing and storage so that images may be viewed and other sequences prescribed during the same session. The computer system is integral to the machine and contains specialized hardware and software for data acquisition and image reconstruction, as well as control of the scanner.

The scanner suite

The room housing the scanner must be specially designed. It must have a reinforced floor and be environmentally controlled to maintain a constant temperature and humidity. The walls and ceiling contain a magnetic shield, which both prevents leakage of the field outside the room and stops FM radio broadcasts from being picked up by the radio-frequency coil.

MRI studies

Case-control design

Structural MRI studies in psychiatry have commonly adopted a cross-sectional or case-control design. This involves scanning two groups of subjects, patients, and matched controls, on a single occasion. The objective is generally to identify anatomical differences in brain structure between cases and controls. In evaluating or planning such a study, it is important to pay attention to several design issues, some of which are summarized below.

Power

What is the power of the study to refute the null hypothesis (zero anatomical difference between the case and control populations) when it is not true? In general, the power of a study is proportional to the sample size (the number of subjects scanned), the effect size (the anatomical difference between populations), and the probability threshold or p value adopted for hypothesis testing. The p value will often be decided in relation to the number of tests conducted—the greater the number of tests, the smaller is the appropriate p value. Therefore, the risk of low power and associated type 2 error is likely to be greatest when differences between two small groups have been multiply tested on the basis of many anatomical variables.

Representativeness

What population is represented by the sample of patients studied, and is this the population of interest? If the ambition of the study is to make inferences about the population of patients with, say, manic-depressive disorder, then it is important that the diagnosis is made according to standard and reliable criteria and that the sampling procedure is such that any patient in that population has an equal chance of being included in the study. This means that the authors of the study will need to sample cases from general practice and the community as well as from hospital clinics and wards. Sampling hospital patients is generally much easier; the patients are already well characterized and hospital treatment facilities are likely to be relatively few in number and close to the scanning unit. However, if only hospital patients are sampled, it follows that inference can only be made about the population of hospital patients, rather than the larger and more general population of individuals with the disorder.

Heterogeneity

Diagnostic categories in psychiatry may subsume considerable heterogeneity in terms of phenomenology and aetiology. For example, patients with a diagnosis of schizophrenia may differ profoundly in terms of positive or negative symptom profiles, cognitive deficit, and genetic or environmental risk factors. These natural sources of heterogeneity may be compounded by differences in treatment. Any or all of these factors may affect brain structure. Studies that simply ignore heterogeneity, or attempt to deal with it by *post hoc* statistical correction, may have less power to detect a group difference than studies which define cases according to refined or subdiagnostic criteria. Thus, studying a sample of schizophrenic patients with high negative symptom scores and marked working memory deficits may be more likely to reveal anatomical abnormalities of frontal cortex than studying an unrefined sample of patients with schizophrenia.

Matching

Ideally, the control or comparison subjects should be indistinguishable from the patients in every characteristic apart from features of the disorder. For example, it is important that cases and controls should be matched for age, handedness, and sex, since all of these factors may affect brain structure. Unfortunately, there are a number of other possible confounding factors that are not so obviously unrelated to presence of the disorder. For example, an unrefined sample of patients with schizophrenia will generally have lower IQ and smaller head size than an age- and sex-matched group of comparison subjects. Should we try to correct these differences as if they were spurious (by either refined sampling or *post hoc* statistical modelling), or should we accept that they represent real features of the disorder? In practice, most published studies tend to correct group differences on global variables by statistical modelling in order to focus attention on regional differences that may be more interesting. A comparable problem arises in relation to medication.

Other designs

A probable future trend in psychiatric MRI research over the next few years is that the hypotheses under investigation will become more concerned with aetiological mechanisms and pathogenetic models and less concerned with the basic question of whether a given group of patients has abnormal brain structure. This shift in hypothetical interest will dictate a shift in design away from case-control or cross-sectional studies.

Longitudinal designs, in which a cohort of volunteers or patients are scanned repeatedly over a period of months or years, are a powerful way of demonstrating normal and abnormal developmental changes in brain structure.[3] They have the obvious disadvantage that they are time consuming to complete and subjects may not attend for multiply repeated examination.

Genetic designs involve subjects that are defined genotypically, rather than phenotypically. Imaging studies of monozygotic twins discordant for a disorder,[4] and of families multiply affected by a disorder, are examples of genetic designs in which there may be complete knowledge about the proportion of genetic information that is shared between subjects but incomplete knowledge about the genetic constitution of each subject.

Overall, there is no single perfect design for an imaging study. Often designing a study will entail finding pragmatic and arguably justifiable solutions to problems. Furthermore, the 'goodness' of a design can really only be judged in relation to the hypothesis under investigation and the methods of analysis applied to the data. The best imaging studies will convey a sense that the design is both ingenious and inevitable, given the hypothesis and available methods, and will also include a frank discussion of the limitations or implications of the particular design adopted.

Data analysis

Clinical analysis

Structural MRI is most often used in clinical practice to exclude non-psychiatric causes for psychopathology. Clinical examination of these cases may also sometimes reveal abnormalities such as hippocampal sclerosis or callosal agenesis which suggest that psychopathology has been determined by birth injury or abnormal development. In assessment of a patient with dementia, MRI may usefully demonstrate signs of vascular disease (such as infarcts or periventricular white matter changes), or a focal pattern of grey matter atrophy suggestive of Pick's disease (frontal cortex) or Huntington's disease (caudate nucleus and frontal cortex). All of these abnormalities may be detected simply by skilled visual examination of the data. However, clinical diagnosis of the subtler abnormalities associated with, say, schizophrenia, obsessive–compulsive disorder, or autism require quantitative analysis of the patient's data and access to normative MRI measurements on appropriately matched samples of the general population. Neither quantitative analysis nor normative databases are widely used in current radiological practice, thus limiting the value of MRI to clinical psychiatry, but this may be expected to change in the future.

Quantitative analysis

There are broadly two requirements for quantitative analysis of structural MRI data. The first is to measure the anatomical structure of the brain (this is often called morphometry). The second is to test hypotheses of interest on the basis of these morphometric variables.[5] Here we shall focus on morphometry.

(a) Morphometry

A widely used method of measuring brain structure from magnetic resonance images is based on the hypothetical expectation that one or more anatomically defined regions of the brain are particularly relevant to the disorder.[6] The area or volume of each of these regions of interest (**ROI**s) can then be measured directly by drawing a line around the region on a computerized display of the data and counting the number of voxels enclosed by the line. Measurements of several ROIs may be combined to produce summary measures of asymmetry[7] or of spatially distributed anatomical systems. The advantages of ROI morphometry are that it is conceptually simple and that it allows measurement of structures (e.g. the hippocampus) or parts of structures (e.g. segments of the corpus callosum) that may be difficult to measure otherwise. Some familiar disadvantages are that it is time consuming and imperfectly reliable. More fundamentally, ROI morphometry is of limited use if it is not obvious in advance what the region of interest is, or there are several possible regions of abnormality (as is likely if the disorder is determined by an insult early in the course of brain development).

Computerized techniques are varied, but commonly adopt an approach summarized in Plate 9. In this scheme, the first step in image processing is removal of extracerebral tissues like skull and scalp leaving an image of the brain alone. The brain image is then segmented into the three main tissue classes. There are a variety of techniques for segmentation or brain tissue classification,[8] but generally the quality of segmentation is improved if multiple images of different contrasts (e.g. proton density, T_2-weighted) are available.

Once the brain image has been divided into its component classes, or even prior to segmentation, it can be automatically registered in a standard anatomical space. The value of making this transformation is potentially two-fold. First, it then becomes possible to compare brain structure between individuals at the spatial resolution of the image, that is in terms of the percentage occupancy of each voxel by each tissue class. This means that one is able ultimately to test anatomical differences between groups over the whole brain in detail, without having to assume a priori that pathological change is located in a particular region.[9,10] The second major advantage of image registration is that the parameters of the transformation used to align the image with the template may themselves be used as measures of brain structure. Thus, if an image represents a brain that is structurally abnormal in some way, the mathematical deformation which must be applied to align it with a standard template may be abnormally great.[11]

(b) Statistical testing

Morphometric variables can be used to test hypotheses by a variety of statistical analyses. For example, the null hypothesis of zero difference in brain structure between two groups can be addressed by a *t*-test of a single ROI or by many thousands of *t*-tests of, say, grey matter occupancy measured at each and every voxel. Similarly hypotheses concerning the relationship between brain structure and psychological function can be addressed by testing the correlation between a morphometric variable and the subjects' scores on a psychometric instrument. Alternatively, the relationships between several morphometric variables may be explored by multivariate methods such as partial least squares.[12]

Further information

MRI theory and practice—McRobbie, D.W., Moore, E.A., Graves, M.J., *et al.* (2007). *MRI from picture to proton* (2nd edn). Cambridge University Press, Cambridge.

Reviews of structural imaging in psychiatry—Symms, M., Jager, H.R., Schmierer, K., *et al.* (2004). A review of structural magnetic resonance neuroimaging. *Journal of Neurology, Neurosurgery, and Psychiatry*, **75**, 1235–44.

Hyman, S.E. (2007). Can neuroscience be integrated into the DSM-V? *Nature Reviews Neuroscience*, **8**, 725–32.

Online repository of methods and data—The Internet Brain Segmentation Repository (http://www.cma.mgh.harvard.edu/ibsr/) provides manually guided expert segmentation results along with magnetic resonance brain image data.

References

1. Basser, P.J. and Jones, D.K. (2002). Diffusion tensor MRI: theory, experimental design and data analysis—a technical review. *NMR in Biomedicine*, **15**, 456–67.
2. Taylor, W.D., Hsu, E., Krishnan, K.R.R., *et al.* (2004). Diffusion tensor imaging: background, potential and utility in psychiatric research. *Biological Psychiatry*, **55**, 201–7.
3. Giedd, J.N., Blumenthal, J., Jeffries, N.O., *et al.* (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neuroscience*, **2**, 861–3.
4. McDonald, C., Bullmore, E.T., Sham, P.C., *et al.* (2004). Association of genetic risks for schizophrenia and bipolar disorder with specific and generic brain structural endophenotypes. *Archives of General Psychiatry*, **61**, 974–84.
5. Bullmore, E.T., Suckling, J., Overmeyer, S., *et al.* (1999). Global, voxel and cluster tests, by theory and permutation, for a difference between two groups of structural MR images. *IEEE Transactions on Medical Imaging*, **18**, 32–42.
6. Wright, I.C., Rabe-Hesketh, S., Woodruff, P.W.R., *et al.* (2000). *The American Journal of Psychiatry*, **157**, 16–25.
7. Shenton, M.E., Dickey, C.C., Frumin, M., *et al.* (2001). A review of MRI findings in schizophrenia. *Schizophrenia Research*, **49**, 1–52.
8. Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximisation algorithm. *IEEE Transactions on Medical Imaging*, **20**, 45–57.
9. Ashburner, J. and Friston, K.J. (2000). Voxel-based morphometry—the methods. *NeuroImage*, **11**, 805–21.
10. Ananth, H., Popescu, I., Critchley, H.D., *et al.* (2002). Cortical and subcortical gray matter abnormalities in schizophrenia determined through structural magnetic resonance imaging with optimized voxel-based morphometry. *The American Journal of Psychiatry*, **159**, 1497–505.
11. Chung, M.K., Worsley, K.J., Robbins, S., *et al.* (2003). Deformation-based surface morphometry applied to gray matter deformation. *NeuroImage*, **18**, 198–213.
12. Nestor, P.G., O'Donnell, B.F., McCarley, R.W., *et al.* (2002). A new statistical method for testing hypotheses of neuropsychological/MRI relationships in schizophrenia: partial least squares analysis. *Schizophrenia Research*, **53**, 57–66.

2.3.8 Functional magnetic resonance imaging

E. T. Bullmore and J. Suckling

Introduction

Functional magnetic resonance imaging (**fMRI**) is a relatively new technique for measuring changes in cerebral blood flow. The first fMRI studies, showing functional activation of the occipital cortex by visual stimulation and activation of the motor cortex by finger movement, were published in the early 1990s.[1–3] In the years since then, fMRI has been used to investigate the physiological response to a wide variety of experimental procedures in both normal human subjects and diverse patient groups. In the next 10 years, fMRI will probably establish a role for itself in radiological and psychiatric practice; currently the clinical role of fMRI is limited to specialized applications such as assessment of hemispheric dominance prior to neurosurgery.[4]

The outstanding advantage of fMRI over alternative methods of imaging cerebral blood flow, such as positron emission tomography (**PET**) and single-photon emission computed tomography (**SPECT**), is that it does not involve exposure to radioactivity. This means that a single subject can safely be examined by fMRI on

many occasions, and that the ethical problems of examining patients are minimized. Functional MRI also has superior spatial resolution (in the order of millimetres) and temporal resolution (in the order of seconds) compared with PET and SPECT.

In this chapter, we provide an introduction to technical issues relevant to fMRI data acquisition, study design, and analysis. An introduction to the basic physical principles of magnetization and nuclear magnetic resonance, and the technology, is given in Chapter 2.3.7.

Many excellent specialist texts covering all aspects of functional magnetic resonance imaging are available for the reader seeking more detailed treatment of the issues.[5–7]

Cerebral activation and blood-flow changes

Try closing your eyes and then opening them again. At the moment that you open your eyes, neurones in the occipital cortex that are specialized for the perception of visual stimuli will show a sudden and dramatic increase in their rate of discharge. There is a short delay (approximately 100 ms) between the stimulus and neural response owing to the propagation of electrical activity from the retina via the optic nerves and tracts to the visual cortex. Later, some 3 to 8 s after stimulus onset, there will be an accompanying change in the local blood supply to the stimulated area of cortex. Blood flow increases without a commensurate increase in oxygen uptake by the visual cortex, leading to a local increase in the ratio of oxygenated to deoxygenated forms of haemoglobin.

The linkage between neural activity and regional cerebral blood flow, sometimes called neurovascular coupling, has been known since Roy and Sherrington first reported 'changes in blood supply in accordance with local variations of functional activity' in 1894. However, the biophysical and biochemical mechanisms for neurovascular coupling are complex and not yet completely defined in detail.[7]

Endogenous contrast agents

The fact that neural activity is linked to local blood flow provides the opportunity for functional MRI. The most common, and non-invasive, approach exploits the paramagnetic properties of iron in deoxygenated haemoglobin as an endogenous contrast agent. Neural activity causes a local reduction in the ratio of deoxygenated to oxygenated haemoglobin, so that the paramagnetic effects of deoxyhaemoglobin are 'diluted'. Since apparent spin–spin relaxation or dephasing is accelerated by microscopic inhomogeneities in the magnetic field due to the presence of paramagnetic contrast agents, the net effect of diluting deoxyhaemoglobin will be to prolong T_2^* times in areas of the brain that receive an increased blood flow as a consequence of neural activity. The haemodynamic effect on spin–spin relaxation can be measured by a T_2^*-weighted signal change (of 3 per cent or less) which is blood oxygen level dependent (**BOLD**).

Imaging sequences for fMRI

Several different pulse sequences can be used to collect MRI data that are sensitive to functionally determined changes in signal strength. Here we will concentrate on gradient echo sequences which, combined with special techniques for very rapid data acquisition, are most widely used to date for fMRI. However, spin echo sequences can also be used for functional MRI data acquisition, and gradient echo sequences can be used for structural MRI.

Gradient echo sequence

The basic principle is similar to spin echo imaging. An initial excitation pulse of radiofrequency energy is supplied at the Larmor frequency to the brain in the presence of a powerful static magnetic field. Protons are excited to a state characterized by increased transverse magnetization and a coherent phase of precession around the axis of the external field. Immediately the radiofrequency pulse has ceased, protons begin to relax back to their equilibrium state of maximum longitudinal magnetization and random phase of precession, emitting a radiofrequency signal by free induction decay.

In gradient echo imaging, the process of spin–spin relaxation (dephasing) is first accelerated by briefly applying a gradient to the magnetic field shortly after the excitation pulse. Then, at some time (TE/2) after the excitation pulse, a second gradient is applied to reverse the process of dephasing, causing rephasing and a signal maximum or echo some time (TE) after excitation. The sequence is repetitively applied with a constant time interval between consecutive excitations (TR).

The objective is to manipulate spin–spin relaxation by brief perturbations of the external magnetic field rather than by supplying additional pulses of radiofrequency energy as in spin echo imaging. Frequency- and phase-encoding gradients are applied to locate the sources of signal in three-dimensional space (see Chapter 2.3.7).

One advantage of gradient echo imaging is that TE and TR can both be shorter than in spin echo imaging, allowing an overall reduction in scanning time. However, if TR is short, spoiler gradients or radiofrequency pulses may be needed to ensure that the protons have returned to equilibrium before the next excitation pulse is supplied. The flip angle α induced by radiofrequency excitation can be adjusted to generate images weighted by different sources of tissue contrast. T_1-weighted images are generated by radiofrequency pulses causing flip angles of the order of 10° to 20°. For functionally sensitive T_2-weighted images, more radiofrequency energy must be supplied in the excitation pulse to give a flip angle approaching 90°.

Echoplanar imaging

The gradient echo sequence equivalent to a fast spin echo sequence is obtained by rapidly applying, or blipping, a series of rephasing gradients following the excitation pulse and dephasing gradient. Gradient blipping is done extremely rapidly (<1 ms), and up to 128 echoes can be generated from a single excitation. Clearly, the advantage of such echoplanar imaging is the speed of acquisition. Multislice images of the entire cortex, with slice thickness of only a few millimetres, can be acquired in 2 s or less. Such high-speed imaging is highly desirable for functional MRI, where we wish to detect physiologically determined changes in magnetic resonance signal with the best possible temporal resolution. However, the hardware required for rapid gradient blipping has only become widely available in the last few years.

Artefacts

The main sources of artefact in functional MRI are the same as for structural MRI (see Chapter 2.3.7).

Movement

Movement of the subject's head during fMRI data acquisition is inevitable, and attempts to eliminate it by fixing the head in the scanner may paradoxically exacerbate the problem. The best approach of minimizing movement is to ensure that the subjects are not unduly anxious about the scanning procedure, that they understand clearly what they are being asked to do, and that they are comfortable in the scanner before data acquisition begins. Experiments should be designed so that they do not require the subject to move extensively; small finger movements required for button pressing do not generally cause severe head movement. However, even very small movements of the head (less than 1 mm) can cause significant artefacts in fMRI data.

Involuntary or physiological movements are mostly due to the cardiorespiratory cycle causing pulsation of the cerebrospinal fluid and vascular spaces. Therefore, these movements often occur at a higher frequency than the frequency of image volume acquisition, and are aliased into the signal as a low frequency confound.

Susceptibility

The susceptibility artefact is exaggerated by gradient echoplanar imaging, typically causing signal loss in inferior temporal and orbitofrontal brain regions close to bone or sinuses. The problem is further compounded if subjects are asked to speak during scanning, since slight deformations of the sinuses associated with overt articulation can cause changes in susceptibility artefact, which can mimic signal changes due to speech-related neural activity. If overt articulation is necessary to monitor the subject's performance on the experimental task, then it is advisable to design the sequence so that images are not acquired while the subject is speaking.

Hardware

The hardware requirements for functional MRI include a superconducting magnet, a radiofrequency coil, computers, and a purpose-built room, as described in Chapter 2.3.7.

Gradient coils

The essential extra prerequisite is gradient coils capable of very rapidly blipping the external magnetic field for echoplanar imaging. The gradients required are small compared with the external field (1 to 10 mT/m) but may need to be applied for less than 1 ms. The speed with which the gradient is switched on, the slew rate, is necessarily fast (up to 200 mT/m/s). Such rapidly changing gradients can cause eddy currents in the gradient coils, adversely affecting the homogeneity of the field. This problem is minimized by actively shielding the coils, which requires yet another set of coils.

Audiovisual equipment

It must be possible to present visual and auditory stimuli to subjects while they are lying with their heads in the bore of the magnet. Headphones are required to clearly present auditory stimuli in the presence of the loud background noise of gradient switching. Visual stimuli can be projected on to a screen, viewed through a periscope. The subject should have access to an alert button. In addition, it is generally useful for subjects to use a button-press device to indicate their response to cognitively demanding experimental tasks. These behavioural data will need to be monitored during scanning.

Experimental design

The basic principle of experimental design in fMRI is to manipulate the subject's experience or behaviour in some way that is likely to produce a functionally specific neurovascular response. It is usually important that the experiment should be designed to allow some other measure of response, for instance a button press, to be monitored simultaneously. We shall illustrate these and other principles by considering how one might design an experiment to identify the regions of the brain that are important in making a (semantic) decision about the meaning of words. As will become clear, no single design is ideal; each has its strengths and weaknesses, and the choice between them should also be considered carefully in the light of the particular hypothesis one is using fMRI to test.

Blocked periodic design

This experimental design, in its simplest form, involves alternately presenting the subject with two conditions: an activation (A) condition and a baseline (B) condition. Each condition is presented for an identical epoch of time. During each epoch, several stimuli are sequentially presented with an interstimulus interval (**ISI**) that is less than the epoch length. The cycle of alternation between A and B conditions is repeated a number of times over the course of each experiment.

For example, during each 30-s epoch of the A condition, we could visually present subjects with a series of 12 common concrete nouns (ISI = 2.5 s) and ask them to decide for each word whether it refers to an animate object (e.g. 'goat') or an inanimate object (e.g. 'bucket'). The subjects could be asked to indicate their decision by pressing one of two buttons: left for living objects and right for non-living objects. During each 30-s epoch of the B condition, we could present words at the same rate, but ask subjects to decide whether they are written in upper- or lowercase letters. This decision could be monitored by button press as in the A condition. The two epochs could be presented alternately, beginning with the B condition, in five cycles for a total experimental time of 5 min. Functional MRI data would be acquired continuously throughout the experiment (Fig. 2.3.8.1).

The rationale for this design is that the two conditions are matched in all respects apart from semantic analysis; words are visually presented at an identical rate, and subjects are asked to signal their decision by an identical device. But while condition A demands semantic analysis of the words (what do they mean?), condition B demands only orthographic analysis (what do they look like?). We assume that only those regions of the brain that are specifically responsible for semantic analysis will show an increased magnetic resonance signal during condition A; those regions responsible for visual perception and motor output will be activated identically under both conditions, and so will not demonstrate a periodic signal change at the frequency of AB alternation. This set of assumptions is sometimes referred to as cognitive subtraction.

Blocked periodic designs can generate robust signal changes in fMRI, as long as the two conditions are not too closely matched. The drawbacks are that it is impossible to assess the response to a single stimulus and the critical assumption of cognitive subtraction may not always be valid. Sometimes two experimental tasks that appear to differ only in terms of one component process may actually invoke entirely different cognitive strategies and so cause activation of entirely different neurocognitive networks.

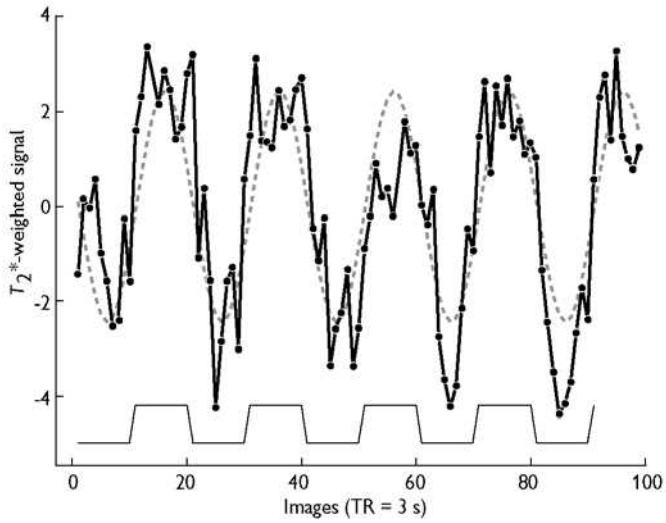


Fig. 2.3.8.1 Design, response, and modelled response for a blocked periodic experiment. The experimental design or input function is represented by a square wave (solid line), which alternates periodically between a baseline (B) and an activation (A) condition. The B condition is presented first, and the BA cycle is repeated five times in the course of the experiment. Images are acquired every 3 s during the experiment, and the T_2^*-weighted signal observed at an activated voxel is shown by points joined by a solid line. There is clearly a signal increase during the A condition. The modelled response is shown by the broken line.

Parametric design

Parametric designs are so called because the same task is presented throughout the experiment but some continuously variable parameter of the task is experimentally manipulated. For example, we could ask subjects to perform the semantic analysis task for 5 min, but continuously vary the interval between consecutive stimuli (words) from 10 s at the start of the experiment to 1 s at the end. Here we are assuming that as the task becomes more difficult, i.e. the interstimulus interval becomes shorter, blood flow to the regions specialized for semantic analysis will increase. The main advantage of this design is that it avoids the assumption of cognitive subtraction; the main disadvantage is that it may lack specificity. Motor and visual cortex, as well as brain regions specialized for semantic analysis, will probably show an increased blood flow as the rate of stimulus presentation is increased.

Event-related design

Event-related designs are composed of a series of individual stimuli. They may be coupled with sequences for very rapid image acquisition so that the temporal pattern of response to a single event can be resolved in detail. An event-related design would be advantageous for our semantic analysis experiment if we were particularly interested in correlating some aspect of the behavioural response to each stimulus, for instance accuracy of decision or reaction time, with the neurovascular response measured using fMRI. A disadvantage of such designs is that signal changes induced by a single trial are generally weak compared with the 2 to 5 per cent signal changes that are typical in blocked periodic experiments.

Beyond a single experiment

Generally, the hypothesis in question demands the investigation of more than a few subjects and/or more than one experimental condition. When designing the studies it is important that randomization should be used appropriately to eliminate confounding effects of the order in which experiments are conducted, and of the order in which different subjects are scanned. Practice on a task may substantially alter the neurovascular response, and so all subjects should receive preliminary training on the task according to a standard protocol. If there is considerable variability between subjects in their ability to perform a task, consider adjusting the difficulty of the task presented in the scanner so that each subject is performing at the same level in terms of accuracy or reaction time. The use of functional MRI to study longitudinal changes by repeated examination of the same subject(s), for instance before and after the administration of a drug, will generally improve the statistical power to detect the effect of interest by controlling for idiosyncratic variability of functional response between subjects.

Data analysis

General principles of data analysis are reviewed here; for more detailed coverage of the issues and the methods implemented in a variety of software packages, see.[5–7]

Movement estimation and correction

The first step in fMRI data analysis is to estimate the extent of head motion during data acquisition and to correct it. Due to the multislice acquisition protocols generally used for fMRI, the magnetic field to which the brain is exposed will change dramatically within the space of a few micrometres at the superior and inferior edges of a selectively irradiated slice. This means that minute head movements (<1 mm) can have disproportionately large effect on magnetic resonance signal.

Head movement occurring at the same time as time as experimental stimuli are present, namely stimulus correlated motion, can artefactually exaggerate the neurovascular response. Head movement occurring randomly with respect to the experimental design is more likely to cause the opposite problem of artefactually attenuating the measured response. Therefore it is essential to use a computerized method for movement estimation and correction.

Statistical models for the neurovascular response

The next step in analysis is to estimate the strength of the experimentally determined signal change in the time series of magnetic resonance signal measurements at each voxel in the image. This requires some sort of model for the response. The simplest model, for a blocked periodic design, is a square wave at the same frequency as the experimental input function. This model assumes that a brain region activated specifically by condition A will show an immediate increase in signal intensity, which is sustained throughout the epoch until the onset of condition B. The problem with this model is that the increase in magnetic resonance signal during condition A is due to changes in blood flow and oxygenation, which are dispersed and delayed by several seconds relative to the onset of condition A. Furthermore, this haemodynamic delay between stimulus onset and measurable response will be variable from one voxel to another. Therefore it is important that the experimental effect should be modelled as an increase in signal intensity that is arbitrarily delayed relative to the onset of the activating stimulus. The most general way of achieving this is to convolve the

experimental input function (the vector coding changes in experimental conditions) with a model of the haemodynamic response. The haemodynamically convolved input function can then be regressed on the fMRI time series at each voxel to estimate the neurovascular response to changing experimental conditions.

In fitting linear regression models to fMRI time series, one important technical issue is that the residuals of the regression will generally not be white noise or serially independent. Rather the residuals will typically have long-range dependency or long memory in time and this will need to be addressed for proper estimation of linear model parameters.[8] There are probably several possible artefactual sources of autocorrelation in fMRI time series residuals—including imperfectly modelled experimental activation, uncorrected head movement, and aliased cardiorespiratory pulsation. However, recently attention has focused on the role of low frequency, spatially coherent, endogenous oscillations of large neuronal ensembles as a source of long memory in fMRI time series.[9] This hypothesis has encouraged studies of the univariate and multivariate properties of fMRI data recorded at rest, i.e. in the absence of experimentally controlled task processing.[10,11]

Activation mapping

The next step in analysis is often to decide which of the several thousand voxels in the image have demonstrated such a strong response to the experiment that it is unlikely to be due to chance. In other words, we want to identify the significantly activated voxels. Let us assume that we have estimated the neurovascular response by the magnitude of a linear model coefficient at each and every voxel, and refer to this as our test statistic. The problem is then to assign a probability to each test statistic under the null hypothesis that the experiment had no effect on the brain. To do this we need to know the probability distribution of our test statistic under the null hypothesis. There are broadly two ways we can know this distribution: we can work it out from mathematical theory, or we can sample it by randomly permuting the data. Theoretical distributions are quicker to evaluate than permutation distributions, but permutation entails many fewer assumptions and is the gold standard against which the validity of theoretical approximations should be checked.

Once we have a probability distribution for the test statistic, we still have to decide what p-value we wish to adopt as our threshold for activation. If we choose a small (conservative) p-value (e.g. <0.00001), only those voxels that demonstrate a very powerful response will be identified as activated. There will be few false-positive or type 1 errors, i.e. almost all the voxels we identify as activated will truly be activated. But there will probably be a large number of false-negative or type 2 errors, i.e. many voxels that are truly activated will not be identified as such. Conversely, if we choose a large (lenient) p-value (e.g. <0.01), there will be a larger number of false-positive errors but a smaller number of false-negative errors. The choice of p-value should be informed by the search volume, or the number of voxels tested for significance. The larger the search volume, the smaller the p-value will need to be for an acceptable degree of type 1 error control. A rule of thumb is that the p-value should be approximately the reciprocal of the search volume. More elaborate methods have been advocated for correcting p-values for large numbers of tests on imaging data.

An alternative approach to testing tens of thousands of voxels against a suitably small probability threshold, with an associated risk of major type 2 error, is to combine information about the experimental response over several voxels. For example, we can initially apply a lenient threshold ($p = 0.05$) to the test statistics estimated at each voxel, and set to zero any voxel that does not have a test statistic greater than the corresponding critical value. The result will be a map of several spatially contiguous clusters, ranging in size from a single voxel to several hundred voxels (Plate 10). We can ascertain the probability distribution for cluster size under the null hypothesis either by theory or permutation. Then we can proceed to identify significantly activated clusters instead of voxels. The advantage of hypothesis testing at cluster level is a greater power to detect significant foci of activation, partly because there will be many fewer clusters than voxels to test, so the p-value can be legitimately increased. The disadvantage is the loss of spatial resolution of activation.

Multivariate approaches

Many of the 'higher-order' cognitive tasks that are likely to be of greatest interest to psychiatric research do not activate a single modular region of the brain. Instead, they typically activate several spatially distinct or distributed regions that together comprise a large-scale neurocognitive network for performance of the task. It may then be of interest to investigate functional integration between different regions or nodes of the network. The simplest way to do this is by estimating the correlation between a pair of fMRI time series observed at different voxels or regions. Large correlations, whether negative or positive, may be described as evidence for functional connectivity. Psychiatric disorders may be characterized by abnormal functional relationships between coactivated regions, or functional dysconnectivity.

More sophisticated techniques for investigating functional relationships between large numbers of voxels or regions include multivariate methods such as principal component analysis, discriminant analysis, and path analysis. These methods are equally applicable to structural MRI data, where they may provide indirect evidence for anatomical connectivity between regions.

Within- and between-group analysis

Once a measure or parameter of experimental response has been estimated in each fMRI time series, the resulting parameter maps can be registered in standard space. There are many possible computational algorithms for spatial registration. The most commonly used at present is an affine transformation, which applies a global and linear rescaling in three dimensions to each individual image. A commonly adopted standard space is that represented in a stereotactic atlas of the brain originally written by Talairach and Tournoux to assist neurosurgeons in locating subcortical structures.[12] In these systems, each voxel is assigned a set of $\{x, y, z\}$ coordinates which define its position. In Talairach–Tournoux space, the coordinates are defined relative to the cerebral midline and a line is drawn between the anterior and posterior commissures (intercommissural or AC–PC line). After registration, parameter maps are usually smoothed by applying a two- or three-dimensional Gaussian filter to accommodate variability in sulcogyral anatomy between subjects and error in spatial registration.

It is then possible to test a wide variety of hypotheses about the response parameters measured over several subjects at each voxel in standard space. For example, one can test the null hypothesis that there is zero mean or median power of experimental response within a group, or the null hypothesis that there is zero difference in the power of response between two groups. It is also possible to

test for correlations between the power of functional response and some behavioural or symptom measure within a group.

Visualization

The final result of fMRI data analysis will often be visualized as a map in standard space. The background for the map will generally be a grey-scale image of cerebral anatomy, such as a structural MRI dataset with fine spatial resolution and good tissue contrast between grey and white matter. In this case, one should beware of the potential discrepancy in geometric distortion between images of the same brain acquired using different sequences.

The background image will often be combined with, or substituted by, a rectangular grid allowing any feature of interest to be referred directly to the appropriate atlas of standard anatomical space. If the image is displayed as a series of two-dimensional slices, the *z* coordinate for each slice in standard space should also be displayed and the left and right sides of the right clearly indicated.

Voxels or clusters that demonstrate a significant effect are generally coloured against the grey-scale background image (Plate 11). A range of colours can be used to encode additional information. For example, the haemodynamic delay of response at each generically activated voxel may be colour coded by a continuous spectrum. Other strategies for visualization include use of three-dimensional rendering to show foci of activation in the context of the sulcogyral anatomy of a whole hemisphere, and 'flat mapping' whereby the template image is deformed to a smooth sphere and then mapped to a plane before activation foci are superimposed on it.

Further information

Comprehensive background information on fMRI physiology, experimental design, and data analysis:

Jezzard, P., Matthews, P.M., and Smith, S.M. (eds) (2003). *Functional magnetic resonance imaging: an introduction to methods*. (2nd edn) Oxford University Press, Oxford.

The physiological origins of the BOLD effect:

Logothetis, N.K., Pauls, J., Augath, M.A., *et al.* (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, **412**, 150–7.

Current perspectives of fMRI applications:

Matthews, P.M., Honey, G.D., and Bullmore, E.T. (2006). Applications of fMRI in translational medicine and clinical practice. *Nature Reviews. Neuroscience*, **7**, 732–44.

Community web site for information about Brain Mapping and methods: www.brainmapping.org

References

1. Kwong, K.K., Belliveau, J.W., Chesler, D.A., *et al.* (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 5675–9.
2. Stehling, M.K., Turner, R., and Mansfield, P. (1991). Echo-planar imaging: magnetic resonance imaging in a fraction of a second. *Science*, **254**, 43–50.
3. Le Bihan, D. and Karni, A. (1995). Applications of magnetic resonance imaging to the study of human brain function. *Current Biology*, **5**, 231–7.
4. Matthews, P.M., Honey, G.D., and Bullmore, E.T. (2006). Applications of fMRI in translational medicine and clinical practice. *Nature Reviews Neuroscience*, **7**, 732–44.
5. Jezzard, P., Matthews, P.M., and Smith S. (2001). *Functional magnetic resonance imaging: an introduction to methods*. Oxford University Press, Oxford.
6. Frackowiak, R.S.J., Ashburner, J.T., Penny, W.D., *et al.* (2002). *Human brain function*. Academic Press, San Diego.
7. Buxton, R.B. (2002). *An introduction to functional magnetic resonance imaging: principles and techniques*. Cambridge University Press, Cambridge.
8. Maxim, V., Sendur, L., Fadili, J., *et al.* (2005). Fractional Gaussian noise, functional MRI and Alzheimer's disease. *NeuroImage*, **25**, 141–58.
9. Leopold, D.A., Murayama, Y., and Logothetis, N.K. (2003). Very slow activity fluctuations in monkey visual cortex: implications for functional brain imaging. *Cerebral Cortex*, **13**, 422–33.
10. Greicius, M.D., Krasnow, B., Reiss, A.L., *et al.* (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 4637–42.
11. Achard, S., Salvador, R., Whitcher, B., *et al.* (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *The Journal of Neuroscience*, **26**, 63–72.
12. Talairach, J. and Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. Thieme, Stuttgart.

2.3.9 Neuronal networks, epilepsy, and other brain dysfunctions

John G. R. Jefferys

Introduction

The dynamics of highly interconnected networks of neurones are fundamental to both normal and pathological functioning of the brain.[1,2] Epilepsy is perhaps the most dramatic example of a dysfunctional neuronal network,[3] characterized by intense and highly synchronous neuronal activity, but more subtle dysfunction is associated with other conditions, such as schizophrenia.[4]

This chapter will largely focus on the hippocampus, and to a lesser degree on the neocortex. The hippocampal formation is implicated in several important psychiatric and neurological problems. The hippocampus and amygdala are often the site of epileptic foci, which can lead to problems in learning and memory, emotion, anxiety, and other problems. This kind of epilepsy is variously known as temporal lobe epilepsy, complex partial seizures, or limbic epilepsy. The hippocampus and associated limbic areas have been linked both to affective disorders and to psychoses. This chapter will consider the cellular organization of the hippocampus and then outline aspects of the emergent properties of neuronal networks in the hippocampus and speculative role in psychiatric disorders. Cellular and network mechanisms of focal epilepsy, and learning impairments associated with limbic epilepsy will be reviewed.

Hippocampal organization

Anatomy

The hippocampus resembles the neocortex in containing a majority of excitatory neurones, the pyramidal cells, and granule cells, which use glutamate as their neurotransmitter (E in Fig. 2.3.9.1).[1,2] Most of the remaining 10–20 per cent of neurones in the hippocampus are inhibitory, and use γ-aminobutyric acid (GABA) as their neurotransmitter (I in Fig. 2.3.9.1). The inhibitory neurones fall into several distinct subtypes according to where their axons go (and hence which cells they inhibit), where their cell bodies are, the shapes of their dendrites, whether they contain more than one transmitter, and whether they contain particular calcium-binding proteins. This chapter will ignore most of the diversity of interneurones.[5]

There are many more excitatory pyramidal cells (E and triangles) than inhibitory interneurones (I and circles). As with most neurones, they receive inputs onto their dendrites and somata (the latter contain the nucleus and are represented by a triangle or circle). The level of simplification is clear from the observation that each pyramidal cell receives tens of thousands of synapses. Axons from other regions, known as afferents, make excitatory synapses (e) with both pyramidal cells and interneurones. Most of the interneurones make inhibitory synapses (i) onto pyramidal cells. The inhibition of the pyramidal cell is called 'feed-forward' (f.f.) when the interneurones were excited by afferent axons and 'feed-back' (f.b.) when they were excited by pyramidal cells. Interneurones also inhibit each other forming a mutually inhibitory network (m.i.); this network is important in some kinds of physiological network oscillation (see text). Finally, pyramidal cells make excitatory synapses onto each other (r.e.), which can lead to epileptic discharges if not held in check by inhibitory mechanisms.

Evoked responses

Each hippocampal (or neocortical) area receives 'afferent' synaptic inputs from other areas. Most afferents are excitatory; stimulating them provides a convenient tool to study the operation of the neuronal circuits involved. The responses evoked in hippocampal neurones typically start with an excitatory postsynaptic potential. If the excitatory postsynaptic potential is strong enough, it will result in an action potential triggered at a low-threshold zone near the cell body, probably a short distance down the axon. The excitatory postsynaptic potential is followed by a fast and a slow inhibitory postsynaptic potential. Both the fast inhibitory and excitatory postsynaptic potentials are due to ligand-gated channels where the transmitter receptor is part of the same molecular structure as the ion channel. In the case of inhibition this is the $GABA_A$ receptor, which allows chloride ions to pass. In the case of excitation it is a variety of glutamate receptors, which are permeable to sodium, potassium, and in some cases calcium ions, and which are further subdivided into α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA)/kainic acid, *N*-methyl-D-aspartate (NMDA) and other classes. The slow inhibitory postsynaptic potential is due to $GABA_B$ receptors, which are G protein coupled and use second messengers to open separate potassium channels. Many other kinds of G-protein couple receptors exist, and often are involved in modulating neuronal excitability or synaptic function.

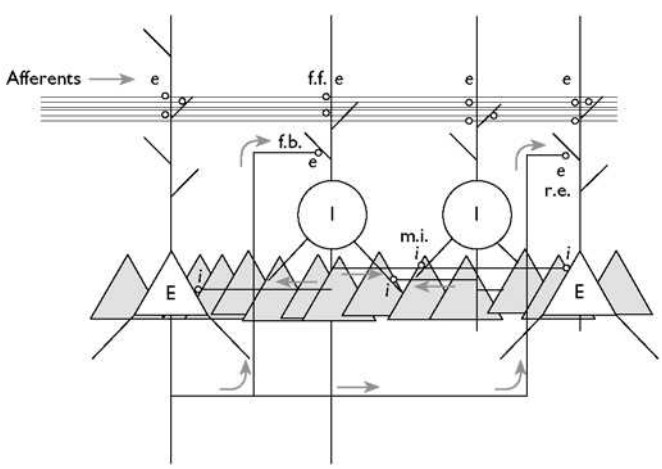


Fig. 2.3.9.1 Schematic illustration of hippocampal neuronal network.

Inhibitory neurones can be triggered both by activity in the principal cells (pyramidal or granule), resulting in recurrent or feedback inhibition, and directly by the incoming afferents, resulting in feed-forward inhibition. Experimentally, the synchrony of the stimulation of the afferent input imposes synchrony on the response with the useful consequence that the extracellular currents generated by the activity of individual pyramidal or granule cells can summate (because the cells are located in tight layers) and produce large 'field potentials' comprising a population excitatory postsynaptic potential, followed by a population spike.

Field potentials evoked by local stimulation are over in 10 to 20 ms and the slowest intracellular components end within a few 100 ms to 1 s. However, stimulation can have much more prolonged effects. The best known of these is long-term potentiation, in which a brief train of stimuli can result in an increase, lasting hours or days, in the response to a fixed test stimulus. The modest conditioning event and the enduring consequence make long-term potentiation an attractive model of learning and memory, although the evidence that it really is the direct cellular substrate for learning remains circumstantial.[6] It is perhaps more likely that long-term potentiation provides an artificial experimental tool that depends on cellular and molecular mechanisms that may also be involved in learning and/or other plastic changes in synaptic strength.

Local circuits

Hippocampal neurones are not just arranged as a simple synaptic relay where afferents excite target cells to produce an output depending on the size of the input and the state of inhibition at that time. Instead there exists a complex synaptic network, or local circuit, interlinking neurones of all kinds (Fig. 2.3.9.1 gives a very much simplified illustration of some of the salient features of hippocampal local circuits). This chapter considers two kinds of emergent network activity that arise from this organization: focal epilepsy and gamma rhythms.

Experimental approaches

Unravelling the cellular and network mechanisms of emergent network phenomena depends on a combination of electrophysiology, pharmacology, anatomy, and realistic computer simulations.[2,7] Two practical issues may need a brief introduction.

Brain slices have played a pivotal role in developing theories on the operation of neuronal networks. Slices about 0.4 mm thick are cut from the brains of deeply anaesthetized or recently killed experimental animals, or sometimes from humans undergoing

neurosurgery. Brain slices can survive many hours *in vitro* in an artificial cerebrospinal fluid, usually equilibrated with 95 per cent oxygen and 5 per cent carbon dioxide (with bicarbonate providing a pH buffer). If the slices are prepared under sterile conditions, they can survive for weeks as 'organotypic slice cultures'. In both cases the visualization of the anatomy of the living slice helps locate electrodes, the mechanical stability greatly simplifies recordings from inside neurones, and the lack of a blood–brain barrier facilitates drug applications and changes in ion concentrations. Brain slices have proved immensely popular and successful, but it is important to remember that they are only one tool in the armoury needed to study brain function, and that ultimately results from them must be put in the context of the whole organism.

Realistic computer simulations provide the means to determine whether what we know at one level, for instance of the properties of individual neurones and their interconnections, are necessary and sufficient to explain the emergent properties at the next level, here of circuits of a few thousand neurones. The most useful models for this purpose are tightly constrained by experimental data, and ideally are used to make experimentally testable predictions. Mostly they consist of several 'compartments' to represent the anatomy of the neurone's dendrites and soma. Each compartment consists of several differential equations representing specific ion channels, which may be gated by membrane potential, extracellular neurotransmitters, or intracellular calcium. Large number of neurones can then be wired together in larger-scale simulations, using biologically realistic anatomical connectivity and synaptic properties.[2,7]

Emergent properties of hippocampal networks

Hippocampal rhythms

The organization of hippocampal networks (Fig. 2.3.9.1) leads to several distinct kinds of oscillation, which can be considered as 'emergent' properties of the network. Perhaps the most prominent rhythm in the hippocampus is theta (3–7 Hz), which, at least in rats, is associated with spatial navigation,[8] and may play a role in memory.[9] Theta results from interactions of the hippocampus with two other limbic structures, the septum and the entorhinal cortex. Often superimposed on theta is a faster rhythm known as gamma (30–100 Hz). The best evidence we have now is that gamma is generated by local circuits in the hippocampus and that inhibitory neurones play a crucial role.[2] The role of gamma in the hippocampus remains unclear; in the neocortex it has been implicated in higher cognitive processes such as the 'binding' of individual sensory features into coherent perceived objects.[10]

Networks for gamma rhythms

The first strong clue that inhibitory neurones played a central role in gamma rhythms came from hippocampal and neocortical slices in which fast excitatory postsynaptic potentials had been blocked by drugs. Excitation by pulses of glutamate or agonist drugs acting at metabotropic (i.e. G-protein-coupled) glutamate receptors resulted in rhythmic inhibitory postsynaptic potentials in the gamma frequency band. A series of experimental tests of predictions from realistic computer simulations showed that this gamma rhythm was generated by the mutual inhibition of inhibitory neurones, which produced a synchronous interruption of the fast discharge the metabotropic glutamate receptor activation would otherwise have evoked. These interruptions lasted for a time, of the order of 25 ms, that depended on the time course of the inhibitory postsynaptic potentials in interneurones. We named this phenomenon 'interneuronal network gamma'.

During interneuronal network gamma pyramidal cells generate rhythmic inhibitory postsynaptic potentials, but do not reach threshold unless they are driven by some other input. Another kind of gamma rhythm occurs when slices are exposed to cholinergic drugs such as carbachol, and/or to non-desensitizing glutamate agonist drugs such as kainic acid. Here each pyramidal cell fires on some cycles of the rhythm,[11,12] so that on average some fluctuating fraction of pyramidal cells fires on each cycle. This is closer to the situation *in vivo*.[13]

Significance of fast coherent cortical oscillations

Rhythms such as gamma have been linked with sensory processing and with perception and other cognitive functions. They may be disrupted in people with some degree of cognitive impairment, for instance normal age-related cognitive decline,[14] or there may be a more general disruption of cortical rhythms in more severe conditions such as schizophrenia.[4] The complexity of the circuits responsible for cortical oscillations means that individuals with apparently normal number of neurones and neuronal organization may still have rather subtle changes on their synaptic networks that can have profound effects on collective oscillations and behaviour.

Gamma rhythms are intimately linked with epilepsy. Coherent neural activity at gamma frequencies is associated with some kinds of epileptic activity. Gamma rhythms are disrupted in at least one chronic model of epilepsy associated with learning impairments. Finally, the ideas behind the synaptic network mechanisms of the two kinds of phenomena have much in common.

Epilepsy—an emergent property of neuronal networks

Epileptic discharges typically involve excessively synchronous activity in principal neurones. In experimental focal epilepsy this excessive synchronization is due to the mutual excitation of pyramidal cells in the hippocampus, neocortex, or related areas. The essential idea is of a chain reaction. Areas that are especially prone to epileptic discharges have strong synaptic interconnections between their principal cells (e.g. the pyramidal cells of the CA3 region of the hippocampus or layers 3 and 5 of the neocortex). Activity in a few pyramidal cells can propagate through the synaptic network to recruit the whole population of neurones. Normally this propagation is held in check by inhibitory neurones; if the control mechanism is ineffective then epileptic discharges result. In experimental models the balance of synchronization versus control is compromised by treatments that weaken inhibition (usually by drugs such as bicuculline or picrotoxin), strengthen excitation (incubating brain slices in solutions lacking magnesium ions) or strengthen synaptic potentials in general (4-aminopyridine). Combined experimental and theoretical studies of such models have led to some general principles.[7] Synchronous epileptic discharges will result under the following conditions.

1 Connections between excitatory neurones are divergent, that is each connects to more than one postsynaptic excitatory neurone.

2 Connections between excitatory neurones are powerful enough to make their postsynaptic cells fire with a high probability. Precisely how high a probability can be depends on factors such as the connectivity and size of the network. The 'intrinsic' electrical properties of the neurones are important. Many epilepsy-prone areas have cells with prominent voltage-sensitive calcium currents, which are more prolonged than the classical voltage-sensitive sodium currents of the axonal action potential, and which cause neurones to fire bursts of fast sodium action potentials. Such intrinsic bursts greatly amplify transmission between pyramidal cells.

3 The network is large enough to allow all the neurones to link together. The critical mass for a network where the probability of any two cells being directly connected is 1 per cent, and the probability of one cell exciting its target cells is approximately 50 per cent, works out at about 1000 to 2000 neurones.

These features explain experimental brief epileptic discharges very effectively. The brain contains inhibitory mechanisms, both synaptic (inhibitory postsynaptic potentials, presynaptic inhibition) and intrinsic (voltage- and calcium-sensitive potassium channels), to terminate hypersynchronous discharges. Other mechanisms are needed to overcome the burst-termination mechanisms for the crucial transition to full-blown seizures lasting tens of seconds to minutes. These include slower synaptic mechanisms (both *N*-methyl-D-aspartate and metabotropic glutamate receptors, GABA, which paradoxically can become depolarizing if present in excess), non-synaptic mechanisms (potassium accumulation, electric fields), and abnormal activity arising in axons (ectopic spikes, gap junctions).

Convulsant drugs can trigger seizures in normal brains. People with epilepsy have a reduced seizure threshold that means they have seizures without an obvious triggering chemical or event. The reasons are far from clear, but may include abnormalities in intrinsic properties of neurones or in the connectivity of the neurones. Improvements in non-invasive imaging and in neuropathology increasingly reveal misplaced neurones and other more or less subtle anatomical malformations in many focal epilepsies, which suggests that the local circuitry is disturbed.

Other kinds of epilepsy have very different mechanisms. Absence epilepsy is the other major class where cellular mechanisms are relatively well understood. They involve the interaction of the thalamus and neocortex, although the received wisdom on the underlying mechanism has recently been challenged by experiments on one of the key animal models of absence epilepsy.[15]

Epilepsy—learning and memory

Patients with temporal lobe epilepsy can have problems with learning and memory. Antiepileptic drugs can have marked side effects, but the observation that chronic animal models of temporal lobe epilepsy also have impairments in learning and memory suggests that this is not the only cause. Memory impairments could result from gross damage, such as hippocampal sclerosis. Gross hippocampal pathology will have effects similar to experimental lesions of the hippocampus, but the observation that at least some of the chronic animal models lack gross hippocampal pathology does not support neuronal death as being the sole cause of impaired learning and memory.

At least two chronic experimental epilepsies have either limited or no cell loss during their induction. These are kindling and stereotaxic injection of a minute dose of tetanus toxin. The tetanus toxin model results in a well-characterized and enduring impairment of learning and memory, which outlasts the active epileptic syndrome in all except a few rats that show relapse. The absence of medication and of gross cell loss suggest that the psychological impairments in this model have some functional cause. Long-term potentiation remains intact, at least over a period of up to an hour. There is an association of learning impairment with the size of population spikes recorded from the same rats *in vivo* and under anaesthesia. Inhibition remains impaired in rats at a stage (>3 months after injection) when they had gained remission from epileptic seizures, but retained learning impairments.[16,17] Abnormalities of the cellular electrophysiology of the postepileptic phase can lead to disruption of network properties, including gamma rhythms, which may, in time, provide a link to the behavioural problem.

Humans with limbic epilepsy often do have substantial hippocampal damage, and this will contribute to learning impairments. The experimental evidence suggests that even in the absence of gross hippocampal damage learning impairments can arise as a result of functional disruption of the hippocampal network.

Conclusions

Understanding the operation of networks of neurones provides valuable insight into a range of neurological and psychiatric diseases. The role of synaptic networks of excitatory neurones in focal epilepsies is now well established. The ways in which brief epileptic discharges transform into events lasting as long as full seizures are starting to be clarified, and may offer new avenues for developing rational therapies. New ideas on the generation of physiological rhythms suggest novel models of psychiatric and neurological problems ranging from impairments in learning and memory in limbic epilepsies to (more speculatively) the disruption of sensory perception in psychoses. Real clinical cases will inevitably be much more complex, but the ideas and models outlined above will aid the understanding of the underlying mechanisms.

Further information

Traub, R.D., Jefferys, J.G.R., and Whittington, M.A. (1999). *Fast oscillations in cortical circuits*. The MIT Press, Cambridge, MA (ISBN 0-262-20118-6).

Jefferys, J.G.R. (2007). Epilepsy *in vitro*: electrophysiology and computer modeling. In *Epilepsy: a comprehensive textbook* (eds. J. Engel Jr., T.A. Pedley, J. Aicardi, *et al.*). Lippincott, Williams & Wilkins, Philadelphia (ISBN 978-0781757775).

Winterer, G. and Weinberger, D.R. (2003). Cortical signal-to-noise ratio: insight into the pathophysiology and genetics of schizophrenia. *Clinical Neuroscience Research*, **3**, 55–66.

http://www.neuroscience.bham.ac.uk/neurophysiology/

References

1. Traub, R.D., Jefferys, J.G.R., and Whittington, M.A. (1999). Functionally relevant and functionally disruptive (epileptic) synchronized oscillations in brain slices. *Advances in Neurology*, **79**, 709–24.

2. Traub, R.D., Jefferys, J.G.R., and Whittington, M.A. (1999). *Fast oscillations in cortical circuits*. The MIT Press, Cambridge, MA.
3. Jefferys, J.G.R. (2003). Models and mechanisms of experimental epilepsies. *Epilepsia*, **44** (Suppl. 12), 44–50.
4. Winterer, G. and Weinberger, D.R. (2004). Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. *Trends in Neurosciences*, **27**, 683–90.
5. Somogyi, P. and Klausberger, T. (2005). Defined types of cortical interneurone structure space and spike timing in the hippocampus. *The Journal of Physiology*, **562**, 9–26.
6. Bliss, T.V.P. and Collingridge, G.L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature*, **361**, 31–9.
7. Jefferys, J.G.R. (2007). Epilepsy *in vitro*: electrophysiology and computer modeling. In *Epilepsy: a comprehensive textbook* (eds. J. Engel Jr and T.A. Pedley). Lippincott, Williams & Wilkins, Philadelphia.
8. O'Keefe, J. and Burgess, N. (2005). Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. *Hippocampus*, **15**, 853–66.
9. Axmacher, N., Mormann, F., Fernandez, G., *et al.* (2006). Memory formation by neuronal synchronization. *Brain Research Reviews*, **52**, 170–82.
10. Usrey, W.M. and Reid, R.C. (1999). Synchronous activity in the visual system. *Annual Review of Neuroscience*, **61**, 435–56.
11. Fisahn, A., Pike, F.G., Buhl, E.H., *et al.* (1998). Cholinergic induction of network oscillations at 40 Hz in the hippocampus *in vitro*. *Nature*, **394**, 186–9.
12. Mann, E.O., Suckling, J.M., Hajos, N., *et al.* (2005). Perisomatic feedback inhibition underlies cholinergically induced fast network oscillations in the rat hippocampus *in vitro*. *Neuron*, **45**, 105–17.
13. Csicsvari, J., Jamieson, B., Wise, K.D., *et al.* (2003). Mechanisms of gamma oscillations in the hippocampus of the behaving rat. *Neuron*, **37**, 311–22.
14. Vreugdenhil, M. and Toescu, E.C. (2005). Age-dependent reduction of gamma oscillations in the mouse hippocampus *in vitro*. *Neuroscience*, **132**, 1151–7.
15. Manning, J.P., Richards, D.A., Leresche, N., *et al.* (2004). Cortical-area specific block of genetically determined absence seizures by ethosuximide. *Neuroscience*, **123**, 5–9.
16. Vreugdenhil, M., Hack, S.P., Draguhn, A., *et al.* (2002). Tetanus toxin induces long-term changes in excitation and inhibition in the rat hippocampal CA1 area. *Neuroscience*, **114**, 983–94.
17. Jefferys, J.G.R. and Mellanby, J. (1998). Behaviour in chronic experimental epilepsies. In *Disorders of brain and mind* (eds. M.A. Ron and A.S. David), pp. 213–32. Cambridge University Press, Cambridge.

2.3.10 Psychoneuroimmunology

Robert Dantzer and Keith W. Kelley

Introduction

Mind-body literature, in the form of magazines and self-help books on stress and healing, is full of definitive claims for the existence of powerful influences of emotions and psychosocial stressors on the immune system, leading to onset or progression of cancers or infectious diseases. This literature often makes explicit reference to research in psychoneuroimmunology to support these claims. Psychoneuroimmunology is a multi-disciplinary field that has grown rapidly during the last three decades at the crossroads of immunology, behavioural neurosciences, neuroendocrinology, and psychology. It studies mechanisms and functional aspects of bidirectional relationships between the brain and the immune system. Although still controversial, there is evidence that psychological events including emotions can and do influence the outcome of infectious, autoimmune, and neoplastic diseases via modulation of cells of the immune system. A surprising finding has been that immune events occurring in the periphery also affect mood, behaviour, and metabolism by modulating brain functions, thereby providing a biologically important link between the immune system and brain. The original discovery that activation of the innate immune system in the periphery causes clinical signs of sickness that are processed in the brain is now being extended to the involvement of the immune system in depressive disorders. This new information has solidified the idea that neurotransmitters, neuropeptides, neural pathways, and immune-derived signals such as cytokines are the minimal essential elements that permit the immune system and brain to communicate with one another. These new data offer the unexpected conclusion that the immune system is likely to be involved in not only how emotions affect health but also how immune events regulate the development and expression of emotions.

Brain influences on immunity

Early investigations

The concept that stressors can have a negative impact on immunity and ultimately induce a reduction in host resistance to infectious pathogens and even to tumour progression is not new. As a follow-up of the early studies of Hans Selye on stress, several scientists demonstrated in the 1950s and 1960s that laboratory rodents exposed to various stressors, including inescapable painful electric shocks, displayed an altered resistance to viral, bacterial, and parasitic infections. These effects of stressors were accompanied by decreases in antibody responses to the specific microbial pathogen under study. In view of the already demonstrated immunosuppressive effects of glucocorticoids and the pivotal role of glucocorticoids in the stress response, pathophysiological mechanisms of the immunosuppressive effects of stressors were easy to determine. However, it was already clear at that time that mechanisms of the effects of stressors on immunity were not that simple since decreases as well as increases in immunity could be observed, depending on the immune response under study, the type of stressor and the time point at which the stressor took place during the mounting of the immune response. Furthermore, administration of glucocorticoids at physiological instead of pharmacological doses had little effect on some aspects of immunity and evolution of the disease process.

There was little innovation in this field until the 1980s when a few pioneer immunologists and neuroscientists decided to work together in order to understand how the central nervous system communicates with the immune system. One impetus for this was the demonstration that immune responses can be submitted to Pavlovian conditioning in apparently the same way as the salivary response, as discovered by Robert Ader at the University of Rochester.[1] Mice exposed to a new taste paired with an immunosuppressive agent such as cyclophosphamide during the development of an antibody response were found to display further decreases in antibody titres when re-exposed to the taste alone in

the absence of any immunomodulating agent. For this to occur, there must be pathways of communication from the brain to the immune system that are activated by the taste paired with cyclophosphamide. The search for these pathways of communication resulted in the demonstration of innervation of the primary (thymus) and secondary (spleen and lymphoid nodes) lymphoid organs by the sympathetic nervous system.[2] Sympathetic efferent nerves enter lymphoid organs with the vasculature but ultimately separate from blood vessels to innervate the parenchyma, where both B and T lymphocytes reside and can proliferate. Sympathetic fibres innervating lymphoid organs contain all the neurotransmitter machinery of other sympathetic neurones, including noradrenaline and neuropeptides such as substance P, neuropeptide Y (NPY), and calcitonin gene-related peptide (CGRP). This implies that the chemical composition of the microenvironment in which lymphocytes are present ultimately depends on activity of the autonomic nervous system. These findings gained in prominence when it was discovered that specific subsets of leukocytes have receptors for these neuronal communication signals.

Receptors within the immune system

In addition to cytoplasmic receptors that bind steroid hormones including glucocorticoids and sex hormones, lymphocytes, and other cells of the immune system have been found to have membrane receptors that bind and respond to most neurotransmitters and neuropeptides and are quasi-identical to brain neurotransmitter and neuropeptide receptors. As supported by an important body of literature, activation of these receptors in leukocytes has functional consequences on immune responses whether immunity is measured in vivo or in vitro.[3] As an typical example, growth hormone (GH), a pituitary hormone known for its growth-promoting activity and with no known immune function, was shown to restore the resistance of hypophysectomized rats to an infection with *Salmonella typhimurium*, with an efficacy comparable to that of a tetracycline antibiotic or the macrophage-stimulating factor interferon-gamma (IFN-γ).[4,5] These results obtained in vivo were replicated in vitro.[6] GH activated highly purified populations of pulmonary macrophages in the same way as IFN-γ. Both factors were able to prime macrophages triggered with opsonized zymosan to secrete superoxide anion O_2^-, an index of macrophage activation, even if GH was less active in this system than IFN-γ. Antibody blocking studies demonstrated that the priming activity of GH was independent of IFN-γ, and vice versa the activity of IFN-γ was distinct from that of GH. This priming had functional consequences since both IFN-γ and GH increased the capability of macrophages to kill *Pasteurella multocida*. Since most of the effects of GH on its target cells are mediated by the local production of insulin-growth factors (IGF), the capability of IGF-I to prime alveolar macrophages in vitro was also tested and found to be similar to that of GH, although the priming effects of GH were independent of the local production of IGF-I.[7] Other studies were showing at the same time that non-stimulated as well as immune-activated leukocytes were able to produce a GH-like peptide that was identical to pituitary GH,[8] conferring credibility to the important hypothesis that communication signals originally identified in the neuroendocrine system can actually be used by immune cells. In the same vein, corticotropin-releasing hormone (CRH), the main regulator of the hypothalamic–pituitary–adrenal axis, has been identified in the immune system in which it functions as an autocrine/paracrine mediator of inflammation.[9] In particular, CRH causes degranulation of mast cells and the release of histamine and several proinflammatory mediators.

The neuropeptides that are contained in sympathetic nerve endings that innervate lymphoid organs can play an important role in the modulation of the fine balance between the different populations of T helper (Th) cells that regulate cellular and humoral immunity. Th1 cells normally produce IFN-γ and interleukin-2 (IL-2), and both promote cellular immunity. In contrast, Th2 cells normally produce IL-4 and IL-10 that down-regulate cellular immunity and promote humoral immunity. CGRP and NPY drive Th1 cells towards the production of IL-4 whereas Th2 cells are driven by somatostatin and CGRP to produce IL-2 and IFN-γ.[10] If these effects that were observed in vitro are also true under in vivo conditions, they provide a possible mechanism by which stress can polarize immune responses in the direction of either Th1 or Th2 cells.

A recent potentially important discovery is that of the inhibition exerted by the parasympathetic nervous system on the production of proinflammatory cytokines by macrophages. Direct electrical stimulation of the peripheral vagus nerve that innervates the liver inhibited the production of proinflammatory cytokines by Kupffer cells in response to a lethal dose of endotoxin and prevented development of septic shock.[11] This vagal function was termed the cholinergic anti-inflammatory pathway[12] and it is mediated by nicotinic acetylcholine receptors containing an alpha-7 subunit.

Neural influences on the immune system

Since the immune system makes use of communication signals and receptors that are identical to those used by the central nervous system, the immune system should be very sensitive to neural influences. Besides the cholinergic anti-inflammatory pathway already mentioned, many data attest to the fact that brain events have an impact on immune responses. For instance, lesions in the neuroendocrine brain have profound influences on immunity. As an example, destruction of the tubero-infandibular region of the hypothalamus in mice persistently abrogates natural killer cell cytotoxic activity without altering T and B cell populations, but cortical and sham lesions had only a short-lived effect.[13] In other studies, ablation of the left sensori-motor cortex decreased cellular immunity whereas ablation of the right sensory-motor cortex increased it, showing that brain influences on immunity are lateralized.[14] This lateralization phenomenon was later demonstrated to exist in the absence of any lesion since left-handed mice, labelled as such based on their predominant use of the left paw to reach a food pellet in a tube that only enabled them to use one paw, displayed higher cellular immune responses than right-handed mice.[15] The mechanisms for this lateralized influence of the brain on immunity are still elusive.

The impact of stressors on immune responses represents another example of the influence of brain events on immunity. At the time these studies were carried out it was already well known that the influence of psychosocial stressors on the hypothalamic–pituitary–adrenal axis are not simply a function of the intensity and duration of the stressors but also depend upon their psychological features. Novelty, predictability, and controllability are the key factors that ultimately determine the neuroendocrine impact of stressors. It was therefore not surprising that the same psychological features were pivotal in the influence of stress on immunity. For instance,

rats exposed to inescapable electric shocks 24 h after injection of syngenic tumour cells displayed more rapid tumour growth and a higher mortality rate than rats exposed to controllable electric shocks, despite the fact that the intensity and duration of electric shocks were exactly the same in both groups.[16] Lack of control had the same influence on the rejection of non-syngenic tumours[17] and cellular immunity as measured by the proliferative response of lymphocytes to T-cell mitogens.[18] It cannot be inferred that uncontrollability is always immunosuppressive. Lack of control over the occurrence of electric shocks was later found to increase rather than decrease humoral immunity, as measured by antibody titres against sheep red blood cells injected into rats that were submitted chronically to controllable or uncontrollable electric shocks.[19] The same difference in the way the immune system responds to an uncontrollable stressor was confirmed in an experiment in which mice were exposed to the odour of a stressed congener. A 24 h exposure to this stressor decreased the cellular immune response, as measured by proliferation of T cells to mitogens and natural killer cell cytotoxicity, but increased antibody titres against keyhole limpet haemocyanin.[20]

Stress and the immune system

Studies of the influence of stress on the immune system have also been carried out in human subjects in experimental settings or in real-life conditions. In these studies, the immune end points are either measured on blood lymphocytes or deduced from the result of an already existing pathological process. The group of Janice Kiecolt-Glaser at Ohio State University in Columbus is certainly the pioneer in this field. For instance, first year medical students were shown to display a reduction in the production of IFN-γ by circulating leukocytes and a reduced cytotoxicity of natural killer cells during the end of the year examination period, and these changes were independent on lifestyles.[21] As a result of extensive studies on different populations at risk, such as spouses experiencing marital conflict, caregivers of patients with Alzheimer's disease, and aged subjects, Kiecolt-Glaser's group proposed that negative emotions and stressful experiences can contribute to prolonged infection and delayed wound healing. In addition, negative emotions were proposed to directly produce the production of proinflammatory cytokines and therefore increase the risk for a spectrum of conditions associated with ageing, including cardiovascular disease, osteoporosis, arthritis, type 2 diabetes, certain cancers, frailty and functional decline, and periodontal disease.[22]

The influence of stressful life events on immune responses also appear to be modulated by coping strategies, as exemplified by a study carried out on susceptibility to upper respiratory tract illness in an adult population sample.[23] In a little less than 30 per cent of a population sample of adults between 18 and 65 years of age, the occurrence of clinical episodes of upper respiratory tract illness over a 15-week period was more frequent in those individuals who experienced high life event stress both before and during the study period. The impact of life events was buffered by an avoidance coping style.

In accordance with the hypothesis that negative emotions negatively impact the immune system, major depressive disorders were initially thought to be associated with depressed immune responses. This association has been confirmed for the number of circulating lymphocytes, proliferative response of lymphocytes to non-specific mitogens, and natural killer cell cytotoxicity.[24,25] However, more recent studies have revealed signs of activation of the innate immune system in at least some forms of depression.[25,26]

Although most of the literature deals with the influence of stressors and negative emotions, positive emotions have also been studied in their relation to immune events. In graduate students vaccinated against hepatitis B, dispositional positive affect was associated with a greater antibody response to vaccination.[27] The same trait was associated with decreased vulnerability to upper respiratory infections. Dispositional optimism, as defined by generalized positive expectations for the future, is positively related to measures of cellular immunity in cancer and HIV patients only when stressors are brief, relatively straightforward and controllable whereas the reverse relationship is observed when stressors are complex, persistent, and uncontrollable.[28]

Mechanisms of the relationship between stressful life events, emotions, and immunity are rarely investigated because of the many biobehavioural pathways that can be implicated in mediating relationship. In view of the postulated immunosuppressive effects of glucocorticoids, it has been important to demonstrate that not all effects of stress on immunity are mediated by activation of the hypothalamic–pituitary–adrenal axis. As an example, implantation of a corticosterone pellet after adrenalectomy in rats that were submitted to inescapable electric shocks so as to prevent the stress-induced increases in plasma corticosterone did not alter the decreased proliferative response of blood lymphocytes observed in stressed rats.[29] Evidence for a role of the sympathetic nervous system was provided by experiments using beta-adrenergic receptor antagonists or sectioning of the sympathetic nerve innervating the spleen. For instance, administration of propranolol prevented the decreased lymphoproliferative response in rats re-exposed to the cage in which they had been previously exposed to inescapable electric shocks.[30] Other possible biological mediators are CRH and endogenous opioids. In the clinic, the search for possible biological mediators of the relationship between stress and immunity is not easily found, and often confounded by the impact of stress and negative emotions on illness behaviour, via for instance deterioration in health-promoting behaviour and alterations in symptom perception.[31]

Susan Lutgendorf at the University of Iowa recently summarized the current view of interactions between health behaviours and psychosocial and biological factors that can combine to affect a multitude of disease outcomes in a biopsychosocial model (Fig. 2.3.10.1).[32]

The immune system as a true sensory organ

In a recent longitudinal study on the relationship between positive affect and clinical signs during a bout of influenza in volunteers inoculated with rhinovirus or influenza virus, production of inflammatory mediators by cells of the innate immune system in the nasal secretions was associated with reduced positive affect.[33] Since 1-day lagged analyses showed that daily production of inflammatory mediators predicted lower positive affect on the next day, it was difficult to interpret these findings in terms of the previously described relationship between emotions and immunity, negative emotions, or decreased positive affect increasing innate immunity. On the contrary, the authors interpreted their findings as supporting a causal association between pathogen-induced local cytokine production and changes in positive affect.

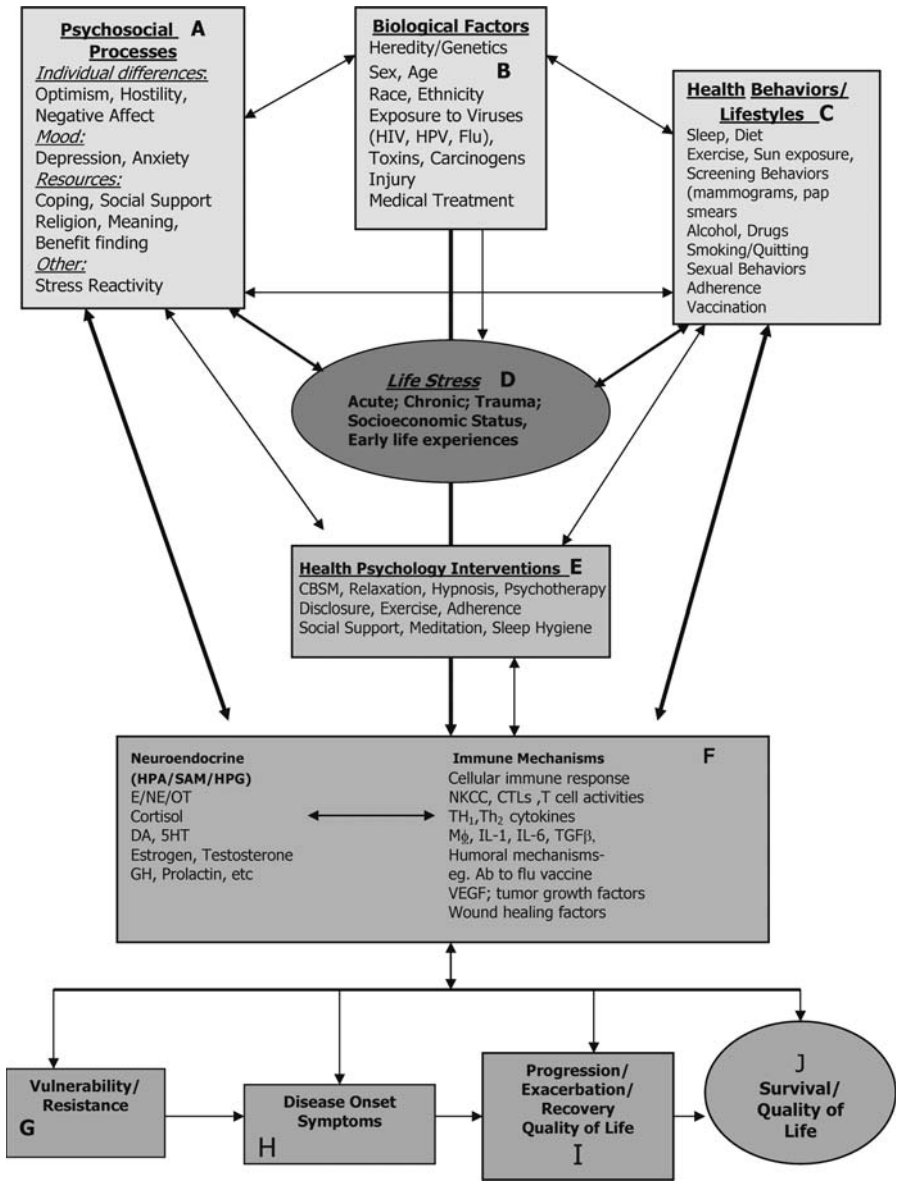


Fig. 2.3.10.1 The figure illustrates the biopsychosocial model in health psychology. The interaction between psychosocial processes (Box A), biological factors (Box B), and health behaviours (Box C) leads to a vulnerability (or resistance) to illness (Box G), disease onset and symptoms (Box H), progression, exacerbation, recovery, with concomitant quality of life (Box I), and survival with concomitant quality of life (Box J) via processes involving neuroendocrine and immune mechanisms (Box F). Effects of life stress (Box D) are filtered through psychosocial processes (Box A) and health behaviours (Box C) in their resultant effects on downstream mechanisms. Health psychology interventions (Box E) can modulate effects of psychosocial processes and health behaviours on neuroendocrine and immune mechanisms and on resultant health outcomes. There are also pathways between biobehavioural factors and disease outcomes not involving neuroendocrine or immune mechanisms, but other pathways are not included in this figure. Psychosocial processes (A) encompass psychological and social factors, particularly those that involve interpretation of and response to life stressors. These include personality variables (e.g., optimism, hostility, and negative affect), mental health and mood variables (e.g., depression and anxiety), coping, social support, spirituality, and sense of meaning. Health behaviours (C) include drug and alcohol use, smoking, sleep, nutrition, exercise, adherence to medical regimens, physical examinations, risk screenings, and risky sexual behaviours, among others. Health psychology interventions (E) can be used to alter psychosocial processes (A: e.g., decrease depression, increase coping) or improve health behaviours (C: e.g,. smoking cessation) to provide a more positive influence on neuroendocrine and immune factors and perhaps slow disease progression/exacerbation. Interventions include cognitive behavioural stress management (CBSM), relaxation, hypnosis, meditation, emotional disclosure, adherence-based interventions, sleep hygiene, exercise, social support groups, psychotherapy, imagery, distraction, behavioural pain management, yoga, massage, biofeedback, drug/alcohol prevention/rehabilitation, psychotherapy, and behavioural conditioning. These interventions can be used at all points of the trajectory of the disease or condition. Box F shows selected mechanisms involved in the bidirectional interactions between neuroendocrine and immune axes that mediate the relationships between biobehavioural factors (A–D) and disease outcomes (G–J). This by no means is an all-inclusive list of mechanisms, but it represents some of the commonly studied factors in this literature. Once vulnerability (G) has been established, continued interaction with positive or negative psychosocial factors (A: e.g., depression/social support), disease factors (B), adaptive/maladaptive health behaviours (C) and stress (D) will contribute to expression (or lack thereof) of disease symptoms (H), disease-free intervals/progression/exacerbation, and quality of life (e.g., functional, physical, emotional, and social well-being) (I), and survival (J). HPA, hypothalamic–pituitary–adrenocortical axis; SAM, sympathoadrenomedullary axis; HPG, hypophyseal pituitary gonadal axis; OT, oxytocin; DA, dopamine; 5HT, seratonin; GH, growth hormone; NKCC, natural killer cell cytotoxicity; CTLs, cytotoxic lymphocytes; $M_{Ð}$, macrophage; IL-1, interleukin 1; IL-6, interleukin 6; TGFÐ, transforming growth factor beta; Ab, antibody; and VEGF, vascular endothelial growth factor. (Reprinted from Brain Behaviour and Immunity, 17(4), ASK Lutgendorf and Es Costanzo, Psychoneuroimmunology and health psychology: an integrative model, 225–32, Copyright 2003, with permission from Elsevier.)

To understand this interpretation, it is necessary to replace the relationship between the immune system and the brain in the context of regulatory immunophysiology. If the brain communicates with the immune system via neuroendocrine factors and autonomic neuronal pathways, it is probably because the immune system needs the brain to regulate its function and do what it cannot do by itself, (i.e., engage the whole organism in the fight against microbial pathogens). If this is the case, then it ensues that the immune system needs to inform the brain of its state of activity. In other words, brain-to-immune communication pathways need to be activated by immune-to-brain communication pathways. This interpretation makes the immune system a true sensory organ, specialized in the detection of the non-self and able to transmit this information to the brain.

Looking through the mirror: immune modulation of emotions and mood

The isolation, cloning, and expression of proinflammatory cytokines and their receptors in the late 1980s, coupled with the discovery of pathogen-associated molecular patterns (PAMPs) and their pathogen recognition receptors (PRRs), are two of the major advances in immunology that set the stage for defining immune-to-brain communication pathways. Interleukin-1 (IL-1) is a prototypical proinflammatory cytokine that is released from activated macrophages following activation of some PRRs. Other proinflammatory cytokines include tumour necrosis factor (TNF-α), IL-6, and IL-8. Injection of IL-1 or TNF-í, either systemically in the form of intraperitoneal administration or centrally via an intracerebroventricular route, induces the classical signs of illness, including fever, activation of the hypothalamic–pituitary–adrenal axis, and sickness behaviour, as evidenced by decreased interaction with the physical and social environment, reduced appetite, disappearance of body care activities, fatigue, malaise, and mild cognitive impairments.[34–37] These actions take place in the brain since pretreatment of rats with the specific antagonist of IL-1 receptors into the lateral ventricles of the brain significantly impairs the ability of systemic IL-1 to cause behavioural deficits.

At the molecular and cellular levels, it is now clear that IL-1 and other proinflammatory cytokines produced by activated innate immune cells in the periphery during the inflammatory response induce expression of the same cytokines in innate immune cells of the brain, including meningeal and perivascular macrophages, microglial cells in the brain parenchyma and mast cells. Peripheral cytokines are relayed to the brain via a humoral pathway involving the action of PAMPs on PRRs located in circumventricular organs at the interface of the brain and periphery and a neural pathway involving activation of the afferent nerves that innervate the body region in which the inflammatory response takes place. At the organism level, this central representation of the peripheral inflammatory response imposes a new mode of functioning on the brain so as to allow the organism to better cope with infection. The elevated body temperature actively maintained by the increased thermogenesis and decreased thermolysis that characterizes the fever response promotes a number of immune events and reduces microbial proliferation. At the same time, zinc and iron that are normally essential for cellular multiplication are sequestered and acute phase proteins including complement proteins are synthesized by hepatocytes to help killing microbial pathogens. The increased glucocorticoid levels that occur as a consequence of the hypothalamic effect of cytokines negatively feed back on activated innate immune cells and significantly limit the intensity and duration of the inflammatory response. Sickness behaviour itself helps to reduce physical activities detrimental to thermogenesis and facilitates the reduction of thermolysis. It also limits the spread of infection within the social group by reducing social activities and reduces pain that is enhanced by proinflammatory cytokines.

Once an infection, and therefore cytokine synthesis, resolves over the time course of a few days, behavioural symptoms of sickness normally disappear. However, if the infection does not resolve, or if there is ongoing autoimmune inflammatory processes of a chronic nature, such as rheumatoid arthritis, multiple sclerosis, inflammatory bowel disease, the synthesis of cytokines and their downstream products continues to be elevated. The chronic action of proinflammatory cytokines acting in the brain has now led to the hypothesis that these very same proteins are somehow involved in development of affective disorders such as depression. In the clinic, this hypothesis is supported by the high prevalence of major depressive disorders that is observed in patients with a chronic inflammatory condition[38] and the observation that chronic administration of recombinant IL-2 and/or IFN-α to cancer or hepatitis C patients induces alterations in mood that are characteristic of depression. In the laboratory, acute or chronic activation of the peripheral immune system induces depression-like behaviour that is apparent even when sickness behaviour has dissipated.[39] A leading candidate protein for this effect is a cytokine-activated enzyme known as indoleamine 2,3 dioxygenase (IDO)[40] which has an ubiquitous cellular localization and is also present in the brain. This enzyme degrades tryptophan, an essential amino acid that is required for the synthesis of the mood-regulating neurotransmitter serotonin. Increased degradation of tryptophan associated with inflammation results in a relative deficit in brain serotonin neurotransmission that can precipitate depression and an increased production of tryptophan neuroactive derivatives, including kynurenine, 3-hydroxy kynurenine, and quinolinic acid. These molecules act as agonists or antagonists of the NMDA receptor. Since brain IDO enzymatic activity is increased during activation of the peripheral immune system, these molecules gain access to the brain because they freely cross the blood–brain barrier. In addition, these metabolites are also produced in the brain.

There is accumulating evidence that activation of immune-to-brain communication pathways is responsible not only for the adaptive sickness behavioural response and the maladaptive syndrome of depression that develops in vulnerable individuals but also for the symptom burden that is experienced by physically ill patients, including fatigue[41] (see Fig. 2.3.10.2), pain[42], sleep disorders,[43] and impaired learning and memory.[44]

Concluding comments

A few years ago, scientists who proffered the idea that mood disorders could be induced by an infective process were considered heretic. That view has changed considerably with the discovery that cytokines from the immune system act as elements that permit active communication between the brain and the rest of the body. Indeed, it was in 1984 when a pioneering scientist in psychoneuroimmunology, J. Edwin Blalock, proposed that the immune system

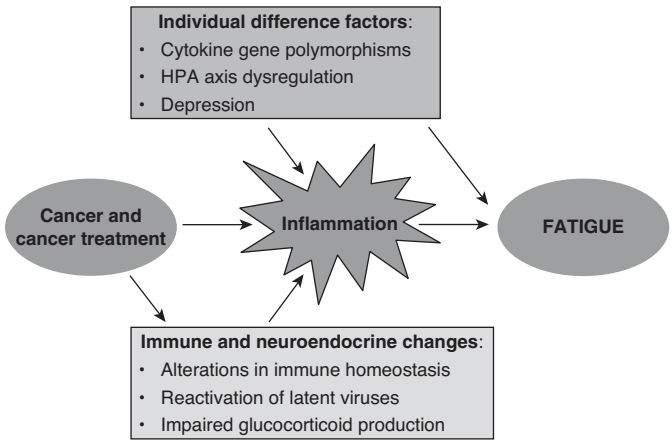


Fig. 2.3.10.2 Potential mechanisms for cancer-related fatigue. Cancer and its treatment can activate the proinflammatory cytokine network, leading to symptoms of fatigue through cytokine effects on the central nervous system. Chronic inflammation may develop when cancer and cancer treatments induce long-term changes in immune homeostasis, including alterations in immune cell subsets, alterations in expression and signaling of Toll-like receptors, and latent virus reactivation. Cancer-related changes in neuroendocrine function may also contribute to chronic inflammation, particularly impairments in glucocorticoid production that result in ineffective control of inflammatory processes. In addition, individual difference factors may increase the risk for chronic inflammation following cancer diagnosis and treatment. Potential risk factors include single nucleotide polymorphisms in cytokine genes, alterations in HPA axis function, and depressive symptoms. Of note, HPA dysregulation and depression may also have direct effects on fatigue. (Reprinted from Brain Behaviour and Immunity, 21(7), JE Bower, Cancer related fatigue, 863–871, Copyright 2007, with permission from Elsevier.)

is really a sixth sensory system.[45] Humans cannot see, smell, touch, hear, or taste pathogenic micro-organisms. However, cells of the immune system, whether they be T cells, B cells, macrophages, microglial cells, or dendritic cells are uniquely endowed with the molecular machinery to detect an endless array of pathogens. One way that the immune system informs the brain that a pathogen has entered the body is by immune cells recognizing these invaders. These leukocytes respond by releasing proinflammatory cytokines, which act as messenger systems to alert the brain that something is amiss in the periphery. The foundation for this conceptual system has strengthened during the last 20 years. It forms the intellectual basis for the notion that emotions are regulated by the immune system, and the immune system affects expression of emotions.

Further information

Ader, R. (2007). *Psychoneuroimmunology*, 2 volumes (4th edn). Elsevier, Amsterdam.

Psychoneuroimmunology Research Society: http://www.pnirs.org

References

1. Ader, R. (2003). Conditioned immunomodulation: research needs and directions. *Brain, Behavior, and Immunity*, **17**(Suppl. 1), S51–7.
2. Felten, D.L. and Felten, S.Y. (1988). Sympathetic noradrenergic innervation of immune organs. *Brain, Behavior, and Immunity*, **2**(4), 293–300.
3. Sanders, V.M. (2006). Interdisciplinary research: noradrenergic regulation of adaptive immunity. *Brain, Behavior, and Immunity*, **20**(1), 1–8.
4. Edwards, C.K. III, Yunger, L.M., Lorence, R.M., *et al.* (1991). The pituitary gland is required for protection against lethal effects of Salmonella typhimurium. *Proceedings of the National Academy of Sciences of the United States of America*, **88**(6), 2274–7.
5. Kelley, K.W., Weigent, D.A., and Kooijman, R. (2007). Protein hormones and immunity. *Brain, Behavior, and Immunity*, **21**(4), 384–92.
6. Edwards, C.K. III, Arkins, S., Yunger, L.M., *et al.* (1992).The macrophage-activating properties of growth hormone. *Cellular and Molecular Neurobiology*, **12**(5), 499–510.
7. Fu, Y.K., Arkins, S., Fuh, G., *et al.* (1992). Growth hormone augments superoxide anion secretion of human neutrophils by binding to the prolactin receptor. *The Journal of Clinical Investigation*, **89**(2), 451–7.
8. Baxter, J.B., Blalock, J.E., and Weigent, D.A. (1991). Expression of immunoreactive growth hormone in leukocytes in vivo. *Journal of Neuroimmunology*, **33**(1), 43–54.
9. Kalantaridou, S., Makrigiannakis, A., Zoumakis, E., *et al.* (2007). Peripheral corticotropin-releasing hormone is produced in the immune and reproductive systems: actions, potential roles and clinical implications. *Frontiers in Bioscience*, **12**, 572–80.
10. Levite, M. (1998). Neuropeptides, by direct interaction with T cells, induce cytokine secretion and break the commitment to a distinct T helper phenotype. *Proceedings of the National Academy of Sciences of the United States of America*, **95**(21), 12544–9.
11. Borovikova, L.V., Ivanova, S., Zhang, M., *et al.* (2000). Vagus nerve stimulation attenuates the systemic inflammatory response to endotoxin. *Nature*, **405**(6785), 458–62.
12. Pavlov, V.A. and Tracey, K.J. (2005). The cholinergic anti-inflammatory pathway. *Brain, Behavior, and Immunity*, **19**(6), 493–9.
13. Forni, G., Bindoni, M., Santoni, A., *et al.* (1983). Radiofrequency destruction of the tuberoinfundibular region of hypothalamus permanently abrogates NK cell activity in mice. *Nature*, **306**(5939), 181–4.
14. Renoux, G., Biziere, K., Renoux, M., *et al.* (1983). A balanced brain asymmetry modulates T cell-mediated events. *Journal of Neuroimmunology*, **5**(3), 227–38.
15. Neveu, P.J., Barneoud, P., Vitiello, S., *et al.* (1988). Brain modulation of the immune system: association between lymphocyte responsiveness and paw preference in mice. *Brain Research*, **457**(2), 392–4.
16. Sklar, L.S. and Anisman, H. (1979). Stress and coping factors influence tumor growth. *Science*, **205**(4405), 513–5.
17. Visintainer, M.A., Volpicelli, J.R., and Seligman, M.E. (1982). Tumor rejection in rats after inescapable or escapable shock. *Science*, **216**(4544), 437–9.
18. Laudenslager, M.L., Ryan, S.M., Drugan, R.C., *et al.* (1983). Coping and immunosuppression: inescapable but not escapable shock suppresses lymphocyte proliferation. *Science*, **221**(4610), 568–70.
19. Mormede, P., Dantzer, R., Michaud, B., *et al.* (1988). Influence of stressor predictability and behavioral control on lymphocyte reactivity, antibody responses and neuroendocrine activation in rats. *Physiology & Behavior*, **43**(5), 577–83.
20. Cocke, R., Moynihan, J.A., Cohen, N., *et al.* (1993). Exposure to conspecific alarm chemosignals alters immune responses in BALB/c mice. *Brain, Behavior, and Immunity*, **7**(1), 36–46.
21. Kiecolt-Glaser, J.K., Glaser, R., Strain, E.C., *et al.* (1986). Modulation of cellular immunity in medical students. *Journal of Behavioral Medicine*, **9**(1), 5–21.
22. Kiecolt-Glaser, J.K., McGuire, L., Robles, T.F., *et al.* (2002). Psychoneuroimmunology and psychosomatic medicine: back to the future. *Psychosomatic Medicine*, **64**(1), 15–28.
23. Cobb, J.M. and Steptoe, A. (1996). Psychosocial stress and susceptibility to upper respiratory tract illness in an adult population sample. *Psychosomatic Medicine*, **58**(5), 404–12.
24. Herbert, T.B. and Cohen, S. (1993). Depression and immunity: a meta-analytic review. *Psychological Bulletin*, **113**(3), 472–86.

25. Zorrilla, E.P., Luborsky, L., McKay, J.R., *et al.* (2001). The relationship of depression and stressors to immunological assays: a meta-analytic review. *Brain, Behavior, and Immunity*, **15**(3), 199–226.
26. Maes, M., Smith, R., and Scharpe, S. (1995). The monocyte-T-lymphocyte hypothesis of major depression. *Psychoneuroendocrinology*, **20**(2), 111–16.
27. Marsland, A.L., Cohen, S., Rabin, B.S., *et al.* (2006). Trait positive affect and antibody response to hepatitis B vaccination. *Brain, Behavior, and Immunity*, **20**(3), 261–9.
28. Segerstrom, S.C. (2005). Optimism and immunity: do positive thoughts always lead to positive effects? *Brain, Behavior, and Immunity*, **19**(3), 195–200.
29. Keller, S.E., Weiss, J.M., Schleifer, S.J., *et al.* (1983). Stress-induced suppression of immunity in adrenalectomized rats. *Science*, **221**(4617), 1301–4.
30. Lysle, D.T., Cunnick, J.E., and Maslonek, K.A. (1991). Pharmacological manipulation of immune alterations induced by an aversive conditioned stimulus: evidence for a beta-adrenergic receptor-mediated Pavlovian conditioning process. *Behavioral Neuroscience*, **105**(3), 443–9.
31. Cohen, S. and Williamson, G.M. (1991). Stress and infectious disease in humans. *Psychological Bulletin*, **109**(1), 5–24.
32. Lutgendorf, S.K. and Costanzo, E.S. (2003). Psychoneuroimmunology and health psychology: an integrative model. *Brain, Behavior, and Immunity*, **17**(4), 225–32.
33. Janicki-Deverts, D., Cohen, S., Doyle, W.J., *et al.* (2007). Infection-induced proinflammatory cytokines are associated with decreases in positive affect, but not increases in negative affect. *Brain, Behavior, and Immunity*, **21**(3), 301–7.
34. Kent, S., Bluthe, R.M., Kelley, K.W., *et al.* (1992). Sickness behavior as a new target for drug development. *Trends in Pharmacological Sciences*, **13**(1), 24–8.
35. Konsman, J.P., Parnet, P., and Dantzer, R. (2002). Cytokine-induced sickness behaviour: mechanisms and implications. *Trends in Neurosciences*, **25**(3), 154–9.
36. Dantzer, R. (2001). Cytokine-induced sickness behavior: mechanisms and implications. *Annals of the New York Academy of Sciences*, **933**, 222–34.
37. Dantzer, R. and Kelley, K.W. (2007). Twenty years of research on cytokine-induced sickness behavior. *Brain, Behavior, and Immunity*, **21**(2), 153–60.
38. Steptoe, A. (ed.) (2007). *Depression and physical illness*. Cambridge University Press, Cambridge.
39. Frenois, F., Moreau, M., O'Connor, J., *et al.* (2007). Lipopolysaccharide induces delayed FosB/DeltaFosB immunostaining within the mouse extended amygdala, hippocampus and hypothalamus, that parallel the expression of depressive-like behavior. *Psychoneuroendocrinology*, **32**(5), 516–31.
40. Capuron, L. and Dantzer, R. (2003).Cytokines and depression: the need for a new paradigm. *Brain, Behavior, and Immunity*, **17**(Suppl. 1), S119–24.
41. Bower, J.E. (2007). Cancer-related fatigue: links with inflammation in cancer patients and survivors. *Brain, Behavior, and Immunity*, **21**(7), 863–71.
42. Wieseler-Frank, J., Maier, S.F., and Watkins, L.R. (2005). Immune-to-brain communication dynamically modulates pain: physiological and pathological consequences. *Brain, Behavior, and Immunity*, **19**(2), 104–11.
43. Irwin, M. (2002). Effects of sleep and sleep loss on immunity and cytokines. *Brain, Behavior, and Immunity*, **16**(5), 503–12.
44. Krabbe, K.S., Reichenberg, A., Yirmiya, R., *et al.* (2005). Low-dose endotoxemia and human neuropsychological functions. *Brain, Behavior, and Immunity*, **19**(5), 453–60.
45. Blalock, J.E. (1984).The immune system as a sensory organ. *Journal of Immunology*, **132**(3), 1067–70.

2.4

The contribution of genetics

Contents

2.4.1 Quantitative genetics

Anita Thapar and Peter McGuffin

Patterns of inheritance

Our understanding of how traits and disorders are passed from one generation to the next began with the work carried out by an Augustinian monk, Gregor Mendel. Although Mendel's published work in 1866 was initially ignored, its rediscovery at the beginning of the twentieth century heralded the beginning of modern genetics. Mendel's experiments on pea plants and his observations of the patterns of inheritance of certain characteristics led to the development of his particulate theory of inheritance. It was only later in 1909 that the units of inheritance that he had described were named genes and alternative forms of a gene were termed alleles. It was also at this time that the terms phenotype, used to describe the observed characteristic, and genotype, used to refer to the genetic endowment, were introduced.

Mendel's laws

Mendel examined clear-cut dichotomous characteristics such as smooth versus wrinkled coats in peas. He first noted that when parents of different types were crossed, the first generation (F1) offspring displayed **uniformity** of that characteristic. He inferred that this uniformity was due to one phenotype being dominant and the other being recessive. Thus, when homozygous parents AA and aa produced heterozygote Aa offspring, these offspring displayed the phenotype of the AA parent rather than manifesting a phenotype intermediate to those of both parents.

Mendel then demonstrated that when the F1 heterozygotes (Aa) were intercrossed (Aa × Aa), **segregation** resulted in the second F2 generation showing recessive and dominant phenotypes in the ratio of 1 to 3. He then inferred that this F2 generation consisted of three types (AA, Aa, and aA, aa) occurring with a probability of 1:2:1.

Finally, Mendel showed that when the transmission of two different phenotypic traits was studied, they showed **independent assortment**. We now know that independent assortment occurs when the genes coding for these traits are either located far apart on the same chromosome or are on different chromosomes (see linkage).

Single-gene disorders

Although disorders showing a simple Mendelian pattern of inheritance are rare, they tend to be clinically severe and collectively impose a significant burden.

For **autosomal dominant** disorders to manifest themselves, only one disease allele is necessary, i.e. both heterozygotes as well as homozygotes (those who carry both disease genes) will be affected. In most instances, where there is one affected parent who is a heterozygote for the disease, approximately 50 per cent of the offspring will show the disorder. Autosomal disorders tend to be severe and manifest themselves in every generation. Huntington's disease and acute intermittent porphyria are examples of autosomal dominant conditions that are often present with psychiatric symptoms.

Autosomal recessive conditions such as phenylketonuria require the presence of two disease alleles to show clinical manifestations of the disorder. Thus, they often appear to skip generations. These disorders usually occur in the offspring of two 'carrier' heterozygote parents and are more common where there is a high rate of consanguinity (e.g. marriages between cousins) as these inbred populations will show greater homozygosity at all loci.

The other group of single-gene disorders consists of **sex-linked** conditions such as fragile X syndrome. Normal females have two X chromosomes whereas normal males have one X chromosome and one Y chromosome. Thus, for recessive disorders on the X chromosome, if the mother is a carrier (X*X) and assuming that the father is unaffected (XY), half of her sons will manifest the disorder (X*Y) and half of her daughters will be carriers (X*X). Where the father is affected by an X-linked recessive condition, all the daughters will be carriers. As sons have to inherit their

X chromosome from their mother, there will be an absence of father to son transmission. X-linked dominant conditions are extremely rare.

Continuous traits

Mendel's laws are based on the transmission of dichotomous characteristics, yet many important human phenotypes such as height, weight, and blood pressure are continuously distributed. However, we are able to show that Mendelian principles can also be applied for these types of quantitative traits.

Let us first consider a phenotype measured on a continuous scale which results from the influence of a single gene with two alleles A_1 and A_2 (see Fig. 2.4.1.1). We can now describe the phenotypes of the three possible genotypes in terms of a quantitative value on the continuous scale. A_1A_1 has a value of $-a$; A_2A_2 has a value of $+a$; and A_1A_2, the heterozygote, has a value of d. When $d = 0$, A_1A_2 lies exactly half way between A_1A_1 and A_2A_2, that is the genetic contribution is entirely additive. When $d = -a$, A_2 is recessive to $A_1$and when $d = +a$, A_2 is dominant to A_1.

At the simplest level, we assume that there are no dominance effects and that there is no mutation, selection, migration, or inbreeding in the population. If p is the frequency of allele $A_1$and q is the frequency of A_2 in the population where $p + q = 1$ then the frequency of genotypes can be expressed as follows:

$$\begin{matrix} A_1A_1 & A_1A_2 & A_2A_2 \\ p^2 & 2pq & q^2 \end{matrix}$$

This is known as the Hardy–Weinberg equilibrium. If we now simplify further and state allelic frequencies where $p = q = 0.5$, then the phenotypic values of A_1A_1, A_1A_2, and A_2A_2 would be distributed in the population with relative frequencies of 1:2:1.

Now if we consider a trait which results from two genes each of which has two alleles of equal frequency and additive effect, there would be five possible phenotypic values with relative frequencies of 1:4:6:4:1. Overall as the number of genetic loci (n) increases, the number of phenotypic values increases ($2n + 1$) and the distribution of phenotypic values more closely approximates a normal distribution. It is thought that most quantitative or continuous traits result from the additive action of genes at many loci which is otherwise known as **polygenic** inheritance. Where familial transmission is explained by environmental factors as well as by multiple genes, we then call this a **multifactorial** mode of inheritance.

Complex disorders and irregular phenotypes

(a) Polygenic/multifactorial threshold models

Most common human psychiatric and medical disorders such as schizophrenia, diabetes, and heart disease do not show a Mendelian pattern of inheritance. Neither can they be considered as continuous traits in that people are described as being affected or unaffected.

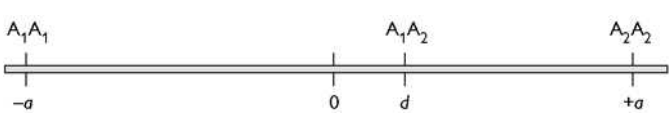


Fig. 2.4.1.1 A phenotype, measured on a continuous scale, resulting from a single gene with two alleles A_1 and A_2.

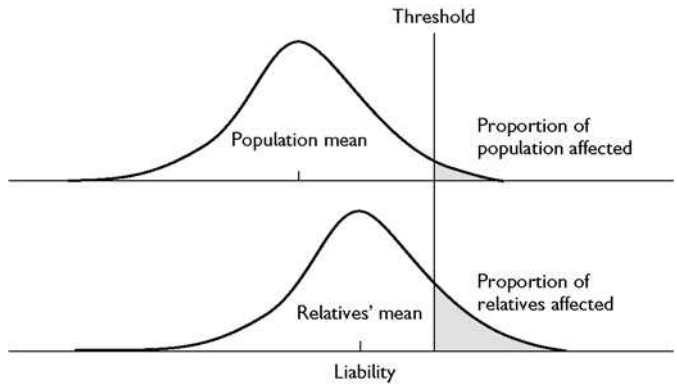


Fig. 2.4.1.2 A polygenic or multifactorial threshold model of disease transmission.

However, these conditions could be regarded as quasi-continuous in that those who are affected can be graded along a continuum of severity. It is possible to extend this to assume that there is an underlying liability to develop the disorder which is continuously distributed in the population. Those who pass a certain threshold manifest the condition. If the underlying liability to develop the disorder is inherited in a polygenic or multifactorial fashion, then we can assume that the distribution will be approximately normally distributed (Fig. 2.4.1.2). The genetic liability of relatives of affected individuals will be increased and their liability distribution will be shifted to the right (Fig. 2.4.1.2). Thus, the proportion of relatives above the disease threshold will be greater compared with the general population. If we know the proportion of affected relatives of probands and the proportion of those affected in the general population, it is possible to calculate the correlation in liability between pairs of relatives using this type of model.

(b) Single major locus model and atypical patterns of Mendelian inheritance

An alternative to a polygenic model of complex disease is a single major locus model. Single-gene disorders do not always show typical Mendelian patterns of inheritance. For example familial transmission can be modified by variable **expression** and **penetrance**. Some conditions can show great variability in terms of clinical expression. For example neurofibromatosis, an autosomal dominant disorder can express itself as the full blown disorder or merely as a few café-au-lait spots. Penetrance is defined as the probability of manifesting the disorder given a particular genotype. For Mendelian disorders this is always 1 or 0, but irregular patterns of inheritance may occur because of incomplete penetrance where the probability of manifesting the disorder is greater than 0 but less than 1.

Finally, there are now molecular explanations for other types of unusual patterns of inheritance for single genes. Anticipation where disorders show a progressively earlier onset and greater severity with subsequent generations is now known to be explained by heritable unstable nucleotide repeat sequences (see later). Huntington's chorea and fragile X syndrome are examples of disorders caused by heritable unstable repeats. For some conditions such as the Angelman syndrome and the Prader–Willi syndrome, manifestation of the disorder depends on the parental origin of the gene. This is known as imprinting.

(c) Other models

Alternative explanations of how complex conditions such as psychiatric disorders are inherited include mixed and oligogenic patterns of transmission. A mixed model includes a major gene and a polygenic/multifactorial contribution. However, for many of these disorders, genes of major effect may not exist. It may be that these irregular phenotypes are best explained by oligogenic models where the co-action or interaction of a small number of genes contributes to the disorder. These issues remain to be resolved by molecular genetic studies (see later).

Components of phenotypic variation

We will now consider the different influences that contribute to phenotypic variation in a population. The total variation in an observed trait (phenotype v_p) at the simplest level (ignoring non-additive effects) can be partitioned into a proportion due to genetic influences (v_g), a component explained by shared environmental factors (v_c), and a remainder accounted by non-shared environmental factors which includes error (v_e):

$$v_p = v_g + v_c + v_e$$

Shared or common environmental influences are aspects of the environment that result in greater similarity of family members for a given phenotype. Non-shared environmental factors refer to environmental influences that have effects which are specific to individuals and that contribute to phenotypic differences between family members.

Although we have so far only considered one type of genetic contribution, the genetic variance v_g can be further subdivided into variance due to additive genetic influences (v_a) and dominance effects (v_d).

The relative influence of genetic factors is expressed as **heritability** and when defined as the proportion of the total phenotypic variance attributable to additive genetic variance, is known as **narrow-sense heritability**:

$$h_n^2 = v_a / v_p.$$

Heritability is also sometimes used to describe the proportion of variance explained by the *total* genetic variance (additive and non-additive genetic variance) and it is then known as **broad-sense heritability**:

$$h_b^2 = v_a / v_g.$$

Similarly we can estimate the proportion of the total phenotypic variance explained by shared environment where $c^2 = v_c/v_p$ and the remaining proportion attributable to non-shared environmental factors and error (e^2).

It is important to remember that the estimate of heritability and the contribution of shared environment and non-shared environment are proportions of total variation within a given population, i.e. these parameters tell us about sources of difference between individuals in a population and have no meaning at an individual level. For example, if an individual was selected from a population where IQ had been shown to have a heritability of 50 per cent, it could not be said that 50 per cent of that individual's IQ was determined by genes. Another important point is that these estimates are specific to the population studied and may differ for other populations. Finally, the contribution of genetic and environmental influences to a phenotype does not allow any inferences about the extent to which that phenotype is modifiable by environmental factors. For example, phenylketonuria is a Mendelian condition that is determined by the presence of a single-gene mutation. Yet the clinical manifestations of the syndrome are prevented by dietary intervention.

Non-additive genetic effects

So far we have simplistically assumed that phenotypic variation is influenced in an additive fashion. However, the contribution of genes and environment is more complex than this. We have already referred to genetic **dominance** effects where there is non-additive interaction of alleles within a locus. Another potential source of influence is the non-additive interaction between alleles at different loci which is known as gene–gene interaction or **epistasis**.

Gene–environment interaction

Gene–environment interplay represents another important form of non-additive genetic contribution to complex phenotypes.[1] The term gene–environment interaction (G × E) is used here to refer to individual genetic differences in response to specific environmental factors. In the presence of gene–environment interaction, individuals who are at genetic risk of a disorder do not manifest the condition unless they are exposed to a specific environmental risk factor. Gene–environment interaction also means that not all those exposed to an environmental risk factor will show disorder. Later, we consider direct investigation of gene–environment interaction through molecular genetic studies. Twin and adoption study designs have also been used to examine G × E, in an indirect way. Here, genetic liability is inferred by virtue of having affected relatives rather than through possession of a specific genetic risk variant.

Gene–environment correlation

Gene–environment correlation further adds to the complexity of interplay between genes and environment. Gene–environment correlation arises when a person's genotype is correlated with the environment that they are exposed to. For example, sociable parents not only endow their children with genes but also provide an environment that encourages greater sociability in their children (passive gene–environment correlation). Moreover, positive gene–environment correlation would result where a sociable child actively seeks out more situations where socializing occurs (active gene–environment correlation) or where he or she evokes friendly responses in others (evocative gene–environment correlation). There is evidence that many important environmental risk factors in psychiatry (for example, life events) do correlate with genetic risk for specific disorders (for example, depression). Where that is the case, genetically sensitive designs are needed to investigate whether environmental risk factors have true environmentally mediated risk effects on disorder or whether the association has arisen because of genetic factors contributing to both the environmental risk exposure and disorder.

The presence of gene–environment interaction and gene–environment correlation highlights that the action of genes and environment must be considered together. Another important point is that in traditional twin study designs, G × E and G − E correlation effects are subsumed within the heritability estimate or in some circumstances the environmental variance component (see twin studies).

Research methods

Family, twin, and adoption studies

So far we have considered the theoretical basis of inheritance and possible sources of phenotypic variation and familial resemblance. Clearly, the investigation of the genetic basis of psychiatric disorders first requires us to examine to what extent genes and environment contribute to a given disorder or trait. Secondly, we need to know how genes and environmental influences exert their risk effects and finally we have to investigate the genetic basis of disorders at a molecular level.

Traditional methods in psychiatric genetics research include family, twin, and adoption studies. Family studies enable us to examine to what extent a disorder or trait aggregates in families. Familiality of a disorder can of course by explained by shared environmental influences as well as by shared genes. Twin and adoption studies allow us to disentangle the effects of genes and shared environment.

Family studies

(a) Methods

Family studies allow us to determine whether a disorder aggregates in families by examining the rate of disorder in the relatives of affected individuals (probands) and comparing this with the rate of disorder in the general population or in a control group. Alternatively we can compare the frequency of disorder in the relatives of probands with the frequency among relatives of a control group of normal individuals or those with another disorder.

There are two types of family studies. The **family history** method is more economical in that the psychiatric history is taken from the proband. However, given that most individuals are unlikely to know as much about family members as about themselves, this method results in an underestimate of diagnoses in relatives. A more thorough but more time-consuming approach is the **family study** method where all available relatives are directly interviewed.

(b) Ascertainment

An important issue is how a family study sample is ascertained. Ideally, probands should be ascertained independently from each other. This is unlikely to pose a problem for rare disorders. However, for more common conditions where a series of cases is collected, for example, by consecutive referrals of the disorder to a particular hospital, it is possible that families included in a family study contain more than one proband. This is known as multiple incomplete ascertainment. Complete ascertainment, where all affected individuals in a given population are included, is rarely possible and in most instances probands are identified after some selection process (e.g. referrals to a particular hospital). Thus, factors influencing selection, such as comorbidity and help seeking may also influence observed patterns of familial aggregation.

(c) Age correction

For genetic studies, we are interested in the proportion of individuals who have ever had the disorder (lifetime prevalence) rather than the proportion who show the disorder at one point in time (point prevalence). However, a difficulty encountered when carrying out family studies is that the observed rates of disorder will also depend on the age of the individual, the risk period for the disorder, and whether or not the individual has lived through that risk period. Thus, some members may not yet have reached the age of risk for the disorder, some are currently unaffected but will become affected at some later point, and others may have died whilst still unaffected. The most appropriate method is to correct for age and express the rate of disorder in relatives as the **morbid risk (MR)** or lifetime expectancy.

There are many methods of age correction, of which the Slater-Stromgen adaption of Weinberg's shorter method is the most straightforward. The MR of the disorder can be estimated as the number of affecteds (A) divided by the *bezugsziffer* (BZ) where the BZ is calculated as:

$$\sum_{i}[a1]w_i + A$$

and where w_i is the weight given to the ith unaffected individual based on their current age. The most accurate approach is to use an empirical age of onset distribution from a large separate sample, for example a national registry of psychiatric disorders, to obtain the cumulative frequency of disorder over a range of age bands, from which weights can be derived.

Another approach is to carry out life table analysis. The distribution of survival times (or times to becoming ill) is divided into a number of intervals. For each of these one can calculate the number and proportion of subjects who entered the interval unaffected and the number and proportion of cases that became affected during that interval as well as the number of cases that were lost to follow-up (because they had died or had otherwise 'disappeared from view'). Based on these, the numbers and proportions 'failing' or becoming ill over a certain time interval (usually taken as the entire period of risk) can be calculated. A further alternative is to use a Kaplan Meier product limit estimator. This allows one to estimate the survival function directly from continuous survival or failure times instead of classifying observed survival times into a life table. Effectively this means creating a life table in which each time interval contains exactly one case. It therefore has an advantage over a life table method in that the results do not depend on grouping of the data.

Twin studies

Identical or monozygotic twins, by virtue of arising from the fertilization of one egg, share 100 per cent of their genes. Non-identical or dizygotic twins are from two fertilized eggs and like full biological siblings share on average 50 per cent of their genes. Thus, assuming that monozygotic twins and dizygotic twins share environment to the same extent, monozygotic twins would share greater similarity than dizygotic twins for a disorder that is genetically influenced. Twin studies are an important method for disentangling the effects of genes and shared environment and can be used to estimate the contribution of genetic influences, shared environmental factors and non-shared environmental factors to the total variation for a given trait or disorder.

For continuous traits, twin similarity is expressed as an intraclass correlation coefficient where:

$$r_{mz} = h^2 + c^2$$

$$r_{dz} = 0.5h^2 + c^2$$

Thus, from observed monozygotic and dizygotic correlations for a given trait we can calculate heritability from the above equations where $h^2 = 2(r_{mz} - r_{dz})$, $c^2 = 2r_{dz} - r_{mz}$, and e^2 is the remaining variance $= 1 - h^2 - c^2$ (see path analysis below).

For dichotomous characteristics (e.g. affected with a disorder and unaffected), twin similarity is expressed as concordance rates. A **pairwise concordance rate** is estimated as the number of twin pairs who both have the disorder divided by the total number of pairs. However, where there has been systematic ascertainment, for example a twin register, it is preferable to report a **probandwise concordance rate** which is calculated as the number of affected twins divided by the total number of cotwins.

(a) Ascertainment

One potential source of bias in twin studies stems from ascertainment procedures. For example, affected twins referred to a specific study or volunteer samples are likely to include more twin pairs who are monozygotic and who are concordant. Ascertainment of twin pairs through hospital registers overcomes this problem to some extent, but for some disorders may be biased by the process of referral. Population-based samples overcome these biases, although when examining disorders rather than traits extremely large sample sizes are required to obtain an adequate number of affected individuals.

(b) Zygosity

A further potential source of error is in the assignment of zygosity. Ideally zygosity should be determined by DNA typing. However, it may be more practical to use a twin similarity questionnaire which includes questions such as whether the twins share the same hair/eye colour, and whether they look alike as two peas in a pod. This method of assigning zygosity is simple and inexpensive with a reported accuracy of over 90 per cent.

(c) Equal environments assumption

It is sometimes argued that a major drawback to the twin study method is that monozygotic twins may experience a more similar environment and may be treated more similarly than dizygotic twins. However, where there is evidence that monozygotic twins share greater environmental similarity than dizygotic twins it is difficult to infer whether this contributes to their similarity for the disorder or whether this is the consequence of greater genetic similarity. There have been several approaches adopted to further explore this issue.

In some studies questionnaire measures of environmental sharing (e.g. being dressed alike as children, sharing friends) have been used. These suggest that environmental sharing is indeed greater for monozygotic twins than for dizygotic twins. However, it appears that for many traits and disorders such as cognitive ability, personality, depressive symptoms, and depressive disorder this degree of similarity for childhood environment does not account for monozygotic twin similarity for the trait. One way of disentangling cause and effect is to use direct observational studies. Although this method has not been much used, one study of young twins suggested that the greater similarity of parental responses to monozygotic twins compared to dizygotic twins appeared to be elicited by the twins themselves.

An alternative method of examining the effects of environmental sharing is to study twins who are mistaken about their zygosity. However, most studies which have used this method suggest that perceived zygosity is a less important influence on twin similarity than true zygosity.

Finally, the most powerful means of examining the effects of environmental sharing is to look at monozygotic twins who have been reared apart. However, such twin pairs are rare and have mostly been ascertained in a biased fashion. Nevertheless, studies of reared-apart twins have informed us that there is a substantial genetic contribution to cognition, personality, and psychosis.

(d) Comparability of twins

The final potential criticism of the twin method is whether twins can be regarded as representative of the general population given some important differences. Twin births are relatively common (1 in 80 births), although the number of dizygotic twins varies in different countries and is influenced by factors such as maternal age and multiparity, a family history of twins and increasingly, the use of fertility drugs. Twins are more likely to experience greater intrauterine and perinatal adversity and the experience of being brought up as a twin is unusual in itself. There is also some evidence that depression is more common in mothers of young twins than among mothers of singletons. However, these differences are only important if they result in different rates of disorder or symptoms in twins compared to singletons. So far there is little evidence to suggest that the rate of psychiatric disorder in twins is any higher than amongst singletons.

Adoption studies

Adoption studies provide another means of teasing apart the effects of genes and environment. The basic method of the adoption study lies in comparing the rates of disorder in biological relatives and adoptive relatives. There are three main types of adoption study.

1 **The adoptee study**: Here the rate of disorder in the adopted-away offspring of affected individuals is compared with the rate of disorder in control adoptees whose biological parents are unaffected.

2 **The adoptee's family study**: In this design, the rate of disorder in the biological relatives of affected adoptees is compared with that among the adopted relatives.

3 **The cross-fostering study**: This allows us to examine gene-environment interaction by comparing the rate of disorder in adoptees who have unaffected biological parents and affected adoptive parents with the rate of disorder in adoptees who have affected biological parents and unaffected adoptive parents.

Although adoption studies allow us to examine the effects of both genes and environment, there are several potential drawbacks to the method. First, adoption is in itself an unusual event and there is a tendency for higher rates of some psychiatric difficulties such as antisocial personality traits amongst adoptees. Second, adoptive placements are not random in that adoption agencies are likely to attempt to match adoptive and biological parents for physical, social, and other characteristics. Nevertheless, despite these difficulties, adoption evidence has given much support to the role of both genes and environment for traits and behaviours such as cognitive ability and criminality.

Methods of analysis

Although the statistical methods used in quantitative genetics may seem complex, the principles are straightforward. We will now consider the methods of analyses that are most commonly used for examining the contribution of genetic and environmental factors, to psychiatric disorders and traits.

Path analysis

Path analysis provides a simple diagrammatic method of estimating the contribution of genetic and environmental factors. The basic path diagram in Fig. 2.4.1.3 shows the sources of resemblance for phenotypes (P_1 and P_2) in a pair of siblings. Using the rules of path analysis, the correlation between the siblings (r_{sib}) is derived by multiplying the path coefficients for each connecting path and then summing these coefficients. Thus

$$r_{sib} = h \times r_g \times h + c \times c.$$

The genetic correlation r_g is 1 for monozygotic twins who share 100 per cent of their genes in common and is 0.5 for full siblings or dizygotic twins. Thus, using path analysis we obtain the equations described earlier where

$$r_{mz} = h^2 + c^2$$
$$r_{dz} = {}^1/_2h^2 + c^2$$

Model fitting

Although we can simply estimate h^2, c^2, and e^2 from the equations above for more complex data, solving multiple linear equations becomes difficult. We may also wish to test alternative models, for example one where there is no genetic contribution or one where shared environmental influences are dropped. Model fitting allows us to first statistically test how well a given model explains the observed data and to then compare different models.

Computer packages such as Mx[2] are all based on the same principles. The raw data are read into the program and the researcher supplies the initial starting values for the unknown parameters (h, c, and e for a full genetic model). The program then iterates with different parameter estimates until values are found which give an optimum fit (usually this involves maximizing a likelihood function or minimizing a χ^2). The goodness of fit of the model is then assessed by examining the χ^2 goodness of fit where a smaller value indicates a better fit.

The fit of a reduced model (R) can then be compared against the full model (F) by subtracting the χ^2 values ($R - F$). Alternatively the fit of models can be compared by using the likelihood ratio test where twice the difference between the log likelihoods for each model (this approximates a χ^2 distribution) is calculated.

(a) Application of model fitting

So far we have considered the influence of genes and environment on variation in a single phenotype using a traditional twin design.

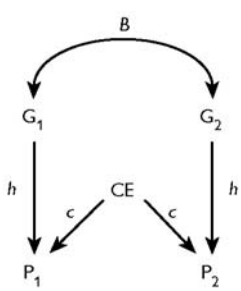


Fig. 2.4.1.3 A single path model of the sources of resemblance between twins or pairs of siblings. G_1 and G_2 are genotypes with correlation B, CE is common environment, P_1 and P_2 are phenotypes, and h and c are path coefficients.

This type of analysis is known as univariate genetic analysis. However the most interesting questions in psychiatry are best addressed by more testing more complex models (See Kendler and Prescott[3] for excellent examples). Designs that involve measuring the same phenotype over time and several phenotypes at the same time (e.g. depression and anxiety) have been used to investigate comorbidity, the aetiological contribution to developmental changes over time and diagnostic boundaries. Studies that have employed such methods have yielded interesting findings. For example, there is evidence that the same set of genes, but different non-shared environmental factors influence anxiety and depressive disorders (and symptoms). The contribution of genetic factors to depressive symptoms and IQ has been found to increase in adolescence compared to childhood. Family and twin studies of autism suggest that there is familial and genetic loading for a broader cognitive and social phenotype in relatives of those with narrowly defined autism.

There have also been imaginative extensions of the twin design that have allowed investigation of different questions. Each type of design has its own set of strengths and weaknesses. Thus having a variety of available methods is invaluable. Examples of different but related designs include studies of the children of twins,[4] studies combining twins, siblings, half siblings, step siblings, and families of twins.[5]

(b) Multiple regression analysis

Another commonly used method of analysing twin data is multiple regression analyses.[6] Here the score of the co-twin C is predicted by the score of the proband twin P, the coefficient of the relationship or zygosity R and an interaction term PR. The partial regression coefficients provide direct estimates of heritability and shared environment. The advantage of this method is that it can then also be used to test whether the magnitude of genetic contribution of extreme scores for a continuous trait differs from scores within the normal range. For example, such analyses suggest that the contribution of genetic factors to variation in very low IQ scores is low, unlike the contribution to IQ variation across the 'normal range'.

Twin and adoption study methods for investigating gene–environment interplay

Traditional twin study designs and analysis of MZ twin similarity and difference that include both measured aspects of the environment (e.g. reported life events) and phenotype (e.g. depression) can be used to test the extent to which the association of environmental factors with psychopathology is mediated by genetic and environmental pathways (gene–environment correlation). For example, one twin study showed environmentally mediated risk effects of childhood maltreatment on antisocial behaviour but different findings for corporal punishment where there were genetically mediated effects.[7] Twin studies have also been used to demonstrate the importance of G × E. Kendler and colleagues (1995)[8] showed that those at higher genetic risk of a major depressive disorder (inferred by lifetime ever diagnosis of major depression in MZ and DZ co-twins) were more sensitive to the depressogenic effects of adverse life events.

Adoption study designs provide a useful method for testing gene–environment interaction as post-natal environmental risk

factors (apart from the very earliest) are experimentally separated from genetic risk. Such studies have shown evidence that the risk of antisocial behaviour is much higher in those who are not only at higher genetic risk (by virtue of having a parent with antisocial behaviour) but who are also exposed to adverse rearing environments in the adoptive home.

Gene mapping

A more direct approach to locating and identifying genes involved in psychiatric disorders is to attempt to map them. Gene mapping technology has advanced at an astonishing pace over the past 20 years. Early studies in psychiatric disorders such as schizophrenia[9] had to rely purely on 'classical' genetic markers such as HLA antigens blood groups and various protein polymorphisms. Capabilities in systematic gene mapping, which involves mounting a search throughout the whole genome (i.e. the 22 pairs of autosomes and the sex chromosomes), only became possible after the discovery of markers based on variation in DNA length. The first of these were the restriction fragment length polymorphisms and these have largely been supplanted by single nucleotide polymorphisms (SNPs). Informative combinations of markers are known as haplotypes. The Human Genome Project has led to a detailed map of common variation in the human genome. This public database of more than one million SNPs[10] provides information about the patterns of linkage disequilibrium between these markers that can guide selection of 'tagging SNPs' for gene mapping studies. Such marker maps allow the genes contributing to traits or diseases to be located. The methods for mapping are linkage and association. Gene mapping is discussed further in Section 2.4.2.

Gene regulation

The human genome is much more complex than previously thought.[11] Gene activity and expression (the process by which proteins are made) is regulated by inherited DNA and RNA, non-inherited mechanisms (epigenetics), endogenous biological factors such as hormones and environmental factors operating externally to the individual organism (e.g. toxins, psychosocial stress).

For example, non-coding sequences of DNA were previously thought to be 'junk DNA'. There is increasing evidence that this is not the case and that such regions may play an important role in regulating gene activity. Gene activity is also altered by environmental factors.[11] Animal studies have now shown that environmental factors can alter the genome through measurable biological mechanisms, that these changes can be passed onto the next generation and be subject to modification. For example, maternal care giving behaviour early in life, notably high levels of licking and grooming of rat pups, has an effect on brain glucocorticoid receptor expression and sensitivity to stress in the offspring.[12] This effect is observable even when non-genetically related pups are fostered to mothers who have low levels of licking. This 'programming effect' appears to be mediated by structural modifications of DNA. The molecular mechanisms explaining this 'epigenetic marking' are beginning to be identified and include DNA methylation and histone acetylation. These non-inherited effects persist over two generations and in animals appear to be possibly reversible.

Linkage

In linkage studies, rather than just studying the segregation of a disease in families, the co-segregation of the disease and a set of genetic markers is investigated. The aims are, first, to detect linkage, indicated by the disorder and the marker co-occurring more often than would be expected by chance (i.e. not showing Mendelian independent assortment) and, second, to estimate the distance between a linked marker and the gene conferring to susceptibility to the disorder.

It is possible to detect linkage only in families containing at least one parent who is a double heterozygote (i.e. heterozygous at both the marker and the disease loci). Technically such families are referred to as double back-cross or intercross/intercross matings but, for simplicity, we will just focus on the double back-cross type (Table 2.4.1.1). The table shows a double heterozygote parent where the alleles *A* and *B* are on one chromosome with *a* and *b* on the other. Consequently offspring of the types *aaBb* or *Aabb* result from recombination or crossing over between the homologous pair of chromosomes during meiosis. These types of offspring are called **recombinants**. Offspring of the same type as the parents (i.e. *AaBb aabb*) are non-recombinant. We can then simply define the recombination fraction, θ, as the number of recombinants divided by total number of offspring.

For two loci that are very widely separated on the same chromosome (and all pairs of loci carried on to two different chromosomes) independent assortment occurs and θ = 1/2. When the two loci are close together dependent assortment may be observed indicated by a recombination fraction of less than a half. The size of the recombination fraction depends on the physical distance between the two loci and (within certain limits) is proportional to it, so that for loci that are very close together recombination rarely occurs and θ tends to zero. Genetic distances estimated by linkage studies are measured in centimorgans (cM) with 1 cM the equivalent of a recombination fraction of 0.01. For reasons that are not fully understood, recombination occurs more frequently in female than in male meioses. Hence, the size of the female human genome expressed in centimorgans is larger than the male genome. The sex-averaged size of the human genome is about 3700 cM. With reasonable sample sizes major gene effects can be confidently detected over distances of around 10 to 15 cM. Hence, a whole genome search can be carried out using 200 to 300 polymorphic markers, provided they are approximately evenly spaced. A **polymorphism** can be defined as a gene or sequence of DNA that occurs in two or more common forms. Classically, 'common' means an allele frequency of at least 1 per cent. SNP markers are common

Table 2.4.1.1 Double back-cross mating

		Parent 1	***x***	**Parent 2**
		AB/ab		ab/ab
Offspring	*AB/ab* Non-recombinant	*Ab/ab* Recombinant	*aB/ab* Recombinant	ab/ab Non-recombinant
No linkage	1/4	1/4	1/4	1/4
Linkage	(1–θ)/2	θ/2	θ/2	(1–θ)/2

and biallelic. Sometimes combinations of allelic variants across different markers (haplotypes) are analysed in gene mapping studies.

Linkage analysis

The standard method of carrying out linkage analysis in humans is the lod score approach devised by Morton.[9,13] Essentially, for a given set of data, lod scores are calculated over a range of values of θ between 0 and 0.5. Where the lod score reaches a maximum, provides the best (or maximum likelihood) estimate of θ. The lod score is so called because it is the common **l**og of the **od**ds that θ has a certain value θ′; rather than a value of 0.5, i.e.

$$\text{lod} = \log_{10}\frac{\text{probability}\ (\theta = \theta')}{\text{probability}\ (\theta = 0.5)}.$$

By convention, a lod of 3 or more is accepted indicating that linkage has been detected, while a lod of −2 or less indicates that linkage can be excluded at that particular value of θ. A lod of 3 corresponds to odds on linkage of 1000:1 and to a nominal *P* value of 0.0001. This therefore seems at first sight to be a very stringent criterion. However, linkage between two loci taken at random is inherently unlikely[13] and Morton's[9] original argument took into account the low prior probability of linkage to arrive at a criterion that gave a posterior probability, or reliability, of 95 per cent. More recently researchers have been concerned about the effects of carrying out many statistical tests in a genome-wide search for linkage and have sought to set an approp riate level of lod score to compensate for this. In fact, as it turns out, the original suggestion of a lod of about 3 is close to recent calculations of what lod is required to conclude in favour of genome-wide significance.[13]

As originally devised, the lod method deals purely with regular Mendelian traits. However, it can be readily adapted for detection of single genes that have incomplete penetrance by applying the general single major locus model discussed earlier. The main drawback is that the model (as specified by the penetrance values, and less critically the gene frequency) must be known accurately. Where the model is mis-specified there is a high risk that linkage will fail to be detected.

A further difficulty is that diseases may show locus heterogeneity, i.e. there may be two or more different (and unlinked) loci where mutations result in similar phenotypes. There are many instances of this among rare Mendelian diseases. A good example is Usher's syndrome causing deafness and retinitis pigmentosa, which can result from mutations in any one of six different genes. Subforms of common diseases can also show locus heterogeneity, the most relevant to psychiatry being early onset familial Alzheimer's disease where autosomal dominant forms can result from mutations in three different genes, called presenilin 1, presenilin 2, and amyloid precursor protein. Although methods exist for detecting linkage in the presence of heterogeneity these have not so far in practice been of great help in psychiatric or other common disorders. Rather, the most frequent general strategy has been to focus on multiplex families (i.e. those containing multiple members with the disorder under study) and to make the following simplifying assumptions.

1 There are major gene subforms of the disorder in at least some families.

2 Although the mode of transmission is unknown, a reasonable guess at the defining parameters can be made.

3 Although there may be locus heterogeneity in the disorder as a whole, within any given family there is likely to be homogeneity.

This has worked very well for several disorders, including, as we have just mentioned, Alzheimer's disease, but it initially produced a rather confusing set of results from studies of schizophrenia and bipolar disorder. The most likely cause of this is that assumption 1 is incorrect and that subforms of these conditions resulting from major genes are very rare or perhaps non-existent. Consequently most recent studies use other linkage methods that do not rely on any assumptions about the mode of transmission. These concentrate on affected siblings or other pairs of relatives both affected by the disorder.

Methods based on relative pairs

The underlying principle of the affected sib-pair approach is simple. For any given locus the probabilities that siblings share 0, 1, or 2 alleles that are identical by descent from their parents is respectively ¼, ½, ¼. On the other hand, if both members of a sib pair are affected by the same disease and we are studying a locus close to a gene that confers susceptibility to that disease, there will be increased allele sharing. This will occur irrespective of the mode of transmission of the susceptibility gene and hence simple non-parametric statistics can be used to test whether there is any perturbation of the expected identical-by-descent proportions. Affected sib-pair methods are therefore robust and are now generally considered to be the method of choice in detecting linkage in oligogenic or polygenic disorders. In order to be certain that a pair of siblings share alleles identical by descent, one needs to know their parents' genotypes. Otherwise it could be that a shared allele identical by state results from one of the pair having inherited it from the father and the other from the mother. However, an advantage of using highly polymorphic single sequence repeat polymorphisms is that it may not always be necessary to genotype parents, i.e. where the population is reasonably homogenous and where gene frequencies can be estimated, it is possible to compute the likelihood that a pair who share one or two alleles identical by state are truly identical by descent. This means that in return for a fairly modest reduction in power (because one is now dealing with a probability rather than a certainty of counting alleles that are identical by descent) there is a halving in the amount of genotyping that needs to be done.[20] In our own experience, the other advantage of being able to make do without parental genotypes is that they are often difficult to obtain in adult-onset disorders such as schizophrenia. Significant regions of genetic linkage for disorders, notably schizophrenia, bipolar disorder, and autism that have been shown across different studies or found in pooled analyses have now been identified. Another use of sib-pair methods is in studying continuous traits (e.g. height, weight, personality test scores) to attempt to detect the quantitative trait loci that contribute to their heritability.[14] One approach is to select probands who have extreme scores on some quantitative measure and investigate the extent to which marker allele sharing by siblings predicts the regression to the population mean of the siblings' scores.[14] This has been successfully used in mapping a quantitative trait locus contributing to reading ability in children. Unfortunately the drawback of such methods and of sib-pair linkage approaches generally is that they are only capable of detecting moderately large effects. This means that a quantitative trait locus contributing less than about

10 per cent of the variance, or a disease susceptibility locus conferring a relative risk of less than 2, will probably require very large samples running into several hundreds, perhaps thousands, of pairs. If we assume that most common diseases and complex behaviours involve the combined action of many genes of small effect, complementary strategies based on allelic association are required.

Association

In their classic form, allelic association studies are more straightforward to carry out than linkage studies. A sample of cases affected by a disorder (or subjects who have scores higher than a given threshold on a quantitative measure) is compared with controls who do not have the disorder (or subject whose scores are near average). The frequency of alleles at the marker locus is then compared in the two groups. The significance of the difference can then be compared in the usual way for contingency table analysis using a χ^2 test (or Fisher's exact test if expected frequencies are small). In addition to significance it is useful to have a measure of the strength of association. A variety of statistics can provide this but probably the most useful and intuitively appealing is the **relative risk**, i.e. the proportion of cases among those carrying the marker allele or risk factor, P_1, divided by the proportion of cases among those not carrying the factor, P_2. As we can calculate from Table 2.4.1.2, RR = $P_1/P_2 = (a/(a + b))/(c/(c + d))$. If the disorder is uncommon, i.e. a and c are small relative to b and d, RR can be approximated by another, easier-to-obtain statistic, the odds ratio, OR = $a \times d/(b \times c)$. If a positive marker disease association has been found the odds ratio will be significantly greater than 1.

Before the current era of molecular genetics many association studies of disease with classical markers were carried out most notably with blood groups and with the HLA system. One of the earliest well-replicated findings was an association between blood group O and duodenal ulcer. The odds ratio was less than 2 in most studies and Edwards[15] pointed out that the proportion of variance in liability to the disorder explained by the association was only about 1 per cent. Even though later disease association studies on HLA, with other diseases such as type I diabetes and various auto immune disorders were stronger, it has been pointed out that here too only a small proportion of variance is accounted for. Although this could in one respect be considered disappointing, it demonstrates that allelic association can detect small gene effects in polygenic or multifactorial traits and may therefore prove to be more useful than linkage.

How does allelic association arise and what does its detection tell us? There are three principal mechanisms of association. The first is linkage disequilibrium. Normally pairs of alleles at two different loci occur together no more often than would be expected by chance (i.e. they are in 'equilibrium'). In most cases this is the result of independent assortment. However, even if loci are linked they will usually approach equilibrium very rapidly with the proportion of associated alleles decreasing by $1 - \theta$ each generation. Only where the two loci are very close together does disequilibrium tend to persist. For example, where the distance is 1 cM corresponding to a recombination rate of one meiosis in 100, the time taken for an association to go half way to equilibrium is 69 generations. For a distance of 0.1 cM the time taken is 693 generations, or about 20 000 years. The second cause of association is when a polymorphism within a gene itself has a functional effect which results in susceptibility to a disease. The third, and in most cases least interesting, phenomenon is population stratification. This occurs where there has been recent admixture of populations or two or more ethnically distinct populations living side by side with little interbreeding. If the populations differ in terms of the frequency of alleles of the genetic markers and in the frequency of the disease being studied, marker disease associations can arise if there is not careful ethnic matching of patients and controls.

Another way of overcoming stratification is not to study unrelated cases and controls but to study families and derive the controls 'internally'. The most familiar method in current use is the transmission disequilibrium test.[16] This requires affected individuals to have at least one parent who is heterozygous at the test locus. The affecteds can therefore each receive one of two alleles from such parents. A 2 × 2 contingency table can then be constructed on whether a particular allele is the transmitted or the non-transmitted allele. This is illustrated in Table 2.4.1.3 for a marker with two alleles, A_1 and A_2. The entries in each cell of the table, a, b, c, and d are counts of the number of parents transmitting or not transmitting each allele to their affected offspring. The significance of the transmission disequilibrium test is simply assessed by a McNemar χ^2 test.

Finally, it is becoming increasingly accepted to actually test for population stratification in the laboratory by examining whether random gene variants not thought to be involved in disease, differ in cases and controls. Assuming that stratification can be overcome there are broadly two ways to proceed with association studies. The first is candidate gene association studies and the second is whole genome association studies.

(a) Candidate gene association studies

Functional candidate gene studies concentrate on polymorphisms in or near genes that encode for proteins that are likely to be involved in the disorder. This has so far been the commonest type of association study in psychiatry as in most other common diseases, and there are some interesting early results relating to, for example, polymorphisms at the serotonin *5-HT$_{2a}$* receptor gene in schizophrenia[17] and the dopamine *D4* receptor gene in attention-deficit-hyperactivity disorder.[18] Positional candidate gene studies involve selecting genes that are in regions implicated by linkage.

Table 2.4.1.2 Case-control allelic association

Marker	*N* affected	*N* unaffected
Present	*a*	*b*
Absent	*c*	*d*

Table 2.4.1.3 Transmission disequilibrium test: affected subjects with at least one parent heterozygous for allele A

	Non-transmitted	
Transmitted	**A_1**	**A_2**
A_1	*a*	*b*
A_2	*c*	*d*

X^2.1 df = $(b–c)^2/(b + c)$

This approach, fine mapping using association to narrow down a region discovered by linkage, has for example resulted in identification of a susceptibility locus, *KIAA0319*,[19] involved in reading disability. Another positional strategy involves obtaining clues about the potential position of a susceptibility gene from individual families who present with a disorder and have a specific region of a chromosome disrupted. For example, a study of Scottish families with translocations involving a region on chromosome 1 and schizophrenia and bipolar disorder led to identification of a susceptibility gene *DISC1* for schizophrenia.[20]

(b) Whole genome association (WGA) studies

The second approach, which has only recently become technically feasible, is to attempt a systematic search through the entire genome with the aim of detecting linkage disequilibrium or direct association. It follows from what we have discussed earlier that a genome-wide search for linkage disequilibrium or direct association has a particular attraction in the study of polygenic disorders in that it should be capable of detecting genes of small effect. Very high throughput genetic analysis involves hydridizing DNA into many thousands of oligonucleotides on microarrays and allows a very large number of biallelic single nucleotide polymorphisms (SNPs) to be tested very rapidly and at comparatively low cost. Cost can be reduced even further by performing the initial screening by DNA pooling. Such methods[21] combine samples from groups of patients and groups of controls processing them in batches. Positive findings can then be followed up by doing individual genotyping. Thus if pools consist of say 100 individuals, the initial cost of genotyping is reduced 100-fold. Against this, some information is lost by DNA pooling and accurate construction of pools is difficult and time consuming. The problem of multiple testing can be overcome either (as in the DNA pooling approach), by carrying out a two-stage analysis with fairly liberal test criteria in the first stage followed by stringent criteria in the second stage, or by simply setting a very stringent criterion at the beginning. It has been shown for example that, even with an alpha level set at 1×10^{-6}, detection of linkage disequilibrium with genes of small effect is feasible with realistic sample sizes.

As we write, WGA studies are being completed using commercially available microarrays containing 500 000 SNPs.[22] Such studies present formidable challenges including the need for very large samples of subjects to provide adequate power of detection and problems of multiple testing and handling of huge amounts of data. Nevertheless, WGA studies have already led to the identification of susceptibility genes for common non psychiatric disorders (e.g. macular degeneration and diabetes) and the initial results of several WGA studies of psychiatric disorders including schizophrenia, bipolar disorder, and ADHD will soon be available or have already been published.[23]

Molecular genetic studies investigating gene–environment interplay and intermediate phenotypes

Genetic variants do not necessarily confer risk for a psychiatric disorder unless in the presence of a specific environmental risk factor. Molecular genetic studies allow a direct test of whether the association of a specific gene variant with disorder is contingent upon exposure to a specific environmental factor. There are a number of important methodological issues that need to be considered when investigating G × E.[24] Most importantly, in testing for G × E, there is the problem of multiple testing given the potential for testing a large number of gene variants and environmental risk factors. Thus, there need to be good *a priori* hypotheses before selecting a specific gene variant and environmental measure. Despite these caveats, there is increasing evidence of the likely importance of G × E in psychiatry as a number of findings have now been replicated. The presence of G × E also provides one potential explanation (amongst others) for non-replication of genetic association findings across different studies.

To date the strongest evidence of G × E has been for a gene variant that affects MAOA enzymatic activity and which appears to increase the risk of antisocial behaviour only in the presence of childhood maltreatment.[25] There is also evidence to suggest that possession of a gene variant that affects function of the serotonin transporter (5HTT) increases the risk of depression in the presence of life events.[26,27] A potentially clinically important type of G × E research is pharmacogenetics. Here the aim is to identify genetic variants that influence clinical response and the risk of side-effects upon exposure to specific types of medication. The hope is that such approaches may lead to a more individually tailored approach to prescribing.

As gene variants associated with psychiatric disorder are identified, and variants that affect gene function recognized, there is an increasing need to identify the intermediate biological pathways and mechanisms. This has led to an increasing amount of research on potential intermediate phenotypes that may account for the link between risk factor and psychiatric outcome. Most interest to date has focused on measures of brain function and structure, assessed through imaging studies[28] as well as neurocognitive, biochemical, and physiological traits. For example, the functional variant in the 5HTT that is associated with depression in the presence of adverse life events appears to be associated with amygdala hyper-responsivity to stress. There is also some evidence that a functional COMT gene variant plays an important role in dopamine clearance in the prefrontal cortex and is associated with prefrontal cortical functioning, assessed through cognitive task performance and fMRI. It is hoped that this new era of imaging genetics together with other areas of neurobiological research will help elucidate the risk pathways that lead from genetic and environmental risk factors to psychiatric disorder.

Further information

Kendler, K.S. and Prescott, C.A. (2006). *Genes, environment and psychopathology: understanding the causes of psychiatric and substance use disorders*. Guilford Publications. New York.

Rutter, M. (2006). *Genes and behavior: nature-nurture interplay explained*. Blackwell, Oxford.

Institute of Medicine Board on Health Sciences Policy. (2006). *Genes, behaviour and the social environment: moving beyond the nature nurture debate*. The National Academies Press, Washington DC (http://www.nap.edu).

References

1. Rutter, M., Moffitt, T.E. and Caspi, A. (2006). Gene-environment interplay and psychopathology: multiple varieties but real effects. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **47**(3–4), 226–61.

2. Neale, M.C. and Cardon, L R. (1991). *Methodology of genetic studies of twins and families*. Kluwer, Dordrecht, The Netherlands.
3. Kendler, K.S. and Prescott, C.A. (2006). *Genes, environment, and psychopathology: understanding the causes of psychiatric and substance use disorders*. Guilford Press, New York.
4. D'Onofrio, B.M., Turkheimer, E.N., Eaves, L.J., *et al.* (2003). The role of the children of twins design in elucidating causal relations between parent characteristics and child outcomes. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **44**(8), 1130–44.
5. Plomin, R., DeFries, J.C., Craig, I.W., *et al.* (eds.) (2003). *Behavioral genetics in the postgenomic era*. APA Books, Washington, DC.
6. De Fries, J.C. and Fulker, D.W. (1988). Multiple regression analysis of twin data: etiology of deviant scores versus individual differences. *Acta Geneticae Medicae et Gemellolologiae*, **37**, 205–16.
7. Jaffee, S.R., Caspi, A., Moffitt, T.E., *et al.* (2004). The limits of child effects: evidence for genetically mediated child effects on corporal punishment but not on physical maltreatment. *Developmental Psychology*, **40**(6), 1047–58.
8. Kendler, K.S., Kessler, R.C., Walters, E.E., *et al.* (1995) Stressful life events, genetic liability, and onset of an episode of major depression in women. *American Journal of Psychiatry*, **152**(6), 833–42.
9. Morton, N.E. (1982). *Outline of genetic epidemiology*. Karger, Basel.
10. International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature*, **437**(7063), 1299–320.
11. Rutter, M. (2006). *Genes and behaviour. Nature-nurture interplay explained*. Blackwell, Oxford.
12. Meaney, M.J. and Szyf, M. (2005). Maternal care as a model for experience-dependent chromatin plasticity? *Trends in Neurosciences*, **28**(9), 456–63.
13. Morton, N.E. (1955). Sequential tests for the detection of linkage. *American Journal of Human Genetics*, **7**, 277–318.
14. Fulker, D.W. and Cardon, L.R. (1994). A sib-pair approach to interval mapping of quantitative trait loci. *American Journal of Human Genetics*, **54**, 1092–103.
15. Edwards, T.H. (1965). The meaning of the associations between blood groups and disease. *Annals of Human Genetics*, **29**, 77–83.
16. Spielman, R.S., McGinnis, R.E., and Ewenst, W.J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics*, **52**, 506–16.
17. Owen, M.J., Craddock, N., and O'Donovan, M.C. (2005). Schizophrenia: genes at last? *Trends in Genetics*, **21**(9), 518–25.
18. Thapar, A., O'Donovan, M., and Owen, M.J. (2005). The genetics of attention deficit hyperactivity disorder. *Human Molecular Genetics*, **14**(Spec No. 2), R275–82.
19. Cope, N., Harold, D., Hill, G., *et al.* (2005). Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *American Journal of Human Genetics*, **76**(4), 581–91.
20. Porteous, D.J., Thomson, P., Brandon, N.J., *et al.* (2006). The genetics and biology of DISC1—an emerging role in psychosis and cognition. *Biological Psychiatry*, **60**(2), 123–31.
21. Daniels, J.P., Holmans, P., Williams, N., *et al.* (1998). A simple method for analysing microsatellite allele image patterns generated from DNA pools and its application to allelic association studies. *American Journal of Human Genetics*, **62**, 1189–97.
22. Carlson, C.S., Eberle, M.A., Kruglyak, L., *et al.* (2004). Mapping complex disease loci in whole-genome association studies. *Nature*, **429**(6990), 446–52.
23. Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14, 000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**(7145), 661–78.
24. Moffitt, T.E., Caspi, A., Rutter, M. (2005). Strategy for investigating interactions between measured genes and measured environments. *Archives of General Psychiatry*, **62**(5), 473–81.
25. Kim-Cohen, J., Caspi, A., Taylor, A., *et al.* (2006). MAOA, maltreatment, and gene-environment interaction predicting children's mental health: new evidence and a meta-analysis. *Molecular Psychiatry*, **11**(10), 903–13.
26. Caspi, A., Sugden, K., Moffitt, *et al.* (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*, **301**(5631), 386–9.
27. Kendler, K.S., Kuhn, J.W., Vittum, J., *et al.* (2005). The interaction of stressful life events and a serotonin transporter polymorphism in the prediction of episodes of major depression: a replication. *Archives of General Psychiatry*, **62**(5), 529–35.
28. Meyer-Lindenberg, A. and Weinberger, D.R. (2006). Intermediate phenotypes and genetic mechanisms of psychiatric disorders. *Nature Reviews. Neuroscience*, **7**(10), 818–27.

2.4.2 Molecular genetics

Jonathan Flint

Introduction

The transformation of the LOD score (an acronym for log of the odds ratio), from obscurity as a footnote in medical genetics, to celebrity as multiple choice test item in professional examinations in psychiatry, epitomizes the invasion of genetics, and particularly molecular genetics into psychiatric research. Moreover, like other celebrities caught up in fast moving fields, LOD scores are likely to return to their humble origins within a few years. As molecular genetic approaches to mental health move away from simply identifying genes and DNA sequence variants towards functional studies of increasing complexity, newcomers to the field have to master an expanding literature that covers diverse fields: from quantitative genetics to cell biology, from LOD scores to epigenetics. This chapter takes on the task of making the reader sufficiently familiar with the broad range of subjects now required to follow the progress of psychiatric genetics in the primary literature.

A number of achievements have to be highlighted. Foremost among these is the completion of the human genome project. Announced annually from 2001[1–3] and thereby begging the question as to what constitutes completion, the human genome project is now an essential biological resource. As expected, the ability to sequence whole genomes has transformed the way genetics is carried out, perhaps most egregiously with the rise of bioinformatics as a core discipline: discovery now takes place using the internet rather than the laboratory. Anyone with an interest in human biology should look at the frequently updated information at http://www.ensembl.org or http://genome.ucsc.edu.

Without the human genome two other critical developments would have been impossible: the ability to analyse the expression of every gene in the genome and the ability to analyse (theoretically at least) every sequence variant. Both developments also depend on miniaturization technologies that enable the manufacture and interrogation of initially thousands and then millions of segments of DNA. In addition, results from the International Haplotype Map (HapMap) project,[4] which catalogues common variation in the human genome have been crucial in making it possible to take apart the genetic basis of common, complex disorders such as depression, schizophrenia, and anxiety.

Few disciplines are more burdened with jargon than molecular genetics. This is partly due to the proliferation of molecular techniques, but it is also partly intrinsic to the subject; the only unifying principle is evolution, which often operates in a very ad hoc fashion. Biological solutions to the problems posed by selection result in the adaptation of existing structures to new uses, rather than to the invention of purpose-built systems. Consequently there are few general lessons to be learnt and the novice simply has to become adept at recognizing the acronyms and neologisms that decorate the literature. The material in this chapter aims to equip the reader with the necessary terminology. It begins with the structure and function of DNA, an essential starting place for a number of reasons.

Nucleic acid structure and function

The chemical constituents of genetic information are deoxyribonucleic acid (**DNA**) and ribonucleic acid (**RNA**). Both molecules consist of linear chains of nitrogenous bases bound to a sugar (ribose) and a phosphate backbone. Because of the way the sugars are joined together, one end of each nucleic acid strand will have a terminal sugar residue in which the carbon atom at position number 5 of the ribose molecule is not linked; the other end has a free carbon atom at position 3. These two ends are termed the 5′ (5 prime) and 3′ (3 prime) ends respectively.

It is usual to describe a DNA or RNA sequence by writing the order of bases in a single strand, in the 5′ to 3′ direction. DNA contains four nitrogenous bases: adenine (A), guanine (G), cytosine (C), and thymine (T). RNA differs in that it contains uracil instead of thymine.

Two structural differences between DNA and RNA are important for understanding nucleic acid function. First, DNA has a hydroxyl group on part of its sugar constituent whereas RNA has a hydrogen atom. The result is that, in most biological environments, RNA is much more unstable than DNA. Second, RNA normally exists as a single molecule, whereas DNA is a double helix in which two strands are held together by weak hydrogen bonds between opposed base pairs (**bp**), C joined to G and A to T. The sequence of one strand can therefore be inferred from the other. The two strands are said to be complementary to each other, and this property is exploited whenever DNA is copied (during meiosis, mitosis, or *in vitro* processes such as amplification of DNA using a polymerase chain reaction).

As befits an unstable molecule, RNA mediates the expression of genetic information; its production and degradation are tightly controlled. RNA is translated into a linear order of amino acids in proteins according to a three-letter code (e.g. GAA encodes the amino acid glutamine). DNA acts as a template for the production of RNA in a process termed as transcription. But DNA is more than a stable repository of encoded protein sequence information; it also contains information that controls the transcription of RNA.

Disorders of the template function of DNA are the molecular basis of inherited dispositions and illness, and are the subject matter of genetics. By contrast, gene expression (the transcription and translation of RNA) is not entirely genetically predetermined. It is highly regulated, but in response to changes in the cellular environment which in turn reflect changes in the state of the organism. Disorders of gene regulation are now emerging as important causes of disease.

Genome organization

DNA within cells is packaged into chromosomes in the cell nucleus, with a tiny amount (16 569 bases containing 37 genes) in the mitochondria. Since mitochondria in the fertilized egg are maternally derived, mitochondrial inheritance is through the female lineage. Although small, mitochondrial disorders contribute substantially to degenerative disorders including ageing. More important is nuclear DNA.

As of 2007, the size of the nuclear genome is 3 253 037 807 bp (3.3 gigabases) containing 21 662 known and 1064 novel genes (http://www.ensembl.org). Packing such a large molecule into a cell is done at a number of levels, with profound consequences for gene function. At the first level, 147 bp lengths of DNA are wrapped around octamers of proteins known as histones. These nucleosomes are the fundamental units of the state of packaged DNA known as chromatin.

Packaged DNA is itself organized into 22 autosomal chromosomes, one inherited from the mother and one from the father, and two sex chromosomes, X and Y. Each chromosome pair exchanges stretches of DNA during sexual division (meiosis) in a process called recombination (without which genetic mapping, the basic method of finding disease genes, would be impossible). Chromosomes have three functional elements: origins of replication, centromeres, and telomeres. Replication origins are required to initiate DNA replication and maintain chromosome copy number. Their molecular structure is unknown. Centromeres are responsible for the segregation of chromosomes during cell division. Their molecular nature is also not understood, but they are visible in light microscopy as a constriction where the duplicated chromosomes (called chromatids) are held together. Chromosomes are said to have two arms, separated by the centromere which, despite its name, is not always at the centre. Short arms are termed p (*petit*) and long arms q (*queue*). Telomeres are the ends of chromosomes and their molecular nature is well understood. They consist of long stretches of the sequence TTAGGG without which the chromosome is unstable, tending to break apart and fuse with itself or to other chromosomes.

No one has found any general principles that organize genetic material within chromosomes. Rather than being an efficiently organized plan of the organism consisting of precise drawings, the genome resembles a working copy written over countless rough drafts and discarded versions, among which there are literally thousands of jottings and scribbles, most irrelevant to the final structure. While there are examples of gene families clustered in the same chromosomal location (for instance genes involved in immune regulation are clustered on chromosome 6p), more commonly the position of genes on chromosomes does not reflect functional similarity. For example, genes expressed only in one tissue, or at one stage of development, are often immediately adjacent to widely expressed housekeeping genes.

Genes and the regulation of gene expression

DNA is transcribed into RNA, which in turn is translated into protein. Transcription involves the excision of large portions of transcribed RNA (by RNA splicing enzymes) and modifications to the ends of the RNA molecule (capping of one end and polyadenylation of the other). The final product, messenger RNA (**mRNA**),

contains a central section, translated into protein, and flanking non-coding regions. The consequence of these manipulations is that DNA and mRNA are not coterminous; sections of DNA that encode mRNA (termed exons) are interrupted by often very large stretches of DNA that are not translated (termed introns) (Fig. 2.4.2.1).

Gene expression is controlled in a number of ways, predominantly at the level of transcription (but posttranscriptional processing and translational control are important for some genes). While knowledge in this area is still rudimentary, but advancing fast, it has already led to new insights in disease aetiology: understanding theories of the neurobiology of depression now requires familiarity with chromatin remodelling; epigenetic effects are often invoked in molecular biology of intellectual disability. Below I summarize information about the relevant molecular processes.

1. **Transcription factors:** transcription factors exercise control over gene expression. Transcription occurs when RNA polymerases manufacture RNA from the template DNA, a process that requires the help of transcription factors, proteins that recognize and bind to specific DNA sequences (note that transcription factors can also repress transcription). Although transcription factor-binding sites are found close to a gene, at a 5′ region known as the promoter (see Fig. 2.4.2.1), they may also be situated far away, sometimes within other genes. The characterization of these control elements remains a major challenge for genome research and is currently a focus of the ENCODE project, a continuation of the human genome project whose aim is the comprehensive identification of all the functional elements in our genome (http://genome.ucsc.edu/ENCODE/).

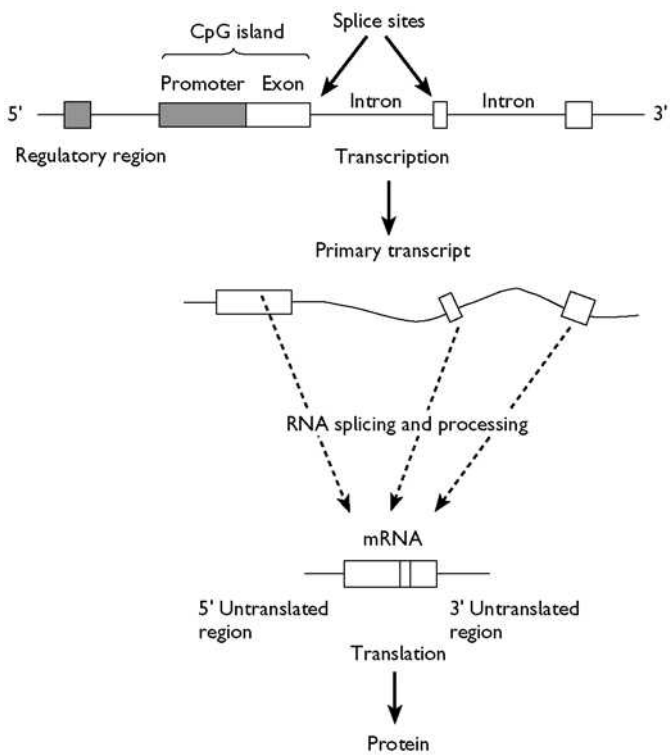


Fig. 2.4.2.1 At the top of the figure the organization of a gene in genomic DNA is shown. Unshaded boxes correspond to coding regions (exons) and the two shaded boxes correspond to control regions. The control region immediately 5′; of the first exon, where transcription is initiated, is known as the promoter and often has a characteristic sequence composition. In almost all ubiquitously expressed genes (and in many tissue-specific genes) it is unmethylated, GC rich, and has a relative excess of the dinucleotide CpG. The region, which typically contains the first exon as well as the promoter, is called a CpG island. The boundaries between exons and introns are called splice sites and are conserved; introns virtually always start with the sequence GT and end with the sequence AG. The entire genomic region is transcribed into a primary transcript (bold arrow) which is then processed to excise the introns. Many human genes undergo alternative splicing to yield a number of different mRNA products. Mature mRNA is then translated into a protein product.

Transcription factors typically control the expression of a number of genes, reflecting the presence of a hierarchical structure of coordinated gene expression. Consequently mutations in transcription factors have effects on different, seemingly unrelated phenotypes, a phenomenon called pleiotropy. The constellation of phenotypic abnormalities seen in intellectual disability syndromes can be explained in this way. For example mutations in the X-linked ATRX gene result in an anaemia (α-thalassaemia), a characteristic facial appearance, profound developmental delay, neonatal hypotonia, and genital abnormalities.[5] The gene contains sequence motifs indicating that it belongs to a group of proteins that bind to chromatin and is involved in chromatin remodelling, discussed below.[6]

2. **DNA methylation:** gene inactivation is associated with DNA methylation, predominantly the addition of methyl groups to cytosine bases. DNA methylation occurs almost always at CpG dinucleotides, and most CpGs in the genome are methylated.[7] DNA methylation represses gene expression in two ways: first, modification of cytosine inhibits the association of some DNA-binding factors with their DNA recognition sequences, and second proteins that recognize methyl-CpG can repress transcription from the methylated DNA. Methylation of DNA does not change the DNA sequence itself, as it is reversible, but the methylation status is maintained when cells divide. Consequently the change is referred to as epigenetic modification.

   DNA methylation is critical for imprinting, a form of gene regulation in which transcripts are expressed from only one of the two parental chromosomes. Although there are relatively few imprinted genes (about 60 are currently documented, although sequence features in the region of known imprinted promoters implicate 600) imprinting is an important phenomenon in neurobiology for three reasons[8]: (i) There is evidence from studies of embryos that maternally expressed genes enhance and paternally expressed genes reduce brain size, indicating that at least some imprinted genes are likely to be involved in neurodevelopment.[9] (ii) Disorders of imprinting are important in intellectual disability syndromes: Rett syndrome, Prader–Willi syndrome, Turner syndrome, and Angelman syndrome.[10] (iii) Imprinting is involved in X chromosome inactivation, the mechanism by which cells compensate for males having just one copy of the X chromosome while females have two.[11] The X chromosome contains a disproportionately high density of loci affecting cognition and since males always inherit their X chromosome from their mother, the presence of X-linked imprinted genes is believed to contribute to sexually dimorphic effects.

   The biology of imprinting is complex and not fully understood. In germ cells and in pre-implantation embryos, all methylation

patterns are removed, and then re-established.[12] About half of identified imprinted genes are clustered within imprinting centres (IC) which carry the allele-specific methylation marks established at this developmental stage. Most maternally silenced imprinted genes are repressed by promoter methylation. No protein-coding imprinted gene has been found that is repressed by paternal methylation derived from the sperm (presumably because of active demethylation of the paternal genome). Importantly, paternal repression is achieved, at least in some cases, by using a transcript on the opposite strand (an antisense transcript). Promoter methylation of the antisense transcript (usually resulting from oocyte-derived methylation) represses its transcription and thus activates the protein-coding gene epigenetically.

Some evidence, obtained by examining females with a single X chromosome (Turner syndrome females), indicates that imprinted genes on the sex chromosomes influence brain structure and function.[13] Since the single X chromosome is inherited from either the mother or father, it is possible to compare the effects attributable to a maternally or paternally imprinted chromosome. Maternally expressed X-linked genes have been reported to influence hippocampal development, while paternally expressed genes influence the normal development of the caudate nucleus and thalamus in females. Using Turner syndrome patients, Skuse and colleagues have suggested that a paternally expressed allele is associated with enhanced social–cognitive abilities.[14] A similar observation has been found in a mouse model of Turner syndrome, lending weight to the view that imprinted genes are involved in cognitive processes.[15]

3. **Chromatin remodelling:** the nucleosome forms a barrier to transcription, primarily because DNA has to be free of nucleosomes for it to be accessible to transcription factors and the large complex of proteins that constitutes RNA polymerase. To some extent the organizational information of where nucleosomes are positioned is embedded within DNA sequence, in a nucleosome code; but the nucleosome is not a static unit. It too has dynamic properties and exerts an effect on transcription. Furthermore, like methylation, the effects are heritable, providing a second form of epigenetic modification to DNA (X inactivation also involves this form of epigenetic modification).

   Histones (the proteins that constitute nucleosomes) are subject to a large number of modifications (acetylation, ubiquitination, methylation, ADP-ribosylation, and sumolation of lysine residues; and phosphorylation of serines and threonines) of which two, lysine acetylation and methylation, have been most heavily studied. Histone modifications can influence each other and may also interact with DNA methylation, in part through the activities of protein complexes that bind modified histones or methylated cytosines. These, and other proteins that remodel chromatin, control genes involved in the development and activity of the central nervous system.

   Chromatin remodelling has attracted attention as a possible mechanism for bringing about persistent change subsequent to an environmental stressor.[16] Two examples are relevant. Meaney and colleagues have reported that heritable differences in stress reactivity in rats depend on variation in parenting, not variation in DNA.[17] Adult offspring of mothers that show higher levels of licking, grooming, and arched-back nursing (high-LG-ABN) are less fearful and show more modest hypothalamic-pituitary axis responses to stress than offspring of 'low-LG-ABN' mothers.[18] How are these maternal effects, or other forms of environmental programming, sustained over the lifespan of the animal?

   Variations in maternal care were found to alter the methylation status of a promoter of the glucocorticoid receptor gene. Central infusion of a histone de-acetylase inhibitor enhanced histone acetylation of the glucocorticoid receptor promoter in the offspring of the low-LG-ABN mothers. Analysis of the promoter showed that CpG dinucleotides were hypomethylated. In consequence, the maternal effect on hippocampal glucocorticoid receptor expression and the hypothalamic-pituitary axis response to stress were both eliminated.[19] This finding suggests that there is a causal relation between epigenetic modifications, glucocorticoid receptor expression, and the maternal effect on stress responses.[20]

   Nestler's group invoked chromatin remodelling as an explanation for the long-lasting behavioural change induced by antidepressants.[21] Chronic defeat stress in rodents is reversed by chronic (but not acute) antidepressant treatment, a model for the action of antidepressant action in our own species. Chronic defeat stress and chronic antidepressant treatment are associated with reciprocal, long-lasting changes in expression levels of brain derived neurotrophic factor (BDNF). This is in turn associated with lasting changes in chromatin architecture at the corresponding BDNF gene promoter. Furthermore, down-regulation of a histone de-acetylase (Hdac5) by chronic antidepressant treatment was necessary for the therapeutic efficacy of the antidepressant.

4. **Small RNAs:** small RNAs include micro RNAs (miRNAs) and small interfering RNAs (siRNAs) directly or indirectly alter gene transcription.[22,23] siRNAs, derived from double-stranded RNAs (dsRNAs), control cleavage of other transcripts and can themselves direct the production of dsRNA by RNA-dependent RNA polymerase; they are also implicated in recruiting heterochromatic modification that leads to transcriptional silencing.

   The extent of siRNA involvement in eukaryotic gene regulation is still unclear. Micro RNAs specify posttranscriptional gene repression by base pairing to the messages of protein-coding genes. They represent nearly 1–5 per cent of all genes in higher eukaryotes and have been implicated in developmental timing and neuronal patterning. They are believed to facilitate the transition between developmental stages and therefore are likely to have an effect on the expression and evolution of most mammalian mRNAs.

   Only 3 per cent of the genome codes for protein.[24] However, that is not to say that the remaining 97 per cent is inactive; in fact the number of transcripts is at least 10 times as great as the number of genes.[25,26] What this means in terms of function is not clear, but along with the discovery of small RNAs it has forced a re-evaluation of what is meant by a gene. The idea that a gene is a section of DNA transcribed into RNA, which in turn encodes a protein, fails to capture the gamut of RNA species, some of known function (such miRNAs), some with only suspected function. The emerging complexity of gene function and the multiple species that need to be included in any definition of a

gene has dramatically increased our understanding of molecular pathogenesis.

Genetics and genotyping

Chromosomes are not stable structures. They rearrange during meiosis, recombining material between the paternal and maternal chromosomes. The mechanisms of recombination are not relevant for understanding neurobiology, but without recombination we would not be able to track mutations. This section describes the basic methodologies currently used. To follow it, and the many reports in the literature, it is essential to be familiar with genetic terminology. A brief reminder is provided next.

A position on a chromosome is called a locus, a general term which can refer to a gene or a segment of DNA with no known function. DNA sequences that differ at the same locus are called allelic variants. Since we have two copies of each chromosome, by definition we have two alleles at each locus. If these alleles are identical the individual is said to be a homozygote; if they are different, the individual is a heterozygote. It follows that for a locus with two alleles (that is one which is di-allelic, as are the single nucleotide polymorphisms that form the basis of almost all genetic mapping experiments), then there are three possible genotypes. For example, if the alleles are either C or T, the possible genotypes are CC, CT, and TT, whose frequencies in a population, in the absence of migration, mutation, natural selection, and assortative mating, are a simple function of allele frequencies (this phenomenon is termed Hardy–Weinberg equilibrium).

The relationship between alleles at different loci is important for genetic mapping. Assume there are two loci on the same chromosome, separated by approximately 1 megabase (Mb). The loci are again di-allelic, the first with alleles C and T, the second with alleles A and G, so that there are nine possible two-locus genotypes: ATCC, TTCG, AAGG etc. Consequently, the chromosomes of an individual with the two-locus genotype ATCG could be any of the following four: A-C, A-G, T-C, or T-G. This combination of alleles along a chromosome is known as a haplotype. If recombination occurs between the two loci it will break-up the haplotype, so that the offspring of someone with the haplotype A-C on one chromosome and T-G on the other may inherit the novel haplotype A-G. The probability at which this occurs depends on the genetic distance between the two loci. For 1 Mb in the human genome this probability is approximately 1 per cent per meiosis, or 1 cM.

Molecular mapping depends on the availability of genetic markers across the genome. The genome is replete with DNA sequence polymorphisms whose only known use is to enable geneticists to map disease genes. On average, every 1000 bp will contain 1.4 bases that differ between two randomly chosen individuals, almost all of which have no phenotypic consequence. In addition, there are small runs of repeated sequence (most commonly CA) which differ in length between individuals. At least one of these short tandem repeats (**STR**s), or **microsatellites**, is found every 50 kilobases (**kb**) and they also have no known phenotypic consequences. There are other more complex sequence polymorphisms, but single nucleotide polymorphisms (**SNP**s) and microsatellites are the most useful for identifying disease genes.

Genotyping of genetic markers almost always starts by amplifying DNA using the polymerase chain reaction (PCR). PCR requires the following reagents: a DNA polymerase; a pair of oligonucleotides (also referred to as primers), which are synthetic single-stranded DNA, usually between 15 and 25 bases long, complementary to two sequences on opposite strands of the target DNA; the target DNA itself; all four nucleotides, usually present in excess; appropriate buffer and cofactors for the reaction. The reaction proceeds in a cycle of three steps: (1) the mixture is heated to over 90°C for 1 min to separate the complementary strands of target DNA, (2) the mixture is cooled, to about 50°C, so that the oligonucleotides anneal to their complementary sequence in the target DNA and allow the DNA polymerase to bind (oligonucleotides are required to prime the polymerase), and (3) the temperature is adjusted to allow the polymerase to function in the extension component of the reaction. Typically the polymerase is from a thermophilic bacterium with a permissive temperature of 72°C. Products from the first cycle serve as targets for a second round of amplification and so on, for up to about 40 cycles.

The method used to genotyping the PCR product depends on the nature of the sequence variant. Microsatellite genotyping involves discriminating the length of the PCR products, usually accomplished by separating the fragments by electrophoresis, in which an electrical current causes smaller fragments to migrate more quickly than larger fragments through a matrix. Typically a fluorescent label is incorporated into the amplified DNA allowing the PCR products to be detected by a laser, a method that allows automated analysis.

SNP genotypes can also be worked out from differences in molecular weight. For example, the nucleotide added to a primer complementary to sequence immediately preceding a SNP in the PCR fragment will depend on the genotype of the individual. The small difference between a primer with a C nucleotide added to its end (reflecting the presence of a G nucleotide in the PCR fragment) compared to the same primer with the addition of a T nucleotide, can be determined using a mass spectrometer. However, most high-throughput genotyping methodologies (required for whole genome association studies) exploit the specificity of DNA hybridization: while two complementary single strands of DNA stick (or hybridize) together (ATTGAC will anneal to TAACTG) a single base mismatch will prevent hybridization (ATTGAC will not anneal to ATAGAC). Consequently, a SNP can be detected by determining whether primers hybridize to the PCR fragment. By labelling the primers, for example, with a fluorescent dye, the hybridization can be visualized, and, using the same technology that builds computer semiconductors, millions of primers can be manufactured on a small piece of glass, allowing the simultaneous detection of millions of SNPs.

Genetic mapping: linkage and association

For genetic disorders that arise from a mutation in a single gene, behaving in a Mendelian fashion (dominant, recessive, or sex-linked inheritance), disease gene identification is conceptually straightforward. Marker alleles follow Mendelian laws of segregation so that it is possible to determine if a marker is co-segregating with a disease in a family and to test the result statistically using methods described in Chapter 2.4.1. The expected result is an estimate of the probability that an allele and a disease locus will recombine; the lower the probability, the closer together the two loci are on the chromosome. The statistical test gives the likelihood that the estimate of recombination distance between a marker and a mutation

is correct. If the likelihood is acceptably high, then the next task is to reduce the genetic interval as much as possible and to identify a causative mutation. Access to the human genome makes it straightforward to identify all genes within a given interval (go to one of the two websites mentioned above) and the ease of DNA sequencing (many companies make this service available) makes it possible to screen candidates by a brute force approach: enough is known about sequence codes to recognize a mutation that will disrupt gene expression, and this can be experimentally verified by looking at the production and structure of mRNA.

A good example of the success of gene mapping in pedigrees is the identification of a mutation in the FOXP2 gene in a family that has a language disorder. The phenotype is complex, including both verbal and non-verbal cognitive impairments. However, the inheritance pattern is straightforward: it fits a model in which a single mutation in one copy of the gene is sufficient to cause disease. By using markers from across the genome, a region on chromosome 7 was identified that co-segregated with the defect and a mutation in a transcription factor, FOXP2, subsequently identified.[27,28]

Pedigree-based linkage works well when the disease follows Mendelian laws of segregation and has led to the identification of many genes involved in intellectual disability and dementia. However, pedigree-based methods for disorders where the genetics does not fit a simple Mendelian pattern have been much less successful. The methods (described in Chapter 2.4.1) typically ask whether affected siblings in a family (usually just a pair) share the same allele at a locus: the more often sharing is observed, the closer the markers locus is inferred to be to the disease gene. The relative failure of the affected sibling pair strategy is almost certainly because of the small contribution that each locus contributes to disease susceptibility.

Neil Risch and Kathleen Merikangas pointed out in 1996 that if the genetic effect size attributable to a single locus is small (that is to say it would increase the risk of developing the disease less than two-fold) then the number of families required using a pedigree method would be impractically large.[29] However, a direct test of association between genetic marker and disease gene had much greater power. Simply by genotyping a marker and determining whether the distribution of genotypes was significantly different between a set of unrelated cases and unrelated controls, small effects could be detected with relatively small sample sizes. The drawback was the need to test variants in every gene, possibly requiring researchers to genotype a million individual markers. Surprisingly, Risch and Merikangas showed that the objection to mounting such a study was not statistical, but technological.

Two advances made whole genome association studies feasible. First, as described above, technologies are available for genotyping on an appropriate scale. Second, an enormous catalogue of SNPs has been compiled. As of April 2007, 11 577 475 SNPs have been identified and mapped on the human genome. Fortunately, not all these variants need to be genotyped for mapping disease genes. It was found that many SNPs were highly correlated with each other, because recombination is less likely to disrupt haplotypes of closely linked markers than of more distantly spaced markers (a phenomenon termed linkage disequilibrium (LD)). Consequently, aetiological variants can be expected to be in LD with one or more SNPs. By genotyping a carefully selected set of markers (called 'tagged SNPs') most of the common variation in the genome can be assayed. However, for this strategy to be practical, a haplotype map of the genome was first required that catalogues the distribution of LD in different populations (since LD patterns are a product of population history, as well as of recombination distance). Genotyping on an immense scale in multiple populations was undertaken by the International Haplotype Map (HapMap) consortium,[4] and the results, regularly updated, are available on the internet (http://www.hapmap.org).

The molecular basis of psychiatric disorders: Mendelian disorders

There are a number of conditions in psychiatry that arise from mutations in single genes (Mendelian mutations). As a rule, such disorders are rare and, importantly, they account for very few, if any, instances of common disorders: depression has never been attributed to a Mendelian mutation (though some single mutations, such as Huntington disease, do include a mood disorder in their phenotype). Nevertheless, rare single gene mutations can provide important clues to the aetiology of more common conditions: for example, mutations in neuroligins, identified as a rare cause of autism, suggest that the pathophysiology lies in abnormalities of synaptic function.

In general, the molecular basis of Mendelian mutations in psychiatry is typical of other human genetic disorders in that they arise primarily from changes to one, or a few, nucleotides. These are described below under the heading point mutations. However, there are two exotica: triplet repeats and imprinting defects.

Point mutations

Changes in a single base pair (e.g. from C to T) of the coding sequence of the gene may alter the function of the protein (**mis-sense mutations**), result in premature termination of the protein product (**non-sense mutations**), or create or destroy a splice site. In addition, deletions (of a single base pair or many megabases of DNA) and insertions (again of any size) disrupt transcription and translation of a gene. Deletions or insertions that do not affect a multiple of three bases alter the way that the message is translated and are known as frame-shift mutations. While none of these mechanisms is special to psychiatric genetics, describing the mutational basis of a genetic disorder is an initial step in understanding how the disorder arises; what the DNA sequence change does to the protein gives a clue to the protein's function (if unknown) and to the pathogenesis of the condition. Analysis of mutants causing dementia, isolated by positional molecular cloning, are an example.

The presenilins are ubiquitously expressed transmembrane proteins in which mutations give rise to about 40 per cent of familial Alzheimer's disease.[30] More than 160 mutations in the presenilin genes have been identified. Almost all are mis-sense mutations. Why are there no frame-shift or non-sense mutants? Mis-sense mutants alter an amino acid in the protein and therefore can alter its function. One clue about the nature of the functional change comes from looking at where mutations occur within the protein, thereby discerning whether some parts of the molecule are more frequently involved than others. The distribution of mutations in the presenilins is indeed non-random; mutations occur at residues which are the same in both presenilin genes, lying on one side of the α helix in transmembrane domains, predominantly in exon 8. Thus the mutational spectrum highlights key residues for

understanding the protein's functions.[31,32] The mutations in presenilin result in a gain of function: they alter the ratio of the two forms of amyloid that constitute neuritic plaques, one of the histological hallmarks of the Alzheimer's disease.[33,34]

Where a gene's function is known, the distribution of mutations may reveal the likely pathogenesis. For example, mutations in the tau gene cause frontotemporal dementia with Parkinsonism (Pick's disease).[35,36] The most common mutation occurs in the 5′ splice site of exon 10, resulting in overproduction and accumulation of one form of tau; mutations in other regions of the gene lead to accumulation of a different form of the protein. The position of the mutations in tau indicate that they cause disruption of tau microtubule binding, which may cause cell death by the degeneration of microtubules or through an increase in unbound tau.

Triplet repeats

One mutational mechanism is unusual and deserves special comment: expansion of trinucleotide repeats.[37] Its importance in psychiatry is that it occurs in at least 16 neurological disorders, some with behavioural phenotypes (such as Huntington disease). No one knows why trinucleotide repeat expansions tend to be found in disorders of the central nervous system. The mechanism was first discovered in 1991 as the cause of fragile X syndrome,[38,39] a common form of inherited X-linked intellectual disability.

The mechanism is important because it explains some otherwise unusual features of the phenotype. Table 2.4.2.1 lists the diseases associated with triplet repeats. At each locus there is a normal range of copy numbers above which the repeat array becomes unstable: the larger the number of copies, the more unstable the allele and in general the more severe the disease. Repeats increase and decrease in size both in somatic and germ line tissues, with two important consequences. First, it may not be possible to infer the severity of the condition from a blood test, since the repeat length in the brain may not be the same as in the lymphocytes (this is an example of somatic mosaicism). Second, as the repeat length increases in successive generations, the age of onset may decrease (this is called anticipation). For instance the age of onset of myotonic dystrophy ranges from birth to adulthood.

Genomic rearrangements and gene dosage effects

A number of disorders, primarily those associated with intellectual disability, have been found to be due to chromosomal rearrangements. Down syndrome (trisomy 21) is by far the most common, accounting for about a third of all cases with moderate to severe retardation.[40] Chromosomal rearrangements can be extremely complex, like the nomenclature used to describe them.

Abnormalities of the number of chromosomes result in aneuploidy. Deletion of part or an entire chromosome is termed a monosomy (or haploinsufficiency); an extra copy of either part of or an entire chromosome is called trisomy. A general term to describe either loss or excess of chromosomal material is aneusomy. Most chromosomal abnormalities involve small regions of aneusomy and consequently are known as segmental aneusomy syndromes or contiguous gene syndromes.[41]

Table 2.4.2.1 Triplet repeat diseases

Disease		Repeat unit	Gene	Normal repeat	Expanded repeat	Mechanism
Fragile X syndrome	FRAXA	(CGC)n	FMRP	6–60	>200	Loss of function
Fragile XE syndrome	FRAXE	(CCG)n	FMR2	4–39	200–900	Loss of function
Friedrich ataxia	FRDA	(GAA)n	Frataxin	6–32	>200	Loss of function
Myotonic dystrophy type 1	DM1	(CTG)n	DMPK	5–37	50–10 000	RNA-mediated
Myotonic dystrophy type 2	DM2	(CCTG)n	ZNF9	10–26	>75	RNA-mediated
Fragile X–associated tremor ataxia syndrome	FXTAS	(CGG)n	FMR1	6–60	60–200	RNA-mediated
Huntington disease	HD	(CAG)n	Huntingtin	6–34	36–121	Polyglutamine expansion
Spinocerebellar ataxia	SCA1	(CAG)n	Ataxin1	6–44b	39–82	Polyglutamine expansion
Spinocerebellar ataxia	SCA2	(CAG)n	Ataxin2	15–24	32–200	Polyglutamine expansion
Spinocerebellar ataxia	SCA3	(CAG)n	Ataxin3	13–36	61–84	Polyglutamine expansion
Spinocerebellar ataxia	SCA6	(CAG)n	CACNA1A	4–19	10–33	Polyglutamine expansion
Spinocerebellar ataxia	SCA7	(CAG)n	Ataxin7	4–35	37–306	Polyglutamine expansion
Spinocerebellar ataxia	SCA17	(CAG)n	TBP	25–42	47–63	Polyglutamine expansion
Spinobulbar muscular atrophy	SBMA	(CAG)n	Androgen receptor	9–36	38–62	Polyglutamine expansion
Dentatorubral-pallidoluysian atrophy	DRPLA	(CAG)n	Atrophin	7–34	49–88	Polyglutamine expansion
Spinocerebellar ataxia	SCA8	(CTG)n	SCA8	16–34	>74	Unknown
Spinocerebellar ataxia	SCA10	(ATTCT)n	10–20	500–4500		Unknown
Spinocerebellar ataxia	SCA12	(CAG)n	PPP2R2B	7–45	55–78	Unknown
Huntington disease-like 2	HDL2	(CTG)n	Junctophilin	7–28	66–78	Unknown

The distribution of rearrangements in the genome is not random, reflecting instead the involvement of higher-order architectural features.[42] Regions susceptible to rearrangement are frequently flanked by, or contain, region-specific repeat sequences, or low-copy repeats. In general, recurrent rearrangements, or those of common size and having clustered breakpoints, most frequently result from homologous recombination between repeats.[43]

Chromosomal rearrangements in psychiatric patients other than those with intellectual disabilities are rare, but their occurrence has led to some important discoveries. Deletions on the end of chromosome 22q cause a syndrome that includes a psychosis often indistinguishable from schizophrenia, thus prompting an intense investigation of this region of the genome.[44] Characterization of patients with translocations and psychosis also led to the identification of a gene called DISC1 (for **d**isrupted **i**n **sc**hizophrenia)[45] and screening for chromosomal rearrangements in patients with Tourette's syndrome led to the identification of mutations in a Slit and Trk-like family member 1 (SLITRK1) gene.[46]

The phenotypes of chromosomal rearrangements are thought to arise because of the loss, in the case of monosomy, or addition, in the case of trisomy, of dosage-sensitive genes, of unrelated function, that happen to lie next to each other on the chromosome. However this is not always the case, as the examples of Prader–Willi syndrome and Angelman syndrome show.

Imprinting defects

A number of diseases can be attributed to a failure to establish, maintain, or recognize methylation. Rett syndrome is a progressive neurodevelopmental disorder that occurs almost exclusively in females, with an incidence of between 1/10 000 and 1/15 000 live births.[47] Most females with Rett syndrome are usually heterozygous for a *de novo* mutation in methyl-CpG-binding protein MeCP2, a protein that induces the recruitment of protein complexes involved in histone modifications and chromatin remodelling. Prader–Willi syndrome and Angelman syndrome are both caused by loss of function of imprinted genes on the proximal long arm of human chromosome 15.[48] Prader–Willi syndrome occurs if the paternal chromosome 15 is missing, Angelman syndrome if the maternal. In a few per cent of patients the disorder is due to aberrant imprinting and gene silencing.

Non-coding RNA

The discovery that small RNA molecules regulate gene expression has occurred too recently for us to know the importance of mutations in this system as a cause of disease. Examples are however beginning to be reported: a point mutation was shown to create a new promoter, driving a novel transcript that in turn silenced a neighbouring gene.[49] There is one example from neuropsychiatric genetics, in which mutations in the binding site for microRNA hsa-miR-189 have been identified as a cause of Tourette's syndrome.[46] These complex mechanisms are difficult to detect without a detailed understanding of gene regulation, which is likely to come in future years.

The molecular basis of psychiatric disorders: complex disorders

Genetic attempts to dissect common psychiatric conditions have been very slow to yield robust results. For every paper that reports a positive association between a genetic variant and schizophrenia, anxiety or alcohol abuse, another can be found that fails to replicate the finding.[50] There are at least three explanations for the difficulties in arriving at agreement. One is that scientific journals afford more importance to positive than negative results; for example, there have been more than 40 studies investigating the relationship between a variant of the dopamine D2 receptor and alcohol abuse. By plotting the effect size of each study against the year of publication, a clear and statistically significant negative correlation is found: studies with the largest effects are published first.[51] Publication bias almost certainly exists in other studies, but has rarely been systematically examined.

Second, a number of commentators have pointed out that the sparse success of genetic linkage and association studies in complex traits can be explained by the low power of individual studies.[52] Meta-analytic techniques that make it possible to combine data from many individual studies have revealed that the odds ratios attributable to individual susceptibility loci are commonly less than 1.3.[53] Analysis of obesity, Type 2 diabetes and breast cancer shows that it is possible to obtain association results that replicate, but that tens of thousands of subjects are required.[54,55] There is no reason to believe that psychiatric disorders will be any different in this respect; it is simply that no study has reported data from enough cases.

A third reason for the inconsistencies is that interaction between genes can obscure the signal attributable to the main effect of the locus.[56] Gene interaction, or epistasis, means that the phenotype depends on the allelic configuration of a number of loci. For example, if there are two alleles (*i* and *j*) at a susceptibility locus on chromosome 1 and two alleles at a susceptibility locus on chromosome 2 (*l* and *m*) then in an epistatic interaction disease will only manifest in those individuals with allele *i* on chromosome 1 and allele *m* on chromosome 2. Imagine a population in which allele *i* on chromosome 1 is common. Genetic association tests of the effect of allele *m* on chromosome 2 will detect an effect. But the same test, when carried out in a population where allele *i* is rare, will not detect an effect, even though allele *m* is present at equivalent frequencies in both populations.

Epistasis is suspected to be important, but there are currently no well-documented examples in psychiatric genetics. More attention, and more information, is available for a similar phenomenon called gene by environment interaction (GXE). Interaction between genes and environment is said to occur if disease manifests only when specific alleles are exposed to a given environment. For example, individuals with one allele (called *s* (for short)) at the promoter of the serotonin transporter may have an increased chance of developing depression, but only if they have been exposed to stressful life events.[57] Individuals with the *l* (for long) allele are protected against the effect of this environmental stressor and are less likely to become depressed. GXE hides genetic association in the same way as epistasis. In the presence of GXE, unless differences in the environmental stresses experienced by subjects are taken into consideration, contradictory results may be obtained from testing a genetic association between the serotonin transporter polymorphism and depression in different populations.

The advent of whole genome association studies applied to large case-control cohorts is beginning to unravel the molecular basis of common complex diseases. These studies are proving to be successful, in that they are identifying small numbers of loci that can be replicated in independent samples. However, since most studies

find at best a handful of loci contributing to disease susceptibility, the results show that the bulk of the genetic predisposition to complex disease is almost certainly due to loci with much smaller effects than those found to date. There are also indications that many of the variants will be extremely rare, possibly even private to individual families.

One important indicator of the importance of rare variants is the poor coincidence between the location of whole genome association hits and those from linkage studies. It is important to bear in mind that linkage and association are detecting different signals: linkage will detect an effect when there are multiple different variants in the same gene (allelic heterogeneity). However, allelic heterogeneity dramatically reduces the power of genetic association, so that signals will be missed. Re-sequencing of genes has also led to a greater appreciation of the role of rare variants. Although individually rare, non-synonymous sequence variants in certain genes are cumulatively frequent and are known to influence quantitative traits, such as plasma lipoprotein levels[58]; it may prove to be true also for behavioural phenotypes.

Finding rare non-synonymous variants in genes should not obscure the importance of aetiological variants that lie outside genes. It had been thought that the spectrum of sequence variants in complex disease would be similar to mutations found from the analysis of Mendelian conditions: that is to say variants in the coding regions of genes, the splice sites, or the promoters. However, SNPs with robust, replicated results for association with complex disease are usually found nowhere near coding regions; they are located outside genes, in regions not known to have any function. Furthermore, when genes implicated by genetic association in psychiatric disease have been entirely sequenced, no obvious abnormalities are seen. This is true, for example, of the genes believed to be involved in schizophrenia (dysbindin and neuregulin).

Genome-scanning technologies have also uncovered an unexpectedly large amount of structural variation in the human genome, including deletions, duplications, and large-scale copy-number variants, as well as insertions, inversions, and translocations. A global survey of copy-number variants identified 1447 regions, covering 360 Mb (a remarkable 12 per cent of the genome), revealing a considerable contribution to overall genetic heterogeneity.[59] Critically, these variants were found in supposedly disease-free populations. It is possible that copy-number variants play an important part in the aetiology of common psychiatric disorders. For example, genome-scanning of autism has found copy-number variants in 10 per cent of cases of sporadic autism and in only 1 per cent of controls.[60]

Functional analysis

The armamentarium of molecular genetics tools wielded in the onslaught on the genetic basis of psychiatric disorders is impressive, expensive, and at the same time relatively uninformative about the neurobiology of the illnesses. As genes are identified an equally complex technology is being applied to working out what those genes do. The last section of this chapter provides a guide to the relevant molecular biology.

One of the most fruitful ways of investigating gene function is through genetic manipulation of animals, for example, by introducing a copy of the gene into a mouse or mutating the mouse version of the gene. Both approaches work because exogenous DNA can integrate into chromosomes. If the site of integration can be targeted, rather than being random, the exogenous DNA can be used to create mutations in specific genes.

Transgenic mice are made by injecting DNA (usually, but not necessarily, human) into fertilized mouse oocytes. Integration is a rare and random event and almost always involves multiple copies entering at a single site, making interpretation of some transgenic experiments difficult. More specific genetic manipulation is achieved by exploiting homologous recombination in embryonic stem (**ES**) cells. DNA containing a mutated copy of the gene of interest in tandem with a selectable marker (e.g. an antibiotic resistance gene) is introduced into ES cells. In a small number of cases the exogenous DNA recombines with, and consequently replaces, the cell's copy of the gene. ES cells are isolated from embryos and can be grown in flasks while retaining the potential to develop into any tissue. Once they have been genetically manipulated and re-injected back into a pregnant mouse, the resulting embryo is a mixture of mutant and normal cells. If some of the ES cells contribute to the germ line of the embryo, then its offspring will be heterozygote mutants, from which homozygotes can be bred.

One important reservation should be borne in mind when assessing studies that have used gene targeting: the genetic background of a mutant can have an effect on the mutant.[61] Most targeting experiments use the ES stem cells derived from substrain 129 mice, which are crossed with another inbred line. The choice of inbred line into which the mutation is bred can be critical because many inbred lines have specific behavioural phenotypes; for instance, the inbred strain DBA/2 shows poor hippocampal-dependent learning and C57BL/6 mice are poor avoidance learners.[62]

Improvements in knockout technology have had a major impact on neurobiology. Rather than simply knockout genes, homologous recombination can be used to change part of a gene, for example, by substituting one amino acid to another, a process referred to as a knock-in technology. The development of binary systems, where two engineered lines are crossed, has given experimenters remarkable control over gene expression: genes can now be turned on, or off, by feeding the animal a compound that penetrates to the cell nucleus, such as tetracycline.[63] The tetracycline transactivator (tTA) is used as a transcriptional switch to drive the expression of a gene of interest. The tTA system requires the use of two lines of transgenic mice. In one line the expression of the tTA protein is driven by a tissue or cell-type-specific promoter; in the second, a gene of interest is placed downstream of the tet operator (tetO) and a promoter. The tTA protein binds to the tetO sequence and induces transcription. However, when the tetracycline is present, it binds to tTA and prevents it from binding to tetO and this halts transcription.

Binary systems also make it possible to generate mutations that are restricted to specific cell or tissue-types. This is done by introducing sequences recognized by an enzyme, cre recombinase, on either side of the gene of interest. Cre is a site-specific DNA recombinase derived from a bacteriophage that recognizes 34 bp sequences termed loxP sites. Cre catalyses the deletion of DNA flanked by a pair of loxP sites (the DNA is said to be floxed). Again the experimental system requires two lines: in one the gene of interest is floxed; in the other, cre is driven by a tissue-specific promoter so that its expression is restricted to the cells of interest. Consequently, the gene is only deleted in that subset of cells.

Targeted homologous recombination, combined with conditional control, has proved so successful a method for investigating gene function, that three major mouse knockout programmes are underway worldwide, working together to mutate every gene in the mouse genome using this technology. I have chosen a few examples on the application of the technology to illustrate the remarkable power of the approach and reveal why it has generated such interest in the neuroscience research community.

One of the earliest was the genetic demonstration of the effect of a brain-subregion-restricted NMDA receptor knockout on spatial memory.[64] Following the discovery that high-frequency stimulation of the hippocampal input fibres can result in long-lasting enhancement of synaptic transmission, long-term potentiation (LTP) has been subject to extensive investigation. The induction of LTP is blocked by amino-phosphonovaleric acid (AP5), an antagonist of the NMDA subset of glutamate receptors. A voltage-dependent magnesium block and high calcium permeability mean that the receptor can be opened by glutamate only when the postsynaptic neurone is depolarized, thereby allowing the receptor to function as a detector and integrator of coincident activity at the synapse. Unsurprisingly, the demonstration that infusion of AP5 into the ventricles of rats impaired spatial learning generated intense interest in the potential role of the NMDAR as a crucial component of memory. Furthermore, since disorders of working memory have been documented in schizophrenia, NMDAR dysfunction has attracted the attention of psychosis researchers. The genes encoding the receptor were identified in the early 1990s. There are seven subunits (NR1, NR2A–D, and NR3A and B), of which the NR1 subunit is the only one that is indispensable for the formation of a functional receptor.

Functional investigation of the NR1 receptor using gene knockout technology proved to be impossible because NR1-knockout mice do not survive for more than a day after birth: the receptor has a crucial role in the midbrain for breathing. This is a general problem with constitutive knockout technology: lethality before, or just after birth is common. Tonegawa and colleagues used the conditional technology to get over this problem. They floxed the NR1 receptor and crossed the mouse to a line in which cre recombinase was under the control of a promoter of a gene expressed in the hippocampus.[65] Because the recombinase was predominantly active in the CA1 region of the hippocampus, the NR1 gene could be knocked out postnatally in CA1 pyramidal cells.

The mouse strain had apparently normal growth and was fertile, but was severely impaired in a test of spatial learning.[65] Furthermore, recording of neuronal activity indicated a loss of coherent spatial representation in the hippocampus.[66] The conditional knockout provided strong evidence that NMDAR activity and NMDAR-dependent synaptic plasticity in the hippocampus are crucial for spatial learning. The importance of CA1 NMDARs in the acquisition of hippocampus-dependent memory was subsequently extended to various non-spatial tasks, including recognition of novel objects, and fear conditioning.

Advances in genetic engineering have also been critical in shaping views of the neurobiology of anxiety and depression. There are a large number of publications reporting anxiety phenotypes associated with constitutive knockouts, implicating so many genes that as one reviewer points out 'the overall message one takes away from these studies is that normal anxiety requires normal neuronal functioning. Disrupt such functioning in any of a number of different ways and anxiety-like behaviour is likely to be disrupted—not a very specific or informative conclusion'.[67] Application of the more focused molecular technology has been more informative.

Benzodiazepines bind the α-subunit of the pentameric receptor, enhancing the efficacy of GABA in activating the receptor. By introducing a histidine to arginine mutation into the genes encoding each of the α1, α2, and α3 subunits and by testing mutants for behavioural abnormalities, it has been possible to determine that the α2 isoform is responsible for the anxiolytic effects, but not the sedative or amnesic, effects of benzodiazepines in mice.[68–70] These results have encouraged a search for isoform-specific medications, since a drug specific for the α2 subunit might reduce anxiety while having little or no sedative effects.

The importance of developmental, as well as tissue-specific effects, has also emerged from the use of conditional knockouts. While knockouts of the 5-HT1A serotonin receptor indicated the importance of this gene in anxiety, the mechanism by which it exerts an effect was not appreciated until Hen and colleagues developed a tissue-specific, inducible rescue of the 5-HT1A knockout.[71] A mutant line, in which the 5-HT1A promoter was replaced with a tTA responsive element, was crossed with a line in which the tTA protein was under the control of a promoter expressed in the hippocampus and cortex (the CaMKII promoter). The tTA protein induced expression of the 5-HT1A receptor in postsynaptic target tissues, but not in the serotonergic neurones of the dorsal raphe nucleus. Rescue of the knockout restored normal anxiety-like behaviour, demonstrating that the lack of postsynaptic receptors, rather than presynaptic, causes the anxiety phenotype. Furthermore, by giving tetracycline only during adulthood, so that the 5-HT1A receptor was expressed only during development, it was found that the mice behaved like wild type animals. By contrast, when 5-HT1A receptor expression was ablated during development, it was not possible to rescue the phenotype in the adult, demonstrating that stimulation of postsynaptic 5-HT1A receptors during a developmental critical period is required to establish normal patterns of anxiety-like behaviour that then persist into adulthood.

The creation of knockouts should not obscure an equally important application of molecular genetic methods in neuroscience: to augment existing functional tools or to create entirely novel ones. The developments can be categorized into visualization tools and, more recently, tools for intervention.

The ability to visualize proteins in cells has transformed cellular neurobiology. Visualization is made possible by attaching a fluorescent tag, usually green fluorescent protein or one of its congeners, to the DNA encoding the relevant protein. The genome projects have made available complete libraries of cloned DNA, so that it is possible to obtain cloned DNA of any gene. By using homologous recombination in bacteria, any gene can be tagged, and then inserted into the mouse genome, either by targeting or random transgenesis.[72,73] This genetic tagging has made it possible to track gene products spatially and temporally at the highest level of resolution (fluorescently tagged genes expressed in the mouse brain can be seen at http://www.gensat.org). One immediate application of this technology is the identification of sequences that confer tissue-specific gene expression, revealed by the expression of the fluorescent protein. Genes expressed in a specific brain region can then be replaced, rather than tagged, making it possible to target proteins to brain regions.[74] This is one of the key technical requirements for mounting an interventionist approach to neurobiology.

Interventionist tools are set to transform neurobiology from a discipline in which function is inferred by observing the connectivity of the nervous system, into one in which a function can be directly tested by activating its component parts. Sensor proteins, that detect changes in the physiological states of neurones through the emission of light, and actuator proteins that effect such changes in response to an exogenous signal, can be genetically engineered and inserted into organisms using genetic targeting technologies.[75,76] The ability to activate specific cell types through genetically encoded sensors responsive to light will undoubtedly facilitate the exploration of neuronal circuits. These techniques, though still in their infancy and currently most useful in invertebrate model organisms, are likely to have broad applications in neurobiology and will be a critical tool for understanding the function of genes involved in psychiatric disease.

Further information

Strachan, T. and Read, A. (2003). *Human molecular genetics* (3rd Revises edn). Garland Publishing Inc., US.

Alberts, B. (ed.) (2002). *Molecular biology of the cell* (4th Revises edn). Garland Publishing Inc., US.

http://www.genome.ucsc.edu/

http://www.ensembl.org/index.html

http://www.hapmap.org/

References

1. Lander, E.S., Linton, L.M., Birren, B., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
2. Venter, J.C., Adams, M.D., Myers, E.W., *et al.* (2001). The sequence of the human genome. *Science*, **291**, 1304–51.
3. International-Human-Genome-Sequencing-Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931–45.
4. Altshuler, D., Brooks, L.D., Chakravarti, A., *et al.* (2005). A haplotype map of the human genome. *Nature*, **437**, 1299–320.
5. Gibbons, R.J., Picketts, D.J., Villard, L., *et al.* (1995). Mutations in a putative global transcriptional regulator cause X-linked intellectual disability with í-thalassemia (ATR-X Syndrome). *Cell*, **80**, 837–45.
6. Gibbons, R.J. and Higgs, D.R. (2000). Molecular-clinical spectrum of the ATR-X syndrome. *American Journal of Medical Genetics*, **97**, 204–12.
7. Klose, R.J. and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences*, **31**, 89–97.
8. Wood, A.J. and Oakey, R.J. (2006). Genomic imprinting in mammals: emerging themes and established theories. *PLoS Genetics*, **2**, e147.
9. Ferguson-Smith, A.C. and Surani, M.A. (2001). Imprinting and the epigenetic asymmetry between parental genomes. *Science*, **293**, 1086–9.
10. Raymond, F.L. and Tarpey, P. (2006). The genetics of intellectual disability. *Human Molecular Genetics*, **15** (Spec. no. 2), R110–16.
11. Reik, W. and Lewis, A. (2005). Co-evolution of X-chromosome inactivation and imprinting in mammals. *Nature Reviews. Genetics*, **6**, 403–10.
12. Reik, W., Dean, W., and Walter, J. (2001). Epigenetic reprogramming in mammalian development. *Science*, **293**, 1089–93.
13. Davies, W., Isles, A.R., Burgoyne, P.S., *et al.* (2006). X-linked imprinting: effects on brain and behaviour. *BioEssays*, **28**, 35–44.
14. Skuse, D.H., James, R.S., Bishop, D. V. *et al.* (1997). Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. *Nature*, **387**, 705–8.
15. Davies, W., Isles, A., Smith, R., *et al.* (2005). Xlr3b is a new imprinted candidate for X-linked parent-of-origin effects on cognitive function in mice. *Nature Genetics*, **37**, 625–9.
16. Tsankova, N., Renthal, W., Kumar, A., *et al.* (2007). Epigenetic regulation in psychiatric disorders. *Nature Reviews. Neuroscience*, **8**, 355–67.
17. Francis, D., Diorio, J., Liu, D., *et al.* (1999). Nongenomic transmission across generations of maternal behavior and stress responses in the rat. *Science*, **286**, 1155–8.
18. Meaney, M.J. (2001). Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annual Review of Neuroscience*, **24**, 1161–92.
19. Weaver, I.C., Cervoni, N., Champagne, F.A., *et al.* (2004). Epigenetic programming by maternal behavior. *Nature Neuroscience*, **7**, 847–54.
20. Meaney, M.J. and Szyf, M. (2005). Maternal care as a model for experience-dependent chromatin plasticity? *Trends in Neurosciences*, **28**, 456–63.
21. Tsankova, N.M., Berton, O., Renthal, W., *et al.* (2006). Sustained hippocampal chromatin regulation in a mouse model of depression and antidepressant action. *Nature Neuroscience*, **9**, 519–25.
22. Zaratiegui, M., Irvine, D.V., and Martienssen, R.A. (2007). Noncoding RNAs and gene silencing. *Cell*, **128**, 763–76.
23. Nilsen, T.W. (2007). Mechanisms of microRNA-mediated gene regulation in animal cells. *Trends in Genetics*, **23**, 243–9.
24. Carninci, P., Kasukawa, T., Katayama, S., *et al.* (2005). The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559–63.
25. Cheng, J., Kapranov, P., Drenkow, J., *et al.* (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*, **308**, 1149–54.
26. Cawley, S., Bekiranov, S., Ng, H.H., *et al.* (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell*, **116**, 499–509.
27. Fisher, S.E., Vargha-Khadem, F., Watkins, K.E., *et al.* (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nature Genetics*, **18**, 168–70.
28. Lai, C.S., Fisher, S.E., Hurst, J.A., *et al.* (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, **413**, 519–23.
29. Risch, N. and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, **273**, 1516–17.
30. Tandon, A. and Fraser, P. (2002). The presenilins. *Genome Biology*, **3**, 1–9.
31. Hardy, J. (1997). Amyloid, the presenilins and Alzhheimer's disease. *Trends in Neurosciences*, **20**, 154–9.
32. Hardy, J., Duf, K., Hardy, K.G., *et al.* (1998). Genetic dissection of Alzheimer's disease and related dementias: amyloid and its relationship to tau. *Nature Neuroscience*, **1**, 355–8.
33. Goedert, M. and Spillantini, M.G. (2006). A century of Alzheimer's disease. *Science*, **314**, 777–81.
34. Shen, J. and Kelleher, R.J. III. (2007). The presenilin hypothesis of Alzheimer's disease: evidence for a loss-of-function pathogenic mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 403–9.
35. Hutton, M., Lendon, C.L., Rizzu, P., *et al.* (1998). Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature*, **393**, 702–5.
36. Spillantini, M.G., Murrell, J.R., Goedert, M., *et al.* (1998). Mutation in the tau gene in familial multiple system tauopathy with presenile dementia. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 7737–41.
37. Orr, H.T. and Zoghbi, H.Y. (2007). Trinucleotide repeat disorders. *Annual Review Neuroscience*, **30**, 575–621.

38. Yu, S., Pritchard, M., Kremer, E., *et al.* (1991). Fragile X genotype characterized by an unstable region of DNA. *Science*, **252**, 1179–81.
39. Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., *et al.* (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, **65**, 905–14.
40. Antonarakis, S.E. and Epstein, C.J (2006). The challenge of Down syndrome. *Trends in Molecular Medicine*, **12**, 473–9.
41. Budarf, M.L. and Emanuel, B.S. (1997). Progress in the autosomal segmental aneusomy syndromes (SASs): single or multi-locus disorders? *Human Molecular Genetics*, **6**, 1657–65.
42. Sharp, A.J., Cheng, Z., and Eichler, E.E. (2006). Structural variation of the human genome. *Annual Review of Genomics and Human Genetics*, **7**, 407–42.
43. Lee, J.A. and Lupski, J.R. (2006). Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron*, **52**, 103–21.
44. Karayiorgou, M. and Gogos, J.A. (2004). The molecular genetics of the 22q11-associated schizophrenia. *Brain Research. Molecular Brain Research*, **132**, 95–104.
45. Mackie, S., Millar, J.K., and Porteous, D.J. (2007). Role of DISC1 in neural development and schizophrenia. *Current Opinion in Neurobiology*, **17**, 95–102.
46. Abelson, J.F., Kwan, K.Y., O'Roak, B.J., *et al.* (2005). Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science*, **310**, 317–20.
47. Bienvenu, T. and Chelly, J. (2006). Molecular genetics of Rett syndrome: when DNA methylation goes unrecognized. *Nature Reviews. Genetics*, **7**, 415–26.
48. Horsthemke, B. and Buiting, K. (2006). Imprinting defects on human chromosome 15. *Cytogenetic and Genome Research*, **113**, 292–9.
49. De Gobbi, M., Viprakasit, V., Hughes, J.R., *et al.* (2006). A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. *Science*, **312**, 1215–17.
50. Ioannidis, J.P., Ntzani, E.E., Trikalinos, T.A., *et al.* (2001). Replication validity of genetic association studies. *Nature Genetics*, **29**, 306–9.
51. Munafo, M.R., Matheson, I.J., and Flint, J. (2007). Association of the DRD2 gene Taq1A polymorphism and alcoholism: a meta-analysis of case-control studies and evidence of publication bias. *Molecular Psychiatry*, **12**, 454–61.
52. Ioannidis, J.P., Trikalinos, T.A., and Khoury, M.J. (2006). Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *American Journal of Epidemiology*, **164**, 609–14.
53. Ioannidis, J.P., Trikalinos, T.A., Ntzani, E.E., *et al.* (2003). Genetic associations in large versus small studies: an empirical assessment. *Lancet*, **361**, 567–71.
54. Wellcome Trust Case Consortium (2007). Genome–wide association study of 14,000 cases of seven common disease and 3,000 shared controls. *Nature*, **447**, 661–78.
55. Cox, A., Dunning, A., Garcia-Closas, M., *et al.* (2007). A common coding variant in CASP8 is associated with breast cancer risk. *Nature Genetics*, **39**, 352–7.
56. Carlborg, O. and Haley, C.S. (2004). Epistasis: too often neglected in complex trait studies? *Nature Reviews. Genetics*, **5**, 618–25.
57. Caspi, A., Sugden, K., Moffitt, T.E., *et al.* (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*, **301**, 386–9.
58. Cohen, J.C., Kiss, R.S., Pertsemlidis, A., *et al.* (2004). Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science*, **305**, 869–72.
59. Redon, R., Ishikawa, S., Fitch, K.R., *et al.* (2006). Global variation in copy number in the human genome. *Nature*, **444**, 444–54.
60. Sebat, J., Lakshmi, B., Malhotra, D., *et al.* (2007). Strong association of *de novo* copy number mutations with autism. *Science*, **316**, 445–9.
61. Gerlai, R. (2000). Targeting genes and proteins in the analysis of learning and memory: caveats and future directions. *Reviews in the Neurosciences*, **11**, 15–26.
62. Crawley, J.N., Belknap, J.K., Collins, A., *et al.* (1997). Behavioral phenotypes of inbred mouse strains: implications and recommendations for molecular studies. *Psychopharmacology (Berl)*, **132**, 107–24.
63. Lewandoski, M. (2001). Conditional control of gene expression in the mouse. *Nature Reviews. Genetics*, **2**, 743–55.
64. Nakazawa, K., McHugh, T.J., Wilson, M.A., *et al.* (2004). NMDA receptors, place cells and hippocampal spatial memory. *Nature Reviews. Neuroscience*, **5**, 361–72.
65. Tsien, J.Z., Huerta, P.T., and Tonegawa, S. (1996). The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. *Cell*, **87**, 1327–38.
66. McHugh, T.J., Blum, K.I., Tsien, J.Z., *et al.* (1996). Impaired hippocampal representation of space in CA1-specific NMDAR1 knockout mice. *Cell*, **87**, 1339–49.
67. Gordon, J.A. and Hen, R. (2004). Genetic approaches to the study of anxiety. *Annual Review of Neuroscience*, **27**, 193–222.
68. Low, K., Crestani, F., Keist, R., *et al.* (2000). Molecular and neuronal substrate for the selective attenuation of anxiety. *Science*, **290**, 131–4.
69. Rudolph, U., Crestani, F., Benke, D., *et al.* (1999). Benzodiazepine actions mediated by specific gamma-aminobutyric acid(A) receptor subtypes. *Nature*, **401**, 796–800.
70. McKernan, R.M., Rosahl, T.W., Reynolds, D.S., *et al.* (2000). Sedative but not anxiolytic properties of benzodiazepines are mediated by the GABA(A) receptor alpha1 subtype. *Nature Neuroscience*, **3**, 587–92.
71. Gross, C., Santarelli, L., Brunner, D., *et al.* (2000). Altered fear circuits in 5-HT(1A) receptor KO mice. *Biological Psychiatry*, **48**, 1157–63.
72. Copeland, N.G., Jenkins, N.A., and Court, D.L. (2001). Recombineering: a powerful new tool for mouse functional genomics. *Nature Reviews. Genetics*, **2**, 769–79.
73. Gong, S., Zheng, C., Doughty, M.L., *et al.* (2003). A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature*, **425**, 917–25.
74. Hatten, M.E. and Heintz, N. (2005). Large-scale genomic approaches to brain development and circuitry. *Annual Review of Neuroscience*, **28**, 89–108.
75. Lima, S.Q. and Miesenbock, G. (2005). Remote control of behavior through genetically targeted photostimulation of neurons. *Cell*, **121**, 141–52.
76. Arenkiel, B.R., Peca, J., Davison, I.G., *et al.* (2007). *In vivo* light-induced activation of neural circuitry in transgenic mice expressing channelrhodopsin-2. *Neuron*, **54**, 205–18.

2.5

The contribution of psychological science

Contents

2.5.1 Developmental psychology through infancy, childhood, and adolescence

William Yule and Matt Woolgar

Introduction

The child is father to the man. This saying seems so obviously true that it may surprise some people that it needs to be analysed and certain assumptions inherent in it need to be challenged if psychiatric practice across the lifespan is to be properly informed by findings from developmental psychology. This chapter examines different conceptualizations of children and childhood through the ages and the ideas and theoretical models that have shaped popular, as well as professional, views on how children develop. It notes that there are no overarching theories of child development, but rather a pot-pourri of smaller models, most of which address disparate aspects of development.

Developmental psychology is not just about charting the norms of development, although knowledge of such is essential in all clinical practice. Rather, there are many issues that need to be critically examined in trying to understand how individuals develop. Taking a developmental perspective is about integrating this knowledge and understanding the patient's presenting problems within such a framework.

The significance that the clinician will place on a particular piece of behaviour will depend not only on the child's sociocultural background, but also on the child's developmental age. Cox and Rutter[1] note four reasons for taking a developmental perspective:

1 Children behave differently at different ages. The clinician must be familiar with the range of behaviours and their age-appropriateness in separating the normal from the abnormal. For instance, simple consonant substitutions are widespread in the speech of pre-school children, but indicate some delay or deviation in the speech of teenagers.

2 Many aspects of behaviour can be viewed as progressing through a normal sequence. Admittedly, discrete stages are overemphasized by stage theorists such as Freud, Piaget, and Bowlby, whereas the continuities in development are more emphasized by social-learning theorists such as Staats, Bijou, and Baer. Either way, an understanding of the normal sequences and ages permits a judgement as to whether the child has deviated in his or her development.

3 Different stages of development are associated with different stresses and different developmental tasks. Bladder and bowel training are normally achieved between the ages of 2 and 4 years. Major stresses on the child or the family at the time may interfere with the achievement of proper bladder and bowel control. Mood swings are very common in adolescence, making it difficult to diagnose the severity of depression at this stage.[2,3]

4 An understanding of the processes which underlie both normal and abnormal development will help in the understanding of how the problems have arisen.[4] Such an historical perspective can help explain to the parents why a particular problem developed, as well as give possible clues for future programmes for prevention. A major implication of this for clinical practice is that it is always necessary to obtain a good account of the child's developmental history.

5 A better understanding of the *processes* underlying a child's development will lead to far better interventions and prevention.

Once we have a better understanding of the *distal* and *proximal* causes of behaviour, better targeted interventions will follow.

Developmental theories and views

There is a bewildering set of mini-models and mini-theories of developmental processes, each trying to deal with changes in children's functioning either at different periods in their lives or in different psychological functions such as perception, language, and memory. By and large, the different theories seem to ignore each other's work—and many also seem keener on theories than on data that might test the theories.

For example, Piaget's theories predominantly address how children develop a cognitive understanding of their world. His was a biological view of development, and his cross-sectional methodology emphasized the separation between the stages he posited. Staats[5] argued that most of the phenomena described by Piaget and his followers could be interpreted within a social learning theory framework that instead emphasized the continuity of development across stages.

Kohlberg's theory of moral judgement is a stage theory that differs radically from Piaget's in that the different forms of reasoning said to typify different stages can coexist. However, the way in which children (or adults for that matter) judge an ethical dilemma does not necessarily determine how they behave. Most financiers would have little difficulty in providing sophisticated moral judgements on Kohlberg type tasks, but many financiers also present the unacceptable face of capitalism in their ruthless dealings. It is not the case that the older we are, the wiser we behave.

In Freud's theory of psychosexual development, children are seen as passively passing through stages, their development being impeded by obstacles or even regressing in the face of trauma. This view owes more to literature than to science, and the evidence on children's psychosexual development clearly shows that whatever Freud was unaware of during the latency period, children are certainly far from inactive.[6]

Apart from being stage theories, these three sets of influential theories really have very little in common. The psychological mechanisms determining growth of cognitive understanding bear little relationship to any that supposedly underlie socio-emotional behaviour. None of the theories take into account all of the work done in perceptual development, language development, development of memory, development of peer relationships, development during adolescence, and so on. They pay little attention to the work on individual differences in personality or temperament, or to biological development generally.

A totally biological, determinist view of development was anathema to the new theorists of behaviour modification and behaviour therapy in the 1960s. It was seen as too pessimistic, offering little hope of change. By ignoring the biological basis of behaviour and seeking explanations solely in the here-and-now (proximal) influences on behaviour, they undoubtedly broke through to a much more optimistic era of interventions.

Simultaneously, child developmentalists were recognizing the contributions the child brought to all aspects of development. The child has increasingly been seen as an active participant in development. The direction of influence was not all one way: the child helped shape the environment. Thus, parents react to individual differences in children. Different children call out different responses from their social environment. As parents have known all along, children do have different temperaments from birth, and these shape how they develop.[7,8]

The implications of this for child psychiatry are many. For example, it implies that clinicians must take into account a child's temperament when planning treatment.[7,9] Children who are extremely introvert react differently to praise and punishment than children who are extremely extrovert.[9,10] They also respond to different teaching styles in the classroom. Such differences need to be accommodated in setting up individualized treatment programmes.

With young infants, it can be very reassuring to a parent to be told that anyone would find their unpredictable child difficult to rear. It can boost parental self-confidence to be told (when true) that their parenting style is perfectly adequate for most children—just not effective with this particular one. This reassurance should greatly alter the way such a parent participates in parent training programmes that are increasingly part of primary and secondary level child mental health services.[11]

All this is not to say that stage theories carry no implications for child mental health services. Far from it. It is very helpful to remember that young children think and reason about their worlds differently from older children. This has to be borne in mind when interviewing children, when trying to elicit their own understanding of their problem, and, equally, when giving them instructions, feedback, or explanations. However, it must again be emphasized that the stages should only ever be regarded as rough guidelines. We know that there are such wide individual differences in the rate at which children develop that we should never make assumptions about the individual child knowing only his chronological age.

Let us take one example that increasingly confronts clinicians—the issue of helping children deal with bereavement. It is not until around the age of 10 or 11 that *most* children appreciate that death is both universal and irreversible.[11–14] This helps explain why some younger children show an almost casual, matter-of-fact interest in death of a loved one and are less upset by it than adults are.[13] But it would be wrong to assume that all younger children fail to have an adult appreciation of the significance of death, and indeed some children as young as 4 years old have been found to have a mature understanding. Knowledge of the broad outline of the development of the conceptualization of death helps clinicians formulate their questions, but the onus must always be on the clinician to check whether or not the individual child conforms to the average. The adult's task may not be finished when they have helped a young child to understand bereavement at the level the child can cope with. That same child will probably want to revisit the issue when she is older and can understand it in a more mature way.[15] What is true of bereavement also holds true for understanding any other major life event and its effects on the child.

Critical issues in development

When one takes a closer look at how children develop, one cannot help but be amazed at the complexities of the process. Children the world over start using words around their first birthday and within a couple of years more, they are talking in complex sentences using complicated ideas. The contrast between the language development of most children and the minority who suffer a severe mental

handicap is devastating. Likewise, blind children start to smile at the same time as sighted children; deaf children start to use a similar range of phonemes; children in Japan, France, and Britain all start uttering the same range of sounds only to have them narrowed down to those they need in their native language—with the later consequence that they may not even be able to discriminate some of the unused sounds, let alone incorporate them when learning a foreign language. The broad developmental trajectory seems very similar across cultural groups, but particular children do not always follow the average in a smooth, predictable way.

Rutter and Rutter[16] draw attention to a number of issues that need to be considered when trying to understand developmental processes. Clinicians are understandably focused on trying to make sense of cases where something has gone wrong in development. Mostly in child psychiatry, abnormal behaviours of children are quantitatively different from normal rather than being qualitatively different. Disorders following brain damage or genetic/chromosomal abnormalities and many involving very severe degrees of mental handicap, including infantile autism, are increasingly recognized as being qualitatively different. Most of the other disorders seen in child and adolescent mental health services are probably best viewed as deviations lying at the extreme of a continuum. But why do some children break down under stress while others do not? Why are some more resilient than others? What factors protect children against environmental and social stressors? Is it really the case that severe depression in late adolescence is just the extreme end of a continuum ranging from happiness through sadness to suicidality? In order to tackle these issues, it is necessary to clarify some of the concepts of development.

1 One should not assume that the same mechanisms underlie both normal and abnormal development.

2 A biological perspective is necessary to understand human development fully. The brain is clearly the most important organ concerned—the genetic inheritance, insults during critical periods of brain growth, hormonal changes—all these have considerable influence on how children develop.

3 One has to expect both *continuities* and *discontinuities* in development. At times, continuities are intrinsic to the particular process as in language development; at other times, continuities—as in academic attainment—are in large part influenced by continuities imposed by the social environment. Parents concerned about education influence the choice of schools and provide support for learning.

4 The *timing* of an experience is as important as its nature. The brain is most vulnerable to insult when it is developing most rapidly, at and shortly after birth. Severe disruptions in caretaking have their greatest effects from around 9 months to 2 or 3 years. Before then, the infant does not show the same quality of selective attachments; after language is well established, the child can better hold the memory of a loved one, and that may act as a protection against the separation.

5 Children are *active* creatures. Not only do they call out responses from others, but as they develop cognitively and linguistically, they actively seek to make sense of their world. They appraise threat from others, even if they do not always get it right. When they are involved in a major catastrophe, their assumptive world[17] can be literally turned upside down and they take a long time to reconstruct the world as a safe place. The way the child interprets experience will come to determine in part how similar experiences are responded to in the future.

6 'Continuity may be heterotypic as well as homotypic'[16] (p. 8). The brilliant idea developed in the New York Longitudinal Study[18] of temperament was that rather than seeking evidence for predictability and continuity in particular infant behaviours across times when behaviour was developing rapidly, the investigators looked instead at how a variety of topographically different behaviours were expressed and found considerable continuities in such aspects as regularity of functions, strength of response, and predominant reaction to new stimuli. Thus, they adduced evidence of temperamental characteristics that were independent of the specific behaviours shown, and moreover, these temperamental characteristics proved to be predictive of later behaviour and adjustment.[16]

7 Both risk and protective factors, and the interactions between them, must be considered. Not all apparently adverse experiences are necessarily wholly bad for healthy development. In the same way that exposure to a virus or infection can boost resistance to infection, so exposure to mild stressors may boost resistance to other stressful experiences later. In part, this is the basis for stress inoculation therapy.[19] Some would argue that young children should have practice in separating from parents under enjoyable conditions so that in the event of a sudden, unexpected, or traumatic separation being necessary, the effects of experience will be mitigated.

8 As noted earlier, continuities may arise indirectly in that the way parents or society in general support attainment and in turn entry to the job market. The moderately high correlations between early attainment and later earning power are thereby in part determined and supported environmentally.

9 Similarly, the achievement of a particular behaviour may set in motion a chain of events. It is important to understand the processes underlying such a sequence. Too often studies are short-term and cross-sectional in nature and despite being aware of the pitfall of confusing correlation with causality, investigators remain prone to identifying a correlate as being a causal agent. For example, in the early days of studies of reading difficulties, it was noted that poor readers did badly on tests of visual perception. It was assumed that they therefore had a visual-perceptual deficit and generations of poor readers were subjected to hours of mindless tracing of lines and walking along benches. The end result was that they performed better on the particular visual-perceptual test but they were no better at reading! A different experimental design was needed to demonstrate causal relationships between psychological processes and poor reading,[20] and when that was understood, the way was open for better remedial work based on a proper understanding of causal mechanisms.

This can also be viewed as an error in confusing a risk *indicator* with a risk *process*. Forty years ago, studies of the dehumanizing effects of institutionalization on adults and children[21] found that poor living conditions and block treatment of residents were related to a greater risk of behavioural and emotional problems. In one set of studies, a good *indicator* of block treatment was whether patients had their own toothbrushes. Clearly, providing individual

toothbrushes to all would not make much difference if all the other aspects of institutionalization remained in force. A fuller understanding of the *process* of institutionalization is needed in order to be able to develop more humane care that improves development.

These critical issues demonstrate just how complicated the relationship between nature, experience, and development can be. But human beings are indeed very complicated, thank goodness, and so a proper appreciation of all these factors is needed in order to be able to understand how a particular child reached a particular point in development; to be able to predict what the future may hold for a child and to be able to develop rational interventions that have a hope of making a real difference to children's lives.

Developmental psychopathology

Developmental psychopathology emerged in the 1980s to bridge the rift between academic and clinical child psychology.[22,23] 'The developmental psychopathologist is concerned with the time course of a given disorder, its varying manifestations with development, its precursors and sequelae, and its relation to non-disordered patterns of behaviour' (p. 18).[23] Developmental psychopathologists, like social learning theorists, look to normal development to illuminate pathological development. They are interested in continuities and changes in behaviour across time. This fits in well with the tradition of risk research[24] and attempts to answer questions not only about why some children are more vulnerable than others, but also about what protective factors operate to lessen the impact of stressors.

Sroufe and Rutter,[23] following Santostefano,[25] articulated several propositions that are broadly agreed across the many different theories alluded to above:

(a) Holism

'The meaning of behaviour can only be determined within the total psychological context' (p. 20).[23] Thus, behaviour such as crying can only be evaluated according to the age of the child and the circumstances in which it occurs. Crying on separation would be seen as usual for a 3-year-old, but unusual in a 15-year-old. One cannot simply judge the significance of a behaviour simply on the basis of its physical, stimulus properties, but one has to evaluate it within the broader social context.

(b) Directedness

Children are not passive reactors to the demands of the environment. Development consists of a reorganization of previous elements, skills, and behaviour, not just a linear addition of skills.

(c) Differentiation of modes and goals

Over time, children's reactions to their environment become both more flexible and increasingly complex in organization. Thus, one sign of pathology is for children to get stuck in a particular way of trying to solve a problem.

(d) Mobility of behavioural functions

Earlier behaviour becomes integrated into later patterns, and 'the individual does not operate only in terms of behaviours that define a single stage. Especially in periods of stress, early modes of functioning may become manifest' (p. 21).[23] In other words, under stress, those patterns of behaviour that have most recently become integrated into the child's repertoire are most susceptible of disruption. This is very different from the unsatisfactory concept of regression in which all skills achieved remain available in the child's repertoire; some earlier ones also manifest at times of stress.

(e) The problem of continuity and change

Above all, development is seen as lawful, even though we are still far from understanding the processes involved in these laws. Sroufe and Rutter[23] emphasize: 'the continuity lies not in isomorphic behaviours over time but in lawful relations to later behaviour, however complex the links' (p. 21). As noted, Thomas, Chess, and Birch[18] were among the first to demonstrate continuities in the *style* of behaviour (temperament) rather than continuities of behaviour *per se*.

It is now recognized that there are many complex ways in which child behaviour is related to later and even adult adjustment.[26] One of the most powerful predictors of later adult psychopathology is inadequate peer relations. The mechanism by which these work may be due to two interacting processes: (1) Poor peer relations are signs of failure to adapt during childhood, and that failure persists; (2) social support later acts as a buffer against adult stressors.[23]

Clearly, this view of development, with its implications for psychopathology, is far removed from the lessons learned from the Skinner box. Yet what has been learned from the paradigms of classical and operant conditioning must also be integrated into ways that child therapists assess children's problems if we are to provide better treatments. This holistic view manages to incorporate ideas on the biological basis for behaviour and the notion of the child as an active participant interacting with his or her effective social environment within a broad social learning framework.[27,28] Understanding how a problem has arisen may provide useful guidance on what aspects to focus on, but the treatment will still focus on the present. There will be implications for maintaining treatment gains and preventing future problems, as well as implications for preventing such problems arising in other children.

For clinicians more used to working with adult patients, it is worth pointing out that children differ in many ways from their grown up counterparts. This has implications for improving diagnostic classificatory systems in that both DSM and ICD are still too adult oriented and pay insufficient attention to developmental aspects of disorders.[29–31]

Garber[32] makes the point that children differ from adults in cognition, language, physiology, and emotions. Such maturational differences may impact children's abilities to experience or express certain affects, cognitions, or behaviours, and thus the manner in which symptoms are expressed may differ over the course of development (p. 32). Recent work on the effects of major disasters and acute stress on children's adjustment, it became evident that children as young as 8 years old showed most of the symptoms of post-traumatic stress disorder (**PTSD**)—with unpleasant thoughts, poor concentration, and sleep disorders predominating.[33,34] Parents and teachers were often unaware of the nature and extent of the children's subjective distress, and only sympathetic but direct questioning elicited the full spectrum of symptomatology.

The criteria for PTSD are less appropriate for children under 8 years of age. Pre-school children often react with more repetitive play and drawing than older ones. Even the youngest children will report very disturbing, intrusive thoughts about the disaster. Scheeringa *et al.*[35] suggest varying criteria for making the diagnosis

of PTSD in young children. Leaving aside the logical problem of altering criteria but keeping the same name for the supposed underlying condition, this clearly is one aspect of the isomorphism mentioned earlier. It is also interesting to speculate whether the repetitive play seen in 6-year-olds is functionally equivalent to the intrusive thoughts seen in 10 year olds, and when the one changes into the other.

Garber also notes that some disorders, such as mental handicap and autism, first manifest in childhood and persist into adulthood. Others, such as encopresis and enuresis manifest in childhood, but rarely persist into adulthood unless part of a more global developmental delay. Some, such as anorexia and bulimia, are more typical of adolescence. Suicide, although rare before puberty, is rapidly becoming the major cause of death in adolescence, but peaks in old age. Major depression and schizophrenia are rare in childhood, although precursors are being more firmly established. While the wish to treat disorders in childhood so as to prevent them continuing to adulthood is laudable, treating them to improve adjustment during childhood is equally valid.

Linking developmental psychopathology to developing children

These exciting ideas need to be brought out of experimental settings and into clinics—the aim of developmental psychopathology theorists. In the second half of this chapter, some of the key aspects of child development relevant to clinical practice will be highlighted in this framework.

(a) Individual differences

Children differ in their personality, character, or temperament. European psychologists have always emphasized these individual differences and adduced evidence that many were based on biologically determined ways of responding to the world. While the three factor structure of personality developed by Eysenck evolved into the big five structure of today, an issue remained as to how one could demonstrate *continuity* in personality from a very early age. The same issue bedevilled studies seeking to establish continuity in differences in intellectual functioning—simply put, little babies show such a different repertoire of behaviour from mobile, talking pre-school children that it was impossible to test the same behaviours at different ages.

In part, the problem was solved in the New York Longitudinal Study[18] (see Box 2.5.1.1 One) by looking at differences in *style* of behaviour rather than differences in content or topography. By avoiding talking of biologically based personality, the findings reawakened interest in the genetics of individual differences.

In the original sample of 78 babies, the investigators were able to demonstrate considerable individual differences in temperament. There was good inter-rater reliability in making these judgements and direct observations agreed well with reports from parents. The temperamental characteristics were found to be stable over both the short (2 years) and medium term and even predicted reasonably well into late childhood.

Three broad types of temperament were characterized—children who were regular, predictable, and showed generally positive reactions—the easy babies; those who were almost the opposite—whom Chess called the mother killers and a sizeable minority who were slow to warm up to new situations but who adjusted eventually. The difficult children were over represented in those who developed behavioural problems in later childhood.

Box 2.5.1.1 Temperamental characteristics

The repertoire of infant behaviour is so different from that of the pre-school child that it proved very difficult to examine whether there were any continuities of behaviour across the ages. Thomas *et al.*[18] and their collaborators in the New York Longitudinal Study were among the earliest to show continuities across the age, but continuities in *style* of behaviour rather than *content*. Through a mixture of observation and exhaustive interviews with mothers, they originally developed nine different categories of behavioural style or *temperament*.

Mainly

1. Active	1. Passive
2. Regular: (e.g. in feeding and sleeping habits)	2. Irregular
3. Reacts intensely (strongly)	3. Reacts mildly
4. Shows approach behaviour to new people, places, toys, foods, etc.	4. Shows withdrawal behaviour
5. Adaptive—adapts fairly easily to change	5. Non-adaptive
6. Reacts easily to small changes	6. Reactions slow
7. Predominantly good moods—happy, contented disposition	7. Predominantly bad disposition—fretful, hard to please
8. Persistent in what she/he is doing as regards time, and in the face of difficulties	8. Non-persistent
9. Easily distracted from whatever she/he is doing	9. Not easily distracted

(b) Cognitive development

One of the major aspects of development that exercises parents and teachers alike is how best to improve the intellectual functioning of children, be that in language, reading, memory, or general intelligence. To what extent individual differences in these areas are predominantly related to heredity or to environment continues to be a popular, if sterile, source of argument. Clearly, the end result comes from an interaction between genetic predisposition and experience, but there remain issues of how best to manipulate the environment so as to help children gain their maximum potential. To that end, an understanding of modern behavioural genetics is essential. Here, some of the methodological issues in assessing babies' cognitive processes and findings in cognitive development, language, and memory are considered.

(c) Getting inside the baby's head

Until a baby starts to speak, it is difficult to know what they are thinking. Fond parents interpret wind-driven grimaces as smiling; every child is seen as recognizing people and being smart from a young age. But how can one tell what really goes on inside a baby's head?

Robert Fantz[36] studied infants' eye movements and used fixation time as a measure of preference for different stimuli. Film recordings were made of light reflected off the baby's eyes. In a study of 30 infants tested weekly from 1 to 15 weeks of age, it was shown that the infants spent longer looking at complex than simple patterns. This demonstration of a clear perceptual preference in the first few weeks after birth gave the lie to the view that all begins as a big, booming confusion. Infants as young as one week show clear preference for human faces over other shapes presented to them some 10 in. away.

Fantz took advantage of the technology of the day. Since then, others have utilized measures of changes in temperature, in galvanic skin response and in heartbeat to provide behavioural indices of preferences. In addition, investigators have used various indices from learning and conditioning paradigms.

Results from such studies highlight the extent to which infants enter the world ready for social interaction. Infants are highly dependent on their parents because of the cortical developments and increase in brain volume required to allow the special human cognitive characteristics to develop that take place *after* the baby has made its way down the birth canal. Infant's readiness to be part of a social interaction, imbued with intentionality, appears to serve a survival function.

The neonate has remarkable hard-wired skills that can be detected from within moments of birth, while the neonate is in a period of alert inactivity. However, there are limits to the extent of the early ability. Eyesight acuity is about 1/30th of adults at birth and develops over the next 4 years, although significant improvements arise by the second month. Despite problems with visual acuity and focusing, even neonates can track objects, albeit jerkily, and scan for simple, visual features, such as linearity, luminance, and symmetry when stimuli are within about 10 in. of their face.[37] These intrinsic visual abilities combine with preferences for certain spatial forms found in the human face, e.g. a preference for vertical over horizontal symmetry. Fagan[38] tested visual preference in 7-month-old babies and later measured their intelligence when they were 3 and 5 years old. The time spent looking at the novel stimulus when a baby correlated 0.42 with performance on the later Picture Vocabulary Test. Thus, it is getting easier to measure various indices of baby's reactions and some of these are found to be usefully predictive of later development and adjustment.

Hearing is pretty much complete by the 5th or 6th week of foetal life, with sophisticated auditory discrimination abilities along the dimensions of pitch, volume, tone, and duration. Hence neonates may come into the world already familiar with soap opera theme tunes, but so too are they familiar with their mother's voice and are able to orientate towards them from the start. Neonates have a preference for women's voices over men's and unlike the specificity shown towards the mother, appear to show little preference for their father's voice compared with other men.[39]

Bathed in amniotic fluid in the womb, it is unclear whether the foetus has strictly speaking *smelled* the mother before birth, although their olfactory system is well-developed *in utero*. Nonetheless, within hours of birth neonates demonstrate a preference for their mother's smell and, within days, can reliably orient their head towards breast pads worn by their mothers over those worn by other women.[40]

Meltzoff and colleagues have demonstrated that a neonate's social sensitivity is not just a passive turning to stimuli, they can also actively imitate them.[41] Indeed, within moments of birth, neonates are able to mirror adult's facial expressions such as tongue protrusions or mouth openings. Quite how this process operates at a cognitive level is unclear. One can speculate on its function as a way to provide a satisfying contingency for a parent that helps imbue the infant with a sense of communication, agency, and personhood. Overall, neonate's abilities to discriminate between stimuli and their hard-wired preferences for some constellations over others lead them to orient towards their caregivers, as the very beginnings of a selective attachment.

(d) Piaget and cognitive development

Piaget was a biologist who studied amoebas for his doctoral work. Biological models found useful for that purpose clearly influenced the way he regarded cognitive development. He held to a sort of moving homeostasis—children develop a model of the world. New information that challenges that model is gradually assimilated and eventually the model accommodates the new ways of thinking—a bit like an amoeba reaching out to a piece of food, surrounding it and assimilating it.

According to this theory, the child passes through three broad stages of thinking (see Box 2.5.1.2): A sensori-motor stage, a long stage where they think in terms of what he called concrete operations, and finally a stage where they can think logically.

Piaget's stage theory has been very influential and helpfully sparked off a great deal of research which has led to a much better understanding of how children develop cognitively. However, his original models were somewhat simplistic and to have seen only three major stages covering a period of such rapid development

Box 2.5.1.2 Piaget's stages of cognitive development

Birth to 18 months: sensori-motor stage

Cognition is based mainly on the child's actions and six sub-stages were described. A key concept at this stage is that of object permanence—the ability to understand that an object continues to exist even when it is out of sight. Infants of 12 months will continue to look for an object where an experimenter hides it, even when they watch the experimenter move it. It is as if the object belongs only in a particular place.

18 months to 12 years: concrete operations

This is the stage of concrete operations. At the early stage, language develops rapidly. Children begin to demonstrate symbolic play, showing that they have memories and internal representations. Many ingenious little experiments were developed to illustrate how children's thinking about their world develops. Various other conservation tasks—of length, mass, and number for instance—convince teachers that children think differently about the world than adults and this has had a major—if not always beneficial—effect on ways of teaching.

12 years and over: formal operations

From around puberty onwards, the child is able to formulate and test hypotheses about the world. The child realizes that mass and volume can be altered in many ways, but they remain essentially unchanged. Children can examine their own thought processes and can begin to reason more logically.

strikes one as inadequate. Moreover, for the clinician, the question one often wants to raise is how can one use this understanding to bolster the reasoning of a child who is developmentally delayed—what Piaget witheringly dismissed as the American question. Piaget tended to argue that children could not be hurried through the stages. However, critics soon argued that some of the regularities that were apparently replicated in his work owed more to the manner in which the tasks were presented to the child than to any necessary underlying cohesion in types of thinking.

A typical experiment is to give a child two pieces of clay that are identical. Then one is rolled out into a sausage and the child is asked if they remain the same. Alternatively, the child is shown two test tubes of differing diameters. The same amount of liquid is poured into each, but, of course, reaches different heights. Which test tube has more liquid? Not surprisingly, younger children make more errors than older ones and it has been shown that conservation of mass (seeing that the quantity of material remains the same in spite of changes in shape) is acquired around 7 to 8 years, while conservation of *weight* is not achieved until 9 or 10 years. It is not until 11 or 12 that the child typically thinks that each shape also occupies the same amount of space (i.e. achieves conservation of *volume*).

Bruner was one of the earliest to demonstrate that children's judgements could be radically manipulated by small changes in the ways the tasks were presented. For example, in the studies on conservation of volume, a screen was placed between the child and the test tubes so that only the tops showed but not the levels reached by the liquid. This simple change meant that many younger children now understood that by pouring liquid from one container to another—perhaps something done daily in the bath—nothing had changed. Some children as young as 4 years were able to perform the revised task, but once they had to make the judgement again without the screen, they reverted to the more primitive form of reasoning. In other words, young children are more at the mercy of their perceptual impressions, a finding that needs to be taken on board in many circumstances.

(e) Language development

Children communicate from the beginning. They signal their basic needs, often by crying. Parents soon learn to respond to the differing signals. Their own child's crying can be very aversive to most parents and so they quickly learn how to switch the noise off! Autistic and brain damaged children produce grossly abnormal cries, but often parents with little experience do not recognize the unusual nature of the cry. Whilst lower primates and other animals can communicate, none use language in the flexible way that human infants come to do. Language—both spoken and written—is truly the most human of attributes.

Sophisticated social interaction is a precursor to the development of language. Though the infant is hard-wired for social interaction, it is not clear that these behaviours are strictly communicative in the sense of being intentional acts. What appears to be an early communicative competence may be best understood in terms of a *social releaser* model, where caregivers' interactions are triggers that release hard-wired reflexes from the infant, little different from the familiar palmar grasp reflex of the first month that causes an infant to grip an adult's thumb, even in their sleep. In this formulation there is no sense of a two-way communicative interaction being shared between carer and infant.

From approximately 2 months of age, infants appear to be active participants in the interactions that caregivers and their babies regularly play. When carers adjust their speech to the infant's preferences for high-pitched, rhythmic, and repetitive speech, often referred to as *motherese* and hold their face about 12 in. away, within the infants optimal focal range, a dynamic dance of turn-taking between carer and infant can be observed. The infant's attention is held, shared smiling can occur, the pitch and tone of utterances becomes matched and the infant can be observed to make distinctive formations with their lips referred to as 'pre-speech'. Caregiver and infant certainly appear to be experiencing social reciprocity. Some clever experiments based on perturbing the reciprocal nature of the infant caregiver interaction have challenged the notion of a social releaser model and given support to the infant's sophisticated early developing sensitivity to social reciprocity.

The still face paradigm requires the infant's partner to disturb the relationship by substituting their dynamic interaction by holding a still, expressionless face.[42] Presented with this perturbation, infants will initially increase their gesturing, as if to draw the adult back in, but become increasingly distressed, crying, or grimacing as this unnatural state continues. Of course, this is a relatively gross perturbation and technological advances have allowed developmentalists to alter more subtle aspects of the real-time contingency between an infant and its mother using a video link. Murray and colleagues filmed mothers and infants interacting over a video link and then altered the contingency by replaying a segment of the mother's interaction to the infant.[43] Infants as young as 8 weeks were able to detect this perturbation, and initially showed reduced levels of engagement followed by distress and protest. The levels of engagement rose again when the link was switched back to a live interaction. The replay condition demonstrated that simply presenting caregiver behaviours that had been previously adequate to elicit sustained interaction were no longer appropriate when they were taken out of the context of the dynamic flow between infant and caregiver. Thus it was not simply the presentation of maternal signals that triggered the release of infant communicative behaviours, but like a dance or a conversation, it was the to and fro sequence of behaviours between infant and caregiver that set the context of the flow of behaviours.

Language has not been the subject of serious study until the past 50 years. Beginning with simple descriptive studies of the acquisition of words, the complexities of grammar and cross-cultural comparisons, studies of language development now encompass many other dimensions from neuropsychological, brain imaging, and genetics. What stands out is just how difficult it is to affect the onset of language by manipulating the environment. It takes extraordinarily environmental deprivation to interfere with language development and even then, when the environment is normalized, considerable catch up occurs.

The grunts and single syllables of the first couple of months soon give way to the production of the full range of phonemes and babbling between 2 and 4 months. States of feelings are communicated clearly by 3 to 7 months. From 6 months, the baby begins to imitate simple sounds and unreinforced phonemes disappear from the vocabulary. By 8 months, the baby begins to utter two syllable combinations such as ma-ma and da-da—amongst the easiest to produce physically and, perhaps not coincidentally, the most emotionally evocative words parents want to hear. At around 12 months, the first true word appears, usually as the infant takes his first

Box 2.5.1.3 Language development

Vocabulary grows astronomically from 1 to 3 years and beyond. Children's sentences get longer and more complex. A good working approximation is that the average length of sentences is in keeping with the number of years of age:

Average length of sentences—2–5 years

Age in months	24	30	36	42	48	54	60
Average number of words per sentence	1.7	2.4	3.3	4.0	4.3	4.7	4.6

From Smith.[46]

step—girls reaching the milestone slightly ahead of boys. Vocabulary grows astronomically from one to three years and beyond. Children's sentences get longer and more complex. A good working approximation is that the average length of sentences is in keeping with the number of years of age.[46]

Studies in the 1960s and 1970s established that language development followed complicated underlying rules. The idea that children learned language by successive approximations to adult speech being differentially reinforced was quickly laid to rest. Such techniques have an important place in remedial intervention for children with deviant language development, but for ordinary children the sheer inevitability, speed, and beauty of acquisition is overwhelming. This led many to postulate that there is a genetically encoded Language Acquisition Device that guides communication, although not the particular form of language that a particular child will develop. That still depends on what language they are exposed to.

Early sentences are telegraphic with some words acting as pivots on which other words hang to form flexible sentences. Thus, the pivot, 'all-gone' can have 'sock', 'milk', or 'daddy' added to create a whole range of meaningful simple sentences. Children develop rules for expression. A common error is for them to extract a rule and then overgeneralize it to a situation that is an exception in their particular language. For example, the present tense is used before the past tense. Instead of saying 'I went' children often form past participles by adding -ed to a stem and come up with the often heard 'I goed'. This shows they are learning a rule, even though they make some mistakes on the way.

Different ways of describing and classifying language disorders have been proposed in the past few years.[44] Bishop's[45] twin study finds that when language disorder is defined as a discrepancy between non-verbal IQ and language score the heritability is far less than when language delay and disorder are considered without reference to IQ. This is important for clinicians in identifying children with such difficulties .

(f) Memory

One of the most intriguing observations in the current child development literature is the contrast between the ever increasing evidence of just how complicated children's cognitive development is and the phenomenon known as infantile amnesia. Basically, people have very few memories before the age of 3 years. Clearly from all that has been described earlier about the differential reactions of babies to specific stimuli, to their recognizing their mother's voice or holding out their arms to their father rather than to a stranger, children increasingly have some form of central representations that they can work on. Yet, these early memories are not accessible in later life. It is really not until language is well established that people have what is ordinarily termed memory for past events.

Clearly, infantile amnesia poses a major challenge to any theory of child development or personality that tries to link very early experiences with later adjustment. Yet, early experience does affect the way in which relationships are formed, so what are the mechanisms? As the different types of memory (See Box 2.5.1.4) are better understood, so better assessment of these functions is possible.

Some very recent work has looked at implicit and explicit memory as well as attentional processes in children with generalized anxiety disorder, PTSD and depression. Broadly speaking, the preliminary findings are in accord with the voluminous findings with adult patients, namely that depressed children tend to have biases in memory for sad things, while anxious children do not. In contrast, children with anxiety disorders (including PTSD) have biases in attention that make them attend more to threatening cues in their environment—or at least to threatening words projected on computer screens. Studies using these adult-generated paradigms but utilized within a developmental framework should greatly increase our understanding of why some children break down under stress and others do not. Biases in cognitive processing of

Box 2.5.1.4 Memory

Goswami[48] summarizes many ingenious experiments that establish the parameters of children's memory. While parents and teachers often talk about children having problems with memory and even short-term memory, developmental psychologists have worked on much more complex paradigms and identified a number of different memory systems:

Recognition memory is simply the ability to realize that a particular stimulus has been encountered before. Recognition is always easier than recall—as those learning a foreign language can testify. Once established in the first year of life, recognition memory does not change much.

Implicit memory is another term for memory without awareness. Although not able to put it into words, people can act differently to previously exposed stimuli than to novel ones. This seems to be fully developed by around 4 years of age.

Episodic memory involves awareness. It is this memory system that organizes memories into stories or scripts concerning similar activities. These scripts contain both temporal and causal information. The ability to learn sequences in a particular chain of events does develop with age. It is now that one realizes that there needs to be some mechanism to get rid of many of the memories for everyday activities, otherwise the whole memory will get clogged up with non-essential information. In other words, memory processes are seen as being very active with some memory traces remaining in (technically) short-term memory for only a few seconds unless operated upon and stored in long-term store.

Eye witness memory has taken on a special importance as children are expected to testify in court on things they have witnessed happening to themselves or to others. Children can

recall things fairly accurately, as long as deliberate leading questions are not put to them.[49] Three-year olds are more suggestible than 5-year olds. Experienced adults—such as psychiatrists or judges—evaluate children's responses to questioning about a real event using the child's behaviour while giving their answer. Where children gave firm answers with lots of supporting details, they were judged to have clear and accurate memories. Where children were uncertain and hesitant, they were seen as fabrication whereas they were hesitating because the questioner was asking the wrong questions—ones that were in conflict with what had actually happened—they were the ones who were actually telling the truth. The younger, confident children were often telling the adult what they wanted to hear! A great deal more needs to be done in relation to helping children recall what has happened to them without using leading questions.

Working memory was seen by Baddeley and his co-workers as consisting of a central executive processing linked to two separate subsystems: a phonological loop and a visuospatial sketch pad. Information decays in the phonological loop in 1 or 2 s, unless it is actively rehearsed. It is thought that children predominantly use a visuospatial encoding until they switch to the phonological/verbal system around the age of 5 years. Deaf children continue to rely on the visual encoding for much longer.

emotional reactions are implicated and can now be studied more readily.[47]

Social and emotional development

Alongside cognitive development, children are developing both socially and emotionally. It has been recognized for years that children brought up in institutions, away from their natural parents, often develop serious and subtle problems in social interactions and emotional development.

Arising in part from his studies of infants in institutions and his collaborative studies of children's reactions to being in hospital, as well as his dissatisfaction with contemporary psychoanalytic theory, Bowlby turned to ethology for an understanding of early infant relationships. He came to view the intense relationship between the infant and the caretaker, usually but not always the biological mother, as serving a biological survival value and as having been produced by natural selection. Bowlby proposed an attachment system that served to keep the child safe during the extended dependency of human infancy, by ensuring proximity to specific and reliable caregivers.[50]

As memory develops in the first year of life, and as the infant becomes more able to express emotion and to move independently, so there is evidence for selective attachment. This is shown round about 6 to 8 months onwards by the upset at leaving the attachment figure, by seeking comfort when threatened and by a general wariness of strangers.

The idea that attachments were simply associative learning—the baby comes to love the person who feeds him—was quickly dismissed by the evidence from Harlow's studies of infant monkeys. They attached to the cuddly terry-towelling surrogate rather than the wire surrogate where they were fed. Rather, selective attachments in the human served to protect the infant during the prolonged period of helplessness. The function of attachment changes over the years, with children using an attachment figure as a secure base from which to explore. Thus, almost paradoxically, the well attached toddler may move away from the attachment figure more than the insecurely attached counterpart. Attachment is not the same as clinginess.

The attachment system is just one amongst several innate systems proposed to operate in infancy, and it is the interplay between competing behavioural systems that led to the gold standard measure of attachment in infancy. Mary Ainsworth's based her Strange Situation Procedure (**SSP**) around observations of infants in Uganda.[51] The SSP pits the attachment system against the fear and exploration systems, in a structured, unfamiliar situation comprised of increasing levels of stress (e.g. strange room, strange adult, and separation from the caretaker). Observations are focused primarily on proximity seeking and contact maintaining behaviours during reunions with the caregiver. This measure is made possible by the development of stranger fear and mobility around the age of 9 months, and loses its validity by about 18 months as the infant's cognitive development permits different responses to separation, e.g. symbolic representation and language. After this point attachment assessments become less about observed behaviours and more about language and play.[52,53]

The SSP originally suggested three types of infant classification, representing organized responses to the prevailing environment: Insecure Avoidant (A) infants whose behaviour implied an attempt to minimize the importance of the attachment relationship and whose impoverished play and exploration functioned as a distraction from their need of comfort from the caretaker; Secure (B) infants who sought sufficient comfort from the caretaker to be able to quickly return to play and exploration; Insecure-Ambivalent, or Insecure-Resistant (C) infants whose behaviour suggested a maximization of the attachment relationship and an ambivalent

Box 2.5.1.5 Attachment

Ainsworth developed the theory and a method for detecting individual differences in attachment. She introduced the strange situations test in which an infant is left in the care of a stranger for a few minutes. The observer then notes how the infant copes both with the separation and with the reunion. These observations led to a tripartite classification of attachments:

- **Secure attachment:** The infant tends to seek proximity and contact with the attachment figure, shows a preference for the mother over a stranger, and shows very little distress before and after separation.
- **Avoidant insecure:** The baby does not cling when held, avoids the mother (or caretaker) during the reunion, and does not differentiate greatly between the caretaker and a stranger.
- **Resistant insecurity:** The infant resists contact and interaction with the mother. There is great distress at reunion.
- **Disorganized or avoidant insecurity:** This category had to be introduced when it was found that infants of severely depressed and abusive mothers showed a mixed pattern of attachments. The child shows contradictory patterns and unusual patterns of negative emotions.

preoccupation with the caretaker, at the expense of returning to play and exploration. Mary Main's observations of children in high-risk environments, particularly those exposed to maltreatment or parental psychopathology, led to an additional coding category, independent of, but complementing the tripartite Ainsworth system. Main noted a high rate of bizarre infant behaviours in the SSP of some children, which seemed to denote a lack of a clear, organized adaptation to separation from the caregiver. There was also evidence of apprehension and fearful behaviour in the presence of the caregiver, these infants seemed disorganized/disoriented, and were categorized as an Insecure-Disorganized (D) attachment pattern.[54]

A second innovation that Bowlby's attachment theory brought to the study of child development was the notion that the experience of the caregiving environment becomes represented as a cognitive heuristic or *internal working model of attachment* (**IWM**). This model is a dynamic, lifelong *work-in-progress*, which evolves as environmental experience accumulates but which also guides the selection of current behaviour on the basis of past experiences. Thus the infant's model of attachment is shaped by the experience of their caregiving environment, but simultaneously influences interactions with the infant's environment by selecting the behaviours most likely to achieve the goals, of say, proximity, based on what was previously successful. The IWM reflects an adaptation to the current environment, but the flexibility to adapt to changes in the environment is weighted by the accumulation of previous experience. So the possibility of fundamental change in attachment style is always possible but more difficult the longer the old environment prevailed.

Bowlby proposed that as time goes on the IWM becomes less about selecting the appropriate behaviours to protect the dependent infant via maintaining proximity to a safe and reliable caretaker, but moves to a level of representation of the self in relation to others. There is evidence to support the idea that the quality of attachment in infancy has some degree of influence over a range of social and emotional outcomes in early childhood, with peers, teachers and even stretching into later adolescence and early adulthood with partners and, as parents, their own offspring.[55] There is good evidence of some degree of continuity across the lifespan, but as Bowlby's original thesis would predict, expectations of continuity should not be overstated, and the degree of continuity is particularly low in high-risk samples.[56]

By the age of 3 to 4 years, most children show good evidence of having multiple attachment figures. By this age, their memories are so much greater that they are less dependent on the physical presence of the attachment figure to provide security and comfort. Bowlby saw good attachments in infancy as laying the basis for future social and intimate relationships and there is currently an explosion of work re-examining psychiatric conditions from an attachment perspective. Thus, Bowlby saw the child as developing a cognitive model of his effective social world, in keeping with the views of other cognitive theorists such as George Kelly, Aaron Beck, and Ronnie Janoff-Bulmann.

(a) Attachment, psychopathology, and 'reactive attachment disorder'

There is sometimes confusion surrounding the ideas of the relationship between insecure attachment and psychopathology. An organized-insecure attachment is an appropriate response to a particular environment. At most an organized-insecure attachment might be a vulnerability factor for later problems, in the context of other risks, but by no means would it be a major risk factor for pathology in itself. In terms of outcomes such as behaviour problems, the greatest influences still appear to be social risk factors, although attachment may confer some small additional risk.[57] There is stronger evidence for insecure-disorganized attachments predicting psychopathology. Aggressive behaviour problems in preschool are associated with insecure-disorganized classifications in infancy.[58] One study has suggested some degree of homotypic continuity into late adolescence with elevated rates of dissociative phenomena echoing the bizarre stilling and freezing behaviours typical of a insecure disorganized classification in infancy.[59] Of course, disorganized classifications most often occur in high-risk environments, so it is not always obvious if consequent psychopathology is a result of the attachment classification or of the continuing impact of a high-risk environment. Clearer evidence for the enduring impact of disorganized attachment could be collected when the high-risk environment that has led to the disorganization is terminated.

A negative aspect of the focus on attachments has been the emergence of an ill defined disorder described as reactive attachment disorder which seems to be diagnosable by the presence of any or all of a long list of symptoms and signs that haunted previous generations under such labels as Minimal Brain Dysfunction and the like. This has been associated in some countries with the use of assaultative holding therapies intended to break the child's resistance to forming attachments. Clearly, more careful studies need to be undertaken to try to pinpoint the subtle social difficulties presented by children whose early lives have been disrupted in fostering and adoption, but care also needs to be taken to adduce evidence for appropriate interventions.

Having good, supportive social relationships has been shown to be a major protective factor in the aetiology and maintenance of many psychiatric disorders. The ability to make and maintain friends—initially of the same age and later of any age—is often related to the existence of disorders such as personality disorders, social anxiety disorders, depression, and even post-traumatic stress disorders. The emphasis on social skills training for socially inadequate persons points to the early basis for such deficits even though they may have their greatest impact in adulthood. Some children are less sensitive to social cues than others, and some children misinterpret the intentions of other people. Both lead to difficulties, albeit of different sorts.

Boys and girls tend to develop different types of social relationships. It may be inconveniently politically incorrect to note that children tend to prefer playing with others of the same gender when freed from adult influence. Boys tend to play in larger, looser groups in which issues of dominance and play fighting predominate; girls relate in smaller groups of more intense relationships, with best friends often changing.

Concluding comments

Taking a developmental perspective to mental health issues should apply across the lifespan. Psychiatrists working with adults need to understand where their clients are coming from and where they are going to. They need to understand the pleasures and pressures that children bring to their parents, and where appropriate they should be considering the impact of parental illness on the children. The

institutionalized separation of child and adult psychiatry (in terms of service delivery) should not lead to a separation in ways of considering the developmental context of presenting problems.

This chapter has shown that there are many small, focused models of development that deal with discrete areas of development. Stage theories emphasize differences at different stages; social learning theories emphasize continuities on processes of development. As long as practitioners are aware that when they say a child is at a particular stage, this is but a rough guide to describing the child, which may be fine. It is when such models are taken literally, that oversimplification leads to poor practice. There is no one overarching theory of child development and while this may be inconvenient for examiners, it truly reflects the rich diversity of human development. By paying more attention to the interactions between biological, social, and psychological factors, a better understanding of healthy, normal development will emerge. Empirical studies will help identify risk and protective factors which in turn will lead to better mental health promotion and more effective interventions when mental disorders manifest.

Further information

Field, T. (2007). *The amazing infant*. Oxford, Blackwell.

Grossman, K.E., Waters, E., and Grossman, K. (2006). *Attachment from infancy to adulthood: the major longitudinal studies*. Guilford, New York.

http://www.apa.org/topics/topicchildren.html

http://www.devpsy.org/links/index.html

References

1. Cox, A. and Rutter, M. (1985). Diagnostic appraisal and interviewing. In *Child psychiatry: modern approaches* (2nd edn) (eds. M. Rutter and L. Hersov). Blackwell, Oxford.
2. Rutter, M., Graham, P., Chadwick, O., *et al.* (1975). Adolescent turmoil: fact or fiction? *Journal of Child Psychology and Psychiatry*, **17**, 35–56.
3. Harrington, R. (1993). *Depressive disorder in childhood and adolescence.* Wiley, Chichester.
4. Rutter, M. (1975). *Helping troubled children*. Penguin Books, Harmondsworth.
5. Staats, A.W. (1971). *Child learning, intelligence and personality: principles of a behavioral interaction approach.* Harper & Row, New York.
6. Rutter, M. (1980). Psychosexual development. In *Scientific foundations of developmental psychiatry* (ed. M. Rutter), pp. 322–39. Heinemann Medical, London.
7. Berger, M. (1985). Temperament and individual differences. In *Child psychiatry: modern approaches* (2nd edn) (eds. M. Rutter and L. Hersov). Blackwell, Oxford.
8. Kagan, J. (1997). Temperament. In *Handbook of child and adolescent psychiatry*, Vol. 1: *Infants and preschoolers* (ed. J.D. Noshpitz), pp. 268–75, Chapter 27. Wiley, New York.
9. Keogh, B.K. (1982). Children's temperament and teachers' decisions. In *Temperamental differences in infants and young children* (eds. R. Porter and G.M. Collins), pp. 269–78. Pitman, London.
10. Eysenck, H.J. and Eysenck, S.B.G. (1969). *Personality structures and measurement*. Routledge and Kegan Paul, London.
11. Scott, S. (2003). Integrating attachment theory with other approaches to developmental psychopathology. *Attachment and human development*, **5**, 307–12.
12. Childers, P. and Wimmer, M. (1971). The concept of death in early childhood. *Child Development*, **42**, 1299–301.
13. Koocher, G.P. (1974). Talking with children about death. *The American Journal of Orthopsychiatry*, **44**, 404–11.
14. Wolff, S. (1969). *Children under stress*. Allen Lane, Penguin Press, London.
15. Monroe, B. and Kraus, F. (eds.) (2005). *Brief interventions with bereaved children*. Oxford University Press, Oxford.
16. Rutter, M. and Rutter, M. (1993). *Developing minds: challenge and continuity across the life span.* Basic Books, New York.
17. Janoff-Bulman, R. (1985). The aftermath of victimization: rebuilding shattered assumptions. In *Trauma and its wake*, Vol. 1 (ed. C.R. Figley). Brunner/Mazel, New York.
18. Thomas, A., Chess, S., and Birch, H.G. (1968). *Temperament and behavior disorders in children*. University of London Press, London.
19. Meichenbaum, D. (1993). Stress inoculation training: a 20-year update. In *Principles and practice of stress management* (2nd edn) (eds. P.M. Lehrer and R.L. Woolfolk). Guilford, New York.
20. Bradley, L. and Bryant, P.E. (1983). Categorizing sounds and learning to read **B** a causal connection. *Nature*, **301**, 419–21.
21. King, R.D., Raynes, N.V., and Tizard, J. (1971). *Patterns of residential care: sociological studies in institutions for handicapped children.* Routledge & Kegan Paul, London.
22. Cicchetti, D. (1984). The emergence of developmental psychopathology. *Child Development*, **55**, 1–7.
23. Sroufe, L.A. and Rutter, M. (1984). The domain of developmental psychopathology. *Child Development*, **55**, 17–29.
24. Garmezy, N. (1974). Children at risk: the search for the antecedents of schizophrenia. I: conceptual models and research methods. *Schizophrenia Bulletin*, **8**, 14–90.
25. Santostefano, S. (1978). *A biodevelopmental approach to clinical child psychology*. John Wiley & Sons, New York.
26. Rutter, M. (1983). Continuities and discontinuities in socioemotional development: empirical and conceptual perspectives. In *Continuities and discontinuities in development* (eds. R. Harmon and R. Emde). Plenum Press, New York.
27. Bandura, A. (1973). *Aggression: a social learning analysis*. Prentice-Hall, London.
28. Yule, W. (1978). Behavioural treatment of conduct disorder. In *Child psychiatry: modern approaches* (eds. M. Rutter and L. Hersov). Blackwell, Oxford.
29. Cantwell, D.P. (1988). DSM-III Studies. In *Assessment and diagnosis in child psychopathology* (eds. M. Rutter, A.H. Tuma., and I.S. Lann), pp. 3–36. David Fulton Publishers, London.
30. Rutter, M. and Shaffer, D. (1980). A step forward or back in terms of the classification of child psychiatric disorders? *Journal of the American Academy of Child Psychiatry*, **19**, 371–94.
31. Yule, W. (1981). The epidemiology of child psychopathology. In *Advances in clinical child psychology*, Vol. 4 (eds. B.B. Lahey and A.E. Kazdin), pp. 1–51. Plenum, New York.
32. Garber, J. (1984). Classification of childhood psychopathology: a developmental perspective. *Child Development*, **55**, 30–48.
33. Yule, W. and Williams, R. (1990). Post traumatic stress reactions in children. *Journal of Traumatic Stress*, **3**, 279–95.
34. Yule, W., Perrin, S., and Smith, P. (1999). Post-traumatic stress disorders in children and adolescents. In *Post-traumatic stress disorders: concepts and therapy* (ed. W. Yule), pp. 25–50. Wiley, Chichester.
35. Scheeringa, M.S., Zeanah, C.H., Drell, M.J., *et al.* (1995). Two approaches to the diagnosis of Posttraumatic Stress Disorder in infancy and early childhood. *Journal of the American Academy of Child and Adolescent Psychiatry*, **34**, 191–200.
36. Fantz, R.L. (1961). *The origin of form perception*. Scientific American, Reprint 459.
37. Morton, J. and Johnson, M. (1991). CONSPEC and CONLERN: a two-process theory of infant face-recognition. *Psychological Review*, **98**, 164–81.

38. Fagan, J.F. (1984). The relationship of novelty preferences during infancy to later intelligence and later recognition memory. *Intelligence*, **8**, 339–46.
39. DeCasper, A.J. and Spence, M.J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior & Development*, **9**(2), 133–50.
40. Porter, R.H., Makin, J.W., Davis, L.B., *et al.* (1992). Breast-fed infants respond to olfactory cues from their own mother and unfamiliar lactating females. *Infant Behavior & Development*, **15**(1), 85–93.
41. Meltzoff, A. and Moore, M. (1989). Imitation in newborns: exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology*, **25**, 954–62.
42. Tronick, E.Z., Als, H., Adamson, L., *et al.* (1978). The infants' response to entrapment between contradictory messages in face-to-face interactions. *Journal of the American Academy of Child Psychiatry*, **16**, 1–13.
43. Murray, L. and Trevarthen, C. (1986). The infant's role in mother-infant communications. *Journal of Child Language*, **13**(1), 15–29.
44. Yule, W. and Rutter, M. (eds.) (1987). *Language development and disorders*. MacKeith Press, London.
45. Bishop, D.V.M. (1998). In *Perspectives on the classification of specific developmental disorders* (eds. J. Rispens, T. van Yperen, and W. Yule). Kluwer Academic, Dordrecht, The Netherlands.
46. Smith, M.E. (1926). An investigation of the development of the sentence and extent of vocabulary in young children. *University of Iowa Studies in Child Welfare*, **3**(5).
47. Dalgleish, T., Taghavi, M.R., Neshat-Doost, H.T., *et al.* (2003). Patterns of processing bias for emotional information across clinical disorders: a comparison of attention, memory and prospective cognition in children and adolescents with depression, generalized anxiety and Posttraumatic Stress Disorder. *Journal of Clinical Child and Adolescent Psychology*, **32**(1), 10–21
48. Goswami, U. (1998). *Cognition in children*. Psychology Press, Hove.
49. Bruck, M., Ceci, S., and Hembrooke, H. (1998). Reliability and credibility of young children's reports. *American Psychologist*, **53**(2), 136–51.
50. Bretherton, I. (1992). The origins of attachment theory: John Bowlby and Mary Ainsworth. *Developmental Psychology*, **28**, 759–75.
51. Ainsworth, M.D.S., Blehar, M.C., Waters, E., *et al.* (1978). *Patterns of attachment: a psychological study of the strange situation.* Lawrence Erlbaum, Hillsdale NJ.
52. Main, M. and Cassidy, J. (1988). Categories of response to reunion with the parent at age 6: predictable from infant attachment classifications and stable over a 1-month period. *Developmental Psychology*, **24**, 415–26.
53. Green, J., Stanley, C., Smith, V., *et al.* (2000). A new method of evaluating attachment representations in the young school-age children: the Manchester Child Attachment Story Task (MCAST). *Attachment & Human Development*, **2**(1), 48–70.
54. Main, M. and Solomon, J. (1990). Procedures for identifying infants as disorgainzed/disoriented during the Ainsworth Strange Situation. In *Attachment in the preschool years: theory, research and intervention* (eds. M.T. Greenberg, D. Cicchetti, and E.M. Cummings), pp. 121–60. University of Chicago Press, Chicago.
55. Grossmann, Klaus E., Grossmann, Karin, Waters, and Everett. (2005). Attachment from infancy to adulthood: the major longitudinal studies. Guilford Press, New York, US.
56. Weinfield, N.S., Sroufe, L.A., and Egeland, B. (2000). Attachment from infancy to early adulthood in a high-risk sample: continuity, discontinuity, and their correlates. *Child Development*, **71**(3), 695–702.
57. Belsky, J. and Fearon, R.M.P. (2002). Infant-mother attachment security, contextual risk, and early development: a moderational analysis. *Development and Psychopathology*, **14**(2), 293–310.
58. Lyons Ruth, K., Easterbrooks, M.A., and Cibelli, C.D. (1997). Infant attachment strategies, infant mental lag, and maternal depressive symptoms: predictors of internalizing and externalizing problems at age 7. *Developmental Psychology*, **33**(4), 681–92.
59. Carlson, E.A. (1998). A prospective longitudinal study of attachment disorganization/disorientation. *Child Development*, **69**(4), 1107–28.

2.5.2 Psychology of attention

Elizabeth Coulthard and Masud Husain

Introduction

Attention is generally taken to be the process by which people are able to concentrate on certain information or processes, while ignoring other events. It appears to be a fundamental attribute of human brain processing, although difficult to pin down in terms of mechanism. Psychologists have attempted to fractionate attention in many different ways, using ingenious behavioural paradigms. In this section we, too, will consider different aspects of attention: selective, phasic and sustained, divided and executive control of attention. However, it would be fair to say that all these aspects of attention do not normally operate in isolation. Instead they interact, and deficiencies in one aspect of attention, for example, in a patient population, often to do not occur in isolation. Functional imaging and lesion studies of attention have proliferated in recent years, attempting to place a neurobiological framework to these varied processes. In general, these studies also tend to confirm the view that attention is likely an emergent property of widespread brain networks, with a special emphasis on frontal and parietal regions of the human brain (Fig. 2.5.2.1). In this discussion we illustrate several aspects of attention with examples particularly from literature on visual attention, which is the most widely studied area, but it should be appreciated that many of the concepts discussed here extend to other domains. In fact, there is a good deal of evidence to suggest that several aspects of attention operate at a supra- or cross-modal level allowing integration of information from different sources.

Recent studies suggest there are two fronto-parietal networks: (Fig. 2.5.2.1) a *dorsal* parieto-frontal network involving the superior parietal lobe (SPL) and dorsal frontal regions such as the frontal eye field (FEF); and a *ventral* network involving the inferior parietal lobe (IPL), temporoparietal junction (TPJ) and inferior frontal gyrus (IFG). In addition, dorsomedial frontal areas, including the anterior cingulate cortex (ACC) and pre-supplementary area (pre-SMA) may play a key role in flexible control of attention for strategic behaviour.

Selective attention

Selective attention refers to the processes involved in selecting relevant information and filtering out irrelevant items from the vast array of information we are exposed to. The brain has limited capacity: it simply cannot process everything it is exposed to. Nor would it be sensible for it do so because the majority of sensory input to which it is exposed is not behaviourally relevant. Therefore there is a need for mechanisms to select the

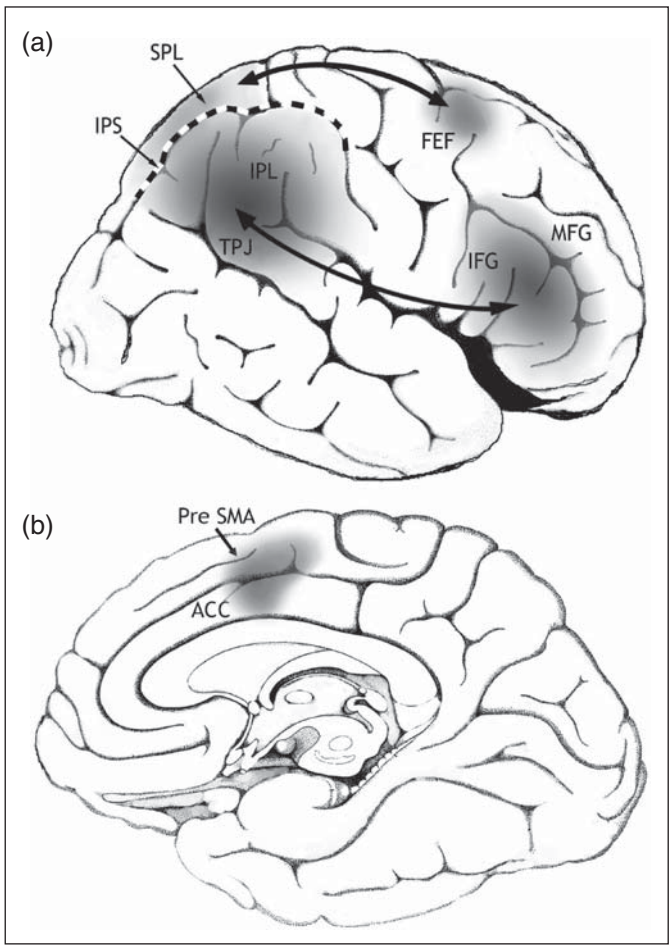


Fig. 2.5.2.1 Lateral (a) and medial (b) regions in the right hemisphere involved in attention. (IPS intraparietal sulcus; MFG, middle frontal gyrus.)

most behaviourally significant and important material, and dispose of trivial, unimportant information. Such selection may occur 'bottom-up', driven by competition between sensory inputs, or 'top-down',
guided by the goals that an individual might have at any moment in time.[1]

Thus if a ball is unexpectedly hurled towards an observer, his attention is likely to be captured—'bottom-up'—by the sensory input of a projectile moving at high velocity towards his head. In this case, selection has been driven by the most perceptually salient item in the external environment. 'Top-down' attention mechanisms, on the other hand, concern selection that is biased by internal goals. For example, consider the processes involved in looking for a friend in a busy, crowded train station. Here the selection process is driven by the features you are searching for: your friend's hair colour, facial features, height all play a role in guiding this search. Importantly, under these circumstances, even perceptually very salient items may be filtered out—or not attended—if they are irrelevant to the task.

But when does selection occur? Is it early or late in the processing of sensory information? This is an issue that dominated attention research for many years. Some investigators proposed that selection occurs early, directly after analysis of the physical characteristics or features of sensory stimuli, but before they are fully identified. According to this view, unattended information receives little or no further processing from this point on. By contrast, others argued that *all* stimuli are analysed up to the point that they are identified. Selection occurs only after this, late in the processing stream. So items that are eventually ignored or unattended, i.e. those that are not selected, actually receive considerable processing before they are discarded. Note that this late selection model allows for the possibility that items which are eventually ignored may nevertheless be processed to a deep level. Thus, even though they may not be attended to, they have the potential to influence our actions subliminally. Most researchers would now agree that there is good evidence for *both* early and late selection systems in the human brain. In fact, whether selection occurs early or late is likely to be influenced by the specific demands of the task.

Two highly influential experimental paradigms used to study selective attention in healthy humans and patients have both focused on *spatial* attention in the visual system. Posner first developed a spatial cueing task in which subjects view a display consisting of a central cross on which they are asked to maintain fixation throughout the trial. On either side of the cross, there are two square boxes. Participants are instructed to ignore a cue which consists of transient illumination of either the left or right box. At varying intervals after the cue, a target stimulus (an asterisk) appears in either left or right box and subjects are required to press a response button as quickly as possible. In 80 per cent of trials the cue is 'valid' in that it occurred at the location of the subsequent target. However, in the remaining 20 per cent of trials the cue is 'invalid', appearing in the box opposite that in which the target subsequently appeared. In healthy volunteers, reaction times to targets appearing where valid cues appear are significantly shorter than when invalid cues are presented.

This critical finding suggests that attention can be spatially localized like a beam or 'spotlight'. Moreover, attention can be captured by the abrupt onset of the cue so that visual processing is selectively oriented towards it, thereby improving responses if a target subsequently appears there. On invalid trials, Posner argued, attention would first have to disengage from the invalidly cued location and then shift to the correct location before engaging it. Note that such shifts occur covertly in the absence of overt eye movements; they represent shifts of visual processing from one location to another across a representation of space in the brain. Subsequent studies have shown that the orienting of attention to a spatial location not only appears to speed up detection of a visual stimulus, as measured by reaction times, but also can improve discrimination of items from non-targets. The neurophysiological mechanisms underlying such boosting of performance are currently the subject of intense scrutiny.[2]

The second experimental task that has proven to be extremely important in selective attention research is the visual search paradigm developed by Treisman. In this task, subjects have to find a target shape embedded among distractors. In simple so-called 'feature search', a target may be defined by a unique feature, e.g. a red circle among green circles is defined uniquely by its colour and therefore 'pops-out' among the distractors. Treisman has considered that such feature searches can occur *pre-attentively* in parallel across the visual scene, without the need for a spotlight of visual attention. She argued that attention needs to be deployed in more complex tasks where a target may share one or more features with distractors, e.g. finding a red circle among green circles and

red squares. In this case, a single feature (colour or shape) is not sufficient to define the target. Instead, the visual system has to find the unique conjunction of red colour and circular shape, and the target does not pop-out to the observer. Such 'conjunction searches', Treisman argued, requires the spotlight of attention to shift serially from one location to the next, inspecting each item in turn. In her model, spatial attention acts to bind or glue together features occupying the same location in space, e.g. the colour, form, luminance, and other attributes that belong to an object at one location in space.

Both the Posner cueing task and Triesman's conjunction task have been used in neuroimaging studies of healthy volunteers and patients with focal lesions on the brain.[3] The results of such studies suggest that regions within the parietal and frontal cortex play a critical role in deploying spatial attention (Fig. 2.5.2.1). In general, dorsal regions of this parieto-frontal network have been implicated in the spatial shifts of processing attention that occupy such a key role in the models of selective attention developed by Posner and Treisman.

Phasic, sustained, and vigilant attention

Several groups have made a distinction between phasic alertness, sustained attention, and vigilance. In general, 'phasic alertness' is used to refer to improvements in performance that follow a warning signal, e.g. an auditory tone or visual cue. The tasks used in such studies are often very simple, comparing the time to respond to a particular stimulus when it is preceded, or not, by a warning tone. Over the course of a few hundred milliseconds after such a warning signal, performance can alter appreciably, with reaction times first declining in the interval from 100 to 500–1000 ms after the tone, and then increasing again thereafter. Thus, this attentional facilitation is limited to a very narrow window in time and refers to the ability of the brain to respond better when warned to expect an upcoming cue to act.

By contrast, 'vigilance' is the term used to refer to the ability to be in a state of readiness to detect small, infrequent changes in the environment which occur at random intervals over *prolonged* periods of time (often hours or tens of minutes). The early studies on vigilance were conducted to assess the ability of radar operators to detect infrequent signals on their screens in the Second World War. Over long 'vigils' the ability to detect rare changes falls—the so-called 'vigilance decrement'—and this may be modulated by the frequency of target and non-target (distractor) stimuli as well as the memory load imposed on observers.

'Sustained attention' is best considered to be at the other end of the continuum from vigilance. Here, too, observers may have to respond over a prolonged period, but the information flow is rapid, with a high frequency of events to monitor and respond to. An everyday, extreme example may be the interpreter who has to give a simultaneous translation of a press conference. In the laboratory, sustained attention may be studied using stimuli presented at a rapid rate, as in many continuous performance tests (CPTs). Such experiments reveal that even healthy subjects may show 'lapses of attention' under such circumstances, with errors of omission (failing to respond to a target stimulus) or errors of commission (responding to a non-target), or increase in mean response time. Another measure that has attracted interest, of late, particularly in patient populations, is the variability of reaction times in such tests.

Functional imaging and lesion studies of patients have emphasized the role of the right lateral frontal lobe in aspects of vigilance and sustained attention. However, it is clear that the inferior parietal lobe of the right hemisphere also has a critical role to play in these functions. The distinction between frontal and parietal contributions remains to be established but, unlike spatial attention studies, investigations of non-spatial vigilance and sustained attention suggest that *ventral*—and not dorsal—frontal and parietal regions of the right hemisphere have a special role in these processes (Fig. 2.5.2.1).[4]

Divided attention

Divided attention is the ability to concentrate on more than one activity at once. Although we often execute two tasks simultaneously ('dual task'), performance on one or both tasks may be impaired compared to doing either alone, e.g. when driving and having a phone conversation. Experimental findings that demonstrate such decrements in performance raise important questions regarding the mechanisms underlying attention. Many investigators have proposed that if a second task leads to deterioration in performance of the first, one may conclude that the two tasks depend on the same brain resources. Because such resources are limited, there will be a decrement in performance once resource limits are reached. Some authors have argued for a single, central resource while others have presented evidence for multiple resources, but the basic concept is the same across these accounts: tasks compete for finite brain resources and if they share those resources, performance suffers.

Other investigators have raised the possibility that it is not just 'resources' that we need to consider but also processing limitations. Two tasks might be difficult to perform simultaneously because they both use a single processing channel or because the tasks interfere with each other. Different paradigms have suggested 'bottlenecks'—mechanisms that are dedicated serially to only one task at a time or have limited resources to spread over two tasks. Such bottlenecks may exist at the level of attentional focus, storage in visual short-term memory (VSTM), and motor preparation.[5] Consider first the issue of processing bottlenecks.

The attentional blink paradigm is used to investigate temporal limitations of attentional processing. Subjects view a stream of individually presented letters and are required to report when they see either of the two target letters (say X and Y). People generally struggle to report a second target if it falls within 360 ms of the first, despite being able to report the second target when not attending to the first. Processing the first object before being free to process the next item appears to require up to 360 ms. Thus speed of attentional processing may act as a limiting factor in dual task processes. Functional imaging studies suggest, once again, that a fronto-parietal network is critical in this regard. Activity in fronto-parietal areas occurs only when the second stimulus is reported. In contrast, even when the second stimulus is undetected, there is still activity in the early visual areas. These findings point to a bottleneck in attentional processes which occurs after basic visual processing has begun, but before conscious perception, consistent with pre-attentive and attentive stages of visual processing as discussed in the section on selective attention above.

The amount of information we can encode at any one time is also limited. The capacity for holding objects in visual short-term memory (VSTM) is around four objects. Diversion of attention

away from an object reduces the accuracy with which it can be encoded suggesting that attention is an important limiting factor in the number of items held in VSTM. Both electrophysiological and functional imaging experiments have suggested that the capacity for storage of objects in VSTM is related to posterior parietal and occipital cortex activity.

As well as these perceptual and memory limitations in capacity, there are limits at the motor selection stage of processing. The psychological refractory period (PRP) refers to the delay in the second response when a subject has to respond to two stimuli one after the other. The shorter the time difference between the two stimuli, the greater the delay in the second response. This bottleneck may be at the level of response selection rather than a pure perceptual or attentional slowing. However, although it is often the case that performing two simultaneous responses with a single effector (e.g. the hand) may be impossible, the PRP cannot simply be explained in terms of a motor selection bottleneck.

When subjects are asked to make two responses to the same attribute of an object (name the colour and press a button for the colour), they can select the appropriate responses at the same time. In addition the PRP is much longer when required to make an incongruent response (e.g. saying 'A' in response visual presentation of the number 1) than a congruent response (e.g. saying 'one' in response to visual presentation of number 1) to the second stimulus. Pashler has proposed that at any time we have a number of response selection rules, with each rule specifying condition-action linkages or associations, i.e. which condition is associated with which action. According to this view, the bottleneck is at the level of rule application, but each rule can specify multiple motor responses.

But how and where rules are selected and maintained? Many studies have suggested a critical role for frontal—so-called 'executive'—control systems when rules need to be applied, as we shall discuss in the next section. Performing two *well learned* or relatively *automatic* tasks together might not lead to performance impairment because these tasks do not require input from such 'executive' control systems to implement rules and select appropriate responses.

In summary, there are multiple possible levels at which processing capacity may be limited. Both the speed of attentional processing and the number of items that can be attended to at once and held in memory are limited. In addition, supervisory or executive control regions may regulate motor output perhaps by implementing goals or strategies. Whether these bottlenecks are due to one resource being serially applied to each stimulus or a limited capacity system shared simultaneously by two processes, but perhaps with a bias toward one, is currently unclear.

Executive control of attention

How are the subtypes of attention mentioned above organized so that they are activated at the appropriate time? Events in our environment reflexively engage attentional networks and this type of stimulus-driven or bottom-up activation may underlie some processes such as rapid shifts in spatial attention to highly salient events. However, everyday experience tells us that rather than always reacting to our surroundings, we are able to generate and implement plans to complete complex tasks and sometimes ignore highly salient events. This has led to the idea of a supervisory or executive attentional control system. Evidence for the presence of such a system comes from patients with brain damage, particularly those with frontal damage.

Patients with frontal lesions have difficulty maintaining attention on a task. They may be highly distractable, or find it difficult to divide attention between competing task demands in an optimal way. Some patients also encounter problems shifting from one task type or task rule to another, often demonstrating perseveration—

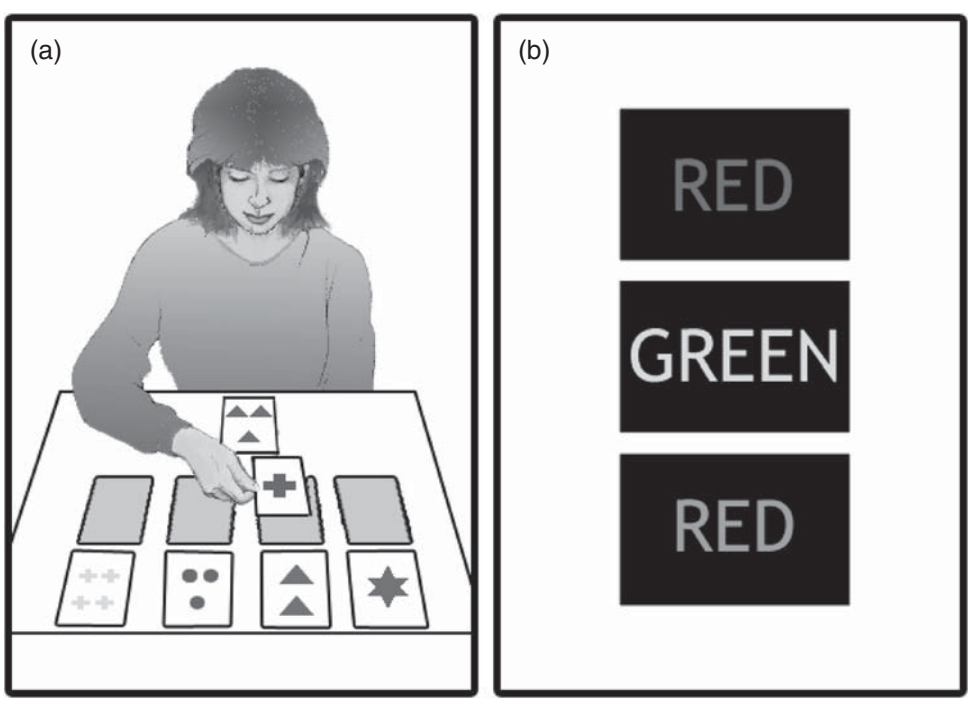


Fig. 2.5.2.2 In the Wisconsin card sorting test (a), the patient has to place a new card alongside one of the visible cards according to a rule to match colour, suit or number. The rule changes during the task and the learning occurs by trial and error. Patients with frontal lobe damage fail to update the new rule. The Stroop task (b) requires subjects to give the colour of the text. In the first box this is easy because the colour of the text and the meaning of the word are congruent (both red). The second and third boxes illustrate the incongruent condition. This is more difficult because the text colour is different from the meaning of the word. Normal individuals are slower to respond in the incongruent than congruent conditions.

applying the 'old' task rule inappropriately when they should be applying the 'new' one. For example, on tasks such as the Wisconsin card sorting task, where subjects establish the criteria for grouping playing cards into sets (according to either colour, or suit, or number) and then have to change to a new grouping rule, such individuals have great difficulty shifting to a new rule (Fig. 2.5.2.2a). Multiple components of executive control including flexible task-switching and response inhibition, as well as vigilance and error monitoring, are required to successfully sort the cards when the rule is switched.

Norman and Shallice developed a model of attention control called the supervisory attention system that conceptualized how this behaviour might be organized. They proposed that routine or well-learned behaviour occurs because perceptual information activates a set of schemas that then triggers appropriate motor output. However, these perceptual-motor associations would no longer be appropriate when the rule changed, for example when one had to sort according to suit rather than number in the Wisconsin card sorting task (Fig. 2.5.2.2a). Under these circumstances, they proposed that the supervisory attention network alters the bias of the schemas so that different motor outputs are triggered by perceptual events. Therefore, rather than processing simple stimulus-response associations such as always signalling the right foot to move from the brake to the accelerator when a green traffic light appears, brain regions within the supervisory network would respond to information at a more abstract level. When driving, for example, one might want to *inhibit* the prepotent response associated with the green traffic light (to press the accelerator) if confronted by a variety of visual inputs such as someone still crossing the road or a car stalled in front. *Response inhibition* in this and other contexts is thought to be one of many abstract functions undertaken by the supervisory control systems.

What are these abstract processes and which regions of the brain perform these supervisory operations? Both lesion studies and functional imaging work have contributed to understanding executive functions. Functional imaging studies provide information about areas activated in association with certain processes while lesion investigations show which regions are critical for normal behaviour. In paradigms such as the Stroop task, subjects are asked to report either the colour of print used to type a word or read the word which can be the same (congruent) or a different (incongruent) colour to the print (Fig. 2.5.2.2b). To avoid making errors in the incongruent condition one is required to focus attention on the target information and suppress the unwanted response. This is usually associated with a reaction time delay in the incongruent compared to the congruent condition. The delay is greater when subjects have to report the colour of the print and inhibit their reading of the word, presumably because word reading is a more hard-wired or automatic than colour naming. Functional imaging and lesion data from subjects performing the Stroop tasks show that medial and lateral frontal as well as parietal areas appear to form a network for executive control of attention.

There is recent evidence that each of these regions has a distinct role within the executive control network. Specifically, the left lateral frontal cortex is thought to maintain and flexibly update task rules, whereas, right lateral prefrontal cortex is critically involved in *inhibiting* the prepotent response associated with a stimulus. Dorsomedial frontal regions (Fig. 2.5.2.1b) reliably activate when executive control is required and are thought perhaps to play a key role in monitoring errors. In addition, part of this area—the anterior cingulate cortex—is considered by some investigators to be important for mediating the physiological autonomic response to demanding circumstances. More posterior regions including parietal cortex, may also contribute to attentional control but, to date, their role has been less well investigated. Research into how areas involved in executive control act and interact to modulate attention is a rapidly expanding area of cognitive neuroscience, likely to yield important insight into mechanisms behind flexible and efficient behaviour.

Acknowledgements

This work is funded by the Wellcome Trust.

Further information

Johnson, A. and Proctor, R.W. (2004). *Attention theory and practice*. Sage Press, California.

Styles, E.A. (2006). *The psychology of attention*. (2nd edn). Psychology Press, New York.

Ward, A. (2004). *Attention a neuropsychological approach*. Psychology Press, Hove.

References

1. Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, **18**, 193–222.
2. Treue, S., (2001). Neural correlates of attention in primate visual cortex. *Trends in Neurosciences*, **24**(5), 295–300.
3. Corbetta, M. and Shulman, G.L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, **3**, 215–29.
4. Husain, M. and Nachev, P. (2007). Space and the parietal cortex. *Trends in Cognitive Sciences*, **11**, 30–36.
5. Marois, R. and Ivanoff, J. (2005) .Capacity limits of information processing in the brain. *Trends in Cognitive Sciences*, **9**, 296–305.

2.5.3 Psychology and biology of memory

Andreas Meyer-Lindenberg
and Terry E. Goldberg

Memory in psychiatric practice

Memory is the ability to store, retain, and retrieve information. This cognitive function plays a key role in psychiatry. Dementia and the amnesic disorders have memory dysfunction as a defining feature. Intrusive and recurrent emotional memories are one of the most distressing symptoms in post-traumatic stress disorder. Although not as obvious, problems with memory are also commonly revealed on testing in schizophrenia. Remembered episodes are often a focus in psychotherapy, as is the acquisition of new habits and response patterns. An ability to understand and assess memory is therefore important for the practising psychiatrist. In this chapter, basic neurobiological and psychological information

on memory will be reviewed. We have tried to cover a very broad field in a concise manner and give the interested reader a sense of the key memory systems and subsystems that are thought to be important for human information processing in health and in disease. We have emphasized the conceptual over the theoretical and key findings over the experimental details where possible. At times, we have not carefully separated the cognitive and neuroanatomical levels of analysis, both because they are sometimes almost inextricably bound and because it made our explanations clearer not to do so. Necessarily but not happily, we have omitted many important and active areas of investigation.

Forms of memory

One of the key discoveries of cognitive neuroscience is that 'memory' is not an unitary function, but consists of several forms that can be dissociated neurally and are differentially impacted by psychiatric disorders.[1] Several approaches can be taken to subdivide memory. One of the most straightforward is by the duration over which information is retained. In this way, ultrashort-term, short-term, and long-term memory can be distinguished. **Ultrashort-term**, also called **sensoric** or **echoic/iconic** memory, lasts from milliseconds to seconds and consists of a brief and modality-specific retention of sensory information. For example, most people are able to 'replay' the auditory trace of the last second or so of a conversation, or briefly maintain a scene visualized after they close their eyes. In contradistinction, **short-term memory** has been shown to be relevant to a large number of psychiatric disorders. In short-term memory, information is briefly (over a period from seconds to minutes) held in mind, often through a process of rehearsal. A typical example is remembering a phone number from reading it to dialing without writing it down. A key feature of short-term memory is capacity limitation: most people are able to retain about seven items in short-term memory.[2] A specific form of short-term memory that has received considerable interest is **working memory**: the ability to hold information in mind that is necessary for a task at hand, but not present in the environment. This faculty is often regarded as a 'mental workspace' that is critical for information manipulation and goal-directed adaptive behavior, and the association of working memory with specific brain systems and psychiatric disorders has been widely studied. Rehearsal is also one important mechanism by which material is being transferred into **long-term memory**, which refers to the ability to retain information for time periods lasting from minutes up to the life span of the individual. This form of memory, which is also of major clinical importance, is not clearly capacity limited and is thought to depend on more enduring changes in neuronal structure and connectivity, raising the question of where in the brain these enduring memory traces, or 'engrams', are stored, how they are encoded for storage and how they are retrieved from it.

A second important subdivision is whether the content of memory can be consciously and intentionally retrieved (this is called explicit or **declarative** memory) or not (nondeclarative or **implicit** memory) (**Fig. 2.5.3.1**). Declarative memory is further subdivided in memory of facts and memory of events. Memory of events, which often includes recollection of temporal, spatial and emotional circumstances, is called **episodic** memory. Questions such as 'What did you have for breakfast this morning?' or 'Where did you go to school?' access episodic memory. Several features of episodic memory bear mention for their clinical relevance. For example, people can often say that they have seen a specific item before or that it 'feels familiar' without being able to recall the specifics of where and when (the episodic context). As discussed below, some evidence suggests that familiarity and recall may be supported by different brain regions. Intense feelings of familiarity without recall are experienced as *déjà vu* in a psychiatric context. People can also be convinced to remember events that did not, in fact, happen. These so-called **false memories** are also encountered in psychiatry. There has been much work on paradigms that produce false memories in normal individuals that address how they develop and how and why they might be successfully rejected. In one canonical account of episodic memory, the medial temporal lobe system (MTL) stores or indexes contextual markers that serve to bind feature information of a memory into an episodic configuration.

In contradistinction to episodic memory, memory of facts is not connected to specific experience. It is called **semantic** memory, and recollection of facts or vocabulary are examples of information that has become independent of episodic memory. Over time such material is thought to be stored in neocortex and can be retrieved without engagement of medial temporal lobe structures (MTL). Various models of the distinctions between semantic and episodic memory have proposed that while learning in the episodic system is rapid and can be based on a single trial or exposure, storage of information in the semantic system occurs slowly over time and only after multiple exposures or activations [3]. Some accounts suggest that the MTL may also be involved in semantic memory, for example for separation or 'decompression' of stimuli previously learned as a unit.[4,5] We have chosen not to review semantic memory in further detail because of space limitations and its complex overlap with psycholinguistics.

Whereas the subdivision of declarative memory is comparatively simple, implicit memory encompasses quite a heterogeneous group of functions that are supported by different brain systems. Among them are **procedural memory**, which refer to the gradual acquisition of sensorimotor, perceptual or cognitive skills through repeated exposure, **priming** (the facilitation of a response to an item if it was previously encountered), **conditioning** as well as various phenomena wherein a previously acquired response is gradually reduced or lost, such as **extinction**. Since the distinction between declarative and implicit depends on whether or not a memory process supports conscious recollection, clinicians must be careful not to confuse this with the properties of a given neuropsychological test; for example, a test of sentence completion requires conscious

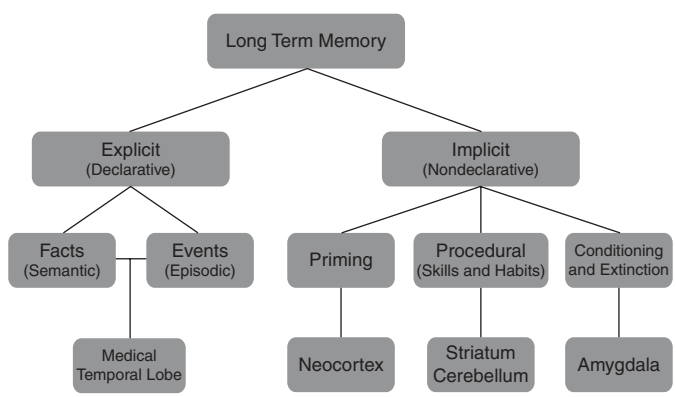


Fig. 2.5.3.1 Classification of memory and associated neural systems

production of words, but performance may strongly depend on implicit processes such as priming.

Although a taxonomy of memory systems is useful, many everyday tasks require functionality from several memory domains, and even relatively subtle changes in task demands may disrupt the balance among those cognitive systems. More refined analysis of learning tasks suggests that a variety of learning systems may mediate performance,[6] and that a given neural system participates in several forms of memory. For example, the hippocampal formation (HF), which is critical for episodic memory, is also thought to play a role in learning sequences so that indirect relations can be specified.[7] Thus, the HF becomes critical not during a>b, b>c, and c>d discriminations, but for the discrimination of the critical indirect b>d probe (e.g. if John is taller than Bill and Bill is taller than Mary and Mary is taller than Ellen, then Bill must be taller than Ellen).

Cellular and molecular mechanisms of memory

Memory is one of the most impressive examples of neural plasticity: the ability of the nervous system for enduring change triggered by external events. Arguably the best-studied cellular mechanisms underlying plasticity are long-term potentiation (LTP) and long-term plasticity (LTD), which mediate enduring changes on the level of the synapse.[8] Both have been best characterized in the hippocampus, one of the key structures for declarative memory.

By stimulating presynaptic fibres in the hippocampus (especially the CA1 section) with a brief pulse of high-frequency electric impulses, a long-lasting increase in responsiveness of the postsynaptic cells to low-frequency stimulation is reliably observed that can last for weeks (**Fig. 2.5.3.2**). This is called LTP. Initiation of this process depends on multiple second messenger mechanisms (**Fig. 2.5.3.2**). One of the best studied pathways starts with calcium influx into the presynapse through a glutamate receptor, NMDA, which activates further molecular cascades involving cAMP and protein kinases such as CamKII (other receptors, such as the glutamate receptors AMPA and mGluR, also play a role). LTP is then maintained by changes in gene transcription factors, such as CREB, and changed patterns of protein synthesis and phosphorylation, probably also dependent on protein kinase cascades. The time course of these processes can be used for a distinction between early phase LTP (the cellular signature of learning that occurs over seconds to minutes) and late phase LTP, which involves protein synthesis and occurs over minutes to hours and is thought to be critical for consolidation of new memoranda and would be linked to memory consolidation.[9,10]

LTD is a closely related process that is triggered when presynaptic stimulation is lower, causing less calcium influx (again through NMDA receptors) and preferential activation of calcineurin, a protein phosphatase. Together, LTP and LTD allow bidirectional enduring modulation of synaptic strength that could underlie formation and reversal of experience-dependent coupling in neural assemblies. Linking these synaptic changes to validated mechanisms underlying the complexities of human memory remains a challenge. However, the active field of neural network modeling has shown that, in principle, such changes in synaptic efficacy can produce efficient mechanisms to encode, store and retrieve information, a proposal first made by the neurophysiologist Donald Hebb.

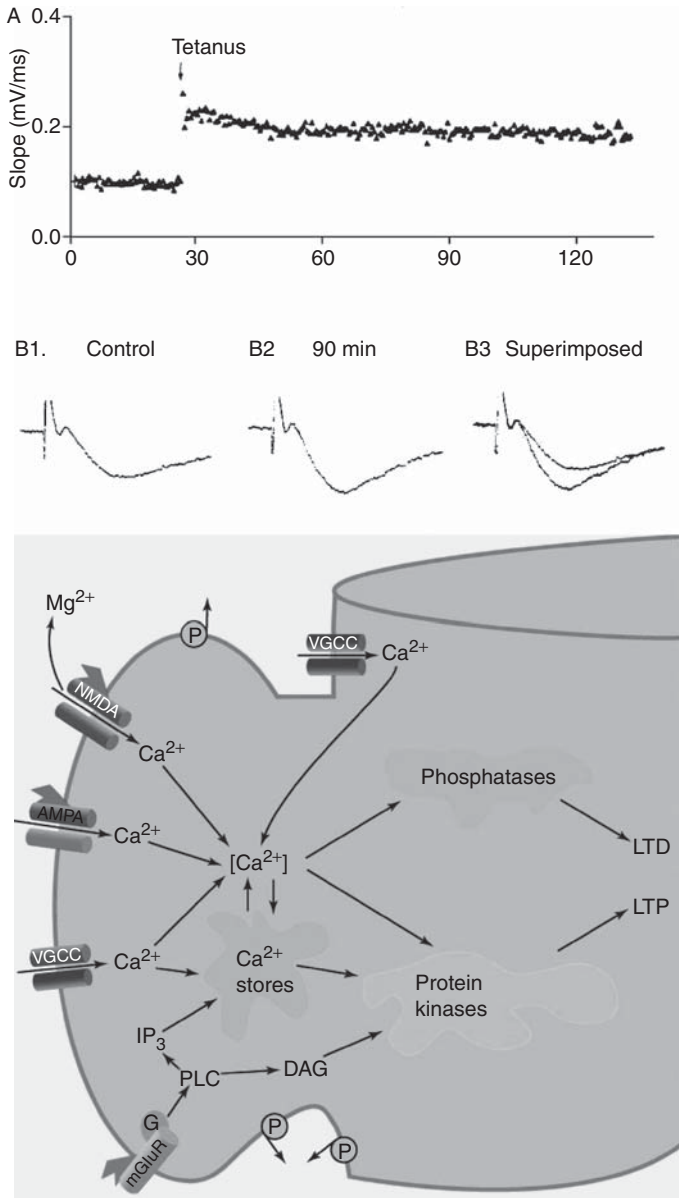


Fig. 2.5.3.2 Top: Long-term potentiation after a tetanic stimulation in the HF. Bottom: molecular mediators of long-term potentiation.

Declarative/episodic memory

(a) Neural systems

Current evidence indicates that the HF and linked regions of the medial temporal lobe (MTL), in interactions with parts of the prefrontal cortex, play a critical role in the encoding and retrieval of episodic memories, whereas engrams are stored in neocortex[11] (**Fig. 2.5.3.3**). Interactions of the HF with amygdala are important for emotional memories. The HF consists of the hippocampus proper, the entorhinal cortex, which provides the main port of entry for connections with the cortex, and the adjacent perirhinal and parahippocampal cortices, which interact with the entorhinal cortex and in turn receive projections from all other neocortical areas, with parietal, dorsal occipital and prefrontal regions primarily projecting to parahippocampal gyrus and temporal cortex to

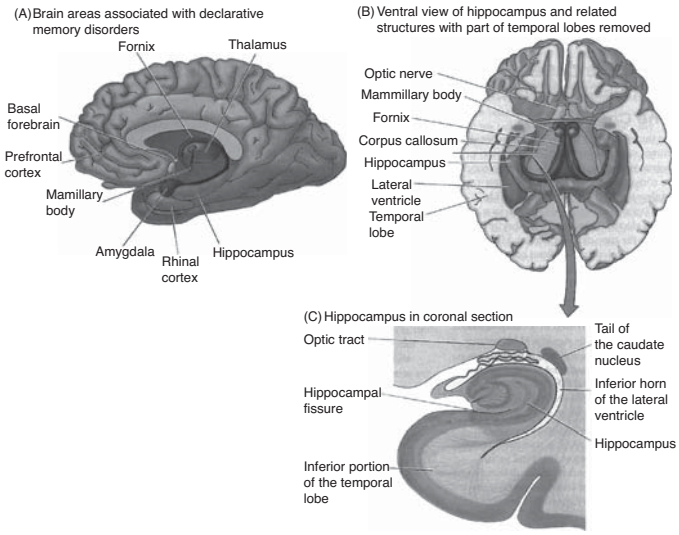


Fig. 2.5.3.3 (A) brain regions associated with dysfunction in episodic memory. (B) the hippocampal formation, view from below. (C) coronal section through the hippocampal formation.

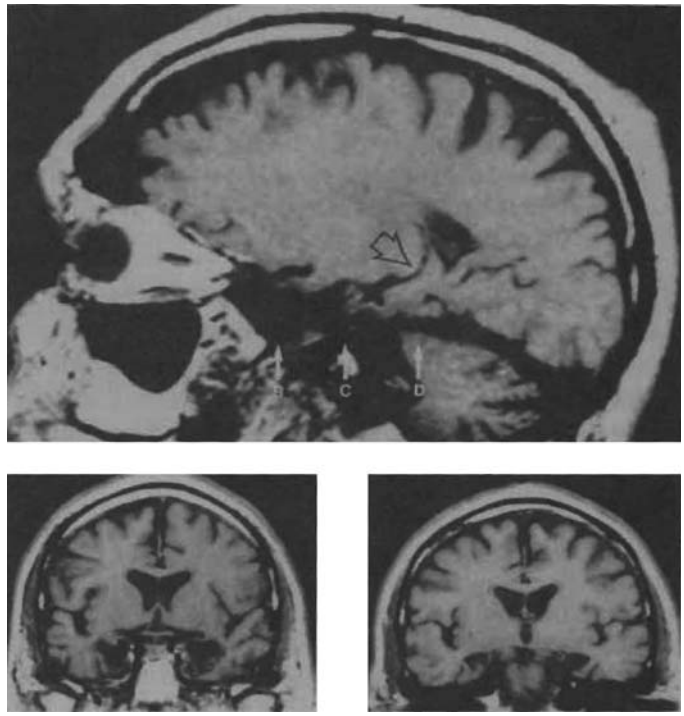


Fig. 2.5.3.4 Hippocampal damage in patient H.M., shown on a sagittal (top) and two coronal slices (below) of the patient's MRI.

perirhinal cortex. In this way, the HF is bidirectionally connected with the rest of the brain. Although pertinent observations were already made at the end of the 19th century by Bechterew, the importance of the hippocampal formation for episodic memory was dramatically shown in 1953 by the case of patient H.M., in whom the HF and amygdala were resected bilaterally as a treatment for drug-resistant epilepsy[12] (**Fig. 2.5.3.4**). This led to complete and enduring **anterograde** amnesia (inability to form new episodic memories). In addition, he has some degree of **retrograde** amnesia (i.e. an inability to retrieve episodic information stored before the operation), while his working and procedural memory, as well as priming, is unimpaired. A similar pattern of memory impairment is observed in Wernicke encephalopathy/Korsakov's syndrome or other neurological processes impacting on diencephalic structures, including the medial thalamus, mamillary body and the fornix, which project to the HF. This indicates that these structures may be viewed as a system that is critical for assigning a spatiotemporal (episodic) context. In rodents, a major role of the HF is indeed to function as a neural map of the environment;[13] it is controversial to what degree this applies to humans. Neuroimaging studies confirm that activation of the HF is observed during successful encoding and retrieval of episodic information.[14]

A neurocognitive model system for examining the computational role of MTL subsystems involved in episodic memory may help the reader gain a sense of the interplay among various subprocessing modules.[15] In this model, the MTL binds memories and their instance-specific context and then stores their code for later retrieval. The model described here is based on studies showing that the bulk of hippocampal (HC) cortical input is segregated over two pathways. One of these may convey spatial information; the other may convey information regarding items and objects. The two streams are interconnected at various levels within entorhinal cortex, which likely contributes to the integration of cortical inputs into a representation of their co-occurrence. The hippocampus proper may quickly associate a code to the conjunction of cortical inputs, such that similar entorhinal patterns come to be separated via their associated hippocampal patterns. These hippocampal patterns are thus not directly associated with individual features, but serve to separate the large number of overlapping entorhinal patterns, which is important to ensure that retrieval will be unambiguous. The representational overlap in entorhinal cortex, combined with the pattern separation system in the hippocampus proper, enormously increases the storage capacity of the memory store and allows accurate recall of episodic memories. In this model, retrieval in this memory system can be sampled using cues, consisting of *partial* input patterns; for instance, part of a context representation from a previously experienced episode. Initially, such a cue may activate only part of an associated entorhinal pattern, but if the set of activated entorhinal nodes sufficiently resembles a stored representation, their combined firing will tend to activate associated hippocampal nodes, through previously strengthened connections with these nodes. The hippocampal nodes, in turn, will recruit missing nodes of the entorhinal representation. This pattern completion process will reinstate the original pattern in the entorhinal layer, which, in turn, can reinstate associated information in the input layers, namely, item representations that have been experienced in that particular context (feature extraction). Thus, all features of an episode can be recalled, even when only part of the input layers is cued.

The model was shown to be sensitive to various types of simulated lesions including reductions of nodes ('neurones'), addition of noise to the system, and perhaps most relevant for the modeling of schizophrenia, marked reductions in connectivity between the various modules. At the network level, this reduced connectivity led to compromised cross-association of episodic features (i.e. item and context) and a superimposed, mild reduction of pattern separation in the system. The latter malfunction made some

patterns irretrievable, affecting all memory tasks including recognition, albeit to a mild degree. The cross-association problem also attenuated 'searching' of the memory store, particularly with single-source cues. This preferentially affected tasks with a large retrieval demand, such as free recall. From an information processing standpoint, what appeared to be disproportionate failures in retrieval were due to compromised encoding.

Recent work has indicated that recollection is predominantly mediated by the hippocampus proper, whereas familiarity has been linked to perirhinal cortex.[16] While this distinction is supported by functional neuroimaging studies of healthy individuals (e.g. Eldridge *et al.*[17]), studies of patients with circumscribed lesions of the hippocampus have nevertheless found severe familiarity-based recognition memory impairments.[18] These results do not necessarily contradict a dual process distinction, but could rather suggest that MTL regions comprise an integrated network that supports both processes.

There has been much interest in the precise role of episodic memory systems in the formation of associations between items that are entering memory. Several new studies suggest that the hippocampus is engaged preferentially when inter-item associations are formed in memory. While there have been several compelling accounts that specific subsystems in the MTL complex play different roles in single item encoding and associative encoding, the distinction between the two may be relative, not absolute.[11,19]

Neuroimaging and lesion studies have suggested that interactions between the HF and the amygdala are relevant for emotional memories, especially if these have a fearful or aversive character.[20] Neuroimaging has also demonstrated that encoding and especially retrieval is associated with activation of the lateral prefrontal cortex, as well as with increased functional interactions of these regions with the HF that are supported by anatomical tracts such as the uncinate fascicle. In some studies, left prefrontal cortex is differentially more involved than right in encoding information into episodic memory, whereas right prefrontal cortex is differentially more involved than left in episodic memory retrieval.[21] Compared to encoding and retrieval, the evidence is much less clear with regard to the storage of the engrams themselves. Cases such as H.M. and the clinical picture of Korsakov's syndrome show that the HF and diencephalic structures cannot be the store, since most episodic and semantic memories laid down before the onset of illness are spared. Current evidence suggests that the neocortex is the ultimate store of memories and that engrams reside in regions that are also specialized in processing stimuli to which they pertain. For example, circumscribed cortical lesions can result in category-specific impairments in retrieving object information, the so-called anomias, which have been described for classes such as people, tools, or living things, and neuroimaging studies show that similar regions are differentially activated during naming of these object classes. In each case, engrams are assumed to be stored in a distributed pattern of synaptic connections over a large group of neurones. It is an open question how the interaction of the HF and cortex accomplishes the encoding, and retrieval of information from these neural assemblies (but see above for a model). On the molecular level, glutamatergic neurotransmission is crucial to support the LTP mechanisms that support the neural plasticity essential for memory formation. In addition, acetylcholine is a neurotransmitter associated with declarative memory function since it is known that muscarinic receptor blockade impairs episodic memory and degeneration of cholinergic neurones in the basal nucleus (of Meynert) is a prominent finding in Alzheimer's disease. It is likely that cholinergic mechanisms act on declarative memory by modulating glutamate-dependent LTP and LTD in MTL regions.

(b) Assessment and neuropsychology

Clinical assessment of memory systems is an important facet of neuropsychiatric and neuropsychologic test batteries. Assessments may vary widely in their depth and breadth and systematic approach. There are numerous tests of episodic memory that can be used for clinical purposes. Well known batteries include the Wechsler Memory Scale-III, which involves verbal memory for stories, verbal paired associates, and word lists, and visual memory for scenes, faces, and designs. Immediate and delayed recall of these tests is assessed. There are also many standardized verbal list learning tests that involve differing degrees of semantic relatedness among words (and perhaps requiring different degrees of strategic encoding) and minor differences in administration. These include the California Verbal Learning Test, the Hopkins Verbal Learning Test, and the Selective Reminding Test. These tests have alternate but equivalent forms that reduce practice effects when tests are administered repeatedly. A comprehensive evaluation of memory should include multiple trials of a word list to assess learning rate or slope; measures of immediate and delayed memory (e.g. recall after 30 minutes); a recognition test in which a subject must decide whether an item had been studied or not (was old or new in order to minimize effortful retrieval); and tests of visual memory and verbal memory that preferentially engage the right or left MTL systems, which are thought to be material specific, at least to some degree.

Cognitively, there are several distinctions or processes in episodic memory that bear special comment. Most are still being actively investigated. One of the earliest and best replicated findings in processing oriented theories of memory suggests that the level at which an item is encoded is an important predictor of later recall.[22] Thus, words that are encoded deeply (for example, after a semantic judgment about animacy) are remembered better than words encoded superficially (for example, after a judgment about whether the word contains a specific orthographic feature). The exact cognitive mechanism by which this occurs is unclear, but could involve the creation of more cue-item associations so that a wider variety of different searches might have yield.

Another important area involves the distinction between a sense of familiarity with an item at recall versus the recollection of an item, which implies knowledge about the spatiotemporal context in which the item was encoded. It is thought that memories may be retrieved through either of these two processes: recollection of a memory that involves adjunctive contextual information; or the feeling of knowing that an item, face, thing, etc. has been encountered before without memory of the surrounding context.[23,24]

Rate of forgetting reflects the degree to which memories that were once successfully retrieved can no longer be retrieved after a delay. For healthy individuals savings (the inverse of rate of forgetting) may be at 80–90 per cent for several hours or more after initial recall. Several amnesic conditions, as well as a form of frontal temporal dementia, have been associated with increased rates of forgetting. In Alzheimer's disease and Korsakov's syndrome, savings may be less than 50 per cent after delays of several minutes. Nevertheless, the situation is undoubtedly more complex psychometrically and cognitively than presented here.[25,26]

Implicit memory

(a) Neural systems

(i) Procedural memory

Clinical experience shows that the acquisition of new visuomotor skills is often unimpaired in patients with deep amnesia due to medial temporal lobe lesions. Conversely, skill learning, but not declarative memory, is often impaired in patients with degenerative or vascular lesions of the basal ganglia (for example in Huntington's) or cerebellum. Neuroimaging has confirmed the importance of basal ganglia and cerebellum for procedural memory[1] and has also demonstrated time-variant activation of primary and secondary motor cortex during skill learning. The basal ganglia receive excitatory glutamatergic projections from the cortex and thalamus, integrate them with monoaminergic inputs and sends them via the globus pallidus and substantia nigra pars reticulata to the thalamus, which projects back to prefrontal cortex.[27] These parallel processing loops are critical for the integration of sensorimotor, cognitive and emotional information.[27] It is usual to distinguish between a dorsal and a ventral subdivision of the striatum, and it is the dorsal striatum (caudate nucleus, putamen and globus pallidus) that is strongly interconnected with cortical areas relevant for motor planning and execution. The learning of repetitive sequence, as well as so-called 'open loop' tasks, in which visual feedback is delayed, especially depends on the integrity of the dorsal striatum and its interactions with cortex.[28] This extends to skills that are not motor, for example the prediction of probabilistic sequences or the planning of complex tasks. Conversely, the importance of the cerebellum lies with 'closed loop' tasks that require continuous visuomotor feedback, as well as fast error control. The cerebellum has also been proposed to play a role in creating new stimulus-response mappings.

(ii) Priming

Neuroimaging suggests that the neural substrate of priming lies in neocortex. Specifically, a reduction of activation to a primed stimulus is consistently found, either in modality-specific regions (such as visual areas for visual repetition priming) or in 'amodal' cortical areas such as lateral temporal cortex for semantic priming. It has been hypothesized that this reduced activation represents neural assemblies that are optimized, by 'pruning' of unnecessary connections, for easier activation, and that this may underlie the facilitated response to a primed stimulus.[29] Clinically, this leads to the prediction that priming should be altered in disorders that impair the integrity of the cortical regions involved, such as semantic priming in Alzheimer's disease.

(iii) Fear conditioning and extinction

Conditioned fear is of high relevance in psychiatry in disorders ranging from simple phobias and generalized anxiety disorder to major depression. In conditioning, a fear response to an unconditional stimulus (for example, an electric shock) is transferred to a conditional stimulus with which it is paired (CS, for example, a tone regularly preceding this shock). A large body of research has established a key role for the amygdala in this memory process.[30] Different subnuclei of this complicated structure are implicated in establishing and storing fear conditioning memory traces. An area of recent research interest concerns extinction, the process in which a conditioned fear response is gradually lost if the conditioned stimulus is repeatedly presented without adverse consequences. It is now clear that extinction is not a passive process, but depends on interactions between amygdala and the cingulate cortex.[31] This circuit has been implicated in depression and anxiety in humans, in which dysphoric mood and affect are abnormally maintained, i.e. not extinguished.[32]

(b) Assessment and neuropsychology

Despite at least 20 years of cognitive science research in implicit memory, procedural memory, or habit formation, there are no commercially available versions of these tests. In part, this may have to do with lack of psychometric evaluation of test-retest reliability, ceiling or floor effects, etc. or unclear relations to functional status or outcome. It might be possible to adapt some instrumentation used experimentally (e.g. rotor pursuit) if there are adequate local normative data for the test. There are several forms of implicit learning, all of which are thought to involve learning or memory without conscious awareness of recall. Various motor skills can be learned incrementally, such as rotor pursuit or mirror tracing. Others involve motor sequences. Probabilistic learning can occur when there is acquisition of information or representations that reflect underlying structural regularities in the input, i.e. when there are statistical regularities between stimuli and responses.[33] This can occur in tasks as seemingly disparate as the so-called weather prediction task (characterized by probabilistic relationships between specific stimulus configurations and a response, in this case 'sunshine' or 'rain') and artificial grammar. Critically, implicit learning can occur even when the episodic system of recollection is dysfunctional.

Other types of implicit memory may be item specific, including some types of priming. Priming is thought to be an instance of memory without awareness. Such priming can be reflected in improvements in accuracy or reaction time during testing. It can be demonstrated in a variety of tasks, some rather rarefied, like word stem completion, and some rather simple and robust, like so-called repetition priming. In the latter paradigm, an item is repeated and access to it (usually measured in reaction time) is speeded at the second presentation, while concomitantly, physiological measures, ranging from single cell activity to BOLD activation, demonstrate reductions in neocortical areas (e.g. for words, inferior prefrontal cortex). Some of these effects may be quite long, lived.

The degree to which episodic memory may also support some types of priming is sometimes unclear and may depend on the paradigm and experimental manipulations. Perhaps the best evidence that priming reflects a dissociable memory system comes from amnesic patients in whom there is little chance that the episodic system is supporting priming (e.g. studies of priming in the amnesic patient HM [34]). Additionally, an important and critical review[35] has proposed that some priming may not reflect changes in the abstract representation of an item, but rather response learning.

Finally, conditioning (especially classical or Pavlovian) has also been considered a type of implicit learning. The neurobiological literature on this phenomenon is quite extensive.[36] The phenomenon itself might be relevant for understanding a wide range of behaviors, including anxiety (as a result of fear conditioning) and preferences.

(c) Working memory

(i) Neural systems

A large body of work has established the importance of dorsolateral prefrontal cortex (DLPFC) for working memory. Both the

simple maintenance of information over a delay and the manipulation of that information require DLPFC function. In the influential model of Baddeley, a 'central executive' component of working memory works together with modality-specific storage systems, the 'visuospatial scratchpad' for visual information and the 'phonological loop' for auditory information.[37] While the details of how this cognitive account reflects neural organization are being debated, it is clear from a multitude of studies that DLPFC activation is usually observed in conjunction with activity of posterior cortical areas that receive input from a variety of specialized sensory cortices.[38] Chief among those is the inferior parietal lobule, a brain region strongly and bidirectionally connected with DLPFC that is likely to be important for item storage during working memory. Within DLPFC proper, some models propose a regional differentiation based on the modality of information stored (with more dorsal activation associated with visual, ventral with semantic items), while others propose a specialization based on cognitive operations (manipulation of memory items being performed more dorsally than pure maintenance). Clinically, large lesions of DLPFC are invariably associated with working memory impairment, however, problems of similar magnitude are also observed in schizophrenia, where only subtle structural abnormalities are found in this (as in any other) brain region.[39] The association of working memory impairment with schizophrenia has driven an extensive programme of research aimed at understanding mechanisms underlying memory impairments in this disorder. One well-validated finding from this work is the importance of the neurotransmitter, dopamine, for DLPFC function, which has been found to exhibit an 'inverted u' shaped relationship with working-memory related activation of DLPFC neurones and dopaminergic, especially D1-receptor, stimulation.[40] It is believed that dopaminergic tone is essential for optimizing signal to noise ratio, or tuning, in DLPFC, an essential network property for working memory maintenance. This is further supported by a modulation of working memory by a functional variant in the COMT gene which alters the protein's thermolability and hence its ability to degrade dopamine in cortex, the impact of dextroamphetamine on N Back driven BOLD activation in fMRI, and the COMT inhibitor tolcapone's impact on N Back RT.[41–43]

Interactions between DLPFC and hippocampus may also be disturbed in schizophrenia and reflect an inability to disengage episodic memory processes during working memory.[44]

(ii) Assessment and neuropsychology

The classic test of simple working memory is digit span, which involves repetition of short sequences of digits. Span is assessed by increasing the length of the sequence that can be recalled. Nonverbal working memory often involves short sequences of locations (as in Visual Span in the WMS-R and the so-called Corsi blocks.) Tests thought to depend on simultaneous storage and manipulation of information are generally considered executive in nature. The Letter-Number Span is a good example of this class of tests. In it, the subject is asked to order a short random sequence of letters or numbers numerically and alphabetically. Interestingly, the test is highly correlated with a nonverbal, formally dissimilar, problem solving task, the Wisconsin Card Sort, providing evidence that both these tests engage executive processes. The Card Sort itself may be the best known executive test administered in clinical test batteries. It calls upon such executive abilities as abstraction, set shifting, and response to examiner feedback.

An exceptionally well validated computerized battery of 'frontal lobe tests' called the CANTAB is based on the comparative and pharmacologic challenge literature, as well as human lesion studies. All tests are nonverbal and may involve problem solving (as in Tower of London), various levels of set shifting (in the ID/ED task), and self-ordered pointing that demands that the subject remember his/her own actions.

A widely used test of cognitive control is the Stroop, for which there are several commercially available versions. In the critical interference condition, the subject must respond to words printed in incongruent colors (e.g. red) by naming the colour of the ink and simultaneously suppressing the prepotent response to read the word.

There are many experimental tests of working memory and executive that could in theory be adapted for clinical purposes provided these have adequate local normative data and adequate validity and psychometric characteristics. In general, many of the paradigms that have interested cognitive neuroscientists have not yet become part of routine clinical assessment.

The working memory (WM) system is thought to be a limited capacity system that holds information on line when the stimulus is no longer present (perhaps up to 40 seconds, as demonstrated in densely amnesic patients). The idea that WM is a capacity limited system comes from the work of G. Miller.[2] Several compelling accounts now suggest that it may under some circumstance be smaller than Miller's canonical '7±2.' For instance, in studies of visual stimuli, Luck *et al.*[45] have suggested that a visual store may hold only four items. Subsystems include a slave system for short term memory of phonological information. This system includes an articulatory rehearsal mechanism and a phonological store. A visual spatial scratchpad processes visual, non-linguistic information. An episodic memory buffer is thought to play a role in the interface between working and episodic memory. A central executive is involved in the allocation of cognitive resources during dual tasks and in the manipulation or transformation of information. The central executive may be involved in cognitive control, such that when there is response conflict (e.g. during response selection in the Stroop task or Eriksen flanker task) more resources can be made available for biasing decisions.[46] More specifically, in this account, cognitive conflict is detected (by the anterior cingulate) and signals are sent to executive areas (in DLPFC) that increase processing resources. The increase in resources is used to bias a response to one or another aspect of the stimulus' features. Importantly, various computational models of this process have shown that a homunculus (i.e. a director or decider) is not necessary for a correct response to be made. There is ongoing interest in how best to characterize operations in the WM system. One view that posits a single algorithm that maintains a stable representation over a delay (although subsystems may be dedicated to spatial, object or verbal information) can be used to characterize the basic function of this system. Goldman Rakic and colleagues provided much neurobiological evidence in favor of this argument. Another view holds that multiple computational algorithms perform a variety of tasks.[47,48] Such functions as attentional set shifting, planning, and monitoring of sequences of responses (self or externally generated) can be brought to bear on a task, depending on its demands and are separable from basic mnemonic maintenance functions.

Updating of information in WM (i.e. registering information and dumping information from a buffer) and suppression of inter-

ference may also engage the executive system and may be critical for refining goal-directed behavior.[49] In one version of a task that makes continual demands for updating and resistance to interference, a restricted set of numbers are displayed successively (and approximately every two seconds) and the subject views one while responding by pressing a button corresponding to a stimulus 'one back'. Thus, the subject must continuously update his/her working memory buffer in that a target-to-be must be shifted in status to a target prior to being 'dumped' from a computational buffer and interference from other similar stimuli must be suppressed. Recent data (see below) suggest that dopaminergic tone at the cortical level may be particularly important in this specific aspect of cognitive control.

Finally, there are several integrative accounts of goal directed behaviour that are dependent on various working and executive functions. In perhaps the best known of these, a 'script' that engages multiple short-term and long-term memories bridges the temporal gap for projects that may be quite long temporally.[50]

Future directions

As our understanding of neural systems involved in memory dysfunction in a variety of psychiatric disorders matures, it will be important to leverage that knowledge to identify new molecular treatment targets. One important recent development in this direction is the identification of genes impacting on memory function. Genetic variation in the neurotrophic factor BDNF, for example, has been shown to impact on hippocampal function and episodic memory,[51] as have several risk genes for schizophrenia.[52] COMT, a gene that impacts on cortical dopamine concentrations, affects working memory performance and prefrontal activation.[41,53] Risk genes for depression and anxiety may affect amygdala-cingulate and amygdala-hippocampal interactions that are important for fear extinction and emotional memory, respectively, providing neural mechanisms for gene by environment interactions mediating the effects of early adverse experience on risk for psychiatric disorders.[32,54] Translational strategies for drug development can arise from these findings. The COMT inhibitor tolcapone has been shown to improve working memory in normals.[43] Data suggesting an impact of stress hormones such as corticosteroids on hippocampal integrity in posttraumatic stress disorder suggest neuroprotective strategies to prevent development of disease before anticipated exposure, for example in battle. Finally, it has become clear that memory impairment is an important predictor for treatment response in psychiatry, schizophrenia being an important example, and neurorehabilitative programs to improve memory are being explored. The study of the neurobiology of memory is therefore a promising avenue towards improved and specific therapies in psychiatry.

Further information

Miyake, A. and Shah, P. (eds.) (1999) *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control.* Cambridge University Press, New York.

Squire, L.R. and Schachter, D.L. (eds.) (2002), *Neuropsychology of Memory*. The Guildford Press, New York.

Hippocampus (the journal).

Wikipedia (wikipedia.org) entry 'memory'.

References

1. Gabrieli, J.D. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology*, **49**, 87–115.
2. Miller, G.A. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review*, **63**(2), 81–97.
3. O'Reilly, R.C. and Norman, K.A. (2002). Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends in cognitive sciences*, **6**(12), 505–10.
4. Gluck, M.A., Meeter, M. and Myers, C.E. (2003). Computational models of the hippocampal region: linking incremental learning and episodic memory. *Trends in cognitive sciences*, **7**(6), 269–76.
5. Moscovitch, M., *et al.* (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: a unifi ed account based on multiple trace theory. *Journal of Anatomy*, **207**(1), 35–66.
6. Ashby, F.G. and Maddox, W.T. (2005). Human category learning. *Annual Review of Psychology*, **56,** 149–78.
7. Eichenbaum, H. (2006). Remembering: functional organization of the declarative memory system. *Current Biology*, **16**(16): R643–5.
8. Miller, S. and Mayford, M. (1999). Cellular and molecular mechanisms of memory: the LTP connection. *Current Opinion in Genetics & Development*, **9**(3), 333–37.
9. Hawkins, R.D., Kandel, E.R. and Bailey, C.H. (2006). Molecular mechanisms of memory storage in Aplysia. *The Biological Bulletin*, **210**(3), 174–91.
10. Izquierdo, I., *et al.* (2006). Different molecular cascades in different sites of the brain control memory consolidation. *Trends in Neurosciences*, **29**(9), 496–505.
11. Squire, L.R., Stark, C.E. and Clark, R.E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, **27**, 279–306.
12. Scoville, W.B. and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, **20**(1), 11–21.
13. O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, **34**(1), 171–75.
14. Squire, L.R., *et al.* (1992). Activation of the hippocampus in normal humans: a functional anatomical study of memory. *Proceedings of the National Academy of Sciences of the United States of America*, **89**(5), 1837–41.
15. Talamini, L.M., *et al.* (2005). Reduced parahippocampal connectivity produces schizophrenia-like memory deficits in simulated neural circuits with reduced parahippocampal connectivity. *Archives of General Psychiatry*, **62**(5), 485–93.
16. Suzuki, W.A. and Eichenbaum, H. (2000). The neurophysiology of memory. *Annals of the New York Academy of Sciences*, **911**, 175–91.
17. Eldridge, L.L., *et al.* (2000). Remembering episodes: a selective role for the hippocampus during retrieval. *Nature Neuroscience*, **3**(11), 1149–52.
18. Wais, P.E., *et al.* (2006). The hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron*, **49**(3), 459–66.
19. Davachi, L. and Wagner, A.D. (2002). Hippocampal contributions to episodic encoding: insights from relational and item-based learning. *Journal of Neurophysiology*, **88**(2), 982–90.
20. Cahill, L., *et al.* (1996). Amygdala activity at encoding correlated with long-term, free recall of emotional information. *Proceedings of the National Academy of Sciences of the United States of America*, **93**(15), 8016–21.
21. Habib, R., Nyberg, L. and Tulving, E. (2003). Hemispheric asymmetries of memory: the HERA model revisited. *Trends in Cognitive Sciences*, **7**(6), 241–5.
22. Craik, F.I.M. and Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, **104**, 268–94.

23. Cipolotti, L. and Bird, C.M. (2006). Amnesia and the hippocampus. *Current Opinion in Neurology*, **19**(6), 593–98.
24. Yonelinas, A.P. (2001). Components of episodic memory: the contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, **356**(1413), 1363–74.
25. Greene, J.D., Baddeley, A.D. and Hodges, J.R. (1996). Analysis of the episodic memory deficit in early Alzheimer's disease: evidence from the doors and people test. *Neuropsychologia*, **34**(6), 537–51.
26. Murre, J.M., Graham, K.S. and Hodges, J.R. (2001). Semantic dementia: relevance to connectionist models of long-term memory. *Brain*, **124**(Pt 4), 647–75.
27. Alexander, G.E., DeLong, M.R. and Strick, P.L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, **9**, 357–81.
28. Packard, M.G. and Knowlton, B.J. (2002). Learning and memory functions of the Basal Ganglia. *Annual Review of Neuroscience*, **25**, 563–93.
29. Wiggs, C.L. and Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, **8**(2), 227–33.
30. Phelps, E.A. and LeDoux, J.E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron*, **48**(2), 175–87.
31. Quirk, G.J., *et al.* (2003). Stimulation of medial prefrontal cortex decreases the responsiveness of central amygdala output neurons. *The Journal of Neuroscience*, **23**(25), 8800–807.
32. Pezawas, L., *et al.* (2005). 5-HTTLPR polymorphism impacts human cingulateamygdala interactions: a genetic susceptibility mechanism for depression. *Nature Neuroscience*, **8**(6), 828–34.
33. Forkstam, C. and Petersson, K.M. (2005). Towards an explicit account of implicit learning. *Current Opinion in Neurology,* **18**(4), 435–41.
34. Corkin, S. (2002). What's new with the amnesic patient H.M.? *Nature Reviews. Neuroscience*, **3**(2), 153–60.
35. Schacter, D.L., Dobbins, I.G. and Schnyer, D.M. (2004). Specificity of priming: a cognitive neuroscience perspective. *Nature Reviews. Neuroscience*, **5**(11), 853–62.
36. Thompson, R.F. (1990). Neural mechanisms of classical conditioning in mammals. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, **329**(1253), 161–70.
37. Baddeley, A. (1992). Working memory. *Science*, **255**(5044), 556–9.
38. Smith, E.E. and Jonides, J. (1997). Working memory: a view from neuroimaging. *Cognitive Psychology*, **33**(1), 5–42.
39. Weinberger, D.R., *et al.* (1992). Evidence of dysfunction of a prefrontal-limbic network in schizophrenia: a magnetic resonance imaging and regional cerebral blood flow study of discordant monozygotic twins. *The American Journal of Psychiatry*, **149**(7), 890–97.
40. Williams, G.V. and Goldman-Rakic, P.S. (1995). Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. *Nature*, **376**(6541), 572–75.
41. Goldberg, T.E., *et al.* (2003). Executive subprocesses in working memory: relationship to catechol-O-methyltransferase Val158Met genotype and schizophrenia. *Archives of General Psychiatry*, **60**(9), 889–96.
42. Mattay, V.S., *et al.* (2003). Catechol O-methyltransferase val158-met genotype and individual variation in the brain response to amphetamine. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(10), 6186–91.
43. Apud, J.A., *et al.* (2007). Tolcapone improves cognition and cortical information processing in normal human subjects. *Neuropsychopharmacology*, **32**(5), 1011–20.
44. Meyer-Lindenberg, A.S., *et al.* (2005). Regionally specific disturbance of dorsolateral prefrontal-hippocampal functional connectivity in schizophrenia. *Archives of General Psychiatry*, **62**(4), 379–86.
45. Luck, S.J. and Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, **390**(6657), 279–81.
46. Botvinick, M.M., Cohen, J.D. and Carter, C.S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*, **8**(12), 539–46.
47. Petrides, M. (2005). Lateral prefrontal cortex: architectonic and functional organization. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, **360**(1456), 781–95.
48. Robbins, T.W. (1996). Dissociating executive functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, **351**(1346), 1463–70; discussion 1470–1.
49. Smith, E.E. and Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, **283**(5408), 1657–61.
50. Wood, J.N., *et al.* (2005). Neural correlates of script event knowledge: a neuropsychological study following prefrontal injury. *Cortex*, **41**(6), 796–804.
51. Egan, M.F., *et al.* (2003). The BDNF val66met polymorphism affects activitydependent secretion of BDNF and human memory and hippocampal function. *Cell*, **112**(2), 257–69.
52. Callicott, J.H., *et al.* (2005). Variation in DISC1 affects hippocampal structure and function and increases risk for schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(24), 8627–32.
53. Egan, M.F., *et al.* (2001). Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, **98**(12), 6917–22.
54. Meyer-Lindenberg, A., et al. (2006). From the Cover: Neural mechanisms of genetic risk for impulsivity and violence in humans. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(16), 6269–74.

2.5.4 The anatomy of human emotion

R. J. Dolan

Introduction

Emotions, uniquely among mental states, are characterized by psychological and somatic referents. The former embody the subjectivity of all psychological states. The latter are evident in objectively measurable stereotyped behavioural patterns of facial expression, comportment, and states of autonomic arousal. These include unique patterns of response associated with discrete emotional states, as for example seen in the primary emotions of fear, anger, or disgust often thought of as emotion proper. Emotional states are also unique among psychological states in exerting global effects on virtually all aspects of cognition including attention, perception, and memory. Emotion also exerts biasing influences on high level cognition including the decision-making processes that guide extended behaviour. An informed neurobiological account of emotion needs to incorporate how these wide ranging effects are mediated.

Although much of what we can infer about emotional processing in the human brain is derived from clinic-pathological correlations, the advent of high resolution, non-invasive functional neuroimaging techniques such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) has

greatly expanded this knowledge base. This is particularly the case for emotion, as opposed to other areas of cognition, where normative studies have provided a much richer account of the underlying neurobiology than that available on the basis of observations from pathology as in classical neuropsychology.

Emotion has historically been considered to reflect the product of activity within the limbic system of the brain. The general utility of the concept of a limbic-based emotional system is limited by a lack of a consensus as to its precise anatomical extent and boundaries, coupled with knowledge that emotion-related brain activity is, to a considerable degree, configured by behavioural context. What this means is that brain regions engaged by, for example, an emotion of fear associated with seeing a snake can have both distinct and common features with an emotion of fear associated with a fearful recollection. Consequently, within this framework emotional states are not unique to any single brain region but are expressed in widespread patterns of brain activity, including activity within early sensory cortices, shaped by the emotion eliciting context. This perspective emphasizes a global propagation of emotional signals as opposed to a perspective of circumscribed limbic-mediated emotion-related activity.

The amygdala and emotion

The above considerations aside, the structure most closely affiliated with emotional processing is the amygdala. This structure is an anatomically and functionally heterogeneous, bilateral, collection of nuclei located in anterior medial temporal cortex. The importance of the amygdala in emotional control was first highlighted by reports that rhesus monkeys with bilateral temporal lobe ablations no longer show appropriate fear or anger responses.[1] The role of the amygdala in emotion has been subsequently extended by findings that humans with lesions to this structure have impaired emotional recognition, particularly for fear, and no longer acquire Pavlovian conditioned responses[2] (see below). Finally, functional neuroimaging findings show activation of amygdala in responses to face stimuli that depict a range of emotions, particularly fear but also other primary emotions.[3, 4]

The importance of the amygdala in emotion derives in part from its extensive anatomical connections with all sensory processing cortices, as well as hippocampus, basal ganglia, cingulate cortex and the homeostatic regulatory regions of hypothalamus and brain stem.[5] This widespread anatomical connectivity means that this structure can access information processing in multiple brain regions and, in turn, can exert diffuse modulatory influences, including influences on effector autonomic and motor output systems. In this way activation of the amygdala by a sensory based emotional stimulus influences widespread brain regions including those that mediate homeostatic regulatory responses as expressed in altered autonomic state, such as change in heart rate, blood pressure and respiration.

Learning predictive emotional responses

A central role for emotion is to index value, specifically whether present or future sensory events or states of the environment that are likely to be associated with reward or punishment. From this perspective, all emotions are to a greater or lesser degree valanced. For example, an emotion of joy signals a likelihood of reward while an emotion of fear signals a likelihood of punishment. The fact that signals that predict such emotional occurrences are to some degree arbitrary means that the brain must have some means of associating sensory cues with potential emotional outcomes, an ability that seems crucial for adaptive behaviour.

Associative learning provides a phylogenetically highly conserved means to predict future events of value, such as the likelihood of food or danger, on the basis of predictive sensory cues. The amygdala plays a crucial role in mediating this form of emotional learning as evidenced by deficits seen with animal lesion data and learning-related effects seen in human functional neuroimaging experiments.[6, 7] In its simplest form, Pavlovian conditioning is expressed when a previously neutral sensory stimulus (the conditioned stimulus, or CS+) acquires emotional predictive significance through pairing with a biologically salient reinforcer (the unconditioned stimulus, or UCS). With conditioning, the predictive stimulus (CS+) comes to elicit behaviour previously associated with the UCS, but in the absence of UCS presentation. There is a wealth of animal and human data which now shows that the amygdala has a key role in this form of associative learning, for both appetitive and aversive outcomes.

How the brain updates predictions of emotional outcomes

While contingencies acquired on the basis of associative learning provide a basis for generation of predictions of future event of value in response to sensory cues, this form of learning lacks flexibility in optimizing future behaviour. For example, the value of future states associated with predictive cues may change in the absence of subsequent pairing with these cues. Thus, a cue that is associated with a particular food that is valued when a person is hungry has diminished relevance when the person is sated with that same food. Consequently, it is important for optimal adaptive behaviour to be able to maintain an updated representation of the current value of such sensory-predictive cues that does not slavishly depend on new learning in relation to that cue.

Reinforcer devaluation is a standard experimental methodology for examining how value representations accessed by predictive cues are updated. As indicated, in the case of food, its value can be decreased through what is termed sensory-specific satiety. In this type of manipulation, the reward value of a food eaten to satiety is reduced (devalued) relative to foods that are not eaten to satiety. In humans, functional neuroimaging measured brain responses elicited by predictive stimuli (such as a CS+), that have been subject to devaluation, are associated with significant response decrements in the OFC paralleling the behavioural effects of satiation.[8] This response pattern within OFC indicates that this region is involved in representing reward value of predictive stimuli in a flexible manner, observations that also accord with extensive evidence from animal lesion data.[9, 10]

The observation that neural responses evoked by a food predictive conditioned stimulus (a CS+) in OFC are directly modulated by hunger states can inform an understanding of the behavioural impact of pathologies that impact on orbital-frontal cortex, especially the feeding abnormalities observed in both the Kluver-Bucy syndrome and fronto-temporal dementias. Patients with these conditions frequently show increased appetite, indiscriminate eating, food cramming, and change in food preference, hyperorality, and even attempts to eat non-food items. A dysfunctional network

involving OFC and amygdala would mean that food cues, and other predictive cues, are unable to recruit motivationally appropriate representations of food-based reward value.

A computational account of emotional learning

Learning to predict reward or danger is a basic and highly conserved form of learning, as embodied in Pavlovian or associative learning. However, to be maximally adaptive, it is important that this form of learning is used not only to predict but also to shape optimal actions. The computational principles that underpin what is now referred to as value learning, involving prediction and optimization of action with respect to likely future outcomes, is more than an abstract issue and speaks to the critical issue of optimal control in decision-making.

One classical solution as to how associative learning is implemented is by means of a signal, referred to as a prediction error, which registers a difference between a predicted and actual outcome. This type of solution to predictive learning has been formalized within what is known as the Rescorla-Wagner learning rule. Temporal difference learning (TD) provides a more sophisticated computational extension of this learning rule that accounts in a precise manner for how an organism learns to make predictions, as well as select optimal actions, in response to states of the environment so as to maximize long-term reward or avoid long-term punishment.[11] As in the case of the Rescorla-Wagner model, when a positive (or negative outcome) is not predicted there is a large prediction error which reduces to zero when this same outcome is fully predicted. The function of the prediction error is to act as a teaching signal that can both update future predictions as well as shape optimal policies or action choices. In temporal difference learning (TD), credit is assigned by means of the difference between temporally successive predictions, rather than between a predictive stimulus and an outcome, such that learning occurs whenever there is a change in prediction over time.

The importance of the above theoretical considerations rests upon empirical observations that TD error-like responses have now been demonstrated in the response pattern of dopamine neurones recorded in monkeys during associative learning.[12] Consequently, in a classical conditioning context where a stimulus is followed by an unexpected reward it can be shown that dopamine neurones respond with a burst of action potentials after actual reward receipt. Over the course of learning, with repeated presentations of a predictive stimulus and reward, dopamine neurones no longer respond to receipt of the reward. In this latter case, the reward is accurately predicted because of the occurrence of the preceding predictive stimulus. What is now observed is a prediction error at the time of the earliest predictor of this reward, for example at time of presentation of a predictive CS stimulus. Prediction error type brain responses have also been shown to occur in the human striatum and orbital-prefrontal cortex during both Pavlovian and Instrumental learning in humans, as measured by fMRI.[13,14] Indeed, a crucial link between a dopamine prediction error signal, human striate activity and reward-related choice behaviour in humans has also been shown using fMRI techniques. In this latter case, a reward outcome prediction error signal was enhanced by boosting the impact of dopamine using L-dopa (a precursor of dopamine), while a dopaminergic blocker Haloperidol led to an attenuation of a prediction error signal. Crucially, the former manipulation was associated with enhanced reward learning while the latter was associated with impaired reward learning in a manner that indicates that a reward outcome prediction error is involved in shaping optimal behaviour.[15]

How emotion influences memory

The cognitive domain where the modulatory influences of emotion have been best characterized is with respect to episodic memory, the type of memory that underpins autobiographical experience. Emotion enhances episodic memory function as seen in an enhancement for material that encompasses personal autobiographical, picture, and word based-items, an effect best seen in free recall tasks.[16] The critical role played by the amygdala in this modulation is illustrated by functional neuroimaging experiments where amygdala activity during encoding predicts a benefit in later recall of emotional material relative to neutral material.[17] Thus, enhanced amygdala activity at encoding for both positive and negative stimuli is predictive of later episodic memory function, during free recall tasks.

During encoding of emotional items there are bi-directional interactions between amygdala and hippocampus, the latter structure being a region essential for episodic memory formation. The bi-directional interaction between amygdala and hippocampus is inferred from the fact that an enhanced amygdala response, measured using functional neuroimaging, to presentation of emotional items is dependent on influence from hippocampus. Conversely, an enhanced hippocampal response to emotional items is dependent on influences from the amygdala.[18] While these studies were carried out at encoding it is important to acknowledge a role for the amygdala during retrieval of emotional items and contexts.

How emotion influences perception

Emotion often signals an environmental event of value. From an evolutionary perspective, it is important that such occurrences are amenable to privileged perceptual processing. There appears to be two distinct mechanisms by which emotion can influence perception of such event. One of these is through emotion grabbing attention, leading to enhanced deployment of attention to an emotional eliciting stimulus. This would result in preferential detection of emotional events enabling appropriate adaptive responses to be enacted.

There is also evidence for a second means by which emotion can influence perception that appears to operate independent of attention. For example, in visual backward masking paradigms, a target presented for a brief instance can be rendered invisible if it is immediately followed by a second 'masking stimulus'. In situations where the hidden target stimulus is an emotional item, for example, a conditioned angry face or a spider, there is preserved processing. This is evident in differential skin conductance responses (SCRs) to fear-relevant compared to fear-irrelevant unseen targets.[19] Similar findings are reported using what is referred to as an attentional blink paradigm. The latter refers to a situation where detection of an initial target stimulus, in a stimulus stream, leads to impaired awareness, or inattentional blindness, for a successive second target. Critically, when this second target has emotional content there is an increased probability of its detection as opposed to the default attentional blindness.[20]

In terms of anatomical substrates of these modulatory effects, there is compelling evidence to implicate the amygdala. In functional

neuroimaging experiments, using visual backward masking paradigms, an amygdala response discriminates between unseen emotional and unseen non-emotional target.[19] In other experiments that involve overt stimulus presentation, but where attention is systematically manipulated, such that emotional items are presented out of the window of attention, an amygdala response to emotional stimuli is independent of the concurrent attentional focus.[21] Likewise, in studies of patients with either blindsight (loss of primary visual cortex resulting in visual field blindness) or visual extinction (a situation following a lesion to the right inferior parietal cortex whereby subjects cannot consciously represent stimuli in the contra-lesional visual field) demonstrate an amygdala response to emotional stimuli presented out of awareness in the damaged hemifield.[22]

How pre-attentive processing of emotional events influence, and enhance, perception is an important mechanistic question. One possibility is that inputs from emotional processing regions, in particular the amygdala, modulate the very regions involved in object perceptual processing, specifically when this relates to an emotion eliciting object or event. Anatomically, the amygdala receives visual inputs from ventral visual pathways and sends feedback projections to all levels of this pathway. Neuroimaging data provide evidence for enhancement of the strength of connectivity between amygdala and extra-striate visual regions during processing of an emotional visual input. In patients with amygdala lesions, the enhancement of activity seen in early extra-striate visual areas during encoding of emotional items, for example faces, is no longer expressed. Crucially, neuropsychological data from patients with amygdala damage indicate that a perceptual enhancement seen in extra-striate visual cortex for emotional items is abolished following damage to this structure.[23] This type of evidence is consistent with a proposal that boosting of activity in early sensory cortices, when an emotional stimulus is encountered, reflects a direct modulatory influence from the amygdala.

The neurobiology of subjective feeling states

Human emotion research often conflates the neurobiological mechanisms that index the perception or occurrence of an emotional event (representational aspects of emotion) with their subjective experiential counterparts, usually referred to as feeling states. Feelings can be formally defined as mental representations of physiological changes that characterize, and are consequent upon, processing an emotion eliciting object event or image.[24] This definition assigns an important causal role in the genesis of subjective feeling states to afferent feedback to the brain from the body, both sensory and neurochemical. At a broader level, feeling states can be thought of as reflecting the operation of homeostatic mechanisms that underlie survival of the organism. In a recent theoretical model, based on neurological observations, prime emphasis is given to the cerebral representation of bodily states as providing the substrate for the conscious awareness of feeling states.[25]

A key neurobiological question is whether brain systems supporting emotional perception are distinct from those supporting feelings states. Candidate structures that mediate feeling states encompass those involved in bodily homeostasis and that process information regarding the bodies internal milieu including brain stem peri-acqueductal grey (PAG) and parabrachial nuclei, tegmentum, hypothalamus, insula, somatosensory and cingulate cortices. Functional neuroimaging provides strong evidence that feeling states are mediated by distinct neuronal systems to those that support emotional perception.[26] Thus, functional neuroimaging studies of volunteer subjects have shown that the central generation and re-representation of peripheral autonomic states involve structures such as anterior cingulate and insular cortex. For example, recall of subjective feeling states associated with past emotional experiences engages regions encompassing the upper brainstem nuclei, hypothalamus, somatosensory, insular and orbitofrontal cortices. In subjects with pure autonomic failure (PAF), where there is absence of visceral afferent and information regarding the peripheral body state due to selective acquired peripheral autonomic damage, there is attenuation of subjective emotional feelings as well as emotion evoked neuronal activity in regions implicated in mediating feeling states, such as anterior cingulate and insula cortex.[27]

Among the regions most strongly implicated in mediating subjective feeling states is the insula cortex, an extensive region of cortex enfolded from the cortical surface within temporal lobes. Direct evidence for its role in representing subjective feeling states comes from investigations that tap awareness of internal bodily states, such as that required in performing a heartbeat detection task.[28] In this task, subjects who have the ability to detect and accurately report their own heartbeat, which is seen as evidence of somatic awareness, show enhanced activity in the anterior insula cortex when performing such a task.[29]

The proposal that the insula cortex area mediates subjective feeling states is bolstered by evidence that empathetic awareness of the subjective feeling states of others, for example, that engendered when one observes another person receiving pain, is reflected in enhanced activity within anterior insula and cingulated cortex.[30] These same regions are also engaged when a subject is exposed to a pain eliciting stimulus suggesting the same neural matrix that represents subjective feeling states is engaged when representing the subjective feeling states of another person.

(a) Imaging emotional influences on decision-making

Emotion is frequently invoked as influencing decision-making, often detrimentally, a view that tends to pit a hot 'irrational' emotional decision system in opposition to a cold 'rational' cognitive decision-making system. This dichotomy almost certainly represents a simplification and there are compelling neurobiological reasons to suggest multiple decision-making systems in the human brain with emotion in many instances facilitating optimal decision-making.

Real-life decision-making often involves choices between actions which yield potential rewards or punishments, albeit with some element of uncertainty, for example, as manifest in varying probabilities and magnitudes of outcomes. Adaptive decisions that seek to optimize goal-oriented behaviour require an estimation of expected future reward that will follow from choosing a particular action. This behaviour can be described as utility maximization. As outlined previously, reward prediction based on expected reward value can be studied through classical conditioning, in which an arbitrary cue (or conditioned stimulus) takes on predictive value by association with subsequent delivery of an affectively significant or unconditioned stimulus (which can be a reward or punishment, or strictly speaking, an appetitive or aversive stimulus).

Neuroimaging studies implicate OFC alongside structures such as amygdala and ventral striatum in prediction for reward and punishment.[31] As described above, human neuroimaging studies of classical conditioning for reward have highlighted a prediction error signal in prominent target areas of dopamine neurones, including the striatum and OFC. The finding that a neural reward prediction error signal is expressed, present in OFC and striatum, and indeed throughout the reward network, is consistent with the idea that this signal provides a basis for flexible learning and updating of stimulus-reward associations.[33]

Accounts of human decision-making emphasize rationalistic perspectives which invoke analytic processes mediated by an executive prefrontal cortex. An emotional or value based contribution to high level decision-making is evident following ventromedial prefrontal cortex damage where, despite the absence of intellectual deficits, such patients often make real life decisions that are disadvantageous.[33] The types of deficits seen in these patients have been conceptualized as a myopia for the future, in which current needs (as opposed to an integration of current and future needs) dominate decision-making. Observations from patients with this type of lesion has led to the suggestion that this ventromedial OFC provides access to feeling states evoked by past decisions during contemplation of future decisions of a similar nature. Thus, evocation of past feeling states biases the decision-making process, towards or away from a particular behavioural option.[34] However, alternative frameworks that might explain behavioural deficits seen following damage to this region include an inability to represent the value of competing options for action or extreme discounting of future rewards (a myopia for the future), leading to an overvaluing of current as opposed to future rewards.

It is well recognized that normative human decision-making does not always accord with rationalistic perspectives of utility maximization. An influence of prior emotional experience on decision processes is captured by the consequences of an emotion such as regret. Regret is an emotion generated by counter-factual thinking involved in comparing an obtained and foregone outcome which indicates to the subject that the latter, if chosen, would have been more advantageous. In this sense, regret is also a prototypical example of a secondary or higher order emotion, meaning that it emerges out of cognitive or higher order processing as opposed to being stimulus elicited as in the case of fear or disgust (prototypical exemplars of primary emotions). It is known that subjects who experience regret as a consequence of a choice show a subsequent bias away from a rationalistic imperative that invokes a maximization of expected value when making similar choices, an effect that can be explained as regret minimization. This behavioural bias is associated with engagement of the amygdala and orbitofrontal cortices regions, that are also engaged by the actual experience of regret.[35] This pattern of brain response is consistent with theories that suggest evocation of past emotions in the context of decision-making, providing a biasing influence on rational decision processes.

An additional tenet of rational behaviour is the idea that human decisions should be consistent regardless of how choices are presented. One notable deviation from this axiom is described as a framing effect. In simple terms, the framing effect describes a bias in decision-making observed when choices are presented in terms of gain, leading to choices of a sure as opposed to a risky option, versus the same choices presented in a loss where subjects are biased to choose a risky option. Functional neuroimaging data show that a framing engendered bias in human decision-making, risk aversiveness in the gain and risk seeking in the loss frame, is associated with enhanced amygdala activity at the point of decision-making.[36] The suggestion here is that an emotional heuristic, mediated via key emotion processing brain regions, is invoked when humans make decisions under situations where information is incomplete or overly complex.

Conclusions

The neurobiology of human emotion has now undergone a radical revision with the development of sophisticated neuroimaging technologies. There is a clear evidence that it no longer makes sense to think of the brain in terms of simple dichotomies such as limbic and non-limbic. Emotion engages widespread regions of the brain with the precise regions being dynamically configured as a function of behavioural context. Thus, patterns of brain activity evoked by the seeing a fear eliciting stimulus, such as a snake, are distinct from those evoked when seeing another person in pain. The former situations involve activation of the amygdala and through its modulatory effects it influences widespread interconnected regions, including early sensory cortices. The latter situation results in engagement of distinct structures such as the insula and cingulate cortex. Learning about likely emotional occurrences involves a distinct teaching signal, a prediction error, expressed in widespread brain regions including the striatum and OFC, the latter region mediating a flexible representation of the value of emotional occurrences including reward.

Further information

Glimcher, P. W. and Rustichini, A. (2004). Neuroeconomics: the consilience of brain and decision. *Science,* **306**, 447–52.

Dayan, P. and Balleine, B.W. (2002). Reward, motivation, and reinforcement learning. *Neuron,* **36**, 285–98.

Schultz, W. and Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review Neuroscience,* **23**, 473–500.

References

1. Weiskrantz, L. (1956). Behavioural changes associated with ablation of the amygdaloid complex in monkeys. *Journal of Comparative Physiology and Psychology*, **49**, 381–391.
2. Bechara, A., Tranel, D., Damasio, H., *et al.* (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science*, **269**, 1115–8.
3. Morris, J.S., Frith, C.D., Perrett, D.I., *et al.* (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, **383**, 812–5.
4. Winston, J.S., O'Doherty, J., Dolan, R.J. (2003). Common and distinct neural responses during direct and incidental processing of multiple facial emotions. *Neuroimage*, **20,** 84–97.
5. Amaral, D.G., Price, J.L., Pitkanen, A. *et al.* (1992). Anatomical organization of the primate amygdaloid complex. In *The amygdala: neurobiological aspects of emotion, memory and mental dysfunction* (ed. J. Aggleton), New York, Wiley_Liss.
6. Buchel, C., Morris, J., Dolan, R.J., *et al.* (1998). Brain systems mediating aversive conditioning: an event-related fMRI study. *Neuron*, **20**, 947–57.
7. LaBar, K.S., LeDoux, J.E., Spencer, D.D., *et al.* (1995). Impaired fear conditioning following unilateral temporal lobectomy in humans. *J Neurosci*, **15**, 6846–55.

8. Gottfried, J.A. and Dolan, R.J. (2004). Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. *Nat Neurosci*, **7**, 1144–52.
9. Hatfield, T., Han, J.S., Conley, M., *et al.* (1996). Neurotoxic lesions of basolateral, but not central, amygdala interfere with Pavlovian second-order conditioning and reinforcer devaluation effects. *J Neurosci*, **16**, 5256–65.
10. Malkova, L., Gaffan, D. and Murray, E.A. (1997). Excitotoxic lesions of the amygdale fail to produce impairment in visual learning for auditory secondary reinforcement but interfere with reinforcer devaluation effects in rhesus monkeys. *J Neurosci*, **17**, 6011–20.
11. Sutton, R. and Barto, A. (1998). *Reinforcement Learning; An Introduction* (Adaptive Computation & Machine Learning). Cambridge, MA: MIT Press.
12. Schultz, W. (1997). Dopamine neurons and their role in reward mechanisms. *Curr Opin Neurobiol*, **7**, 191–7.
13. O'Doherty, J., Dayan, P., Schultz, J., *et al.* (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, **304**, 452–4.
14. O'Doherty, J.P., Dayan, P., Friston, K., *et al.* (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, **38**, 329–37.
15. Pessiglione, M., Seymour, B., Flandin, G., *et al.* (2006). Dopaminedependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, **442**, 1042–5.
16. Cahill, L., Prins, B., Weber, M., *et al.* (1994). Beta-adrenergic activation and memory for emotional events. *Nature*, **371**, 702–4.
17. Canli, T., Zhao, Z., Brewer, J., *et al.* (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *J Neurosci*, **20**, RC99.
18. Richardson, M.P., Strange, B.A. and Dolan, R.J. (2004). Encoding of emotional memories depends on amygdala and hippocampus and their interactions. *Nat Neurosci*, **7**, 278–85.
19. Morris, J.S., Ohman, A. and Dolan, R.J. (1998). Conscious and unconscious emotional learning in the human amygdala [see comments]. *Nature*, **393**, 467–70.
20. Anderson, A.K. and Phelps, E.A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, **411**, 305–9.
21. Vuilleumier, P., Armony, J.L., Driver, J., *et al.* (2001). Effects of Attention and Emotion on Face Processing in the Human Brain. An Event-Related fMRI Study. *Neuron*, **30**, 829–41.
22. Vuilleumier, P., Armony, J.L., Clarke, K., *et al.* (2002). Neural response to emotional faces with and without awareness: event-related fMRI in a parietal patient with visual extinction and spatial neglect. *Neuropsychologia*, **40**, 2156–66.
23. Vuilleumier, P., Richardson, M.P., Armony, J.L., *et al.* (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nat Neurosci*, **7**, 1271–8.
24. Damasio, A. (1999). *The Feeling of What Happens*. New York, Harcourt Brace.
25. Damasio, A.R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philos Trans R Soc Lond B Biol Sci*, **351**, 1413–20.
26. Critchley, H.D., Mathias, C.J. and Dolan, R.J. (2001). Neuroanatomical basis for fi rst- and second-order representations of bodily states. *Nat Neurosci*, **4**, 207–12.
27. Damasio, A.R., Grabowski, T.J., Bechara, A., *et al.* (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nat Neurosci*, *3*, 1049–56.
28. Critchley, H.D., Mathias, C.J. and Dolan, R.J. (2002). Fear conditioning in humans: the infl uence of awareness and autonomic arousal on functional neuroanatomy. *Neuron*, **33**, 653–63.
29. Critchley, H.D., Wiens, S., Rotshtein, P., *et al.* (2004). Neural systems supporting interoceptive awareness. *Nat Neurosci*, **7**, 189–95.
30. Singer, T., Seymour, B., O'Doherty, J., *et al.* (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, **303**, 1157–62.
31. Knutson, B., Adams, C.M., Fong, G.W., *et al.* (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J Neurosci*, **21**, RC159.
32. Elliott, R., Dolan, R.J. and Frith, C.D. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: evidence from human neuroimaging studies. *Cereb Cortex*, **10**, 308–17.
33. Anderson, S.W., Bechara, A., Damasio, H., *et al.* (1999). Impairment of social and moral behaviour related to early damage in human prefrontal cortex. *Nat Neurosci*, **2,** 1032–7.
34. Bechara, A., Damasio, H., Tranel, D., *et al.* (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, **275**, 1293–5.
35. Coricelli, G., Critchley, H.D., Joffily, M., *et al.* (2005). Regret and its Avoidance: A Neuroimaging Study of Choice Behaviour. *Nature Neuroscience*, **8**, 1255–1262.
36. De Martino, B., Kumaran, D., Seymour, B., *et al.* (2006). Frames, biases, and rational decision-making in the human brain. *Science*, **313**, 684–7.

2.5.5 Neuropsychological basis of neuropsychiatry

L. Clark, B. J. Sahakian, and T. W. Robbins

Introduction

Neuropsychology makes an essential contribution to neuropsychiatry. It seeks objectively to characterize mental competence in component cognitive functions such as perception, attention, spatial cognition, memory, learning, language, thinking, and 'executive' function. Executive function is often associated with the functions of the frontal lobes, although these are not at all synonymous; we will pay special attention to this domain below, as it may be crucial to the understanding of several neuropsychiatric disorders. Neuropsychology is often conveniently divided into clinical neuropsychology and cognitive neuropsychology. The former is primarily concerned with the methodology and psychometric theory that lies behind the selection, administration, and interpretation of standardized psychological tests aimed at assessing deviation from the norm and an individual patient's profile of strengths and weaknesses with a view to optimizing functional outcome and quality of life. Cognitive neuropsychology, by contrast is more concerned with the elucidation of cognitive processes through the study of patients, using both classical and newly devised tasks.[1] Neuropsychology also forms part of cognitive neuroscience, which has as its major goal the understanding of normal, as well as abnormal cognitive function, not only through the neuropsychological study of patients and healthy controls, but also using other techniques, including functional neuroimaging and the use of transcranial magnetic stimulation or psychopharmacology. In practical

terms, a neuropsychological assessment is often made together with a psychiatric examination; in addition to contributing to diagnosis it also helps to define the functional status of the patient.

Functions of neuropsychological assessment

These are perhaps easiest to define when there has been an organic brain injury causing a lesion; for example, due to a stroke, removal of a tumour, or a closed or penetrating head injury. It is clearly vital to have an accurate evaluation of a patient's cognitive status so that his or her care can be optimized. This applies equally in neuropsychiatric disorders. In conditions like attention-deficit/hyperactivity disorder (ADHD) or prodromal Alzheimer's disease, the cognitive examination provides essential information for making a diagnosis. Whilst acquired brain damage can lead to neuropsychiatric symptoms (e.g. depression or apathy), it may also be associated with specific patterns of cognitive deficit. However, most neuropsychiatric disorders are not associated with clearly defined brain injuries, but are instead hypothesized to result from the cumulative effects of neurodegenerative disease, neurotransmitter malfunctions, developmental hypoplasias, diffuse white matter lesions, or brain volume gains and losses. This is an important point, as it may require more sensitive new methodologies to characterize deficits produced by regional *overactivity*, or deficits associated with changes in regional *connectivity*, as distinct from brain lesions per se. In these examples, neuropsychology offers the opportunity to examine underlying pathophysiological mechanisms, in combination with other methods such as structural and functional brain imaging and evoked brain potentials.

Cognitive deficits (e.g. of memory) can exist in parallel to, and independently of, psychiatric symptoms (e.g. of melancholia) but they may also be intrinsic to them. For example, part of the psychiatric description of anxiety may emphasize abnormal attentional biases paid to threatening stimuli, which can be objectively assessed using cognitive testing.[2] Furthermore, some psychological factors may influence performance across a number of distinct cognitive domains, producing a broad profile of impairments from a relatively specific form of deficit. For example, patients with depression may show a 'catastrophic response' to receiving error feedback during testing, such that they are then more likely to respond incorrectly on the subsequent trial.[3] This interaction of emotional and social factors with cognitive processes forms an especially important part of neuropsychology as applied to psychiatry. Neuropsychology also enables the impact of neuropsychiatric symptoms to be assessed on functional status—whether the patient will be able to function in everyday life, and return to paid employment, and how rehabilitation may be best achieved. It is becoming clear that the effective treatment of certain symptoms (e.g. psychotic symptoms) can sometimes unmask profound cognitive impairments that are actually the main barrier to rehabilitation, as has recently been shown in schizophrenia.[4,5] Taking into account the corollary finding that cognitive status may be the best predictor of functional outcome and return to paid employment, this has made cognitive disabilities a new target for pharmaceutical treatment. Of course, with schizophrenia again in mind, it is also necessary to ascertain that medication is not associated with significant cognitive toxicity, for example in terms of sedation, distractibility, or impaired judgement. Neuropsychology must play a major role in providing such evidence. Sometimes neuropsychology may substantially contribute to the diagnosis itself of a psychiatric state, for example in dissociative disorders and in the study of fatigue disorders and malingering.

Principles of neuropsychological testing

Neuropsychological scores on most tests are standardized with respect to overall age, IQ, and ideally gender. Parallel forms of tests exist in different languages although cultural and ethnic factors are still difficult to take account of adequately. Testing is generally done in a quiet room without distraction by an experienced clinical neuropsychologist. Most tests have generally been shown to give consistent results when given by different testers and on different occasions to the same patient, when using standardized instructions. These 'inter-tester' and 'test–retest' forms of reliability are often critical factors in situations where testing has to be repeated, for example, in drug trials, or epidemiological studies.[6] Good test–retest reliability (i.e. $r > 0.8$) can be hindered by practice effects that markedly change how subjects approach the tasks. Such factors typically affect measures of executive function, where the subject may evolve strategies for dealing with the task over repeated test sessions. The various cognitive domains are usually tested in one or two sessions that contain an assortment of tests drawn from the types described above to provide a cognitive profile of the patient. The duration of the test sessions should be carefully considered in view of the concentration span and distractibility, or apathy of many patients. Neuropsychological test batteries normally comprise a selection of paper and pencil tests which are being supplemented increasingly with computerized elements. Computerized tasks benefit from the ease and accuracy of recording and analysing complex data (e.g. reaction times), and of standardizing the presentation of the test materials and trial-by-trial feedback. The Cambridge Neuropsychological Test Automated Battery (CANTAB), for example, utilizes a touch sensitive screen that allows the subject to interact directly with the test materials and obviates the need for divided attention between a video monitor screen and keyboard or desktop. Other possible advantages of computerized testing include the removal of the 'confrontational' or 'interrogative' elements of conventional testing, which may be especially advantageous when testing, for example, schizophrenic patients. The construction of batteries may also be affected by other factors such as the need to translate findings, presumably through the use of non-verbal tests, across species, based on the extensive knowledge of underlying neurobiological mechanisms gained through studies of non-human primates. Alternatively, there is a trend to customize neuropsychological batteries so as to focus on typical deficits in a given disorder, as illustrated by the recent derivation of the MATRICS battery for cognitive deficits in schizophrenia.[7]

Domains and neuropsychological tests of cognitive function

Neuropsychological assessment is made generally with respect to the overall profile of cognitive performance. For example, it is difficult to interpret an apparent memory deficit if the subject has a profound impairment in perception or attention. An important index of overall performance is the intelligence quotient, or IQ. The structure of intelligence is still being debated; whether there is a distinction for example between Cattell's fluid and concrete

intelligence, and the possible existence of a general factor, Spearman's *g* versus Thurstone's more specific components.[8] Regardless of these theoretical issues, it is still useful to classify an individual in terms of their overall intelligence with a mean scaled score of 100 and a standard deviation of 15. IQ is often measured using sub-tests from the Wechsler Adult Intelligence Scale (or the child equivalent, the WISC), which broadly subdivides into verbal and non-verbal ('performance') components. The individual sub-tests include such categories as vocabulary, information, comprehension, arithmetic, digit span, similarities, block design, picture arrangement, picture completion, object assembly, and digit symbol, which thus probe a range of abilities from general knowledge and basic language skills to memory span, working memory, visuo-spatial construction, and psychomotor speed. As the time taken to make all these assessments can be prohibitive, a 'prorated' score based on a smaller selection of the tests is often employed, justified by the relatively high inter-correlation of performance among the 12 sub-tests. Intelligence is also measured effectively by the Raven's Matrices, a set of visuo-spatial problems based on analogies.

Frequently, it is useful to be able to gauge the patient's premorbid intelligence level, before the onset of psychiatric symptoms. One way of estimating this is from years of education, as frequently employed in the United States. A second method depends on the National Adult Reading Test (NART), an instrument that depends on the subject's ability to pronounce infrequent words; this correlates with educational level, and its utility was realized when it became apparent that patients with Alzheimer's disease showed relative sparing of reading abilities, thus enabling their premorbid IQ to be captured by this test.[9] Alongside its US analogue, the Wechsler Test of Adult Reading (WTAR), the NART is now widely used to estimate premorbid intelligence in neuropsychiatric disorders including schizophrenia.

Superimposed on this general assessment of IQ is performance on tests of more specific abilities. The Wechsler Memory scale provides a method by which memory can be assessed in the context of overall intelligence. Some of its components, such as the Logical Memory test are still much used. However, with our burgeoning theoretical understanding of the components of cognition has come the introduction of ever more sophisticated instruments for measuring sub-components of cognitive performance. This brief chapter can but summarize some of the consequences of this, and encyclopaedic compilations of the various tests are now available, together with details of their mode of administration and interpretation.[10] However, the main domains of function that are generally evaluated are now outlined.

Over and above basic clinical sensory testing, measures of higher order perception in either the auditory or the visual modalities are available—for example, the Visual Object and Space Perception Battery.[11] Memory function is a controversial and complicated area of assessment. Most rapid batteries include tests of recognition memory, simply deciding whether or not a stimulus is familiar or not. Recent evidence, for example from the animal literature, suggests that such memory is relatively independent of hippocampal function, and depends instead on such regions as the perirhinal cortex.[12] Verbal recall, particularly a free recall, is considered to be a more demanding form of memory, thought to require the coordinated functioning of medial temporal lobe and frontal lobe structures. The frontal lobes are also implicated in retrieval of the temporal sequence or order of events, or the source of a particular memory in the past. Tulving's distinction between episodic and semantic memory is still influential.[13] Episodic memory generally refers to the subjective reminiscence of an event that has 'what, where, and when' qualities, almost invariably associated with a person's autobiographical memory. By contrast, semantic memory reflects memory for facts that may well have a different form of representation within the temporal lobe and is importantly influenced by verbal processes. In addition to these various modalities of memory material, are specific memory processes that have to be evaluated, such as encoding and retrieval. The latter is often assessed efficiently by the so-called verbal fluency tests, in which subjects have to retrieve words from a semantic category (e.g. animals) or beginning with a specific letter (e.g. F) in a set time period, usually of 60 s. Standardized tests of learning of verbal or non-verbal forms of material are provided by the California Verbal Learning Test (CVLT), the Rey Auditory Verbal Learning Test (RAVLT), and the CANTAB Paired Associate Learning (PAL) test. Emotional memory, which almost certainly implicates such structures as the amygdala, is another area of memory research that requires urgent development for application to psychiatry, for such conditions as Post-Traumatic Stress Disorder (PTSD). However, it is difficult to see how the current tests of traditional neuropsychology can be of much assistance here when the problem is to quantify the disruptive effect of a specific memory that has become over-salient rather than inaccessible.

Working memory is another important component of memory that refers to the coordination of various short-term memory stores to provide more permanent representations and aid in the solution of ongoing activities such as planning and discourse. Working memory comprises short-term visuo-spatial and verbal memory stores, with postulated rehearsal processes (e.g. the articulatory 'loop') and a 'central executive' system. Perhaps a simple way of understanding what working memory accomplishes is to contrast digit span in a forward and backward modes (e.g. repeat back the sequence '5 2 3 7 1' in either the same order, or in the reverse order). In the former case, it represents the buffer capacity of the verbal store, whereas in the latter case, the response sequence requires manipulation at output. Working memory reflects the active ('conscious') processing of memory traces whether at the encoding or retrieval stages and is clearly dependent on such factors as attention and arousal level.

Attention is a complex theoretical construct that has gained much in recent years from its analysis in terms of underlying neural systems.[14] Unfortunately, despite the interest in attention from experimental psychology, there are not many standardized tests of attention in general use. A common distinction is between the automatic forms of 'covert' attention mediated by posterior cortical structures and voluntary forms engaged by anterior cortical (cingulate and prefrontal cortices).[15] In practical terms, tests are usually provided of *sustained attention*, which might include detecting infrequent visual or auditory targets in 'continuous performance' tasks over a protracted period of several minutes. *Vigilance*, the capacity to detect rare targets over very long time periods, is rarely assessed, for obvious practical reasons. The capacity for *selective attention* is very important to assess in psychiatry because of the evident distractibility of patients with anxiety, psychosis, or attentional deficit/hyperactivity disorder. Selective

attention can be assessed with a variety of tests including the Stroop task, the Eriksen Flanker paradigm, or the Posner covert spatial attentional task. In the Stroop test, for example, the subject is required to name the ink colour of congruent or incongruent colour words (e.g. the word GREEN prinked in red ink). Resolution of response conflict is a key feature of this test, which has been associated with anterior cortex function through extensive investigation with functional brain imaging.[16] The Stroop test has been much used in psychiatry, both in its standard form and in various adapted forms with disorder-relevant stimuli. In the Alcohol Stroop, for example, alcohol-related stimuli (e.g. BOTTLE in red ink) may also interfere with naming of the ink colour in patients with alcohol-dependency problems.

There is, in fact, considerable overlap between tests of working memory, attention, and what is termed executive function. The notion of a set of control functions that optimize performance is controversial but seemingly necessary in accounting for performance on complex tests involving planning and decision-making. Planning is generally measured by such tests as the 3-disc Tower of London task, which requires subjects to move discs in one array to match the goal arrangement of another. The test is especially useful in its form where the subjects are not allowed to move the discs but are required to visualize and plan the solution in their 'mind's eye'. Validation of this test as being sensitive to frontal lobe function has come from the study of patients with neurosurgical lesions of the frontal lobe and also functional brain imaging.[17] However, it should be realized that there is no such thing as a 'frontal lobe' test; this is evident for example from studies of patients with basal ganglia lesions and also from the demonstration of an extensive neural network that is activated during spatial planning performance. The Tower of London is a planning task where there is one goal only, whereas many real life situations that overwhelm psychiatric patients require the optimization of performance on several task simultaneously. This requires high level planning and scheduling of behaviour, as tapped by the 'Six Elements Test' in which six different tasks have to be completed.

A further cluster of executive tasks tap into higher-level emotional and affective processes, which appear to be associated with the inferior parts of the frontal lobes including the ventromedial prefrontal cortex. Famous neurological case studies like Phineas Gage and EVR displayed profound changes in personality, social behaviour, and judgement after lesions to this brain area. These processes are also often defective in psychiatric disorders such as bipolar disorder and schizophrenia, which resemble in some respects, the problems in everyday life encountered by patients with ventromedial prefrontal cortex lesions. The Iowa Gambling Task is a measure of emotional decision-making that was developed for the assessment of such patients. The Iowa task allows subjects to 'play' with four decks of cards which vary in their pay-offs, such that two risky decks are associated with attractive short-term gains but with heavy occasional penalties over time. Typically, patients with ventromedial frontal lesions persist in selecting from these 'dangerous' decks, despite accruing considerable debt, and comparable neuropsychological deficits have also been seen in neuropsychiatric conditions, particularly drug dependence.[18] These tasks also draw on concepts of inhibitory control over behaviour, as there is a requirement to suppress responses that have either become dominant through repeated practice, or that are superficially attractive by virtue of immediate reinforcement.

Specific applications of neuropsychological tests to psychiatric disorders

(a) Standardization of neuropsychological assessment in schizophrenia

Now that cognitive deficits have been realized to be a core problem in schizophrenia, there is increasing attention being focused on the nature of these cognitive deficits and how to remediate them, for example by novel pharmacological treatments. In order to facilitate this process, NIH funded researchers to perform a meta-analysis of the cognitive deficits in schizophrenia in order to determine a profile of deficits and to construct a battery that would be sensitive to those deficits and to possible remediation.[19] They provided a principal component analysis suggesting cognitive impairments to be present in seven major domains: attention and vigilance, visual long-term memory, verbal long-term memory, working memory, reasoning and problem solving, psychomotor speed, and social cognition.[7] In fact, there was not very much published evidence for prominent deficits in the latter, but the investigators felt that it was justified to include it on the basis of clinical judgement. This led secondarily to the construction of a test battery for those domains, with a major requirement being its test–retest reliability. This battery is being used now in clinical trials and so it is too early to assess its utility. As with many other initiatives to provide a standard battery, this carries with it both advantages and disadvantages. A consensus battery is clearly of enormous value and aids the collection of data across multiple sites. Hard decisions do have to be made about the trade-off of reliability with validity and sensitivity and committees have to make compromises because of the difficulties of addressing these sometimes intangible issues. For example, the Wisconsin Card Sort test is clearly sensitive to deficits in schizophrenia, but it does have relatively poor test–retest reliability because of its intrinsic changes of contingency and requirement for cognitive shifting, and so has been excluded. However, it is possible that performance on such a test would be sensitive to cognitive enhancing drugs. Similarly, all tests using verbal material are problematic because of the need to provide translated versions whose reliability and relationship to the existing test has to be reassessed to determine its validity. Once batteries such as MATRICS are adopted it is sometimes difficult for innovative methods based on other theoretical perspectives, such as contemporary cognitive neuroscience to be developed. However, the recognition of a domain of cognitive function, social cognition, for which there are relatively few standardized instruments available has been useful for motivating new research and test development in that area. The final crucial issue, which has a much more general applicability than the MATRICS battery, is the need to relate the profile of deficits obtained to sensitive and reliable measures of functional outcome.

(b) Early detection of Alzheimer's disease

Alzheimer's Disease (AD) is a chronic and severe form of dementia that is increasing in prevalence as improved health care extends life expectancy. AD patients typically present with deficits in episodic memory; that is, memory for specific events or experiences that can be defined in time and space. AD poses an important challenge to neuropsychologists, to facilitate the early detection of the illness, so that future generations of pharmacological interventions may be administered to patients at the earliest opportunity, so that the

rapid decline may be slowed, or eventually, halted. Individuals in the prodrome of AD often report subjective memory problems but do not fulfil clinical criteria for AD, in that activities of daily living and non-memory cognitive faculties are intact. This condition is referred to as Mild Cognitive Impairment (MCI) or 'Questionable Dementia'. There is accumulating evidence that the Paired Associates Learning test in the CANTAB assessment may be sensitive to the early stages of AD. PAL is a non-verbal test of learning and memory that requires the subject to associate abstract visual patterns with a series of box locations on a computer screen. Successful performance relies on learning the conjunction of shape and location; the task cannot be solved by learning either shape or location alone. The task increases in difficulty from one to eight patterns, and the task is terminated if the subject fails repeatedly at a given stage.

PAL performance was previously shown to be defective in patients with mild Alzheimer's disease, and test scores could discriminate AD from other neuropsychiatric conditions with similar symptom presentation, including Fronto-Temporal Dementia and unipolar depression.[20] Recent studies of 'at risk' older adults with MCI or Questionable Dementia have reliably isolated a subgroup of cases who have an increased likelihood of converting to a full AD diagnosis by the time of a follow-up assessment. PAL performance in the baseline assessment could be used to accurately predict the conversion from Questionable Dementia to an AD diagnosis.[21] Through a combination of age, PAL score, and performance on the Graded Naming test, subjects with Questionable Dementia could be classified with 100 per cent accuracy into those who did or did not convert to probable AD at a 32-month follow-up.[22] These neuropsychological measures display both sensitivity, in reliably detecting the progression to AD in a subclinical sample, as well as specificity, in differentiating AD from depression. A strong programme of translational research with the PAL task suggests this utility may stem from the critical role of medial temporal lobe structures in visuo-spatial associative learning.

(c) Relating neuropsychiatric disorders to specific impairments in executive function: ADHD, OCD, and mania

Recent biological analyses of such disorders as bipolar illness, ADHD, and OCD have focused on the importance of providing objective measures of performance that relate strongly to the clinical symptoms themselves, but that also provide sensitive indices of the underlying pathophysiology. These markers are often referred to as endophenotypes, a term that draws a further link to the putative genetic susceptibility to these conditions. One example is the profound deficit in thinking and decision-making that is a characteristic symptom of mania—a recent study has been able to demonstrate deficits in decision-making in patients with mania in the context of a gambling task.[23] What was especially striking about that study was that one of the main measures of quality of decision-making correlated significantly with symptom ratings on the Young Mania scale.

ADHD is a spectrum disorder with several sub-types that are characterized by the prominence of different clusters of symptoms that may be related to different underlying psychological constructs, neural and neurochemical substrates, and genetic factors.[24] Prominent among these is the so-called hyperactive/impulsive sub-type, which then requires sensitive and reliable measures of behavioural inhibition. One commonly used test is the Go-No-Go type task. In perhaps its most sophisticated form, this task measures the capability to countermand a speeded response (the 'stop signal reaction time (SSRT) task). The speed of the SSRT has been related to the volume of damage in the right inferior frontal gyrus in patients with frontal damage.[25] The task is also robustly impaired in children and adults with ADHD, and this deficit can be remediated by psychostimulant treatment with drugs such as methylphenidate.[26]

OCD may also relate to pathology in executive functions. The compulsive component in particular may arise from the repeated selection of a response option long after that option has ceased to be beneficial or contextually appropriate. This symptom may be operationalized in a neuropsychological setting in two complementary ways. First, there may be a failure to suppress a previously reinforced response, comparable to the deficit reported above in ADHD. Second, there may be a failure to flexibly shift responding to the newly-relevant mode. These processes can be assessed with some further tasks that are frequently employed in the assessment of psychiatric patients. For example, the Wisconsin Card Sorting test requires subjects to learn a rule on the basis of trial-and-error feedback and then to shift that rule according to altered feedback. OCD patients, like neurosurgical patients with frontal lobe lesions, display perseveration where they continue to sort cards according to the previously reinforced rule. Performance on a similar set-shifting test has been shown to be impaired not only in patients with OCD, but also in their first-degree relatives, implying that this capacity for cognitive flexibility may provide a suitable endophenotype for this disorder. Structural and functional neuroimaging data support a neuropsychological account that implicates the orbitofrontal region and interconnected (ventral) striatal circuitry in the pathophysiology of this condition.[27]

Conclusion

Psychiatry is the science of psychopathology, and as such, the measurement of behaviour and cognition is central to its theory and methodology. Neuropsychology provides such measures, which can be used to augment the psychiatric interview and other clinical instruments, as well as to provide an interface with other important approaches including functional brain imaging (see Dolan, this volume), functional genomics, and clinical psychopharmacology. Clinical neuropsychology has developed via the need to assess brain-damaged patients, whereas in most neuropsychiatric disorders, such damage is much less well defined if it is present at all. Thus, whilst there is growing information about specific brain abnormalities in many forms of neuropsychiatric disorder, the lesion model is not necessarily the most appropriate. Moreover, some of the deficits in disorders such as depression and anxiety involve subtle interactions between specific emotional and attentional mechanisms with cognitive function. Therefore, the study of neuropsychiatric patients has also enriched our understanding of clinical neuropsychology. We predict that these aspects of the discipline will develop considerably in the next few years, particularly in combination with data from other domains such as functional brain imaging and pharmacogenetics. Indeed, the specification of specific neural systems implicated in core behavioural or cognitive processes may well aid the enterprise of psychiatric genetics by providing more precise definitions of phenotypes (or endophenotypes)

than are currently feasible in nosology (e.g. as defined by the Diagnostic and Statistical Manual).

A further theme for the future that can readily be envisaged is the development of neuropsychological methods for children and adolescents that are more suitable in a psychiatric context. We can predict that there will be an increasing emphasis on 'lifespan' studies that potentially will enable the origins of many psychiatric disorders and prodromal states to be identified. Thus, it will be necessary to examine cognitive function, including under-researched areas such as 'social cognition', in longitudinal terms with tests that can be administered appropriately at different points in the lifespan. Allied to this will be pressure to make tests less 'laboratory-bound' and more predictive of everyday functioning at school or in the workplace. One technological advance that may facilitate all of these requirements would be through the use of 'virtual reality' software, and also the collection of norms on a massive scale by utilizing web-based data collection. It is to be hoped that psychiatry will encourage and embrace such developments, rather than rely on the traditional methods.

Further information

Baddeley, A.D. (2007). *Working memory, thought and action*. Oxford University Press, New York.

Gottesman, I.I. and Gould, T.D. (2003). The endophenotype concept in psychiatry: etymology and strategic intentions. *The American Journal of Psychiatry*, **160,** 636–45.

Lezak, M.D., Howieson, D.B., and Loring, D.W. (2004). *Neuropsychological assessment* (4th edn). Oxford University Press, New York.

Robbins, T.W. (2007). Shifting and stopping: fronto-striatal substrates, neurochemical modulation and clinical implications. *Philosophical Transactions of the Royal Society of London (Biological Sciences)*, **362**, 917–32.

Ron, M.A. and Robbins, T.W. (2004). *Disorders of brain and mind*, Vol. 2. Cambridge University Press, Cambridge.

References

1. Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge University Press, Cambridge.
2. Williams, J.M., Mathews, A., and MacLeod, C. (1996). The emotional stroop task and psychopathology. *Psychological Bulletin*, **120**, 3–24.
3. Elliott, R., Sahakian, B.J., Herrod, J.J., *et al.* (1997). Abnormal response to negative feedback in unipolar depression: evidence for a diagnosis specific impairment. *Journal of Neurology, Neurosurgery, and Psychiatry*, **63**, 74–82.
4. Green, M.F. (1996). What are the functional consequences of neurocognitive deficits in schizophrenia? *The American Journal of Psychiatry*, **153**, 321–30.
5. Turner, D.C. and Sahakian, B.J. (2006). Analysis of the cognitive enhancing effects of modafinil in schizophrenia. *Progress in Neurotherapeutics and Neuropsychopharmacology*, **1**, 133–47.
6. Ferris, S.H., Lucca, U., Mohs, R., *et al*. (1997). Objective psychometric tests in clinical trials of dementia drugs. Position paper from the international working group on harmonization of dementia drug guidelines. *Alzheimer's Disease and Associated Disorders*, **11**(Suppl. 3), 34–8.
7. Green, M.F., Nuechterlein, K.H., Gold, J.M., *et al*. (2004). Approaching a consensus cognitive battery for clinical trials in schizophrenia: the NIMH-MATRICS conference to select cognitive domains and test criteria. *Biological Psychiatry*, **56**, 301–7.
8. Mackintosh, N.J. (1998). *IQ and human intelligence*. Oxford University Press, New York.
9. Crawford, J.R., Parker, D.M., and Besson, J.A. (1988). Estimation of premorbid intelligence in organic conditions. *The British Journal of Psychiatry*, **153**, 178–81.
10. Spreen, O. and Strauss, E. (1991). *A compendium of neuropsychological tests*. Oxford University Press, New York.
11. Warrington, E.K. and James, M. (1991). *The visual object and space perception battery*. Harcourt Assessment, San Antonio, TX.
12. Murray, E.A., Bussey, T.J., and Saksida, L.M. (2007). Visual perception and memory: a new view of medial temporal lobe function in primates and rodents. *Annual Review of Neuroscience*, **30**, 99–122.
13. Tulving, E. (2002). Episodic memory: from mind to brain. *Annual Review of Psychology*, **53**, 1–25.
14. Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, **18**, 193–222.
15. Posner, M.I. and Petersen, S.E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, **13**, 25–42.
16. Bush, G., Luu, P., and Posner, M.I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, **4**, 215–22.
17. Robbins, T.W. (1998). Dissociating executive functions of the prefrontal cortex. In *The prefrontal cortex: executive and cognitive functions* (eds. A.C. Roberts, T.W. Robbins, and L. Weiskrantz). Oxford University Press, Oxford.
18. Bechara, A. and Damasio, H. (2002). Decision-making and addiction. I. Impaired activation of somatic states in substance dependent individuals when pondering decisions with negative future consequences. *Neuropsychologia*, **40**, 1675–89.
19. Kern, R.S., Green, M.F., Nuechterlein, K.H., *et al.* (2004). NIMH-MATRICS survey on assessment of neurocognition in schizophrenia. *Schizophrenia Research*, **72**, 11–19.
20. Swainson, R., Hodges, J.R., Galton, C.J., *et al.* (2001). Early detection and differential diagnosis of Alzheimer's disease and depression with neuropsychological tasks. *Dementia and Geriatric Cognitive Disorders*, **12**, 265–80.
21. Blackwell, A.D., Sahakian, B.J., Vesey, R., *et al.* (2004). Detecting dementia: novel neuropsychological markers of preclinical Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, **17**, 42–8.
22. De Jager, C., Blackwell, A.D., Budge, M.M., *et al.* (2005). Predicting cognitive decline in healthy older adults. *The American Journal of Geriatric Psychiatry*, **13**, 735–40.
23. Murphy, F.C., Rubinsztein, J.S., Michael, A., *et al.* (2001). Decision-making cognition in mania and depression. *Psychological Medicine*, **31**, 679–93.
24. Castellanos, F.X. and Tannock, R. (2002). Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. *Nature Reviews Neuroscience*, **3**, 617–28.
25. Aron, A.R., Fletcher, P.C., Bullmore, E.T., *et al.* (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nature Neuroscience*, **6**, 115–16.
26. Lijffijt, M., Kenemans, J.L., Verbaten, M.N., *et al.* (2005). A meta-analytic review of stopping performance in attention-deficit/hyperactivity disorder: deficient inhibitory motor control? *Journal of Abnormal Psychology*, **114**, 216–22.
27. Chamberlain, S.R., Blackwell, A.D., Fineberg, N.A., *et al.* (2005). The neuropsychology of obsessive compulsive disorder: the importance of failures in cognitive and behavioural inhibition as candidate endophenotypic markers. *Neuroscience and Biobehavioral Reviews*, **29**, 399–419.

2.6

The contribution of social sciences

Contents

2.6.1 Medical sociology and issues of aetiology

George W. Brown

Introduction

David Mechanic, in his pioneering textbook, *Medical Sociology*,[1] views human activity within an adaptive framework—as a struggle of human beings to control their environment and life situation. While this view informs the research to be outlined, there are a number of ways it differs in emphasis from much medical sociology. First, by its concern with particular disorders defined in medical terms. Second, by its use of the investigator rather than respondent to characterize phenomena—to decide, for example, whether an incident should be classified as a life event. Third, by the importance placed on context. In order, for example, for an investigator to make a judgement about the likely meaning of an event such as a loss of a job it is essential to know whether it cast the person in a bad light; its impact on the person's family; her chance of getting another job, and so on. It is such circumstances surrounding an event that usually give it meaning via the emotion they create. Finally, by recognizing that where appropriate such emotion should be taken into account: 'A world experienced without any affect would be a pallid, meaningless world, and it is what gives us feedback about what is what is good or bad about our lives'.[2]

Context and measurement

Concern with context in the social sciences was central to the problem of meaning discussed widely in Germany in the late nineteenth century; and the ideas were introduced into sociology by Max Weber[3] and into psychiatry by Karl Jaspers,[4] although no one showed how to apply the methods systematically to concrete examples.[5] Jaspers, in his *Allegemeine Psychopathologie*, emphasized the way in which *Verstehen*, or understanding, on the part of an investigator 'depends primarily on the tangible facts' (i.e. verbal contents, cultural factors, people's acts, ways of life, and expressive gestures) in terms of which the connection is understood, and which provides the objective data.[4] While this view influenced the approach to meaning in what follows, there is a critical difference. No attempt has been made to make a judgement about the presence of a causal link between a set of circumstances and a psychiatric episode. The investigator has to judge only the likely meaning of a set of circumstances in the light of 'tangible facts' about a person's past and present. Any link with disorder is explored using established scientific procedures.

As noted by Jaspers, it is possible to take note of cultural factors; for example, when rating the likely implications of a birth, as part of research among women in a black township in Zimbabwe, investigators took into account the cultural importance on a wife producing a male child for her husband and his family.[6]

A second, more limited, use of context deals with the actual observation of emotion. For example, the Camberwell Family Interview, by taking account of vocal (in contrast to verbal) aspects of speech, for example, establishes how far a parent's talk about a child conveys 'criticism' rather than 'dissatisfaction'.[7] The approach can be extended to deal with core sociological concepts such as role commitment. For example, mothers in North London have been shown to differ substantially in commitment to roles such as 'mother' or 'wife' judged by how enthusiastically associated activities are discussed.[8] The relevant context is limited to the interview itself and what this conveys about a person's emotional style. In everyday life we automatically make allowances for the fact, for example, that some people show warmth in a more open way and, by taking this into account, makes it is possible for different expressive styles to be treated as equivalent.

Some methodological considerations

These developments have enabled 'soft' variables to be quantified. It has also been possible to make a reasonably persuasive case that significant bias has not followed the use of the investigator as the measuring instrument. For example, Creed[9] in a study of appendectomy found a relationship between severely threatening events rated contextually and the onset of non-inflamed but not inflamed conditions. This result was persuasive for two reasons. First, on the basis of a detailed description of the event and surrounding circumstances a consensus rating team reached agreement about the likely degree of threat, blind to what the person conveyed she felt and to whether she was a patient. Second, Creed, who provided the team with this edited account, was blind to clinical details. (He consulted medical records after the ratings had been made.) Such flexibility is difficult, if not impossible, with a questionnaire-based instrument which hands over the task of measurement to the respondent.

There are now a number of investigator-based instruments developed to deal with psychiatric issues covering areas ranging from 'expressed emotion' (e.g. critical comments, warmth, etc.), attitudes to self (e.g. self-esteem), plans and concerns (e.g. commitment to various roles), behavioural systems (e.g. styles of attachment), experience of adversity in childhood (e.g. sexual and physical abuse), and characteristics of non-family groups (e.g. restrictiveness of a psychiatric ward regimen). Around each a fair amount of replicable and theoretically relevant findings have emerged. Particularly important was the development in the 1960s of the clinically informed interview-based Present State Examination (**PSE**) by Wing *et al.*[10] later amended to deal with a 12-month period,[11] and psychosocial measures such as that of 'expressed emotion'[12] and the Life Events and Difficulties Schedule.[13] Part of the strength of the resulting research has been due to the levels of inter-rater reliability achieved and the ability of the approach to deal with time order. It is often overlooked that even with longitudinal designs it can be important to be able to establish what has happened between interviews.

Life events and building aetiological models

The characteristic features of the aetiological studies that have emerged can be illustrated by those dealing with life events. The significance of findings concerning depression is enhanced by the fact that most studies have produced broadly consistent findings about the role of events.[14–16] Indeed, for some years the challenge has been not so much to establish the presence of an effect, but to learn more about the nature of the events involved and to integrate findings into a more comprehensive aetiological model.[17]

The role of life events in the aetiology of depression

(a) Measurement and meanings

The original version of the Life Events and Difficulties Schedule was developed to study schizophrenic episodes[18] and there has since been a good deal of research dealing with psychotic patients.[19] An early achievement in the study of depression was to make clear that the amount of change in activity as such appears to be irrelevant and that the impact of events results from their meaning.[13] It has also been clear that attention needs to be given to ongoing difficulties that can either be brought about by an event (e.g. the death of husband leading to financial problems) or lead to an event (e.g. a marital difficulty eventually ending in a separation).

In dealing with meaning, two perspectives have proved productive. The first is summed up by the statement that we cannot fully know the meaning of an event until we relate it in some manner to our concerns. One way of conceiving of these is in terms of the impact of a particular event on plans and purposes that stem from role activity caught up in the crisis: how, for example, being turned down for rehousing by a local authority thwarts a woman's wish to move from an overcrowded and damp flat to give her children 'a better start in life'.

A second perspective concerning meaning assumes the likely presence of evolutionary-based response patterns that help to guide us in terms of what to want or to avoid, and that such systems are sensitive to a particular range of stimuli. The attachment system and fear responses are obvious examples.[20, 21] Of course, such responses will be influenced by individual differences of various kinds and by cultural display rules concerning emotions, but there is good reason to believe that such systems are often involved in the development of psychiatric disorders. For example, the central importance in a number of cultures of 'critical comments' of a close relative rather than 'dissatisfaction' in a schizophrenic relapse probably reflects an evolutionary-based sensitivity to emotionally toned criticism interacting with some constitutional predisposition to the disorder.[22]

The Life Events and Difficulties Schedule deals with both kinds of meaning. Blind consensus ratings usually based on four-point scales, made by several members of a team are employed to rule out reporting artifacts using 'edited' accounts supplied by the person who carried out the interview as discussed earlier in relation to the study of appendectomy. General as well as specific kinds of threat are rated in this way. They are contextual in the sense of taking into account a person's likely concerns of relevance for the event insofar as these can be assessed from a person's current circumstances and biography. In making such ratings no account is taken of reported feelings or whether a disorder followed the event. It deals not only with possible bias on the part of raters, but also with the problem that the cognitive processes involved in the appraisal of an event are not necessarily ones a person is willing or able to report.[23] General guidelines for rating severity of threat are given in an extensive manual containing thousands of examples listed in terms of a number of event categories (such as 'demotion at work' and 'unplanned pregnancy'). A similar procedure is followed for ongoing difficulties.

Some findings concerning depression

The first use of the Life Events and Difficulties Schedule to study depressive conditions took place in the early 1970s and involved a patient series seen at the Maudsley Hospital together with a sample of women from the local Camberwell population. A threshold of 'caseness' reflected what an outpatient psychiatrist would accept as a 'case'.[13] This enquiry, and a number made later, have established that the majority of episodes of clinical relevance are preceded by a severely threatening event.[14–16] These at a minimum had to be judged to continue to convey threat for at least a further 10 to 14 days. Nothing emerged to suggest that events with only short-term threat play a role.

Table 2.6.1.1 gives a typical result from a prospective enquiry of 400 women living in Islington in North London with at least one child at home. The table shows that 29 of the 32 onsets in the first

Table 2.6.1.1 Onset of depression within 6 months of a severe event or a severe difficulty among 303 women in Islington

	Percentage onset
Severe event	22 (29/130)
Severe difficulty and no severe event	5 (1/20)
Neither	1 (2/153)
Total	11 (32/303)

follow-up year were preceded by at least one severe event in the prior 6 months with most occurring within a matter of weeks.[8,23] For example, a woman experiencing a second miscarriage after persistent attempts to have her first child would probably have the event rated severe, but a first miscarriage shortly after marriage would most likely be rated upsetting but not severely so.

This finding emerged despite the use of contextual ratings that, as made clear, are based on a limited amount of information and deliberately designed to be approximate and probabilistic. It was also possible to obtain more direct evidence about the relevance of plans and purposes by a measure of emotional commitment to various roles made at the time of the first interview based on how they were talked about. Where a severe event (e.g. a child's delinquency) in the follow-up year 'matched' an area of high emotional commitment (e.g. to motherhood), risk of an onset was considerably increased when compared with a non-matching severe event.[8]

The contextual approach has also been used to take account of more specific aspects of meaning. Severe events preceding an onset of depression generally involve loss, if this is defined broadly not only in terms of loss of a person but loss of a role or a cherished idea—the latter about oneself or someone close.[24] (In contrast, events preceding the onset of anxiety tend to involve 'danger'—the threat of future loss.[25]) However, although loss is typically present it may not be the factor of central aetiological importance. Table 2.6.1.2 illustrates this by the development of a more comprehensive rating scheme—again carried out by the investigator. Four overall types of meaning are considered, covering in all nine categories. The ratings are hierarchical. Where more than one rating is possible the highest on the scale is taken. The first three categories concern possible types of **humiliation**, i.e. the likelihood of the event provoking a sense of being put down or a marked devaluation of self. The first category, for example, covers separating from a partner or a lover where they either took the initiative or the respondent was 'forced' to leave or break off a relationship because of violence or the discovery of infidelity.

Table 2.6.1.2 Onset by type of severe event over 2-year period in the Islington community series

Hierarchical event classification	No. of onsets	Percentage onset rate
(a) All 'humiliation' events	31/102	30
Humiliation: separation	12/34	35
Humiliation: other's delinquency	7/36	19
Humiliation: put down	12/32	38
(b) All 'trapped' alone events (i.e. not (a))	10/29	34
(c) All 'loss' alone events (i.e. not (a) or (b))	14/157	9
Death	7/24	29
Separation: subject initiated	2/18	11
Other key loss	4/58	7
Lesser loss	1/57	2
(d) All 'danger' alone events (i.e. not (a), (b), or (c))	3/89	3
All severe events	58/377	15

Events associated with **entrapment**, the second main type, had to have failed to meet criteria for one of the three humiliation categories. Such events emphasized the fact of being imprisoned in a punishing situation that had gone on for some time. The third type deals with four kinds of **loss** (in the absence of humiliation or entrapment) with the final type, **danger**, involving threat of a future loss.[24]

The table shows whether a particular severe event (or sequence of closely related events) was followed by an onset of depression, taking the event (or sequence) nearest to the onset when there was more than one event within 6 months of onset. Using a 2-year period for the Islington women, it shows that there were large differences in risk by event type. If events involving humiliation are combined with those of entrapment, risk was increased threefold.[24] The relatively low risk of depression associated with loss alone, except following a severe event involving a death, suggests that while the majority of events involve loss, something more than this is usually involved and that the experience of humiliation or entrapment associated with the loss is often critical.

Diagnostic issues

So far I have discussed only studies dealing with depressive onsets in the general population, almost entirely of a 'neurotic' kind. In the Camberwell enquiry of psychiatric patients, while events were rather less frequent before 'melancholic' than before 'neurotic' depression, there was considerable overlap between the two types. This lack of a clear link between the presence of a provoking life event and type of diagnosis had been reported earlier[26] and also in several subsequent studies.[27,28]

A recent study of North London psychiatric depressive patients has thrown possible light on this somewhat unexpected picture. When episode number was taken into account, those patients with both a melancholic/psychotic diagnosis and a prior episode of depression had a much smaller chance of experiencing a severe event before onset.[29] A patient series from Pittsburgh produced consistent findings.[30] These results, if confirmed, may also help to explain inconsistencies in published results since the proportion with a melancholic/psychotic picture and a prior episode is bound to vary by type of treatment centre. It is also of note that this same London study concluded that, despite detailed questioning, as many as one-tenth of *patients* with a 'neurotic' depressive disorder gave no hint of being provoked by social adversity of any kind.[24]

It is of interest that the smaller number with provoking events as episode number rises has also been found to relate to the course of bipolar conditions where there is some evidence for a sensitization or kindling mechanism.[31] Important research continues to emerge about psychosocial risk factors and bipolar conditions[32] and other psychiatric conditions.[33] However, for this brief review I will continue to focus on common depressive disorders.

Course and remission

Life-event research has also thrown some light on the processes involved in remission from depression. Evidence has begun to emerge that these often involve the reverse of the process leading to onset. A 'positive' event or the reduction in the level of a severe difficulty (with or without an event) is commonly found to have been present in the 20-week period before any remission (or marked improvement). However, it is of interest that although the events involved were rated contextually as likely to have given renewed hope about the future, one-third were at the same time judged as severely threatening.[34]

In the Islington general population series somewhat over half of the remissions of episodes that had lasted 20 weeks or more were preceded by such an event. There was no such link with episodes lasting less than this. In the patient series the result was much the same, although the chances of a positive event or difficulty reduction for those on antidepressant medication was somewhat less.[35]

A different approach to the issue of outcome concerns determinants of the length of a particular episode. Here the presence of severe interpersonal difficulty at the point of onset (but no other difficulty) was an important predictor of whether a depressive episode would go on to last for at least 1 year, and this held for a general population and patient series.[36, 37] Ongoing difficulties therefore need to be taken into account in terms both remission and in terms of whether an episode takes a chronic course.

Psychosocial vulnerability

The part played by severe events in depression has proved to be a particularly effective platform for exploring psychosocial vulnerability. The importance of this question was, in fact, illustrated in Table 2.6.1.1 where, despite the fact that the majority of onsets were preceded by a severe event, only one-fifth of those who experienced a severe event developed depression. While, as already discussed, taking account of event type increases this to one-third, it is still necessary to ask why only a minority go on to develop clinically relevant depression following a severe event.

In the Islington series, two background factors present at the time of the first interview proved highly predictive of onset: a negative environmental factor (negative interaction with others in the home and in addition, for single mothers, lack of a close confidant seen fairly often) and a negative psychological factor (negative evaluation of self). (The presence of chronic anxiety or subclinical depressive symptoms has often been included in the latter index, but findings are broadly similar with only negative evaluation of self.)

Their predictive power can be judged by the fact that, while only 23 per cent of the women without depression at the time of first contact had both, three-quarters of onsets in the 12-month follow-up occurred among these women.[38] The result has recently been confirmed in a second prospective enquiry.[39]

In terms of an overall aetiological model, the predictive power of the two indices appears to be largely a result of relating to a greater chance of a severe event occurring and to a lack of effective emotional support from a close confidant once the event has occurred.[40–43] Consideration of social support, of course, links the research, with social science concepts such as 'social integration', 'social bond', and 'social alienation'.[5] The topic is too complex to pursue in any detail in the present brief review. Research has so far underlined the need to recognize that support at one point in time does not necessarily predict what will occur in a subsequent crisis and, indeed, that a significant aspect of a number of severe events is the fact they involve the withdrawal of social support that up to that point had largely been taken for granted.[43, 44] There is also some indication that the ability of a woman to make supportive ties that can be used in a future crisis, and also to use them in such circumstances, is adversely influenced by the early experience of childhood abuse and neglect.[45, 46]

Figure 2.6.1.1, dealing with the Islington series, sums up research on the issue of vulnerability. It takes into account both event type and vulnerability and it shows that a severe event, however threatening, was not enough to provoke depression without the presence of at least one of the two vulnerability risk factors. Subsequent research has confirmed this general picture.

Gender differences

Most research on depression using the Life Events and Difficulties Schedule has dealt with women, although in the original Camberwell enquiry a small series of men gave similar results, as did a subsequent population enquiry.[24] However, recent research has gone further to suggest that the well recognized greater risk of a depressive onset among women may relate to their greater sensitivity to severe events involving their role as mothers. In a study of couples experiencing a severe event in common, women were twice as likely to develop a depressive episode following events involving procreation, children, and housing; for other events, risk was much the same.[47] While this study requires replication, it does support the increasingly held opinion that the large gender difference in the experience of 'neurotic' depression is likely to have an essentially psychosocial explanation.[48]

A lifespan perspective

So far only current environmental circumstances have been taken into account. The Camberwell research also identified loss of a mother before the age of 11 as a risk factor.[43] While there has been a good deal of controversy about this finding,[49, 50] two further population studies produced equally clear evidence with the added suggestion that such a loss between 11 and 17 may also increase

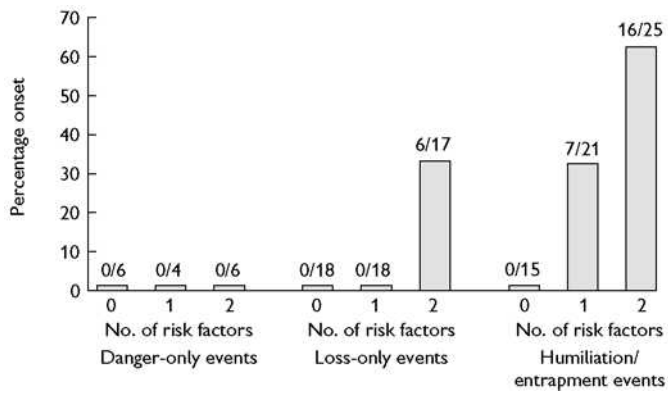


Fig. 2.6.1.1 Onset rate among 130 Islington women with a severe event by type of event and number of background risk factors.

risk.[51,52] The mode of the impact of such an early loss of mother for women is undoubtedly complex, but several studies have now established the critical importance of untoward experiences during childhood *after* the loss and have downplayed the role of loss as such.[51] More important than the loss of the mother itself was the quality of replacement parental care (in terms of an index of parental indifference or lax control). Risk of adult depression was doubled for those positive on the index.In order to understand its link with later adult depression it has been necessary to trace the history of a woman from the loss itself to such depression in a way that makes it possible to gain some sense of a life trajectory. Certain early experiences are particularly associated with the chance of experiencing the kind of adult factors already discussed.[53]

A factor playing a critical mediating role was the experience of a premarital pregnancy, and, like the care index itself, this was found to be associated with the subsequent experience of severe events. What seemed to be crucial about these premarital pregnancies was that they often trapped women in relationships which they might well not otherwise have chosen and which became a source of ongoing problems—such as severe housing and financial difficulties consequent upon a couple starting a family too young to have built up adequate savings, or marital difficulties with undependable partners. These women also emerged as less upwardly mobile, in terms of social class, than their peers without such pregnancies. In interpreting this complex of experiences, a conveyor belt of adversities was outlined, on which some women often appeared to move inexorably from one crisis to another, starting with lack of care in childhood and passing via premarital pregnancy to current working-class status, lack of social support and high rates of severe events.[53]

Although it was often hard to see from the women's accounts of their lives how they could have left this conveyor belt once their childhood had located them on it, a more personal element is likely to have played a role in many instances. Here the work of Quinton and Rutter[54–57] has been particularly significant in developing a lifespan perspective in this regard, and over this issue the results of the two research programmes have largely complemented each other.

The kind of early adverse experience just outlined was subsequently incorporated into a broader index of childhood adversity that included severe physical abuse within the family and also severe sexual abuse in any setting.[58] This index not only relates to a doubling of the risk of a depressive onset in adult life, but also to a number of other adverse outcomes, for example to the quadrupling of risk of an episode taking a chronic course[59] and also to the risk of a depressive condition comorbid with anxiety disorder defined by DSM-IIIR criteria (excluding simple phobias and mild agoraphobia).[60] It can be added that the retrospectives measures of early maltreatment used in these enquiries appear to be sufficiently free of bias to enable the findings to be taken seriously.[61, 62]

A population perspective

It is useful also to consider 'neurotic' conditions that form the bulk of depressive disorders, even in patient series, in terms of a population perspective. Figure 2.6.1.2 summarizes the findings of six population studies of women aged between 18 and 65 carried out in a comparable manner, using the same semi-structured interview-based measures as in the Islington survey, including the Present State Examination. The bottom half of Fig. 2.6.1.2 shows the rate of depressive 'caseness' in a 12-month period. Between the two extreme populations there is a 10-fold difference—3 per cent in a rural Basque-speaking population in Spain[63] and 30 per cent in a black urban population in Zimbabwe.[6, 64] In addition these rates are fairly closely paralleled by differences in the experience of severe events particularly likely to provoke depression (see top half of Fig. 2.6.1.2).

One of the implications of these results becomes clear in the context of a behavioural genetic perspective. A key point about the concept of heritability is that it is specific to a particular population and based on consideration of individual discrepancies from a population mean. While there may well be a genetic contribution in each of the six populations, this would reflect individual variability in risk within each population. Even if large, the genetic contribution would only be likely to be of relevance for explaining the substantial population differences in rates of depression if there were large population differences in the frequency of relevant genes. On present evidence the most plausible interpretation of such differences is that they are the result of psychosocial factors.[65]

There is, however, no inherent conflict between the two perspectives—they refer to different ways of looking at the 'variance' of a condition.[66] The general point is that the study of individual variability within particular populations cannot rule out the possibility that the mean level of disorder is largely under environmental control; that it can be increased or decreased markedly by external changes quite uninfluenced by the genetic make-up of a population.

A population perspective is also concerned with variability in rates of disorder within populations in terms of social categories such as socio-economic status. Thus, the survey in Camberwell in South London in the early 1970s found that, while the rate of severe events was related to social class position, this explained comparatively little of the large class difference in prevalence of depression; of greater importance were background vulnerability factors such as an unsupportive marriage.[13] However, the picture has recently become more complex with the finding that severe events involving

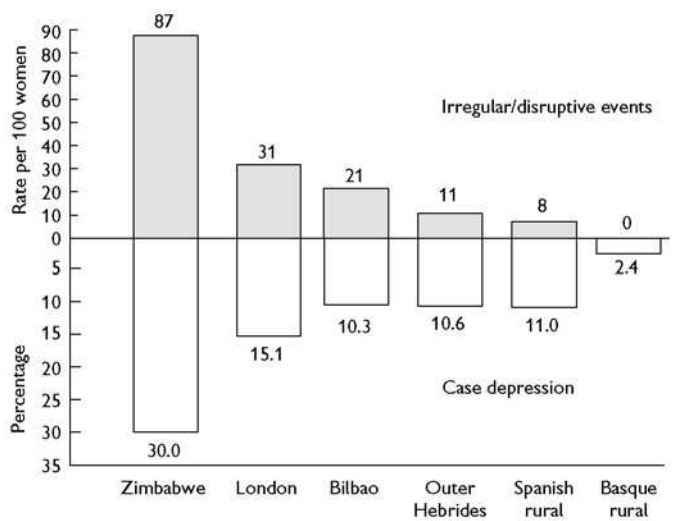


Fig. 2.6.1.2 Yearly rate of irregular or disruptive severe events per 100 women in six populations and prevalence of caseness of depression in year.

humiliation and entrapment are not only especially depressogenic but particularly common in high-risk populations (such as Harare) and within populations in high-risk subgroups (such as working-class women in London). In Islington such events were common among single mothers, a social category that has expanded dramatically in most western populations in recent years and among whom there is a high risk of depression.[67]

Final comments

There are two ways of looking at the findings that have been reviewed. First, that the study of life events has been an effective way of opening up wider issues concerning psychosocial factors and the aetiology of depression. This has been possible because, given the presence of a substantial causal link, a platform is provided for the study of a whole range of other experiences. As research has progressed it has pushed back in time to consider the aetiological role of early experiences of neglect and abuse which can have event-like characteristics. In more general terms the study of events has led to consideration of issues of vulnerability and protection, event production, chronicity, and course of particular episodes, and also issues not covered in this account such as coping and social support. It can be added that the work has also led to a good deal of research on other psychiatric conditions that have not been reviewed.[9,18,68,69]

The second contribution of the findings reviewed involves the stimulation of cross-disciplinary research concerning the depression—event link itself. This is a complex issue because life events correlate with factors ranging from genetic/personality[70] to macrolevel/societal.[65] The growing literature on the role of the serotonin transporter gene in depression is of particular interest as evidence has emerged for an important interplay with life events. A recent study has documented an interaction of the *s* allele variant of 5-HTTLPR polymorphism and life events with young adults. It also reported that childhood maltreatment predicted depression only among those with the *s* allele.[71] A somewhat similar finding with adult twins concerning life events has also been reported.[72] It is intriguing that there is a possibility of an important developmental contribution: that the critical gene-environmental interplay occurs early in life perhaps involving experiences such as parental maltreatment.[73] However, other studies have had more mixed findings with some showing no evidence of an interactive effect.[74,75] It is clearly too soon to review findings with any confidence—for example, some of the mix findings may well relate to the general failure to utilize the kind of detailed investigator-based measure of life events reviewed earlier and a failure to restrict consideration to events occurring not long before onset. In most studies the environmental measures fall a good way short of current best practice, although optimizing such assessments is essential for the detection of gene x environment interaction.[76,77] Nonetheless the research is an exciting development and it seems possible that well-established findings will emerge. But here my earlier comments about the likely critical importance of psychosocial factors in explaining differences in *rates* of disorder across populations should be kept in mind, although here there is some evidence that the genotype for 5-HTTLPR polymorphism may differ across populations.[78] It will also be important to take into account diagnostic issues. For example, it would be interesting to explore the fact, mentioned earlier, that it has been possible to isolate a small group of 'endogenous' neurotic depressive episodes.[24] Also it should be borne in mind that while this short review has largely restricted itself to work on depression, the life-event approach has been successfully employed with a number of other psychiatric and physical conditions and the possibility for collaborative research may well be much wider. For example, the *s* variant of 5-HTTLPR polymorphism also appears to be relevant for other psychiatric conditions.[74] There are other relevant biological considerations. The findings concerning life events involving humiliation and entrapment may need to be viewed from an evolutionary perspective, that is, that in some way a response pattern closely linked to issues surrounding defeat and exclusion, which has developed in group-living animals, may have functioned to promote survival.[78,79] It is possible that clinically relevant depressive conditions are often a complication of essentially non-pathological submission and appeasement responses to defeat in group-living mammals. Therefore, the high rates of clinically relevant depression that appear to be possible in some populations may well be a result of the more highly developed cognitive developments of *Homo sapiens* together with the event-creating potential of many societies experiencing periods of marked social change due to factors such as war, industrialization, technological development, urbanization, changing sexual mores.[80]

Further information

Brown, G.W. and Harris, T.O. (1989). *Life events and illness*. Guilford Press, New York.

Goldberg, D. and Goodyer, I. (2005). *The origins and course of common mental disorders*. Routledge, London and New York.

Monroe, S.M. and Hadjiyannakis, K. (2002). The social environment and depression: focusing on severe events. In *Handbook on depression* (eds. I.H. Gotlib and C.L. Hammen), pp. 314–40. Guilford Press, New York.

References

1. Mechanic, D. (1978). *Medical sociology* (2nd edn). Free Press, New York.
2. Tomkins, S.S. (1979). Script theory: differential magnification of affects. In *Nebraska symposium on motivations*, Vol. 26 (eds. H.E. Howe and R.A. Dienstbier), pp. 201–36. University of Nebraska Press, Lincoln, NE.
3. Weber, M. (1964). *The theory of social and economic organisation* (ed. and trans. T. Parsons). Collier-Macmillan, London.
4. Jaspers, K. (1962). *General psychopathology* (trans. J. Hoenig and M.W. Hamilton). Manchester University Press, Manchester.
5. Scheff, T.J. (1990). *Microsociology: discourse, emotion and social structure*. University of Chicago, Chicago.
6. Broadhead, J.C. and Abas, M.A. (1998). Life events, difficulties and depression among women in an urban setting in Zimbabwe. *Psychological Medicine*, **28**, 29–38.
7. Brown, G.W. (1985). The discovery of 'Expressed Emotion': induction or deduction? In *Expressed emotion in families: its significance for mental illness* (eds. J. Leff and C. Vaughn), pp. 7–27. Guilford Press, New York.
8. Brown, G.W., Bifulco, A., and Harris, T.O. (1987). Life events, vulnerability and onset of depression: some refinements. *The British Journal of Psychiatry*, **150**, 30–42.
9. Creed, F. (1981). Life events and appendectomy. *Lancet*, **i**, 1381–5.
10. Wing, J.K., Cooper, J.E., and Sartorious, N. (1974). *The measurement and classification of psychiatric symptoms: an instruction for the Present State Examination and CATEGO Programme*. Cambridge University Press, Cambridge.

11. Finlay-Jones, R., Brown, G.W., Duncan-Jones, P., *et al.* (1980). Depression and anxiety in the community. *Psychological Medicine*, **10**, 445–54.
12. Brown, G.W., Birley, J.L.T., and Wing, J.K. (1972). The influence of family life on the course of schizophrenia. *Journal of Health and Social Behavior*, **9**, 203–14.
13. Brown, G.W. and Harris, T.O. (1978). *Social origins of depression: a study of psychiatric disorder in women*. Tavistock Publications, London.
14. Jenaway, A. and Paykel, E.S. (1997). Life events and depression. In *Depression: neurobiological psycho-pathological and therapeutic advances* (eds. A. Honig and H.M. van Praag), pp. 279–96. Wiley, Chichester.
15. Bebbington, P., Hurry, J., Tennant, C., *et al.* (1981). Epidemiology of mental disorders in Camberwell. *Psychological Medicine*, **11**, 561–79.
16. Surtees, P.G., Miller, P.M., Ingham, J.G., *et al.* (1986). Life events and the onset of affective disorders: a longitudinal general population study. *Journal of Affective Disorders*, **10**, 37–50.
17. Bebbington, P., Wilkins, S., Sham, P.I., *et al.* (1996). Life events before psychotic episodes: do clinical and social variables affect the relationship? *Social Psychiatry and Psychiatric Epidemiology*, **31**, 122–8.
18. Brown, G.W. and Birley, J.L.T. (1968). Crises and life changes and the onset of schizophrenia. *Journal of Health and Social Behaviour*, **9**, 203–14.
19. Ventura, J., Nuechterlein, K.H., Subotnik, K.L., *et al.* (2000). Life events can trigger depressive exacerbations in early course of schizophrenis. *Journal of Abnormal Psychology*, **109**, 139–44.
20. Gilbert, P. (1989). *Human nature and suffering*. Erlbaum, Hove.
21. Ohman, A. (1986). Face the beast and fear the face: animal social fears as prototypes for evolutionary analysis of emotion. *Psychophysiology*, **23**, 123–45.
22. Martins, C.M, de Lemos, A.I., and Bebbington, P.E. (1992). A Portuguese/Brazilian study of expressed emotion. *Social Psychiatry and Psychiatric Epidemiology*, **27**, 22–7.
23. Brown, G.W. (1989). Depression. In *Life events and illness* (eds. G.W. Brown and T.O. Harris), pp. 49–94. Guilford Press, New York.
24. Brown, G.W., Harris, T.O., and Hepworth, C. (1995). Loss, humiliation and entrapment among women developing depression. A patient and non-patient comparison. *Psychological Medicine*, **25**, 7–21.
25. Finlay-Jones, R. and Brown, G.W. (1981). Types of stressful life event and the onset of anxiety and depressive disorders. *Psychological Medicine*, **11**, 803–15.
26. Paykel, E.S., Prusoff, B.A., and Klerman, G.L. (1971). The endogenous-neurotic continuum in depression, rater independence and factor distributions. *Journal of Psychiatric Research*, **8**, 73–90.
27. Katschnig, H., Pakesch, G., and Egger-Zeidner, E. (1986). Life stress and depressive subtypes: a review of present diagnostic criteria and recent research results. In *Life events and psychiatric disorders: controversial issues* (ed. H. Katschnig), pp. 201–45. Cambridge University Press, Cambridge.
28. Bebbington, P. and McGuffin, P. (1989). Interactive models of depression. In *Depression, an integrative approach* (eds. E. Paykel and K. Herbst), pp. 65–80. Heinemann Medical Books, London.
29. Brown, G.W., Harris, T.O., and Hepworth, C. (1994). Life events and endogenous depression: a puzzle re-examined. *Archives of General Psychiatry*, **51**, 525–34.
30. Frank, E., Anderson, B., Reynolds, C.F., *et al.* (1994). Life events and research diagnostic criteria endogenous subtype. *Archives of General Psychiatry*, **51**, 519–24.
31. Post, R.M., Rubinow, D.R., and Ballenger, J.C. (1986). Conditioning and sensitisation in the longitudinal course of affective illness. *The British Journal of Psychiatry*, **149**, 191–201.
32. Kim, E.Y., Miklowitz, D.J., Biuckians, A., *et al.* (2007). Life stress and the course of early-onset bipolar disorder. *Journal of Affective Disorder*, **99**, 37–44.
33. Hatcher, S. and House, A. (2003). Life events, difficulties and dilemmas in the onset of chronic fatigue syndrome: a case-control study. *Psychological Medicine*, **33**, 1185–92.
34. Brown, G.W., Lemyre, L., and Bifulco, A. (1992). Social factors and recovery from anxiety and depressive disorders: a test of the specificity hypothesis. *The British Journal of Psychiatry*, **161**, 44–54.
35. Brown, G.W. (1993). Life events and affective disorder: replications and limitations. *Psychosomatic Medicine*, **55**, 248–59.
36. Brown, G.W. and Moran, P. (1994). Clinical and psychosocial origins of chronic depressive episodes. I. A community survey. *The British Journal of Psychiatry*, **165**, 447–56.
37. Brown, G.W., Harris, T.O., Hepworth, C., *et al.* (1994). Clinical and psychosocial origins of chronic depressive episodes. II. A patient enquiry. *The British Journal of Psychiatry*, **165**, 457–65.
38. Brown, G.W., Bifulco, A., and Andrews, B. (1990). Self-esteem and depression. 3. Aetiological issues. *Social Psychiatry and Psychiatric Epidemiology*, **25**, 235–43.
39. Bifulco, A., Brown, G.W., Moran, P., *et al.* (1998). Predicting depression in women. The role of past and present vulnerability. *Psychological Medicine*, **28**, 39–50.
40. Harris, T.O. (1992). Some reflections of the process of social support: and nature of unsupportive behaviors. In *The meaning and measurement of social support* (eds. H.O.F. Veiel and U. Baumann), pp. 171–89. Hemisphere, Washington, DC.
41. Edwards, A.C., Nazroo, J., and Brown, G.W. (1998). Gender difference in marital support following a shared life event. *Social Science & Medicine*, **46**, 1077–85.
42. Veiel, H.O.F. and Baumann, U. (eds.) (1992). *The meaning and measurement of social support*. Hemisphere, Washington, DC.
43. Brown, G.W., Andrews, B., Harris, T.O., *et al.* (1986). Social support, self-esteem and depression. *Psychological Medicine*, **16**, 813–31.
44. Andrews, B. and Brown, G.W. (1998). Marital violence in the community. A biographical approach. *The British Journal of Psychiatry*, **153**, 305–12.
45. Harris, T.O., Brown, G.W., and Bifulco, A. (1986). Loss of parent in childhood and adult psychiatric disorder. The role of lack of adequate parental care. *Psychological Medicine*, **16**, 641–59.
46. Harris, T.O., Brown, G.W., and Bifulco, A. (1990). Loss of parent in childhood and adult psychiatric disorder: a tentative overall model. *Development and Psychopathology*, **2**, 311–28.
47. Nazroo, J.Y., Edwards, A.C., and Brown, G.W. (1997). Gender differences in the onset of depression: a study of couples. *Psychological Medicine*, **27**, 9–19.
48. Bebbington, P. (1996). The origins of sex differences in depressive disorder: bridging the gap. *International Review of Psychiatry*, **8**, 295–332.
49. Tennant, C., Bebbington, P., and Hurry, J. (1980). Parental death in childhood and risk of adult depressive disorders; a review. *Psychological Medicine*, **10**, 289–99.
50. Harris, T.O. and Brown, G.W. (1985). Interpreting data from aetiological studies: pitfalls and ambiguities. *The British Journal of Psychiatry*, **147**, 5–15.
51. Harris, T.O., Brown, G.W., and Bifulco, A. (1987). Loss of parent in childhood and adult psychiatric disorder: the role of social class position and premarital pregnancy. *Psychological Medicine*, **17**, 163–83.
52. Bifulco, A., Brown, G.W., and Harris, T.O. (1987). Childhood loss of parent, lack of adequate parental care and adult depression: a replication. *Journal of Affective Disorders*, **12**, 115–28.
53. Harris, T.O. (1988). Psycho-social vulnerability to depression: the biographical perspective of the Bedford college studies. In *Handbook of social psychiatry* (eds. A.S. Henderson and G.D. Burrows), pp. 55–71. Elsevier, Amsterdam.

54. Quinton, D. and Rutter, M. (1984). Parents with children in care. 1. Current circumstances and parenting skills. *Journal of Child Psychology and Psychiatry*, **25**, 211–29.
55. Quinton, D. and Rutter, M. (1984). Parents with children in care. 2. Intergenerational continuities. *Journal of Child Psychology and Psychiatry*, **25**, 231–50.
56. Pawlby, S.J., Mills, A., and Quinton, D. (1997). Vulnerable adolescent girls: opposite-sex relationships. *Journal of Child Psychology*, **38**, 909–20.
57. Quinton, D., Pickle, A., Maughan, B., *et al.* (1993). Partners, peers and pathways: assertive pairing and continuities in conduct disorder. *Development and Psychopathology*, **5**, 763–83.
58. Bifulco, A., Brown, G.W., and Harris, T.O. (1994). Childhood Experience of Care and Abuse (CECA): a retrospective interview measure. *Child Psychology and Psychiatry*, **35**, 1419–35.
59. Brown, G.W., Craig, T.K.J., Harris, T.O., *et al.* (2007). Development of a retrospective interview measure of parental maltreatment using the Childhood Experience of Care and Abuse (CECA) instrument—a life-course study of adult chronic depression—1. *Journal of Affective Disorders*, **103**, 205–15.
60. Brown, G.W. and Harris, T.O. (1993). Aetiology of anxiety and depressive disorders in an inner-city population. 1. Early adversity. *Psychological Medicine*, **23**, 143–54.
61. Brown, G.W. (2006). Childhood maltreatment and adult psychopathology: some measurement options. In *Relational processes and DSM-V: Neuroscience, Assessment, Prevention and Intervention* (eds. S.R.H Beach, M. Wamboldt, N. Kaslow, *et al.*), pp. 107–22. American Psychiatric Association, Washington, DC.
62. Brown, G.W., Craig, T.K.J., Harris, T.O., *et al.* (2007). Validity of retrospective measures of early maltreatment and depressive episodes using CECA (Childhood Experience of Care and Abuse)—a life-course study of adult chronic depression—2. *Journal of Affective Disorders*, **103**, 217–24.
63. Gaminde, I., Uria, M., Padro, D., *et al.* (1993). Depression in three populations in the Basque country—a comparison with Britain. *Social Psychiatry and Psychiatric Epidemiology*, **28**, 243–51.
64. Abas, M.A. and Broadhead, J.C. (1997). Depression and anxiety among women in an urban setting in Zimbabwe. *Psychological Medicine*, **27**, 59–71.
65. Brown, G.W. (1998). Genetic and population perspectives in life events and depression. *Social Psychiatry and Psychiatric Epidemiology*, **33**, 363–72.
66. Weizmann, F. (1971). Correlational statistics and the nature-nurture problem. *Science*, **171**, 589.
67. Brown, W. and Moran, P. (1997). Single mothers, poverty and depression. *Psychological Medicine*, **27**, 21–33.
68. Craig, T.K.J., Drake, H., Mills, K., *et al.* (1994). The South London somatisation study. II. Influence of stressful life events, and secondary gain. *The British Journal of Psychiatry*, **165**, 248–58.
69. Harris, T.O. (1989). Disorders of menstruation. *Life events and illness* (eds. G.W. Brown and T.O. Harris), pp. 261–94. Guilford Press, New York.
70. Owens, M.J. and McGuffin, P. (1997). Genetics and psychiatry. *The British Journal of Psychiatry*, **171**, 201–2.
71. Caspi, A., Sugden, K., Moffitt, T.E., *et al.* (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*, **301**, 386–9.
72. Kendler, K.S., Kuhn, J.W., Vittum, J., *et al.* (2005). The interaction of stressful life events and a serotonin transporter polymorphism in the prediction of episodes of major depression: a replication. *Archives of General Psychiatry*, **62**, 529–35.
73. Sibille, E. and Lewis, D.A. (2006). SERT-ainly involved in depression, but when. *The American Journal of Psychiatry*, **163**, 8–11.
74. O'Hara, R., Schroder, C.M., Mahadevan, R., *et al.* (2007). Serotonin transporter polymorphism, memory and hippocampal volume in the elderly: association and interaction with cortisol. *Molecular Psychiatry*, **12**, 544–55.
75. Surtees, P.G., Wainwright, N.W.J., Willis-Owen, S.A.G., *et al.* (2006). Social adversity, the serotonin transporter (5-HTTLPR) polymorphism and major depressive disorder. *Biological Psychiatry*, **59**, 224–9.
76. Kaufman, J. (2007). Failure to replicate gene-environment interactions in psychopathology: a reply. *Biological Psychiatry*, **62**, 545.
77. Nestler, E.J., Alreja, M., and Aghajanian, G.K. (1999). Molecular control of locus coeruleus neurotransmission. *Biological Psychiatry*, **46**, 1131–9.
78. Williams, R.B., Marchuk, D.A., Gadde, K.M., *et al.* (2003). Sertonin-related gene polymorphisms and central nervous system serotonin function. *Neuropsychopharmacology*, **28**, 533–41.
79. Gilbert, P. (2006). Evolution and depression: issues and implications. *Psychological Medicine*, **36**, 287–97.
80. Brown, G.W. (2002). Social roles, context and evolution in the origins of depression. *Journal of Health and Social Behavior*, **43**, 255–76.

2.6.2 Social and cultural anthropology: salience for psychiatry

Arthur Kleinman

Social and cultural anthropology

One of the social sciences (together with history, economics, political science, sociology, and social psychology), social and cultural anthropology is principally concerned with the study of society, in almost all of its aspects. Together with linguistics, archaeology, and biological anthropology, social and cultural anthropology formed the classic (and now considered overly ambitious) four-field base of anthropology, the science of man. Yet still, in many universities, anthropology departments bridge the traditional divisions of the humanities, social sciences, and natural sciences. From the outset, anthropologists defined their subject in holistic terms meant to contextualize women and men in a nested hierarchy of influential environments that ran from the human body to the social body, and that assumed that these levels were related to each other, so that individual and collective processes (biological, psychological, social relational, and cultural) intersected in some way. Social and cultural anthropology, in particular, took as its subject matter studies of communities, ranging from small-scale preliterate groups to neighbourhoods or institutions in megacities. Comparison of different societies, or different structures and processes in those societies, is still seen as a defining approach, as is the analysis of cultural symbol systems (from languages to aesthetics), history of kinship, and other systems of social relationship, as well as research on large-scale social changes such as our era's globalization, ethnonationalism, and resurgence of religious fundamentalism.

Anthropology's chief research methodology is ethnography, the close study of a local world—a village, an urban neighbourhood, an institution, a network. Ethnography privileges local language, conceptual categories, values, and practices. Its procedure is to begin with local definitions and perceptions of reality (sometimes

called 'emics', from phonemics), and only when these experience—near patterns are understood in a particular context of everyday life (with the larger political, economic, and cultural forces that influence it) are comparisons made with other local worlds in the framework of experience-distant scientific definitions of reality (referred to as 'etics', from phonetics). Knowledge is generated by participant observation, informal interviews, and the use of more formal procedures from structured interviews to questionnaires. Cross-cultural comparison is another core mode of knowledge production. Both ethnography and cross-cultural comparisons draw on empirical data to engage larger questions in social theory, which itself is constantly being reorganized in this dialectical engagement.

In this century, social and cultural anthropology's division of labour has spun off at least two subfields that are of particular relevance for psychiatry: psychological anthropology and medical anthropology.

Psychological anthropology

This subfield grew out of the culture and personality school (ca. 1930–1950), when psychoanalysts and anthropologists sought to collaborate to understand how mental processes differed or were similar across greatly different societies. Margaret Mead, Ruth Benedict, and Irving Hallowell are those anthropologists most often associated with this school. Although, most anthropologists became critical of the basis of the field in psychoanalysis and a correlation of individuals with entire cultures, a small group of social and cultural anthropologists continue, none the less, to pursue this direction, and over time they have developed broader ties with psychology, as can be readily seen by their leading research interests in cognition, lifecycle development, and ethnopsychological categories. Anthropologists working in this tradition have studied self-concepts and self-images, emotion terms, interpersonal processes and their relation to personhood, as well as experiences of childhood, child rearing, adolescence, midlife, and ageing. Psychological anthropology has been influential in recent years in psychology, where a sister subdiscipline called cultural psychology has started up in close connection to it.

Medical anthropology

Physicians were among the founders of anthropology, and some, like the British polymath W.H.R. Rivers, combined medicine and anthropology. Another source of medical anthropology was social medicine and public health; indeed the great German pathologist and social medicine advocate, Rudolph Virchow, was one of the first to use the term 'medical anthropology'. Thus, medical anthropology's early roots were applied. After the Second World War, the field took off as anthropologists developed an interest in the theoretical and empirical aspects of non-Western medical traditions, religious healing and its relation to medicine, and increasingly in experiences of suffering. In more recent years, medical anthropologists, of whom there are several thousand worldwide, have developed special interests in infectious diseases (especially diarrhoeal disease, malaria, tuberculosis, and AIDS), female reproductive lifecycle problems, the health problems of children and the aged, substance abuse, cancer, diabetes, disabilities, medical ethics, and the economic and social transformation of health care. One of the earliest and abiding interests has been in psychiatric diagnosis, disorders, and treatments. This subfield of social and cultural anthropology has many ongoing relationships to cultural psychiatry (see Chapter 2.6.1) and has been active in recent years in the effort to introduce mental health concerns into international health (see Chapter 1.3.2 and 7.3). Indeed, the cultural sections of the DSM-IV were contributed by a taskforce that included both medical anthropologists and cultural psychiatrists, in equal numbers.

Major contributions of anthropology to psychiatry

Cultural critique of biomedicine

One of the crucial contributions of anthropology is theoretical, namely a critique of the theoretical biases inherent in psychiatric science and clinical practice. This may seem self-evident because unlike any other branch of medicine, there is no blood test, biopsy, or radiograph to diagnose psychiatric disorder (leaving aside Alzheimer's disease, which is after all a neurological disorder). That means that psychiatric diagnosis is based on the establishment of symptom and syndromal criteria, which are based in turn in language, lay categories, and everyday social experience. Cultural bias can enter this process in several ways. Anthropologists have shown that this can happen when diagnostic criteria that have been developed in one society are applied to another where they lack validity. This is called a 'category fallacy', a term introduced by Kleinman.[1] Classic examples include trance and possession states in many non-Western societies, which are frequently normative and normal experiences. Failure to recognize this phenomenon, and therefore the diagnosis of persons in religious trance as psychotic, creates a category fallacy in the application of the diagnostic criteria of psychosis to normal people. The cultural critique has been applied to personality disorders as well, because this category of disorder models self-processes on a Euro-American, middle-class, and usually male behavioural type and lifestyle. Anthropologists argue for a much more flexible and interactive understanding of subjectivity that changes in basic ways in response to different social circumstances.

In the 1990s, cultural critique has been important in highlighting the influence of institutional racism in psychiatric diagnosis, referral, and treatment. Leading examples are the overdiagnosis of African-Americans and African-Caribbean Britons with schizophrenia, the tendency to perceive them as more dangerous and less amenable to psychotherapy, and differences in the way their discharge and aftercare are organized. Anthropologists have examined how racism is unwittingly built into psychiatric categories and infiltrates the model cases used to illustrate diagnostic criteria, and also the way that psychiatrists are trained to replicate such patterns in the practice of triage.

Cultural critique, informed by the cross-cultural and international data, is the basis for anthropologists' doubts about the validity of many of the psychiatric conditions detailed in DSM-IV and ICD-10. The ethnographic database strongly suggests that, apart from brain tumours and infections, Alzheimer's disease, metabolic encephalopathy, substance abuse, and other well-documented brain-based disorders such as certain sleep disorders, only six psychiatric syndromes of adults can be found cross-culturally; i.e. only these have stability as syndromes outside the cultural mainstream of Euro-American societies. The conditions are schizophrenia, brief

reactive psychoses, major depression, bipolar disease, a range of anxiety disorders from panic states through phobias to obsessive–compulsive disorder, and trauma, whether understood as PTSD or in other categories. Most of the other hundreds of conditions described in DSM-IV, for example, are culture bound to Euro-America.

Related to these contributions of cultural critique, anthropologists have also contributed to the development of culturally informed diagnostic criteria, questionnaires, structured interviews, and guidelines for working with translators. Globalizing and indigenizing psychiatric approaches is an even more general emphasis in anthropology. Anthropology contains numerous concepts and methods that might be tried out, but relatively few have been experimented with or adopted. Besides those described below, several examples of the concepts, methods, and findings from anthropology that await trial in psychiatry are listed in Table 2.6.2.1.

Local moral worlds: interpersonal basis of illness experience

Ethnographies—hundreds of them, including many on psychiatric topics—demonstrate, with great consistency, that most people and most patients are not isolated individuals but rather live their lives as active members of local worlds. By local worlds, ethnographers mean villages, neighbourhoods, networks, and families, as well as particular social institutions, including hospitals and outpatient systems. These local worlds are differentiated by class, ethnicity, gender, age cohort, political faction, religious ties, and still other social differences. In any given local world crucial things are at stake that orient the attention and actions of participants. What is at stake may be shared (status, resources, survival, transcendence), but it also can be as distinctive as the different meanings of ultimacy that make religions distinguishable from each other.

Illness experience and experiences of treatment are as much caught up these stakes as experiences of normality. Thus, for anthropology how a person's illness is encountered, coped with, understood, and lived is crucial for understanding the illness and the treatment. Therefore, anthropologists write about the social course of illness: meaning that local worlds shape the course of illness so thoroughly that the same disease process (diabetes, AIDS, depression, schizophrenia) can take different trajectories. When sick people go for treatment, who they first seek out, whether they comply with the therapeutic regime, how they assess their experience of treatment—all are in one way or another influenced by what is most at stake for communities, families, networks, and individuals. The anthropological contribution here is to highlight the processes through which individuals relate to collectives. Thus, Estroff[2] shows that collective and individual definitions of identity affect how schizophrenic patients live their schizophrenia as an illness identity, which in turn affects their careers as patients and their experiences in other domains (family, workplace, etc.).

Table 2.6.2.1 Anthropological concepts, methods, and findings that await trial in psychiatry

Ethnography Ethnography as a research strategy in clinical and epidemiological research. For example, as a means of studying the clustering of psychiatric conditions with social problems. Ethnography also has uses in evaluation research, in generating categories and questions in epidemiological studies, and in sociosomatic research. It is also a means of training researchers.
Ethnographic database Ethnographic database as a routine source of knowledge about communities (foreign and domestic) for clinicians, mental health planners, and researchers.
Cross-cultural comparisons Cross-cultural comparisons as a routine form of knowledge production. For example, cross-cultural comparisons of psychotherapy might help (a) to clarify what is common among religious, moral, and medical healing, (b) to determine how culture influences psychotherapeutic practice, and (c) to develop psychotherapeutic techniques for use with patients and families from different minority, ethnic, and international populations.
Social theory Systematic reading of social theory as a source of hypotheses for research in social psychiatry and as a means of preparing clinicians to practise community psychiatry. Concepts such as social and symbolic capital, globalization, marginalization. ethnic identity, and institutional racism, among many others, can be used to frame psychiatric research and clinical intervention.

Practical clinical relevance

Immigration processes have so altered national demographic patterns that most nation states today have plural populations representing distinctive ethnic backgrounds. In 1900, the population of the United States, for example, included only 13 per cent categorized as ethnic minority members. By 1990 that figure was greater than 25 per cent. The percentage is projected to be one-third in 2010, and by mid-century to reach an astonishing 50 per cent. In California, the largest American state, non-Hispanic white Americans are already in a minority.

Ethnic background has been shown empirically to influence epidemiological rates of disease, patterns of access to health care, help-seeking, and patient–doctor interactions, often with negative outcomes such as delayed treatment, misdiagnosis, noncompliance, and treatment failure. Taking ethnicity into account in the provision of services means a variety of things, such as making translators available, putting up signs in several languages, holding clinics at times when working-class patients can attend, and paying attention to differences in cultural meanings and practices. The now popular idea of providing culturally informed and sensitive care is premised on anthropological concepts and methods. Several of these have been elaborated in the literature.

1 The distinctions between illness and disease: for medical anthropologists illness is the patient's experience of symptoms in the context of family, work, and community; disease is the practitioner's model of the pathological process. Help-seeking is usually orientated around the illness experience with respect to what is most at stake for the patient and significant others. Care can founder when the patient's primary concerns with the illness experience conflicts with or is entirely different from the physician's focus on disease. Thus, many patients with chronic pain experience interrogation of the disease process by the sceptical physician as delegitimizing their illness experience. This leads to high rates of dissatisfaction with care among this group of patients. When patients and families are from ethnic minority backgrounds, differences in cultural meanings and practices intensify conflicts between patient and physician models.

2 Medical anthropology sponsors a method to reduce this explanatory gap and thereby to improve clinical relationships. Called the explanatory models' methodology, it involves three steps.[3]

(a) Elicitation of patient and family explanatory models of the illness experience and treatment, which can be accomplished by asking the following questions:

- What do you call your problem?
- What do you think caused it?
- Why did it start when it did?
- What course will it take from here on?
- How does it work in your body?
- What do you most fear about the illness?
- What kind of treatment do you desire for this illness?
- What do you most fear about the treatment?

(b) Presentation of the clinician's explanatory model of the disease process.

(c) Negotiation of a mutually acceptable understanding of the clinical problem across patient, family, and physician models.

Closely related to this technique is the development of a mini-ethnography. This is a brief description, based on interviewing the key parties about the impact of family and work context on the illness experience and vice versa. The mini-ethnography and the explanatory models' elicitation generally give rise to patient stories of the illness experience. The influence of cultural categories, values and practices on the illness and treatment can be assessed from this standpoint.[4]

Revised cultural formulation

Appendix I to DSM-IV contains an outline of how psychiatric cases can be culturally formulated. This has recently been updated[6] and includes:

- the ethnic identity of the individual
- what is at stake as patients and their loved ones face an episode of illness
- the illness narrative
- psychosocial stresses and social supports
- cultural elements of the relationship between the individual and the clinician
- an examination of the efficacy of a cultural competency approach in the particular case

This is a feasible approach to routine patient care with members of ethnic minorities and recent immigrants and refugees that has a high likelihood of making that care culturally informed and culturally sensitive. Key to it, as it is to anthropology's core methodology, ethnography, is the display of genuine respect for patients, families, and their meanings and practices. That respect for the person and his or her illness experience is the indispensable condition of anthropologically informed care. It includes, as its first step, the ethical act of acknowledging the suffering of the other in his or her own terms as the basis for diagnosis and treatment. In this sense, it reverses the cultural preoccupation of the biomedical practitioner with the disease process, and establishes the interpersonal relationship as the grounds of knowledge as well as caregiving.

Conclusions

Anthropology's chief contribution to psychiatry is to emphasize the importance of the social world in diagnosis, prognosis, and treatment, and to provide concepts and methods that psychiatrists can apply (the appropriate cross-disciplinary translation first being made, however). But that is not the only contribution that anthropology offers. Ethnographers are aware that knowledge is positioned, facts and values are inseparable, and experience is simply too complex and robust to be easily boxed into tight analytical categories. Hence a sense of the fallibility of understanding, the limitation of practice, and irony and paradox in human conditions is the consequence of ethnography as a method of knowledge production.

Anthropology also complements the idea of psychosomatic relationships with evidence and theorizing about sociosomatic relationships. Here moral processes—namely what is at stake in local worlds—are shown to be closely linked with emotional processes, which are frequently about experiences of loss, fear, vexation, and betrayal of what is collectively and individually at stake in interpersonal relationships. Change in the former can change the latter, and this can at times work in reverse as well. Examples include the way symptoms intensify or even arise in response to fear and vexation concerning threats perceived as serious dangers to what is most at stake.

The relationship of poverty to morbidity and mortality is a different example of sociosomatic processes. Poverty correlates with increased morbidity and mortality. Psychiatrists have often had trouble getting the point that public health and infectious disease experts have long understood. But it is not just diarrhoeal disease, tuberculosis, AIDS, heart disease, and cancer that demonstrate this powerful social epidemiological correlation—so do psychiatric conditions. Depression, substance abuse, violence, and their traumatic consequences not only occur at higher rates in the poorest local worlds, but also cluster together (much as do infectious diseases), and those vicious clusters define a local place, usually a disintegrating inner-city community. Hence the findings of the National Co-Morbidity Study in the United States of America that most psychiatric conditions occur as comorbidity is a step toward this ethnographic knowledge—that in the most vulnerable, dangerous, and broken local worlds, psychiatric diseases are not encountered as separate problems but as part of these sociosomatic clusters.

Finally, anthropology is also salient for policy and programme development in psychiatry. Against an overly narrow neurobiological framing of psychiatric conditions as brain disorders, anthropology in psychiatry draws on cross-national, cross-ethnic, and disintegrating community data to emphasize the relationship of increasing rates of mental health problems, especially among underserved, impoverished populations worldwide, and increasing problems in the organization and delivery of mental health services to fundamental transformations in political economy, institutions, and culture that are remaking our epoch. In so doing, anthropology projects a vision of psychiatry as a discipline central to social welfare and health policy. It argues as well against the profession's ethnocentrism and for the field as a larger component of international health. Anthropology (together with economics, sociology,

and political science) also provides the tools for psychiatry to develop policies and programmes that address the close ties between social conditions and mental health conditions, and social policies and mental health policies.[5] In this sense, anthropology urges psychiatry in a global direction, one in which psychiatric knowledge and practice, once altered to fit in more culturally salient ways in local worlds around the globe, have a more important place at the policy table.[6,7]

Further information

Jenkins, J. and Barrett, R. (eds.) (2004). *Schizophrenia, culture, and subjectivity: the edge of experience*. Cambridge University Press, Cambridge.

Kleinman, A. and Good, B. (eds.) (1985). *Culture and depression*. University of California Press, Berkeley, CA.

Kleinman, A. (ed.) (2006). *What really matters: living a moral life amidst uncertainty and danger.* Oxford Press, Oxford.

Luhrmann, T. (2000). *Of two minds: the growing disorder in American psychiatry.* Knopf, New York.

References

1. Kleinman, A. (1977). Depression, somatization and the new cross-cultural psychiatry. *Social Science and Medicine*, **11**, 3–10.
2. Estroff, S.E. (1993). Disability and schizophrenia. In *Knowledge, power and practice: the anthropology of medicine and everyday life* (eds. S. Lindenbaum and M. Lock), pp. 247–86. University of California Press, Berkeley, CA.
3. Kleinman, A. (1988). *The illness narratives*. Basic Books, New York.
4. Kleinman, A. (1988). *Rethinking psychiatry: from cultural category to personal experience.* Free Press, New York.
5. Lee, S. and Kleinman, A. (1997). Mental illness and social change in China. *Harvard Review of Psychiatry*, **5**, 43–5.
6. Kleinman, A. and Benson, P. (2006). Anthropology in the clinic: The problem of cultural competency and how to fix it. *PLoS Medicine*, **3**, e294. www.plosmedicine.org.
7. Desjarlais, R., Eisenberg, L., Good, B., *et al.* (ed.) (1995). *World mental health: problems and priorities in low-income countries*. Oxford University Press, New York.

2.7

The contribution of epidemiology to psychiatric aetiology

Scott Henderson

Introduction

Epidemiology deals with the overall patterns of disease. On one hand, people are unique with their own genetic endowment and life experiences. This idiographic paradigm is balanced by the nomothetic in which recurrent and predictable patterns are sought in the whole of humankind. It is the business of psychiatric epidemiology to determine the distribution of mental disorders in populations, the factors determining that distribution, and measures that may help in their prevention.

From their undergraduate years onward, clinicians see patients who have a disorder and who at the same time give a history of certain experiences from birth to their present. It may be tempting for both patient and doctor to accept that the patient's recent experiences have some role in the onset of symptoms. But if the principles of epidemiology are brought into play, some questions need to be asked first. Being unwell may itself bias the recall of recent or distant experiences. What proportion of the general population have had the same experiences but not developed the disorder? What proportions have the same disorder but have not had these experiences? What proportions have the same disorder but have not reached health services? A simple two-by-two table is the simplest way to think this through (Table 2.7.1).

The columns are made up of persons in a population who have or do not have a particular disorder. The rows are the numbers who have or have not had a certain exposure. That exposure is being considered as a putative risk factor. It may be biological or psychosocial and may have taken place at any time from conception to the present.

Table 2.7.1 Cases and exposure: a two-by-two table

		Case	
		Yes	**No**
Exposed	Yes	*a*	*b*
	No	*c*	*d*

Letters refer to numbers of persons.

To establish a causal link between some factor and a disorder is a demanding but most engaging exercise. It is well worth reading the classic expositions by Hill[1] and Susser[2] on how a cause can reasonably be inferred from the data.

The uses of epidemiology

In his celebrated monograph bearing this title, Morris[3] described seven uses of epidemiology. It continues to give us a framework for assessing the state of psychiatric epidemiology in relation to the biological and psychosocial conditions of the contemporary world. Morris's list can be reinterpreted for our use as follows.

Completing the clinical picture

This means knowing about all the ways in which a disorder may present and what its usual course is. But it also means relating subclinical cases to fully developed ones. An excellent example here would be the anxiety, depressive, or somatization states seen in general practice or field surveys compared to the more severe syndromes specified in the international criteria and encountered by psychiatrists.

Community diagnosis

Here one obtains estimates of morbidity as it occurs at the general population level, not just in persons who have reached primary care or mental health services. Only by having such estimates of prevalence or incidence for whole populations can the size of the nation's disease burden be determined. This is because community-based measures of morbidity include not only persons with treated conditions but also those who are symptomatic yet have not reached services.

Secular changes in incidence

This refers to the rise and fall of diseases in populations. For example, there is some evidence that schizophrenia has been dropping in incidence and becoming more benign in its clinical course; and it is likely that in many Western countries, depressive disorder has become more frequent in persons born since the Second World

War.[4] The suicide rate of young persons has indisputably increased in many industrialized countries. It is likely that eating disorders have increased in frequency and it is certain that the use of illegal drugs and AIDS are new arrivals and will be a continuing burden.

The search for causes

Here, epidemiology is looking for aetiological clues. It is the substance of this chapter.

Applying population data to individual risk

In this, the focus moves from the population back to the individual. For example, if the annual incidence rate for schizophrenia is known in a population and if this information is age-specific, it is possible to estimate the probability that a person in a given age group will develop the disorder within the next year. This is the base rate, before one starts to consider risk factors such as family history. Next, by aggregating data on the course of schizophrenia, it is possible to estimate the chances of recovery for persons who are currently having their first episode. The common principle is that data based on large numbers of persons are used to make probability estimates for individuals.

Delineation of syndromes

This is done by examining the distribution of clinical phenomena as they occur in the population. It fits well with recent experience of repetitive strain injury, chronic fatigue syndrome, and post-traumatic stress disorder or its congeners.

Health services research

This begins with a determination of needs and of resources, then an analysis of services currently in action, and ends with attempts to evaluate them, including the costs. Research activity in this area has expanded greatly in recent years, driven by the forces of economic rationalism.

Prevention

To Morris's seven uses of epidemiology should be added prevention, which Gruenberg[5] said was its 'ultimate service'. All other uses are subsidiary to this. Examples are the current activity in the prevention of suicide in young persons and of alcohol or drug abuse. In these, the traditional medical approach of targeting high-risk groups should be contrasted with the epidemiological and population-based approach described by Rose.[6] One under-recognized fact is that knowledge about factors that determine the duration of a disorder can lead to prevention. This is because prevalence is the incidence rate times the duration. So shortening the duration will lower the prevalence. For example, if people with depressive disorder were treated earlier in primary care, the prevalence should fall. The prevention of a mental disorder is greatly helped by knowledge about aetiology, but it is not essential. For example, Snow did not know about the cholera vibrio when he had the water supply changed. Prevention is discussed further by Bertolotte in Chapter 7.4, prevention in child psychiatry by Lenroot in Chapter 9.1.4, and in intellectual disability by Kaski in Chapter 10.3.

Research on aetiology: three levels

Epidemiological methods can be applied at any of the three levels: to disorders as these present in hospitals and specialist health services, in primary care, or in the general population. These are represented diagrammatically in Fig. 2.7.1 as a three-dimensional cone, derived from the seminal volume on pathways to care by Goldberg and Huxley.[7] The base of the cone consists of those in the general population who have clinically significant psychological symptoms (Stage 1). When these symptoms become unmanageable to self or to others, people seek help from their doctor or other health practitioners (Stage 2). But only some of them are recognized by the professional to have significant mental health problems (Stage 3). A small proportion may be referred to mental health services (Stage 4), of whom an even smaller fraction are admitted to inpatient care (Stage 5). Note that most teaching and the diagnostic criteria in international use are largely based on their authors' experience in Stages 4 and 5, where patients are more severely ill!

There are two very different ways to express morbidity. The most common, and easiest to obtain, is prevalence, either at the time of assessment (point prevalence), 1-month, 1-year, or lifetime prevalence. The other is incidence, the number of new or fresh onset cases in a given period. For aetiological research, incident cases are emphatically preferable.

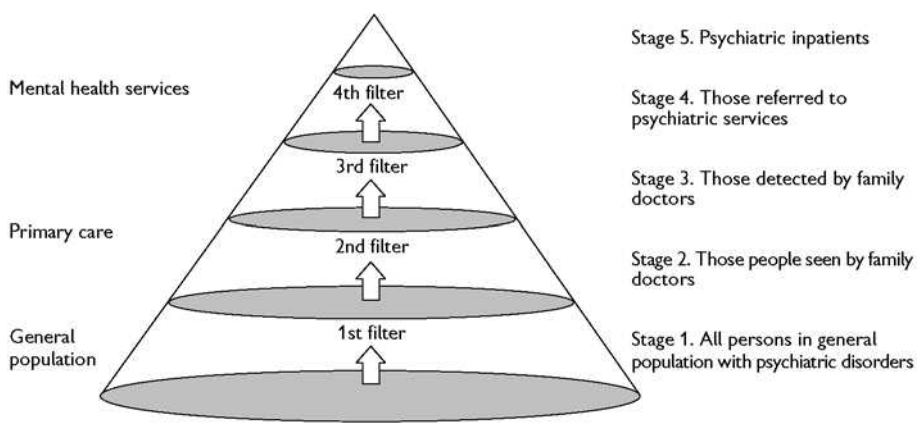


Fig. 2.7.1 Pathways to care. (Reproduced from D.P. Goldberg and P. Huxley, *Mental illness in the community,* Copyright © 1980, The Tavistock Institute.)

Three main designs

At any of these levels, research directed at aetiology uses one of three designs: cross-sectional, prospective longitudinal, or case-control. A cross-sectional study is often an excellent start, because it provides a picture of how much morbidity is present at one point in time and the variables most closely associated with this. But because it is only a 'snapshot', the cross-sectional study can rarely allow much to be said about causes. For example, in a community sample of several thousand adults, the data will show that persons with symptoms of anxiety or depression will tend to report having had more adversities. But it would be unwise to conclude that adversity contributes to the onset of symptoms. First, persons with anxiety or depression may be more inclined to report that they have had many troubles. This may be through selective recall of unpleasant events, because it is known that depressed people are more likely to remember bad times than good times.[8] Another mechanism is effort after meaning, whereby people try to account for feeling psychologically unwell. Next, symptomatic persons may be more likely to have unpleasant things happen to them as a consequence of their mental state. Lastly, persons with anxiety or depression may have certain personality traits or lifestyles that make them more likely to have troubled lives and also be vulnerable to common mental disorders.

Such problems in methodology can be resolved to some extent by using a *prospective longitudinal design* or cohort study. In this, a population sample is assessed at the start when most persons are psychologically well. In one type of cohort study, the sample may deliberately include a group who have had a particular exposure, such as a head injury or disaster, and an equivalent number who have not. At the start, data are obtained on personality, lifestyle, past health, and family history. The cohort is then re-examined at least once after an appropriate interval. Some will have developed symptoms. The research question is whether the putative risk factors that were assessed at the start were more frequently present in those who later developed symptoms. A design of this type yields considerably more information about the causal processes likely to be at work, either those leading to mental disorders or protecting against them. It also overcomes the problem of a putative risk factor really being a consequence rather than an antecedent of a disorder. But it is obviously very demanding in resources—human, administrative, and financial. It also takes a long time. For these reasons, epidemiologists often use the *case-control* method as a more practicable alternative.

Case-control designs have been underused in psychiatric research, but they can be a powerful strategy for identifying risk factors for a specified disorder.[9,10] The essence of the case-control design lies in obtaining data to complete the cells in Table 2.7.1. The aim is to find a sample of all persons in a population who have reached case level for a particular disorder and an equivalent number of persons who are similar in age, gender, and other variables, but who do not have the disorder, at least not yet. The cases should ideally be 'incident' or recent in onset. If instead, the study has to have recourse to all the cases of the disorder known to the service, that is, the prevalent cases, some will be long-standing and some more recent. This could lead to misleading results because a putative risk factor may show up as 'positive' not because it is a cause or true risk factor, but because it is associated with chronicity through prolonging the duration of the disorder. This problem can be avoided only by recruiting recent-onset or incident cases for case-control studies. The cases and controls are then asked about the various possible exposures. If the cases are unable to give information because they are cognitively impaired (as in dementia), at least one informant has to be found for each case, usually a partner or close family member.

In Table 2.7.1, the important question is whether there are more persons in cell *a* than would occur by chance. We do not know the incidence of the disorder in all persons in the population who were exposed to each risk factor, nor do we know the number not exposed. Likewise, we do not know how many people in the population have recently developed the disorder. As a consequence, we cannot compare the incidence in those exposed and not exposed for the whole population. All we have are the data from the cases examined, who are necessarily only a fraction of all incident cases in the population; and data from a fraction of all healthy persons. But we can proceed as follows. First, the relative risk is calculated from Table 2.7.1. The relative risk is $a/(a + b)$ divided by $c/(c + d)$

i.e. $$\frac{a/(a+b)}{c/(c+d)}$$

By simple algebra, this becomes

$$\frac{a(c+d)}{c(a+b)}$$

Then something very helpful can be done. Where a disorder is fairly uncommon in the general population, *a* will be very small compared with *b*, and *c* will be small compared with *d*. If we assume a negligible contribution by *a* in the term $a + b$, and by *c* in the term $c + d$, the relative risk will be nearly equal to

$$\frac{ad}{bc}$$

This is the odds ratio, which is an expression of the strength of a risk factor.

Whom to study: principles of sampling

The essential principle is that everyone in the true denominator (usually the total population within a defined geographical or administrative area) *must have an equal probability* of being included in the numerator. If this is not achieved, there is a likelihood of bias whereby the achieved sample may be systematically different in ways that could be important in the analysis. For example, the sample of cases should not differ from all the incident cases in that population in attributes such as level of education, age, or likelihood of having been exposed to a candidate exposure or risk factor. So in a study of the association between, say, sexual abuse in childhood and depressive disorder in adult life, the cases of depression should ideally be representative of all those with depressive disorder in that community and not just those reaching a particular service. See also Chapter 2.2 by Dunn.

Sample bias

In field surveys, it has long been accepted that not everyone who is in the 'target sample' will agree to be interviewed or will be available at the time the interviewer calls. It is common to find that

only 70 to 90 per cent are actually assessed. Furthermore, those who refuse or are repeatedly not available are known to be more likely to have the mental disorder under investigation. For this reason, the prevalence that is found will often be an underestimate. A putative risk factor may itself increase the chances of a person's not being in a sample in the first place, of dropping out, or of dying during the study. Statistical methods are available for estimating how much error may have occurred due to refusals and how to correct for this in the conclusions drawn.

The other occasion when non-response is a problem is in longitudinal studies, where a sample is followed over several years. If a disorder with an increased mortality is the topic, such as dementia or schizophrenia, it is recognized that some cases will be lost at follow-up. This means that those who are successfully re-examined are a survival élite and are different in important ways from the original cohort. These distortions could lead to mistaken conclusions if the losses are not allowed for. Various techniques have been developed to handle these difficulties, including Bayesian methods which adjust final estimates on the basis of prior knowledge.[11]

Specifying the disorders

Diagnostic categories

The epidemiology of mental disorders could have made no real progress without methods for specifying the disorders to be investigated, then measuring these. Only in this way can research data be comparable between research teams, within and between countries. Having consistency in diagnosis has been made much easier through the development of the diagnostic criteria now in wide international use. The two systems are the International Classification of Diseases (10th Revision) (**ICD-10**) with its *Classification of Mental and Behavioural Disorders*; and the *Diagnostic and Statistical Manual*, fourth edition (**DSM-IV**) of the American Psychiatric Association. These classifications are described further in Chapter 1.9. Both are under revision.

Continuous measures of morbidity

Reliable and valid case ascertainment might be assumed to be the *sine qua non* for any progress in the epidemiology of mental disorders. But to use the traditional expression 'case ascertainment' nicely illustrates the very problem that has to be re-thought, because it implies a categorical structure in the morbidity that we wish to study. In a population, there are traditionally cases and non-cases. But this is not really how morbidity shows itself. As expressed by Pickering,[12] 'Medicine in its present state can count up to two, but not beyond'. He was referring to hypertension, but others have argued that mental disorders also have dimensional properties.[6] The frequency distribution of their component symptoms such as anxiety, depression, or cognitive impairment is usually a reversed J-shape, with most people having none or only a few symptoms and progressively fewer persons having higher counts. A committee of clinicians in Geneva or Washington, whose experience is largely derived from teaching hospitals, has decided by consensus where the cut-point should be placed for persons to be 'cases'. While this is entirely appropriate for some purposes, it may not always be a true representation of the underlying pathology. In statistical terms, it loses information.

It is not disputed that mental disorders exist in categorical states and that these have some utilitarian value: a depressive episode, Alzheimer's disease, anorexia nervosa, or alcohol dependency are clinically realistic entities. What is proposed here is that, in epidemiological studies at the general population level, hypotheses about the aetiology require large numbers of respondents, solely because the base rates for such conditions are not large. But it is possible to identify persons with *some* symptoms of depression, of cognitive impairment, of abnormal eating, or of alcohol misuse. The score on a scale of these symptoms can become the dependent variable in an analysis of candidate risk factors. So it is usually more powerful statistically to look for associations between a putative risk factor and morbidity expressed as a continuous variable, rather than as a dichotomy of cases and non-cases.

When a continuous measure of common psychological symptoms such as the General Health Questionnaire (**GHQ**)[13] (*vide infra*) is applied to a population, a unimodal distribution curve is found, with no break between so-called cases and normals. Rose[6] argued that there are three important consequences from this approach to studying morbidity. First, a characteristic of the community as a whole emerges. This is the mean and standard deviation of its GHQ scores. Second, this collective characteristic may show significant differences between men and women, geographical regions, social strata, and income groups. These differences are based on shifts of the entire distribution. The third consequence is that differences between these groups in the prevalence of probable cases (those with a score above a threshold) are related to different average scores in these groups. As Rose[6] concisely put it, 'The visible part of the iceberg (prevalence) is a function of its total mass (the population average)'.

He suggested that '*Psychiatrists, unlike sociologists, seem generally unaware of the existence and importance of mental health attributes of whole populations, their concern being only with sick individuals*' (emphasis added). It is an appealing notion that populations, while they are made up of individuals, take on properties of their own, much as molecules acquire attributes not found in their constituent atoms. The concept of populations having different frequency distributions of dimensional morbidity, not just different prevalence rates for clinical cases, carries with it the implication that some factors are shifting the overall distribution in some populations but not in others. An example is the societal forces that Durkheim[14] proposed were related to national suicide rates.

Disablement

There is yet a further advantage in considering morbidity as a continuum in a population. Morbidity refers to symptoms, syndromes, or disorders. But there is a universe of discourse closely linked to this, namely disablement. This is the collective noun now used to refer to the impairment, disability, and social role handicap in daily life that disorders bring with them. The main categories of mental disorder, especially the psychoses, affective disorders, and dementias, are almost always associated with substantial disablement. But subclinical levels of mental disorders also carry with them a certain amount of disablement. From the point of view of a whole population, the cumulative amount of disablement from subclinical or milder conditions is considerable because such conditions have a high point prevalence. Therefore, from a public health perspective, the significance of milder mental disorders is not trivial.

Measurement of disablement

The most comprehensive measure is the Disability Assessment Schedule (DAS)[15] that assesses an individual's functioning in daily life. Its short form is suitable for survey use. Self-completion instruments are the Brief Disability Questionnaire (BDQ)[16] and the SF-36[17] or its briefer version, the SF-12.

The measurement of psychiatric symptoms

Instruments for epidemiological research fall into two types: self-completed questionnaires and standardized interviews.

Questionnaires

The more simple type is a symptom scale which can be completed by respondents themselves or administered by an interviewer. The best-known instrument is the GHQ. The briefest version, the GHQ-12, is a highly efficient screening tool. A score of 2 or more indicates that the person is likely to have one of what Goldberg and Huxley[18] have usefully termed common mental disorders. Another screening instrument is the Hopkins Symptom Checklist.[19] For depressive states specifically, examples are the Center for Epidemiologic Studies Depression Scale (**CES-D**)[20] and the Beck Depression Inventory.[21]

The Alcohol Use Disorders Identification Test (**AUDIT**) was developed by the WHO for population screening. This 10-item test has been shown to have satisfactory psychometric and predictive properties. The total score is 40 and a score of 8 or more (7 for women) is recommended for identifying persons likely to have adverse consequences of drinking.[22]

For cognitive impairment, the Mini-Mental State Examination (**MMSE**)[23] gives a score for a person's current cognitive function, or by applying a cut-point to that score, it can be used to identify persons who are likely to have a dementia. Like any other cognitive test, it cannot itself make a diagnosis of dementia. It detects cognitive impairment, not cognitive decline, which requires a history. The MMSE is known to be sensitive to education, in that persons with limited intelligence or education may have low scores without their having had any cognitive decline.

All these instruments can be used in two ways: by applying a cut-point to identify persons who are likely to be clinically significant cases; or by using the score as a continuous variable. When a questionnaire or self-rated instrument is used in research, the investigators have to be confident about its psychometric properties. In addition to validity and reliability, there are two others: its sensitivity and specificity. These refer to its performance when compared with a criterion or 'gold standard', such as a comprehensive psychiatric examination or a consensus diagnosis amongst experts. In Table 2.7.2, a sample of persons has been examined with both the screening instrument and a full examination. We consider the number of persons who are 'cases' according to both assessments (*a*), according to one but not the other (*b* or *c*), and according to neither (*d*).

Sensitivity is the proportion who screen positive and who are indeed cases by the criterion; that is: $a/(a + c)$, while specificity is the number who screen negative and who are indeed not cases: $d/(b + d)$.

The sensitivity and specificity of a test are expressed as a percentage and vary according to where the cut-point is placed on the scores.

Table 2.7.2 Screening a population sample

		Criterion (cases by full psychiatric assessment)		
		Yes	**No**	
Cases by screening test	Yes	*a*	*b*	*a+b*
	No	*c*	*d*	*c+d*
		a+c	*b+d*	

Letters refer to numbers of persons.

As sensitivity increases, specificity tends to decrease, so that an appropriate balance between the two has to be determined. For some purposes, such as in screening for depression, it is more important to identify as many as possible of the true cases, but it does not matter if there are quite a few false-positives because these can be corrected by a more extensive second-stage examination. Under these conditions, one would want a highly sensitive screening test that placed most of the true cases in cell *a* and few in *c*. It matters rather less if quite a few of the true non-cases are incorrectly placed in *b*. See also Chapter 2.2 by Dunn.

Standardized psychiatric interviews

Even the best-designed questionnaires with the best psychometric properties cannot be a substitute for a psychiatric interview. Only it can obtain the information that leads to a diagnosis, reached according to the international criteria. What is termed information variance is reduced by having interviewers ask about symptoms in the same way. Next is the reduction of criterion variance, where the symptoms or signs elicited are, like building bricks, assembled in exactly the same way, both within and across studies. This is achieved by applying to the data an algorithm that is a precise expression of the diagnostic criteria in ICD or DSM. The algorithm can be computerized so that the responses to each item in the interview are assembled automatically and invariably.

There are two types of standardized psychiatric examination. The Schedule for Clinical Assessment in Neuropsychiatry (**SCAN**) is a clinician's instrument, requiring familiarity with the phenomenology of mental disorders. It assumes that interviewers are comfortable in examining persons with a mental disorder. The second type, exemplified by the Composite International Diagnostic Interview (**CIDI**) is fully scripted and has been automated for interviewing on laptop computers so that it can be administered by laypersons after only a few days' training. A very large body of survey data has now been collected using it. These instruments are fully described by Cooper and Oates in Chapter 1.8.1.

Validity issues

The quality of information these instruments obtain is clearly of central importance. We know that they do not obtain the same information as each other, nor do they identify the same persons as cases of a particular diagnostic category in the same sample. Their validity and the practical significance is therefore important for administrators using the data obtained in large surveys to guide planning. But it is also a matter of interest for aetiological research.

Typical prevalence estimates

The above instruments have been used in a large number of surveys of the prevalence of the main mental disorders in the general population. Examples are shown in Table 2.7.3. The rates vary markedly, but little is known about why this is so. The differences may be due to differences in methods of case ascertainment; or they may indicate actual variation between countries. What is important is that they all *attempt* to estimate the prevalence in the community of syndromes familiar to psychiatrists working in clinics and hospitals. Prevalence estimates provide evidence for the public health significance of mental illnesses and their associated disablement. Their administrative impact has been considerable. So far, however, they have brought little to advance knowledge about aetiology.

(a) Possible causes of mental disorders: the domains

Having specified the psychiatric disorder to be studied and having developed methods to measure it, the next task is to identify whatever factors might contribute to its onset or to its course. These lie in the traditional three domains: the biological, social, and psychological. But in epidemiology, any variable is rarely specific to one domain. Some biological, some social, and some psychological factors are often conflated. For example, gender expresses biological differences but also different social experiences in the past, different social contexts in the present, and different psychological or intrapersonal differences in personality traits and behaviour. Likewise, age groups can reflect social role, educational opportunities in the past, marital status, financial concerns, and physical health. It is therefore important not to be misled into thinking that a variable is tapping only what one is primarily interested in. Confounding is ever present.

(b) Sociodemographic variables

The level of psychiatric morbidity in a population may differ significantly between age groups, gender, marital status, ethnic background, and socio-economic or educational level. Depressive disorders are consistently more prevalent in women[24] and dementia has a higher prevalence, not only with increasing age, but also in those with lower education.[25] Furthermore, there may be important interaction effects between variables in their relation to morbidity. For example, the average age of onset in schizophrenia is different in men and in women. This has proved to be a clue to causal processes.

Table 2.7.3 Some 12-month prevalence rates per 100 of population, all by DSM-IV criteria

	Any disorder* (%)	Anxiety disorder	Mood disorder	Alcohol or substance use disorder
Australia	20.3	5.6	6.6	7.9
Canada	10.9	5.2	5.8	_
China (Beijing)	9.1	3.2	2.5	2.6
Colombia	17.8	10.0	6.8	2.8
European countries (six)	9.6	6.4	4.2	1.0 (alcohol only)
Iran†	17.1	8.4	4.3	nil
Lebanon	16.9	11.2	6.6	1.3
New Zealand	20.7	14.8	8.0	3.5
Nigeria	4.7	3.3	0.8	0.8
Ukraine	20.5	7.1	9.1	6.4
USA	26.4	18.2	9.6	3.8

*Includes disorders other than those listed.
†Only lifetime estimates were published.

(c) The social environment

This can be considered in two parts: first is the individual's immediate social environment—what the sociologist Cooley[26] called the primary group—consisting of those around a person with whom there is both interaction and commitment. There is then the wider community with its standard of living, lifestyle, and cultural values. Plausibly, both may have some influence on the incidence of mental disorders and on their course. The hypothesis that social support protects against depression and other common mental disorders has proved hard to investigate.[27, 28] This is because social support is probably influenced by some intrapersonal factors rather than being a product solely of the individual's environment. Here is a good example of confounding: a major variable concerning the social environment of individuals turns out to be determined not solely by environmental factors, but partly by their own personality attributes. The evidence suggests that social support, stripped of these confounding factors, is not a powerful factor in aetiology.[29] A separate issue is whether social support influences the outcome of psychiatric disorders once these have developed.

Societal (macrosocial) variables have long been suspected of playing an important role in aetiology. It was such a hypothesis that was investigated in the celebrated Stirling County Study in Canada by Dorothea and Alexander Leighton *et al.*[30] with their concept of sociocultural disintegration. The current increase in depression and suicide in the young is popularly attributed to such macrosocial variables. On the other hand, work opportunities, diet and use of drugs and alcohol have also changed appreciably over the last 50 years. There is no certain explanation so far.

Experiential variables

'Experiential' is a useful term to refer inclusively to all that individuals have been exposed to, from conception to death. In epidemiological research, it includes intrauterine exposures—such as maternal influenza or malnutrition—and perinatal events. In infancy and childhood, social and interpersonal experiences have been the main focus of research. Maternal deprivation was intensively studied in both clinical and community samples for two decades. The expectation from Bowlby's attachment theory was that loss or separation from the mother would be pathogenic for depression and possibly personality disorders.[31] This hypothesis has proved very hard to test because of confounding by other factors. Rutter[32] concluded that '. . . the residual effects of early experiences on adult behaviour tend to be quite slight because of both the maturational changes that take place during middle and later childhood and also the effects of beneficial and adverse experiences during all the years after infancy . . . '

Migration and schizophrenia

Migration is an excellent example of how epidemiological data lead to aetiological clues, straddling both the biological and social. People who have moved from one country to another have offered opportunities for aetiological research on schizophrenia for a long time, starting with the celebrated study in 1932 by Ödegaard. He found that Norwegian migrants to Minnesota had more mental disorders than those at home. Since then, migration has become increasingly common and some striking findings have been made. An integrative review[33] has found a relative risk of 2.7 for first-generation migrants and 4.5 for the second-generation. While a family history is the largest risk factor for schizophrenia, migration is the next compared with most others. It is not due to an artefact in diagnosis. The effect is unlikely to be due to any single factor. Both biological and social factors are probably implicated. In the latter, perception of social inequality is thought to be a plausible mechanism.

Parental style

Promising findings have been obtained on the association between parenting style and depressive disorders in adulthood. Parker[34] developed the Parental Bonding Instrument (**PBI**) to measure two fundamental dimensions of the manner in which parents behave towards their children: care and affection as one dimension and protectiveness as the other. The PBI is too lengthy for epidemiological research on community samples where the interview is often already extensive. Parker and his colleagues have subsequently developed a briefer instrument, the Measure of Parenting Style (**MOPS**), which includes the experience of physical and sexual abuse.[35] MOPS is likely to prove useful in case-control and prospective studies of psychiatric disorders for systematically obtaining information on exposures of theoretical relevance.

(a) Childhood abuse

It seems intuitively likely that children who have been physically or sexually abused have an increased risk of having anxiety, depression, or other psychiatric disorders in adulthood. The findings from epidemiological studies on unreferred samples (people in the community) point to the many other adverse experiences that accompany childhood sexual abuse, including physical violence, unstable and untrustworthy relationships with parents, and emotional deprivation. This topic is further considered in Chapter 9.3.3.

(b) Recent exposure to adversity

Adverse experiences have been very extensively studied for their contribution to the onset and course of psychiatric disorders. In epidemiological research, much attention has been accorded to issues that arise in the measurement of adversity. Some of these issues are equally relevant in clinical practice. They include the following:

- The duration of the stressor: acute or long-standing.
- Its magnitude and how to determine this independently of the person's reaction to it.
- The independence of the event from the individual: some events are entirely independent while others may have come about because of the individual's own behaviour or psychiatric state.
- The personal context of the experience may augment or reduce its psychological impact.
- Confounding by personality traits that may be independently associated with psychiatric morbidity.
- The additive effect of multiple events, some of which may be causally linked in a chain.
- The quality of the event itself: a loss or a threat.
- Effort after meaning, whereby patients and doctors may attribute symptoms to a particular experience as a way to explain the onset of illness.

These issues are fully discussed in Chapter 2.6.1.

Personality variables

Although personality traits may contribute to how vulnerable individuals are to adverse experiences, it has not often been possible to measure personality traits in general populations, then follow the sample prospectively to demonstrate if the incidence of specific disorders is indeed higher in some types. Many measures of personality are too lengthy to be used in surveys. One exception is the Eysenck Personality Questionnaire (**EPQ-R**) in which the trait of neuroticism has been found to confer increased risk of anxiety, depression or later schizophrenia.[36] The assessment of personality is considered in Chapter 1.8.2.

Molecular genetics and epidemiology

In the next few years, it can be expected that some personality traits will be found to occur more often in people with particular alleles in genes related to brain function. Amongst these, a preferred group of candidates are genes with known polymorphisms that alter the function of neurotransmitter systems, either by affecting the metabolism of a transmitter or some aspect of its function such as transport, receptor binding, or signal transduction. The attraction for psychiatric epidemiology is twofold: the promise of introducing to population studies a biological variable of fundamental significance; and the possibility of looking for interaction between biologically based vulnerability and life experiences. The finding by Caspi *et al.*[37] sparked great interest by reporting an association between a 5-HTT polymorphism and depression, but only on persons who had been exposed to recent adversity. The finding has not been consistently replicated. Furthermore, the relevant polymorphisms may be rather more complicated than at first assumed. What is significant is that it is now feasible to study genotype and environment at the population level.

How epidemiologists think

We can now look back on the strategies used to find causes of mental disorders and how epidemiologists go about the task. To find aetiological clues, they look firstly for associations, some of which may later be shown to have a causal influence. This is often best done by working not with individual patients, nor even with a series of patients in one clinic. Instead it is better to have data that represent *all* cases of a disorder in a defined population. The way in which a candidate risk factor comes to be proposed is itself very interesting. There are three:

Table 2.7.4 The search for causes: a matrix for epidemiological studies

		Child psychiatric disorders	Anxiety disorders	Affective disorders	Schizophrenia	Dementias	Personality disorders	Eating disorders	Alcohol and substance abuse	Suicide	Parasuicide	Intellectual disability	Psychological well being
Sociodemographic factors	Gender Age Marital status Social class Education Employment status Urban/rural World region												
Experiential factors	Season of birth Parental age Childhood separation Parental style Cultural or subcultural beliefs and attitudes Adverse experiences Extreme experiences Bereavement Expressed emotion (for relapse) Social support Secular changes in society Other macrosocial factors (economic depression, war, social cohesion) Migration Climate and daylight Noise Environmental toxins Diet Alcohol and drugs Medication Infections Physical illness												
	Interaction of two or more factors Comorbidity Genetics: major genes quantitative trait loci												

No attempt has been made to be exhaustive in either the classes of morbidity or the putative causal factors.

1 **The inspired hypothesis**. A sharp-eyed clinician develops the idea that some factor—in any of the three domains—is present more often than by chance in a certain disorder; and that factor may contribute to the onset of the disorder. This is exactly what happened to the ophthalmologist Sir Norman Gregg in Sydney when he noticed that many disabled children had mothers who had contracted rubella during the pregnancy. The association was later confirmed and shown to be causal. So here a hypothesis arising in the course of clinical work was taken out of the clinic and tested by epidemiological methods, then the knowledge applied in prevention.

2 **A coarse observation**. A second pathway to a hypothesis starts with a coarse observation of a link between a disorder and, say, some demographic variable. Here are two examples. In the aetiology of schizophrenia, a slight excess in winter birthdates was noticed in persons who later developed schizophrenia. Work on the 1946 birth cohort in Britain showed that persons who later developed schizophrenia had been recorded as children to have had more speech and educational problems, more social anxiety, and a preference for solitary play. Both observations point to a neurodevelopmental hypothesis for schizophrenia.

3 **The enquiry is theory driven**. A good example is what came from attachment theory. In epidemiological research on Alzheimer's disease, it was theory that led to a search for putative risk factors such as a family history of Down syndrome, aluminium in drinking water, maternal age at the patient's birth, and a history of previous head injury. It was also theory that led Jenkinson *et al*. to find an inverse association between Alzheimer's disease and rheumatoid arthritis. This subsequently led to studies of people who had been treated with long-term anti-inflammatory drugs.

4 **The search for causes: a matrix for epidemiological studies**. All these three approaches can be brought together, and then used as the building material to construct a matrix. This can then drive a systematic search. In Table 2.7.4, the main categories of mental disorders are listed across the top to form the columns, while the rows are made up of those variables that may contribute to the onset or course of morbidity. The matrix proves to be a tidy way of organizing what information is already available; but it also acts heuristically by proposing associations that call for investigation but might not otherwise have been considered. The variables can be placed in categories: sociodemographic, experiential, intrapersonal (psychological), and biological. The alert observer will notice that interactions between two or more variables should be considered, because these can be of the greatest importance.

Conclusion

The epidemiology of psychiatric disorders has shown the extent to which these are present in all human populations. It has also provided a large body of knowledge about aetiology. Now that biological including genetic information can be added to data on environmental exposures, opportunities for further advancement carry much promise.

Further information

Burger, H. and Neeleman, J. (2007). A glossary on psychiatric epidemiology. *Journal of Epidemiology Comm Health*, **61**, 185–9.

Hybels, C.F. and Blazer, D.G. (2003). Epidemiology of late-life mental disorders. *Clinics in Geriatric Medicine*, **19**, 663–96.

Henderson, A.S. (1988). *An introduction to social psychiatry*. Oxford University Press, Oxford.

References

1. Hill, A.B. (1965). The environment and disease: association or causation? *Proceedings of the Royal Society for Medicine*, **58**, 295–300.
2. Susser, M. (1973). *Causal thinking in the health sciences*. Oxford University Press, New York.
3. Morris, J.N. (1964). *Uses of epidemiology*. Williams and Wilkins, Baltimore, MD.
4. Klerman, G.L. (1988). The current age of youthful melancholia: evidence for increase in depression among adolescents and young adults. *British Journal of Psychiatry*, **152**, 4–14.
5. Gruenberg, E.M. (1966). Epidemiology of mental illness. *International Journal of Psychiatry*, **2**, 78–134.
6. Rose, G. (1993). Mental disorder and the strategies of prevention. *Psychological Medicine*, **23**, 553–5.
7. Goldberg, D.P. and Huxley, P. (1980). *Mental illness in the community: the pathway to psychiatric care*. Tavistock Publications, London.
8. Neugebauer, R. and Ng, S. (1990). Differential recall as a source of bias in epidemiologic research. *Journal of Clinical Epidemiology*, **43**, 1337–41.
9. Schlesselman, J.L. (1982). *Case-control studies: design, conduct, analysis*. Oxford University Press, New York.
10. Anthony, J.C. (1988). The epidemiologic case-control strategy, with applications in psychiatric research. In *Handbook of social psychiatry* (eds. A.S. Henderson and G.D. Burrows), pp. 157–71. Elsevier, Amsterdam.
11. Best, N.G., Spiegelhalter, D.J., Thomas, A., *et al*. (1996). Bayesian analysis of realistically complex models. *Journal of the Royal Statistical Society*, **159**, 323–42.
12. Pickering, G.W. (1968). *High blood pressure*. Churchill Livingstone, London.
13. Goldberg, D.P. and Williams, P. (1988). *A user's guide to the GHQ*. NFER/Nelson, London.
14. Durkheim, E. (1897). *Le Suicide: Etude de Sociologie*. Felix Alcan, Paris.
15. World Health Organization. (1988). *Psychiatric disability assessment schedule* (*WHO/DAS*). World Health Organization, Geneva.
16. Von Korff, M., Üstün, T.B., Ormel, J., *et al*. (1996). Self-report disability in an international primary care study of psychological illness. *Journal of Clinical Epidemiology*, **49**, 297–303.
17. Brazier, J.E., Harper, R., Jones, N.M.B., *et al*. (1992). Validating the SF-36 health survey questionnaire, new outcome measure for primary care. *British Medical Journal*, **305**, 160–4.
18. Goldberg, D. and Huxley, P. (1992). *Common mental disorders*. Tavistock Publications/Routledge, London.
19. Derogatis, L.R., Lipman, R.S., Rickels, K., *et al*. (1974). The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behavioral Science*, **19**, 1–15.
20. Radloff, L.S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, **1**, 385–401.
21. Beck, A.T. (1967). *Depression: clinical, experimental and theoretical aspects*. Harper and Row, New York.
22. Conigrave, K.M., Hall, W.D., and Saunders, J.B. (1995). The AUDIT questionnaire: choosing a cut-off score. *Addiction*, **90**, 1349–56.
23. Folstein, M.F., Folstein, S.E., and McHugh, P.R. (1975). 'Mini-Mental State': a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **12**, 189–98.
24. Bebbington, P. (1998). Sex and depression. *Psychological Medicine*, **28**, 1–8.

25. White, L., Katzman, R., Losonczy, K., *et al*. (1994). Association of education with incidence of cognitive impairment in three established populations for epidemiologic studies of the elderly. *Journal of Clinical Epidemiology*, **47**, 363–74.
26. Cooley, C.H. (1909). *Social organization: a study of the larger mind*. Scribner, New York.
27. Henderson, S., Byrne, D.G., and Duncan-Jones, P. (1981). *Neurosis and the social environment*. Academic Press, Sydney.
28. Henderson, A.S. and Brown, G.W. (1988). Social support: the hypothesis and the evidence. In *Handbook of social psychiatry* (eds. A.S. Henderson and G.D. Burrows), pp. 73–85. Elsevier, Amsterdam.
29. Henderson, A.S. (1998). Social support: its present significance for psychiatric epidemiology. In *Adversity, stress and psychopathology* (ed. B.P. Dohrenwend), pp. 390–7. Oxford University Press, New York.
30. Leighton, D.C., Harding, J.S., Macklin, D.B., *et al*. (1963). *The character of danger*. Basic Books, New York.
31. Bowlby, J. (1977). The making and breaking of affectional bonds. I. Aetiology and psychopathology in the light of attachment theory. *British Journal of Psychiatry*, **130**, 201–10.
32. Rutter, M. (1981). *Maternal deprivation reassessed*, p. 197. Penguin, Harmondsworth.
33. Cantor-Graae, E. and Selten, J.P. (2005). Schizophrenia and migration: a meta-analysis and review. *The American Journal of Psychiatry*, **162**, 12–24.
34. Parker, G. (1992). Early environment. In *Handbook of affective disorders* (2nd edn) (ed. E.S. Paykel), pp. 171–83. Guilford Press, New York.
35. Parker, G., Roussos, J., Hadzi-Pavlovic, *et al*. (1997). The development of a refined measure of dysfunctional parenting and assessment of its relevance in patients with affective disorders. *Psychological Medicine*, **27**, 1193–203.
36. Fergusson, D.M., Horwood, L.J., and Lawton, J.M. (1989). The relationships between neuroticism and depressive symptoms. *Social Psychiatry and Psychiatric Epidemiology*, **24**, 275–81.
37. Caspi, A., Sugden, K., Moffitt, T.E., *et al*. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*, **301**(5631), 386–9.

This page intentionally left blank

SECTION 3

Psychodynamic Contributions to Psychiatry

This page intentionally left blank

3.1

Psychoanalysis: Freud's theories and their contemporary development

Otto F. Kernberg

Psychoanalysis is:

1 A personality theory, and, more generally, a theory of psychological functioning that focuses particularly on unconscious mental processes;

2 A method for the investigation of psychological functions based on the exploration of free associations within a special therapeutic setting;

3 A method for treatment of a broad spectrum of psychopathological conditions, including the symptomatic neuroses (anxiety states, characterological depression, obsessive–compulsive disorder, conversion hysteria, and dissociative hysterical pathology), sexual inhibitions and perversions ('paraphilias'), and the personality disorders.

Psychoanalysis has also been applied, mostly in modified versions, i.e. in psychoanalytic psychotherapies, to the treatment of severe personality disorders, psychosomatic conditions, and certain psychotic conditions, particularly a subgroup of patients with chronic schizophrenic illness.

All three aspects of psychoanalysis were originally developed by Freud[1–3] whose theories of the dynamic unconscious, personality development, personality structure, psychopathology, methodology of psychoanalytic investigation, and method of treatment still largely influence the field, both in the sense that many of his central ideas continue as the basis of contemporary psychoanalytic thinking, and in that corresponding divergencies, controversies, and radical innovations still can be better understood in the light of the overall frame of his contributions. Freud's concepts of dream analysis, mechanisms of defence, and transference have become central aspects of many contemporary psychotherapeutic procedures.

Freud's ideas about personality development and psychopathology, the method of psychoanalytic investigation, and the analytic approach to treatment gradually changed in the course of his dramatically creative lifespan. Moreover, the theory of the structure of the mind that he assumed must underlie the events that he observed clinically changed in major respects, so that an overall summary of his views can hardly be undertaken without tracing the history of his thinking. The present overview will lead up to summaries of his final conclusions as to the structure of the mind and how this is reflected in personality development and psychopathology. Psychoanalysis will then be described as a method of treatment, as seen from the point of view of resolution of conflict between impulse and defence, and from that of object-relations theory. We shall explore significant changes that have occurred in all these domains, and conclude with an overview of contemporary psychoanalysis, with particular emphasis upon the presently converging tendencies of contemporary psychoanalytic approaches, and new developments that remain controversial.

Freud's theory of the mental apparatus: motivation, structure, and functioning

Unconscious mental processes: the topographic theory; defence mechanisms

Freud's starting point[4] was his study of hysterical patients and the discovery that, when he found a way to help these patients piece together a coherent account of the antecedents of their conversion symptoms, dissociative phenomena, and pathological affective dispositions, all these psychopathological phenomena could be traced to traumatic experiences in their past that had become unconscious. That is, these traumatic experiences continued to influence the patients' functioning despite an active defensive mechanism of 'repression' that excluded them from the patient's conscious awareness. In the course of a few years, Freud abandoned his early efforts to recover repressed material by means of hypnosis, and replaced hypnosis with the technique of 'free association', an essential aspect of psychoanalytic technique until the present time. Freud instructed his patients to eliminate as much as possible all 'prepared agendas', and to try to express whatever came to mind, while attempting to exert as little censorship over this material as they could. He provided them with a non-judgemental and stable setting in which to carry out their task, inviting them to recline on a couch while he sat behind it. The sessions lasted for an hour and were conducted five to six times a week. There has been little change in the essentials of this format, except that sessions have been shortened to 45 to 50 min and are carried out three to

five times a week. The method of free association led to the gradual recovery of repressed memories of traumatic events. Originally, Freud thought that the recovery of such events into consciousness would permit their abreaction and elaboration, and thus resolve the patients' symptoms.

Practicing this method led Freud to several lines of discovery. To begin, he conceptualized unconscious mechanisms of defence that opposed the recovery of memories by free association. He described these mechanisms, namely, repression, negation, isolation, projection, introjection, transformation into the opposite, rationalization, intellectualization, and most important, reaction formation. The last of these involves overt chronic patterns of thought and behaviour that serve to disguise and disavow opposite tendencies linked to unconscious traumatic events and the intrapsychic conflicts derived from them. The discovery of reaction formations led Freud to the psychoanalytic study of character pathology and normal character formation, and still constitutes an important aspect of the contemporary psychoanalytic understanding and treatment of personality disorders (for practical purposes, character pathology and personality disorders are synonymous concepts).

A related line of development in Freud's theories was the discovery of the differential characteristics of conscious and unconscious thinking. Freud differentiated conscious thinking, the 'secondary process', invested by 'attention cathexis' and dominated by sensory perception and ordinary logic in relating to the psychosocial environment, from the 'primary process' of the 'dynamic unconscious'. That part of the unconscious mind he referred to as 'dynamic' exerted constant pressure or influence on conscious processes, against the active barrier constituted by the various defensive operations, particularly repression. The dynamic unconscious, Freud proposed, presented a general mobility of affective investments, and was ruled by the 'pleasure principle' in contrast to the 'reality principle' of consciousness. The 'primary process' thinking of the dynamic unconscious was characterized by the absence of the principle of contradiction and of ordinary logical thinking, the absence of negation and of the ordinary sense of time and space, the treatment of a part as if it were equivalent to the whole, and a general tendency towards condensation of thoughts and the displacement of affective investments from one to another mental content.

Finally, Freud proposed a 'preconscious', an intermediate zone between the dynamic unconscious and consciousness. It represented the storehouse for retrievable memories and knowledge and for affective investments in general, and it was the seat of daydreaming, in which the reality principle of consciousness was loosened, and derivatives of the dynamic unconscious might emerge. Free association, in fact, primarily tapped the preconscious as well as the layer of unconscious defensive operations opposing the emergence of material from the dynamic unconscious.

This model of the mind as a 'place' with unconscious, preconscious, and conscious 'regions' constituted Freud's[1] 'topographic theory'. He eventually replaced it with the 'structural theory' namely, the concept of three interacting psychic structures, the ego, the superego, and the id.[5] This tripartite structural theory is still the model of the mind that dominates psychoanalytic thinking. A major determinant of the shift from the topographic to the structural model was Freud's recognition that the 'regions' of conscious, preconscious, and unconscious were fluid, and that the defence mechanisms directed against the emergence in consciousness of the dynamic unconscious were themselves unconscious. Another consideration was Freud's[6] discovery of a specialized unconscious system of infantile morality, the superego. What follows is a summary of the characteristics and contents of these structures, an analysis that will lead us directly into contemporary psychoanalytic formulations.

The structural theory, the dual-drive theory, and the Oedipus complex

The id: infantile sexuality and the Oedipus complex

The id is the mental structure that contains the mental representatives of the 'drives', that is, the ultimate intrapsychic motivations that Freud[7] described in his final, 'dual-drive theory' of libido and aggression, or metaphorically, the sexual or life drive and the destruction or death drive to be examined below. Behind this categorical formulation lies a complex set of discoveries regarding the patients' unconscious experiences that Freud came across in the course of the application of the psychoanalytic method to the treatment of neurotic and characterological symptoms. In exploring unconscious mental processes, what at first appeared to be specific traumatic life experiences turned out to reflect surprisingly consistent, repetitive intrapsychic experiences of a sexual and aggressive nature.

Freud[4] was particularly impressed by the regularity with which his patients reported the emergence of childhood memories reflecting seductive and traumatic sexual experiences on one hand, and intense sexual desires and related guilt feelings, on the other. He discovered a continuity between the earliest wishes for dependency and being taken care of (the psychology, as he saw it, of the baby at the mother's breast) during what he described as the 'oral phase' of development; the pleasure in exercising control and struggles around autonomy in the subsequent 'anal phase' of development (the psychology of toilet training); and, particularly, the sexual desire towards the parent of the opposite gender and the ambivalent rivalry for that parent's exclusive love with the parent of the same gender. He described this latter state as characteristic of the 'infantile genital stage' (from the third or fourth to the sixth year of life) and called its characteristic constellation of wishes and conflicts the positive Oedipus complex. He differentiated it from the negative Oedipus complex, i.e. the love for the parent of the same gender, and the corresponding ambivalent rivalry with the parent of the other gender. Freud proposed that Oedipal wishes came to dominate the infantile hierarchy of oral and anal wishes, becoming the fundamental unconscious realm of desire.

Powerful fears motivated the repression of awareness of infantile desire: the fear of loss of the object, and later of the loss of the object's love was the basic fear of the oral phase, directed against libidinal wishes to possess the breast; the fear of destructive control and annihilation of the self or the object was the dominant fear of the anal phase directed against libidinal wishes of anal expulsion and retentiveness, and the fear of castration, 'castration anxiety', the dominant fear of the Oedipal phase of development, directed against libidinal desire of the Oedipal object. Unconscious guilt was a dominant later fear, originating in the superego and generally directed against drive gratification (see under superego). Unconscious guilt over sexual impulses unconsciously equated with Oedipal desires constitute a major source of many types of pathology, such as sexual inhibition and related character pathology.

Prototypical intrapsychic infantile experiences linked to the Oedipus complex were fantasies and perceptions around the sexual intimacy of the parents (the 'primal scene'), and unconscious fantasies derived from experiences with primary caregivers ('primal seduction'). In all these phases of infantile development of drive motivated wishes and fears, powerful aggressive strivings accompanied the libidinal ones, such as cannibalistic impulses during the oral phase of physical dependency on the breast and psychological dependency on mother, sadistic fantasies linked to the anal phase, and parricidal wishes and phantasies in the Oedipal stage of development.

Freud described the oral phase as essentially coinciding with the infantile stage of breast feeding, the anal phase as coinciding with struggles around sphincter control, and the Oedipal stage as developing gradually during the second and through the fourth years, and culminating in the fourth and the fifth years of life. This latter phase would then be followed by more general repressive processes under the dominance of the installation of the superego, leading to a 'latency phase' roughly corresponding to the school years, and finally, to a transitory reactivation of all unconscious childhood conflicts under the dominance of Oedipal issues during puberty and early adolescence.

The id: drives

The drives represent for human behaviour what the instincts constitute for the animal kingdom, i.e. the ultimate biological motivational system. The drives are constant, highly individualized, developmentally shaped motivational systems. Under the dominance of the drives and guided by the primary process, the id exerts an ongoing pressure towards gratification, operating in accordance with the pleasure principle. Freud initially equated the drives with primitive affects. After discarding various other models of unconscious motivation, he ended up with the dual-drive theory of libido and aggression.

He described the libido or the sexual drive as having an 'origin' in the erotogenic nature of the leading oral, anal, and genital bodily zones; an 'impulse' expressing the quantitative intensity of the drive by the intensity of the corresponding affects; an 'aim' reflected in the particular act of concrete gratification of the drive; and an 'object' consisting of displacements from the dominant parental objects of desire.

The introduction of the idea of an aggressive or 'death' drive, arrived at later in Freud's[7, 8] writing, stemmed from his observations of the profound self-destructive urges particularly manifest in the psychopathology of major depression and suicide, and of the 'repetition compulsion' of impulse-driven behaviour that frequently seemed to run counter to the pleasure principle that supposedly governed unconscious drives. He never spelled out the details of the aggressive drive as to its origins. This issue was taken up later by Klein,[9] Fairbairn,[10] Winnicott,[11] Jacobson,[12] and Mahler and her colleagues.[13] Freud described drives as intermediate between the body and the mind; the only thing we knew about them, Freud suggested,[14] were 'representations and affects'.

The structure and functions of the ego

While the id is the seat of the unconscious drives, and functions according to the 'primary process' of the dynamic unconscious, the ego, Freud[5] proposed, is the seat of consciousness as well as of unconscious defence mechanisms that, in the psychoanalytic treatment, appear as 'resistances' to free association. The ego functions according to the logical and reality-based principles of 'secondary process', negotiating the relations between internal and external reality. Guided by the reality principle, it exerts control over perception and motility; it draws on preconscious material, controls 'attention cathexes' and permits motor delay as well as selection of imagery and perception. The ego is also the seat of basic affects, particularly anxiety as an alarm signal against the danger of emergence of unconscious, repressed impulses. This alarm signal may turn into a disorganized state of panic when the ego is flooded with external perceptions that activate unconscious desire and conflicts, or with overwhelming, traumatic experiences in reality that resonate with such repressed unconscious conflicts, and overwhelm the particularly sensitized ego in the process. The fact that the ego was seen by Freud as the seat of affects, and that affects had previously been described by him as discharge phenomena reflecting drives (together with their mental representations) tended to dissociate affects from drives in psychoanalytic theory, in contrast to their originally being equated in Freud's early formulations. As we shall see, this issue, the centrality of affects in psychic reality and interactions, has gradually re-emerged as a major aspect of contemporary psychoanalytic thinking.

Freud originally equated the 'I', i.e. the categorical self of the philosophers, with consciousness; later, once he established the theory of the ego as an organization of both conscious and unconscious functions, he at times treated the ego as if it were the subjective self, and at other times, as an impersonal organization of functions. Out of this ambiguity evolved the contemporary concept of the self within modern ego psychology as well as in British and American object relations and cultural psychoanalytic contributions.[15] An alternative theory of the self was proposed by Kohut[16] the originator of the self-psychology approach within contemporary psychoanalysis.

Nowadays, an integrated concept of the self as the seat of subjectivity is considered an essential structure of the ego, and the concept of 'ego identity' refers to the integration of the concept of the self: because of developmental processes in early infancy and childhood better understood today, an integrated self-concept usually goes hand-in-hand with the capacity for an integrated concept of significant others. An unconscious tendency towards primitive dissociation or 'splitting' of the self-concept and of the concepts of significant objects runs counter to such integration: we shall return to this process later. Already Freud,[17] in one of his last contributions, described a process of splitting in the ego as a way of dealing with intolerable intrapsychic conflict, thus opening up the road for considering splitting processes of the ego as an alternative, pathological defence against intolerable intrapsychic conflict (alternative, that is, to the repression of that conflict and to drawing important related ego functions into repression as well).

Character, from a psychoanalytic perspective, may be defined as constituting the behavioural aspects of ego identity (the self-concept) and the internal relations with significant others (the internalized world of 'object relations'). The sense of personal identity and of an internal world of object relations, in turn, reflect the subjective side of character. It was particularly the ego psychological approach—one of the dominant contemporary psychoanalytic schools—that developed the analysis of defensive operations of the ego, and of pathological character formation as a stable defensive organization that needed to be explored and resolved in the

psychoanalytic treatment. In the process, ego psychology contributed importantly to the psychoanalytic treatment of personality disorders.

Personality disorders reflect typical constellations of pathological character traits derived from abnormal developmental processes under the influence of unconscious intrapsychic conflicts. The description of 'reaction formation' as one of the defences of the ego led Freud to the description of the 'oral', 'anal', and 'genital' characters, particularly to the description of the obsessive–compulsive personality as a typical manifestation of reaction formations against anal drive derivatives. This was followed by the description by Abraham[18] of the hysterical personality as a consequence of multiple reaction formations against the female castration complex. Over the years, psychoanalytic explorations led to the description of a broad spectrum of pathological character constellations, which today are a part of the spectrum of personality disorders.

Perhaps the most important psychoanalytic contribution to character pathology and the personality disorders is the clinical description of the narcissistic personality disorder. While Freud provided the basic elements that led to its eventual description, psychoanalytic understanding and treatment, it was not he who crystallized the concepts of normal and pathological narcissism. Freud[19] conceptualized narcissism as the libidinal investment of the ego or self, in contrast to the libidinal investment of significant others ('objects'). In proposing the possibility of a withdrawal of libidinal investment from others with an excessive investment in the self as the basic feature of narcissistic pathology, he pointed to a broad spectrum of psychopathology, and thus first stimulated the contribution of Abraham,[20] and later those of Klein,[21] Rosenfeld,[22] Grunberger,[23] Kohut,[16] Jacobson,[12] and Kernberg.[24] Thus, crystallized the description of the narcissistic personality as a disorder derived from a pathological integration of a grandiose self as a defence against unbearable aggressive conflicts, particularly around primitive envy.

The superego in normality and pathology

In his analysis of unconscious intrapsychic conflicts between drive and defence, Freud regularly encountered unconscious feelings of guilt in his patients, reflecting an extremely strict, unconscious infantile morality, which he called the superego. This unconscious morality could lead to severe self-blame and self-attacks, and particularly, to abnormal depressive reactions, which he came to regard as expressing the superego's attacks on the ego. It was particularly in studying normal and pathological mourning, where Freud[6] arrived at the idea of excessive mourning and depression as reflecting the unconscious internalization of the representation of an ambivalently loved and hated lost object. In unconsciously identifying the self with that object introjected into the ego, the individual now attacked his or her own self in replacement of the previous unconscious hatred of the object; and the internalization of aspects of that object into the superego reinforced the strictness of the individual's pre-existing unconscious infantile morality.

Freud traced the origins of the superego to the overcoming of the Oedipus complex via unconscious identification with the parent of the same gender: in internalizing the Oedipal parent's prohibition against the rivalry with him or her and the unconscious death wishes regularly connected with such a rivalry, and against the incestuous desire for the parent of the other gender, this internalization crystallized an unconscious infantile morality. The superego, thus based upon prohibitions against incest and parricide, and a demand for submission to, and identification with the Oedipal rival, became the guarantor of the capacity for identification with moral and ethical values in general. In simple terms, the little boy renounces mother out of fear and love of father, takes father's fantasized prohibition against the little boy's sexuality into the superego as a fundamental prohibition, and establishes an identification with his father in the consolidation of his character structure. The little boy thus enacts the unconscious fantasy that, in identifying with father, he will gradually grow into his role, and satisfy his sexual desire in the distant future, by choosing another woman who, unconsciously and symbolically, will represent mother. The superego thus introduces a new time perspective into the functioning of the psychic apparatus.

Freud also described the internalization of the idealized representations of both parents into the superego in the form of the 'ego ideal'. He suggested that the earliest sources of self-esteem, derived from mother's love, gradually fixated by the baby's and small child's internalizations of the representations of the loving mother into the ego ideal, led to the parental demands becoming internalized as well. In other words, normally self-esteem is maintained both by living up to the expectations of the internalized idealized parental objects, and by submitting to their internalized prohibitions. This consideration of self-esteem regulation leads to the clinical concept of narcissism as normal or pathological self-esteem regulation, in contrast to the theoretical concept of narcissism as the libidinal investment of the self.

The superego, in summary, is a mental structure constituted by the internalized demands and prohibitions from the parental objects of childhood, the 'heir to the Oedipal complex'. This unconscious structure is of fundamental importance in determining unconscious 'fixations' to infantile prohibitions against drive derivatives and the corresponding unconscious motivation for the activation of a broad spectrum of ego defences against them, thus preventing the ego from responsibility-examining and reintegrating unresolved pathogenic conflicts from early childhood. In health, this internal sense of unconscious morality is the underpinning of moral and ethical systems. Excessive superego severity, usually derived from excessive parental strictness, determines excessive repressive mechanisms and ego inhibitions, irrational moralistic behaviour, or pathological activation of depression and loss of self-esteem.

Having thus summarized the basic psychoanalytic theory of motivation (drives), of development (the stages of development from the early oral phase to the dominance of the Oedipal complex), of structure (the tripartite model), and their implications for psychopathology, I shall now describe more specifically the contemporary psychoanalytic theory of psychopathology and of psychoanalytic treatment.

Psychoanalytic treatment

The psychoanalytic theory of psychopathology

The psychoanalytic theory of psychopathology proposes that the clinical manifestations of the symptomatic neuroses, character pathology, perversions, sexual inhibitions, and selected types of psychosomatic and psychotic illness reflect unconscious intrapsychic conflicts between drive derivatives following the pleasure principle,

defensive operations reflecting the reality principle, and the unconscious motivations of the superego. Unconscious conflicts between impulse and defence are expressed in the form of structured conflicts between the agencies of the tripartite structure: there are ego defences against impulses of the id; the superego motivates inhibitions and restrictions in the ego; at times the repetitive, dissociated expression of id impulses ('repetition compulsion') constitutes an effective id defence against superego pressures. The resolution of unconscious conflicts implies the analysis of all these intersystemic conflicts.

All these conflicts are expressed clinically by three types of phenomena:

1 inhibitions of normal ego functions regarding sexuality, intimacy, social relations, work, and affect activation;

2 compromise formations between repressed impulses and the defences directed against them;

3 dissociative expression of impulse and defence.

The last category implies a dominance of the splitting mechanisms referred to before; these have acquired central importance in the understanding of severe character pathology as reflected in contemporary psychoanalytic thinking.

The structural formulation of the psychoanalytic method

Psychoanalytic treatment consists, in essence, in facilitating the reactivation of the pathogenic unconscious conflicts in the treatment situation by means of a systematic analysis of the defensive operations directed against them. This leads to the gradual emergence of repressed impulses, with the possibility of elaborating them in relation to the analyst, and their eventual adaptive integration into the adult ego. Freud[25] had described the concept of 'sublimation' as an adaptive transformation of unconscious drives: drive derivatives, converted into a consciously tolerable form, are permitted gratification in a symbolic way while their origin remains unconscious. The result of this process is an adaptive, non-defensive compromise formation between impulse and defence. In analysis, the gradual integration into the patient's conscious ego of unconscious wishes and desires from the past and the understanding of the phantasized threats and dangers connected with them, facilitates their gradual elaboration and sublimatory expression in the consulting room and in everyday life as well.

The object-relations theory formulation of psychoanalytic treatment

In the light of contemporary object-relations theory, the formulation based upon the structural theory (resolution of unconscious conflicts between impulse and defence) has changed, in the sense that all unconscious conflicts are considered to be imbedded in unconscious internalized object relations. Such internalized object relations determine both the nature of the defensive operations and of the impulses against which they are directed. These internalized object relations constitute, at the same time, the 'building blocks' of the tripartite structure of id, ego, and superego. Object-relations theory proposes that the gradual analysis of intersystemic conflicts between impulse and defence (structured into conflicts between ego, superego, and id) decomposes the tripartite structure into the constituent conflicting internalized object relations. These object relations are reactivated in the treatment situation in the form of an unconscious relation between self and significant others replicated in the relation between patient and analyst, i.e. the 'transference'.

The transference is the unconscious repetition in the 'here and now' of unconscious, conflicting pathogenic relationships from the past. The transference reflects the reactivation of the past conflict not in the form of a memory, but in the form of a repetition. This repetition provides essential information about the past, but constitutes, at the same time, a defence in the sense that the patient repeats instead of remembering. Therefore, transference has important informative features that need to be facilitated in their development, and defensive features that need to be therapeutically resolved once their nature has been clarified. Transference analysis is the fundamental ingredient of the psychoanalytic treatment.

The psychoanalytic treatment process

The psychoanalytic treatment consists of the creation of an atmosphere of safety in which a patient is willing to try to express whatever comes to mind. In 45 to 50 min sessions, three to five times per week, the patient usually reclines on a couch while the analyst, generally sitting behind the patient, helps the patient become aware of his or her defensive operations ('resistances') by means of interpretations. The systematic interpretation of resistances gradually permits an ever-growing freedom of free association, and helps the patient to become aware of his or her unconscious desires and fears, phantasies and terrors, traumatic situations, and unresolved mourning. Defensive operations are usually classified as ego defences (in the form of the mechanisms listed earlier), superego defences in the form of excessive guilt feelings activated during the treatment, id resistances in the form of repetition compulsion, the development of secondary gain from symptoms as a powerful resistance, and, last and most importantly, the transference as the dominant resistance and source of information.

Gill,[26] in a classical definition that is still relevant today, proposed the definition of psychoanalysis as a treatment that facilitates the development of a 'regressive transference neurosis' and its resolution by means of interpretation alone, carried out by the analyst from a position of technical neutrality. Let us define these concepts.

'Regression' refers to the patient's return to earlier experiences (temporal regression), and modes of functioning (structural and formal regression) under the effect of the analysis of resistances, and is an expression of the reactivation of his unconscious conflicts from the past in the transference. In essence, the patient activates or enacts earlier object relations in the transference. Certain past stages of development where particular traumatic experiences occurred act as gathering points ('fixations') that foster regression towards them. The concept of a regressive transference neurosis refers to the gradual gathering into the relationship with the analyst of the patient's most important past pathogenic experiences and unconscious conflicts. The concept of a regressive transference neurosis has been largely abandoned in practice because, particularly in patients with severe character pathology, transference regression occurs so early and consistently that the gradual development of a regressive transference neurosis is no longer a useful concept.

Gill's proposal that the resolution of the transference be achieved 'by interpretation alone', refers to 'interpretation' as a set of the psychoanalyst's interventions that starts with 'clarification' of the

patient's subjective experiences communicated by means of free association, expands with the tactful 'confrontation' of aspects of the patient's patterns of behaviour that are expressed in a dissociated or split-off manner from his subjective awareness, and thus complements the total expression of his intrapsychic life in the treatment situation, and finally evolves into 'interpretation per se'. Interpretation per se implies the formulation of hypotheses regarding the unconscious meanings in the 'here and now' of the patient's material, and the relation of these unconscious meanings with the 'there and then' of the patient's unconscious, past pathogenic experiences. The analysis of the transference is 'systematic', in the sense that all emerging transference dispositions are interpreted, ideally, in the natural sequence of their emergence in the analytic situation. Gill's phrase, 'by interpretation alone', implies that the psychoanalyst abstains from measures other than helping the patient to fully understand the unconscious conflicts activated in the here and now. Thus, providing guidance about life decisions, or attempting to modify the patient's behaviour or state by means of praise, prohibition, or reward is not part of the psychoanalytic method of treatment.

The concept of 'technical neutrality' refers to the analyst's impartiality regarding both impulse and defence, with a concerned objectivity that provides a helpful collaboration with the patient's efforts to come to grips with his intrapsychic conflicts.

This definition of the nature of psychoanalytic treatment needs to be complemented with the contemporary concepts of 'transference', 'countertransference', 'acting out', and 'working through'.

An object-relations theory model of the transference and countertransference

Modern object-relations theory, further explored below, and presented in more detail in terms of particular schools in Chapter 3.2, proposes that, in the case of any particular conflict around sexual or aggressive impulses, the conflict is imbedded in an internalized object relation, i.e. in a repressed or dissociated representation of the self ('self-representation') linked with a particular representation of another who is a significant object of desire or hatred ('object representation'). Such units of self-representation, object representation, and the dominant sexual, dependent or aggressive affect linking them are the basic 'dyadic units', whose consolidation will give rise to the tripartite structure. Internalized dyadic relations dominated by sexual and aggressive impulses will constitute the id; internalized dyadic relations of an idealized or prohibitive nature the superego, and those related to developing psychosocial functioning and the preconscious and conscious experience, together with their unconscious, defensive organization against unconscious impulses, the ego. These internalized object relations are activated in the **transference** with an alternating role distribution, i.e. the patient enacts a self-representation while projecting the corresponding object representation onto the analyst at times, while at other times projecting his self-representation onto the analyst and identifying with the corresponding object representation. The impulse or drive derivative is reflected by a dominant, usually primitive affect disposition linking a particular dyadic object relation; the associated defensive operation is also represented unconsciously by a corresponding dyadic relation between a self-representation and an object representation under the dominance of a certain affect state.

For example, a conflict between unconscious aggression and unconscious guilt feelings, respectively located in id and superego, is clinically represented by manifestations of a guilt-provoking object representation relating to a guilty self (the superego defence), and an enraged self-representation attempting to attack a threatening or frustrating object representation (the id impulse). The development of the transference, therefore, consists of a sequence of activation of such impulsively determined and defensively determined internalized object relations and their systematic clarification, confrontation, and interpretation by the analyst.

The concept of **countertransference**, originally coined by Freud as the unresolved, reactivated transference dispositions of the analyst is currently defined as the total affective disposition of the analyst in response to the patient and his or her transference, shifting from moment to moment, and providing important data of information to the analyst. The countertransference, thus defined, may be partially derived from unresolved problems of the analyst, but stems as well from the impact of the dominant transference reactions of the patient, from reality aspects of the patient's life, and sometimes from aspects of the analyst's life situation that are emotionally activated in the context of the transference developments. In general, the stronger the transference regression, the more the transference determines the countertransference; thus the countertransference becomes an important diagnostic tool. The countertransference includes both the analyst's empathic identification with a patient's central subjective experience ('concordant identification') and the analyst's identification with the reciprocal object or self-representation ('complementary identification') unconsciously activated in the patient as part of a certain dyadic unit, and projected onto the analyst.[27] In other words, the analyst's countertransference implies an identification with what the patient cannot tolerate in him- or herself, and must dissociate, project, or repress.

At this point, it is important to refer to certain primitive defensive operations that were described by Klein[9] and her school in the context of the analysis of severe character pathology. Primitive defensive operations are characteristic of patients with severe personality disorders, and emerge in other cases during periods of regression. They include splitting, projective identification, denial, omnipotence, omnipotent control, primitive idealization, and devaluation (contempt). All these primitive defences centre around splitting, i.e. an active dissociation of contradictory ego (or self) experiences as a defence against unconscious intrapsychic conflict. They represent a regression to the phase of development (the first 2 to 3 years of life) before repression and its related mechanisms mentioned are established.

Primitive defensive operations present important behavioural components that tend to induce behaviours or emotional reactions in the analyst, which, if the analyst manages to 'contain' them, permit him to diagnose in himself projected aspects of the patient's experience. Particularly 'projective identification' is a process in which:

1 the patient unconsciously projects an intolerable aspect of self-experience onto (or 'into') the analyst;

2 the analyst unconsciously enacts the corresponding experience ('complementary identification');

3 the patient tries to control the analyst, who now is under the effect of this projected behaviour;

4 the patient meanwhile maintains empathy with what is projected.

This scenario is in contrast to the more mature mechanism of 'projection', secondary to repression, where there is no longer any conscious emotional contact with what is projected. Such complementary identification in the countertransference permits the analyst to identify him- or herself through his own experience with the aspects of the patient's experience communicated by means of projective identification. This information complements what the analyst has discovered about the patient by means of clarification and confrontation, and permits the analyst to integrate all this information in the form of a 'selected fact' that constitutes the object of interpretation. Interpretation is thus a complex technique that is very much concerned with the systematic analysis of both transference and countertransference.

Contemporary trends of the psychoanalytic method

Contemporary psychoanalytic technique can be seen as having evolved from a 'one person psychology' into a 'two person psychology' and then into a 'three person psychology'. The concept of 'one person psychology' refers to Freud's original analysis of the patient's unconscious intrapsychic conflicts by analysing the intrapsychic defensive operations that oppose free association. The 'two person psychology' refers to the central focus on the analysis of transference and countertransference. In the views of the contemporary intersubjective, interpersonal, and self-psychology psychoanalytic schools, the relationship between transference and countertransference is mutual, in the sense that the transference is at least in part a reaction to reality aspects of the analyst, who therefore must be acutely mindful of his contribution to the activation of the transference. The so-called 'constructivist' position regarding transference analysis assumes that it is impossible for the analyst to achieve a totally objective position outside the transference/countertransference bind.

In contrast, the contemporary 'objectivist' position, represented by the 'three person psychology' approaches of the Kleinian school, the French psychoanalytic mainstream, and significant segments of contemporary ego psychology proposes that the analyst has to divide him- or herself between one part influenced by transference and countertransference developments, and another part that, by means of self-reflection, maintains him- or herself outside this process, as an 'excluded third party', who, symbolically, provides an early triangulation to the dyadic regression that dominates transference developments. This triangulation in the treatment situation becomes particularly important in the treatment of severe personality disorders.

The 'enactment' of pathogenic past internalized object relations in the form of both transference and countertransference developments needs to be differentiated from **acting out**, the replacement of self-awareness by often dramatic, and at times, violent action. It is characteristic of patients with severe character pathology, and may occur in both patient and analyst under the influence of regression. Acting out may occur both during and outside the sessions. While it reflects an intense defensive operation and resistance, it also offers the opportunity for a very fundamental exploration of a primitive conflict, if dealt with by consistent interpretations in as much depth as possible. Acting out may also be considered an extreme, behavioural manifestation of 'enactment' as the usual experience of transference/countertransference manifestations.

The **repetition compulsion** as a resistance of the id is most probably a form of acting out as a defence against emotional containment of an extremely painful or traumatic set of experiences. **Working through** refers to the repeated elaboration of an unconscious conflict in the psychoanalytic situation. It is a major task for the analyst, who has to be alert to the subtle variation in meanings and implications of what on the surface may appear to be an endless repetition of the same conflict in the transference. The patient elaboration of the conflict that presents itself with these repetitive characteristics also implies the function of 'holding' originally described by Winnicott.[11] It consists of the analyst's capacity to withstand the onslaught of primitive transferences without retaliation, abandonment of the patient, or a self-devaluing giving up, and the maintenance of a working relationship (or 'therapeutic alliance') that addresses itself consistently to the healthy part of the patient, even when the latter is under the control of his most conflicting behaviours. Bion's concept[28] of 'containing' is complementary to 'holding', in the sense that holding deals mostly with the affective disposition of the analyst, and containing with his cognitive capacity to maintain a concerned objectivity and focus on the 'selected fact', permitting the integration in the analyst's mind what the patient can only express in violently dispersed or split-off behaviour patterns.

Dream analysis developed in the context of the method of free association, and constituted, in Freud's[29] view, a 'royal road to the unconscious'. Freud's discovery of primary process thinking derived from his method of dream analysis. By now, psychoanalytic thinking has evolved into the view that there are many 'royal roads' to the unconscious. The analysis of character defences, for example, or of particular transference complications, may be equally important avenues of entry into the patient's unconscious mind.

The technique of dream analysis consists, in essence, in asking the patient to free associate to elements of the 'manifest content' of the dream, in order to arrive at its 'latent' content, the unconscious wish defended against and distorted by the unconscious defensive mechanisms that constitute the 'dream work', and have transformed the latent content into the manifest dream. The latent content is revealed with the help of the simultaneous analysis of the way in which the dream is being communicated to the analyst, the 'day residuals' that may have triggered the dream, the unconscious conflicts revealed in it, and the dominant transference dispositions in the context of which the dream evolved. Dreams also provide some residual, universal symbolic meanings that may facilitate the total understanding of the latent content.

The **analysis of character** may be the single most important element of the psychoanalytic method in bringing about fundamental characterological change. Character analysis is facilitated by the patient's use of reaction formations, i.e. his defensively motivated character traits, as transference resistances. Thus, the activation of defensive behaviours in the transference, reflecting the patient's characterological patterns in all interpersonal interactions, facilitates both the analysis of the underlying unconscious conflicts, and in the process, the resolution of pathological character patterns. The result is an increase in the patient's autonomy, flexibility, and capacity for adaptation. Character analysis was originally developed by Reich[30] within an ego-psychology perspective, but has re-emerged in the work of Rosenfeld[31] and Steiner[32] in the analysis of 'pathological organizations' in the transference, within the Kleinian school. Gray[33] and Busch[34] within an ego-psychological perspective, have

enriched further the technique of character analysis by means of detailed exploration of particular characterological defences in the transference.

Character analysis, although not always referred to under this specific heading, constitutes a major focus of contemporary psychoanalytic treatment. In essence, its technique addresses repetitive, ego syntonic behaviour patterns in the transference, raising the patient's curiosity about their function in the relationship with the analyst, and inviting the patient to associate about this behaviour. Gradually, their exploration makes character resistances ego dystonic, and facilitates the discovery of the underlying internalized object relations condensed in these pathological character traits, both in their defensive and impulsive meanings. The question, to what extent such rigid behaviours should be analysed first, in order to free the patient's capacity for analytic work, or to what extent they should be left for later, until more fluid conflicts have been resolved, has been settled in favour of the general psychoanalytic technical principle of focusing interpretations upon what is affectively dominant in each hour.[35] Affective dominance refers once more to the 'selected fact',[36] to be interpreted. All interpretations are usually carried out from surface to depth, which in practice means first analysing the object relation activated by the need for defence before analysing the corresponding object relation activated by impulse.

The overall objective of psychoanalytic treatment is not only the resolution of symptoms and pathological behaviour patterns or characteristics, but fundamental, structural change, that is the expansion and enrichment of ego functions as the consequence of resolution of unconscious conflict and the integration of previously repressed and dynamically active id and superego pressures into ego potentialities. Such change is reflected in the increasing capacity for both adaptation to and autonomy from psychosocial demands and expectations, and an increased capacity for gratifying and successful functioning in love and work.

Derived modalities of treatment

One of the most important contributions of psychoanalytic theory and technique to the contemporary treatment of a broad spectrum of patients with severe psychopathology who, for various reasons, cannot benefit from psychoanalytic treatment proper, is the development of psychoanalytic psychotherapy, also called expressive or exploratory psychotherapy, and of supportive psychotherapy (SP) based on psychoanalytic principles. These treatments are explored below.

Psychoanalytic psychotherapy

Psychoanalytic psychotherapy may be characterized by the same basic techniques as psychoanalysis, but with quantitative modifications that, in combination, result in a qualitative shift in the nature of the treatment. Any given session of psychoanalytic psychotherapy may be indistinguishable from a psychoanalytic session, but over time, the differences emerge quite clearly. Psychoanalytic psychotherapy utilizes interpretation, but with patients with severe psychopathology, a good deal of time must be devoted to clarification and confrontation before interpretation can be effective; and interpretations of unconscious meanings in the 'here and now' occupy the foreground until late in the treatment, when genetic interpretations in the 'there and then' become useful.[15, 37]

In the treatment of patients with severe character pathology, transference analysis is the essential focus of psychoanalytic psychotherapy from the very beginning; it must be modified, however, by active interpretive connection of transference analysis with exploration in depth of the patient's daily life situation, an approach made necessary by the predominance of primitive defence operations in these patients. Splitting operations in particular tend to dissociate the therapeutic situation from the patient's external life, and may lead to severe, dissociated acting out either in the sessions or outside the sessions. Therefore, interpretive linkage between the patient's external reality and transference developments in the hours becomes central.

In order to enable the therapist to analyse transference developments in sufficient depth, psychoanalytic psychotherapy requires a minimum frequency of two sessions per week. It is usually carried out in 'face-to-face' sessions.

Technical neutrality is an essential feature of analysis in general, but in the treatment of patients with severe character pathology, the need to set limits may necessitate abandoning neutrality again and again, in order to control life- or treatment-threatening acting out. The self-perpetuating nature of acting out in these cases may prove impossible to resolve interpretively without such structuring or setting limits. Whenever the analyst has to abandon technical neutrality to protect the patient or the treatment, it is essential to explore the episode immediately. The transference implications of the therapist's structuring behaviour must be laid out, followed by the analysis of the transference implications of the patient's behaviour that necessitated the imposition of limits or the initiation of a new structure in the treatment; this in turn is followed by the gradual resolution of the structure or limit setting by interpretive means, thus restoring technical neutrality. In short, technical neutrality in psychoanalytic psychotherapy is an ideal working state that is again and again preventively abandoned and interpretively reinstated.[15, 37, 38]

Supportive psychotherapy

Supportive psychotherapy based on psychoanalytic theory may also be defined in terms of the three major techniques of interpretation, transference analysis, and technical neutrality. Supportive psychotherapy utilizes the preliminary steps of interpretive technique, i.e. clarification and confrontation, but rarely uses interpretation per se. It seeks to strengthen the ego by bolstering adaptive compromises between impulse and defence through the provision of cognitive support in the form of information, persuasion and advice, and emotional support via suggestion, reassurance, encouragement, and praise. Supportive psychotherapy may call upon direct environmental intervention by the therapist, relatives, or other mental health personnel engaged in auxiliary therapeutic functions.[39]

While the transference is seldom interpreted in supportive psychotherapy, it is not ignored either. Careful attention to transference developments helps the therapist to analyse any maladaptive transference developments, to call the patient's attention to the reproduction with the therapist of pathological interactions the patient generally engages in with significant others, and to encourage the patient to reduce such pathological behaviours. Pointing out the distorted, unproductive, destructive, or confusing nature of the patient's behaviour is accompanied by clarifying the patient's conscious reasons for his behaviour, followed by the transfer or 'export' of the knowledge thus achieved to the patient's relationships outside the treatment. In short, supportive psychotherapy includes the clarification, reduction, and 'export' of the transference, thus contributing to

the re-educative functions of supportive psychotherapy together with the direct cognitive and affective support of adaptive combinations of impulse and defence, and direct supportive environmental interventions.

Technical neutrality is systematically abandoned in supportive psychotherapy, the therapist taking a stance alternatively on the side of the ego, superego, id, or external reality, according to which agency represents, at a certain point, the more adaptive potential for the patient. The main dangers, of course, in supportive psychotherapy are, on the one hand, infantilizing the patient by an excessively supportive stance, and, on the other, countertransference acting out as a consequence of the abandonment of the position of technical neutrality. The therapist carrying out supportive psychotherapy, therefore, needs a heightened awareness of the risk of these complications. Like psychoanalytic psychotherapy, supportive psychotherapy is carried out in 'face-to-face' sessions. It has the advantage of considerable flexibility regarding its frequency, from several sessions per week, to one session a week, or one or two sessions per month, according to the urgency of the patient's present difficulties, the long-range objectives of the treatment, and the patient's ability to tolerate and use the relationship with the therapist.

Indications and contraindications for psychoanalysis and derived psychotherapies

The **indications** for these three modalities of treatment remain controversial: with the recognition of the limitations of psychoanalysis in many cases with severe, chronic, life-threatening self-destructive behaviour, such as chronic suicidal behaviour, severe eating disorders, dependence upon drugs or alcohol, and severely antisocial behaviour, psychoanalytic psychotherapy has proven to be a highly effective treatment for many but by no means all patients with these conditions. The differential diagnosis of a spectrum of severity of antisocial behaviour and those cases of severe self-destructive and antisocial behaviour who are amenable to treatment with psychoanalytic psychotherapy has been one of the important side-products of the psychoanalytic exploration of these cases.[37]

Supportive psychotherapy, originally conceived of as the treatment of choice for patients with severe personality disorders, now may be considered the alternative treatment for those patients with severe personality disorders who are unable to participate in psychoanalytic psychotherapy. The Menninger Foundation Psychotherapy Research Project showed that patients with the least severe psychopathological disturbances tend to respond very positively to all three modalities derived from psychoanalytic theory, although best to standard psychoanalysis.[40]

Standard psychoanalysis is the treatment of choice for patients with neurotic personality organization, that is with good identity integration and a repertoire of defences centring on repression along with sufficient severity of illness to warrant such a major therapeutic intervention. Psychoanalysis has also expanded its scope to some of the severe personality disorders, particularly a broad spectrum of patients with narcissistic personality disorders, patient with mixed hysterical-histrionic features, and selected cases of patients with severe paranoid, schizoid, and sado-masochistic features.

We are still lacking systematic studies of the relationship between particular types of psychopathology and outcome with the various psychotherapeutic treatments derived from psychoanalytic theory. As a tentative generalization it may be stated that there is a definite relationship between outcome and the severity of illness in any diagnostic category. The least severe cases will respond favourably to either brief psychoanalytic psychotherapy, supportive psychotherapy, or psychoanalysis. Psychoanalysis represents the opportunity for most improvement if the severity of the case warrants psychoanalytic treatment. For cases of neurotic personality organization of moderate severity, psychoanalysis is the treatment of choice; definitely less can be expected in these cases from psychoanalytic psychotherapy. For the most severely ill patients (those with severe identity diffusion, predominance of primitive defences centring on splitting, and general 'ego weakness') psychoanalytic psychotherapy is the treatment of choice, with supportive psychotherapy a second choice if psychoanalytic psychotherapy is contraindicated. A few such cases may be able to participate in psychoanalysis and benefit from it.

In all cases, individualized **contraindications** for the respective treatment are important: in the case of psychoanalysis, individual contraindications depend on the questions of ego strength, motivation, introspection or insight, secondary gain of illness, intelligence, and age. In the case of psychoanalytic psychotherapy, secondary gain, the impossibility of control of life- or treatment-threatening acting out, limited intelligence, significant antisocial features, and a desperate life situation may constitute individual contraindications, particularly when they occur in combination. When psychoanalytic psychotherapy is contraindicated for such reasons, supportive psychotherapy becomes the treatment of choice. Participation in supportive psychotherapy requires a sufficient capacity for commitment to an ongoing treatment arrangement, and the absence of severe antisocial features as minimal individual requirements. This is not meant to be a complete list, but an illustration of the kind of criteria that become dominant in the individual decisions regarding the selection of the treatment and its contraindications.

Psychoanalytic object-relations theories: overview and critique

Given the centrality of object-relations theory in practically all contemporary psychoanalytic formulations and treatment approaches, the following summary is included. It should help the reader to further clarify the references made earlier to this theory.

Psychoanalytic object-relations theories may be defined as those that place the internalization, structuralization, and reactivation in the transference and countertransference of the earliest dyadic object relations at the centre of their clinical formulations, and of their thinking about motivation, pathogenesis, development, and psychic structure. Internalization of object relations refers to the concept that, in all interactions of the infant and child with the significant parental figures, what the infant internalizes is not merely an image or representation of the other ('the object' of fear, hatred, or desire), but the relationship between the self and the other, in the form of a self-image or self-representation linked to an object image or object representation by the affect that dominates their interaction. This internal structure replicates in the intrapsychic world both real and phantasied relationships with significant others.

Several major issues separate different object-relations theories, the most important of which is the extent to which the theory is perceived as harmonious with or in opposition to Freud's traditional drive theory: i.e. whether object relations are seen as replacing drives as the motivational system for human behaviour. From this perspective, Klein,[9,21] Mahler *et al.*,[13] and Jacobson[12] occupy one pole.

They combine Freud's dual-drive theory with an object-relations theory. For Fairbairn,[10] and Sullivan,[41] on the other hand, object relations themselves replace Freud's drives as the major motivational system. Here, the establishment of gratifying object relations in itself constitutes the major motivational system. Contemporary interpersonal psychoanalysis as represented by Greenberg and Mitchell,[42] based upon an integration of principally Fairbairnian and Sullivanian concepts, asserts the essential incompatibility between drive- and object relations-based models of psychic motivational systems. Winnicott,[11] Loewald,[43] Sandler,[44] and Sandler and Sandler,[45] (each for different reasons) maintain an intermediate posture; they perceive the affective frame of the infant–mother relationship as a crucial determinant in shaping the development of drives. While adhering to Freud's dual-drive theory, Kernberg[15] considers drives supraordinate motivational systems, while affects are their constituent components.

A related controversy has to do with the origin and role of aggression as motivator of behaviour. Those theoreticians who reject the idea of inborn drives,[41] or equate libido with the search for object relations,[10] conceptualize aggression as secondary to the frustration of libidinal needs, particularly traumatic experiences in the early mother–infant dyad. Theoreticians who adhere to Freud's dual-drive theory, in contrast, believe aggression is inborn and plays an important part in shaping early interactions: this group includes Klein in particular, and to some extent Winnicott, and ego-psychology object-relations theoreticians such as Kernberg.[37] Finally, contrast may be made between object-relations theories and French approaches, both Lacanian and (non-Lacanian) mainstream psychoanalysis. The French psychoanalytic mainstream,[46, 47] has maintained close links with traditional psychoanalysis, including the British object-relations theories. Insofar as Lacan[48] conceptualizes the unconscious as a natural language and focuses on the cognitive aspects of unconscious development, he underemphasizes affect—a dominant element of object-relations theories. At the same time, however, in postulating a very early Oedipal structuralization of all infant–mother interactions, Lacan emphasizes archaic Oedipal developments, which implicitly links his formulations with those of Kleinian object-relations theory in general. French mainstream analysis also focuses on archaic aspects of Oedipal developments, but places a traditional emphasis on Freud's dual-drive theory and on the affective nature of the early ego-id. As neither French mainstream nor Lacanian psychoanalysis spells out specific structural consequences of dyadic internalized object relations, however, neither would fit the definition that frames the field of object-relations theory as proposed in this chapter.

All object-relations theories focus heavily on the enactment of internalized object relations in the transference, and on the analysis of countertransference in the development of interpretive strategies. They are particularly concerned with severe psychopathologies, including those psychotic patients who are approachable with psychoanalytic techniques, borderline conditions, severe narcissistic character pathology, and the perversions ('paraphilias'). Object-relations theories explore primitive defensive operations and object relations both in cases of severe psychopathology and at points of severe regression with all patients, regarding such exploration as essential in facilitating transference analysis and conflict resolution.

The contemporary re-evaluation of Freud's dual-drive theory that has occurred mostly in France is relevant to the relationship between object-relations theory and drive theory. Perhaps particularly the work of Laplanche[49] and Green[46] has emphasized the central importance of unconscious destructive and self-destructive drive manifestations in the form of attacks on object relations, and the central role of unconscious erotization in the mother–infant relationship in libidinal development, all of which tends to link drive theory and object-relations theory in intimate ways.

Another important development within psychoanalytic theory has been the growing emphasis on affects as primary motivators, and the centrality of the communicative functions of affects in early development, particularly the infant–mother relationship.[50] This emphasis has linked affect theory and object-relations theory quite closely, despite the persistent controversy between those who see affect, particularly peak affect states, as essential representatives of the drives,[50] and those who stress the psychophysiological nature of the affective response, and attempt to replace drive theory with an affect theory.[51]

The basic (self-representation–object representation) units of internalized object relations include the representative affects, or else, the constituent affective components of the drives. One might say that the affect of sexual excitement is the central affect of libido, in the same way as the affect of primitive hatred constitutes the central affect of the aggressive or death drive. The id is conceptualized in this object-relations theory model as the sum total of repressed, desired, and feared primitive object relations. The gradual integration of successive layers of persecutory and idealized, prohibitive and demanding, internalized object relations become part of the primitive superego, while internalized object relations activated in the service of defence consolidate as part of an integrated self-structure surrounded by integrated representations of significant others. In short, the id or dynamic unconscious, the superego, and the ego are constituted by different constellations of internalized object relations, so that the development of the drives and the development of the psychic apparatus—the tripartite structure—occur hand in hand.

Perhaps the most important practical implication of object-relations theory is the conception of identification as a series of internalization processes of dyadic units of self-representation and object representation linked by a dominant affect state, ranging from earliest introjections to identifications per se, to the development of complex identity formation. Each step includes the internalizing of both self and object representations and their affective interactions under the conditions that prevail at different developmental levels.

In the transference of healthier patients, with a well-consolidated ego identity, the diverse self-representations are relatively stable in their coherent mutual linkage. This fosters the relatively consistent projection onto the analyst of the object representation aspect of the enacted object relationship. In contrast, patients with severe identity diffusion lack such linkage of self-representations into an integrated self. They tend to alternate rapidly between projection of self and object representations in the transference, so that the analytic situation seems chaotic. Systematic interpretation of how the same internalized object relation is enacted again and again with rapid role reversals between patient and analyst makes it possible to clarify the nature of the unconscious object relation, and the double splitting of self-representation from object representation and idealized from persecutory object relations. This process of interpretation promotes integration of the split representations, which characterize severe psychopathology and account for the marked instability of the emotions, behaviour, and interpersonal relationships of these patients.

Kernberg[37] proposes that affects are the primary motivational system and that, internalized or fixated as the very frame of internalized 'good' and 'bad' object relations, affects are gradually integrated into libidinal and aggressive drives to form hierarchically supraordinate motivational systems. In other words, primitive affects are the 'building blocks' of the drives. He sees unconscious intrapsychic conflicts as always between the following:

1 certain units of self and object representations under the impact of a particular drive derivative (clinically, a certain affect disposition reflecting the drive derivative side of the conflict);
2 contradictory or opposing units of self and object representations and their respective affect dispositions reflecting the defensive structure.

Unconscious intrapsychic conflicts are never simply between impulse and defence; rather, both impulse and defence find expression, respectively, through certain internalized object relations.

In patients with borderline personality organization and severe conflicts around early aggression, splitting mechanisms stabilize such dynamic structures within an ego-id matrix and permit the contradictory aspects of these conflicts to remain at least partially conscious, in the form of primitive, mutually split-off, idealized, and persecutory transferences. In contrast, patients with neurotic personality organization present impulse–defence configurations that contain specific unconscious wishes of an integrated though infantile self, reflecting sexual and aggressive drive derivatives embedded in unconscious phantasies relating to the Oedipal objects. Repressed unconscious wishes, however, always come in the form of corresponding units composed of self-representation and object representation and affect linking them.

Patients with neurotic personality organization present well-integrated superego, ego, and id structures; within the psychoanalytic situation, the analysis of resistances brings about the activation, in the transference, first of relatively global characteristics of these structures, and later, the internalized object relations of which they are composed. Oedipal conflicts dominate the dynamic unconscious of these patients. The analysis of drive derivatives occurs in the context of the analysis of the relation of the patient's infantile self to significant parental objects as projected onto the analyst.

Patients with severe personality disorders or borderline personality organization, in contrast, show a predominance of psychic representations of pre-Oedipal conflicts, with pre-Oedipal aggression, in particular, condensed with representations of the Oedipal phase. Conflicts are not predominantly repressed and therefore unconsciously dynamic: rather, they are avoided by being represented in mutually dissociated ego states reflecting the defence of primitive dissociation or splitting. The activation of primitive object relations that predate the consolidation of ego, superego, and id is manifest in the transference as apparently chaotic affect states, which have to be analysed in sequential steps as follows:

1 the clarification of a dominant primitive object relation in the transference, with its corresponding self and object representation, and the dominant affect linking them;
2 the analysis of the alternative projection of self and object representation onto the therapist, while the patient identifies with a reciprocal self or object representation of this object relationship, leading to the patient's gradual capacity to become aware of his identification with an object in that relationship;
3 the interpretive integration of mutually split-off, idealized and persecutory 'part object' relations with the characteristics mentioned.

This analysis may gradually bring about a transformation of mutually split, ('part object') relations into 'total object' relations, or of primitive transferences (largely reflecting Mahler's stages of development that predate object constancy) into the advanced transferences of the Oedipal phase. In other words, a gradual integration of self-representations into an integrated self-concept, and a parallel integration of significant object representations into integrated concepts of significant others develop first in the transference, and later generalize in the patient's relations with significant others. The analyst's exploration of his or her countertransference, including concordant and complementary identifications in the countertransference,[27] facilitates transference analysis; and the analysis of primitive defensive operations, particularly splitting and projective identification in the transference, also contributes to strengthening the patient's ego.

Treatment results: research on outcome

The psychoanalytic profession has been slow in developing systematic research on treatment process and results, let alone controlled randomized comparison of treatment methods evaluating efficacy and efficiency. The reasons are multiple: the complexity of the psychoanalytic treatment, and the changes in its technique; the long duration of treatment, making systematic research, and controlled comparison with other treatment methods difficult; the private nature of psychoanalytic exploration in the context of patients' regression, and the related concerns over disturbing the therapeutic relationship by recording or direct observation. In addition, the general methodology of psychotherapy research evolved to a degree of sophistication applicable to the evaluation of psychoanalytic treatment only in recent decades. With all these reservations, significant progress has been made, and outcome studies are beginning to be available.

The Menninger Psychotherapy Research Project, a naturalistic study comparing psychoanalysis, psychoanalytic psychotherapy, and SP, showed psychoanalysis to be the most effective of these approaches with patients presenting relatively good ego strength, while patients with severe ego weakness—what nowadays would be described as presenting severe personality disorders or borderline personality organization—improved most with psychoanalytic psychotherapy.[39] This research also showed how important supportive elements were throughout all modalities of treatment.[52] A comprehensive review of outcome studies on psychoanalytic psychotherapy and psychoanalysis by Bachrach *et al.*[53] concluded that the improvement rates are in the 60 to 90 per cent, but it also pointed to limitations and problems in the methodology utilized.

Recently, studies regarding the treatment process and outcome of psychoanalysis and psychoanalytic psychotherapy have become more precise in defining the specific treatment variables of psychotherapeutic and psychoanalytic treatments, and several systematic studies on psychoanalytic psychotherapies and psychoanalysis are in progress.[54] A recent study by the Stockholm Outcome of Psychoanalysis and Psychotherapy Project has found, on the basis of a relatively large patient population, that psychoanalytic treatment, in comparison with psychoanalytic psychotherapy, obtained a significantly higher degree long-range symptomatic improvement.[55] The extent to which the psychotherapist had years of experience linked with

appropriate, long-term supervisory experiences, i.e. an 'experiential learning cluster', was related to treatment outcome, in the sense that those therapists with long experiences in doing teaching or supervision of psychotherapy had a significantly better outcome than therapists who only had been in supervision or personal therapy for long periods. It also appeared that the maintenance of a rigid 'psychoanalytic' attitude as part of a psychoanalytic psychotherapy was not as effective as a more flexible shift in techniques in psychotherapy, but not in analysis proper.[56] A manualized psychoanalytic psychotherapy for a specific patient population, namely, the psychotherapy research project of the Cornell Personality Disorders Institute's manualized treatment for borderline patients has provided evidence for the efficacy of the treatment with severely ill patients. This treatment, called Transference Focused Psychotherapy (TFP) was found to be more effective than treatments as usual (TAU) for borderline patients,[38] and, compared to dialectic behaviour therapy (DBT), and supportive psychotherapy (SP) in a randomized controlled study, proved as effective as DBT and SP in improving depression, anxiety, global functioning, and social adjustment at the end of 1 year of treatment. It also was more effective in reducing aggression than DBT and SP, and the only one to improve reflective function (RF), an index of mentalization, that is the patient's capacity for self-reflection and appropriate assessment of others in depth.[57,58] (Bateman and Fonagy[59,60] have found that mentalization-based therapy (MBT), another form of psychoanalytic psychotherapy was more effective than treatment as usual (TAU) in the treatment of borderline patients in a day hospital setting. Further developments of MBT research will be referred to in Sections 3, 5, and 6.

In summary, process research has predated outcome research on psychoanalysis and derived psychotherapies; major efforts at outcome research are being made, and should contribute to clarify the effects, not only of psychoanalysis proper, but also of the derived psychotherapeutic approaches now being carried out in clinical practice.

Further information

Cooper, A.M. (ed.) (2006). *Contemporary psychoanalysis in America: leading analysts present their work*. American Psychiatric Publishing, Washington, DC.

Kernberg, O.F. (2004). *Contemporary controversies in psychoanalytic theory, techniques, and their applications*. Yale University Press, New Haven.

Laplanche, J. and Pontalis, J.-B. (2006). *The language of psychoanalysis*. Karnac Books, London.

Skelton, R.M. (ed.) (2006). *The Edinburgh international encyclopedia of psychoanalysis*. Edinburgh University Press, Edinburgh.

References

1. Freud, S. (1963). *Introductory lectures on psycho-analysis. Standard edition*, Vol. 16, pp. 243–463. Hogarth Press, London.
2. Freud, S. (1964). *An outline of psycho-analysis. Standard edition*, Vol. 23, pp. 141–207. Hogarth Press, London.
3. Breuer, J. and Freud, S. (1955). *Studies on hysteria. Standard edition*, Vol. 2, pp. 3–311. Hogarth Press, London.
4. Freud, S. (1953). *Three essays on the theory of sexuality. Standard edition*, Vol. 7, pp. 125–245. Hogarth Press, London.
5. Freud, S. (1961). *The ego and the id. Standard edition*, Vol. 19, pp. 3–66. Hogarth Press, London.
6. Freud, S. (1957). *Mourning and melancholia. Standard edition*, Vol. 14, pp. 237–58. Hogarth Press, London.
7. Freud, S. (1955). *Beyond the pleasure principle. Standard edition*. Hogarth Press, London.
8. Freud, S. (1961). *Civilization and its discontents. Standard edition*, Vol. 21, pp. 57–145. Hogarth Press, London.
9. Klein, M. (1952). Notes on some schizoid mechanisms. In *Developments in psycho-analysis* (ed. J. Riviere), pp. 292–320. Hogarth Press, London.
10. Fairbairn, W.R.D. (1954). *An object-relations theory of the personality*. Basic Books, New York.
11. Winnicot, D. (1965). *The maturational processes and the facilitating environment*. International Universities Press, New York.
12. Jacobson, E. (1964). *The self and the object world*. International Universities Press, New York.
13. Mahler, M., Pine, F., and Bergman, A. (1975). *The psychological birth of the human infant*. Basic Books, New York.
14. Freud, S. (1957). *Instincts and their vicissitudes. Standard edition*, Vol. 14, pp. 109–40. Hogarth Press, London.
15. Kernberg, O. (1984). *Severe personality disorders. Psychotherapeutic strategies*. Yale University Press, New Haven.
16. Kohut, H. (1971). *The analysis of the self*. International Universities Press, New York.
17. Freud, S. (1964). *Splitting of the ego in the process of defense. Standard edition*, Vol. 23, pp. 273–4. Hogarth Press, London.
18. Abraham, K. (1920). Manifestations of the female castration complex. In *Selected papers on psycho-analysis*, pp. 338–69. Brunner/Mazel, New York.
19. Freud, S. (1957). *On narcissism. Standard edition*, Vol. 14, pp. 69–102. Hogarth Press, London.
20. Abraham, K. (1979). A particular form of neurotic resistance against the psychoanalytic method. In *Selected papers on psycho-analysis*. (1919), pp. 303–11. Brunner/Mazel, New York.
21. Klein, M. (1957). *Envy and gratitude*. Basic Books, New York.
22. Rosenfeld, H. (1964). On the psychopathology of narcissism: a clinical approach. *International Journal of Psycho-Analysis*, **45**, 332–7.
23. Grunberger, B. (1979). *Narcissism: psychoanalytic essays*. International Universities Press, New York.
24. Kernberg, O.F. (1975). *Borderline conditions and pathological narcissism*. Jason Aronson, Inc., New York.
25. Freud, S. (1953). *Fragment of an analysis of a case of hysteria. Standard edition*, Vol. 7, pp. 3–122. Hogarth Press, London.
26. Gill, M. (1954). Psychoanalysis and exploratory psychotherapy. *Journal of American Psychoanalytic Association*, **2**, 771–97.
27. Racker, H. (1957). The meaning and uses of countertransference. *Psychoanalytic Quarterly*, **26**, 303–57.
28. Bion, W.R. (1967). *Second thoughts. Selected papers on psychoanalysis*. Basic Books, New York.
29. Freud, S. (1953). *The interpretation of dreams. Standard edition*, Vol. 4, pp. 1–338; Vol. 5, pp. 339–625. Hogarth Press, London.
30. Reich, W. (1972). *Character analysis*. Farrar, Straus, & Giroux, New York.
31. Rosenfeld, H. (1987). *Impasse and interpretation*. Tavistock, New York.
32. Steiner, J. (1993). *Psychic retreats*, Routledge, London.
33. Gray, P. (1994). *The ego and analysis of defense*. Aronson, Northvale, NJ.
34. Busch, F. (1995). *The ego at the center of clinical technique*. Jason Aronson, Northvale, NJ.
35. Fenichel, O. (1941). *Problems of psychoanalytic technique*. Psychoanalytic Quarterly, Inc., Albany.
36. Bion, W.R. (1967). *Second thoughts. Selected papers on psychoanalysis*. Basic Books, New York.
37. Kernberg, O.F. (1992). *Aggression in personality disorders and perversion*. Yale University Press, New Haven.
38. Clarkin, J.F., Foelsch, P.A., Levy, K.N., *et al.* (2001). The development of a psychodynamic treatment for patients with borderline personality

disorder: a preliminary study of behavioral change. *Journal of Personality Disorders*, **15**, 487–95.
39. Rockland, L.H. (1989). *Supportive therapy: a psychodynamic approach*. Basic Books, New York.
40. Kernberg, O.F., Burnstein, E., Coyne, L., *et al.* (1972). Psychotherapy and psychoanalysis: final report of the menninger foundation's psychotherapy research project. *Bulletin Menninger Clinic*, **36**, 1–275.
41. Sullivan, H. (1953). *The interpersonal theory of Psychiatry*. Norton, New York.
42. Greenberg, J.R. and Mitchell, S.A. (1983). *Object relations in psychoanalytic theory*. Harvard University Press, Cambridge, MA.
43. Loewald, H.W. (1960). On the therapeutic action of psycho-analysis. *International Journal of Psycho-Analysis*, **41**, 16–33.
44. Sandler, J. (1987). *From safety to superego: selected papers of Joseph Sandler*. Guilford Press, New York.
45. Sandler, J. and Sandler, A.M. (1998). *Internal objects revisited*. Karnac Books, London.
46. Green, A. (1993). *Le Travail du Négatif*. Paris: Les Editions de Minuit.
47. Laplanche, J. (1992). *Seduction, translation, drives*. Psychoanalytic Forum. Institute of Contemporary Arts, London.
48. Lacan, J. (1966). *Ecrits*. Editions du Seuil, Paris.
49. Laplanche, J. (1987). *Nouveaux fondements pour la psychanalyse*. Presses Universitaires de France.
50. Krause, R. (1998). *Allgemeine Psychoanalytische Krankheitslehre*, Vols 1 and 2. Kohlhammer, Stuttgart.
51. Mitchell, S. (1988). *Relational concepts in psychoanalysis*. Harvard University Press, Cambridge, MA.
52. Wallerstein, R. (1986). *Forty-two lives in treatment: A study of psycho-analysis and psychotherapy*. Guilford Press, New York.
53. Bachrach, H.M., Weber, J.J., and Murray, S. (1985). Factors associated with the outcome of psychoanalysis. Report of the Columbia Psychoanalytic Research Center (IV). *International Review of Psychoanalysis*, **12**, 379–89.
54. Fonagy, P., Kächele, H., Krause, R., *et al.* (eds.) (1998). *An open door review of outcome studies in psychoanalysis*. International Psychoanalytic Association, London.
55. Sandell, R., Blomberg, J., and Lazar, A. (1997). When reality doesn't fit the blueprint: doing research on psychoanalysis and long-term psychotherapy in a public health service program. *Psychotherapy Research*, **7**, 333–44.
56. Sandell, R., Schubert, J., Blomberg, J., *et al.* (1997). The influence of therapist factors on outcomes of psychotherapy and psychoanalysis in the Stockholm Outcome of Psychotherapy and Psychoanalysis Project (STOPP). Paper presented at the Annual Meeting of the Society for Psychotherapy Research, Geilo, Norway.
57. Clarkin, J. F., Levy, K.N., and Schiavi, J.M. (2005). Transferenced focused psychotherapy: development of a psychodynamic treatment for severe personality disorders. *Clinical Neuroscience Research*, **4**, 379–86.
58. Clarkin, J.F., Yeomans, F., and Kernberg, O.F. (2006). *Psychotherapy for borderline personality: focusing on object relations*. American Psychiatric Publishing, Washington, DC.
59. Bateman, A. and Fonagy, P. (2004a). *Psychotherapy for borderline personality disorder: Mentalization-based treatment*. Oxford University Press, New York.
60. Bateman, A. and Fonagy, P. (2004b). Mentalization-based treatment of BPD. *Journal of Personality Disorders*, **18**, 36–51.

3.2

Object relations, attachment theory, self-psychology, and interpersonal psychoanalysis

Jeremy Holmes

Despite many splits and schisms, dating back to Adler and Jung's early break with Freud, there has been an enduring attempt within psychoanalysis to hold to a central psychodynamic vision and to find common ground between differing theoretical and clinical approaches. The aim of this chapter is to describe the work of some of the major figures who have extended and developed Freud's ideas, pointing to areas of both conflict and convergence, and, wherever possible, to relate their concepts to the everyday practice of psychiatry.

From drive theory to object relations

Psychoanalysis started its life as a 'drive theory' or 'dual instinct' theory—the idea that mental life and its pathologies could be understood in terms of the interplay between the erotic and death drives, and the ways in which these were repressed, or expressed either covertly via 'conversion', or directly. As Freud's thought evolved, so new paradigms began to emerge. Drive theory had little to say about relationships: other people appear merely as satisfiers or thwarters of an individual's instinctual needs. Freud began to ask how children, and later adults, reconciled their own wishes and desires—their drives or instincts—with those of their caregivers and peers. Struggling with this problem, while remaining within the confines of drive theory, he now differentiated between self-love, or narcissism, and other, or 'anaclitic', love, directed outwards. In this model, the individual gradually emerges from egg-like self-absorption and healthy narcissism into the world of relationships.

A further push towards a more relational theory came from Abraham, later to become Melanie Klein's analyst, who noticed the parallels between the phenomena of grief and depression. The intense psychic pain and disruption associated with a loss suggested a much more intimate connection between relationships and the architecture of the psyche than drive theory would allow. 'The unconscious' is not so much a repository of drives and desires, but an inner world populated by significant others or 'objects'. The self is forged out of these 'objects' with whom the individual has or has had important relationships: 'the shadow of the object falls on the ego'.[1] A further theoretical move arose from considering the origins of conscience and ideals. It is a matter of observation that much of development depends on processes of imitation and identification. The developing child internalizes, or 'introjects', his or her parent's values and standards. How, and where in the psyche, does this process take place? In Freud's 'tripartite model', the 'superego', alongside the ego (i.e. executive and experienced self) and the id (the locus of desire and dreaming), is the focus for these internalized parental values and aspirations. The inner world now contained not just 'objects', but value-based relations between them: prohibitions, encouragements, injunctions, and gratifications. Much of post-Freudian theory consists of attempts to develop and elaborate these ideas.

Object relations 1: Klein, Fairbairn, and their successors

This was the state of theoretical play in psychoanalysis when **Melanie Klein** first burst on to the scene in the late 1920s. Like Freud, her work can be divided into a number of phases.[2,3]

Psychoanalysis is concerned with early mental life, which it sees as the basis for much adult psychopathology. But how do we gain access to the thought processes of small children, whose verbal and introspective capacities are limited or non-existent? Klein's great technical innovation was the introduction of play therapy. She provided her little patients with play materials—paper and pencils, a doll house with figures, a sandpit, and farmyard animals—and observed the pictures and games which the children set-up, making her interpretations around them. She used the methods of dream interpretation to formulate her ideas. What she observed in play—movement of figures in and out; bringing things together, often violently; separation and disruption—she took to represent the workings of the child's mind. Still deeply influenced by drive theory, and by Freud's insistence on the pre-eminence of sexuality and castration anxiety, she found sexual and aggressive meanings in all that was presented to her. Every vertical line or orifice-shaped circle drawn had a sexual significance; every conjoining or emitted sound stood for parental intercourse, by which the child was both fascinated and frightened. Exploration and the drive to know were seen

as an expression of the desire to possess the mother's body, and inhibitions of learning as manifestations of castration anxiety.

Here Klein began to depart from Freud. For him the Oedipus complex arose around the age of three, when the child begins to observe his or her parents' relationship and to feel such emotions as passionate love, envy, fear, and jealous vengefulness. Klein, by contrast, saw Oedipal phenomena as arising much earlier in development. For example, the infant may experience weaning as a punishment or symbolic castration, and believe that his mother's breast in his mouth has been displaced by the paternal penis in her vagina. Two other aspects of Kleinian thought emerge from this. First, in Klein's schema the infant has an instinctual knowledge of the body and its relationships. There appears to be a reservoir of unconscious phantasy, which she saw as the mental accompaniment of bodily function: phantasies about the breast, the mouth, the penis, the vagina, and their relationships that could not have arisen from direct observation, and therefore must be present from within, as correlates of the child's bodily sensations, which Klein saw as dominating the early years of life. Unconscious phantasies are akin to Jungian archetypes or perhaps the 'language acquisition devise' postulated by linguists: preformed mental constructs unconsciously shaping experience and patterns of relationship.

Second, and closely related to unconscious phantasy, is the idea of internal objects—initially body parts, and later 'whole objects' that are salient to emotional life—the mother and her breast, the father and his penis, bellies and their contents such as unborn babies, faeces, and sphincters. These objects are endowed with motivational properties reflecting the infant's emotional life, which Klein saw as dominated by persecutory fears. The 'death instinct' ensures that the child reacts to frustration with overwhelming feelings of hatred and destructiveness. These feelings are then projected outwards on to the objects in the child's emotional environment, which are in turn reintrojected to populate the inner world. To preserve good feelings from these terrifying bad objects, the child also projects goodness outwards. Thus a radical split arises between good and bad experiences, which are attributed to good and bad objects: 'in the very earliest stage every unpleasant stimulus is related to the 'bad', denying, persecuting breasts, every 'good' experience to the 'good' gratifying breasts'.[2]

Klein depicts early emotional life as dominated by the infant's fears of annihilation from without, and the use of the mechanisms of splitting and projection to reduce these fears. She postulated the onset of a new type of anxiety towards the end of the first year of life. Here the infant is beginning to bring the image of the 'good' and the 'bad' breast together, and to realize that they are one and the same. With weaning, the child experiences his first major loss. Now 'depressive' anxiety comes into the picture. The child believes that he is responsible for the loss, and that he has destroyed the good object with his aggression and sadism. He begins to feel guilt and remorse, and wants to repair the damage he believes he has inflicted on his objects. His attempts at creation, the gifts he offers, and the charm with which he approaches his caregivers are all motivated by this sense of depressive despair and the wish to make reparation.

Klein thus described a developmental sequence: inherent aggression, annihilation anxiety, projection and splitting of the object into good and bad, loss, bringing together the split objects, depressive despair, concern for the object, and finally reparation. For her this was a description of normal development, and she saw pathological states as resulting from developmental arrest along this line. The fulcrum of this sequence is the movement from what, drawing on Fairbairn's term (see below) Klein now called the '**paranoid–schizoid' position** (**PSP**) to the '**depressive' position**' (**DP**), a movement from splitting, blaming, and avoidance, to integration, responsibility, and concern for the object (see Hobson *et al.*[4] for objective evidence of the validity of the PSP–DP distinction). Klein saw the struggle between PSP and DP as a lifelong process, an equilibrium driven one way or the other depending on life experience and constitutional endowment.

Klein was generally rather unconcerned about the impact of external reality on psychological development (a point which, as we shall see, stimulated Bowlby's divergence from her ideas). To the extent that she did consider the real as opposed to the phantasized role of the parents, it was as benign figures whose job it is to mitigate the strength of the infant's need to hate, project, and split. An important late theoretical contribution, however, concerned the role of envy in psychic life. One of the strengths of a psychoanalytical approach to psychotherapy is that it takes seriously the phenomenon of resistance, and the fact that psychic growth is usually hard-won, often with much backsliding and self-defeatingness. With her emphasis on the dark side of human nature, Klein realized that the infant may feel persecuted not just by frustration and separation, but also by the very capacity of the caregiver to satisfy his needs. The breast upon which the baby depends for satisfaction and pleasure can also be a source of envy and hatred in its plenitude and ability to create dependency. This **envy** then becomes a basis for destructiveness within psychotherapy, and more generally: an explanation, perhaps, for the graffiti which inevitably appear on beautiful buildings, or, at times, the fact that patients attack and seem to want to destroy the very help that is offered to them.

Another key Kleinian concept is that of **projective identification** (**PI**), a difficult and perhaps misnamed concept, coined almost casually by Klein in an attempt to describe how parts of the ego may be split-off and projected not just *on to* objects in the environment as visualized in Freud's notion of projection, but *into* them. As originally conceived by Klein PI referred to the solipsistic world of the infant described above, in which unbearable feelings of rage and hatred are split-off, projected into the breast, which is then perceived by the child as 'having' properties that in fact originated in the self. Projective identification here is a form of misperception or delusional perception, which can be used both to explain the fact that normal adults' experience of the world is inevitably coloured by their emotional state (the gloomy or rose-tinted spectacles with which we view the world), and to account for delusional ideas in psychosis, such as paranoid feelings of persecution which, it is hypothesized, originate in the subject's own aggressive phantasies but are attributed, via projective identification, to persecutors.

Projective identification differs from simple projection in that the objects of PI are induced or controlled by the projection in such a way that they then *enact* the phantasy, which has been transferred into them. Paranoid people have the capacity to make those around them behave in suspicious or hostile ways, and thus projective identification can be thought of as a form of communication in which the recipient of the projection is induced to think or feel in ways that properly 'belong' to the projector. Post-Kleinian authors, notably Bion,[5] Heimann,[6] and Ogden,[7] have extended the concept of projective identification, with an emphasis on this communicative aspect, in that PI requires a recipient as well as a projector.

Bion, an analysand of Klein, realized that projective identification also underlies normal empathy and fellow feeling. PI is 'primitive' in the sense that preverbal children rely on it almost exclusively to communicate their feelings, but this denotes immaturity rather than pathology. Bion went on to develop his **container-contained** theory of early emotional communication. Here the mother, or 'breast', acts, via PI, as a recipient or container for the infant's unmanageable feelings of fear, hatred, annihilation, etc. These feelings are contained or held by the mother, and 'detoxified' before they are 'returned' to the infant through her understanding and empathic handling. She knows intuitively—through projective identification—when her child cries whether it is hungry or cold, or bored or wet, etc., and responds appropriately. In this way the infant begins to build-up a sense of himself through the **reflective awareness** of the mother. Disruptions of this process, for example through maternal depression or the violent use by the parent of the infant as a container (**role reversal**) as occurs in child abuse, may sow the seeds of disorders of identity found in borderline personality disorder in later life.

PI is important in the contemporary understanding of **counter-transference**. Paula Heimann pointed out that the therapist's reactions to the patient, while no doubt coloured to some extent by her own conflicts (Freud's classical conceptualization of counter-transference), also represent feelings induced by contact with the patient, that is to say they are a manifestation of projective identification. By attending to these thoughts and feelings the therapist gains clues about the patient's state of mind, which can then be put into words as interpretations. Here the therapist's mind is the container for the patient's split-off feelings. Sometimes this container-contained relationship fails, and the therapist is induced to enact some aspect of the patient's inner world, for instance by forgetting an appointment with a patient who has felt neglected and overlooked as a child, or by expressing anger or boredom in his tone of voice, being himself moved by feelings which properly belong to the patient.

The firm boundaries of psychotherapy are, in part, designed to minimize these occurrences (although they are unavoidable, and often, if reflected on, can be put to good use in the form of deepened understanding), but in the much more uncontained setting of general psychiatric wards or community mental health centres such enactments are widespread. A common example would be the polarization which disturbed people with borderline personality disorder can induce in their carers, some seeing the patient as manipulative and demanding, others feeling intense sympathy, and the wish to repair past hurts on the patient's behalf. Each perspective represents a split-off aspect of the patient's inner world that has been picked up via PI by different staff members. This is an essentially interactive process, since, no doubt, what determines which aspect depends on the carers' own developmental history and defensive strategies.

Working in the relative isolation of Scotland, and coming to essentially similar conclusions to Klein about the importance of splitting, W.R.D. Fairbairn[8] further developed this interpersonal perspective. For him drives were 'a signpost to the object', the glue that held human beings together. Sex is what gets us close to those who matter, rather than vice versa, as originally conceived by Freud. Like Bion later, Fairbairn also placed great emphasis on the role of the mother and of environmental failure as a source of psychopathology. Frustration plays a central part in his schema. With a perfectly responsive mother, the child has no need to think or develop an inner world. When separation and frustration come into play, the child then builds up an image of the object, which is split into three parts: the **ideal object** (one that would never cause frustration), the **libidinal object** (one that could satisfy the child's drive-related needs), and the **anti-libidinal object** (the one that frustrates). This in turn sets up a split of the self into three corresponding parts—ideal self, libidinal self, and anti-libidinal self. The Fairbairnian model provides clarity in understanding some typical phenomena found in severe personality disturbance: the swing between idealization and denigration of therapists and partners (who become the anti-libidinal withholding object at that point), the self-destructiveness of the anti-libidinal self, or 'internal saboteur', and the split-off search for pure libidinal satisfaction unrelated to persons represented by substance abuse and promiscuity.

Fairbairn's notion of schizoid withdrawal was conceptualized as a typical interpersonal strategy in the face of frustration. John Steiner[9] has developed a similar idea in his notion of the psychic retreat, an inner place to which individuals with borderline personality may repair in the face of environmental trauma, and which may make them relatively inaccessible in therapy. Another important neo-Kleinian development has been Ronald Britton's[10] attempt to link the Oedipus complex with the tolerance of separateness and loss implicit in the depressive position. Britton sees the ability, at times, to let go of the mother as the Oedipal stage is successfully negotiated—in which the child comes to see that his mother and father are sexually involved with one another and he is necessarily excluded—as an important developmental step towards the establishment of an inner world and the ability to see things from varying perspectives. This can be linked with Bion's idea of creative thought in which ideas are brought together to create 'conceptions', in contrast to the destructiveness of schizoid thinking in which, as a way of reducing anxiety, the links between things and ideas are attacked, and the world emptied of meaning. The restoration of meaning is a central task of psychotherapy. The dialectic of close involvement and repeated separation inherent in the therapeutic relationship fosters this capacity, enabling disturbed patients first to **find their experience mirrored** by the responsive therapist, then gradually to **tolerate loss and envy**, and so to gain the capacity to think and to feel more autonomously.

Object relations 2: Balint and Winnicott

The 'Object relations' school of psychoanalysis is a broad church. Klein's view of the mind and of psychopathology was essentially a **conflictual** model: difficulty arises out of the inherent conflict in an immature mind between love and hate, and attempts to avoid the inevitability of loss. For her, such conflict was characteristic of normal development, and pathology merely an exaggeration of normal conflict in which the environment has failed to mitigate its potentially destructive effects. By contrast, the non-Kleinian members of the 'object relations' school tend to espouse some variety of a **deficit** model, in which normal and abnormal development are more sharply differentiated, and the basis for psychopathology is a failure of the environment to provide the conditions needed for healthy psychic growth.

Michael Balint[11] is perhaps best known for his work in raising psychological awareness among general practitioners through the

use of 'Balint groups', but he was also a significant figure in psychoanalysis, introducing a number of key terms and concepts. In contrast to Klein, who saw the newborn infant as wracked with fear and conflict, Balint proposed a state of **primary love** characterizing the early mother–infant relationship—which he described as a 'harmonious interpenetrative mix-up'. Where, however, parenting was inadequate, due to neglect, overintrusiveness, aggression, or abuse he claimed that the child would be permanently scarred at the level of the '**basic fault**'. His model of therapy implied a **remedial**, rather than purely **interpretative** approach, with the therapist's role including both quiet acceptance, and on occasion therapeutic 'acting in': Balint would sometimes gently hold the patient's hand, and, famously, once encouraged a patient who stated that she had never had the courage to do a somersault to try one out in the consulting room then and there (behaviour therapy meets psychoanalysis!).

Donald Winnicott,[12] visualized an intermediate zone in the early years of life that was neither the realm of pure phantasy (as described by Klein), nor that of reality (to which adaptation by the ego was required, as described by Freud), although it partakes of both. In this intermediate, or **transitional** zone the infant learns, with the help of the mother, to play (another key Winnicottian theme). Here phantasies can become reality, at least for the duration of the interactive play. In this transitional space Winnicott saw the origins of creativity and culture generally, and of a nascent sense of self. He suggested that the mother's face is a kind of mirror in which the child sees his own feelings reflected, and through this recognition begins to gain a sense of who he is. This process is disrupted if the mother is depressed or abusive, and here perhaps are the germs of borderline personality disorder, characterized by a deficient sense of self, and feelings of inner emptiness and sterility. Winnicott saw 'learning to play' as a key task in therapy in helping patients to regain their sense of self.

A related phenomenon is that of the **transitional object**—the special handkerchiefs, teddy bears, and precious playthings that toddlers often need for comfort and to help them sleep. Winnicott saw these as buffers against loss, objects that are invested with the properties of the primary object (the mother and her breast) but remain under the control of the child. They are 'transitional' in the sense that they lie between the ideal object of phantasy and the real, but potentially unreliable, objects of external reality.

The subtlety of Winnicott's thought is exemplified by his notion of the **good-enough mother**. Unlike some psychoanalytical writers he did not attribute all the evils of mankind to parental failure. Winnicott realized that a 'perfect' mother, intrusively aware of her infant's needs could inhibit rather than foster the development of a sense of oneself as a separate and autonomous being. Mothers (and presumably fathers) should be 'good enough', not perfect, not least because through healthy protest about parental failure the child learns his own strength and finds limits, which reassure him that his parents can withstand his aggression and still love him.

Winnicott realized that developmental deficit does not always take the form of neglect or overt violence. He noticed the ways in which parents, driven by their own unconscious needs, may subtly impose their will on a compliant child, thereby inhibiting the growth of a robust and distinct sense of self. The **false-self-real-self** distinction tries to capture the ways in which children, and later personality disordered adults, may present an acceptable face to the world that is radically at variance with inner feelings of terror, emptiness, or rage. In his seminal, but today largely forgotten, classic *The Divided Self*, R.D. Laing[13] took Winnicott's false-self-real-self distinction as a central theme in his psychodynamic account of schizophrenia, seeing delusions as representing a way of holding together, albeit 'falsely', a disintegrating 'real' self and its inner world.

John Bowlby and attachment theory

Winnicott's contemporary John Bowlby[14] life's work was an attempt to bring logical and scientific rigour to psychoanalytical thought. Attachment theory, an empirically validated version of object relations theory, starts from Freud's[15] revised theory of anxiety, in which, rather than viewing it as the result of incomplete repression of incestual wishes, **anxiety is conceptualized in interpersonal terms as a response to the threat of the loss of a loved one**. Based on his observations of delinquent youths, many of whom had suffered the loss of a parent during early childhood, and the depressive reactions of small children to separation from their parents on entering hospital, Bowlby saw that protection from danger was a key component of the parent–child relationship, and that there were built-in psychological mechanisms to ensure the maintenance of attachment bonds.

Attachment theory[16] postulates that, when faced with threat, illness, or exhaustion, children will seek proximity to their caregivers, or '**secure base**'. A protective response from the caregiver assuages the child's attachment needs, who can then return to play or exploratory behaviour, secure in the knowledge that help will once more be at hand if needed. This provides the conditions for **secure attachment**, and the child builds up an **internal working model** (Bowlby's preferred term for the inner world) of a secure robust self and responsive others.

Secure attachment arises out of **responsive and sensitive parenting** and is contrasted with **insecure attachment**, which Bowlby saw as a factor predisposing to adult neurosis. Bowlby's collaborator, Mary Ainsworth, and her students, have researched different patterns of insecure attachment and the conditions under which they arise.[17] They delineate three main types of insecure attachment: **insecure-avoidant, insecure-ambivalent, and insecure-disorganized**. The avoidant child has experienced brusque or aggressive parenting, and tends to avoid close contact with people, hovering near caregivers rather than openly expressing need when faced with threat. The ambivalent child clings to his inconsistent parents, and finds exploratory play difficult, even when the danger has past. Disorganized children behave in bizarre ways when threatened, and tend to have parents who are either emotionally intrusive or absent, often in the context of a parental history of abuse in their childhood. Disorganization is thought to be a severe form of insecure attachment and a possible precursor of severe personality disorder and dissociative phenomena in adolescence and early adulthood.

Mary Main[18] has developed a psychodynamic interview schedule, the Adult Attachment Interview (**AAI**), which is rated for the interviewee's narrative style, and, in long-term follow-up studies of children whose attachment patterns have been classified in infancy, yields significant links with these earlier patterns of attachment. As with response to threat in childhood, adults' ways of talking about themselves and their lives vary enormously. Some, in

the **secure-autonomous** style, talk freely about themselves and their past pain in a coherent and apposite way. The **insecure-dismissive** style minimizes problems and is characterized by unelaborated speech lacking in metaphor or vividness. The **insecure-preoccupied** style is rambling and emotionally laden, while an **insecure-unresolved** pattern has evident breaks in continuity and logical flow. These insecure speech patterns are, it is suggested, manifestations of the underlying psychobiological relational dispositions, which the various theories of object relations attempt to capture. The way we speak about ourselves reveals the state of our inner world. Peter Fonagy[19] has suggested that the capacity to represent experience, which he calls **reflexive function**, (a contemporary version of the classical psychoanalytic notion of 'insight'), is a buffer against psychiatric disturbance. Once pain is represented in the mind the sufferer can distance himself from it, and consider alternative ways of responding. Enhancement of reflexive function is a generic psychotherapeutic strategy and applies as much to cognitive therapy (becoming aware of negative cognitions and automatic thoughts) as psychodynamic therapies.

Bowlby objected to what he saw as the hijacking of the term 'biological' by organic psychiatry, since he believed the attachment relationship and its vicissitudes, adaptively shaped by evolutionary pressures, was no less 'biological' than the neurochemistry which presumably mediates it. For him human psychology was fundamentally relational. He saw attachment needs as existing throughout the life cycle, and put separation and loss as central to his view of the origins of psychiatric disturbance. In the attachment model, separation from a caregiver is a threat: we are biologically programmed to respond with shock, denial, anger, and searching behaviours when separated from a loved person or object. Loss is an irrevocable separation, and the early phases of the bereavement response are all vain attempts to restore the status quo. Despair and depression come with the recognition that separation is final, and, beyond that, reorganization of internal working models, the recognition that although the loved one is lost in reality, good memories live on in the inner world.

The attachment perspective has implications for the day-to-day practice of psychiatry. One function of the psychiatric facilities and of mental health workers is to provide the patient with a 'secure base', which in itself goes some way to reducing anxiety. Appropriate dependency is integral to the supportive psychotherapeutic relationship, which is such a key part of the psychotherapeutic dimension of psychiatry. Short-term, unresponsive, or rejecting relationships with psychiatrists and other mental health workers reinforce insecure attachment and may lead patients to redouble their efforts to cling on to the psychiatric institution—an all too familiar vicious circle.

The ego and its defences: Anna Freud, Hartmann, and Lacan

The role of the ego and of defence mechanisms was a particular concern of Anna Freud,[20] who represented a parallel tendency to the object relations school. She elaborated a taxonomy of defences used by the ego to maintain its integrity in the face of both internal threat from the id, and the demands and impingements of external reality. Valliant[21] groups Anna Freud's defences into those that are **immature** (like projective identification and splitting), **neurotic defences** (which include intellectualization, reaction formation, and identification with the aggressor), and **mature** ones (such as humour and sublimation).

Reaction formation describes the ways in which the ego counteracts unconscious desires or impulses that threaten its equilibrium by consciously held views directly contrary to these: the militant pacifist who is out of touch with any feelings of aggression for example. **Identification with the aggressor** is frequently invoked in discussions of the psychological effects of childhood abuse. One way of dealing with the horror of abuse is to 'dis-identify' with oneself (a form of dissociation), and to put oneself in the place of the person who is attacking, thereby reducing feelings of pain and helplessness. This idea helps to explain how those who have been abused in childhood may become abusers themselves in adult life. A frequent experience in working with severely disturbed patients, many of whom are abuse survivors, is that health care workers may themselves feel attacked or symbolically 'abused' by these patients—seeing how the patient may have unconsciously identified with their aggressor can help carers to a greater understanding of their patients' problems and to respond less defensively to these attacks.

Valliant has found that men who use more mature defence mechanisms are less vulnerable to physical and psychological illness, and an important aim of psychotherapy would be to help the patient move from the use of more- to less-primitive defence mechanisms. Defences are therefore legitimately seen as adaptive, and, from a developmental perspective, the earlier the presumed psychic trauma, the more likely are primitive defence mechanisms to be employed.

David Malan's[22] **triangular model** of anxiety, defence, and 'hidden impulse' is another variety of ego psychology, which has found favour in psychiatric circles. It provides a clear formula for thinking about neurotic difficulties: for example people suffering from agoraphobia commonly defend against anxiety by avoidance and dependency; underlying this there may be hidden feelings of dissatisfaction and aggression, immediately towards a spouse, and in the past towards a controlling but unaffectionate mother. For Malan, the task of therapy is to allow the ego to tolerate and express the hidden feelings; note that cognitive therapy (q.v.) similarly helps the patient to become aware of and then counteract the automatic thoughts (equivalent to hidden feelings) that undermine the ego's attempts to achieve conflict-free functioning.

The self, meaning, and interpersonal psychoanalysis: Sullivan, Horney, and Kohut

Freud's models of the mind were essentially intrapsychic, and couched in quasi-scientific language. Object relations retained this perspective but introduced a relational dimension never fully developed by Freud. **Interpersonal psychoanalysis** in the United States was even more radically interpersonal than object relations. Harry Stack Sullivan[23] was a free thinker who emphasized this existential aspect of psychotherapy, while remaining within the psychoanalytical tradition. He worked particularly with people suffering from schizophrenia. Sullivan believed in close involvement with his psychotic patients. His mission was always to find meaning in their experience, rather than dismiss it as an unintelligible manifestation of organic illness. He was a major influence on a generation of psychoanalytically informed psychiatrists including Harold **Searles**, Freida **Fromm-Reichman**, and Karen Horney.[24]

The latter, like Sullivan, was critical of the patriarchal bias of psychoanalysis. For her, castration complex and penis envy were social rather than biological phenomena, manifestations of social relations that subjugated women, and from which, by appropriate action, including psychotherapy, they could be liberated. Horney's contemporary, Eric **Fromm**, brought a Marxist influence into psychoanalysis, emphasizing the part played by capitalist production methods in contributing to isolation and anomie of modern men and women and their psychological troubles.

Although conventional psychoanalytical treatment for schizophrenia is now largely discredited, there is increasing interest in the role of psychosocial interventions in psychosis. Here the Sullivanian principles of respect for the patient's experience and its meaning, the need for a long-term supportive psychotherapeutic relationship, attention to the social precipitants of psychosis, and a focus on the ways in which the therapist may, through countertransference, foster recovery or reinforce pathology, are all highly relevant to contemporary psychiatry.

Heinz Kohut[25] was concerned not so much with schizophrenia, but with that intermediate world between neurosis and psychosis which psychoanalysts call 'borderline' pathology, and which has entered the DSM as Borderline Personality Disorder. Like Sullivan, Kohut puts **self-esteem and its disorders** at the centre of his psychology, seeing the origins of self-esteem in the empathic responsiveness of caregivers in the early years of life. For him there is a core of **healthy narcissism**, which is based on the grandiosity and omnipotence of the young child ('his majesty the baby', as Freud put it), which is both accepted and fostered by effective parenting. Parents at this stage are '**self-objects**', a concept akin to Winnicott's transitional objects, who partake both of the self and of the responsive environment, and which the infant believes, in his state of healthy delusion, to be there exclusively for his benefit.

Like Winnicott, Kohut emphasizes 'mirroring' as a key interpersonal theme. For Winnicott parental mirroring helps the child to own his emotions and begin to know who he is. Kohut, by contrast, takes up the narcissistic aspect of the mirror: the child sees his reflected glory in the eyes of his admiring parents, and this contributes towards his own positive self-regard. As development proceeds there is a process of 'optimal disillusionment', similar to the resolution of the Oedipus complex, in which the child gradually learns that his objects have a life of their own. By this time, however, his sense of a valued and effective self will be sufficiently developed, and residual narcissism will serve the useful functions of ambition, aspiration to success and admiration, a sense of duty and concern for others, and the capacity to invest in ones offspring.

Where the environment is unempathic, mirroring is deficient, fragile grandiosity squashed, or disillusionment traumatic, the stage is set for borderline pathology, in which self-absorption, and the use of others as self-objects, appropriate to the infantile years, persists into adult life. Self-injurious behaviour such as drug abuse, eating disorders, and deliberate self-harm are 'breakdown products' of a disintegrated self, trying to use the environment as a self-object that will provide momentary and illusory satisfaction and self-affirmation.

A therapeutic implication of Kohut's approach is that the therapist is more supportive than in classical analysis, tolerant of the patient's grandiose designs, especially in the early stages of treatment. This contrasts with the approach of Otto Kernberg[26] who synthesizes classical and Kleinian concepts, and advocates rigorous interpretation, especially of destructive and self-defeating behaviour in borderline patients. Kohut's and Kernberg's theories reflect the typical polarization that such patients evoke in clinical settings, perhaps mirroring an inner world rigidly split into good and bad objects. Effective treatment requires a synthesis: empathy and tolerance is needed to form a working alliance, but firm limit-setting and confrontation of destructiveness is also essential. Another interpersonal synthesis is to be found in the work of Stephen Mitchell[27] whose approach reminds the therapist of the reciprocity of the therapeutic relationship in what Robert Lang calls the 'bipersonal field': patient and therapist form a system of mutual influence, the job of the therapist being both to participate in this, and at the same time to be sufficiently detached to be able to reflect upon it.

The shift in interpersonal psychoanalysis away from the analyst as an objective and privileged observer to a co-participant has given rise to a contemporary interest in narrative or hermeneutic explanations in psychotherapy,[28] in contrast to the scientific psychology, which Freud originally hoped to establish. These authors argue the case for psychoanalysis as a hermeneutic discipline whose aim is to explore meaning rather than objective truth. If Freud is one of the intellectual founding fathers of modernism, their approach is 'postmodern' in the sense that it stresses the relativism of values and meanings, and the importance of power in determining one's view of the world. Here is a link—albeit so far rather distant—with the emerging 'user' movement in psychiatry, and the importance of giving as much weight to the client's voice as to that of the professional. Psychological truths are inherently contextual, and without awareness of the social context they can be obfuscatory.

Conclusions

Many new perspectives have emerged in the century since psychoanalysis was conceived. Emphasis has shifted from the intrapsychic to the interpersonal. Kleinian psychoanalysis offers a unique vision of the ways in which interpersonal reality is inescapably coloured by the emotional state of the participants. Attachment theory provides an account of human psychological development that both takes account of meaning and is empirically based. Psychoanalysis is emerging from its isolation and bridges have begun to be built with cognitive science: the inner world of phantasy is not unlike the world of schemata and assumptions that are the focus of cognitive therapy. There are, through modern neuroimaging techniques links to be forged with neurobiology: the impact of effective therapeutic interventions on brain architecture can now be visualized. Progress will depend on further theoretical syntheses and technological advances, while holding firm to the humanistic emphasis on personal meaning and inner experience that is the fundamental contribution of psychoanalysis to contemporary psychiatry.

Further information

Bateman, A. and Holmes, J. (1995). *Introduction to psychoanalysis: contemporary theory and practice*. Routledge, London.

Cassidy, J. and Shaver, P. (eds.) (2007). *Handbook of attachment* (2nd edn). Guilford, New York.

Fonagy, P. and Target, M. (2002). *Psychoanalytic theories. Perspectives from developmental psychopathology*. Whurr, London.

Gabbard, G., Beck, J., and Holmes, J. (2005). *Oxford textbook of psychotherapy*. Oxford University Press, Oxford.

Mayes, L., Fonagy, P., and Target, M. (eds.) (2007). *Developmental science and psychoanalysis*. Karnac, London.

References

1. Freud, S. (1917). Mourning and melancholia. In *Standard edition of the complete psychological works of Sigmund Freud*, Vol. 14 (ed. J. Strachey), pp. 239–58. Hogarth Press, London, 1957.
2. Klein, M. (1986). *The selected Melanie Klein* (ed. J. Mitchell). Penguin, Harmondsworth.
3. Spillius, E. (1988). *Melanie Klein today*. Routledge, London.
4. Hobson, R.P., Patrick, M., and Valentine, J. (1998). Objectivity in psychoanalytic judgements. *The British Journal of Psychiatry*, **173**, 172–7.
5. Bion, W. (1988). Attacks on linking. In *Melanie Klein today* (ed. E. Spillius), pp. 178–86. Routledge, London.
6. Heimann, P. (1950). On countertransference. *International Journal of Psychoanalysis*, **31**, 81–4.
7. Ogden, T. (1979). On projective identification. *International Journal of Psychoanalysis*, **60**, 357–73.
8. Fairbairn, W.R. (1952). *Psychoanalytic studies of the personality*. Routledge, London.
9. Steiner, J. (1993). *Psychic retreats*. Routledge, London.
10. Britton, R., Feldman, M., and O'Shaughnessy, E. (1989). The Oedipus complex todsay. Karnac Press, London.
11. Balint, M. (1968). *The basic fault*. Tavistock Press, London.
12. Winnicott, D. (1971). *Playing and reality*. Penguin, Harmondsworth.
13. Laing, R. (1961). *The divided self*. Penguin, Harmondsworth.
14. Bowlby, J. (1988). *A secure base: clinical applications of attachment theory*. Routledge, London.
15. Freud, S. (1926). Inhibitions, symptoms, and anxiety. In *Standard edition of the complete psychological works of Sigmund Freud*, Vol. 20 (ed. J. Strachey). Hogarth Press, London.
16. Cassidy, J. and Shaver, P. (eds.) (2007). *Handbook of attachment* (2nd edn). Guilford, New York.
17. Holmes, J. (2001). The search for the secure base: attachment theory and psychotherapy. Routledge, London.
18. Main, M. (1995). In *Attachment theory: social, development and clinical perspective* (eds. S. Goldbery, R. Muir and J. Kerr), pp. 407–74. Analytic Press. New York.
19. Fonagy, P. (1995). In Attachment theory: social, development and clinical perspective (eds S. Goldbery, R. Muir and J. Kerr), pp. 233–78. Analytic Press. New York.
20. Freud, A. (1936). *The ego and the mechanisms of defense*. Hogarth Press, London.
21. Valliant, G. (1992). *Ego mechanisms of defense: a guide for clinicians and researchers*. American Psychiatric Press, Washington, DC.
22. Malan, D. and Coughlin Della Selva, P. (2006). *Lives transformed*. Karnac, London.
23. Sullivan, H.S. (1962). *Schizophrenia as a human process.* Norton, New York.
24. Horney, K. (1939). *New ways in psychoanalysis.* Norton, New York.
25. Kohut, H. (1977). *The restoration of the self*. International Universities Press, New York.
26. Kernberg, O. (1996). A psychoanalytic theory of personality disorders. In *Major theories of personality disorder* (eds. J. Clarkin and M. Lenzenweger), pp. 116–137. Guilford Press, New York.
27. Mitchell, S. (1988). *Relational concepts in psychoanalysis: an integration*. Harvard Universities Press, Cambridge, MA.
28. Roberts, G. and Holmes, J. (eds.) (1998). *Healing stories: narrative in psychiatry and psychotherapy*. Oxford University Press, Oxford.

3.3

Current psychodynamic approaches to psychiatry

Glen O. Gabbard

Psychodynamic psychiatry is broadly defined today. In fact, the term psychodynamic is now used almost synonymously with psychoanalytical. Freud originally used the term psychodynamic to emphasize the conflict between opposing intrapsychic forces: a wish was opposed by a defence, and different intrapsychic agencies, such as ego, id, and superego, were in conflict with one another. Indeed, for much of the twentieth century psychoanalytical theory was dominated by the drive-defence model, often referred to as ego psychology.

In the last decades of the twentieth century, however, psychoanalytical theory expanded beyond the notion of conflict among intrapsychic agencies. Internal object relations became paramount in models deriving from these sources. In addition, a deficit model of symptomatology arose from the work of the British object-relation theorists, such as Balint and Winnicott. In the United States, Kohut's self-psychology also developed a model based on developmental deficits. In other words, disturbed patients who came to treatment were seen as suffering from absent or weakened psychic structures based on developmental failures by parents or caretakers in the early childhood environment. (See Chapter 3.1 for an account of the development and modern practice of psychoanalysis.)

The typical psychodynamic psychiatrist then uses multiple models to assist in the understanding of a particular patient. Developments in neuroscience must also be taken into account. Moreover, the diagnostic and treatment approach to an individual patient is psychodynamically informed even when a decision has been made to forego psychodynamic psychotherapy. Psychodynamic thinking provides a conceptual framework within which all treatments are prescribed, including pharmacotherapy, psychotherapy, inpatient or partial hospital treatment, and group or family modalities. Psychodynamic psychiatry is not synonymous with psychodynamic psychotherapy.

A comprehensive definition of current psychodynamic psychiatry is the following:[1]

Psychodynamic psychiatry is an approach to diagnosis and treatment characterized by a way of thinking about both patient and clinician that includes unconscious conflict, deficits, and distortions of intrapsychic structures, and internal object relations, and that integrates these element with contemporary findings from the neurosciences.

Basic principles

A set of time-honoured basic principles, all derived from psychoanalytical technique and theory, define the overall approach of the dynamic psychiatrists (Table 3.3.1).

The unconscious

A fundamental premise of psychodynamic psychiatry is that mental activity going on outside our awareness can be profoundly influential. Freud saw signs of the unconscious in two major types of clinical evidence: parapraxes and dreams. Parapraxes, commonly referred to as slips of the tongue or 'Freudian slips', involve substituting one word for another. For example, a patient who intends to say 'Protestant', may unwittingly say 'prostitute'. Parapraxes may also involve actions, such as forgetting, or executing one action when intending to do another.

Freud regarded dreams as the 'Royal Road' to the understanding of the unconscious. Another primary way that the unconscious manifests itself in the clinical setting is the patient's behaviour toward the clinician. Certain characteristic patterns of relatedness to others set in childhood become internalized and are manifested automatically and unconsciously as part of the patient's character. Hence certain patients may consistently act deferentially toward the clinician, while others will behave in a highly rebellious way. This type of procedural memory is closely linked to Squire's[2] notion of implicit memory, which occurs outside the realm of verbal narrative memory.

While declarative or autobiographical memory involves remembered events and narratives of one's life, procedural memory

Table 3.3.1 Basic principles of psychodynamic psychiatry

The unconscious
Psychic determinism
Developmental orientation
Emphasis on the uniqueness of the individual rather than how the individual is like others
Transference
Countertransference
Resistance

stores the 'how' of executing sequences of actions, such as motor skills. Once guitar-playing or bicycle-riding has been mastered, no conscious recall is necessary when one sits down with a guitar or jumps on a bicycle. The schema referred to as unconscious internal object relations are to some extent procedural memories repeated again and again in a variety of interpersonal situations. They are non-conscious, but not dynamically unconscious, in the sense of being defensively banished from conscious awareness.

The notion that much of mental life is unconscious is one that is often challenged by psychoanalytical critics, but it is also one that is extensively validated by literature from experimental psychology.[3] Repression of memory has even been demonstrated in fMRI research.[4] The active effort to 'forget' unwanted past experiences involves a novel form of reciprocal interaction between the prefrontal cortex and the hippocampus. When subjects control unwanted memories, there is increased dorsolateral prefrontal activation associated with reduced hippocampal activation. The magnitude of forgetting is predicted by prefrontal, cortical, and right hippocampal activations.

Psychic determinism

The notion of psychic determinism is intimately linked with the construct of the unconscious. Freud felt that behaviour and mental life were related to multiple and complex causation.[5] The term overdetermination implies that a variety of intrapsychic and unconscious factors come together to produce specific symptoms or behaviours. The notion of multiple causation implies that there can be alternate sets of sufficient conditions, some involving primarily unconscious conflicting forces, others stemming from biological and environmental influences that ultimately produce similar symptoms or behaviours.

Developmental orientation

Regardless of which psychoanalytical theory seems to fit best with a particular patient, the dynamic psychiatrist always thinks in terms of developmental models. Patterns of relatedness established in childhood are repeated in adult relationships. Modern dynamic psychiatrists avoid the early psychoanalytical reductionism that attempted to link an adult psychopathological syndrome to a specific developmental arrest or fixation in childhood. Today, full account is taken of genetic contributions to personality and to psychiatric disorders. Environmental influences and genetic factors interact with one another reciprocally to shape the human being in health and illness. Still, the wisdom of the psychodynamic approach is that within each of us is a child yearning to complete some unfinished business from earlier in life.

Emphasis on the uniqueness of the individual

In much of descriptive psychiatry the major focus is on taxonomy—specifically: How do groups of patients fit together under one classification? In psychodynamic psychiatry, by contrast, there is great interest in how a particular patient is unique—in other words, different from others. The subjective experience of the individual has been forged through an idiosyncratic narrative that is different from all other life stories and involves a specific interaction between genetic predisposition, intrapsychic factors, and environmental influence.

Transference

Intrinsic to the developmental model of mental organization is that adults are constantly repeating childhood patterns in the present. Transference is the best-known example of this phenomenon. The patient unconsciously experiences the doctor as a significant figure from the past and reacts to the doctor based on a set of unconscious attributions based on those past experiences. Transference has undergone considerable revision in more recent writings, so that today much more emphasis is placed on the clinician's contributions to the patient's transference. In other words, if a clinician is silent and remote, the patient may experience that clinician as disengaged and cold. While an internal template of past experiences with authority figures may correlate with that perception, we would also recognize that the clinician's real behaviour contributes to that precise transference paradigm. In that regard, a more contemporary view of transference would be that every treatment relationship is a mixture of new features based on real characteristics of the clinician and old experiences from the patient's past. Psychodynamic clinicians also recognize a bidimensional quality to transference: while one dimension involves repetition of the past, another dimension is seeking an experience with a new object to facilitate further emotional growth.

Countertransference

Central to the psychodynamic viewpoint is that the clinician and the patient bring their own separate subjectivities to an encounter, and mutually influence one another. Countertransference, in this respect, is the counterpart of transference. In other words, as Freud originally used the term, it referred to the analyst's attribution of certain qualities to the patient based on the analyst's past experiences with similar figures. This perspective, often referred to as the narrow view of countertransference, regarded the phenomenon as an obstacle to be removed because it interfered with the analyst's objectivity.

Subsequent contributors to the literature on countertransference[6, 7] noted that countertransference with severely disturbed patients often involves an objective component. The patient behaves in such a provocative manner that virtually anyone would respond with a certain set of emotional reactions to that patient. This way of looking at countertransference is often regarded as the broad or totalistic view. Inherent in this perspective is that the clinician's reaction has much less to do with his or her own individual past than with the specific characteristics of the patient and that patient's capacity to induce strong reactions in others.

As the definition has continued to evolve, countertransference is now generally regarded as involving both the narrow and the broad characteristics. In other words, most theoretical perspectives view countertransference as entailing a jointly created reaction in the clinician that stems, in part, from contributions of the clinician's past and, in part, from feelings induced by the patient's behaviour.[8] In some cases the emphasis may be more on the contributions of the clinician than the patient, while in other cases the reverse may be true. This model also regards countertransference as something of a unique construction that varies depending on the two subjectivities involved (see Box 3.3.1). In this contemporary perspective, countertransference is both a source of valuable information about the patient's internal world and something of an interference with the treatment.

Box 3.3.1 Changing views of countertransference

Narrow The original Freudian view connoting the analyst's transference to the patient: an obstacle to be removed through careful analysis of the clinician.

Broad or totalistic All the feelings experienced towards the patient, some of which are induced by the patient's behaviour.

Joint creation The contemporary perspective that emphasizes mutual contributions from the patient's behaviour toward the clinician and the clinician's past experiences with similar figures. This perspective emphasizes countertransference as a source of valuable information in addition to being an interference.

Resistance

In 1912 Freud[9] wrote, 'The resistance accompanies the treatment step by step. Every single association, every act of the person under treatment must reckon with the resistance and represents a compromise between the forces that are striving towards recovery and the opposing ones'. The patient's resistance defends the patient's illness from the clinician's attempt to treat it and change it. Resistance may be conscious, preconscious, or unconscious. It may take many forms, including not taking medication as prescribed, forgetting appointments with the psychiatrist, changing the subject in the middle of an appointment to something trivial, and discounting every insight the psychiatrist offers. The patient's characteristic defence mechanisms are often transformed into resistances in the treatment situation. The dynamic psychiatrist knows that all progress will be accompanied by some degree of resistance, and the exploration of resistance is a major part of therapeutic work. Resistance is intimately related to transference because the patient often rebels against the doctor resulting from unconscious transference configurations that lead the patient to oppose the doctor's help.

The mind–brain interface

The psychodynamic psychiatrist eschews reductionism. Recognizing that mental life and psychiatric symptoms are both overdetermined and multiply caused, psychodynamic clinicians are always interested in the interface between the biological and the psychosocial. Psychodynamic psychiatry is not antibiological. The psychodynamic psychiatrist is the integrator par excellence. Avoiding Cartesian dualism, the mind is seen as the expression of the activity of the brain.[10] Subjective experience affects the brain just as mental phenomena arise from the brain. Every treatment intervention is seen as being biopsychosocial in nature. Medications have psychological effects. Psychotherapeutic interpretations affect the brain. Moreover, psychodynamic psychotherapy and medications may work synergistically to provide better outcomes for patients. For example, a patient with a bipolar disorder who is denying that he has an illness and refusing to take lithium may ultimately have better compliance with the medication if the clinician explores the meaning of his denial and his reluctance to consider himself as someone requiring treatment.

The comprehensive mind–brain strategy of the contemporary psychodynamic psychiatrist fits well with our growing knowledge of the interaction between genes and the environment. In an inspired series of experiments with the marine snail *Aplysia*, Kandel[11,12] has demonstrated that synaptic connections are strengthened and permanently altered through regulation of gene expression connected with learning from the environment. In *Aplysia* the number of synapses actually double or triple as a result of learning. Kandel has suggested that psychotherapy might make similar neuroanatomical changes in the synapses. He argues that just as representations of self and others are malleable, the brain itself is a dynamic and plastic structure. He postulates that psychotherapy is a form of learning that produces alteration of gene expression and thereby alters the strength of synaptic connections. While the template function or the sequence of the gene is not affected by environmental experience, the transcriptional function of the gene (namely the ability of a given gene to direct the manufacture of specific proteins) is highly regulated and responsive to environmental factors.

Antisocial personality disorder may be a model disorder with which to examine the interaction of genes and environment. In a perspective study based in Dunedin,[13] a birth cohort of 1037 children was followed prospectively. By the age of 26, 96 per cent of the sample was contacted and evaluated. Between the ages of 3 and 11 years, 8 per cent experienced 'severe' maltreatment, 28 per cent experienced 'probable' maltreatment, and 64 per cent experienced no maltreatment. The investigators determined that a functional polymorphism in the gene responsible for the neurotransmitter metabolizing enzyme monoamine oxidase-A (MAO-A) was found to moderate the effect of maltreatment. Males with low MAO-A activity genotype who were maltreated in childhood had elevated antisocial scores. Males with high MAO-A activity did not have elevated antisocial scores, even when they had experienced childhood maltreatment. Of males with both low MAO-A activity genotype and severe maltreatment, 85 per cent developed antisocial behaviour.[13]

The research summarized here points to the dynamic interplay between genetic expression and the environment. Gene expression cannot be considered static. It is a dynamic phenomenon that interacts with and reacts to environmental experiences. Heritable characteristics of children actually shape their relationships with their parents and siblings.[14] In turn, the response of family members to the child affect the genetic expression. Hence genetic influences on some types of psychopathology may be dependent on the mediation of social processes. A child's genetic endowment will influence the way parents relate to a child, and the way the parents treat the child will then influence that child's developing brain. Biological and psychosocial processes are constantly intertwined, and neither is prior.

In many major psychiatric disorders, such as depression, genetic factors appear to influence whether a stressor produces an episode of illness.[15] From a psychodynamic perspective, the meaning of stressors must also be incorporated. Some stressors that may seem mild to one individual are overwhelming to another because of their idiosyncratic conscious or unconscious meaning. In addition, the presence of biologically generated symptoms in no way diminishes the importance of meaning. Pre-existing psychodynamic conflicts may attach themselves to biologically driven symptoms, and the symptoms then function as a vehicle for the expression of the conflicts.[16] Auditory hallucinations are generated by alterations in neurotransmitters in persons with schizophrenia, but

the content of the hallucination often has specific meanings based on the patient's psychodynamic conflicts. Hence a patient who is being told that he is a failure and should kill himself by a hallucinated voice may be tormented by a sense that his life is shattered by his illness and that he no longer has any purpose in living.

Development of personality

Another key component of the psychodynamic approach is that the clinician treats the person and not just the illness. In practice, that perspective means taking the personality into account in every case. The interface of the biological and the psychosocial is particularly apparent in the area of personality. The psychobiological model of personality developed by Cloninger *et al.*[17] recognizes an equal contribution of biological and environmental factors (see Table 3.3.2). The four dimensions of temperament are roughly 50 to 60 per cent heritable independently of one another. They all manifest themselves early in life, and they involve preconceptual biases, habit formation, and perceptual memory. They include the following:

1 novelty-seeking: characterized by active avoidance of frustration, quick loss of temper, impulsive decision-making, frequent exploratory activity in response to novelty, and extravagance in the approach to cues and rewards
2 harm-avoidance: which involves pessimistic worry about the future, passive avoidant behaviour such as fear of uncertainty, shyness regarding strangers, and rapid fatiguability
3 reward-dependence: characterized by sentimentality, social attachment, and dependence on the approval of others
4 persistence: which refers to the capacity to persevere despite fatigue and frustration

Certain of these temperament dimensions appear to correlate with specific types of personality disorders. The cluster A personalities in DSM-IV, for example, are strongly associated with low reward-dependence. Cluster B personality disorders have been shown to be high in novelty-seeking, while cluster C personality disorder patients tend to rate high in harm-avoidance.

The other component of personality in this model is character. While temperament is genetically based, character is shaped by environmental experiences, such as family relationships, peer relationships, trauma, and neglect. These dimensions appear to make up about 50 per cent of personality. There have been three dimensions of character identified that appear to mature in adulthood. These dimensions influence social and personal effectiveness by insight-learning about self-concepts. The three character dimensions are self-directedness, cooperativeness, and self-transcendence.

Low self-directedness and low cooperativeness are associated with all categories of personality disorder in the DSM-IV system.[17] Self-transcendence, on the other hand, does not differentiate patients with personality disorders from those without personality disorders.

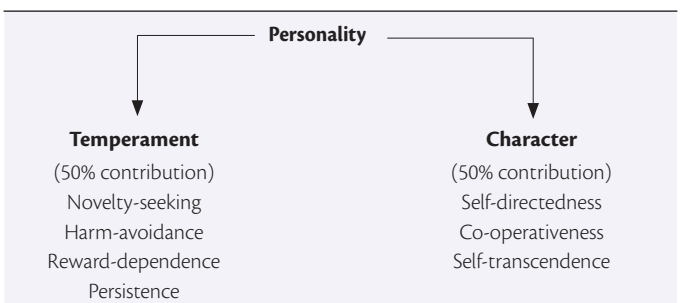
Self-directedness and cooperativeness reflect two fundamental tasks in personality development as defined by Blatt *et al.*[18]: the achievement of a stable, differential, realistic, and positive identity, and the establishment of enduring, mutually gratifying relationships with others. These two dimensions evolve in a dialectical and synergistic relationship to one another throughout the life cycle. Patients with character pathology tend to divide into two groups: introjective types, who are primarily focused on self-definition; and anaclitic types, who are more concerned about relatedness.

The character dimensions readily lend themselves to typical psychodynamic constructs. The self-directedness dimension is closely linked to what are often called ego functions or self-structures. The dimension of cooperativeness is a direct measure of a person's characteristic pattern of internal object relations as they are externalized in relationships with others. In one's assessment of a patient's personality, the transference–countertransference dimensions of the clinical interaction provide a privileged glimpse of the typical patterns of relatedness that cause difficulties in the patient's outside relationships.[19] The patient is involved in an ongoing attempt to actualize certain patterns of relatedness that reflect various wishes in the patient's unconscious. Through the patient's behaviour, he or she subtly tries to impose on the clinician a certain way of responding and experiencing.[20]

An individual internalizes a self-representation in interaction with an object representation connected by an affect through a series of repetitive interactions in childhood. This pattern ultimately leads to an internalized set of self- and object representations in interaction with one another. The adult individual repeats these patterns again and again as an effort to fulfill an unconscious wish. Even abusive or painful relationships involving a 'bad' or tormenting object may be wished for because of the safety and affirmation such relationships provide. In other words, a child who has been abused has internalized a highly conflictual abusive relationship as a predictable and familiar pattern. Having an abusive object may be preferable to having no object at all or being abandoned. Many patients with histories of an abusive childhood become convinced that the only way to remain connected to a significant person is to maintain an abuser–victim relationship.

The repetitive interactions seen in patients with personality disorder may reflect actual relationships with real objects in the past, but they may also involve wished-for relationships, such as those often seen in patients with childhood trauma who seek a rescuer. Clinicians who are influenced by the patient's interpersonal pressure to respond in a particular way may unconsciously accept the role in which they have been cast. When this phenomenon occurs, it is often referred to as projective identification.[8] In other words, the patient may 'nudge' the therapist into assuming the role of an abuser in response to the patient's 'victim' role, and the therapist may feel countertransference hate or anger and begin to make sarcastic or demeaning comments to the patient.

Table 3.3.2 Development of personality



Reproduced from Cloniger, C.R., Svarik, D.M. and Pryzbeck, T.R. (1993). A Psychobiological model of temperament and character. *Archives of General Psychiatry*, **50**, 975–90. Copyright 2003, The American Medical Association.

In addition to this pattern of object relations, the other major component of character, from a psychodynamic perspective, is the particular constellation of defence mechanisms that characterizes the individual patient.[21] While defences were traditionally regarded as intrapsychic mechanisms designed to prevent awareness of unconscious aggressive or sexual wishes, the current understanding of defence mechanisms has been expanded far beyond Freud's dual-drive theory. We now understand that defences also preserve a sense of self-esteem in the face of narcissistic vulnerability, assure safety when one feels dangerously threatened by abandonment, and serve to insulate one from external dangers through, for example, denial or minimization.

Different personality types or disorders use characteristic sets of defence mechanisms. For example, the paranoid personality may typically use projection as a way of disavowing unacknowledged feelings and attributing them to others. Patients with obsessive–compulsive personality disorder may use defensive operations such as isolation of affect, intellectualization, and reaction formation to control affective states that are highly threatening. In the relationship with the clinician, as noted previously, these defences will manifest themselves as resistances. Hence, if a patient with an obsessive–compulsive personality disorder uses intellectualization as a defence against painful affects, when the patient comes to treatment, intellectualization will be used as a resistance to avoid getting at feelings in psychotherapy.

Dynamic pharmacotherapy

The commonly used psychodynamic constructs, such as therapeutic alliance, transference, countertransference, and resistance apply to all modalities of psychiatric treatment, even though their usage is generally associated with psychodynamic psychotherapy. In a study of the relationship between the therapeutic alliance and the outcome of 250 depressed outpatients in the National Institute of Mental Health Treatment of Depression Study,[22] the therapeutic alliance was found to be of extraordinary importance. The patients had been randomly assigned to one of four conditions: brief interpersonal therapy, brief cognitive behavioural therapy, imipramine plus clinical management, or placebo plus clinical management. The researchers found that the therapeutic alliance was just as important for drug therapy as for psychotherapy. In all four treatment cells, the therapeutic alliance counted for more of the variance of treatment outcome than the treatment method itself. This was the first empirical study to show the importance of the therapeutic alliance in psychotherapy, pharmacotherapy, and placebo outcome.

Non-compliance is one of the most challenging problems facing psychiatric practitioners. Many factors go into compliance. Although many patients blame side effects, they often unconsciously undermine the treatment plan. The patient may have a negative transference to the prescribing clinician related to attitudes toward parents and other authority figures that lead the patient to rebel and defy the doctor's orders. Some clinicians may have countertransference reactions to specific patients that lead them to prescribe in a highly authoritarian manner or a tentative and ambivalent manner, giving unconscious messages to the patient that reflect the doctor's attitude about the medication. Patients who feel the doctor is bullying them to take the medication may not comply. Similarly, patients who sense their doctor is ambivalent about the value of the medication may also choose not to fill the prescription. Unconscious resistance is frequently a major factor in non-compliance. Medications may have idiosyncratic meanings to patients based on unconscious identifications with family members who have taken the same medication, views of psychiatric illness as moral weakness, or fears about the effects of the medication. Sometimes a pill or capsule may serve as a transitional object that substitutes for the person of the prescriber when the physician is unavailable. The colour or shape of a tablet may take on special significance for some patients, making them reluctant to change dosage or switch to another medication.

Multiple-treater settings

In inpatient units and partial hospital settings, psychodynamic concepts are of considerable value in understanding the patient's psychopathology as it unfolds in a group setting. Patients re-create their internal object relations in the inpatient or partial hospital milieu.[23] The conflicts that occur in their family context will re-emerge in their relationships with hospital staff members. Through projective identification the patient subtly pressures various staff members to play the roles that are in keeping with the patient's internal world. Hence a patient who has been physically and/or sexually abused by a parent will behave in such a way toward a certain nurse, for example, that the nurse begins to feel abusive toward the patient. The same patient may treat another nurse as an idealized rescuer figure, eliciting loving and protective feelings from that nurse. This form of splitting[23] may create extraordinary conflicts between staff members over the best treatment approach to the patient. Therefore failure to attend to the transference–countertransference dimensions of the milieu treatment may lead to a total disruption of the staff members' capacity to be effective with certain patients.

Moreover, individual patients often act out covert staff conflicts. Psychodynamically informed hospital treatment may help to identify these conflicts and allow staff to process them in such a way that the patients no longer need to enact them. When covert conflicts between staff members become overt and open to discussion, the patient's disruptive behaviour often settles down.[23] This observation reflects how the dynamics of the patient group and staff group often parallel each other. The psychodynamic clinician also understands that individuals behave differently in groups than they do alone or in a one-to-one context. Powerful group forces, such as scapegoating, can be recognized and processed so they do not become destructive to treatment. Similarly, the patient's recurrent problems in groups can be diagnosed, in part, by a careful study of the transference and countertransference responses.

Two-person context of treatment

One of the major shifts in psychodynamic thinking in recent years has been a greater acknowledgement of the influence of the clinician's perspective on the observations about the patient. Postmodern contributions from intersubjectivity and social constructivism have challenged the view that the clinician assesses the patient from a detached and objective frame of reference. Fundamental to this perspective is that clinicians can never transcend their irreducible subjectivity.[24] The psychiatrist in this context can never fully know how his or her subjectivity is influencing the diagnostic

assessment or the treatment process. Countertransference is viewed as both unconscious and continuous, so that a therapist cannot possibly be capable of keeping up with every emotional reaction of the patient.[25]

This two-person model of the treatment situation has contributed to the demise of the classical psychoanalytical view of the therapist or analyst as a blank screen or a dispassionate observer. The influence of the clinician's biases and unconscious feelings toward the patient may have far-reaching implications for a variety of situations in psychiatry. Frustration about a patient's nonresponsiveness to treatment, for example, can lead a clinician to recommend electroconvulsive therapy as a reaction to despair, rather than as a result of systematic decision trees or algorithms about refractory depression. Even in the case of physician-assisted suicide, countertransference may play a major, though hidden, role.[26] Within this context the patient's wish to die may stem from a self-concept as worthless and a burden to others that is, in part, a reflection of what the physician brings to the encounter. Similarly, the doctor's death-anxiety might underlie an omnipotent need to triumph over death through the prescription of physician-assisted suicide that strives to preserve an illusion of control and mastery. In the worst scenario, a clinician's intense countertransference hate toward a patient may lead to a wish to kill that is transposed into a recommendation for physician-assisted suicide.

Psychodynamic psychotherapy for specific disorders

A psychodynamic approach is relevant to the treatment of the vast majority of psychiatric disorders encountered in clinical practice. Depending on the nature of the illness, the setting in which the illness is treated, and the psychological mindedness of the patient, psychodynamic strategies may be the major emphasis in the treatment plan or a relatively minor contribution. Psychodynamic psychotherapy per se is generally divided into short-term psychodynamic psychotherapy (STPP) and long-term psychodynamic psychotherapy (LTPP). The former is generally regarded as involving fewer than 24 sessions or 6 months' duration, while the latter is viewed as a therapy lasting more than 6 months.[27] Psychodynamic psychotherapy, whether long-term or short-term, is often defined as 'a therapy that involves careful attention to the therapist–patient interaction, with thoughtfully timed interpretation of transference and resistance embedded in the sophisticated appreciation of the therapist's contribution to the two-person field'.[28] This form of psychotherapy is also conceptualized as operating on an expressive-supportive continuum. The highly expressive forms of psychodynamic psychotherapy offer more interpretation of unconscious conflict, while the forms that are more supportive focus on bolstering adaptive defences and building self-esteem. The continuum of interventions from the most expressive to the most supportive (see Table 3.3.3) guide the psychotherapist in how to intervene with any given patient.

Short-term psychodynamic psychotherapy

A recent meta-analysis of STPP[29] found that this modality made significant changes in general psychiatric symptoms, target problem, and social functioning. The treatment also yielded significant and large pre-treatment—post-treatment effect sizes. The effect sizes were stable and tended to increase at follow-up. No significant differences were found between STPP and other forms of psychotherapy. Evidence from randomized controlled trials supports the use of STPP for major depressive disorder, panic disorder, social phobia, and post-traumatic stress disorder. In addition, the treatment has also been efficacious in somatoform disorders, bulimia nervosa, and substance-related disorders in association with drug counseling.

Long-term psychodynamic psychotherapy

Research on LTPP has been more limited than for STPP because the gold standard of the randomized controlled trial is more difficult to implement when studying LTPP. One must find a suitable control group where an alternative extended treatment or a placebo condition can be implemented. The most rigorous controlled condition is an alternative extended treatment, although some investigations have used treatment as usual as well.[30] The dropout rate can also be problematic in long-term studies. In addition, intervening life events, Axis I conditions, and medication shifts can influence outcome. Nevertheless, despite the obstacles to designing and implementing rigorous research on LTPP, there are a number of studies that have appeared in recent years that suggest that LTPP is an efficacious treatment.

Two randomized controlled trials have shown LTPP to be efficacious with Cluster C personality disorders. Included in this group are avoidant, dependent, and obsessive–compulsive personality disorders. One study compared 40 sessions of psychodynamic therapy to control patients on a waiting list and found substantially better outcomes than those who received the dynamic therapy.[31] In a more rigorously designed study,[32] 40 sessions of dynamic psychotherapy were compared to 40 sessions of cognitive behaviour therapy. While both treatments were effective, the dynamic therapy resulted in continued improvement after termination of treatment, suggesting that patients internalized the therapeutic dialogue and used it to deal with problems as they arose.

Borderline personality disorder has also been subjected to rigorous trials of long-term psychodynamic psychotherapy. In one head-to-head comparison between a form of LTPP known as transference-focused psychotherapy (TFP), dialectical behaviour therapy (DBT), and supportive psychotherapy (SPT),[33,34] all three modalities

Table 3.3.3 An expressive–supportive continuum of interventions

Expressive						Supportive
Interpretation	Confrontation Observation	Clarification	Encouragement to Elaborate	Empathic Validation	Advice and Praise	

(Data modified from Gabbard, G.O. (2004), *Long-term psychodynamic psychotherapy: a basic text*, American Psychiatric Publishing, Arlington, VA.)

showed general improvement. However, TFP showed improvements that were not demonstrated by either SPT or DBT. Participants in the study who received TFP were more likely to move from an insecure attachment classification to a secure one, show greater changes in mentalizing capacity, and have more extensive symptomatic improvement than the other two groups. Only TFP made significant changes in impulsivity, irritability, verbal assault, and direct assault. Suicidality was reduced to an equal extent by TFP and DBT.

A psychoanalytically oriented partial hospitalization treatment for patients with borderline personality disorder was compared to a treatment-as-usual approach.[35,36] In the treatment group, the major difference was the provision of individual and group psychotherapy compared to the control condition. The treatment lasted a maximum of 18 months, and was significantly superior to standard psychiatric care, both at the end of therapy and at the 18-month follow-up.

A randomized controlled trial for children with learning disabilities[37] compared intensive psychodynamic therapy (four times a week) to once-a-week sessions. This trial on featured treatments had lasted longer than 1 year. In the follow-up assessment, children who had sessions four times a week showed much greater improvement.

Future directions

The psychodynamic model continues to enrich the patient's understanding and the psychiatrist's practice. The time-honoured principles elaborated here serve as windows into the murky recesses of the unconscious and illuminate human motivation. They also provide the clinician with a 'second sight' that helps make sense out of bewildering and complex clinical situations.

The evidence regarding the impact of psychotherapy on the brain opens up new lines of investigation to enhance our understanding of psychopathology and treatment. These include the following:

1 the mechanisms of action of psychotherapy

2 the interrelationships between the mechanisms of action of psychotherapy and medication

3 a clearer understanding of pathogenesis itself and the malleability of some components of the pathogenetic mechanisms of major psychiatric disorder

Research is sorely needed on psychodynamic treatments because there is only a modest empirical base for psychodynamic therapy. Many more studies are needed, especially those with a randomized controlled design targeted at specific disorders. Studies investigating extended dynamic therapy of a year or more are needed to demonstrate which patients benefit from the additional investment of time and money. In the current climate of cost containment, practitioners of psychodynamic therapy must take cost-effectiveness into account.

The optimal treatment for many psychiatric patients involves a combination of medication and psychotherapy, but research support for this view is also rather modest. Controlled studies of combined treatment versus single modalities are needed for personality disorders and anxiety disorders. In addition, the role of psychodynamic thinking in compliance problems needs rigorous investigation.

In the meantime, the psychodynamic model continues to focus on the uniqueness of the individual patient. Psychodynamic psychiatry, above all, is interested in the person with illness rather than the illness alone.

Further information

Gabbard, G.O. (2004). *Long-term psychodynamic psychotherapy: a basic text*. American Psychiatric Publishing, Arlington, VA.

Gabbard, G.O. (2005). *Psychodynamic psychiatry in clinical practice* (4th edn). American Psychiatric Publishing, Arlington, VA.

Leichsenring, F., Rabung, S., and Liebing, E. (2004). The efficacy of short-term psychodynamic therapy in specific psychiatric disorders: a meta-analysis. *Archives of General Psychiatry*, **61**, 1208–16.

References

1. Gabbard, G.O. (2005). *Psychodynamic psychiatry in clinical practice* (4th edn). American Psychiatric Publishing, Arlington, VA.
2. Squire, L.R. (1987). *Memory and brain*. Oxford University Press, New York.
3. Westen, D. (1999). The scientific status of unconscious processes: is Freud really dead? *Journal of the American Psychoanalytic Association*, **47**, 1061–106.
4. Anderson, M.C., Ochsner, K.N., Kuhl, B., *et al.* (2004). Neural systems underlying the suppression of unwanted memories. *Science*, **303**, 232–5.
5. Sherwood, M. (1969). *The logic of explanation in psychoanalysis*. Academic Press, New York.
6. Winnicott, D.W. (1949). Hate in the countertransference. *International Journal of Psycho-analysis*, **30**, 69–74.
7. Kernberg, O.F. (1965). Notes on countertransference. *Journal of the American Psychoanalytic Association*, **13**, 38–56.
8. Gabbard, G.O. (1995). Countertransference: the emerging common ground. *International Journal of Psycho-analysis*, **76**, 475–85.
9. Freud, S. (1912). The dynamics of transference. In *The standard edition of the complete works of Sigmund Freud*, Vol. 12 (ed. J. Strachey), pp. 97–108. Hogarth Press, London, 1958.
10. Andreasen, N.C. (1997). Linking mind and brain in the study of mental illness: a project for a scientific psychopathology. *Science*, **275**, 1587–93.
11. Kandel, E.R. (1983). From metapsychology to molecular biology: explorations into the nature of anxiety. *The American Journal of Psychiatry*, **140**, 1277–93.
12. Kandel, E.R. (1998). A new intellectual framework for psychiatry. *The American Journal of Psychiatry*, **155**, 457–69.
13. Caspi, A., McClay, J., Moffitt, T.E., *et al.* (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, **297**, 851–4.
14. Reiss, D., Plomin, R., and Hetherington, E.M. (1991). Genetics and psychiatry: an unheralded window on the environment. *The American Journal of Psychiatry*, **148**, 283–91.
15. Kendler, K.S., Kessler, R.C., Walters, E.E., *et al.* (1995). Stressful life events, genetic liability, and onset of an episode of major depression in women. *The American Journal of Psychiatry*, **152**, 833–42.
16. Gabbard, G.O. (1992). Psychodynamic psychiatry in the decade of the brain. *The American Journal of Psychiatry*, **149**, 991–8.
17. Cloninger, C.R., Svrakic, D.M., and Pryzbeck, T.R. (1993). A psychobiological model of temperament and character. *Archives of General Psychiatry*, **50**, 975–90.
18. Blatt, S.J., Ford, R.Q., Berman, W.H. Jr., *et al.* (1994). *Therapeutic change: an object relations perspective*. Plenum Press, New York.
19. Gabbard, G.O. (1997). Finding the 'person' in personality disorders. *The American Journal of Psychiatry*, **154**, 891–3.
20. Sandler, J. (1981). Character traits and object relationships. *The Psychoanalytic Quarterly*, **50**, 694–708.
21. Gabbard, G.O. (1999). Psychoanalysis and psychoanalytic therapy. In *The DSM-IV personality disorders* (ed. W.J. Livesley). Guilford Press, New York.

22. Krupnick, J.L., Sotsky, S.M., Simmens, S., *et al*. (1996). The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: findings in the National Institute of Mental Health treatment of depression collaborative research program. *Journal of Consulting and Clinical Psychology*, **64**, 532–9.
23. Gabbard, G.O. (1989). Splitting in hospital treatment. *The American Journal of Psychiatry*, **146**, 444–51.
24. Renik, O. (1993). Analytic interaction: conceptualizing technique in light of the analyst's irreducible subjectivity. *The Psychoanalytic Quarterly*, **62**, 553–71.
25. Hoffman, I.Z. (1998). *Ritual and spontaneity in the psychoanalytic process: the dialectical constructivist view*. Analytic Press, Hillsdale, NJ.
26. Varghese, F.T. and Kelly, B. (1999). Countertransference and assisted suicide. In *Review of psychiatry*, Vol. 18 (eds. J. Oldham and M. Riba), *Countertransference in psychiatric treatment* (ed. G. Gabbard), pp. 85–118. American Psychiatric Press, Washington, DC.
27. Gabbard, G.O. (2004). *Long-term psychodynamic psychotherapy: a basic text*. American Psychiatric Publishing, Arlington, VA.
28. ibid., p. 2.
29. Leichsenring, F., Rabung, S., and Liebing, E. (2004). The efficacy of short-term psychodynamic therapy in specific psychiatric disorders: a meta-analysis. *Archives of General Psychiatry*, **61**, 1208–16.
30. Gabbard, G.O., Gunderson, J.G., and Fonagy, P. (2002). The place of psychoanalytic treatments within psychiatry. *Archives of General Psychiatry*, **59**, 505–10.
31. Winston, A., Laikin, M., Pollack, J., *et al*. (1994). Short-term psychotherapy of personality disorders. *The American Journal of Psychiatry*, **151**, 190–4.
32. Svartberg, M., Stiles, T.C., and Seltzer, M.H. (2004). Effectiveness of short-term dynamic psychotherapy and cognitive therapy for cluster C personality disorders: a randomized controlled trial. *The American Journal of Psychiatry*, **161**, 810–17.
33. Clarkin, J.F., Levy, K.N., Lenzenweger, M.F., *et al.* (2007). Evaluating three treatments for borderline personality disorder: a multiwave study. *The American Journal of Psychiatry*, **164**, 922–28.
34. Levy, K.N., Meehan, K.B., Clarkin, J.F., *et al*. (2006). Change in attachment patterns and reflective function in a randomized control trial of transference-focused psychotherapy for borderline personality disorder. *Journal of Consulting and Clinical Psychology*, **74**, 1027–74.
35. Bateman, A. and Fonagy, P. (1999). The effectiveness of partial hospitalization in the treatment of borderline personality disorder: a randomized control trial. *The American Journal of Psychiatry*, **156**, 1563–9.
36. Bateman, A. and Fonagy, P. (2001). Treatment of borderline personality disorder with psychoanalytically oriented partial hospitalization: an 18-month follow-up. *The American Journal of Psychiatry*, **158**, 36–42.
37. Heinicke, C.M. and Ramsey-Klee, D.M. (1986). Outcome of child psychotherapy as a function of frequency of sessions. *Journal of the American Academy of Child Psychiatry*, **25**, 247–53.

SECTION 4

Clinical Syndromes of Adult Psychiatry

This page intentionally left blank

4.1

Delirium, dementia, amnesia, and other cognitive disorders

Contents

4.1.1 Delirium

David Meagher and Paula Trzepacz

Introduction

Delirium is an acute or subacute, usually reversible syndrome of impaired higher cortical functions hallmarked by generalized cognitive disturbance and caused by one or more aetiologies. It is most common in medical-surgical patients, especially in intensive care units, and those in hospice and nursing homes. The term 'delirium' derives from the Latin '*lira*' meaning literally to wander from the furrow. Prior to DSM-III (1980) such disturbances were described by a plethora of labels (acute organic brain syndrome, acute confusional state, brain failure, toxic encephalopathy, intensive care psychosis), before Lipowski advocated for the umbrella term delirium to subsume these multiple synonyms. This engendered a more scientific research effort and consistent approach to detection and management. Though delirium has been recognized for at least two millenia, it is only now beginning to receive the attention that it warrants, with increasing appreciation of the considerable impact upon outcomes and independent need for treatment as a brain disorder beyond only treating its underlying aetiological precipitants.

Inattention is the cardinal disturbance, including distractibility, reduced vigilance or concentration, and impaired environmental awareness. This contrasts with dementia, another disorder of generalized cognitive deficits, where memory deficits are cardinal. While full-blown episodes are easier to diagnose, its prodrome, subclinical presentation, and potential persistence present unresolved dilemmas regarding diagnostic boundaries of delirium. Further, comorbidity with dementia presents challenges for detection and attribution of progressive impairments in the elderly. Though delirium occurs at any age, there is a dearth of research in younger age groups such that it is unclear whether research findings from geriatric studies can be generalized to other age groups (e.g. regarding risk factors and outcomes). Studies that clarify common features such as phenomenology, neural circuitry or electrophysiology are thus critical.

Clinical features

Delirium is a complex neuropsychiatric syndrome with a broad range of cognitive and neurobehavioural symptoms which is why it can be misattributed to other psychiatric disorders by nonspecialists. Symptoms involve cognition, thought, language, sleep-wake cycle, perception, affect, and motor behaviour. The constellation of symptoms—along with the cardinal symptom of inattention and acute onset and fluctuating temporal course—are characteristic of delirium and when comorbid with dementia, delirium dominates the clinical presentation.[1] Phenomenology studies (mostly cross-sectional) suggest that 'core' symptoms occur with greater frequency while other less consistent 'associated' symptoms may reflect the biochemical influence of particular aetiologies or genetic, neuronal or physiological vulnerabilities (see Table 4.1.1.1). Accumulating evidence indicates three core domains of delirium phenomenology: 'Cognition' comprising of inattention and other cognitive deficits; 'Higher Level Thinking Processes' including impaired executive function, semantic expression, and comprehension; and 'Circadian Rhythm' including fragmented sleep-wake cycle.[2] The underlying neural support for these domains is consistent with neuroanatomical findings in lesion and functional neuroimaging studies that implicate certain brain regions and neural circuitry.

Table 4.1.1.1 Symptoms of delirium

Diffuse cognitive deficits
Attention
Orientation (time, place, person)
Memory (short- and long-term; verbal and visual)
Visuoconstructional ability
Executive functions
Temporal course
Acute/abrupt onset
Fluctuating severity of symptoms over 24-hr period
Usually reversible
Subclinical syndrome may precede and/or follow the episode
Psychosis
Perceptual disturbances (especially visual), including illusions, hallucinations, metamorphosias
Delusions (usually paranoid and poorly formed)
Thought disorder (tangentiality, circumstantiality, loose associations)
Sleep-wake disturbance
Fragmented throughout 24-hr period
Reversal of normal cycle
Sleeplessness
Psychomotor behavior
Hyperactive
Hypoactive
Mixed
Language impairment
Word-finding difficulty/dysnomia/paraphasia
Dysgraphia
Altered semantic content
Severe forms can mimic expressive or receptive aphasia
Altered or labile affect
Any mood can occur, usually incongruent to context
Anger or increased irritability common
Hypoactive delirium often mislabeled as depression
Lability (rapid shifts) common
Unrelated to mood preceding delirium

(Reproduced from Trzepacz, P. T., Meagher, D.J. *Delirium, Chapter 11 in American Psychiatric Publishing Textbook of Neuropsychiatry*, (5th edn) eds. S. Yudofsky and R. Hales, copyright 2007, American Psychiatric Publishing, Inc., Washington DC.)

Delirium occurs as a stage of consciousness in the continuum between normal awakeness/alertness and stupor or coma. During the 20th century, delirium was described as a 'clouding of consciousness' but this rather nebulous concept has been replaced by a better understanding of the components of phenomenology that culminate in severely impaired higher order brain functions. Specifically, a disproportionate disturbance of attentional processes, including environmental awareness difficulties, along with impaired higher level thinking reflected in irrelevant, unfocused or illogical thought processes and impaired abstraction and comprehension (i.e. executive cognition and semantic language function) typifies the delirious state. Sleep-wake cycle fragmentation belies a circadian disturbance that may contribute to the abnormal level of consciousness and alterations in motor behaviour, where hypoactivity contributes to difficulties in differential diagnosis of delirium from stupor. Delirium is distinguished from stuporose states by the presence of arousability. The majority of intensive care unit patients emerging from comatose states experience a period of diagnosable delirium.[3]

Inattention is the cardinal and required symptom to diagnose delirium and is noticeable on interview by distractibility, spatial inattention, and inability to sustain attention. More formal testing can be assessed using months of the year backwards or digit span. The Cognitive Test for Delirium (CTD)[4] allows for separate visual digit span and vigilance testing. Memory impairment—of both short and long term—can be affected by inattention but appears to be independently impaired. Visuospatial impairment can be assessed by observing patient behaviour in their immediate environment e.g. losing their way or getting lost. Constructional ability can be tested formally by copying figures; clock face drawing assesses not only proportions and details but also involves prefrontal executive functions for placing the minute hand correctly. Delirious patients have executive dysfunction affecting abstraction, initiation/perseveration, switching mental sets, working memory, temporal sequencing and organization, insight and judgment. Poor performance on the Trailmaking Part B test distinguishes delirious from nondelirious patients and requires not only spatial attention and concentration but also switching mental sets. Though none of these cognitive deficits is specific to delirium, the array and pattern is highly suggestive.

Thought process abnormalities in delirium range from circumstantiality and tangentiality to frank loose associations in more severe cases. Naming impairment is common though more severe cases can mimic fluent dysphasia with semantic deficits being characteristic such that communication is wrought with deficits of meaningfulness. Interestingly, the dysphasia can be mistaken for Wernicke's aphasia and possible stroke but is reversible when the delirium clears. Careful assessment can usually distinguish between semantic deficits (language impairment) and loose associations (thought process disorder) except with word salad.

Disruption of sleep-wake cycle is essentially ubiquitous in delirium except in the briefest episodes (e.g. concussion) and often predates the appearance of a full-blown episode. Minor disturbances with insomnia or excessive daytime somnolence may be hard to

distinguish from other medically ill patients without delirium, but more substantial alterations involve sleep fragmentation or even complete sleep-wake cycle reversal that reflect disturbed circadian rhythm regulation. The relationship of circadian disturbances to the characteristic fluctuating severity of delirium symptoms over a 24 hr period or to motor disturbance is unknown.

Motor activity alterations are very common in delirium. They have been used to define clinical subtypes (hypoactive, hyperactive, mixed) though studies are inconsistent as to the prevalence of these subtypes and often include nonmotor symptoms in descriptions. Cognitive impairments and EEG slowing are comparable in hyperactive and hypoactive patients though other symptoms may vary. Psychotic symptoms occur in both although the prevailing stereotype suggests that they only occur in hyperactive cases. Hypoactive cases are prone to non detection or misdiagnosis as depression. A range of studies suggest that motor subtypes differ regarding underlying pathophysiology, treatment needs, and prognosis for function and mortality though inconsistent subtype definitions and delayed/poorer detection of hypoactives impacts interpretation of these findings.

Psychotic symptoms occur in up to 50 per cent of patients with delirium. Thought content abnormalities include suspiciousness, overvalued ideation and frank delusions. Delusions are typically poorly-formed and less stereotyped than in schizophrenia or Alzheimer's disease. They usually relate to persecutory themes of impending danger or threat in the immediate environment (e.g. being poisoned by nurses), and less commonly to grandiose themes. Misperceptions include depersonalization, delusional misidentifications, illusions and hallucinations. Hallucinations and illusions are primarily visual though they can be tactile and auditory whereas auditory modalities tend to dominate in psychosis in mood and primary psychotic disorders. Formications suggest dopamine or anticholinergic toxicity.

Delirium may be abrupt as with concussion, drug intoxication or stroke, or can be preceded by a prodromal period characterized by anxiety, sleep disturbance, cognitive impairments and increased levels of perceived distress. Symptom profile appears similar across age groups[5] but delirium is understudied in pediatric patients. The propensity for particular aetiologies to shape clinical presentation is also understudied. Unfortunately, delirium tremens (with florid psychosis and agitation) is the dominant clinical stereotype even though many cases present with relative hypoactivity especially in the elderly, those with concomitant dementia, or where delirium is related to metabolic causes or organ failure. This misleading stereotype is one of the reasons for the poor recognition of delirium where typically 50 per cent of cases are missed in routine clinical practice.

Diagnosis and differential diagnosis

ICD-10 and DSM-IV share key features used to diagnose delirium (i.e. acute onset, fluctuating course, inattention, and disorganized thinking) although the ICD-10 description gives better account of the breadth of symptoms that can occur (e.g. disturbances of sleep and motor activity). DSM IV is more inclusive and preferred in research studies, though may be less rigorous than ICD-10 when used by less skilled clinicians. DSM-IV classifies cases according to presumed aetiological cause, though a single aetiology occurs in less than half of cases and no aetiology is identified in around 10 per cent. Delirium is also subclassified according to its relationship to dementia.

Delirium is poorly detected in clinical practice by nonpsychiatrists with more than 50 per cent of cases missed, misdiagnosed, or diagnosed late. This is due to multiple factors: the complex and fluctuating nature of delirium symptoms, inadequate education and interview skills of nonpsychiatrists, underappreciation of the prognostic significance of delirium, and inadequate routine cognitive screening in real world practice. Delirium can be the first indicator of serious physical morbidity (e.g. stroke) and represents a medical emergency. It is not surprising therefore that nondetection is associated with poorer outcomes that include elevated mortality.[6] Poor outcomes in hypoactive patients may be in part due to nondetection.

The course of delirium is highly variable reflecting the heterogeneity of aetiology and patient populations in which delirium occurs, with recent studies emphasizing that it is frequently not the benign and transient condition that was previously thought. While in many cases, delirium is brief (hours to days), represents a transitional state from unconsciousness or is a benign reaction to treatment exposures, in other cases it can be more prolonged (e.g. after traumatic brain injury) or associated with serious complications and persistent cognitive difficulties where differentiation from dementia becomes difficult. Rudberg *et al.*[7] studied elderly medical-surgical inpatients with delirium and found that episode duration was 24 hrs or less in over two-thirds of patients. Conversely, Sylvestre *et al.*[8] studied elderly medical admissions over two-month follow-up and identified five separate patterns of recovery, with fast improvement in only 11 per cent of patients. Greater clarity regarding the factors that shape these varying courses is needed.

Delirium diagnosis is complicated by *comorbidity* where over 50 per cent of cases are superimposed on dementia or other pre-existing cognitive impairments. Distinguishing delirium from the neuropsychiatric symptoms of dementia can be challenging but acute onset, fluctuating course, temporal relationship to an identifiable physical precipitant, prominent inattention and altered level of consciousness usually allow differentiation. Third party informants and previous medical charts can be crucial in clarifying the trajectory of cognitive impairment. Studies comparing symptoms of delirium and dementia indicate that where they coexist, delirium symptoms dominate the clinical picture. Given the poor prognostic implications of delirium, a management hierarchy applies with delirium taking diagnostic precedence over other neuropsychiatric disorders so that any acute alteration in mental state is presumed to be delirium until otherwise established.

Some symptoms of delirium also *overlap with primary psychiatric disorders*. Major depressive disorder can be misdiagnosed in hypoactive presentations or when affective lability includes tearfulness and sad mood. Agitated depression or severe mania ('Bell's mania') can mimic hyperactive delirium but the affective lability and incongruent moods of delirium contrast with more sustained alterations in mood and effect in major mood disorders. The character of psychotic symptoms in delirium differs from primary psychotic illness (see above). Acute schizophrenic psychosis involves disorganized thoughts with delusions and hallucinations but inattention is less prominent. Acute schizophrenia can include marked cognitive impairment with perplexity that can mask or be mistaken for comorbid delirium and in such cases the EEG can be helpful. Table 4.1.1.2 describes key clinical features for differentiating delirium from other neuropsychiatric conditions.

Table 4.1.1.2 Differential diagnosis of delirium vs other common neuropsychiatric conditions

	Delirium	Dementia	Depression	Schizophrenia
Onset	Acute	Insidious[a]	Variable	Variable
Course	Fluctuating	Often progressive	Diurnal variation	Variable
Reversibility	Frequently[b]	Not usually	Usually but can be recurrent	Chronic relapsing and remitting course typical
Level of consciousness	Impaired	Unimpaired until late stages	Generally unimpaired	Unimpaired (perplexity in acute stage)
Attention/memory	Inattention is primary with poor memory	Poor memory without marked inattention except in end-stage illness	Mild attention problems, inconsistent pattern – depressive pseudodementia, memory intact with formal testing	Poor attention, inconsistent pattern, memory intact
Affect	Lability	No clear pattern	Flattening	Incongruity
Hallucinations	Usually visual; can be auditory, tactile, gustatory, olfactory	Can be visual or auditory	Usually auditory	Usually auditory
Delusions	Fleeting, fragmented, and usually persecutory often relate to immediate environment or impending danger	Paranoid, often fixed, relate to misconceptions	Complex and mood congruent e.g. themes of guilt or nihilism	Frequent, complex, systematized, and often paranoid

[a]Except for large strokes that can be abrupt and Lewy Body Dementia which can be subacute.

[b]Can be chronic (paraneoplastic syndrome, central nervous system adverse events of medications, severe brain damage).

Nondetection of delirium is particularly common in older patients with comorbid dementia, multiple medical problems and hypoactive motor presentations. Chronic subsyndromal delirium in the elderly is commonly related to low grade infections or medication adverse effects, where adjusting medications can significantly improve cognition. Monitoring for any acute deterioration from baseline function coupled with regular formal assessment with simple cognitive tests such as the digit span, months of year backwards, serial sevens or clock drawing enables delirium detection. Unfortunately, over-reliance on orientation as a measure of cognition precludes more accurate detection. The emphasis on orientation, inconsistent administration, and ceiling effects limit the usefulness of the (MMSE) in measuring delirium. The Cognitive Test for Delirium was designed specifically for delirium and emphasizes attention, semantic comprehension, and nonverbal, right hemisphere cognitive functions. It is particularly useful in critically medically ill persons.

The *Confusion Assessment Method* (CAM)[9] is a screening tool to assess the presence or absence of four items from DSM-III-R delirium criteria to make a provisional diagnosis of delirium. It is especially suited to epidemiological studies and screening in high risk populations where neuropsychiatric differential diagnosis or broad phenomenological measurements are not needed. Its accuracy is enhanced if ratings are anchored by formal testing as in the CAM-ICU[10] but is substantially reduced when used by nurses because they frequently miss inattention when it is present.

Psychiatrists and delirium specialists use *more detailed instruments* for more specific and sensitive detection of a broader range of symptoms. The Delirium Rating Scale-Revised-98 (DRS-R98)[11] and the Memorial Delirium Rating Scale (MDAS)[12] are the most widely used rating scales, and include measures of a wide breadth of symptoms. The MDAS is a severity scale used in conjunction with a DSM or ICD diagnosis whereas the DRS-R98 is both a diagnostic and severity instrument where each item rating is anchored by phenomenological descriptions.

In clinically challenging situations, an *EEG* can be used to help differentiate delirium from other neuropsychiatric disorders where generalized slowing of the dominant posterior rhythm is characteristic. Additionally, clues for specific disorders like complex partial status epilepticus can be identified.

Epidemiology and outcomes

Epidemiological studies have focused on elderly hospitalized populations with far less research in younger age groups or the general population. Delirium occurs in all age groups but those at age extremes, with pre-existing cognitive impairment, cancer, and the critically ill have especially high rates. It is estimated that around 10 per cent to–15 per cent of general hospital patients have delirium upon admission with a further 10–40 per cent developing delirium during hospitalization. Overall frequency is estimated at 11 per cent to–42 per cent[13] with the clinical rule of thumb that one in five general hospital inpatients experience delirium at some time during hospitalization. Delirium incidence is expected to increase as demographics of the general population shift toward older ages and with higher prevalence of dementia and cerebrovascular and cardiovascular disorders, although improved medical care for elderly persons may offset this pattern. Presence of APOE-4 alleles may confer increased risk for poorer recovery from delirium, reflecting vascular and neurodegenerative influences.

Delirium episodes are associated with elevated morbidity, longer hospital stays, greater costs of care, and higher frequency of complications.[14] Moreover, in the elderly, reduced post hospital independence and elevated one year mortality rates occur. Although the latter is partly due to the effects of age, frailty, comorbid dementia, severity of medical comorbidities, and medication exposure, some

epidemiological studies identify delirium as an independent predictor of poorer outcomes in the elderly. However, most studies have not adequately accounted for premorbid vulnerability and cognition, burden of medical problems, and pharmacological effects such that delirium may simply be a marker of underlying pathology causing poor outcomes. Delirium incidence or severity in the elderly can be improved by earlier specialist intervention including the judicious use of haloperidol and more comprehensive delirium care, though the magnitude of this effect remains unclear.

Many report a new diagnosis of *dementia after an episode of delirium*. Some evidence suggests that persistent cognitive impairment can occur after delirium even in those thought to be premorbidly cognitively intact[15,16], while others find that baseline status or medical burden predicts outcome and follow-up reveals progression of an incipient dementing process. Contributors to post delirium cognitive decline include baseline CNS vascular or neurodegenerative pathology, physical problems that also caused the delirium, toxic effects of or inability to comply and benefit from treatments, unresolved delirium, an accelerating effect of delirium on cognitive decline, or possibly a direct neurotoxic effect of the delirious state itself. However, it is still unclear whether delirium plays a causal role or is simply a marker of medical morbidity and preexisting baseline vulnerabilities both of which are associated with delirium. Pharmacological effects (beneficial or adverse effects) are unaccounted for in most studies. The possibility that delirium itself is neurotoxic remains unproven, though could theoretically involve the effects of neurochemical abnormalities associated with delirium (e.g. dysfunctional cellular metabolism or glutamatergic surges). Studies in younger age groups are needed to disentangle confounds from aging.

Elevated mortality rates (ranging from 4 per cent to–65 per cent) during the index admission may reflect a variety of factors including the impact of underlying physical causes of delirium, the consequences of reduced ability to cooperate with medical care, the complications that occur due to the delirium symptoms (e.g. pressure sores, infections, falls). Critical review of methodology suggests that delirium mostly carries an associated increased mortality risk during the year following an episode. Mortality risk is related to agedness, severity of underlying physical illness, presence of dementia, timing of diagnosis, motor presentation and delirium symptom severity. Mortality is elevated even when the confounding effects of age, medical morbidity, and medication exposure are accounted for.[17,18] Mortality is also elevated in hospitalized patients with up to four selected delirium symptoms ('subsyndromal') but without meeting syndromal criteria,[19] highlighting the need for careful assessment and monitoring of patients at risk.

Risk factors

Delirium is a multifactorial condition. A typical episode reflects the cumulative effects of predisposing risk factors, individual patient vulnerabilities (including genetic), and precipitating aetiological insults. A wide range of patient, illness, and treatment variables increase the likelihood of developing delirium but preexisting cognitive impairment, any CNS disorder, age extremes, low serum albumin, and exposure to particular medications are particularly robust predictors of delirium across populations. Geriatric and paediatric medically ill patients may share risk factors such as more vulnerable cholinergic neurotransmission—related to aging effects and developmental immaturity, respectively. Some risk factors are also considered aetiologies (e.g. UTI, anticholinergic medications). Table 4.1.1.3 lists a variety of reported risk factors across a number of reports.

Table 4.1.1.3 Factors associated with an increased risk for delirium

1. Patient vulnerabilities
Age extremes
Pre-existing cognitive impairment e.g. dementia
CNS disorder
Genetic factors e.g. APOE genotype
Visual deficit
Hearing deficits
Poor nutritional status
Previous episode of delirium
2. Environmental
Social isolation
Sensory extremes
Immobility
Novel environment
Stress
Use of restraints
ICU stay
3. Medical
Severe medical illness
Burns
HIV / AIDS
Organ insufficiency
Infection (e.g. UTI)
Hypoxemia
Fracture
Hypothermia / fever
Metabolic disturbances
Dehydration
Elevated BUN
Low serum albumin
Nicotine withdrawal
Increased blood-brain barrier permeability
Uncontrolled pain
4. Procedure-related
Peri-operative
Type of surgery (e.g. hip)
Emergency procedure
Duration of operation
Urinary Catheterization
Artificial respiration
5. Drug-related
Polypharmacy
Drug / alcohol dependence
Psychoactive drug use
Specific agents (e.g. anticholinergics / opiates/ benzodiazepines)

The interaction between predisposition (baseline vulnerability) and precipitating insults account for delirium incidence. Inouye and Charpentier[20] developed a model of four common predisposing and five precipitating factors that predicted a 17-fold variation in delirium risk in elderly medical patients, which has been replicated in post-operative elderly patients.[21] To date most genetic studies have focussed on genotyes related to increased risk of alcohol withdrawal delirium, though APOE-4 allele genotype has been linked to longer duration of delirium in ICU patients.[22]

Aetiology

Single-aetiology delirium is the exception with typically 3–4 significant causative factors relevant during any single episode which interact and overlap sequentially to produce or sustain delirium symptoms. It is crucial that potential aetiologies are constantly reevaluated throughout a delirium episode and even after a single cause is identified, efforts to unearth other factors should continue. Categories of delirium aetiologies include drug intoxication, drug withdrawal, metabolic/endocrine, traumatic brain injury, seizures, intracranial infection, systemic infection, intracranial neoplasm, extracranial neoplasm, cerebrovascular disorder, organ insufficiency, other CNS disorder, and other systemic factors (heat stroke, radiation, hypothermia, etc). Table 4.1.1.4 lists clinical investigations recommended in routine evaluation of delirium and additional tests indicated in particular cases.

Among the most common causes are infections and those related to illicit and prescribed drugs and alcohol, either in toxicity or withdrawal. Additionally, when serum albumin levels are low, more unbound drug is available to cause adverse events. Increased blood-brain barrier permeability (e.g. uremia, sepsis) can allow passage of drugs that ordinarily do not cross into the brain and delirium can result.

Neuropathogenesis

Delirium reflects a generalized disturbance of brain function as evidenced by the broad range of neuropsychiatric symptoms, diffuse slowing on EEG, and widespread alterations in cerebral blood flow.[23] Despite the range of underlying aetiologies, delirium presents with a relatively consistent clinical profile and is thus considered a unitary syndrome reflecting a final common neural pathway for multiple diverse causes and pathophysiologies.[24,25] When studying delirium pathophysiology, it is important to distinguish between physiological mechanisms of aetiologies and neural pathology in the CNS that leads to characteristic delirium symptoms. Figure 4.1.1.1 offers examples of areas to distinguish.

Many functions are typically not disturbed in delirium (e.g. primary motor or sensory functions) and certain neuropsychological functions are disproportionately impaired (e.g. attention) suggesting that particular neurobiological underpinnings are relevant to delirium neuropathogenesis. Neuroimaging and neuropsychological studies suggest involvement of prefrontal cortex, thalamus, nondominant posterior parietal and fusiform cortices, and subcortical regions, especially right-sided pathways.[24] Additionally, anterior and posterior portions of cingulate cortex may be involved in Cognition and Higher Level Thinking domains, while subcortical regions including thalamus, hypothalamus, basal forebrain and brainstem may be involved in the Circadian Rhythm domain. Other features may be related indirectly to the underlying brain disturbances that cause domain abnormalities.

Evidence from preclinical studies, causation (e.g. exposure to anticholinergic and dopaminergic deliriogenic agents), direct studies of pathophysiology, and treatment with dopamine blockers point to a relative cholinergic deficit and dopaminergic excess as the principal neurochemical disturbances underpinning delirium although other neurochemical systems (e.g. serotonergic, glutamateric, GABAergic, noradrenergic) are clearly implicated in delirium due to particular aetiologies, perhaps through their interactions with dopaminergic and cholinergic systems. Synaptic, axonal and glial abnormalities are implicated.

Altered oxidative metabolism, stress axis activation, and neuroinflammatory mechanisms may acutely impact neurotransmission. Further, neurostructural derangements previously thought to occur in chronic neurodegenerative disorders can occur acutely and transiently, and may underlie delirium. These include traumatic hyperphosphorylation during anaesthesia or traumatic elevations of Abeta inducing synaptic morphological and functional alterations.

Table 4.1.1.4 Clinical investigations recommended for delirium

1. Mandatory (recommended for all patients)
Full blood count and differential
Urea and electrolytes to include Mg, Ca, Po4
Renal function
Liver function tests to include serum albumin
Urinalysis
Random blood glucose
Electrocardiogram
Chest X-Ray
2. As indicated (according to particular clinical circumstances—list not exhaustive)
Drug screen (therapeutic and illicit)
Blood alcohol concentration
Blood cultures
Cardiac enzymes
Arterial blood gases
Serum Folate / B12
Thyroid function tests
Erythrocyte sedimentation rate
Cerebrospinal fluid examination
Syphilis serology
CT brain
MRI brain
Electroencephalography (with nasopharyngeal leads)
Polysomnography
Prothrombin time
Urinary porphyrins
Screen for heavy metals and insecticides

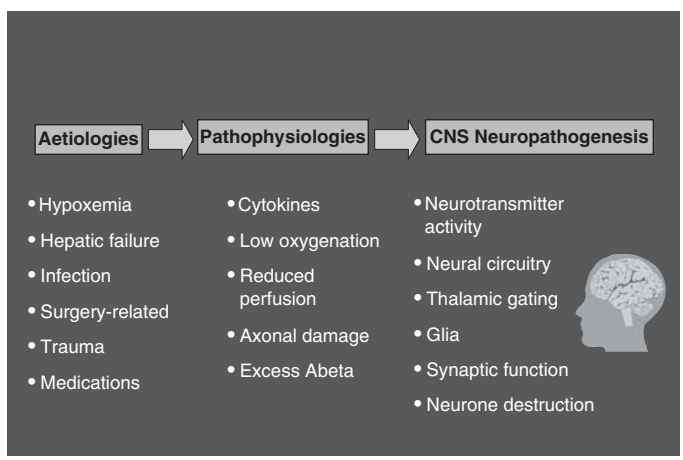


Fig. 4.1.1.1. Examples of different aetiologies for delirium and a variety of pathophysiological mechanisms that can then alter brain function. The neuropathogenesis of delirium involves dysfunction of brain regions and circuitry which may ultimately result in characteristic symptoms of delirium despite a wide variety of aetiologies and pathophysiological insults to the brain.

The broad disturbance of CNS function that occurs in delirium inevitably involves alterations to many cortical and subcortical regions and their neural circuitry, via both direct and indirect (diaschesis) effects.

Management

Delirium is a medical urgency, yet the value of timely intervention in limiting the many deleterious effects of a delirium episode is underappreciated. Optimal management requires the collaborative efforts of primary treating physicians and nursing staff with delirium specialists. Both the underlying causes and the brain disorder need simultaneous assessment and treatment. Careful attention to reorientation strategies (e.g. clearly visible clock/calendar), safety in immediate surroundings and optimal level of environmental stimulation (e.g. natural levels of diurnal lighting) are fundamental to the management of delirium across treatment settings and populations.[26] Relatives / loved ones can report changes in behaviour and mental state ('not themselves') and provide collateral information about baseline cognitive and independent functioning and risk factor exposure.

Delirium prevention using multicomponent interventions to address modifiable risk factors can reduce the frequency and severity of delirium in elderly medical and post-operative populations.[27,28] Common elements include elimination of unnecessary medications, careful attention to hydration and nutritional status, pain relief, correction of sensory deficits, sleep enhancement, early mobilization, and cognitive stimulation. Careful attention to reorientation strategies (e.g. clearly visible clock/calendar), safety in immediate surroiundings and optimal level of environmental stimulation (e.g. natural levels of diurnal lighting) are key elements of delirium care.[26]

Recent studies of pharmacological prophylaxis of delirium in high-risk populations have been encouraging. Controlled studies using haloperidol,[29] olanzapine,[30] donepezil[31] and rivastigmine[32] report significant reductions in delirium incidence, severity and/or duration.

(a) Pharmacological management

Pharmacological management of delirium addresses the brain dysfunction itself while the underlying medical problems are being separately considered. This is akin to acute heart failure where treatment of cardiac function is concurrent with management of the aetiologies for the organ failure. Pharmacological management of delirium is based on empirical knowledge drawn from case reports, open label prospective studies and a small number of randomized trials some of which are comparator studies. Adequately powered double blind randomized placebo-controlled efficacy trials are needed because there is currently no medication with an indication for delirium treatment by any regulatory authority. The inherent fluctuating nature, varying duration, spontaneous recovery rate, and impact of medical treatments upon underlying causes render placebo-controlled studies especially important in evaluating therapeutic interventions. Nevertheless, there are over 20 well-conducted prospective studies of antipsychotic agents used in acute treatment of delirium where more than two-thirds of treated delirious patients experience clinical improvement, typically within a week.[33] A randomized controlled trial of haloperidol vs olanzapine vs non-drug treatment in elderly patients indicated similar response rates in those receiving haloperidol (87.5 per cent) and Olanzapine (82 per cent), which was significantly greater than those in the non-drug treatment group (31 per cent).[34] Treatment response includes improved cognitive and noncognitive symptoms of delirium and does not appear to be closely linked to antipsychotic effect or sedative action. Younger patients with hyperactive presentations and patients without comorbid dementia respond more robustly to antipsychotics, but hypoactive patients also improve and deserve treatment given that delirium comprises many serious symptoms besides challenging motor behaviour.[35,36]

Most pharmacological strategies are based upon the prevailing notion of a relative dopaminergic excess and cholinergic deficiency as the principal neurochemical aberrations underlying delirium, and agents with either procholinergic or antidopaminergic effects are favoured. *Haloperidol* remains the standard agent used to treat delirium and is available in oral, intramuscular, and intravenous preparations. However, intravenous haloperidol does not have any indicated use by a regulatory agency and carries a high risk for QTc prolongation and torsades de pointes tachyarrhythmia that can lead to sudden death. Suggested haloperidol doses are 1–2 mg every four hours as needed but with lower doses (e.g. 0.25–0.5 mg every four hours) in the elderly, very frail, or populations with neuroleptic sensitivity—APA guidelines, 1999.[37] Uncontrolled agitated delirium can be life threatening especially in critically ill intensive care unit patients where substantially higher doses have been reported without major adverse effects. Careful monitoring of ECG (telemetry with intravenous haloperidol), and maintaining normal serum potassium, calcium, and magnesium levels are recommended.

Accumulating evidence supports the use of *atypical antipsychotic agents* in delirium (risperidone, olanzapine, quetiapine, aripirazole, ziprasidone, amisulpiride). Comparison studies suggest similar response rates to haloperidol for both risperidone and olanzapine but with reduced extrapyramidal side effects. In highly agitated patients where sedation is desirable, more sedative agents or combination with lorazepam may be considered. Given the importance of sleep-wake cycle disturbances in delirium, dose scheduling should encourage recovery of normal sleeping patterns.

Benzodiazepines can act as either an alleviating or as a risk factor for delirium depending on the circumstances of exposure. Benzodiazepines can be associated with worsening of mental state[38] and increase delirium risk in ICU[39] and cancer patients.[40] Moreover, therapeutic effects vary from anxiolytic to sedative to hypnotic with ascending doses. Conversely, benzodiazepines are first line treatment for delirium related to sedative and alcohol-withdrawal or seizures. Benzodiazepines can allow for lower neuroleptic doses where intolerance is a problem or where extra sedation is desired. Lorazepam is preferred due to its short acting nature and relatively predictable bioavailability when given intramuscularly. Lower doses are required in the elderly and those with respiratory or hepatic compromise or receiving drugs that undergo extensive hepatic oxidative metabolism. Unwanted effects of benzodiazepines can be rapidly reversed with flumazenil.

The use of *procholinergic agents* such as intravenous physostigmine has long been advocated for delirium due to toxicity with anticholinergic drugs, but routine delirium use is limited by gastrointestinal side effects, cardiac arrythmia and seizures. To date there has been limited study of newer procholinergics in part because their long half-lives preclude reaching steady state for use

in acute conditions, in contrast to parenteral physostigmine's fast onset of action.

Anecdotal evidence also exists for the use of mianserin, trazadone, melatonin, psychostimulants, and even ECT in the treatment of delirium but these strategies are not well-studied or applied in routine clinical practice.

(b) Risks, benefits, and dosing

Adequate drug treatment of delirium is limited by concerns over potential toxic effects in highly morbid, frail elderly whose delirium may actually herald bad outcome from medical illness. In assessing the risk-benefit ratio of medication use one must consider the risk of nontreatment of delirium given its grave consequences. Dosing needs to take into account structural degenerative changes, reduced neurochemical flexibility, less robust counter-regulatory homeostatic mechanisms, reduced renal and hepatic function, lower water to fat ratio with reduced muscle mass, and reduced plasma esterase activity. In hypoactive or mechanically ventilated patients careful dose titration can be assisted by regular monitoring of sedation. With sedation, a beneficial effect of catching up from sleep deprivation also needs to be considered. Adverse events such as Parkinsonism or akathisia can be misattributed to agitation of delirium, though are uncommon in treatment studies perhaps reflecting a protective effect of the hypocholinergic state that frequently underpins delirium. The potential for cardiac arrythmias can be reduced with ECG monitoring in high dose or intravenous haloperidol use or where patients have a cardiac history or baseline ECG shows QTc interval >450 mSec.

(c) Patients with concomitant dementia

Both pharmacological[35] and non-pharmacological strategies[28] appear less effective in patients with concomitant dementia perhaps reflecting the inherently poor outcome of elderly demented populations with high physical comorbidity. There are concerns regarding the small but increased risk of cerebrovascular events in demented patients chronically receiving neuroleptics, but the relative risks of short-term use in delirium must be proportionalized against potential benefits. Lower doses should be used with careful monitoring for adverse effects and more prolonged use should not occur in the absence of clear benefits.

(d) Management after recovery

Depending on the degree of memory imprinting during an episode, patients experience significant emotional distress after recovery[41] and many continue to have distressing recollections of delirium six months later, especially where the episode has involved psychotic symptoms. For families, disturbing recollections of final contacts with loved ones can be enduringly distressing and associated with complicated bereavement reactions. Some patients minimize their experiences fearing that they represent emerging senility or loss of competence. Explicit recognition allows clarification of the causes of delirium and reduction of future risk factors.

Conclusion

Delirium is a complex neuropsychiatric syndrome that occurs commonly across all age groups and healthcare settings. Significant adverse outcomes of delirium are increasingly recognized and can be reduced by a more consistent approach to detection that emphasizes disturbances of attention. Optimal management requires the collaborative efforts of carers and healthcare staff and judicious use of pharmacological and nonpharmacological strategies that concurrently manage underlying physical causes and the delirium itself. Greater clarity is needed regarding the prognostic relationship to dementia, phenomenology of prodromal, subsyndromal and syndromal delirium, and how risk factors, vulnerabilities, and treatment may vary across populations and treatment settings.

Further information

The *European Delirium Association* advocates for better research and aims to foster research activity across all disciplines. Annual meeting details, educational materials and a discussion forum are available at www.europeandeliriumassociation.com

A US-based delirium organization aims to foster delirium care and research in critically ill patients and includes various teaching resources including protocols for assessment and treatment at www.icudelirium.org

The American Psychiatric Association website includes detailed delirium treatment guidelines, a quick reference guide as well as a patient and family guide (see www.psych.org/psych_pract/treatg/quick_ref_guide/DeliriumQRG_4-15-05.pdf)

References

1. Trzepacz, P.T., Mulsant, B.H., Dew, M.A., *et al.* (1998). Is delirium different when it occurs in dementia? A study using the Delirium Rating Scale. *J Neuropsychiatry Clin Neurosci*, **10**, 199–204.
2. Meagher, D.J., Moran, M., Raju, B., *et al.* (2007): Phenomenology of 100 consecutive adult cases of delirium. *British Journal of Psychiatry*, **190**, 135–41.
3. McNicoll, L., Pisani, M.A., Zhang, Y. *et al.* (2003). Delirium in the intensive care unit: occurrence and clinical course in older patients. *J Am Geriatr Soc*, **51**, 591–98.
4. Hart, R.P., Levenson. J.L., Sessler, C.N., *et al.* (1996). Validation of a cognitive test for delirium in medical ICU patients. *Psychosomatics*, **37**, 533–46.
5. Turkel, S.B., Trzepacz, P.T., Tavare, C.J. (2006). Comparison of delirium in adults and children. *Psychosomatics*, **47**, 320–24
6. Kakuma, R., du Fort, G.G., Arsenault, L., *et al.* (2003). Delirium in older emergency department patients discharged home: effect on survival. *J Am Geriatr Soc*, **51**, 443–50.
7. Rudberg, M.A., Pompei, P., Foreman, M.D., *et al.* (1887). The natural history of delirium in older hospitalized patients: a syndrome of heterogeneity. *Age Ageing*, **26**, 169–74.
8. Sylvestre, M.P., McCusker, J., Cole, M., *et al.* (2006). Classification of patterns of delirium severity scores over time in an elderly population. *Int Psychogeriatr*, **18**, 667–80.
9. Inouye, S.K., van Dyke, C.H., Alessi, C.A., *et al.* (1990). Clarifying confusion: the Confusion Assessment Method. *Ann Intern Med*, **113,** 941–48.
10. Ely EW, Gordan S, Francis J, *et al.* (2001). Evaluation of delirium in critically ill patients: validation of the Confusion Assessment Method for the intensive care unit (CAM-ICU). *Crit Care Med*, **29**, 1370–79.
11. Trzepacz, P.T., Mittal, D., Torres, R., *et al.* (2001). Validation of the Delirium Rating Scale–Revised-98: comparison to the Delirium Rating Scale and Cognitive Test for Delirium. *J Neuropsychiatry Clin Neurosci*, **13**, 229–42.
12. Breitbart, W., Rosenfeld, B., Roth, A., *et al.* (1997).The Memorial Delirium Assessment Scale. *J Pain Symptom Manage*, **13**, 128–37.
13. Siddiqi, N., House, A.O., Holmes, J.D., *et al.* (2006). Occurrence and outcome of delirium in medical in-patients: a systematic literature review. *Age Ageing*, **35**, 350–64.

14. Fick, D.M., Kolanowski, A.M., Waller, J.L., *et al.* (2005). Delirium superimposed on dementia in a community-dwelling managed care population: a three-year retrospective study of occurrence, costs and utilization. *J Gerontol Med Sci*, **60A**, 748–53.
15. Gruber-Baldini, A.L., Zimmerman, S., Morrison, R.S., *et al.* (2003). Cognitive impairment in hip fracture patients: timing of detection and longitudinal follow-up. *J Am Geriatr Soc*, **51**, 1227–36.
16. Lundstrom, M., Edlund, A., Bucht, G., *et al.* (2003). Dementia after delirium in patients with femoral neck fractures. *J Am Geriatr Soc*, **51**, 1002–6.
17. Ely, E.W., Shintani, A., Truman, B., *et al.* (2004). Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit. *JAMA*, **291**, 1753–62.
18. Pitkälä, K.H., Laurila, J.V., Strandberg, T.E., *et al.* (2005). Prognostic signifi cance of delirium in frail older people. *Dement Geriatr Cogn Disord*, **19**, 158–63.
19. Cole, M., McCusker, J., Dendukuri, N., *et al.* (2003). The prognostic signifi cance of subsyndromal delirium in elderly medical inpatients. *J Am Geriatr Soc*, **51**, 754–60.
20. Inouye, S.K. and Charpentier, P.A., *et al.* (1996): Precipitating factors for delirium in hospitalized elderly patients: predictive model and interrelationships with baseline vulnerability. *JAMA,* **275**, 852–57.
21. Kalisvaart, K.J., Vreeswijk, de Jonghe, J., *et al.* (2005b) Risk factors and prediction of post-operative delirium in elderly hip surgery patients. Implementation and validation of the Inouye risk factor model. *Int Psychogeriatrics*, **17** (S2), 261.
22. Ely, E.W., Girard, T.D., Shintani, A.K., *et al.* (2007). Apolipoprotein E4 polymorphism as a genetic predisposition to delirium in critically ill patients. *Crit Care Med*, **35**, 112–7.
23. Yokota, H., Ogawa, S., Kurokawa, A., *et al.* (2003). Regional cerebral blood fl ow in delirious patients. *Psychiatry Clin Neurosci,* **57**, 337–9.
24. Trzepacz, P.T. (1994). Neuropathogenesis of delirium: a need to focus our research. *Psychosomatics*, **35**, 374–91.
25. Trzepacz, P.T., Meagher, D.J. (2007). Delirium. In *American Psychiatric Publishing Textbook of Neuropsychiatry* (5th edn). (eds S. Yudofsky and R. Hales). American Psychiatric Publishing, Inc, Washington, DC.
26. Meagher, D. (2001). Delirium: optimising management. *British Medical Journal,* **7279**, 144–9.
27. Inouye, S.K., Bogardus, S.T., Charpentier, P.A., *et al.* (1999). A multicomponent intervention to prevent delirium in hospitalized older patients. *N Engl J Med*, **340**, 669–76.
28. Marcantonio, E.R., Flacker, J.M., Wright, R.J., *et al.* (2001). Reducing delirium after hip fracture: a randomized trial. *J Am Geriatr Soc*, **49**, 516–22.
29. Kalisvaart, K.J., de Jonghe, J.F.M., Bogaards, M.J., *et al.* (2005). Haloperidol prophylaxis for elderly hip surgery patients at risk for delirium: a randomized, placebo-controlled study. *J Am Geriatr Soc*, **53**, 1658–66.
30. Larsen, K., Kelly, S., Stern, T., *et al.* (2007). A double-blind, randomized, placebocontrolled study of perioperative administration of olanzapine to prevent postoperative delirium in joint replacement patients. Academy of Psychosomatic medicine 54th Annual meeting proceedings.
31. Sampson, E.L., Raven, P.R., Ndhlovu, P.N., *et al.* (2006). A randomized, double-blind, placebo-controlled trial of donepezil hydrochloride (Aricept) for reducing the incidence of postoperative delirium after elective total hip replacement. *Int J Geriatr Psychiatry,* **22**, 343–9.
32. Moretti, R., Torre, P., Antonello, R.M., *et al.* (2004). Cholinesterase inhibition as a possible therapy for delirium in vascular dementia: a controlled, open 24-month study of 246 patients. *Am J Alzheimers Dis Other Demen,* **19**, 333–9.
33. Meagher, D. and Leonard, M. (2008). The active management of delirium: improving detection and treatment. *Advances in Psychiatric Treatment*, **14**, 292–301.
34. Hua, H., Wei, D., Hui, Y., *et al.* (2006). Olanzapine and haloperidol for senile delirium: a randomised controlled observation. *Chinese Journal of Clinical rehabilitation*, **10**, 188–90.
35. Breitbart, W., Tremblay, A., Gibson, C., *et al* (2002). An open trial of olanzapine for the treatment of delirium in hospitalized cancer patients. *Psychosomatics*, **43**, 175–82.
36. Platt, M.M., Breitbart, W., Smith, M., *et al.* (1994). Efficacy of neuroleptics for hypoactive delirium. *J Neuropsychiatry Clin Neurosci,* **6**, 66–67.
37. American Psychiatric Association (1999). Practice guidelines for the treatment of patients with delirium. *Am J Psychiatry*, **156**(suppl), 1–20.
38. Breitbart, W., Marotta, R., Platt. M.M., *et al.* (1996). A double-blind trial of haloperidol, chlorpromazine, and lorazepam in the treatment of delirium in hospitalized AIDS patients. *Am J Psychiatry*, **153**, 231–37.
39. Pandharipande, P., Shintani, A., Peterson, J., *et al* (2006). Lorazepam is an independent risk factor for transitioning to delirium in intensive care unit patients. *Anesthesiology*, **104**, 21–6.
40. Gaudreau, J.D., Gagnon, P., Harel, F., *et al.* (2005). Psychoactive medications and risk of delirium in hospitalized cancer patients. *J Clin Oncol*, **23**, 6712–8.
41. Breitbart, W., Gibson, C. and Tremblay, A., *et al.* (2002b). The delirium experience: delirium recall and delirium-related distress in hospitalized patients with cancer, their spouses/caregivers, and their nurses. *Psychosomatics*, **43**, 183–94.

4.1.2 Dementia: Alzheimer's disease

Simon Lovestone

Introduction

Alzheimer's disease (**AD**) and other dementias incur huge costs to society, to the families of those affected, and to the individuals themselves. Costs to society include both direct costs to health and social services and indirect economic costs in terms of lost productivity, as carers are taken out of the workplace, and the economic costs to those families caring for or funding the care of their relative. Increasingly, as treatments become available, these costs are targets for change and are part of the cost–benefit analysis of new compounds, especially the largest single direct cost, that of the provision of nursing and other forms of continuing care. Apart from the financial cost to families there is the emotional impact resulting in distress and psychiatric morbidity.

As the population ages, these costs pose substantial social and economic problems. Although lifespan itself has remained static, the numbers of elderly in both developed and developing societies is increasing rapidly. In the developed world the sharpest projected growth is in the very elderly cohort—precisely the one that is at most risk of AD. Within the developing world, the total number of elderly people is projected to rise substantially, reflecting to a large part better child health and nutrition. For countries in South America and Asia, with large and growing populations, the costs involved in caring for people with dementia in the future will become an increasing burden on health and social services budgets. In the absence of such services families will inevitably shoulder the main part of providing care, although the very process of development is associated with increasing urbanization and, to some

degree, a diminution of the security provided by extended family structures.

From discovery towards understanding

In the early part of the twentieth century, Alois Alzheimer described his eponymous disorder in a middle-aged woman who suffered not only cognitive deterioration and functional decline but psychotic experiences, including delusions and auditory hallucinations. Neuropathology included gross atrophy and plaques and tangles on microscopy. Although all the important features of AD were described at this stage, two important developments came much later. First, in the 1960s with the studies of Roth and colleagues in Newcastle[1] and others elsewhere, it was appreciated that much dementia in the elderly has an identical neuropathological appearance to that of AD in younger people. The other development was the rediscovery that AD has a rich phenomenology. The non-cognitive symptomatology of AD is integral to the clinical manifestation of this disease, and is a major cause of carer burden and medical intervention. This second phase of research—the recognition that both the neuropathology and clinical phenomenology described by Alzheimer occur in what had previously been though of as senile dementia or, worse, just ageing, was accompanied by a growing understanding of the neurotransmitter deficits in AD. The cholinergic hypothesis provided the first glimpse of possible interventions, and remains the most important finding from this period of AD investigations. The third phase of AD research encompasses the use of molecular approaches to understanding pathogenesis. The techniques of molecular biology have been applied to understanding the formation of plaques and tangles, to a growing understanding of the genetic aetiology of much of AD, and, through the use of transgenic approaches, to developing animal and cellular models of pathogenesis.

Just as research can broadly be seen to have three phases—discovery, neuropathology, and molecular aspects—so too does the clinical response to AD. For many years cognitive impairment in the elderly was perceived as senility. As a process thought to be an inevitable consequence of ageing it was difficulty to establish medical-care models. Hence the needs of the elderly with AD were not seen as requiring specialist intervention, carers needs were not realized, and public appreciation of the impact of dementia on the elderly themselves or on the family was negligible. The change in perception of AD from 'just ageing' to a disease was accompanied, and to some degree led, by the development of 'old-age psychiatry' as a specialism on the one hand and by the rapid growth of the AD societies on the other. During this second phase of AD treatment, the goals have been to ensure that the care needs of patients are met, that families' concerns are addressed, and that behavioural disturbance is minimized. The third phase of AD treatment began with the arrival of specifically designed interventions. Compounds have been introduced that were designed to ameliorate some of the deficits incurred by the disease process, and other approaches are being developed to treat those disease processes themselves.

Clinical features

Cognitive impairment

Dementia is acquired cognitive decline in multiple areas resulting in functional impairment; AD is one cause of dementia and the core clinical symptom of AD is cognitive loss. However, as noted above, AD is clinically heterogeneous and includes diverse non-cognitive symptoms. Cognitive decline is manifested as **a**mnesia, **a**phasia, **a**gnosia, and **a**praxia (the 4As).

(a) Amnesia

Memory loss in AD is early and inevitable. Characteristically, recent memories are lost before remote memories. However, there is considerable individual variation, with some patients able to recall specific and detailed events of childhood and others apparently having few distant memories accessible. With disease progression, even remote and emotionally charged memories are lost. The discrepancy between recent and remote memory loss suggests that the primary problem is of acquisition or retrieval of memory rather than a destruction of memory, and this is confirmed in early AD,[2] although as the disease progresses it is likely that all memory processes are impaired. Retrieval of remote memory is assumed to be preserved for longer because of rehearsal over life.

(b) Aphasia

Language problems are found in many patients at presentation, although the language deficits in AD are not as severe as those of the fronto-temporal degenerations[3] and may only be apparent on detailed examination. Word-finding difficulties (nominal dysphasia) are the earliest phenomena observed and are accompanied by circumlocutions and other responses, for example repetitions and alternative wordings. As the disorder progresses, syntax is affected and speech becomes increasingly paraphasic. Although harder to assess, receptive aphasia, or comprehension of speech, is almost certainly affected. In the final stages of the disorder, speech is grossly deteriorated with decreased fluency, preservation, echolalia, and abnormal non-speech utterances.

(c) Agnosia

Patients with AD may have difficulty in recognizing as well as naming objects. This can have implications for care needs and safety if the unrecognized objects are important for daily functioning. One particular agnosia encountered in AD is the loss of recognition of one's own face (autoprosopagnosia). This distressing symptom is the underlying cause of perhaps the only clinical sign in AD—the mirror sign. Patients exhibiting this will interpret the face in the mirror as some other individual and respond by talking to it or by apparent fearfulness. Autoprosopagnosia can present as an apparent hallucinatory experience, until it is realized that the 'hallucination' is fixed in both content and space, occurring only when self-reflection can be seen.

(d) Apraxia

Difficulties with complex tasks that are not due to motor impairment become apparent in the moderate stages of AD. Typically, difficulties with dressing or tasks in the kitchen are noticed first, but these are inevitably preceded by loss of ability for more difficult tasks. Strategies to avoid such tasks are often acquired as the disease progresses, and it is only when these fail that the dyspraxia becomes apparent.

Other cognitive impairment

There appear to be no cognitive functions that are truly preserved in AD. Visuospatial difficulties commonly occur in the middle stages of the disorder and may result in topographical disorientation,

wandering, and becoming lost. Difficulties with calculation, attention, and cognitive planning all occur.

Functional impairment

Although the cognitive decline in AD is the core symptom, it is the functional deterioration that has the most impact on the person themselves and it is the functional loss that necessitates most of the care needs of patients with AD, including nursing-home residency.[4] Increasingly, abilities to function in ordinary life (activities of daily living (**ADLs**)) are lost, starting with the most subtle and easily avoided and progressing to the most basic and essential. In general, functional abilities decline alongside cognitive abilities. However, the precise correlation between these functions is not perfect, suggesting that factors other than disease severity account for part of the variance between patients.[5] Functional abilities are related to gender; for example, cooking abilities are rehearsed more frequently in women, and home-improvement skills in men. However, the overall pattern shows some similarities between groups of patients with similar disease severity. This is exploited in the Functional Assessment Staging (**FAST**) Scale;[6] in the original form, this is a seven-point scale of functional impairment, with stage 1 as no impairment and stage 7 as severe AD. A sequential decline is mapped by descriptions of the abilities that are lost: stage 2, difficulties with language and finding objects; stage 4, difficulties with finances; stage 6, incontinence and inability to dress or wash oneself.

ADLs are divided into those that relate to self-care and those that concern instrumental activities. Instrumental ADLs, those related to the use of objects or the outside world, are lost first and can be subtle.[7] A change in the ability to use the telephone properly or to handle finances accurately may not be apparent. Self-care ADLs include dressing and personal hygiene and are also lost gradually; for example, untidiness in clothing progresses to difficulties in dressing. Personal hygiene becomes poor as dentures are not cleaned and baths taken less often, before finally assistance is required with all self-care tasks.

Neuropsychiatric symptoms

(a) Mood

The relationship between AD and depression is complex. Depression is a risk factor for AD, depression can be confused with dementia (pseudodementia), depression occurs as part of dementia, and cognitive impairments are found in depression. Depression occurring as a symptom of dementia will be considered here. Assessing the mood of a person with dementia is difficult for obvious reasons. However, psychomotor retardation, apathy, crying, poor appetite, disturbed sleep, and expressions of unhappiness all occur frequently. The rates of depression found in cohorts of patients with AD vary widely, reflecting changes in prevalence at different levels of severity and difficulties in the classification of symptoms suggestive of depression in those with cognitive loss. A major depressive episode is found in approximately 10 per cent of patients, minor depressive episode in 25 per cent, some features of depression in 50 per cent, and an assessment of depression by a carer in up to 85 per cent.[8–12] It is commonly believed that depression is more common in the early than in the later stages of AD, although this may reflect the difficulties of assessing depression in the more severely affected and least communicative patients. Indeed, severely affected patients in nursing homes may be particularly prone to depression.[13] Elation, disinhibition, and hypomania all occur in AD but are relatively infrequent, elevated mood being found in only 3.5 per cent of patients by Burns *et al.*[10]

The underlying cause of mood change in AD is not known. However, loss of serotonergic and noradrenergic markers accompanies cholinergic loss; some studies have found a greater loss of these markers at post-mortem in AD patients with depression than in non-depressed patients.[14–16]

(b) Psychosis

Psychotic symptoms occur in many patients, although, as with depression, there is an inherent difficulty in determining the presence of delusions or hallucinatory experiences in the moderately to severely demented. In community-based surveys, between 20 and 70 per cent are reported to suffer from some form of psychotic symptom with delusions being more common than hallucinations.[11,17,18] Delusions are frequently paranoid and the most common delusion is one of theft. In the context of the confusion and amnesia of dementia, it is easy to appreciate how the experience of mislaying an object becomes translated into conviction of a theft. Other patients become convinced that someone, often a family member, is trying to harm them.

Hallucinations are only somewhat less frequent than delusions—the median of one series of studies being 28 per cent.[19] Visual hallucinations are reported more commonly than auditory ones, and other modalities are rare. Most studies of the non-cognitive symptomatology of AD precede the wide recognition and accepted criteria of dementia with Lewy bodies, one of the cardinal symptoms of which is visual hallucinations. It is probable that a large number of those AD patients experiencing visual hallucinations reported in the studies would now be classified as having dementia with Lewy bodies.

Phenomena falling short of delusions or hallucinations, such as persecutory ideas or intrusive illusionary experiences, are common in AD as are misidentification syndromes. Capgras' syndrome may occur, but frequently the symptom is less fully evolved with the patient mistaking one person for another. Failure to recognize one's own face may be due to visuospatial difficulties or to a true misidentification syndrome—distinguishing between the two is difficult.

Various factors have been associated with psychosis in AD, but few have been substantiated in multiple studies. Burns *et al.*[20] found that more men than women suffered delusions of theft, although others find that psychosis occurs more often or earlier in women. An association with polymorphic variation in serotonin receptors has been reported.[21–23] The relationship between psychosis and dementia severity is not as clear cut as that between functional ability and dementia severity. Psychosis can occur at any stage of the disease process, although most studies find the maximal rate of psychosis in those with at least moderate dementia.

Although the biological basis of psychosis within AD is not fully understood,[14] it is clear that psychosis symptoms impact upon carers causing increased distress,[24,25] and that underlying psychosis accounts for much of the behavioural disturbance and aggression encountered in AD.[26]

(c) Personality

Changes in personality are an almost inevitable concomitant of AD. Indeed, it is difficult to envisage how profound cognitive impairment resulting in the loss of recognition of loved ones, and

an understanding of and ability to react with the outside world, could not result in a change in personality. Family members have described the loss of personality as a 'living bereavement'—the body remains, but the person once known has gone. Personality change is most frequently one of loss of awareness and normal responsiveness to the environment. Individuals may become more anxious or fearful, there is a flattening of affect, and a withdrawal from challenging situations. Catastrophic reactions are short-lived emotional reactions that occur when the patient is confronted, and cannot avoid, such a challenging situation. Less commonly, personality changes may be of disinhibition with inappropriate sexual behaviours or inappropriate affect. Aggressiveness is, as noted above, often accompanied by psychosis, but it may be part of a more general personality change.

(d) Other behavioural manifestations

Behavioural complications in AD have become a target of therapy. However, the term encompasses a wide range of behaviours, some of which include neuropsychiatric syndromes, some caused by neuropsychiatric syndromes, and some of which have little apparent relationship to mood or to thought content. Behavioural complication is itself a largely subjective term that relies to a great extent on informer evaluation: but behaviour may be a complication in one context, although not in another.

Behaviours exhibited in AD include wandering, changes in eating habit, altered sleep or circadian rhythms, and incontinence. These behaviours are closely linked to disease severity and occur to some extent in the majority of patients with AD. Wandering may be a manifestation of topographical confusion, a need for the toilet, or it may reflect hunger, boredom, or anxiety. Sleep is frequently disturbed, with many patients exhibiting altered sleep–wake cycles and others experiencing increased confusion towards evening ('sundowning'). A central defect in the regulation of circadian rhythms underlying these phenomena is postulated.[27] Excessive or inappropriate vocalizations (grunting and screaming) occur in the late stages.

Classification

Alzheimer's disease is classified, as with all other disorders, by DSM-IV and by ICD-10. In addition, it also has a specialized classification system resulting from the National Institute of Neurological and Communicative Disorders and Stroke–AD and Related Disorders Association (**NINCDS–ADRDA**).[28] This clinical diagnostic system is internationally accepted and widely observed. There are also classification systems for neuropathological diagnosis, most notably the Consortium to Establish a Registry for AD (**CERAD**) criteria.[29]

DSM-IV stipulates that a dementia syndrome is characterized by deterioration in multiple cognitive deficits, including amnesia, resulting in functional impairment. A gradual onset and decline in the absence of other conditions sufficient to cause dementia indicates AD. ICD-10 shares with DSM-IV the definition of a dementia syndrome and notes that an insidious onset and slow decline in the absence of other disorders sufficient to cause dementia indicates AD. The NINCDS–ADRDA criteria defines possible, probable, and definite categories; the latter being restricted to neuropathological confirmation of a clinical diagnosis.[28] It is important to note that both clinical and neuropathological data are required—no single neuropathological lesion is pathognomonic of AD, and it is still uncertain how often or to what extent the neuropathological lesions of AD also occur in normal ageing.[30] Probable AD, according to NINCDS–ADRDA, requires a dementia with progressive decline in memory and other cognitive areas, cognitive impairment established by formal testing, no disturbance of consciousness, and absence of other disorders sufficient to cause dementia. Supporting features include decline in function, change in behaviour, positive family history, and decline in specific cognitive areas including aphasia, apraxia, and agnosia. Non-specific change on electroencephalography (**EEG**) and progressive changes on CT are supporting, but not necessary, features. Possible AD should be diagnosed if there are variations in the clinical presentation, another disorder sufficient to cause a dementia (even if it is not thought to do so in this case), or a restricted cognitive decline.

A number of studies have attempted to determine the accuracy of diagnostic criteria against post-mortem diagnosis. One of the difficulties in these studies is that because AD is the most common dementia (by some way), such studies are very likely to find a high-positive predicative value. Kukull *et al.*[31] found the specificity of DSM-III to be higher than NINCDS–ADRDA (0.8 versus 0.65), but NINCDS–ADRDA had a higher sensitivity (0.92 versus 0.76), Mok *et al.* find broadly similar findings for both primary care physicians and for neurologists;[32] some others find an even lower specificity.[33]

Diagnosis

Alzheimer's disease is the most common of the dementias, occurring in some 60 to 70 per cent of cases. However, this oft-stated figure must be treated with some caution for two reasons. First, cases that come to post-mortem represent a biased sample and in the community a large proportion (up to a third) of non-demented individuals have pathological signs of AD such as neuritic plaques.[30] Second, even at post-mortem the distinction between different dementias is not clear cut—many AD brains show the presence of Lewy bodies and others have considerable evidence of vascular damage. The proportion of mixed pathologies is actually rather high, between 15 and 30 per cent of all dementias. Thirdly, even the gold-standard of neuropathological diagnosis is not infallible. Neuropathologists show a very high degree of inter-rater agreement on the diagnosis of probable AD and Dementia with Lewy Bodies (DLB) but a rather lower rate of agreement when there is vascular damage, when the diagnosis is of fronto-temporal dementia (FTD) and on the more equivocal cases of AD.[34]

History

Making a clinical diagnosis of AD is a positive process and not one of exclusion. The most valuable diagnostic assessment is a careful informant history, paying attention to the pattern and timing of onset and progression. In the research context, a family history interview conducted by telephone provides a degree of accuracy compatible with a full clinical assessment.[35] Detailed semi-structured family informant diagnostic schedules are available, such as Cambridge Mental Disorders of the Elderly Examination (CAMDEX).[36] A history should be taken for the presence of risk factors for AD (e.g. a positive family history) and vascular and other risk factors (e.g. hypertension and head injury). Taking a family history for late-onset disorders such as AD requires special

attention. Because of attrition due to other illness, many elderly people have had too few relatives reach the age of onset of dementia to make a pedigree analysis informative. The ages at death of all relatives should be established, together with cause of death and the presence or absence of dementia or memory problems in late life. The term 'sporadic' dementia should be avoided, and is misleading when applied to an individual with a dementia where one parent died young and where no sibling reached the age of 65 to 70 years. The history should also screen for the presence of other illnesses sufficient to cause a dementia and for systemic health in general. The presence of any significant physical illness, from chronic pain to delirium, may significantly alter cognitive abilities in the elderly, and especially so in those with AD.

A careful history should also establish the presence of any behavioural disturbance that has occurred. The relationship of aggression, wandering, agitation, or other behaviours to care tasks and other recent changes in the provision of the care package should be established. As the mainstay of the management of behavioural disturbance in all dementias is behavioural, establishing the antecedents to behaviour is an absolute prerequisite to effective management.

Examination

In addition to an examination of the mental state to establish the presence of disorders of mood and thought content, the examination will establish the specific pattern of cognitive impairment and the degree of impairment. Screening tests used to establish the presence of cognitive impairments include the Mini Mental State Examination;[37] this is a 30-point scale routinely used in all clinical trials of drugs for the treatment of AD, which is also a useful proxy measure for severity. It should be accompanied by other cognitive testing, including supplementary examination for aphasia and apraxias. Other cognitive and physical examinations will be necessary where the differential diagnosis is between a lobar dementia (e.g. FTD) or a subcortical dementia (e.g. that accompanying Huntington's disease).

In addition to the cognitive examination, a physical examination should be conducted in all patients with AD, although this might not be most effectively and conveniently performed at the initial assessment. Physical illness, including chronic pain, infection, cardiac insufficiency, or anaemia are all common in the elderly and can both complicate the diagnosis of AD and increase confusion in those known to have AD.

Assessment of function

Clinical assessment of function can be performed by informant history and by direct observation. Key to an assessment of function is a careful informant history seeking to establish where there has been a functional decline and remembering that instrumental ADLs are lost before basic activities. Instrumental ADLs are highly individual and require careful interviewing to assess—one patient may have a modest decline in their ability to use information technology whilst another may have trouble using all the appropriate settings on the central heating. The occupational therapist fulfils an invaluable role in establishing the detailed functional ability of those with AD, in addition to implementing changes in the home designed to maximize function. The FAST Scale[6] is based on the premise that the pattern of decline in function is relatively uniform in AD, and hence establishes a staging of severity on function rather than cognition. As in most instances functional severity is of more relevance for the provision of services, there is much to recommend such an approach. Scales used in research that can also be usefully employed in the clinic include the Bristol ADL Scale[38] and the Disability Assessment for Dementia.[39]

Global assessment

Driven largely by the United States Food and Drugs Administration, global assessment has become part of the assessment of all patients with AD in clinical trials and is finding its way into clinical practice. The underlying premise is that an assessment by a clinician, often supplemented by an informant history, provides information on severity that neither a cognitive assessment nor a functional assessment alone can provide. Two scales, the Clinicians Interview of Change,[40] and the Clinical Dementia Rating,[41] have become widely used in this context.

Investigations

At the initial assessment, patients with dementia should be investigated for other disorders that could complicate, exacerbate, or be confused with AD. A dementia screen might include routine biochemistry, thyroid function tests, vitamin B_{12} and folate estimations, and a full blood count; many would also include syphilis serology, although the frequency of abnormal findings is low. Neuroimaging is recommended in all cases by expert guidelines,[42] and serves two purposes—to exclude reversible causes of dementia and to contribute towards a definitive diagnosis. Thus using structural imaging with CT or, increasingly, with magnetic resonance imaging (MRI), the hippocampal atrophy of AD, the frontal predominant atrophy of FTD and the lesions of vascular dementia can be identified, adding to the specificity of diagnosis. In practice, neuroimaging is often omitted particularly when patients present with a typical history of a slowly progressive dementia of many years standing. Functional scanning (single-photon emission CT (SPECT) in particular) can be useful where regional dementias are suspected, and MRI should be the imaging modality of first choice where vascular dementia is a possibility. An EEG is nearly always non-specifically abnormal even in the early stages of AD, in contrast with fronto-temporal degenerations where an EEG remains unaffected at a broadly equivalent severity. This can help to distinguish the conditions, particularly where there is neuroimaging evidence of regional insufficiency.

Aetiology and molecular neurobiology

Alzheimer's disease is the most common dementia, affecting more than 20 per cent of the population over the age of 85 years. Epidemiological evidence has suggested risk factors and putative protective factors, but the greatest advances in understanding its pathogenesis have come from the combination of molecular and epidemiological approaches.

Neuropathology

At post-mortem, the brain in AD is lighter than aged-unaffected controls with more prominent sulci and a larger ventricular volume. Microscopic examination reveals the most prominent lesions described by Alzheimer—the extracellular plaque and intracellular neurofibrillary tangle. No consensus has developed regarding which of these lesions is responsible for the cognitive impairment

of AD. Plaques, or more precisely amyloid load, might correlate with the degree of cognitive impairment,[43] although a significant amyloid deposition is also found in normal, unimpaired, aged individuals.[30] However, there is a high degree of correlation between dementia severity and neurofibrillary tangle formation,[44] although it is possible that some of the features of AD are more stable than others; for example, extracellular neurofibrillary tangles persist after the neurone has died, whereas extracellular Lewy bodies are not found.

The plaque consists of an amyloid core surrounded by dystrophic neurites, which are themselves filled with highly phosphorylated tau protein. Studies of Down syndrome brains have suggested a temporal course to plaque formation. First, peptides derived from the amyloid precursor protein (**APP**) are deposited in a diffuse plaque.[45] Over time this becomes organized as the amyloid peptides become fibrillar and form the amyloid deposit, neuritic change then occurs, and the plaque becomes fully mature.

Neurofibrillary tangles are composed of paired helical filaments, structures which are also found in the dystrophic neurites around mature plaques, and together with straight filaments, in neuropil threads. These filaments are themselves composed of the microtubule-associated protein, tau, which is present in a stably and highly phosphorylated state.[46,47] Tau is a neuronal-specific protein, found predominantly in the axon that functions to stabilize microtubules, a property that is regulated by phosphorylation. Phosphorylated tau is less effective in promoting tubulin polymerization into microtubules; normal adult brain a proportion of tau is highly phosphorylated, but this proportion is considerably greater in AD. Tau deposits are a feature of other disorders, such as progressive supranuclear palsy and some fronto-temporal degenerations[48] and together these tau-related disorders have been grouped together as the 'tauopathies'. Mutations have been found in fronto-temporal degenerations with parkinsonism (FTDP-17), and in other tauopathies thereby emphasizing the importance of this molecule to neurodegeneration.[49]

Braak and Braak[50] studied large numbers of brains from individuals who died at various ages and at different stages of dementia severity, which has resulted in the wide acceptance of the neuropathological staging of AD. The very earliest stages, before the clinical manifestation of dementia, are characterized by the appearance of highly phosphorylated tau in the hippocampus. In later stages, neurofibrillary tangles appear in the same brain regions and then become more widely distributed.

(a) The cholinergic hypothesis

The pathological changes in AD are localized both structurally and functionally. Plaques and tangles first occur in the hippocampus before spreading to involve other regions. Some areas of the brain are relatively preserved—the occipital lobe is affected relatively late and the cerebellum appears to be spared from neuritic change (neurofibrillary tangles and the fully matured plaques, although diffuse amyloid deposits do occur). Functional localization was demonstrated by evidence of the relatively greater and earlier loss of cholinergic neurotransmission. At post-mortem there is evidence of significantly greater neuronal loss in the cholinergic nucleus basalis of Meynert and loss of cholinergic markers.[51] These observations led to the cholinergic hypothesis, which stated that the cognitive impairment of AD was due to a disorder predominantly affecting cholinergic neurones. It was this hypothesis that led to the development of pharmacological strategies to rectify cholinergic loss and the introduction of the first compounds specifically designed for and efficacious in AD. However, the cholinergic hypothesis was something of a simplification as other neurotransmitter systems (e.g. serotonergic and noradrenergic) are also affected in AD.

(b) The amyloid cascade hypothesis

In 1984, the protein deposited in blood vessels (congophilic angiopathy) in AD was shown to be a 4-kDa peptide known as β-amyloid. This peptide, which is identical to the amyloid in plaques, is derived from a larger peptide, APP, the gene for which is coded on chromosome 21. Subsequently, mutations in the *APP* gene were found in a family with autosomal dominant early onset AD. These two discoveries—the identification of β-amyloid and the discovery of mutations in the parent *APP* gene—led the way to the amyloid cascade hypothesis, which has remained the dominant molecular model of the disorder (reviewed in Refs[52–54]). Many subsequent molecular observations have been consistent with this model, which posits the formation of β-amyloid as the initiating, or at least early event, leading to all the other changes observed including tau aggregation and phosphorylation, neuronal loss, cholinergic deficits, and clinical symptoms. Perhaps the most convincing evidence that there is such a unidirectional cascade comes from the observation that mutations in the *APP* gene give rise to plaque formation and also to neurofibrillary tangle pathology, whereas mutations in the tau gene give rise to tangle formation but not to plaque formation in the tauopathies.

Much subsequent research has concentrated upon understanding the metabolism of APP and the formation of β-amyloid peptide. APP is a ubiquitous single-pass cell-membrane protein expressed in many cell lines with a high degree of evolutionary conservation. At least three putative secretases cleave APP[55] and the metabolic products can be detected in individuals unaffected by AD; the processing is not pathological in AD, but the balance between different metabolic routes may be shifted in the disease state. α-Secretase cleaves APP at the outer cell-membrane surface at a site within the β;-amyloid moiety itself. Clearly, α-secretase cannot therefore yield intact β-amyloid, and this metabolic route, resulting in a secreted product, APPs, and other fragments, is termed non-amyloidogenic. On the other hand, amyloidogenic metabolism is the result of β-secretase (also known as Beta Amyloid Cleaving Enzyme or BACE) cleaving APP beyond the amino terminus of β-amyloid and of γ-secretase cleaving the resulting peptides at the carboxy terminus in the cell. The β-amyloid products vary in length, with predominant species having a length of 40 or 42 amino acids. The longer peptides are somewhat more prone to forming aggregates *in vitro* and it is probable that a relative increase in the longer peptides is critical in pathogenesis, and that mutations in the *APP* gene increase these longer amyloid peptides. Transgenic mice overexpressing the mutated *APP* gene also produce more β-amyloid peptide and have amyloid deposits in brain.[56] Interestingly, these animals do not develop other aspects of AD pathology, in particular they lack tangle formation. The nature of the toxicity of β-amyloid peptide is not fully understood but increasingly it appears that it is small aggregates (oligomers) that damage neurones rather than the longer fibrils that form the core of the plaque.[57]

(c) The presenilin genes

Mutations in *presenilin-1 (PS-1)* and *presenilin-2 (PS-2)*, two very similar genes on chromosome 14 and chromosome 1 respectively, also cause early onset autosomal dominant AD. The proteins encoded by these genes are part of the γ-secretase complex that metabolizes APP to β-amyloid. In fact these unique proteins turn out to have very many substrates and function in relation to many of these to release an intracellular component of a membrane bound protein that then translocates to the nucleus and triggers gene transcription events.[58] This is certainly the case for APP and also for Notch protein, which is also implicated in some other neurodegenerative conditions, for example. Mutations in the presenilin genes result in an increase in the production of β-amyloid probably through interfering with the normal γ-secretase complex.

Tangle formation and tau phosphorylation

Tangles are composed of paired helical filaments, themselves composed of aggregated and highly phosphorylated tau.[59–62] There are other post-translational modifications in tau, including truncation, and it is not fully determined which of these are primary events. However, in post-mortem studies, neuropathological evidence does suggest that highly phosphorylated tau accumulates in the brain before the formation of tangles, and before the clinical manifestation of AD suggesting that it is a early change in the pathological process.[63]

Protein phosphorylation is a product of kinase and phosphatase activity. It is likely that many such enzymes may participate in the regulation of tau phosphorylation in the brain, but two have been shown to be predominant. In cells, and *in vitro*, glycogen synthase kinase-3 and cyclin-dependent kinase 5 (CDK5) seem to be the predominant tau-kinases and protein phosphatase 2A is probably the predominant tau phosphatase.[64,65]

Molecular genetics

Mutations in three genes have been found to cause early onset familial AD, which is inherited in an autosomal dominant fashion.[66,67] Mutations in the *APP* gene (on chromosome 21) are the least common, only affecting perhaps 20 families worldwide. Mutations in *PS-1* (on chromosome 14) are somewhat more frequent, although are still a rare cause of AD. Mutations in *PS-2* (on chromosome 1) appear to be largely restricted to an ethnic German people residing in the United States, suggesting an individual founder effect. Individuals with Down's syndrome are at extremely high risk of AD, with neuropathological evidence being present in virtually all individuals living to middle age, probably because of trisomy APP (on chromosome 21). Mutations in other genes gives rise to disorders showing similarities to AD and much has been learnt from these findings about the overlap between neurodegenerative disorders. These genes include *tau* and *progranulin*, mutations in which give rise to FTD and related disorders.

The genetic component of late-onset AD has been demonstrated by epidemiological studies, showing that a family history of dementia is the largest single risk factor for AD.[68] However many, perhaps most, patients with AD do not have a positive family history, thus giving rise to the idea of 'sporadic' AD with a separate aetiology to 'familial' AD. For late-onset AD this concept is outmoded and redundant. Many patients with AD do not have a family history because of attrition of family members due to death by other causes. For the cohort currently suffering from AD their parents were born in the latter part of the nineteenth century or early years of the twentieth, lived through two major world wars, and reached adulthood before the discovery of antibiotics. It is not surprising that few patients with late-onset AD have two parents and more than one sibling living to the age of onset of AD, and if one parent died young and there are no elderly siblings then the family history is non-informative. Risk figures for relatives of probands with late-onset AD have been calculated and can be useful in counselling families.[69]

One gene has been unequivocally associated with late-onset AD, although even this gene accounts for only something like 50 per cent of the genetic variance.[70] The *apolipoprotein E* gene (*APOE*, gene; apoE, protein) on chromosome 19 has three common alleles, coding for three protein isoforms that differ by the substitution of an amino acid at just two positions. Of the three alleles ε3 is the most frequent and ε2 the least; following linkage to chromosome 18 it was demonstrated that the ε4 allele confers risk, whilst the ε2 may be protective.[71,72]

The mechanism of action of the *APOE* gene in increasing the risk of AD is not known. As *APOE* variation is a major genetic influence on serum cholesterol (people with the *APOE* ε4/* genotype have higher serum cholesterol levels), it is possible that an altered lipid metabolism—either peripherally or locally—might affect the pathogenesis of AD.[73,74] Alternative theories arise from *in vitro* studies, which show a differential binding of APOE protein isoforms both to amyloid protein and to tau protein.

Other genes have been associated with AD, but none have been replicated in as many studies as *APOE*. It is likely that a combination of linkage and association studies using large populations will identify the other genes that influence AD, either alone or in interactions with other genes or the environment.

Treatment

For many conditions the goals of treatment or intervention are self-evident—cure, prevention of relapse, and resolution of symptoms. For AD, however, the goals of treatment can be less obvious and differ between patients and for individual patients over time. Ultimately, the quality of life of the patient should be improved, but assessing quality of life is difficult in those with dementia, and given the early loss of insight who is to judge such issues?[75] Quality of life may appear poor—patients may have diminished emotional repertoires, few pleasurable activities, and considerable handicap—but they may share none of the negative cognitions experienced by others with a similarly questionable quality of life induced by different illnesses. Other patients may appear content or happy, despite the loss of the autonomy and self-awareness normally considered an essential component of a good quality life. The needs of the patient can be difficult to ascertain.[76] Equally, the treatment unit in AD includes carers, and there are times when the patient's quality of life is in conflict with the quality of life for other members of the family.[77] Resolving such conflicts of interest and other moral and ethical issues is part of the treatment process in AD. With the arrival of specific treatments for AD and the prospect of disease-modifying therapies, an even harder question arises regarding prolonging life for those with dementia: if quality of life appears poor to observers, is it right to prolong the process, can quality of life in those with dementia truly be assessed, or should carers and families be allowed to assess for themselves the benefits and costs of treatment?

There is no single model of management of patients with AD. In many countries management is the role of the gerontologist or neurologist. In others, as in the United Kingdom, the old-age psychiatry team provides the core specialist services. Many, perhaps even the majority, of those with AD are managed within primary care with the support of social services. Referral from primary care to specialist services will be according to local agreements, but most would concur that behavioural disturbance or the use of specific drugs to treat AD warrant referral to secondary care. Interventions for AD, whether provided in primary or secondary care, can be thought of as directed towards the patient, the patient's family, and the patient's environment. Guidelines on the identification and management of patients with dementia have been produced and may be a constructive approach to ensuring best clinical practice.[78]

Managing the patient

Management of the patient with dementia is discussed in greater detail in Chapter 4.1.13. Management starts with the assessment and diagnosis, and perhaps the difficult dilemma is how much of the diagnosis and prognosis to discuss with the patient.[79–81] Most practitioners do not discuss the diagnosis with the patient themselves, although practice is changing and especially in the early stages a frank consultation can be beneficial.

A large part of managing the patient is directed towards managing mood and behavioural disturbance. Accurate assessment of the disturbance is critical, and includes determining the antecedents and responses to the behaviour as well as a full description of the behaviour and any associated abnormalities in the mental state. Treatments of behavioural disturbance in AD are most often behavioural and sometimes restricted to giving information to carers. Evidence overwhelmingly suggests that anti-psychotic medication is relatively ineffective and has frequent adverse effects in dementia.[82,83] They should be a treatment of last resort, if at all.

Specific treatments for AD have been developed, concentrating in clinical trials on ameliorating the core symptom of cognitive impairment. The first to be licensed were the cholinesterase inhibitors followed by memantine. Drug treatments for AD are described in Chapter 6.2.7.

Managing the family

Although patients may not appreciate or be able to follow a detailed discussion of the diagnosis and prognosis, their relatives, spouses, and other carers will. This is an important part of the treatment process; as the carer provides the main interventions for much of the period of the disease process, care should be taken to ensure that appropriate and sufficiently complete information is given.

Caring for a patient with AD can be difficult and stressful and some carers suffer accordingly and need, and may benefit from support.[84] The characteristics of both carers and patients influence the impact that this 'burden' of caring has on the carers themselves. Men in general, and husbands in particular, seem to be less vulnerable to the adverse effects of caring, possibly because of the response seen in many male carers of rapidly and effectively recruiting outside help.[85] Women may be socialized into accepting more caring roles themselves and therefore seek less help. Non-white carers appear to suffer from less adverse consequences of caring, perhaps because of cultural differences in the perception of family bonds.[86] Patient characteristics that increase the burden of caring include behavioural disturbances,[25] depression, and unawareness of cognitive impairment but not the cognitive impairment itself. Although the core outcome variable in clinical trials of AD drugs is the severity of cognitive impairment, it is not the variable that induces most stress in relatives nor is it the variable that predicts entry to residential care. Other variables are almost certainly protective, and caring for a loved one with dementia is not a universally negative experience. Much caring is done willingly, effectively, with love, and without complaint.

Carer support groups offer much to a person with a relative afflicted by AD. Through support groups, and especially through the national AD societies and the umbrella group—AD International—carers can obtain up-to-date and useful information regarding all aspects of AD. A support group can help individuals practically and emotionally through difficult times. Many carers talk of the support group as a lifeline, although little empirical evidence exists as to the impact on carer well being.

One particular intervention for the family is that of genetic counselling. Many relatives are worried about inheriting AD. This concern might arise from two sources—the frequent discussion of genes 'for' AD in the media and the observation of familial occurrence of AD in many individual families. For families with clinically apparent familial AD, advice, information, and where appropriate, genetic testing can be arranged through a genetics centre. Where predictive testing is contemplated for genes causing autosomal dominant, early onset AD this will adhere to guidelines established for Huntington's disease. Genetic testing in late-onset AD is not recommended at the present time but is the subject of an ongoing research programme.[87]

Managing the environment

The mainstay of interventions for AD are provided by social services. The goal of the provision of social care in people with AD is to provide an environment that is comfortable, stimulating, and, above all, safe. For most patients, and for all patients in the early stages, this means care at home, perhaps with the support of home-meal delivery and home-helps to provide shopping and cleaning assistance. Further home care may become necessary as the patient requires assistance with basic self-care tasks such as washing and dressing. The carer may require a sitting service, either for periods during the day to allow them time to themselves or in the evening to allow them to attend a carers group or for socializing. Safety issues are especially important for those with dementia living alone. There are inherent risks to the patient themselves if they wander out of the home and risks to others if the gas can be left on or fires started.

Day care is appropriate for many patients, ideally in a specialist unit. In a generic facility for elderly people those with early dementia can receive little input and those with moderate or advanced dementia can necessitate too much input from the day-centre staff. A good dementia specialist day-care facility will provide the staffing ratio appropriate to patients with a range of 70s, in addition to providing a varied programme of group and recreational facilities to maintain interest and stimulation. Day centres, where patients are arrayed around the edge of the room with a television as a focal point, are, or should be, consigned to history. Day care provides essential respite to many carers, and longer periods of occasional or regular respite can prolong the period a patient can remain in their own home.

The multidisciplinary team consisting of care workers, social services, community psychiatric nurse occupational therapist, and psychologist can maintain patients at home more effectively and for longer periods than can clinicians alone. However, long-term care becomes a necessity for many patients at some point. The costs of providing nursing-home care are huge and far outweigh the costs of providing relatively intensive community care or relatively costly drugs. If treatments were shown to reduce the total length of stay in nursing homes then this would affect the cost–benefit ratio of these compounds considerably.

Translational research in AD

The rapid and comprehensive advances in understanding the molecular basis of AD has led to the promise of advances in health care—translational or bench to bedside research. Most importantly are potential disease modifying therapies. These are distinguished from symptomatic therapies in that they are designed to halt or slow down the disease process itself. Designing trials to assess efficacy of a potential treatment that might only slow down deterioration, and differentiating symptomatic effects from disease-modifying effects is not easy.[88] Two broad approaches are suggested—either a comparison of slopes of decline in which case a disease-modifying therapy would result in divergent slopes whereas a symptomatic therapy would result in parallel slopes, or strategies such as delayed start where in the case of a disease-modifying therapy the treatment arm of the delayed group never do quite as well as the early start group.[89]

Many approaches to disease modification are being pursued including therapeutics designed to alter APP processing—BACE inhibitors and γ-secretase inhibitors for example—or therapeutics designed to prevent β-amyloid from aggregating or for increasing the clearance of β-amyloid.[90] Many such compounds are in early stages of development and some were in phase III trials in 2007. Other potential therapies attempt to reduce tau phosphorylation or aggregation and yet other approaches are predicated on epidemiological findings such as the observation that non-steroidal anti-inflammatory drugs reduce risk of AD or that diabetes increases risk. Primary preventative therapies are probably even further away than disease-modification therapies but modifying cardiovascular risk or other approaches have been suggested. Secondary prevention, possibly in those with memory impairments not amounting to dementia (minimal cognitive impairment), is a more realistic prospect rendering the determination of the very earliest signs of disease or evidence of a prodromal state a high priority.

A second significant translational target in AD research close to clinical utility is that of biomarkers.[91,92] A marker is sought that might help in diagnosis, prediction or disease monitoring. For diagnosis a biomarker is sought that would make early or differential diagnosis more accurate, for prediction a biomarker that would help in predicting which elderly people were more likely to suffer from dementia or, more likely, which of those with mild cognitive impairment are more likely to convert to dementia. A marker of disease progression is sought that could supplement clinical assessments of deterioration. Of the many approaches to biomarkers, biochemical assays of tau and β-amyloid in CSF[91,93] and serial quantitative MRI[94] are the most promising although markers in plasma appear promising.[95,96]

Conclusions

For the foreseeable future, AD will remain a disorder afflicting a large proportion of the world's elderly. The impact on developing countries especially will be considerable. Care for these patients will continue to be provided from many sources, with specialist services being necessary to compliment primary and generic services, particularly for those patients exhibiting the complex psychiatric phenomenology described by Alzheimer and for those patients where specific drugs are indicated. As the molecular pathogenesis of AD is increasingly understood it is to be hoped that this is translated into treatments ever more effective in modifying or preventing the disease process itself.

Further information

http://www.alzforum.org/
http://www.alzheimers.org.uk/
http://alzheimers-research.org
http://www.alz.org/

References

1. Tomlinson, B.E., Blessed, G., and Roth, M. (1970). Observation on the brains of demented old people. *Journal of the Neurological Sciences*, **11**, 205–42.
2. Petersen, R.C., Smith, G.E., Ivnik, R.J., *et al.* (1988). Memory function in very early Alzheimer's disease. *Neurology*, **44**(5), 867–72.
3. Neary, D., Snowden, J.S., Northen, B., *et al.* (1988). Dementia of frontal lobe type. *Journal of Neurology, Neurosurgery, and Psychiatry*, **51**, 353–61.
4. Riter, R.N. and Fries, B.E. (1992). Predictors of the placement of cognitively impaired residents on special care units. *The Gerontologist*, **32**, 184–90.
5. Reed, B.R., Jagust, W.J., and Seab, J.P. (1989). Mental status as a predictor of daily function in progressive dementia. *The Gerontologist*, **29**, 804–7.
6. Reisberg, B. (1988). Functional assessment staging (FAST). *Psychopharmacology Bulletin*, **24**, 653–9.
7. Green, C.R., Mohs, R.C., Schmeidler, J., *et al.* (1993). Functional decline in Alzheimer's disease: a longitudinal study. *Journal of the American Geriatrics Society*, **41**, 654–61.
8. Levy, M.L., Cummings, J.L., Fairbanks, L.A., *et al.* (1996). Longitudinal assessment of symptoms of depression, agitation, and psychosis in 181 patients with Alzheimer's disease. *The American Journal of Psychiatry*, **153**(11), 1438–43.
9. Rovner, B.W., Broadhead, J., Spencer, M., *et al.* (1989). Depression and Alzheimer's disease. *The American Journal of Psychiatry*, **146**, 350–3.
10. Burns, A., Jacoby, R., and Levy, R. (1990). Psychiatric phenomena in Alzheimer's disease. III: disorders of mood. *The British Journal of Psychiatry*, **157**, 81–6, 92.
11. Tractenberg, R.E., Weiner, M.F., Patterson, M.B., *et al.* (2003). Comorbidity of psychopathological domains in community-dwelling persons with Alzheimer's disease. *Journal of Geriatric Psychiatry and Neurology*, **16**(2), 94–9.
12. Weiner, M.F., Doody, R.S., Sairam, R., *et al.* (2002). Prevalence and incidence of major depressive disorder in Alzheimer's disease: findings from two databases. *Dementia Geriatric and Cognitive Disorders*, **13**(1), 8–12.
13. Payne, J.L., Sheppard, J.M., Steinberg, M., *et al.* (2002). Incidence, prevalence, and outcomes of depression in residents of a long-term care facility with dementia. *International Journal of Geriatric Psychiatry*, **17**(3), 247–53.

14. Meeks, T.W., Ropacki, S.A., and Jeste, D.V. (2006). The neurobiology of neuropsychiatric syndromes in dementia. *Current Opinion in Psychiatry*, **19**(6), 581–6.
15. Zubenko, G.S., Moossy, J., and Kopp, U. (1990). Neurochemical correlates of major depression in primary dementia. *Archives of Neurology*, **47**, 209–14.
16. Forstl, H., Burns, A., Levy, R., *et al.* (1992). Neuropathological correlates of behavioural disturbance in confirmed Alzheimer's disease. *The British Journal of Psychiatry*, in press.
17. Holtzer, R., Tang, M.X., Devanand, D.P., *et al.* (2003). Psychopathological features in Alzheimer's disease: course and relationship with cognitive status. *Journal of the American Geriatrics Society*, **51**(7), 953–60.
18. Steinberg, M., Sheppard, J.M., Tschanz, J.T., *et al.* (2003). The incidence of mental and behavioral disturbances in dementia: the cache county study. *Journal of Neuropsychiatry and Clinical Neurosciences*, **15**(3), 340–5.
19. Wragg, R.E. and Jeste, D.V. (1989). Overview of depression and psychosis in Alzheimer's disease. *The American Journal of Psychiatry*, **146**, 577–87.
20. Burns, A., Jacoby, R., and Levy, R. (1990). Psychiatric phenomena in Alzheimer's disease. I: disorders of thought content. *The British Journal of Psychiatry*, **157**, 72–6, 92.
21. Assal, F., Alarcon, M., Solomon, E.C., *et al.* (2004). Association of the serotonin transporter and receptor gene polymorphisms in neuropsychiatric symptoms in Alzheimer disease. *Archives of Neurology*, **61**(8), 1249–53.
22. Nacmias, B., Tedde, A., Forleo, P., *et al.* (2001). Association between 5-HT(2A) receptor polymorphism and psychotic symptoms in Alzheimer's disease. *Biological Psychiatry*, **50**(6), 472–5.
23. Holmes, C., Arranz, M.J., Powell, J.F., *et al.* (1998). 5-HT_{2A} and 5-HT_{2C} receptor polymorphisms and psychopathology in late onset Alzheimer's disease. *Human Molecular Genetics*, **7**, 1507–9.
24. Coen, R.F., Swanwick, G.R., O'Boyle, C.A., *et al.* (1997). Behaviour disturbance and other predictors of carer burden in Alzheimer's disease. *International Journal of Geriatric Psychiatry*, **12**, 331–6.
25. Donaldson, C., Tarrier, N., and Burns, A. (1998). Determinants of carer stress in Alzheimer's disease. *International Journal of Geriatric Psychiatry*, **13**, 248–56.
26. Gormley, N., Rizwan, M.R., and Lovestone, S. (1998). Clinical predictors of aggressive behaviour in Alzheimer's disease. *International Journal of Geriatric Psychiatry*, **13**, 109–15.
27. Volicer, L., Harper, D.G., Manning, B.C., *et al.* (2001). Sundowning and circadian rhythms in Alzheimer's disease. *The American Journal of Psychiatry*, **158**(5), 704–11.
28. McKhann, G., Drachman, D., Folstein, M., *et al.* (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS–ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, **34**, 939–44.
29. Gearing, M., Mirra, S.S., Hedreen, J.C., *et al.* (1995). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part X. Neuropathology confirmation of the clinical diagnosis of Alzheimer's disease. *Neurology*, **45**, 461–6.
30. MRC CFAS. (2001). Pathological correlates of late-onset dementia in a multicentre, community-based population in England and Wales. Neuropathology Group of the Medical Research Council Cognitive Function and Ageing Study (MRC CFAS). *Lancet*, **357**(9251), 169–75.
31. Kukull, W.A., Larson, E.B., Reifler, B.V., *et al.* (1990). The validity of 3 clinical diagnostic criteria for Alzheimer's disease. *Neurology*, **40**, 1364–9.
32. Mok, W., Chow, T.W., Zheng, L., *et al.* (2004). Clinicopathological concordance of dementia diagnoses by community versus tertiary care clinicians. *American Journal of Alzheimer's Disease and Other Dementias*, **19**(3), 161–5.
33. Nagy, Z., Esiri, M.M., Hindley, N.J., *et al.* (1998). Accuracy of clinical operational diagnostic criteria for Alzheimer's disease in relation to different pathological diagnostic protocols. *Dementia and Geriatric Cognitive Disorders*, **9**(4), 219–26.
34. Halliday, G., Ng, T., Rodriguez, M., *et al.* (2002). Consensus neuropathological diagnosis of common dementia syndromes: testing and standardising the use of multiple diagnostic criteria. *Acta Neuropathologica*, **104**(1), 72–8.
35. Devi, G., Marder, K., Schofield, P.W., *et al.* (1998). Validity of family history for the diagnosis of dementia among siblings of patients with late-onset Alzheimer's disease. *Genetic Epidemiology*, **15**, 215–23.
36. Roth, M., Tym, E., Mountjoy, C.Q., *et al.* (1986). CAMDEX. A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. *The British Journal of Psychiatry*, **149**, 698–709.
37. Folstein, M.F., Folstein, S.E., and McHugh, P.R. (1975). Mini-mental state: a practical method of grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **12**, 189–98.
38. Bucks, R.S., Ashworth, D.L., Wilcock, G.K., *et al.* (1996). Assessment of activities of daily living in dementia: development of the Bristol Activities of Daily Living Scale. *Age Ageing*, **25**(2), 113–20.
39. Gelinas, I., Gauthier, L., McIntyre, M., *et al.* (1999). Development of a functional measure for persons with Alzheimer's disease: the disability assessment for dementia. *The American Journal of Occupational Therapy*, **53**(5), 471–81.
40. Knopman, D.S., Knapp, M.J., Gracon, S.I, *et al.* (1994). The Clinician Interview-Based Impression (CIBI): a clinician's global change rating scale in Alzheimer's disease. *Neurology*, **44**, 2315–21.
41. Hughes, C.P., Berg, L., Danziger, W.L., *et al.* (1982). A new clinical scale for the staging of dementia. *The British Journal of Psychiatry*, **140**, 566–72.
42. Waldemar, G., Dubois, B., Emre, M., *et al.* (2007). Recommendations for the diagnosis and management of Alzheimer's disease and other disorders associated with dementia: EFNS guideline. *European Journal of Neurology*, **14**(1), e1–26.
43. Cummings, B.J. and Cotman, C.W. (1995). Image analysis of β-amyloid load in Alzheimer's disease and relation to dementia severity. *Lancet*, **346**, 1524–8.
44. Nagy, Z., Esiri, M.M., Jobst, K.A., *et al.* (1995). Relative roles of plaques and tangles in the dementia of Alzheimer's disease: correlations using three sets of neuropathological criteria. *Dementia*, **6**, 21–31.
45. Mann, D.M., Brown, A., Prinja, D., *et al.* (1989). An analysis of the morphology of senile plaques in Down's syndrome patients of different ages using immunocytochemical and lectin histochemical techniques. *Neuropathology and Applied Neurobiology*, **15**, 317–29.
46. Lovestone, S. and Reynolds, C.H. (1997). The phosphorylation of tau: a critical stage in neurodevelopmental and neurodegenerative processes. *Neuroscience*, **78**(2), 309–24.
47. Avila, J. (2006). Tau phosphorylation and aggregation in Alzheimer's disease pathology. *FEBS Letters.* **580**(12), 2922–7.
48. Buee, L. and Delacourte, A. (1999). Comparative biochemistry of tau in progressive supranuclear palsy, corticobasal degeneration, FTDP-17 and Pick's disease. *Brain Pathology*, **9**(4), 681–93.
49. Heutink, P. (2000). Untangling tau-related dementia. *Human Molecular Genetics*, **9**(6), 979–86.
50. Braak, H. and Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*, **82**, 239–59.
51. Francis, P.T., Palmer, A.M., Snape, M., *et al.* (1999). The cholinergic hypothesis of Alzheimer's disease: a review of progress. *Journal of Neurology, Neurosurgery, and Psychiatry*, **66**, 137–47.
52. Hardy, J. (2006). Alzheimer's disease: the amyloid cascade hypothesis: an update and reappraisal. *Journal of Alzheimer's Disease*, **9**(Suppl. 3), 151–3.
53. Hardy, J. and Selkoe, D.J. (2002). The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science*, **297**(5580), 353–6.

54. Selkoe, D.J. (2000). Toward a comprehensive theory for Alzheimer's disease. Hypothesis: Alzheimer's disease is caused by the cerebral accumulation and cytotoxicity of amyloid beta-protein. *Annals of the New York Academy of Sciences*, **924**, 17–25.
55. Esler, W.P. and Wolfe, M.S. (2001). A portrait of Alzheimer secretases—new features and familiar faces. *Science*, **293**(5534), 1449–54.
56. Emilien, G., Maloteaux, J.M., Beyreuther, K., *et al.* (2000). Alzheimer disease: mouse models pave the way for therapeutic opportunities. *Archives of Neurology*, **57**(2), 176–81.
57. Walsh, D.M. and Selkoe, D.J. (2007). A beta Oligomers—a decade of discovery. *Journal of Neurochemistry*, **101**(5), 1172–84.
58. Parks A.L. and Curtis, D. (2007). Presenilin diversifies its portfolio. *Trends in Genetics*, **23**(3), 140–50.
59. Lace, G.L., Wharton, S.B., and Ince, P.G. (2007). A brief history of tau: the evolving view of the microtubule-associated protein tau in neurodegenerative diseases. *Clinical Neuropathology*, **26**(2), 43–58.
60. Mi, K. and Johnson, G.V. (2006). The role of tau phosphorylation in the pathogenesis of Alzheimer's disease. *Current Alzheimer Research*, **3**(5), 449–63.
61 Binder, L.I., Guillozet-Bongaarts, A.L., Garcia-Sierra, F., *et al.* (2005). Tau, tangles, and Alzheimer's disease. *Biochimica et Biophysica Acta*, **1739**(2–3), 216–23.
62. Iqbal, K., Alonso, A.C., Chen, S., *et al.* (2005). Tau pathology in Alzheimer disease and other tauopathies. *Biochimica et Biophysica Acta*, **1739**(2–3), 198–10.
63. Braak, E., Braak, H., and Mandelkow, E.-M. (1994). A sequence of cytoskeleton changes related to the formation of neurofibrillary tangles and neuropil threads. *Acta Neuropathologica*, **87**, 554–67.
64. Jope, R.S. and Johnson, G.V. (2004). The glamour and gloom of glycogen synthase kinase-3. *Trends in Biochemical Sciences*, **29**(2), 95–102.
65. Trojanowski, J.Q. and Lee, V.M. (1995). Phosphorylation of paired helical filament tau in Alzheimer's disease neurofibrillary lesions, focusing on phosphatases. *The FASEB Journal*, **9**(15), 1570–6.
66. Blacker, D. and Lovestone, S. (2006). Genetics and dementia nosology. *Journal of Geriatric Psychiatry and Neurology*, **19**(3), 186–91.
67. Bertram, L. and Tanzi, R.E. (2004). The current status of Alzheimer's disease genetics: what do we tell the patients? *Pharmacological Research*, **50**(4), 385–96.
68. van Duijn, C.M., Clayton, D.G., Chandra, V., *et al.* (1994). Interaction between genetic and environmental risk factors for Alzheimer's disease: a reanalysis of case-control studies. *Genetic Epidemiology*, **11**, 539–51.
69. Cupples, L.A., Farrer, L.A., Sadovnick, A.D., *et al.* (2004). Estimating risk curves for first-degree relatives of patients with Alzheimer's disease: the REVEAL study. *Genetics in Medicine*, **6**(4), 192–6.
70. Owen, M., Liddell, M., and McGuffin, P. (1994). Alzheimer's disease. *British Medical Journal*, **308**, 672–3.
71. Raber, J., Huang, Y., and Ashford, J.W. (2004). ApoE genotype accounts for the vast majority of AD risk and AD pathology. *Neurobiology of Aging*, **25**(5), 641–50.
72. Cedazo-Minguez, A. and Cowburn, R.F. (2001). Apolipoprotein E: a major piece in the Alzheimer's disease puzzle. *Journal of Cellular and Molecular Medicine*, **5**(3), 254–66.
73. Hatters, D.M., Peters-Libeu, C.A., and Weisgraber, K.H. (2006). Apolipoprotein E structure: insights into function. *Trends in Biochemical Sciences*, **31**(8), 445–54.
74. Martins, I.J., Hone, E., Foster, J.K., *et al.* (2006). Apolipoprotein E, cholesterol metabolism, diabetes, and the convergence of risk factors for Alzheimer's disease and cardiovascular disease. *Molecular Psychiatry*, **11**(8), 721–36.
75. Ettema, T.P., Droes, R.M., de, L.J., *et al.* (2005). The concept of quality of life in dementia in the different stages of the disease. *International Psychogeriatrics*, **17**(3), 353–70.
76. van der Roest, H.G., Meiland, F.J., Maroccini, R., *et al.* (2007). Subjective needs of people with dementia: a review of the literature. *International Psychogeriatrics*, **19**(3), 559–92.
77. Glozman, J.M. (2004). Quality of life of caregivers. *Neuropsychology Review*, **14**(4), 183–96.
78. Waldemar, G., Dubois, B., Emre, M., *et al.* (2000). Diagnosis and management of Alzheimer's disease and other disorders associated with dementia. The role of neurologists in Europe. European Federation of Neurological Societies. *European Journal Neurology*, **7**(2), 133–44.
79. Derksen, E., Vernooij-Dassen, M., Gillissen, F., *et al.* (2006). Impact of diagnostic disclosure in dementia on patients and carers: qualitative case series analysis. *Aging & Mental Health*, **10**(5), 525–31.
80. Elson, P. (2006). Do older adults presenting with memory complaints wish to be told if later diagnosed with Alzheimer's disease? *International Journal of Geriatric Psychiatry*, **21**(5), 419–25.
81. Lin, K.N., Liao, Y.C., Wang, P.N., *et al.* (2005). Family members favor disclosing the diagnosis of Alzheimer's disease. *International Psychogeriatrics*, **17**(4), 679–88.
82. Schneider, L.S., Tariot, P.N., Dagerman, KS., *et al.* (2006). Effectiveness of atypical antipsychotic drugs in patients with Alzheimer's disease. *The New England Journal of Medicine*, **355**(15), 1525–38.
83. Schneider, L.S., Dagerman, K.S., and Insel, P. (2005). Risk of death with atypical antipsychotic drug treatment for dementia: meta-analysis of randomized placebo-controlled trials. *The Journal of American Medical Association*, **294**(15), 1934–43.
84. Thompson, C.A., Spilsbury, K., Hall, J., *et al.* (2007). Systematic review of information and support interventions for caregivers of people with dementia. *BMC Geriatrics*, **7**, 18.
85. Bedard, M., Molloy, D.W., Pedlar, *et al.* (1997). Associations between dysfunctional behaviors, gender, and burden in spousal caregivers of cognitively impaired older adults. *International Psychogeriatrics*, **9**, 277–90.
86. Connell, C.M. and Gibson, G.D. (1997). Racial, ethnic, and cultural differences in dementia caregiving: review and analysis. *The Gerontologist*, **37**, 355–64.
87. Roberts, J.S., Barber, M., Brown, T.M., *et al.* (2004). Who seeks genetic susceptibility testing for Alzheimer's disease? Findings from a multisite, randomized clinical trial. *Genetics in Medicine*, **6**(4), 197–203.
88. Vellas, B., Andrieu, S., Sampaio, C., *et al.* (2007). Disease-modifying trials in Alzheimer's disease: a European task force consensus. *Lancet Neurology*, **6**(1), 56–62.
89. Leber, P. (1996). Observations and suggestions on antidementia drug development. *Alzheimer Disease and Associated Disorders*, **10**(Suppl. 1), 31–35.
90. Pangalos, M.N., Jacobsen, S.J., and Reinhart, P.H. (2005). Disease modifying strategies for the treatment of Alzheimer's disease targeted at modulating levels of the beta-amyloid peptide. *Biochemical Society Transactions*, **33**(Pt 4), 553–8.
91. Sunderland, T., Hampel, H., Takeda, M., *et al.* (2006). Biomarkers in the diagnosis of Alzheimer's disease, are we ready? *Journal of Geriatric Psychiatry and Neurology*, **19**(3), 172–9.
92. Lovestone, S. (2006). Biomarkers in Alzheimer's disease. *The Journal of Nutrition Health & Aging*, **10**(2), 118–22.
93. Wiltfang, J., Lewczuk, P., Riederer, P., *et al.* (2005). Consensus paper of the WFSBP task force on biological markers of dementia: the role of CSF and blood analysis in the early and differential diagnosis of dementia. *The World Journal of Biological Psychiatry*, **6**(2), 69–84.
94. Fox, N.C., Black, R.S., Gilman, S., *et al.* (2005). Effects of Abeta immunization (AN1792) on MRI measures of cerebral volume in Alzheimer disease. *Neurology*, **64**(9), 1563–72.
95. van, O.M., Hofman, A., Soares, H.D., *et al.* (2006). Plasma Abeta(1-40) and Abeta(1-42) and the risk of dementia: a prospective case-cohort study. *Lancet Neurology*, **5**(8), 655–60.
96. Hye, A., Lynham, S., Thambisetty, M., *et al.* (2006). Proteome-based plasma biomarkers for Alzheimer's disease. *Brain*, **129**(Pt 11), 3042–50.

4.1.3 Frontotemporal dementias

Lars Gustafson and Arne Brun

Introduction

Nosological classification of organic dementia is based on current knowledge and theories of aetiology, including genetics, clinical picture, the pathological substrate, and the predominant location of brain damage. This chapter is concerned with dementia syndromes caused by a degenerative disease primarily affecting the frontal and temporal lobes, named frontal-lobe dementia[1] or frontotemporal dementia (**FTD**).[2] The terminology should be viewed from a historical perspective. The relationship between localized cortical atrophy in dementia and symptoms of aphasia was first reported by Pick in 1892.[3] The pathological account of this lobar degeneration by Alzheimer in 1911 described 'ballooned' neurones (Pick cells) and argentophilic globes (Pick bodies),[4] and the clinicopathological entity was named **Pick's disease**.

In the 1980s, attention was drawn to a larger group of frontal-lobe dementias associated with frontotemporal cortical degeneration.[5–7] The Lund–Manchester consensus of 1994 delineated the prototypical clinical syndrome of FTD with three neuropathological constituents, frontal lobe degeneration of non-Alzheimer type (**FLD**),[5] (alternatively designated '**dementia lacking distinctive histology**'),[8] Pick's disease, and motor neurone disease (**MND**) with dementia (**FTD-MND**).[2] The 1998 consensus on clinical diagnostic criteria for frontotemporal lobar degeneration (**FTLD**)[9] encompassed two additional dementia syndromes; progressive non-fluent aphasia (**PA**),[10,11] and **semantic dementia**.[12] Corticobasal degeneration (**CBD**) and progressive supranuclear palsy (**PSP**) have also been associated with FTLD.[13] A changing clinical classification is shown in Fig. 4.1.3.1. The addition of important genetic and histochemical characteristics has further added to the complex classification of FTD and FTLD with a risk of developing numerous and partly competing definitions. FTLD may be further subclassified into forms positive or negative for tau and ubiquitin. The ubiquitinated form will be referred to as **FTD-U**, which is synonymous to **FLTD-U**.[14]

Neuropathology

On a neuropathological basis about two-thirds of the FTLD cases are of the type with ubiquitinated inclusions (FTD-U) or of FTD type without such inclusions (FLD), both lacking tau pathology.[14,15] In an attempt to classify FTD forms from a structural point of view, FLD might be appointed as a basic form, showing type and distribution of main pathological changes common to the majority of FTLD forms. To this set of alterations are added further features proceeding in the description of frontotemporal degenerative disorders (Table 4.1.3.1).

In FLD, the cortex is the site of a *simple* degenerative process resulting in a cortical atrophy which is frontotemporal with or without asymmetries, moderate or even at times mild. It involves cortical layers 1–3, showing neuronal loss, gliosis, and microvacuolation, seen also in the striatum in a small proportion of cases as well as a mild-to-moderate degeneration of the substantia nigra.[5] DLDH has essentially the same pathology as FLD.[8]

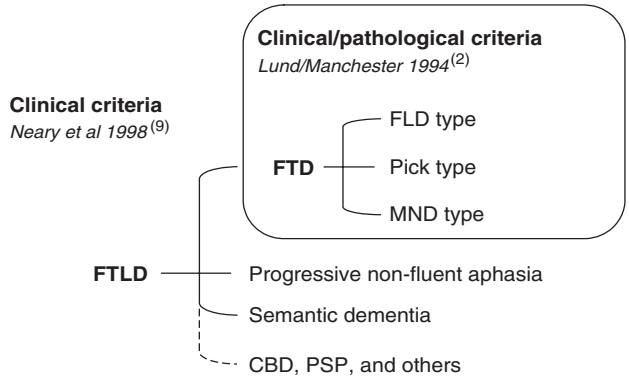


Fig. 4.1.3.1 Clinical classification of frontotemporal dementias.

In FTD-U the ubiquitin-positive inclusions and dystrophic neurites supplement the picture of FLD, as also described in FTD, linked to chromosome 3.[16] The inclusions contain a ubiquitinated protein, also identified in some other varieties of FTD such as FTD-MND.[17] Here there is also a degeneration of motor neurones and paths. The language variants of FTLD show a mainly temporal degeneration in semantic dementia and an asymmetric frontotemporal degeneration with left-sided predominance in PA.

With a mainly frontal or frontotemporal including anterior cingulate gyrus involvement FTD differs markedly from Alzheimer's disease (AD) with its posterior temporal parietal, posterior cingulate, and severe hippocampal involvement and also with Lewy bodies, amyloid, and plaque pathology not seen in FTD. So far mentioned FTD forms belong in the same histopathological group of non-tauopathies and further share a frequent and individually varyingly severe white-matter sclerosis, which is often predominantly frontal and may be primary or secondary.[18] It is also seen in the tauopathy group and in forms of vascular dementia.

The remaining, less common forms comprise the FTD-tau group. In Pick's disease the degenerative process is more intense and partly involves all cortical layers, creating a severe circumscribed or lobar frontotemporal atrophy referred to as 'knife blade atrophy'. Microscopically there are tau-positive neuronal inclusions and sometimes glial tangles and often ballooned nerve cells. The FTD with Parkinsonism (FTDP-17) group with a number of familial disorders shares basic pathological features with the FLD but is

Table 4.1.3.1 Organic dementia with frontotemporal lobar degeneration (FTLD)

Frontal lobe degeneration of non-Alzheimer type (FLD) (Dementia lacking distinctive histology) without ubiquitinated inclusions
FTD with ubiquitin-positive inclusions (FTD-U)
Familial FTD with chromosome 3 mutation (FTD-3)
FTD with motor neurone disease (FTD-MND)
Progressive non-fluent aphasia (PA)
Semantic dementia
Pick's disease
FTD with Parkinsonism (FTDP-17)
Corticobasal degeneration (CBD)
Progressive supranuclear palsy (PSP)

regularly more severe in the basal ganglia and substantia nigra.[19] Glial cells and neurones contain various types of tau-positive inclusions. CBD is structurally and clinically heterogeneous, resulting in overlap with other diseases, especially PSP.[20] These two may represent varieties of the same pathological process, less obviously belonging in the FTD group even if a similar laminar frontal cortical degeneration and related symptoms are part of the presentation. In CBD ballooned neurones, tau-positive inclusions in neuronal and glial cells, white-matter rarefaction and nigral degeneration may be found. For PSP different forms and patterns are noted but it basically affects more widely the striatum, basal ganglia, and hypothalamus. It also involves the brain stem including the substantia nigra as well as the cerebellum, again with silver- or tau-positive inclusions and tangles in neurones and glial cells.[21]

Epidemiology

Most demographic data concern the grouping of FTD, not separating FLD and Pick's disease. Pick's disease is rare, estimated at 24–60/100 000 in Minnesota and 30–60 in Switzerland.[22] The calculated prevalence of FTD in the Netherlands is 10.7 per million between 50 and 60 years of age and 28 between 60 and 70 years.[23] The prevalence of FTD in the province Zuid-Holland in the Netherlands was 3.6/100 000 at age 50–59, 9.4 at age 60–69 years, and 3.8 at age 70–79 years.[24] Pasquier *et al.*[25] diagnosed FTD in 4.8 per cent of all types of dementia. The marked geographic variation of the prevalence might be due to genetic and environmental factors, but also influenced by differences in the diagnostic process and the age group studied. In a clinical study of a total catchments area of 20 million people in Germany the relative proportion of FTLD was 1.9 per cent.[26] The prevalence of dementia in motor neurone disease has been estimated to 2–6 per cent.[27] The proportion of FTD in relation to all types of dementia in different clinicopathological studies varies between 5 and 18.9 per cent.[7, 8, 28–30]

Clinical features

The first clinical manifestations of FTD usually appear in the presenium, in some cases as early as 35 and seldom after 70 years of age. The mean age at onset in post-mortem verified FLD cases is 56 ± 7.6 years with a mean duration of 8 ± 3.4 years (range 3–17 years).[31] The mean age of onset in Pick's disease is similar, 62 years, with a range of 40–80 years and a mean survival of 9.8 years with a range of 4.8–21.2 years.[22] The large variations of the duration of FLD and Pick's disease are similar to that of early-onset AD. The clinical onset of MND dementia is usually in the sixth decade and the mean duration is about 30 months. Age at onset is similar in familial and sporadic cases of FTD and sometimes past 80 years.[32] The Lund–Manchester consensus on clinical criteria for FTD is summarized in Table 4.1.3.2.

Table 4.1.3.2 The Lund–Manchester consensus (1994) on clinical criteria for frontotemporal dementia (slightly modified)[2]

Core diagnostic features
Behavioural disorder
Insidious onset and slow progression
Early loss of insight into changes of own mental state
Early loss of personal and social awareness
Early signs of disinhibition and lack of judgement
Mental rigidity and inflexibility
Stereotyped, repetitive, and imitating behaviour
Hyperorality, oral/dietary changes
Utilization behaviour
Distractibility, impulsivity, and impersistence
Affective symptoms
Depression, anxiety, excessive sentimentality
Hypochondriasis, bizarre somatic complaints
Emotional bluntness, apathy, and lack of empathy
Amimia
Speech disorder
Progressive reduction of speech output
Stereotypy of speech, perseveration
Echolalia
Late mutism
Spatial orientation, receptive speech, and praxis preserved
Physical signs
Early primitive reflexes
Early incontinence
Late akinesia, rigidity, and tremor
Low and labile blood pressure
Investigations
Normal EEG despite clinically evident dementia
Brain imaging (structural and/or functional): predominant frontal and/or anterior temporal abnormality
Neuropsychology: profound failure on 'frontal-lobe' tests in the absence of severe amnesia, or perceptual spatial disorder
Supportive diagnostic features
Onset before 65
Positive family history of similar disorder in a first-degree relative
Bulbar palsy, muscular weakness and wasting, fasciculations (motor neurone disease)

(Reproduced from A. Brun, E. Englund, L. Gustafson, *et al.* Consensus statement—clinical and neuropathological criteria for frontotemporal dementia, *Journal of Neurology, Neurosurgery, and Psychiatry*, **57**, 416–18, copyright 1994, BMJ Publishing Group Ltd.)

Disordered behaviour

The early stage of FLD and Pick's disease is characterized by changes of personality and behaviour, affective symptoms, and a progressive reduction of expressive speech. The clinical onset is insidious with slow progression without ictal events. The changes of personality and behaviour are rather non-specific and easily misinterpreted as a non-organic mental disease such as mood disorder, stress reaction, schizophrenia, or other psychotic reaction. Loss of insight concerning the mental changes is an early and alarming manifestation of the disease. FTD patients may, however, consult a doctor referring to symptoms such as anxiety, tiredness, and strange somatic complaints combined with bizarre hypochondriacal ideas.

The early loss of personal and social awareness is seen as neglect of personal hygiene and grooming, and tactlessness and antisocial behaviour.[6, 9] The impaired control of behaviour is seen as increased sentimentality, inadequate smiling, inappropriate joking, irritability, and acts of aggressiveness, leading to conflicts at home and work. Craving for affection and sexual contact may be easily provoked, but usually expressions of sexual disinhibition are possible to divert. Impulse buying, shoplifting, indecency, and other disinhibited behaviour may, however, lead to rejection by the family and society. Such unpredictable and pseudopsychopathic

behaviour imposes severe strain on the patient's family, leading in some cases to economic problems, divorce, and even suicide in the family.[33] Complications of this type are uncommon in families with an AD patient. FTD patients tend to become inattentive and careless and a danger to traffic. Changes in drinking behaviour are sometimes reported. The patient starts to drink more frequently and in larger quantities than before. The changes of behaviour, which may lead to misdiagnosis of alcohol-induced dementia, can often be controlled by a firm attitude from relatives.

Affective symptoms

The FTD patient becomes emotionally shallow and blunt, showing less concern about family and friends. The patient is described as egocentric, rigid, and lacking empathy. The early emotional changes may be difficult to differentiate from non-organic personality disorders and affective disorder. Mood changes towards euphoria, especially when associated with press of speech and overactivity, may at first be mistaken for a hypomanic or manic state. Slowly developing apathy, in combination with sparse mimical movements and verbal aspontaneity, may be misdiagnosed as depression. During the depressive reactions, which are mostly of short duration, the patient may become dysphoric, and dwell on suicidal thoughts. FTD patients are often diagnosed as depressed and treated with antidepressant medication during the early stage of the disease.[32]

Early symptoms of dementia must be judged against information about the patient's premorbid personality, education, and social background. The vast majority of cases show normal premorbid personality although a few have previously manifested anxiety and restlessness. The emotional features in FTD do not seem primarily related to premorbid personality traits but rather to the distribution of brain pathology as shown at autopsy and brain imaging.[34]

Other symptoms

A striking feature of FTD is the **stereotyped and perseverative behaviour** seen as wandering, clapping, humming, dancing, and hoarding of objects, as well as complex rituals involving washing and dressing. Such behaviour sometimes reaches psychotic intensity. Imitative behaviour is frequent in FTD and occurs more often than in AD.

Hallucinations and delusions are reported in about 20 per cent of FTD and early-onset AD cases. The psychotic symptoms in FTD are often bizarre and the combination with emotional changes and stereotypy of speech and behaviour gives the impression of functional psychosis with schizophrenia as an early tentative diagnosis.[6–8] The psychotic symptoms in early-onset AD seem more strongly related to the cognitive failure with memory failure, impaired recognition, and disorientation, and the degeneration of the temporoparietal association cortex.[6]

The **human counterpart of the Klüver–Bucy syndrome** has been reported in FLD and in Pick's disease. The hyperorality and changes of oral/dietary behaviour are seen as overeating, food fads, excessive smoking, and alcohol consumption. Utilization behaviour, defined as an irresistible impulse to explore and use objects in the visual environmental, shows important similarity to the hypermetamorphosis and distractibility of the Klüver–Bucy syndrome.[35] The Klüver–Bucy syndrome in AD is usually less complete than in FTD with less hypersexuality and utilization behaviour, supporting the suggestion that frontal as well as temporal limbic involvement is needed to produce the syndrome in humans.[6]

Dissolution of language

A core feature of FTLD is progressive impairment of speech and language. In FTD, this has been described as *dissolution du langage* or *Sprachverödung*.[6] Speech becomes aspontaneous with word-finding difficulties and frequent use of stereotyped comments and set phrases. During the early stage there may also be increased pressure of speech. The language dysfunction in FTD is dominated by dynamic expressive failure, which is in agreement with damage in the frontal cortex especially premotor areas. Echolalia is observed in about 50 per cent of FTD and Pick cases.[31] Finally the patients become mute which in combination with the amimia makes communication extremely difficult. The ability to understand information and instructions usually remains until comparatively late in the course of FTD, as does the ability to write. The handwriting may, however, change in magnitude, spelling, and speed of writing. These disturbances are unlike the temporoparietal type of dysgraphia and global dysphasia observed in AD. The symptom constellation of palilalia (stereotypy of speech), echolalia, mutism, and amimia (PEMA syndrome of Guiraud) is typical of FTD and seldom found in AD.

There are important similarities between the speech disorder of early FTD and the clinical spectrum of PA,[10] characterized by effortful speech production in the context of preserved word comprehension and relative preservation of memory and practical abilities. Dementia often develops later in the course, and the underlying degenerative process may be similar to that of FLD, with a predominant and early involvement of the speech-dominant hemisphere. Semantic dementia, the fluent language variant of FTLD is characterized by progressive loss of word retrieval and understanding, and recognition of sensory stimuli.[36] The pathological substrate is bilateral, often asymmetric temporal lobe degeneration.

Physical signs

Few pathological somatic findings including neurological symptom are reported early in FTD. However, primitive reflexes may appear early, while akinesia, rigidity, and tremor may emerge later in the course. Increased muscular tension is, however, significantly more common in AD.[6] The spectrum of FTLD also includes the syndrome of the disinhibition–dementia–parkinsonism–amyotrophy complex linked to chromosome 17 also named FTD-17.[37]

Generalized epileptic seizures may appear in FTD although less prevalent than in AD, and myoclonic twitchings and logoclonia which are prevalent in early-onset Alzheimer's cases are rare in FTD. Urinary incontinence, which is reported early in about 50 per cent of FTD cases, is a comparatively late feature in uncomplicated early-onset AD.

FTD patients in general have low and labile blood pressure with a high prevalence (50 per cent) of orthostatic blood pressure drops and syncopal attacks. These symptoms are, however, also reported in early-onset AD (40 per cent) and late in the course of vascular dementia (50 per cent). The relationship between blood pressure changes and brain damage especially the white-matter changes in FTD is still unclear.[18]

Dementia in motor neurone disease

The clinical picture of the dementia in motor neurone disease is similar to that in FLD with early changes of personality and behaviour, emotional changes such as euphoria and apathy, and signs of disinhibition and hyperorality.[38] Speech becomes stereotyped and perseverative, later developing into mutism. Receptive speech function, orientation, and practicable abilities remain relatively untouched by the degenerative process. The mental changes may appear early and even precede development of typical neurological features.

Investigations

EEG

The EEG may be normal or only slightly pathological in FTD at a stage when dementia is strongly suspected or clinically evident, but it is usually pathological late in the course. This has been shown in FLD, Pick's disease, and FTD-MND.[39] By contrast, EEG is almost always pathological in AD even at an early stage. Quantitative EEG mapping and repeated recordings may strongly improve the differential diagnosis between FTD and AD.[32]

Brain imaging and other investigations

Structural and functional brain imaging has strongly improved the diagnosis and differential diagnosis of FTD, AD, and other dementias. Cortical atrophy with more or less frontal focal accentuation, sometimes asymmetrical, is shown with CT and magnetic resonance imaging (**MRI**).[8,40] MRI may show significantly more prevalent and severe frontal periventricular white-matter lesions in FTD patients than in matched normal controls. The anterior–posterior gradient of the atrophic changes may, however, contribute to the differentiation from AD.[18,40] The differential diagnosis from vascular dementia with frontal subcortical lesions, but lacking large cortical infarctions, may be difficult.

Functional brain imaging measuring regional cerebral blood flow (rCBF) and metabolism with SPECT, PET, and other techniques show frontal and frontotemporal flow pathology in FTD with better preserved perfusion in posterior areas.[41,42] PET studies have indicated the ventromedial frontal cortex as the earliest site of imaging pathology.[43] These changes may at an early stage be mild and asymmetric in accordance with the clinical picture.

Recently several biomarkers in the cerebrospinal fluid (CSF) have been developed for differential diagnosis of dementia. Riemenschneider *et al.*[44] reported significantly higher CSF tau concentrations in FTD compared to healthy controls, but significantly lower than in AD, while CSF Abeta42 levels were significantly lower in FTD than in controls, but significantly higher than in AD. Early diagnosis of FTD might in the future be based on a combination of profile and levels of CSF biomarkers such as tau, β-amyloid, and neurofilaments.[45] Interestingly tau levels are dependent on lobar localization but independent of the degree of cerebral atrophy.[44,46]

Assessment of cognitive impairment

The cognitive changes, which appear early in FTD, may be difficult to evaluate due to the patient's emotional and behavioural changes. Distractibility and slightly reduced recent memory are common findings and remote memory is also impaired although to a lesser extent than in AD. The patients show significant impairment on 'frontal-lobe' tests such as the Wisconsin card sorting test, word fluency test, and the Stroop and trail-making tests. The early test profile is characterized by slow verbal production and relatively intact visuospatial ability, reasoning, and memory, while intellectual and motor speed are reduced.[10,47] Early AD usually shows a relatively preserved verbal ability and simultaneous impairment of reasoning ability, verbal and spatial memory dysfunction, dysphasia, and dyspraxia.[10,48,49] Difficulties in understanding instructions are found early only in a minority of FTD cases. Misspelling and dyscalculia are sometimes reported early in FTD.

Discrimination between FTD and AD can be based on a short test-battery (verbal ability, visuospatial ability, and verbal memory), when used in the context of a neuropsychological evaluation of qualitative as well as quantitative aspects of test performance.[47,49,50] Using a screening instrument based on frontal release signs, awareness of social/ethical dilemma in a short story, and the number of preservation errors, FTD was classified correctly in 83 per cent, validated against clinical diagnosis.[51] The Mini-Mental State Examination does not reflect the FTD patient's true competence because of influence of motivational and behavioural factors.[52]

Differential diagnosis

Differential diagnosis between FTD and AD and other dementias is often possible based on a careful clinical history and examination, supported by diagnostic tests and brain imaging. Detailed neuropathology remains a gold standard for definite diagnosis of FTD and other disorders presenting with FTD-like clinic (Table 4.1.3.3).

The clinical differences between FTD and AD are often obvious at an early stage. The initial stages of FTD are dominated by emotional and personality changes, and progressive reduction of speech. Consequently severe dyspraxia, memory failure, and spatial disorientation develop comparatively late with the relative sparing of the temporoparietal occipital cortical areas. In contrast, early-onset AD is characterized by memory failure, dyspraxia, dysgnosia, and impaired sense of locality, whereas habitual personality traits, social competence, and insight are better preserved in agreement with the consistent pattern of cortical involvement.[6] A minority of AD cases, about 5 per cent, show a marked frontal-lobe involvement at an early stage and consequently also present a frontal-lobe clinical pattern in addition to the temporoparietal symptoms.

The Lund–Manchester consensus[2] is recommended as a guideline for clinical recognition and differential diagnosis of FTD.[53–55]

Table 4.1.3.3 Clinical diagnostic alternatives to FTD

Alzheimer's disease (AD) with frontal emphasis
Vascular dementia with frontal emphasis
Selective white-matter infarction
Binswanger's disease
Multiinfarct dementia with frontal emphasis
Strategic infarct dementia (striatal, thalamic)
Huntington's disease
Creutzfeldt–Jacob disease
General paresis

The NINCDS–ADRDA criteria for AD were originally formulated with the aim to differentiate between AD and vascular dementia. Varma *et al.*[56] found a high sensitivity for probable AD, but also a low specificity since 77 per cent of pathologically confirmed FTD cases fulfilled the NINCDS–ADRDA criteria for AD.

Vascular dementia with frontal emphasis may be caused by selective incomplete white-matter infarction, Binswanger's disease, and frontal and strategic thalamic infarctions. The frontal-lobe dysfunction caused by vascular lesions may closely mimic the course of FTD, when developing gradually without dramatic onset or fluctuations.

The clinical distinction between FTD and Huntington's disease may be difficult when personality changes and psychotic features dominate, and when neurological characteristics are less obvious or appear late in the course. Brain imaging showing striatal involvement and genetic analysis may contribute to the diagnosis.

PSP and the rare progressive subcortical gliosis may also show a frontal-lobe clinical and imaging pathology. CBD may also present with a dementia of frontal-lobe type in addition to the typical asymmetric akinetic-rigid dystonic syndrome.[20] These three diseases have grown increasingly important because of studies suggesting a linkage to chromosome 17.[19]

Dementia of the frontal and frontal subcortical type is also found in Creutzfeldt–Jakob disease, in the AIDS dementia complex, and in general paresis.

Classification dilemma

Classification of frontotemporal dementias must be viewed in a historical perspective. The various clinicopathological entities have been identified by presence of certain and absence of other clinical and pathological features. The current classification and terminology of FTD illustrates, however, the difficulties of most one-dimensional diagnostic systems. A classification on pure clinical grounds is unsatisfactory since symptoms depend more on the starting point in the brain or topography than type of changes, and also on the progression pattern of the disease among brain areas, factors which may vary between cases with structurally identical disorders. A diagnostic classification should be clinically useful and valid, flexible and open to new ideas. There is a need for combination of classifications such as one based on phenomenological syndrome kept apart from, and combined with, aetiological including genotypical classification. Hopefully new generations of classifications and new diagnostic techniques and treatment strategies will further increase the awareness of FTD in clinical practice and research.

DSM-IV[1] and ICD-10[57] do not introduce the concepts of frontal-lobe dementia or FTD. DSM-IV presents Pick's disease as 'One of the pathologically distinct aetiologies among the heterogeneous group of dementing processes that are associated with frontotemporal brain atrophy'. ICD-10 describes 'dementia in Pick's disease' as an early-onset non-Alzheimer degenerative brain disease.

Aetiology and pathogenesis

The FTD is a heterogeneous disease group, but with important clinical features and probably also aetiological factors in common. The clinical and pathological similarities between familial and sporadic cases are, however, striking. About 40–50 per cent of patients with FLD have a history of similar disorder in a first-degree relative.[23,31] A family history study on 478 first-degree relatives of 74 index patients suffering from FTD reported a 10-fold increase in the incidence of FTD compared with the incidence of FTD in a population study.[58] A Swedish pedigree with FTD in 10 out of 21 family members in three generations has been described.[59]

Several genetic loci for FTD have been identified at human chromosomes 3, 9, and 17 in familial forms of the disease. One FTDP-17 locus has been mapped to a region of chromosome 17 where the tau gene is located. So far 35 different mutations in around 100 families have been identified.[36] However, tau mutations seem to be a rare cause of FTD. Dementia of frontal-lobe type has also been linked to chromosome 3p11–12 in a Danish family.[16] Conflicting results exist concerning the relation of FTD-MND to a locus on chromosome 9[60] and to chromosome 19 and the ApoE-allele pattern.[61] There is no solid proof of an autosomal-dominant inheritance in the majority of studies with Pick disease.

Several research groups have recently found mutations in the progranulin (PGRN) gene close to the tau gene to result in a loss of function of the growth or maintenance factor PGRN. This leads to a shortage of PRGN and degenerative dementia of the FTD type, including autosomal-dominant FTLD-U linked to chromosome 17 as well as further tau-positive and tau-negative forms.[17,62] Further research has shown that this mechanism may operate also in other neurodegenerative brain disorders including CBD and MND. MND is often linked to FTD-U by a common ubiquitinated protein[17] and by a frequent concurrence. This protein TDP-43 forms inclusions in a wide variety of functional brain areas in close correspondence to the symptomatology in the FTD spectrum, the clinical expression of which depends on the individual anatomical starting point and spread of the degeneration. The identification of mutations in PGRN explains why multiple families linked to chromosome 17 lack tau mutations.[62]

The pattern of degeneration in FTD may be related to a selective vulnerability of different brain regions to factors such as oxidative stress, environmental toxins, neurotransmitter dysfunction, and certain mutations. A retrospective study of risk for sporadic FTD reported a significantly higher prevalence of head trauma and thyroid disease.[63] A prion aetiology has been excluded in FLD and also in FTDP-17.

Neurochemical post-mortem studies of FTD have indicated abnormalities in serotonin metabolism but no alterations in cholinergic markers have been found.[64,65]

Treatment and care

Early diagnosis is a prerequisite for adequate treatment and care of the patient and support for the family and other carers involved. It is essential to explain the nature of the patients' changed behaviour and current problems. It is important not to forget the children who may still be at school age. It is especially in the social interaction that the earliest signs of the brain disease emerge.[33] Alternative optimal placements must be arranged when the patient can no longer be taken care of at home because of disturbing

symptoms and lack of insight. Temporary and prolonged hospital admissions may be needed to make it possible for the family to cope with the situation. Being a spouse is both physically and emotionally exhausting, causing ill health and socio-economic problems for the family. FTD patients are often restless and stereotyped with a strong need for physical activity such as walking long distances, which, as well as the comparatively preserved memory, spatial and practical abilities should be channeled in a meaningful way rather than restricted. A well-structured programme for daily activities considering the patients premorbid personality and interests, 'routinizing therapy'[66] may be rewarding and minimize the need for pharmacological treatment. Prevailing psychotic features and unpredictable aggressive behaviour should be managed by a special psychiatric or psychogeriatric services[33] also responsible for support to the spouse and family members who often suffer from social isolation and loneliness.

There is presently no specific pharmacological treatment for the underlying degenerative disease in FTD but symptomatic treatment with serotonin-boosting antidepressants may be effective in treating some behavioural disturbances.[67] There are no reports that acetylcholinesterase inhibitors improve cognition or behaviour in FTD, and clinical experience is that the FTD patient may be extremely sensitive to psychotropic medication with disturbing side effects and paradoxical reactions. Various ways to improve progranulin levels may offer future pharmacological treatment for the FTD disorders.

Further information

Neary, D., Snowden, J., and Mann, D. (2005). Frontotemporal dementia. *Lancet Neurology*, **4**, 771–80 [review].

Alzheimer Research Forum: http://www.alzforum.org/

Referencest

1. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
2. Brun, A., Englund, E., Gustafson, L., *et al.* (1994). Consensus statement–clinical and neuropathological criteria for frontotemporal dementia. *Journal of Neurology, Neurosurgery, and Psychiatry*, **57**, 416–18.
3. Pick, A. (1892). Über die Beziehungen der senilen Hirnatrophie zur Aphasie. *Prager Medizinische Wochenschrift*, **17**, 165–7.
4. Alzheimer, A. (1911). Über eigenartige Krankheitsfälle des späteren Alters. *Zeitschrift für die Gesamte Neurologie und Psychiatrie*, **4**, 356–85.
5. Brun, A. (1987). Frontal lobe degeneration of non-Alzheimer type. I. Neuropathology. *Archives of Gerontology and Geriatrics*, **6**, 193–207.
6. Gustafson, L. (1987). Frontal lobe degeneration of non-Alzheimer type. II. Clinical picture and differential diagnosis. *Archives of Gerontology and Geriatrics*, **6**, 209–23.
7. Neary, D., Snowden, J.S., Northern, B., *et al.* (1988). Dementia of frontal lobe type. *Journal of Neurology, Neurosurgery, and Psychiatry*, **51**, 353–61.
8. Knopman, D.S., Mastri, A.R., Frey, *et al.* (1990). Dementia lacking distinctive histologic features. A common non-Alzheimer degenerative dementia. *Neurology*, **40**, 251–6.
9. Neary, D., Snowden, J.S., Gustafson, L., *et al.* (1998). Frontotemporal lobar degeneration. A consensus on clinical diagnostic criteria. *Neurology*, **51**, 1546–54.
10. Snowden, J.S., Neary, D., and Mann, D.M.A. (1996). *Frontotemporal lobar degeneration: frontotemporal dementia, progressive aphasia, semantic dementia*. Churchill Livingstone, London.
11. Mesulam, M.M. (1982). Slowly progressive aphasia without generalised dementia. *Annals of Neurology*, **11**, 592–8.
12. Hodges, J.R., Patterson, K., Oxbury, S., *et al.* (1992). Semantic dementia. Progressive fluent aphasia with temporal lobe atrophy. *Brain*, **115**, 1783–806.
13. Kertesz, A., McMonagle, P., Blair M., *et al.* (2005). The evaluation and pathology of frontotemporal dementia. *Brain*, **128**, 1996–2005.
14. Shi, J., Shaw, C.L., Du Pleiss, D., *et al.* (2005). Histopathological changes underlying frontotemporal lobar degeneration with clinicopathological correlation. *Acta Neuropathologica*, **110**, 501–12.
15. Josephs, K.A., Holton J.L., Rossor, M.N., *et al.* (2004). Frontotemporal lobar degeneration and ubiquitin immunohistochemistry. *Neuropathology and Applied Neurobiology*, **30**, 369–73.
16. Brown, J., Asworth, A., Gydesen, S., *et al.* (1995). Familial non-specific dementia maps to chromosome 3. *Human Molecular Genetics*, **4**, 1625–8.
17. Neumann, M., Sampathu, D.M., Kwong L.K., *et al.* (2006). Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science*, **314**, 130–3.
18. Larsson, E.M., Passant, U., Sundgren, P.C., *et al.* (2000). Magnetic resonance imaging and histopathology in dementia, clinically of frontotemporal type. *Dementia and Geriatric Cognitive Disorders*, **11**, 123–34.
19. Hutton, M., Lendon, C.L., Rizzu, P., *et al.* (1998). Association of missense and 5;-splice-site mutation in tau with the inherited dementia FTDP-17. *Nature*, **393**, 702–5.
20. Boeve, B.F., Lang, A.E., and Litvan, I. (2003). Corticobasal degeneration and its relationship to progressive supranuclear palsy and frontotemporal dementia. *Annals of Neurology*, **54**, S15–19.
21. Cairns, N.J., Lee, V.M., and Trojanowski, J.Q. (2007). Genetics and neuropathology of frontotemporal dementia. In *The human frontal lobes. Function and disorders* (eds. B. Miller and J.L. Cummings), pp. 382–407. The Guilford Press, NY, London.
22. Markesbery, W.R. (1998). Pick's disease. In *Neuropathology of dementing disorders* (ed. W.R. Markesbery), pp. 142–57. Arnold, London.
23. Stevens, M., Van Duijn, C.M., Kamphorts, W., *et al.* (1998). Familial aggregation in fronto-temporal dementia, *Neurology*, **50**, 1541–5.
24. Rosso, S.M., Donker Kaat, L., Baks T., *et al.* (2003). Frontotemporal dementia in the Netherlands: patient characteristics and prevalence estimates from a population-based study. *Brain*, **126**, 2016–22.
25. Pasquier, F., Lebert, F., and Amouyel, P. (1995). In *Les démences frontotemporales, épidémiologie* (eds. F. Pasquier and F. Lebert), pp. 23–9. Masson, Paris.
26. Ibach, B., Koch, H., Koller, M., *et al.* (2003). Hospital admission circumstances and prevalence of frontotemporal lobar degeneration: a multicenter psychiatric state hospital study in Germany. *Dementia and Geriatric Cognitive Disorders*, **16**, 253–64.
27. Lopez, O.L., Becker, J.T., and De Kosky, S.T. (1994). Dementia accompanying motor neuron disease. *Dementia*, **5**, 42–7.
28. Barker, W.W., Luis, C.A., Kashuba, A., *et al.* (2002). Relative frequencies of Alzheimer's disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the state of Florida Bank. *Alzheimer Disease and Associated Disorders*, **16**, 203–12.
29. Giannakopoulos, P., Hof, P.R., and Bouras, C. (1995). Dementia lacking distinctive histopathology: clinicopathological evaluation of 32 cases. *Acta Neuropathologica*, **89**, 346–55.

30. Kertesz, A., McMonagle, P., Blair, M., *et al.* (2005). The evaluation and pathology of frontotemporal dementia. *Brain*, **128**, 1996–2005.
31. Gustafson, L. (1993). Clinical picture of frontal lobe degeneration of non-Alzheimer type. *Dementia*, **4**, 143–8.
32. Passant, U., Rosén, I., Gustafson, L., *et al.* (2005a). The heterogeneity of frontotemporal dementia with regard to initial symptoms, qEEG and neuropathology. *International Journal of Geriatric Psychiatry*, **20**, 983–8.
33. Passant, U., Elfgren, C., Englund, E., *et al.* (2005b). Psychiatric symptoms and their psychosocial consequences in frontotemporal dementia. *Alzheimer Disease and Associated Disorders*, **19**, S15–18.
34. Lebert, F., Pasquier, F., and Petit, H. (1995). Personality traits and frontal lobe dementia. *International Journal of Geriatric Psychiatry*, **10**, 1046–9.
35. Cummings, J.L. and Duchen, L.W. (1981). Klüver–Bucy syndrome in Pick's disease: clinical and pathological correlations. *Neurology*, **31**, 1415–22.
36. Neary, D., Snowden, J., and Mann, D. (2005). Frontotemporal dementia. *Lancet Neurology*, **4**, 771–80.
37. Lynch, T., Sano, M., Marder, K.S., *et al.* (1994). Clinical characteristics of a family with chromosome 17-linked disinhibition–dementia–parkinsonism–amyotrophy complex. *Neurology*, **44**, 1878–84.
38. Neary, D., Snowden, J.S., and Mann, D.M.A. (1990). Frontal lobe dementia and motor neuron disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, **53**, 23–32.
39. Rosén, I., Gustafson, L., and Risberg, J. (1993). Multichannel EEG frequency analysis and somatosensory-evoked potentials in patients with different types of organic dementia. *Dementia*, **4**, 43–9.
40. Förstl, H., Besthorn, C., Hentschel, F., *et al.* (1996). Frontal lobe degeneration and Alzheimer's disease: a controlled study on clinical findings, volumetric brain changes and quantitative electroencephalography data. *Dementia*, **7**, 27–34.
41. Risberg, J., Passant, U., Warkentin, S., *et al.* (1993). Regional cerebral blood flow in frontal lobe dementia of non-Alzheimer type. *Dementia*, **4**, 186–7.
42. Foster, N.L. (2003). Validating FDG-PET as a biomarker frontotemporal dementia. *Experimental Neurology*, **184**, S2–8.
43. Salmon, E., Garaux, G., Delbeuck, X., *et al.* (2003). Predominant ventromedial frontopolar metabolic impairment in frontotemporal dementia. *Neuroimage*, **20**, 435–40.
44. Riemenschneider, M., Wagenpfeil, S., Diehl, J., *et al.* (2002). Tau and Abeta42 protein of patients with frontotemporal degeneration. *Neurology*, **11**, 1622–8.
45. Engelborghs, S., Maertens, K., Vloeberghs, E., *et al.* (2006). Neuropsychological and behavioural correlates of CSF biomarkers in dementia. *Neurochemistry International*, **48**, 286–95.
46. Pijnenburg, Y.A.L., Schoonenboom, S.N.M., Barkhof, F., *et al.* (2006). CSF biomarkers in frontotemporal lobar degeneration: relations with clinical characteristics, apolipoprotein E genotype, and neuroimaging. *Journal of Neurology, Neurosurgery, and Psychiatry*, **77**, 246–8.
47. Johanson, A. and Hagberg, B. (1989). Psychometric characteristics in patients with frontal lobe degeneration of non-Alzheimer type. *Archives of Gerontology and Geriatrics*, **8**, 29–137.
48. Frisoni, G.B., Pizzolato, G., Geroldi, C., *et al.* (1995). Dementia of the frontal type: neuropsychological and (99Tc)-HMPAO SPET features. *Journal of Geriatric Psychiatry and Neurology*, **8**, 42–8.
49. Elfgren, C., Brun, A., Gustafson, L., *et al.* (1994). Neuropsychological tests as discriminators between dementia of Alzheimer type and frontotemporal dementia. *International Journal of Geriatric Psychiatry*, **9**, 635–42.
50. Thompson, J.C., Stopford, C.L., Snowden, J.S., *et al.* (2005). Qualitative neuropsychological performance characteristics in frontotemporal dementia and Alzheimer's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, **76**, 920–7.
51. Gregory, C.A., Orrell, M., Sahakian, B., *et al.* (1997). Can frontotemporal dementia and Alzheimer's disease be differentiated using a brief battery of tests? *International Journal of Geriatric Psychiatry*, **12**, 375–83.
52. Pasquier, F. (1996). Neuropsychological features and cognitive assessment in frontotemporal dementia. In *Frontotemporal dementia* (eds. D. Pasquier, F. Lebert, and P. Scheltens), pp. 46–69. ICG, the Netherlands.
53. Miller, B.L., Ikone, C., Ponton, M., *et al.* (1997). A study of the Lund–Manchester research criteria for frontotemporal dementia: clinical and single-photon emission CT correlations. *Neurology*, **48**, 937–42.
54. Bathgate, D., Snowden, J.S., Varma, A., *et al.* (2001). Behaviour in frontotemporal dementia, Alzheimer's disease and vascular dementia. *Acta Neurologica Scandinavica*, **103**, 367–78.
55. Ikeda, M., Brown, J., Holland, A.J., *et al.* (2002). Changes in appetite, food preference, and eating habits in frontotemporal dementia and Alzheimer's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, **73**, 371–6.
56. Varma, A.R., Snowden, J.S., Lloyd, J.J., *et al.* (1999). Evaluation of the NINCDS-ADRDA criteria in the differentiation of Alzheimer's disease and frontotemporal dementia. *Journal of Neurology, Neurosurgery and Psychiatry*, **66**, 184–8.
57. World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders. Clinical descriptions and diagnostic guidelines*. World Health Organization, Geneva.
58. Gräsbeck, A., Horstmann, V., Nilsson, K., *et al.* (2005). Dementia in first-degree relatives of patients with frontotemporal dementia. *Dementia and Geriatric Cognitive Disorders*, **19**, 145–53.
59. Passant, U., Gustafson, L., and Brun, A. (1993). Spectrum of frontal lobe dementia in a Swedish Family. *Dementia*, **4**, 160–2.
60. Vance, C., Al-Chalabi, A., Ruddy, D., *et al.* (2006). Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3. *Brain*, **129**, 868–76.
61. Pickering-Brown, S.M., Siddons, M., Mann, D.M.A., *et al.* (1995). Apolipoprotein E allelic frequencies in patients with lobar atrophy. *Neuroscience Letters*, **188**, 205–7.
62. Baker, M., Mackenzie, I.R., Pickering-Brown, S.M., *et al.* (2006). Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. *Nature*, **442**, 916–9.
63. Rosso, S.M., Landweer, E.J., Houterman, M., *et al.* (2003). Medical and environmental risk factors for sporadic frontotemporal dementia: a retrospective case-control study. *Journal of Neurology, Neurosurgery, and Psychiatry*, **74**, 1574–6.
64. Francis, P.T., Holmes, C., Webster, M.-T., *et al.* (1993). Preliminary neurochemical findings in non-Alzheimer dementia due to lobar atrophy. *Dementia*, **4**, 172–7.
65. Huey, E.D., Putnam, K., and Grafman, J. (2006). A systematic review of neurotransmitter deficits and treatments in frontotemporal dementia. *Neurology*, **66**, 17–22.
66. Tanabe, H., Ikeda, M., and Komori, K. (1999). Behavioral symptomatology and care of patients with frontotemporal lobe degeneration—based on the aspects of the phylogenetic and ontogenetic processes. *Dementia and Geriatric Cognitive Disorders*, **10**, S50–4.
67. Lebert, F., Stekke, W., Hasenbroekx, C., *et al.* (2004). Frontotemporal dementia: a randomised, controlled trial with trazodone. *Dementia and Geriatric Cognitive Disorders*, **17**, 355–9.

4.1.4 Prion disease

John Collinge

Introduction

The human prion diseases, also known as the subacute spongiform encephalopathies, have been traditionally classified into Creutzfeldt–Jakob disease (**CJD**), Gerstmann–Sträussler syndrome (**GSS**) (also known as Gerstmann–Sträussler–Scheinker disease), and **kuru**. Although rare, affecting about 1–2 per million worldwide per annum, remarkable attention has been recently focused on these diseases. This is because of the unique biology of the transmissible agent or prion, and also because bovine spongiform encephalopathy (**BSE**), an epidemic bovine prion disease, appears to have transmitted to humans as variant CJD (**vCJD**), opening the possibility of a significant threat to public health through dietary exposure to infected tissues.

The transmissibility of the human diseases was demonstrated with the transmission, by intracerebral inoculation with brain homogenates into chimpanzees, of first kuru and then CJD in 1966 and 1968, respectively.[1,2] Transmission of GSS followed in 1981. The prototypic prion disease is **scrapie**, a naturally occurring disease of sheep and goats, which has been recognized in Europe for over 200 years and which is present in the sheep flocks of many countries. Scrapie was demonstrated to be transmissible by inoculation in 1936[3] and the recognition that kuru, and then CJD, resembled scrapie in its histopathological appearances led to the suggestion that these diseases may also be transmissible.[4] Kuru reached epidemic proportions amongst the Fore linguistic group in the Eastern Highlands of Papua New Guinea and was transmitted by ritual cannibalism. Since the cessation of cannibalism in the 1950s the disease has declined but a few cases still occur as a result of the long incubation periods in this condition, which may exceed 50 years.[5] The term Creutzfeldt–Jakob disease was introduced by Spielmeyer in 1922 bringing together the case reports published by Creutzfeldt and Jakob. Several of these cases would not meet modern diagnostic criteria for CJD and indeed it was not until the demonstration of transmissibility allowed diagnostic criteria to be reassessed and refined that a clear diagnostic entity developed. All these diseases share common histopathological features; the classical triad of spongiform vacuolation (affecting any part of the cerebral grey matter), astrocytic proliferation, and neuronal loss, may be accompanied by the deposition of amyloid plaques.

Aetiology

Prion diseases of both humans and animals are associated with the accumulation in the brain of an abnormal, partially protease-resistant, isoform of a host-encoded glycoprotein known as prion protein (PrP). The disease-related isoform, PrP^{Sc}, is derived from its normal cellular precursor, PrP^{C}, by a post-translational process that involves a conformational change. PrP^{C} is rich in α-helical structure while PrP^{Sc} appears to be predominantly composed of β-sheet structure. According to the 'protein-only' hypothesis,[6] an abnormal PrP isoform[7] is the principal, and possibly the sole, constituent of the transmissible agent or prion. PrP^{Sc} is hypothesized to act as a conformational template, promoting the conversion of PrP^{C} to further PrP^{Sc}. PrP^{C} appears to be poised between two radically different folding states, and α- and β-forms of PrP can be inter-converted in suitable conditions.[8] Soluble β-PrP aggregates in physiological salt concentrations to form fibrils with morphological and biochemical characteristics closely similar to PrP^{Sc}. A molecular mechanism for prion propagation can now be proposed.[8] Prion replication, with recruitment of PrP^{C} into the aggregated PrP^{Sc} isoform, may be initiated by a pathogenic mutation (resulting in a PrP^{C} predisposed to form β-PrP) in inherited prion diseases, by exposure to a 'seed' of PrP^{Sc} in acquired cases, or as a result of the spontaneous conversion of PrP^{C} to β-PrP (and subsequent formation of aggregated material) as a rare stochastic event in sporadic prion disease.

The human PrP gene (*PRNP*) is a single copy gene located on the short arm of chromosome 20 and was an obvious candidate for genetic linkage studies in the familial forms of CJD and GSS, which both showed an autosomal dominant pattern of disease segregation. A turning point in understanding the human prion diseases was the identification of mutations in the prion protein *gene* in familial CJD and GSS in 1989. The first mutation to be identified in *PRNP* was in a family with CJD and constituted a 144 bp insertion into the coding sequence.[9] A second mutation was reported in two families with GSS and genetic linkage was confirmed between this missense variant at codon 102 and GSS, confirming that GSS was an autosomal dominant Mendelian disorder.[10] Uniquely, these diseases are therefore both inherited and transmissible. Current evidence suggests that around 15 per cent of prion diseases are inherited and over 30 coding mutations in *PRNP* are now recognized.

With the exception of the rare iatrogenic CJD cases mentioned above, most prion disease occurs as sporadic CJD. While, by definition, there will not be a family history in sporadic cases, mutations are seen in occasional apparently sporadic cases, as with a late-onset disease the family history may not be apparent or non-paternity may occur. However, in the majority of sporadic CJD cases there is neither a coding mutation nor a history of iatrogenic exposure. Human prion diseases can therefore be subdivided into inherited, sporadic, and acquired forms. However, a common PrP polymorphism at residue 129, where either methionine or valine can be encoded, is a key determinant of genetic susceptibility to acquired and sporadic prion diseases, the large majority of which occur in homozygous individuals.[11,12] This protective effect of *PRNP* codon 129 heterozygosity is also seen in some of the inherited prion diseases.

The aetiology of sporadic CJD remains unclear. It has been speculated that these cases might arise from somatic mutation of *PRNP* or spontaneous conversion of PrP^{C} to PrP^{Sc} as a rare stochastic event. The alternative hypothesis, in which such cases arise as a result of exposure to an environmental source of either human or animal prions, is not supported by epidemiological evidence.[13]

A major problem for the 'protein-only' hypothesis of prion propagation has been how to explain the existence of multiple isolates or strains of prions, with distinct biological properties. Understanding how a protein-only infectious agent could encode such phenotypic information has been of considerable biological interest. However, it is now clear that prion strains can be distinguished by differences in the biochemical properties of PrP^{Sc}. Prion strain diversity appears to be encoded by differences in PrP conformation and pattern of glycosylation.[14] A molecular strain

typing approach based on these characteristics has allowed the identification of four main types amongst CJD cases, sporadic and iatrogenic CJD being of PrP^{Sc} types 1–3, while all vCJD cases are associated with a distinctive type 4 PrP^{Sc} type.[14,15] A similar PrP^{Sc} type to that seen in vCJD is seen in BSE and BSE when transmitted to several other species. Such molecular strain typing strongly supported the hypothesis that vCJD was human BSE. This conclusion was strengthened by subsequent transmission studies of vCJD into both transgenic and conventional mice which argued that cattle BSE and vCJD were caused by the same strain.[16,17] Such studies are allowing a molecular classification of human prion diseases. Two such classifications are in use: no internationally agreed classification has yet emerged and it is likely that additional PrP^{Sc} types or strains will be identified.[15,18] Molecular classification may well open new avenues of epidemiological investigation and offer insights into causes of 'sporadic' CJD. The ability of a protein to encode a disease phenotype has important implications in biology, as it represents a non-Mendelian form of transmission. It would be surprising if this mechanism had not been used more widely during evolution such that prion biology may prove to be of far wider relevance.

Transmission of prion diseases between different mammalian species is limited by a so-called 'species barrier'. Early studies of the molecular basis of the species barrier argued that it principally resided in differences in PrP primary structure between the species from which the inoculum was derived and the inoculated host. Transgenic mice expressing hamster PrP were, unlike wild-type mice, highly susceptible to infection with hamster prions.[19] That most sporadic and acquired CJD occurred in individuals homozygous at *PRNP* polymorphic codon 129 supported the view that prion propagation proceeded most efficiently when the interacting PrP^{Sc} and PrP^{C} were of identical primary structure.[12] However, it has been long recognized that prion strain type affects ease of transmission to another species. Interestingly, with BSE prions the strain component to the barrier seems to predominate, with BSE not only transmitting efficiently to a range of species, but maintaining its transmission characteristics even when passaged through an intermediate species with a distinct PrP gene.[20] The term 'species-strain barrier' or simply 'transmission barrier' may be preferable.[21] Both PrP amino acid sequence and strain type affect the 3D structure of glycosylated PrP which will presumably, in turn, affect the efficiency of the protein–protein interactions thought to determine prion propagation.

Mammalian PrP genes are highly conserved. Presumably only a restricted number of different PrP^{Sc} conformations (that are highly stable and can therefore be serially propagated) will be permissible thermodynamically and will constitute the range of prion strains seen in mammals. While a significant number of different such PrP^{Sc} conformations may be possible amongst the range of mammalian PrPs, only a subset of these would be allowable for a given single mammalian PrP. Substantial overlap between the favoured conformations for PrP^{Sc} derived from species A and species B might therefore result in relatively easy transmission of prion diseases between these two species, while two species with no preferred PrP^{Sc} conformations in common would have a large barrier to transmission (and indeed transmission would necessitate a change of strain type). According to such a *conformational selection model*[21] of a prion transmission barrier, BSE may represent a thermodynamically highly favoured PrP^{Sc} conformation that is permissive for PrP expressed in a wide range of different species, accounting for the remarkable promiscuity of this strain in mammals. Contribution of other components to the species barrier are possible and may involve interacting co-factors which mediate the efficiency of prion propagation, although no such factors have yet been identified.

Additional data has further challenged our understanding of transmission barriers.[22] The assessment of species barriers has relied on the development of a clinical disease in inoculated animals. However, it is now clear that *subclinical prion infections* are sometimes established on prion inoculation of a second species.[23] Such animals harbour high levels of prion infectivity but do not develop clinical disease during a normal lifespan. The existence of such subclinical carrier states of prion infection has important potential animal and public health implications and argues against direct neurotoxicity of prions.

The transmission barrier between cattle BSE and humans cannot be directly measured but can be modelled in transgenic mice expressing human PrP^{C}, which produce human PrP^{Sc} when challenged with human prions. Long-term transmission studies have been carried out using such 'humanized' mice to both to characterize the distinct prion strains causing human disease and to model human susceptibility to infection with BSE and other prions.[24] While these transgenic mouse models have been able to faithfully propagate human prion strains[14,16,25] and recapitulate the characteristic neuropathology of vCJD,[26] there are important caveats in extrapolating from such animal models to human susceptibility. However, these studies have found a much higher infection rate in transgenic mice expressing human PrP M129 than mice expressing human PrP V129 when challenged with either BSE or vCJD prions, and demonstrated that BSE prion infection can produce disease phenotypes resembling sporadic CJD infection of these mice and also novel prion strain phenotypes. Most recently, these studies have argued that the vCJD phenotype may only be expressed in the presence of the M form of human PrP.[27] While this would imply that only those humans expressing human PrP M129 may develop the vCJD phenotype, this does not mean that VV individuals are completely resistant to BSE prion infection—but rather that if infected they would show a different phenotype.[27] Modelling of susceptibility of the MV genotype suggests that several different phenotypes, all distinct from vCJD, may be possible when infected with BSE or vCJD prions.[28]

Clinical features and diagnosis

The human prion diseases can be divided aetiologically into inherited, sporadic, and acquired forms with CJD, GSS, and kuru now seen as clinicopathological syndromes within a wider spectrum of disease. Kindreds with inherited prion disease have been described with phenotypes of classical CJD, GSS, and also with other neurodegenerative syndromes including fatal familial insomnia. Some kindreds show remarkable phenotypic variability which can encompass both CJD- and GSS-like cases as well as other cases which do not conform to either CJD or GSS phenotypes and which indeed readily mimic, and are frequently misdiagnosed as, many other neurodegenerative conditions. Inherited prion diseases are a frequent cause of pre-senile dementia and a family history is not always apparent: *PRNP* should be analysed in all suspected cases of CJD, and considered in all early-onset dementia and ataxias. Cases diagnosed by *PRNP* analysis have been reported which are not only

clinically atypical but which lack the classical histological features entirely. Significant clinical overlap exists with familial Alzheimer's disease, Pick's disease, frontal lobe degeneration of non-Alzheimer type, and amyotrophic lateral sclerosis with dementia. Although classical GSS is described below it now seems more sensible to designate the familial illnesses as inherited prion diseases and then to subclassify these according to mutation. Acquired prion diseases include iatrogenic CJD, kuru, and now vCJD. Sporadic prion diseases at present consist of CJD and atypical variants of CJD. Cases lacking the characteristic histological features of CJD have been transmitted. As there are at present no equivalent aetiological diagnostic markers for sporadic prion diseases to those for the inherited diseases, it cannot yet be excluded that more diverse phenotypic variants of sporadic prion disease exist. The key clinical features and investigations for the diagnosis of prion disease are given in the Table 4.1.4.1.

Table 4.1.4.1 Diagnosis of prion disease

Sporadic (classical) CJD
- Rapidly progressive* dementia with two or more of myoclonus, cortical blindness, pyramidal signs, cerebellar signs, extrapyramidal signs, akinetic mutism
- Most cases age 45–75
- Serial EEG shows pseudoperiodic complexes in most cases
- CSF 14-3-3 protein usually positive
- CT normal or atrophy, MRI may show high signal in the striatum and/or cerebral cortex in FLAIR or diffusion-weighted images
- *PRNP* analysis: no pathogenic mutations, most are 129 MM (VV and MV may be longer duration, clinically atypical and EEG less often positive)
- Brain biopsy in highly selected cases (to exclude treatable alternative diagnoses): PrP immunocytochemistry or Western blot for PrP^{Sc} types 1–3

Iatrogenic CJD
- Progressive cerebellar syndrome and behavioural disturbance or classical CJD-like syndrome with history of iatrogenic exposure to human prions (pituitary-derived hormones, tissue grafting, or neurosurgery)
- May be young
- EEG, CSF, and MRI generally less helpful than in sporadic cases
- *PRNP* analysis: no pathogenic mutations, most are 129 homozygotes
- Brain biopsy in highly selected cases (to exclude treatable alternative diagnoses): PrP immunocytochemistry or Western blot for PrP^{Sc} types 1–3

Variant CJD
- Early features: depression, anxiety, social withdrawal, peripheral sensory symptoms
- Cerebellar ataxia, chorea, or athetosis often precedes dementia, advanced disease as sporadic CJD
- Most in young adults; however, age at onset 12–74 years seen
- EEG non-specific slow waves, CSF 14-3-3 may be elevated or normal
- MRI: pulvinar sign usually present (particularly using FLAIR sequence) but may be late feature
- *PRNP* analysis: no mutations, all 129 MM to date
- Tonsil biopsy: characteristic PrP immunostaining and PrP^{Sc} on Western blot (type 4τ)

Iatrogenic vCJD
- Has occurred in recipients of blood transfusion from a donor who subsequently developed clinical vCJD
- Known recipients of implicated blood or blood products in the UK have been notified of their risk status
- Clinical features and investigations as for primary vCJD

Inherited prion disease
- Varied clinical syndromes between and within kindreds: should consider in all pre-senile dementias and ataxias irrespective of family history
- *PRNP* analysis: diagnostic, codon 129 genotype may predict age at onset in pre-symptomatic testing

*Clinical duration typically 6 months or less but high variability: type 1 PrP^{Sc} associated with short duration (~8 weeks); ~10% have duration >2 years.

Sporadic prion disease

CJD

The core clinical syndrome of classic CJD is of a rapidly progressive multifocal dementia usually with myoclonus. The onset is usually in the 45–75-year age group with peak onset between 60 and 65 years. The clinical progression is typically over weeks progressing to akinetic mutism and death often in 2–3 months. Around 70 per cent of cases die in under 6 months. Prodromal features, present in around a third of cases, include fatigue, insomnia, depression, weight loss, headaches, general malaise, and ill-defined pain sensations. In addition to mental deterioration and myoclonus, frequent additional neurological features include extrapyramidal signs, cerebellar ataxia, pyramidal signs, and cortical blindness. About 10 per cent of cases present initially with cerebellar ataxia.

Routine haematological and biochemical investigations are normal although occasional cases have been noted to have raised serum transaminases or alkaline phosphatase. There are no immunological markers and acute phase proteins are not elevated. Examination of the cerebrospinal fluid is normal 14-3-3 protein is usually elevated in CJD and is a useful adjunct to diagnosis in the appropriate clinical context.[29] It is also positive in recent cerebral infarction or haemorrhage and in viral encephalitis, although these conditions do not usually present diagnostic confusion with CJD. It may also be elevated in rapidly progressive Alzheimer's disease, which may be difficult to clinically distinguish from CJD. Neuronal specific enolase (NSE) and S-100b may be also elevated although also are not specific for CJD and represent markers of neuronal injury[30,31] Neuroimaging with CT or MRI is crucial to exclude other causes of subacute neurological illness but MRI has become increasingly useful in diagnosis of sporadic CJD, showing high signal in the striatum and/or cerebral cortex in FLAIR or diffusion-weighted images.[32] Cerebral and cerebellar atrophy may be present in longer duration cases. The electroencephalogram (EEG) may show characteristic pseudoperiodic sharp wave activity, which is very helpful in diagnosis but present only in around 70 per cent of cases. To some extent demonstration of a typical EEG is dependent on the number of EEGs performed and serial EEG is indicated to try and demonstrate this appearance.

Prospective epidemiological studies have demonstrated that cases with a progressive dementia, and two or more of the following: myoclonus; cortical blindness; pyramidal, cerebellar, or extrapyramidal signs; or akinetic mutism in the setting of a typical EEG nearly always turn out to be confirmed as histologically definite CJD if neuropathological examination is performed.

Neuropathological confirmation of CJD is by demonstration of spongiform change, neuronal loss, and astrocytosis. PrP amyloid plaques are usually not present in CJD although PrP immunohistochemistry, using appropriate pre-treatments, will nearly always be positive. Protease resistant PrP, seen in all the currently recognized

prion diseases, can be demonstrated by immunoblotting of brain homogenates. *PRNP* analysis is important to exclude pathogenic mutations. Genetic susceptibility to CJD has been demonstrated in that most cases of classical CJD are homozygous with respect to the common 129 polymorphism of PrP (see aetiology).

Atypical forms of CJD

Atypical forms of CJD are well recognized. Around 10 per cent of cases of CJD have a much more prolonged clinical course with a disease duration of over 2 years. These cases may represent the occasional occurrence of CJD in individuals heterozygous for PrP polymorphisms. Around 10 per cent of CJD cases present with cerebellar ataxia rather than cognitive impairment, so-called ataxic CJD. Heidenhain's variant of CJD refers to cases in which cortical blindness predominates with severe involvement of the occipital lobes. The panencephalopathic type of CJD refers to cases with extensive degeneration of the cerebral white matter in addition to spongiform vacuolation of the grey matter and has been predominately reported from Japan.

Amyotrophic variants of CJD have been described with prominent early muscle wasting. However, most cases of dementia with amyotrophy are not experimentally transmissible and their relationship with CJD is unclear. Most cases are probably variants of motor neurone disease with associated dementia. Amyotrophic features in CJD are usually seen in late disease when other features are well established.

Acquired prion diseases

While human prion diseases can be transmitted to experimental animals by inoculation, they are not contagious in humans. Documented case to case spread has only occurred during ritual cannibalistic practices (kuru) or following accidental inoculation with prions during medical or surgical procedures (iatrogenic CJD).

Kuru

Kuru reached epidemic proportions amongst a defined population living in the Eastern Highlands of Papua New Guinea.[33] The earliest cases are thought to date back to the early part of the century. Kuru affected the people of the Fore linguistic group and their neighbours with whom they intermarried. Kuru predominantly affected women and children (of both sexes), with only 2 per cent of cases in adult males and was the commonest cause of death amongst women in affected villages. It was the practice in these communities to engage in consumption of dead relatives as a mark of respect and mourning. Women and children predominantly ate the brain and internal organs, which is thought to explain the differential age and sex incidence. Preparation of the cadaver for consumption was performed by the women and children such that other routes of exposure may also have been relevant. It is thought that the epidemic related to a single sporadic CJD case occurring in the region some decades earlier. Epidemiological studies provided no evidence for vertical transmission, since most of the children born after 1956 (when cannibalism had effectively ceased) and all of those born after 1959 of mothers affected with or incubating kuru were unaffected. From the age of the youngest affected patient, the shortest incubation period is estimated as 4.5 years, although may have been shorter, since time of infection was usually unknown. The disease has gradually declined in incidence although a small number of cases have been documented in recent years with incubation periods which may exceed 50 years.[5]

Kuru affects both sexes and onset of disease has ranged from age 5 to over 60. The mean clinical duration of illness is 12 months with a range of 3 months to 3 years; the course tends to be shorter in children. The central clinical feature is progressive cerebellar ataxia. In sharp contrast to CJD, dementia is usually absent, even in the latter stages, although in the terminal stages many patients have their faculties obtunded. The occasional case in which gross dementia occurs is in marked contrast to the clinical norm. Detailed clinical descriptions have been given by a number of observers and the disease does not appear to have changed in features at different stages of the epidemic. A prodrome and three clinical stages are recognized:

(a) Prodromal stage

Kuru typically begins with prodromal symptoms consisting of headache, aching of limbs, and joint pains, which can last for several months.

(b) Ambulatory stage

Kuru was frequently self-diagnosed by patients at the earliest onset of unsteadiness in standing or walking, or of dysarthria or diplopia. At this stage there may be no objective signs of disease. Gait ataxia however worsens and patients develop a broad-based gait, truncal instability, and titubation. A coarse postural tremor is usually present and accentuated by movement; patients characteristically hold their hands together in the midline to suppress this. Standing with feet together reveals clawing of toes to maintain posture. This marked clawing response is regarded as pathognomonic of kuru. Patients often become withdrawn at this stage and occasionally develop a severe reactive depression. Prodromal symptoms tend to disappear. Astasia and gait ataxia worsen and the patient requires a stick for walking. Intention tremor, dysmetria, hypotonia, and dysdiadochokinesis develop. Although eye movements are ataxic and jerky, nystagmus is rarely seen. Strabismus, usually convergent, may occur particularly in children. This strabismus does not appear to be concomitant or paralytic and may fluctuate in both extent and type sometimes disappearing later in the clinical course. Photophobia is common and there may be an abnormal cold sensitivity with shivering and piloerection even in a warm environment. Tendon reflexes are reduced or normal and plantar responses are flexor. Dysarthria usually occurs. As ataxia progresses the patient passes from the first (ambulatory) stage to the second (sedentary) stage. The mean clinical duration of the first stage is around 8 months and correlates closely with total duration.

(c) Sedentary stage

At this stage patients are able to sit unsupported but cannot walk. Attempted walking with support leads to a high steppage, wide-based gait with reeling instability, and flinging arm movements in an attempt to maintain posture. Hyperreflexia is seen although plantar responses usually remain flexor with intact abdominal reflexes. Clonus is characteristically short-lived. Athetoid and choreiform movements and Parkinsonian tremors may occur. There is no paralysis, although muscle power is reduced. Obesity is common at this stage but may be present in early disease associated with bulimia. Characteristically, there is emotional lability and bizarre uncontrollable laughter, which has led to the disease being

referred to as 'laughing death'. There is no sensory impairment. In sharp contrast to CJD, myoclonic jerking is rarely seen. EEG is usually normal or may show non-specific changes. This stage lasts around 2–3 months. When truncal ataxia reaches the point where the patient is unable to sit unsupported, the third or tertiary stage is reached.

(d) Tertiary stage

Hypotonia and hyporeflexia develop and the terminal state is marked by flaccid muscle weakness. Plantar responses remain flexor and abdominal reflexes intact. Progressive dysphagia occurs and patients become incontinent of urine and faeces. Inanition and emaciation develop. Transient conjugate eye signs and dementia may occur. Primitive reflexes develop in occasional cases. Brainstem involvement and both bulbar and pseudobulbar signs occur. Respiratory failure and bronchopneumonia eventually lead to death. The tertiary stage lasts 1–2 months.

Iatrogenic CJD

Iatrogenic transmission of CJD has occurred by accidental inoculation with human prions as a result of medical procedures. Such iatrogenic routes include the use of inadequately sterilized neurosurgical instruments, dura mater and corneal grafting, and use of human cadaveric pituitary-derived growth hormone or gonadotrophin. It is of considerable interest that cases arising from intracerebral or optic inoculation manifest clinically as classical CJD, with a rapidly progressive dementia, while those resulting from peripheral inoculation, most notably following pituitary-derived growth hormone exposure, typically present with a progressive cerebellar syndrome, and are in that respect somewhat reminiscent of kuru. Unsurprisingly the incubation period in intracerebral cases is short (19–46 months for dura mater grafts) as compared to peripheral cases (typically 15 years or more). There is evidence for genetic susceptibility to iatrogenic CJD with an excess of codon 129 homozygotes[11] (see aetiology).

Epidemiological studies have not shown increased risks of particular occupations that may be exposed to human or animal prions, although individual CJD cases in two histopathology technicians, a neuropathologist, and a neurosurgeon have been documented. While there have been concerns that CJD may be transmissible by blood transfusion, extensive epidemiological analysis in the UK has found that the frequency of blood transfusion and donation was no different in over 200 cases of CJD and a matched control population.[34] Recipients of blood transfusions who developed CJD had clinical presentations similar to those of sporadic CJD patients and not to the more kuru-like iatrogenic cases arising from peripheral exposure to human prions. Furthermore, experimental transmission studies have shown only weak evidence for infectivity in blood, even when inoculated via the most efficient (intracerebral) route. Iatrogenic (secondary) vCJD related to blood transfusion has however been recognized (see below).

Variant CJD

In late 1995, two cases of sporadic CJD were reported in the UK in teenagers.[35] Only four cases of sporadic CJD had previously been recorded in teenagers, and none of these cases occurred in the UK. In addition, both cases were unusual in having kuru-type plaques, a finding seen in only around 5 per cent of CJD cases. Soon afterwards a third very young sporadic CJD case occurred. These cases caused considerable concern and the possibility was raised that they might suggest a link with BSE. By March 1996, further extremely young onset cases were apparent and review of the histology of these cases showed a remarkably consistent and unique pattern. These cases were named 'new variant' CJD although it was clear that they were also rather atypical in their clinical presentation; in fact most cases did not meet the accepted clinical diagnostic criteria for probable CJD. Extensive studies of archival cases of CJD or other prion diseases failed to show this picture and it seemed that it did represent the arrival of a new form of prion disease in the UK. The statistical probability of such cases occurring by chance was vanishingly small and ascertainment bias seemed most unlikely as an explanation. It was clear that a new risk factor for CJD had emerged and appeared to be specific to the UK. The UK Government advisory committee on spongiform encephalopathy (SEAC) concluded that, while there was no direct evidence for a link with BSE, exposure to specified bovine offal (SBO) prior to the ban on its inclusion in human foodstuffs in 1989, was the most likely explanation. A case of vCJD was soon after reported in France. Direct experimental evidence that vCJD is caused by BSE was provided by molecular analysis of human prion strains and transmission studies in transgenic and wild-type mice (see aetiology). While it is now clear that vCJD is caused by infection with BSE prions, it is unclear why this particular age group should be affected and why none of these cases had a pattern of unusual occupational or dietary exposure to BSE. However, very little is known of which foodstuffs contained high-titre bovine offal. It is possible that certain foods containing particularly high titres were eaten predominately by younger people. An alternative is that young people are more susceptible to BSE following dietary exposure or that they have shorter incubation periods. A possible age-related co-factor could be coexistent infection involving lymphoid tissue, for example tonsillar infection. It is important to appreciate that BSE contaminated feed was fed to sheep, pigs, and poultry and that although there is no evidence of natural transmission to these species, it would be prudent to remain open minded about other dietary exposure to novel animal prions.

vCJD has an insidious clinical onset and its early features are highly non-specific. The clinical presentation is often with behavioural and psychiatric disturbances and in some cases with sensory disturbance. Initial referral has frequently been to a psychiatrist and the most prominent feature is depression but anxiety, social withdrawal, and behavioural change is frequent. Suicidal ideation is infrequent and response to antidepressants poor. Delusions, which are complex and unsustained, are common. Other features include emotional lability, aggression, insomnia, and auditory and visual hallucinations. A prominent early feature in some is dysaesthesiae or pain in the limbs or face or pain, which is persistent rather than intermittent and unrelated to anxiety levels. A minority of cases have been noted to have forgetfulness or mild gait ataxia from an early stage but in most cases overt neurological features are not apparent until some months into the clinical course. In most patients a progressive cerebellar syndrome develops with gait and limb ataxia. Overt dementia then occurs with inevitable progression to akinetic mutism. Myoclonus is seen in most patients, and chorea is often present which may be severe in some patients. Cortical blindness develops in a minority of patients in the late

stages of disease. Upgaze paresis, an uncommon feature of classical CJD, has been noted in some patients. The age at onset in the initial 14 cases reported ranged from 16 to 48 years (mean 29 years) and the clinical course was unusually prolonged (9–35 months, median 14 months). The age range of cases has since broadened, with ages at onset ranging from 12 to 74 years, although the mean remains around 28 years. The EEG is abnormal, most frequently showing generalized slow wave activity, but without the pseudoperiodic pattern seen in most sporadic CJD cases. Neuroimaging by CT is either normal or shows only mild atrophy. The most useful non-invasive investigation in advanced cases is MR neuroimaging, particularly the FLAIR sequence.[36] Early case reports noted bilateral increased signal in the posterior thalamus (pulvinar) on T2-weighted images.[37] A retrospective review of 36 histologically confirmed cases of vCJD suggested that the 'pulvinar sign' occurred frequently in advanced cases of vCJD[38] with a sensitivity and specificity of up to 86 and 96 per cent, respectively. However, this sign appears a late feature of the disease process. Histologically confirmed cases of vCJD with minimal or absent pulvinar changes at a mean 10.5 months during an illness of mean 15 months duration were identified in this series. Figures of 81 per cent sensitivity and 94 per cent specificity have also been reported in a series including 27 cases of vCJD diagnosed by tonsil biopsy.[39] As these studies suggest, the pulvinar sign is not specific for vCJD. These MRI appearances are described in sporadic CJD and paraneoplastic limbic encephalitis, both of which are important considerations in the differential diagnosis of patients with suspected vCJD. Pulvinar signal change on MRI is also reported in a number of rare conditions, which might otherwise be distinguished from vCJD on clinical grounds such as benign intracranial hypertension, status epilepticus associated with cat scratch disease, Alpers' disease, and post-infectious encephalitis. The absence of pulvinar sign does not exclude a diagnosis of vCJD.

Tonsillar biopsy remains the most sensitive and specific diagnostic procedure for vCJD.[39–43] Tonsillar PrP^{Sc} is uniformly present in clinically affected cases of vCJD but not in other forms of human prion disease, including iatrogenic CJD associated with use of human cadaveric-derived pituitary hormones, arguing that this distinctive pathogenesis relates to effect of prion strain rather than to a peripheral route of infection.[41–43] As infection of lymphoreticular tissues is thought to precede neuroinvasion, and indeed has been detected in archived surgical samples removed prior to development of vCJD,[44,45] it is likely to allow firm diagnosis at the early clinical stage or indeed pre-clinically.[46] The PrP^{Sc} type detected on Western blot in vCJD tonsil has a characteristic pattern designated type 4t. A positive tonsil biopsy obviates the need for brain biopsy, which may otherwise be considered in such a clinical context to exclude alternative, potentially treatable diagnoses. CSF 14-3-3 protein may be elevated or normal. *PRNP* analysis is essential to rule out pathogenic mutations, as the inherited prion diseases present in younger patients and may clinically mimic vCJD. It is particularly important to exclude mutations prior to tonsil biopsy. Remarkably, to date all clinical cases of vCJD have been of the *PRNP* codon 129 MM genotype (see aetiology).

The neuropathological appearances of vCJD are striking and relatively consistent, generally allowing differentiation from other forms of prion disease. While there is widespread spongiform change, gliosis and neuronal loss, most severe in the basal ganglia and thalamus, the most remarkable feature is abundant PrP amyloid plaques in cerebral and cerebellar cortex. These consist of kuru-like, 'florid' (surrounded by spongiform vacuoles) and multicentric plaque types. The 'florid' plaques, seen previously only in scrapie, are a consistent feature. There is also abundant pericellular PrP deposition in the cerebral and cerebellar cortex. A further unusual feature is extensive PrP deposition in the molecular layer of the cerebellum. Western blot analysis (molecular strain typing, see aetiology) of brain tissue demonstrates PrP^{Sc} type 4, which is pathognomonic of vCJD.

Some of the features of vCJD are reminiscent of kuru, in which behavioural changes and progressive ataxia predominate. In addition, peripheral sensory disturbances are well recognized in the kuru prodrome. Kuru plaques are seen in around 70 per cent of cases and are especially abundant in younger kuru cases. The observation that iatrogenic prion disease related to peripheral exposure to human prions has a more kuru-like than CJD-like clinical picture may well be relevant and would be consistent with a peripheral prion exposure.

The relatively stereotyped clinical presentation and neuropathology of vCJD contrasts sharply with sporadic CJD. This may be because vCJD is caused by a single prion strain and may also suggest that a relatively homogeneous genetically susceptible subgroup of the population with short incubation periods to BSE has been selected to date.

Secondary (iatrogenic) vCJD

The prominent lymphoreticular involvement raised early concerns that vCJD may be transmissible by blood transfusion. Indeed the tissue distribution is similar to that of ovine scrapie where prionaemia has been demonstrated experimentally. In 2004, two transfusion-associated cases of vCJD prion infection were reported amongst a small cohort of patients identified as having received blood from a donor who subsequently developed vCJD. One patient had a typical clinical course of vCJD although the diagnosis was not made until autopsy, and had the *PRNP* codon 129 MM genotype. The second, who died of an unrelated condition, was found to have prion infection at autopsy. This patient had the *PRNP* codon 129 MV genotype which is associated with relative resistance to prion disease. Subsequently two further patients have been diagnosed with vCJD during life from this group of 23 known surviving recipients of implicated blood. That 4/23 patients have been infected, three dying of vCJD, in each case following transfusion with a single unit of implicated red cells, suggests the risk to recipients of blood from a silently infected donor is very substantial. The incubation period in the clinical cases was 6–7 years. Since 2003, all known recipients of implicated blood have been notified of their status. Over 6000 individuals in the UK have been exposed to blood products prepared from large donor pools containing blood from a donor who went on to develop vCJD. None of these individuals, predominantly haemophiliacs, have yet developed vCJD.

Inherited prion diseases

Gerstmann–Sträussler–Scheinker disease

The first case was described by Gerstmann in 1928 and was followed by a more detailed report on seven other affected members of the same family in 1936. The classical presentation of GSS is with a chronic cerebellar ataxia accompanied by pyramidal features, with dementia occurring later in a much more prolonged clinical course than that seen in CJD. The mean duration is around 5 years, with

onset usually in either the third or fourth decades. Histologically, the hallmark is the presence of multicentric amyloid plaques. Spongiform change, neuronal loss, astrocytosis, and white matter loss are also usually present. Numerous GSS kindreds from several countries (including the original Austrian family described by Gerstmann, Sträussler, and Scheinker in 1936) have now been demonstrated to have mutations in the PrP gene. GSS is an autosomal dominant disorder which can now be classified within the spectrum of inherited prion disease.

Inherited prion diseases

The identification of one of the pathogenic PrP gene mutations in a case with neurodegenerative disease allows not only molecular diagnosis of an inherited prion disease but also its subclassification according to mutation (see Fig. 4.1.4.1). Over 30 pathogenic mutations are reported in the human PrP gene and consist of two groups: (1) point mutations within the coding sequence resulting in amino acid substitutions in PrP or production of a stop codon resulting in expression of a truncated PrP; (2) insertions encoding additional integral copies of an octapeptide repeat present in a tandem array of five copies in the normal protein (octapeptide repeat insertion [OPRI]). A suggested notation for these diseases is 'Inherited prion disease (PrP mutation)', for instance: Inherited prion disease (PrP 6 OPRI) or Inherited prion disease (PrP P102L). Brief details of the more commonly seen types are given below, for a more comprehensive review see Ref.[47] *PRNP* analysis should be considered in all early-onset dementing or ataxic disorders and is available from the UK National Prion Clinic (see websites).

(a) PrP P102L

This mutation was first reported in 1989 in a UK and US family and has now been demonstrated in many other kindreds worldwide. Progressive ataxia is the dominant clinical feature, with dementia and pyramidal features. However, marked variability both at the clinical and neuropathological level is apparent in some families. A family with marked amyotrophic features has also been reported and cases with severe dementia in the absence of prominent ataxia are also recognized.

(b) PrP A117V

This mutation has been described in families from France, United States, and UK. The clinical features are pre-senile dementia associated with pyramidal signs, parkinsonism, pseudobulbar features, and cerebellar signs. Parkinsonian features may predominate in the early stages and mimic Parkinson's disease.

(c) PrP D178N

This mutation was originally described in two Finnish families with a CJD-like phenotype and has since been demonstrated in families in Hungary, the Netherlands, Canada, Finland, France, and the UK. This mutation was also reported in two unrelated families with fatal familial insomnia (FFI).[48] The first case described had a rapidly progressive disease characterized clinically by untreatable insomnia, dysautonomia and motor signs, and neuropathologically by selective atrophy of the anterior-ventral and medio-dorsal thalamic nuclei. Proteinase K treatment of extracted PrP^{Sc} from FFI cases has shown a different sized PrP band on Western blots than PrP^{Sc} from CJD cases suggesting that FFI may be caused by a distinct prion strain type. Goldfarb *et al.*[49] reported that in all the codon 178 families they studied with a CJD-like disease the codon 178 mutation was encoded on a valine 129 allele while all FFI kindreds encode the same codon 178 mutation on a methionine 129 allele. They suggested that the genotype at codon 129 determines phenotype. Insomnia is not uncommon in CJD patients and FFI and CJD may represent extremes of a spectrum of related disease phenotypes. An inherited case with the E200K

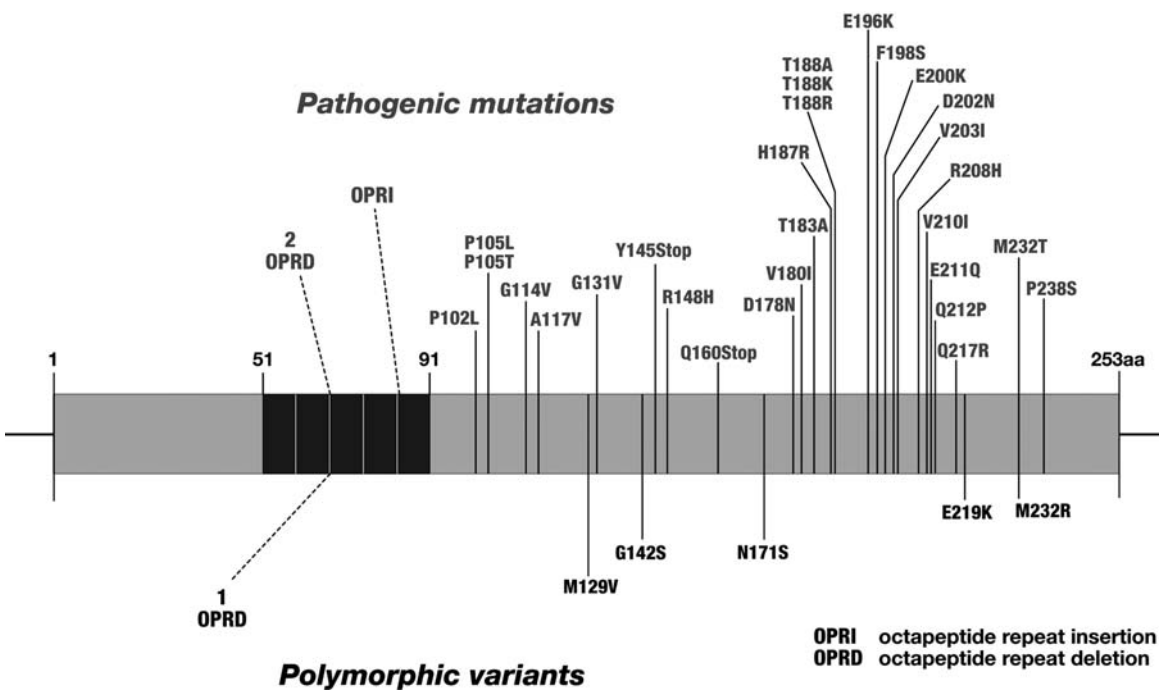


Fig. 4.1.4.1 Pathogenic mutations (above) and polymorphic variants of the human prion protein gene.

mutation, which is normally associated with a CJD-like phenotype, has been reported with an FFI phenotype. An Australian family has also been reported with the FFI genotype but in which affected family members have a range of phenotypes encompassing typical CJD, FFI, and an autosomal dominant cerebellar ataxia-like illness.

(d) PrP E200K

This mutation was first described in families with CJD. Affected individuals develop a rapidly progressive dementia with myoclonus and pyramidal, cerebellar or extrapyramidal signs, and a duration of illness usually less than 12 months. In marked contrast to other variants of inherited prion disease, the EEG usually shows the characteristic pseudoperiodic sharp wave activity seen in sporadic CJD. Interestingly, this mutation accounts for the three reported ethnogeographic clusters of CJD where the local incidence of CJD is around 100-fold higher than elsewhere (amongst Libyan Jews and in regions of Slovakia and Chile).[50,51] Now that cases can be diagnosed by PrP gene analysis, atypical forms of this condition are being detected with phenotypes other than that of classical CJD. Of interest also are reports that peripheral neuropathy can occur in this disease. Elderly unaffected carriers of the mutation have been reported. Patients with this condition have now been reported in several other countries outside the well-recognized clusters, including the UK. At least one of the UK cases does not appear to related to the ethnogeographic clusters mentioned above suggesting a separate UK focus for this type of inherited prion disease.

(e) PrP 6 OPRI

This was the first PrP mutation to be reported and was found in a small UK family with familial CJD[9] now known to form part of a the largest known kindred with an inherited prion disease caused by an OPRI mutation. The diagnosis in the family had been based on an individual who died in the 1940s with a rapidly progressive illness characteristic of CJD. The reported duration of illness was 6 months. Pathologically there was gross status spongiosis and astrocytosis affecting the entire cerebral cortex, and this case is used to illustrate classic CJD histology in Greenfield's *Neuropathology*. However, other family members had a much longer duration GSS-like illness. Histological features were also extremely variable. This observation led to screening of various case of neurodegenerative disease and to the identification of a case classified on clinical grounds as familial Alzheimer's disease.[52] More extensive screening work identified further families with the same mutation which were then demonstrated by genealogical studies to form part of an extremely large kindred.[53–55] Clinical information has been collected on over 80 affected individuals over seven generations. Affected individuals develop in the third to fourth decade onset of a progressive dementia associated with a varying combination of cerebellar ataxia and dysarthria, pyramidal signs, myoclonus and occasionally extrapyramidal signs, chorea and seizures. The dementia is often preceded by depression and aggressive behaviour. A number of cases have a long-standing personality disorder, characterized by aggression, irritability, antisocial and criminal activity, and hypersexuality, which may be present from early childhood, long before overt neurodegenerative disease develops. The histological features vary from those of classical spongiform encephalopathy (with or without PrP amyloid plaques) to cases lacking any specific features of these conditions.[56] Age at onset in this condition can be predicted according to genotype at polymorphic codon 129. Since this pathogenic insertional mutation occurs on a methionine 129 PrP allele, there are two possible codon 129 genotypes for affected individuals, methionine 129 homozygotes or methionine 129/valine 129 heterozygotes. Heterozygotes have an age at onset which is about a decade later than homozygotes.[53]

Pre-symptomatic and antenatal testing

Direct gene testing allows unequivocal diagnosis in patients with inherited forms of the disease and pre-symptomatic testing of unaffected but at-risk family members, as well as antenatal testing.[57] Because of the effect of *PRNP* codon 129 genotype on the age of onset of disease associated with some mutations it is possible to determine within a family whether a carrier of a mutation will have an early or late onset of disease. Most of the mutations appear to be fully penetrant, however experience with some is extremely limited. In some families, for example with E200K or D178N (fatal familial insomnia), there are examples of elderly unaffected gene carriers who appear to have escaped the disease. Genetic counselling is essential prior to pre-symptomatic testing and follows a protocol similar to that established for Huntington's disease. A positive PrP gene analysis has important consequences for other family members, and it is preferable to have discussed these issues with others in the immediate family before testing. Following the identification of a mutation the wider family should be referred for genetic counselling. It is vital to counsel both those testing positive for mutations and those untested but at-risk that they should not be blood or organ donors and should inform surgeons, including dentists, of their risk status prior to significant procedures as precautions may be necessary to minimize risk of iatrogenic transmission.

Prevention

While prion diseases can be transmitted to experimental animals by inoculation, it is important to appreciate that they are not contagious in humans. Documented case-to-case spread has only occurred by cannibalism (kuru) or following accidental inoculation with prions. Such iatrogenic routes include the use of inadequately sterilized intracerebral electrodes, dura mater, and corneal grafting, and from the use of human cadaveric pituitary-derived growth hormone or gonadotrophin. As discussed above, there is now evidence that vCJD prion infection is transmissible by blood transfusion. UK policy for some time has been to leucodeplete all whole blood and to source plasma for plasma products from outside the UK. A further possible route of transmission of vCJD is via contaminated surgical and medical instruments. Prions resist conventional sterilization methods and neurosurgical instruments are known to be able to act as a vector for prion transmission: several cases of iatrogenic transmission of sporadic CJD prions via neurosurgical instruments are documented.[58,59] Recent evidence suggests that classical CJD may also be transmitted by other surgical procedures.[60] The wider tissue distribution of prions in vCJD[42] together with the potential that significant numbers in the population may be silently infected has considerably increased these concerns.

Certain occupational groups are at risk of exposure to human prions, for instance neurosurgeons and other operating theatre staff, pathologists and morticians, histology technicians, as well as

an increasing number of laboratory workers. Because of the prolonged incubation periods to prions following administration to sites other than the central nervous system (CNS), which is associated with clinically silent prion replication in the lymphoreticular tissue, treatments inhibiting prion replication in lymphoid organs may represent a viable strategy for rational secondary prophylaxis after accidental exposure. A preliminary suggested regimen is a short course of immunosuppression with oral corticosteroids in individuals with significant accidental exposure to human prions.[61]

Prognosis and treatment

All recognized prion diseases are invariably fatal following a progressive course.

The duration of illness in sporadic patients is very short with a mean duration of 3–4 months. However, in some of the inherited cases the duration can be 20 years or more. However, there have been significant recent advances in understanding prion propagation and neurotoxicity and clear proof of principle studies of several therapeutic or secondary prophylactic approaches in animal models suggesting effective therapeutics for human disease is realistic.[62]

A variety of drugs have been tried in individual or small numbers of patients over many years. There is no clear evidence of efficacy of any agent, and controlled clinical trials are needed. Such trials are highly challenging. Prion diseases are rare, often rapidly progressive and always fatal which may make randomization to placebo unacceptable. Patterns of disease overall extremely variable with clinical durations varying from weeks to more than 2 years in sporadic CJD, and more than 20 years in some inherited prion diseases. As 'first generation' treatments proposed for prion disease are likely, at best, to have only a modest effect on disease progression, even using survival duration as an outcome measure requires study of large numbers to reliably assess efficacy. There is a lack of systematic natural history studies of disease progression and an absence of biological markers of disease activity. In the United Kingdom, at the request of the Government's Chief Medical Officer, a clinical trial protocol (http://www.controlled-trials.com/ISRCTN06722585/prion1) and infrastructure has been developed to rigorously assess the drug quinacrine[63] and to provide a framework for assessment of novel therapeutics as these become available: the MRC PRION-1 trial. Importantly under these circumstances, a formal consultation with patient's representatives was organized to refine the protocol so that it would be acceptable to the majority of potential participants http://www.mrc.ac.uk/prn/pdf-cjd_workshop.pdf). Pentosan polyphosphate is another candidate anti-prion drug and has shown some efficacy in animal models. Unlike quinacrine, it does not enter the CNS readily and has been administered by intraventricular infusion in several patients. Major toxicity has been reported by this route in animal studies and such treatment was not supported by the UK's Committee of Safety on Medicines or CJD Therapy Advisory Group. A report summarizing clinical experience to date with this treatment has been produced (http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC003453).

While the precise molecular events in prion propagation are not clear, it is clear that PrP^C is the essential substrate. Interference with PrP^C expression in adult brain is without serious effect and blocks onset of neurological disease in animal models.[61] It should be possible to identify small molecules, which penetrate the CNS to bind to PrP^C and to prevent its recruitment into prions, or to use one of a number of emerging technologies to reduce PrP^C expression in brain. If such methods are able to reduce prion propagation rates to below those of natural clearance mechanisms it ought to be possible to cure prion infection. New methods for early diagnosis—and their timely use—will be crucial, as such methods will not reverse neuronal cell loss which is appreciable or severe by the time clinical diagnosis is typically reached. Proof of principle studies in animal models suggest that humanized anti-PrP monoclonal antibodies could be used for passive immunization in the early pathogenesis to block neuroinvasion. This treatment could be considered for known iatrogenically infected individuals.[64]

Further information

UK National Prion Clinic, National Hospital for Neurology and Neurosurgery, London, http://www.nationalprionclinic.org

Medical Research Council Prion Unit, Institute of Neurology, London, http://www.prion.ucl.ac.uk/

UK CJD Surveillance Unit, Western General Hospital, Edinburgh, http://www.cjd.ed.ac.uk/

UK Department of Health, http://www.dh.gov.uk/PolicyAndGuidance/HealthAndSocialCareTopics/CJD/fs/en

CJD Support Network, http://www.cjdsupport.net/

References

1. Gajdusek, D.C., Gibbs, C.J. Jr, and Alpers, M.P. (1966). Experimental transmission of a kuru-like syndrome to chimpanzees. *Nature*, **209**, 794–6.
2. Gibbs, C.J. Jr, Gajdusek, D.C., Asher, D.M., *et al.* (1968). Creutzfeldt-Jakob disease (spongiform encephalopathy): transmission to the chimpanzee. *Science*, **161**, 388–9.
3. Cuillé, J. and Chelle, P.L. (1936). La maladie dite tremblante du mouton est-elle inocuable? *C R Acad Sci*, **203**, 1552–4.
4. Hadlow, W.J. (1959). Scrapie and kuru. *Lancet*, **2**, 289–90.
5. Collinge, J., Whitfield, J., McKintosh, E., *et al.* (2006). Kuru in the 21st century—an acquired human prion disease with very long incubation periods. *Lancet*, **367**, 2068–74.
6. Griffith, J.S. (1967). Self replication and scrapie. *Nature*, **215**, 1043–4.
7. Prusiner, S.B. (1982). Novel proteinaceous infectious particles cause scrapie. *Science*, **216**, 136–44.
8. Jackson, G.S., Hosszu, L.L.P., Power, A., *et al.* (1999). Reversible conversion of monomeric human prion protein between native and fibrilogenic conformations. *Science*, **283**, 1935–7.
9. Owen, F., Poulter, M., Lofthouse, R., *et al.* (1989). Insertion in prion protein gene in familial Creutzfeldt-Jakob disease. *Lancet*, **1**, 51–2.
10. Hsiao, K., Baker, H.F., Crow, T.J., *et al.* (1989). Linkage of a prion protein missense variant to Gerstmann–Straussler syndrome. *Nature*, **338**, 342–5.
11. Collinge, J., Palmer, M.S., and Dryden, A.J. (1991). Genetic predisposition to iatrogenic Creutzfeldt-Jakob disease. *Lancet*, **337**, 1441–2.
12. Palmer, M.S., Dryden, A.J., Hughes, J.T., *et al.* (1991). Homozygous prion protein genotype predisposes to sporadic Creutzfeldt-Jakob disease. *Nature*, **352**, 340–2.
13. Brown, P., Cathala, F., Raubertas, R.F., *et al.* (1987). The epidemiology of Creutzfeldt-Jakob disease: conclusion of a 15-year investigation in France and review of the world literature. *Neurology*, **37**, 895–904.
14. Collinge, J., Sidle, K.C.L., Meads, J., *et al.* (1996). Molecular analysis of prion strain variation and the aetiology of 'new variant' CJD. *Nature*, **383**, 685–90.

15. Hill, A.F., Joiner, S., Wadsworth, J.D., *et al.* (2003). Molecular classification of sporadic Creutzfeldt-Jakob disease. *Brain*, **126**(Pt 6), 1333–46.
16. Hill, A.F., Desbruslais, M., Joiner, S., *et al.* (1997). The same prion strain causes vCJD and BSE. *Nature*, **389**, 448–50.
17. Bruce, M.E., Will, R.G., Ironside, J.W., *et al.* (1997). Transmissions to mice indicate that 'new variant' CJD is caused by the BSE agent. *Nature*, **389**, 498–501.
18. Parchi, P., Giese, A., Capellari, S., *et al.* (1999). Classification of sporadic Creutzfeldt-Jakob disease based on molecular and phenotypic analysis of 300 subjects. *Annals of Neurology*, **46**, 224–33.
19. Prusiner, S.B., Scott, M., Foster, D., *et al.* (1990). Transgenetic studies implicate interactions between homologous PrP isoforms in scrapie prion replication. *Cell*, **63**, 673–86.
20. Bruce, M., Chree, A., McConnell, I., *et al.* (1994). Transmission of bovine spongiform encephalopathy and scrapie to mice: strain variation and the species barrier. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **343**, 405–11.
21. Collinge, J. (1999). Variant Creutzfeldt-Jakob disease. *Lancet*, **354**, 317–23.
22. Hill, A.F., Joiner, S., Linehan, J., *et al.* (2000). Species barrier independent prion replication in apparently resistant species. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 10248–53.
23. Hill, A.F. and Collinge, J. (2003). Subclinical prion infection. *Trends in Microbiology*, **11**, 578–84.
24. Collinge, J. (2001). Prion diseases of humans and animals: their causes and molecular basis. *Annual Review of Neuroscience*, **24**, 519–50.
25. Collinge, J., Palmer, M.S., Sidle, K.C.L., *et al.* (1995). Unaltered susceptibility to BSE in transgenic mice expressing human prion protein. *Nature*, **378**, 779–83.
26. Asante, E.A., Linehan, J.M., Desbruslais, M., *et al.* (2002). BSE prions propagate as either variant CJD-like or sporadic CJD-like prion strains in transgenic mice expressing human prion protein. *The EMBO Journal*, **21**, 6358–66.
27. Wadsworth, J.D., Asante, E.A., Desbruslais, M., *et al.* (2004). Human prion protein with valine 129 prevents expression of variant CJD phenotype. *Science*, **306**, 1793–6.
28. Asante, E.A., Linehan, J.M., Gowland, I., *et al.* (2006). Dissociation of pathological and molecular phenotype of variant Creutzfeldt-Jakob disease in transgenic human prion protein 129 heterozygous mice. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 10759–64.
29. Collinge, J. (1996). New diagnostic tests for prion diseases. *The New England Journal of Medicine*, **335**, 963–5.
30. Otto, M., Stein, H., Szudra, A., *et al.* (1997). S-100 protein concentration in the cerebrospinal fluid of patients with Creutzfeldt-Jakob disease. *Journal of Neurology*, **244**, 566–70.
31. Zerr, I., Bodemer, M., Räcker, S., *et al.* (1995). Cerebrospinal fluid concentration of neuron-specific enolase in diagnosis of Creutzfeldt-Jakob disease. *Lancet*, **345**, 1609–10.
32. Macfarlane, R.G., Wroe, S.J., Collinge, J., *et al.* (2006). A review of the neuroimaging findings in human prion disease. *Journal of Neurology, Neurosurgery, and Psychiatry*.
33. Alpers, M.P. (1987). Epidemiology and clinical aspects of kuru. In *Prions: novel infectious pathogens causing scrapie and Creutzfeldt-Jakob disease* (eds. S.B. Prusiner and M.P. McKinley), pp. 451–65. Academic Press, San Diego.
34. Esmonde, T.F., Will, R.G., Slattery, J.M., *et al.* (1993). Creutzfeldt-Jakob disease and blood transfusion. *Lancet*, **341**, 205–7.
35. Britton, T.C., Al-Sarraj, S., Shaw, C., *et al.* (1995). Sporadic Creutzfeldt-Jakob disease in a 16-year-old in the UK. *Lancet*, **346**, 1155.
36. Collie, D.A., Summers, D.M., Sellar, R.J., *et al.* (2003). Diagnosing variant Creutzfeldt-Jakob disease with the pulvinar sign: MR imaging findings in 86 neuropathologically confirmed cases. *AJNR. American Journal of Neuroradiology*, **24**, 1560–9.
37. Chazot, G., Broussolle, E., Lapras, C.I., *et al.* (1996). New variant of Creutzfeldt-Jakob disease in a 26-year-old French man. *Lancet*, **347**, 1181.
38. Zeidler, M., Sellar, R.J., Collie, D.A., *et al.* (2000). The pulvinar sign on magnetic resonance imaging in variant Creutzfeldt-Jakob disease. *Lancet*, **355**, 1412–8.
39. Siddique, D., Kennedy, A., Thomas, D., *et al.* (2005). Tonsil biopsy in the investigation of suspected variant Creutzfeldt-Jakob disease—a cohort study of 50 pts. *Journal of the Neurological Sciences*, **238**(Suppl. 1), S1–570.
40. Hill, A.F., Zeidler, M., Ironside, J., *et al.* (1997). Diagnosis of new variant Creutzfeldt-Jakob disease by tonsil biopsy. *Lancet*, **349**, 99–100.
41. Hill, A.F., Butterworth, R.J., Joiner, S., *et al.* (1999). Investigation of variant Creutzfeldt-Jakob disease and other human prion diseases with tonsil biopsy samples. *Lancet*, **353**, 183–9.
42. Wadsworth, J.D.F., Joiner, S., Hill, A.F., *et al.* (2001). Tissue distribution of protease resistant prion protein in variant CJD using a highly sensitive immuno-blotting assay. *Lancet*, **358**, 171–80.
43. Hilton, D.A., Sutak, J., Smith, M.E., *et al.* (2004). Specificity of lymphoreticular accumulation of prion protein for variant Creutzfeldt-Jakob disease. *Journal of Clinical Pathology*, **57**, 300–2.
44. Hilton, D.A., Fathers, E., Edwards, P., *et al.* (1998). Prion immunoreactivity in appendix before clinical onset of variant Creutzfeldt-Jakob disease. *Lancet*, **352**, 703–4.
45. Hilton, D.A., Ghani, A.C., Conyers, L., *et al.* (2004). Prevalence of lymphoreticular prion protein accumulation in UK tissue samples. *The Journal of Pathology*, **203**, 733–9.
46. Collinge, J. (2005). Molecular neurology of prion disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, **76**, 906–19.
47. Collinge, J. (2005). Creutzfeldt-Jakob disease and other prion diseases. In *Dementia* (ed. J.T. O'Brien), pp. 763–76.
48. Medori, R., Montagna, P., Tritschler, H.J., *et al.* (1992). Fatal familial insomnia: a second kindred with mutation of prion protein gene at codon 178. *Neurology*, **42**, 669–70.
49. Goldfarb, L.G., Petersen, R.B., Tabaton, M., *et al.* (1992). Fatal familial insomnia and familial Creutzfeldt-Jakob disease: disease phenotype determined by a DNA polymorphism. *Science*, **258**, 806–8.
50. Goldfarb, L.G., Korczyn, A.D., Brown, P., *et al.* (1990). Mutation in codon 200 of scrapie amyloid precursor gene linked to Creutzfeldt-Jakob disease in Sephardic Jews of Libyan and non-Libyan origin. *Lancet*, **336**, 637–8.
51. Brown, P., Galvez, S., Goldfarb, L.G., *et al.* (1992). Familial Creutzfeldt-Jakob disease in Chile is associated with the codon 200 mutation of the PRNP amyloid precursor gene on chromosome 20. *Journal of Neurological Sciences*, **112**, 65–7.
52. Collinge, J., Harding, A.E., Owen, F., *et al.* (1989). Diagnosis of Gerstmann-Straussler syndrome in familial dementia with prion protein gene analysis. *Lancet*, **2**, 15–17.
53. Poulter, M., Baker, H.F., Frith, C.D., *et al.* (1992). Inherited prion disease with 144 base pair gene insertion. I. Genealogical and molecular studies. *Brain*, **115**, 675–85.
54. Mead, S., Poulter, M., Beck, J., *et al.* (2006). Inherited prion disease with six octapeptide repeat insertional mutation—molecular analysis of phenotypic heterogeneity. *Brain*, **129**(Pt 9), 2297–317.
55. Collinge, J., Brown, J., Hardy, J., *et al.* (1992). Inherited prion disease with 144 base pair gene insertion. II. Clinical and pathological features. *Brain*, **115**, 687–710.
56. Collinge, J., Owen, F., Poulter, M., *et al.* (1990). Prion dementia without characteristic pathology. *Lancet*, **336**, 7–9.
57. Collinge, J., Poulter, M., Davis, M.B., *et al.* (1991). Presymptomatic detection or exclusion of prion protein gene defects in families with inherited prion diseases. *American Journal of Human Genetics*, **49**, 1351–4.

58. Bernoulli, C., Siegfried, J., Baumgartner, G., *et al.* (1977). Danger of accidental person-to-person transmission of Creutzfeldt-Jakob disease by surgery [letter]. *Lancet*, **1**, 478–9.
59. Blattler, T. (2002). Implications of prion diseases for neurosurgery. *Neurosurgical Review*, **25**, 195–203.
60. Collins, S., Law, M.G., Fletcher, A., *et al.* (1999). Surgical treatment and risk of sporadic Creutzfeldt-Jakob disease: a case-control study. *Lancet*, **353**, 693–7.
61. Aguzzi, A., and Collinge, J. (1997). Post-exposure prophylaxis after accidental prion inoculation. *Lancet*, **350**, 1519–20.
62. Mallucci, G. and Collinge, J. (2005). Rational targeting for prion therapeutics. *Nature Reviews. Neuroscience*, **6**, 23–34.
63. Korth, C., May, B.C., Cohen, F.E., *et al.* (2001). Acridine and phenothiazine derivatives as pharmacotherapeutics for prion disease. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 9836–41.
64. Wroe, S., Pal, S., Siddique, D., *et al.* (2006). Clinical presentation and pre-mortem diagnosis of variant Creutzfeldt-Jakob disease associated with blood transfusion: a case report. *Lancet*, **368**, 2061–7.

4.1.5 Dementia with Lewy bodies

I. G. McKeith

Introduction

Lewy bodies are spherical neuronal inclusions, first described by the German neuropathologist Friederich Lewy while working in Alzheimer's laboratory in Munich in 1912. In 1961, Okazaki published case reports about two elderly men who presented with dementia and died shortly after with severe extrapyramidal rigidity. Autopsy showed Lewy bodies in their cerebral cortex.[1] Over the next 20 years, 34 similar cases were reported, all by Japanese workers. Lewy body disease was thus considered to be a rare cause of dementia, until a series of studies in Europe and North America, in the late 1980s, identified Lewy bodies in the brains of between 15 and 20 per cent of elderly demented cases reaching autopsy.[2,3] Dementia with Lewy bodies (DLB) is unlikely to be a newly occurring disorder, since re-examination of autopsy material collected from elderly demented patients in Newcastle during the 1960s, reveals cortical Lewy bodies in 17 per cent of cases. The recent recognition of DLB as the second most common form of degenerative dementia in old age is largely due to the widespread use of improved neuropathological techniques, initially antiubiquitin immunocytochemistry, and more recently specific staining for alpha-synuclein which is a core constituent of Lewy bodies and related lesions.

The spectrum of Lewy body disease

The presence of Lewy bodies probably indicates neuronal dysfunction which is usually indicative of neurological disease. The clinical presentation varies according to the site of Lewy body formation and associated neuronal loss. Three main clinicopathological syndromes have been described.

- Parkinson's disease, an extrapyramidal movement disorder—associated with degeneration of subcortical neurones, particularly in substantia nigra.
- Dementia with Lewy bodies, a dementing disorder with prominent neuropsychiatric features—associated with degeneration of cortical neurones, particularly in frontal, anterior cingulate, insular, and temporal regions.
- Autonomic failure with syncope and orthostatic hypotension—associated with degeneration of sympathetic neurones in spinal cord.

In clinical practice, elderly patients often have heterogenous combinations of parkinsonism, dementia, and autonomic failure, reflecting pathological involvement at multiple locations.

Clinical features

Dementia is usually, but not always, the presenting feature of DLB. A minority of patients present with parkinsonism alone, some with psychiatric disorder in the absence of dementia, and others with orthostatic hypotension, falls, or transient disturbances of consciousness. Episodes of confusion, progressive cognitive decline, and dementia follow in due course. Fluctuation in cognitive performance and functional abilities, which is based in variations in attention and level of consciousness, is the most characteristic feature of DLB and the one which causes greatest diagnostic difficulties. It is usually evident on a day-to-day basis, and often apparent within much shorter periods. The marked amplitude between best and worst performance distinguishes it from the minor day-to-day variations that commonly occur in dementia of any aetiology. Transient disturbances of consciousness, in which patients are found mute and unresponsive for periods of several minutes, may represent the extreme of fluctuation in attention and arousal and are often mistaken for transient ischaemic attacks despite a lack of focal neurological signs. Repeated visual hallucinations are present in about two-thirds of patients. They take the form of vivid, colourful, and sometimes fragmented figures of people and animals, which are usually described in great detail. Emotional responses vary from intense fear to indifference or even amusement. Although patients may respond to their hallucinations, for example, trying to feed an imaginary dog, they later often have good insight into their unreality. Others develop elaborate systematized delusions, usually persecutory or of a phantom boarder. Auditory hallucinations are much less frequent, and only a minority of patients have olfactory or tactile experiences. Depressive symptoms are common and about 40 per cent of patients will have a major depressive episode, similar to the rate in Parkinson's disease and significantly greater than in Alzheimer's disease (AD). The frequency and severity of spontaneous motor features of parkinsonism varies from one clinical setting to another due to referral biases. Postural instability and gait difficulty are the most common manifestations, tremor dominant symptoms occurring in only 20 per cent or less.[4] Less than half of DLB cases have parkinsonism at presentation and a quarter continue to have no evidence at any point in their illness. Clinicians must therefore be prepared to make the diagnosis of DLB in the absence of extrapyramidal motor features. If they do not, their case detection rates will be unacceptably low. Severe neuroleptic sensitivity reactions can precipitate irreversible parkinsonism, further impair consciousness level, and induce autonomic disturbances reminiscent of neuroleptic malignant syndrome. They occur in 40 to 50 per cent of neuroleptic-treated DLB cases and are associated with a two- to threefold

increased mortality.[5] Acute D_2 receptor blockade is thought to mediate these effects; and, despite initial reports, atypical antipsychotics seem to be as likely to cause neuroleptic sensitivity reactions as older drugs.[6] Sleep disorders have more recently been recognized as common in DLB with daytime somnolence and rapid eye movement sleep behaviour disorder as prodromal features.[7] Recurrent falls and syncope occur in up to a third of DLB cases, reflecting autonomic nervous system involvement which may also be evident as early urinary incontinence, constipation, and sexual dysfunction.

Pathological classification

Lewy bodies are composed of intermediate neurofilament proteins, which are abnormally truncated and phosphorylated. Their presence indicates that a neurone is attempting to eliminate damaged proteins from its cytoplasm, a process which is usually followed by cell death. Ubiquitin, α-synuclein, α- and β-crystalin, and associated enzymes are the main chemical constituents. Subcortical Lewy bodies have a dense hyaline core surrounded by a halo of radiating filaments, and are easily seen with conventional histopathological techniques. Cortical Lewy bodies are more easily visualized using antiubiquitin staining but this lacks specificity and immunohistochemical staining for alpha-synuclein, is now the most sensitive and specific method currently available for detecting Lewy bodies and Lewy-related pathology. Current thinking is that Lewy bodies form within neurones as a cytoprotective response in an attempt to sequester toxic alpha-synuclein oligomers. Widely distributed aggregates of alpha-synuclein (Lewy neurites) probably represent an earlier stage in the neurodegenerative process than Lewy body formation itself. Lewy neurites are seen in the substantia nigra, hippocampal region CA2/3, dorsal vagal nucleus, basal nucleus of Meynert, and transentorhinal cortex. Ubiquitin immunocytochemistry and α-synuclein-specific monoclonal antibody stains are beginning to reveal the extensive nature of these neuritic changes, which are probably more relevant for symptom formation than the relatively sparsely distributed Lewy bodies. The presence of Lewy neurites in presynaptic terminals is thought to have a particularly severe impact on synaptic function.[8]

Recommendations have been made[9] about which brain regions to examine for the presence of Lewy bodies and Lewy neurites and a simple semi-quantitative scoring system devised. These scores are added to generate three pathological categories:

1 Brainstem-predominant DLB: predilection sites are substantia nigra, locus coeruleus, and dorsal nucleus of vagus.

2 Limbic (or transitional) DLB: predilection sites are anterior cingulate and transtentorhinal cortex.

3 Neocortical DLB: predilection sites are frontal, temporal, and parietal cortex.

Of DLB cases presenting via psychiatric clinics, 69 per cent have extensive neocortical Lewy body pathology,[10] but this is not essential for the development of dementia or other psychiatric symptoms, both of which may occur in the presence of disease limited to limbic structures (24 per cent of cases) or the brainstem (7 per cent).

Interpretation of the significance of coexistent Alzheimer-type pathology is a major issue in the pathological assessment of DLB cases. High senile plaque counts are found in 80 to 90 per cent of DLB cases, diffuse and neuritic β-amyloid plaques occurring in similar proportions as in pure AD. Significant tau pathology is absent, however, in 80 to 90 per cent whether measured biochemically or by counting neocortical neurofibrillary tangles. Most DLB cases are therefore classified as 'the Lewy body variant of AD'[2] if AD is defined by increased plaque density. Conversely, if AD is defined by frequent neocortical neurofibrillary tangles, equivalent to Braak stages 5 and 6, then 85 to 90 per cent of DLB cases will not fulfil such criteria.[11] (The pathological classification of AD is also discussed in Chapter 4.1.2.) The most recent revision of pathological diagnostic criteria for DLB suggests that both Lewy and Alzheimer pathologies should be fully reported. A probability matrix (see Table 4.1.5.1) is then used to predict the likelihood of the patient having presented with a DLB clinical syndrome, this being directly related to the severity of Lewy-related pathology, and inversely related to the severity of concurrent AD-type pathology.[9] Minor vascular pathology is additionally present in 30 per cent of DLB cases[10] and this is also likely to impact upon the clinical manifestations.

Table 4.1.5.1 Pathological criteria for DLB taking into account the relative contributions of Lewy body and Alzheimer type pathology as predictors of a probable DLB clinical presentation–high, intermediate, or low probability.

		Alzheimer type pathology		
		NIA–Reagan Low (Braak stage 0-II)	**NIA–Reagan Intermediate (Braak stage III-IV)**	**NIA–Reagan High (Braak stage V-VI)**
Lewy Body type pathology	**Brainstem-predominant**	Low	Low	Low
	Limbic (transitional)	High	Intermediate	Low
	Diffuse neocortical	High	High	Intermediate

The relationship between DLB and Parkinson's disease dementia

There has been extended debate about the classification of patients who present with motor symptoms of Parkinson's disease and later develop the typical features of DLB, sometimes after many years of severe motor disability. This is a common outcome reported in up to 78 per cent of PD patients followed over an 8-year period. No major differences between DLB and Parkinson's disease dementia have been found in any variable examined including cognitive profile, neuropsychiatric features, sleep disorders, autonomic dysfunction, type and severity of parkinsonism, neuroleptic sensitivity, and responsiveness to cholinesterase inhibitors. It has been suggested that DLB should be diagnosed when dementia occurs before or concurrently with parkinsonism and Parkinson's disease dementia should be used to describe dementia that occurs in the context of well-established Parkinson's disease.[9,12] This distinction between DLB and Parkinson's disease dementia has two distinct clinical phenotypes, based solely on the temporal sequence of appearance of symptoms that has been criticized by those who regard the different clinical presentations as simply representing

different points on a common spectrum of LB disease, itself underpinned by abnormalities in alpha-synuclein metabolism.

The neurobiological basis of dementia in Parkinson's disease is discussed in detail in Chapter 4.1.6.

Clinical diagnosis of DLB

Patients with DLB may present to psychiatric services (cognitive impairment, psychosis, or behavioural disturbance), internal medicine (acute confusional states or syncope), or neurology (movement disorder or disturbed consciousness). The details of clinical assessment and differential diagnoses will, to a large extent, be shaped by these symptom and specialty biases. In all cases, a detailed history from the patient and reliable informants should document the time of onset of relevant key symptoms, the nature of their progression, and their effects on social, occupational, and personal function.

The recent consensus criteria for the clinical diagnosis of DLB are shown in Table 4.1.5.2. Particular emphasis needs to be given to recognizing the characteristic dementia syndrome. Attentional deficits and prominent frontosubcortical and visuo-perceptual dysfunction are the main features—symptoms of persistent or prominent memory impairment are not always present early in the course of illness, although they are likely to develop in most patients with disease progression. Patients with DLB perform better than Alzheimer's disease on tests of verbal recall, but relatively worse on tests of copying and drawing. With the progression of dementia, the selective pattern of cognitive deficits may be lost, making differential diagnosis based on clinical examination difficult during the later stages.

It is the evaluation of fluctuation which causes greatest difficulty in clinical practice.[13] Questions such as, 'are there episodes when his/her thinking seems quite clear and then becomes muddled?' were previously suggested as useful probes, but two recent studies found 75 per cent of both AD and DLB carers to respond positively.[14,15] More detailed questioning and qualitative analysis of carers' replies is therefore needed. The Clinician Assessment of Fluctuation Scale[16] requires an experienced clinician to judge the severity and frequency of 'fluctuating confusion' or 'impaired consciousness' over the previous month. The semi-structured One Day Fluctuation Assessment Scale[16] can be administered by less experienced raters and generates a cut-off score to distinguish DLB from AD or VaD. The Mayo Fluctuations Composite Scale[15] requires three or more 'yes' responses from caregivers to structured questions about the presence of daytime drowsiness and lethargy, daytime sleep >2 h, staring into space for long periods or episodes of disorganized speech, as suggestive of DLB rather than AD. Recording variations in attentional performance using a computer-based test system offers an independent method of measuring fluctuation, which is also sensitive to drug treatment effects.[17] The assessment of extrapyramidal motor features may be complicated by the presence of cognitive impairment. A simple, five-item subscale of the Unified PD Rating Scale[18] contains only those items that can reliably be assessed in DLB independent of severity of dementia (tremor at rest, action tremor, body bradykinesia, facial expression, rigidity). Standardized methods of assessing visual hallucinations and other visual pathologies in DLB are under development.[19]

(a) Consensus criteria for DLB

Probable DLB can be diagnosed (Table 4.1.5.2) if any two of the three core features (fluctuation, visual hallucinations, spontaneous

Table 4.1.5.2 Consensus criteria for the clinical diagnosis of probable and possible dementia with Lewy bodies (DLB) (Reproduced from McKeith I., Dickson, D., Emre, M., *et al.* Dementia with Lewy bodies, 3rd report of the dementia consortium, *Neurology*, **65**, 1863–72, copyright 2005, AAN Enterprises, Inc.)

1	**Central feature** (*essential for a diagnosis of possible or probable DLB*) Dementia defined as progressive cognitive decline of sufficient magnitude to interfere with normal social or occupational function. Prominent or persistent memory impairment may not necessarily occur in the early stages but is usually evident with progression. Deficits on tests of attention, executive function and visuo-spatial ability may be especially prominent.
2	**Core features** (*two core features are sufficient for a diagnosis of probable DLB, one for possible DLB*) Fluctuating cognition with pronounced variations in attention and alertness Recurrent visual hallucinations that are typically well formed and detailed Spontaneous features of parkinsonism
3	**Suggestive features** (*if one or more of these is present in the presence of one or more core features, a diagnosis of probable DLB can be made. In the absence of any core features, one or more suggestive features is sufficient for possible DLB. Probable DLB should not be diagnosed on the basis of suggestive features alone*) REM sleep behaviour disorder Severe neuroleptic sensitivity Low dopamine transporter uptake in basal ganglia demonstrated by SPECT or PET imaging
4	**Supportive features** (*commonly present but not proven to have diagnostic specificity*) Repeated falls and syncope Transient, unexplained loss of consciousness Severe autonomic dysfunction e.g. orthostatic hypotension, urinary incontinence Hallucinations in other modalities Systematized delusions Depression Relative preservation of medial temporal lobe structures on CT/MRI scan Generalised low uptake on SPECT/PET perfusion scan with reduced occipital activity Abnormal (low uptake) MIBG myocardial scintigraphy Prominent slow wave activity on EEG with temporal lobe transient sharp waves
5	**A diagnosis of DLB is less likely** In the presence of cerebrovascular disease evident as focal neurological signs or on brain imaging In the presence of any other physical illness or brain disorder sufficient to account in part or in total for the clinical picture If parkinsonism only appears for the first time at a stage of severe dementia
6	**Temporal sequence of symptoms** DLB should be diagnosed when dementia occurs before or concurrently with parkinsonism (if it is present). The term Parkinson's disease dementia (PDD) should be used to describe dementia that occurs in the context of well-established Parkinson's disease. In a practice setting the term that is most appropriate to the clinical situation should be used and generic terms such as LB disease are often helpful. In research studies in which distinction needs to be made between DLB and PDD, the existing one-year rule between the onset of dementia and parkinsonism DLB continues to be recommended. Adoption of other time periods will simply confound data pooling or comparison between studies. In other research settings that may include clinico-pathologic studies and clinical trials, both clinical phenotypes may be considered collectively under categories such as LB disease or alpha-synucleinopathy.

motor features of parkinsonism) are present. Probable DLB can also be diagnosed if one core feature is accompanied by one or more suggestive features (REM sleep behaviour disorder, severe neuroleptic sensitivity, low dopamine transporter uptake in basal ganglia demonstrated by SPECT or PET imaging). Possible DLB can be diagnosed if there is one core feature alone or one or more suggestive features in the absence of any core features. Suggestive features are not in the light of current knowledge considered sufficient, even in combination, to warrant a diagnosis of probable DLB in the absence of any core feature.

Differential diagnosis

There are four main categories of disorder that should be considered in the differential diagnosis of DLB (Table 4.1.5.3).

(a) Other causes of dementia

Of autopsy-confirmed DLB cases, 65 per cent meet the NINCDS-ADRDA clinical criteria for probable or possible AD,[20] and this is the most frequent clinical misdiagnosis of DLB patients presenting with a primary dementia syndrome. This suggests DLB should routinely be excluded when making the diagnosis of AD. Up to one-third of DLB cases are additionally misclassified as vascular dementia by virtue of items such as the fluctuating nature and course of illness. Pyramidal and focal neurological signs are, however, usually absent. The development of myoclonus in patients with a rapidly progressive form of DLB may lead the clinician to suspect Creutzfeldt–Jakob disease.

(b) Other causes of delirium

In patients with intermittent delirium, appropriate examination and laboratory tests should be performed during the acute phase to maximize the chances of detecting infective, metabolic, inflammatory, or other aetiological factors. Pharmacological causes are particularly common in elderly patients. Although the presence of any of these features makes a diagnosis of DLB less likely, comorbidity is not unusual in elderly patients and the diagnosis should not be excluded simply on this basis.

Table 4.1.5.3 Conditions to be considered in the differential diagnosis of dementia with Lewy bodies

Other causes of dementia
Alzheimer's disease
Vascular dementia
Creutzfeldt–Jakob disease
Other causes of delirium
Infective/pharmacological/metabolic/inflammatory
Other neurological syndromes
Parkinson's disease
Progressive supranuclear palsy (Steele–Richardson–Olszewski syndrome)
Multisystem atrophy
Corticobasal degeneration
Rapid eye movement sleep behaviour disorder (RBD)
Recurrent syncope/unexplained falls
Transient disturbances of consciousness
Comlex partial seizures
Late-onset psychiatric disorders
Delusional disorder (late paraphrenia/late-onset schizophrenia)
Depressive psychosis

(c) Other neurological syndromes

In patients with a prior diagnosis of Parkinson's disease, the onset of visual hallucinations and fluctuating cognitive impairment may be attributed to side-effects of antiparkinsonian medications, and this must be tested by dose reduction or withdrawal. Other atypical parkinsonian syndromes associated with poor levodopa response, cognitive impairment, and postural instability include progressive supranuclear palsy and multi-system atrophy. Syncopal episodes in DLB are often incorrectly attributed to transient ischaemic attacks, despite an absence of focal neurological signs. Recurrent disturbances in consciousness accompanied by complex visual hallucinations may suggest complex partial seizures (temporal lobe epilepsy), and vivid dreaming with violent movements during sleep may meet criteria for REM sleep behaviour disorder. Both these conditions have been reported as uncommon presenting symptoms of autopsy-confirmed DLB.[21,22]

(d) Late-onset functional psychiatric disorder

DLB should be considered if a patient spontaneously develops parkinsonian features or cognitive decline (or shows excessive sensitivity to neuroleptic medication) in the course of late-onset delusional disorder, depressive psychosis, or mania.[21]

Laboratory investigations including neuroimaging

Systemic and pharmacological causes of delirium need to be excluded. The standard EEG may show early slowing, epoch by epoch fluctuation, and transient temporal slow wave activity.[23] There are as yet no clinically applicable genotypic or CSF markers to support a DLB diagnosis.[24] There are, however, sufficient studies to conclude that neuroimaging investigations may be helpful in supporting the clinical diagnosis. Changes associated with DLB include preservation of hippocampal and medial temporal lobe volume on MRI[25,26] and occipital hypo-perfusion on SPECT.[27] Other features such as generalized atrophy, white matter changes,[28] and rates of progression of whole brain atrophy[29] appear to be unhelpful in differential diagnosis. Dopamine transporter loss in the caudate and putamen, a marker of nigro-striatal degeneration can be detected by pre-synaptic dopaminergic SPECT. Preliminary studies suggesting high specificity and sensitivity for predicting clinical[30] and pathological diagnoses of DLB have been confirmed in a large multi-centre trial which found 78 per cent sensitivity and 90 per cent specificity for identifying probable DLB versus non-DLB dementia.[31] Diagnostic sensitivity based upon the presence of the three core clinical features alone has been estimated at below 50 per cent and specificity at >90 per cent which suggests that dopaminergic imaging is most useful when significant clinical diagnostic uncertainty exists.

Epidemiology

In population-based clinical studies, prevalences of around 0.7 per cent for DLB in the 65+ age group have been reported suggesting that it could account for up to 10 per cent of all dementia cases, a figure consistent with DLB rates of 10–15 per cent from hospital-based autopsy series. A recent community study of 85+ year olds found 5.0 per cent to meet consensus criteria for DLB (3.3 per cent probable, 1.7 per cent possible) representing 22 per cent of all demented cases,[32] similar to other clinical estimates and consistent with estimates of Lewy body prevalence in a dementia case

register followed to autopsy.[33] One population-based, autopsy study found Lewy bodies to be evenly distributed between the demented and the non-demented, and this may be interpreted as evidence of a substantial pool of pre-clinical cases.[34] Classical epidemiological studies to determine age and sex variation and potential risk factors for DLB have not yet been reported.

Genetics

It is clear from several case studies that familial cases of DLB occur[35,36] and that Lewy bodies are commonly seen in familial cases of Alzheimer's disease.[37] There are recent reports that triplication of the alpha-synuclein gene (SNCA) can cause DLB, Parkinson's disease, and Parkinson's disease dementia whereas gene duplication is associated only with motor Parkinson's disease suggesting a gene dose effect.[38] However, SCNA multiplication is not found in most Lewy body disease patients.

Course and prognosis

Rate of cognitive decline in DLB is generally reported as similar to Alzheimer's disease[39] and survival from onset to death is reduced with self-reports of depression and the presence of extrapyramidal signs as important adverse predictors.[40] The end stage is typically one of profound dementia and parkinsonism. Even in the early stages, personal and social function and performance in daily living skills may be markedly impaired by a combination of cognitive, psychiatric, and neurological disability to a degree significantly greater than in patients with Alzheimer's disease and comparable mental test scores.[41] Psychotic symptoms, particularly visual hallucinations, tend to be very persistent throughout the whole course of illness. There have been three overlapping stages of the illness described.[21]

The first stage is often recognized only in retrospect, and may extend back 1 to 3 years' prepresentation with occasional minor episodes of forgetfulness, sometimes described as lapses of concentration or 'switching off'. A brief period of delirium is sometimes noted for the first time, often associated with genuine physical illness and/or surgical procedures. Disturbed sleep, nightmares, and daytime drowsiness often persist after recovery.

Progression to the second stage frequently prompts psychiatric or medical referral. A more sustained cognitive impairment is established, albeit with marked fluctuations in severity. Recurrent confusional episodes are accompanied by vivid hallucinatory experiences, visual misidentification syndromes, and topographical disorientation. Extensive medical screening is usually negative. Attentional deficits are apparent as apathy, and daytime somnolence and sleep behaviour disorder[17] may be severe. Gait disorder and bradykinesia are often overlooked, particularly in elderly subjects. Frequent falls occur due to either postural instability or syncope.

The third and final stage often begins with a sudden increase in behavioural disturbance, leading to requests for sedation or hospital admission by perplexed and exhausted carers. The natural course from this point is variable and obscured by the high incidence of adverse reactions to neuroleptic medication. For patients not receiving, or not tolerating, neuroleptics a progressive decline into severe dementia with dysphasia and dyspraxia occurs over months or years, with death usually due to cardiac or pulmonary disease. During this terminal phase patients show continuing behavioural disturbance including vocal and motor responses to hallucinatory phenomena. Lucid intervals with some retention of recent memory function and insight may still be apparent. Neurological disability is often profound, with fixed flexion deformities of the neck and trunk and severe gait impairment. Parkinsonian signs and paraplegia in flexion may also occur in advanced AD and other dementias. Parkinsonism occurring for the first time late in the course of a dementia is therefore consistent with a diagnosis of DLB, but not specific for it.

Advice about management

Patient management in DLB is complex and includes: early detection, investigation, diagnosis, and treatment of cognitive impairment; assessment and management of neuropsychiatric and behavioural symptoms; treatment of the movement disorder and monitoring and management of autonomic dysfunction, and sleep disorders. The evidence base for making recommendations about the management of DLB is limited and what follows is based upon consensus opinion of clinicians experienced in treating DLB.[9] The most important practice point in the management of a patient with DLB is caution in (or preferably avoidance of) the use of neuroleptic medications, which are the mainstay of antipsychotic treatment in other patient groups. Severe neuroleptic sensitivity reactions[5,6] can precipitate irreversible parkinsonism, further impair consciousness level, and induce autonomic disturbances reminiscent of neuroleptic malignant syndrome. They occur in 40 to 50 per cent of neuroleptic-treated DLB cases and are associated with a two- to threefold increased mortality. Acute D_2 receptor blockade is thought to mediate these effects; and despite initial reports, atypical antipsychotics seem to be as likely to cause neuroleptic sensitivity reactions as older drugs. A scheme for the management of the neuropsychiatric symptoms of DLB is suggested in Fig. 4.1.5.1.

Until safe and effective medications become available, there is no doubt that the mainstay of clinical management is to educate patients and carers about the nature of their symptoms and to suggest coping strategies. The clinician must ascertain which symptoms are most troublesome for the sufferer and explain the risks and benefits associated with changes in medication. In these circumstances where the clinician is walking a therapeutic tightrope between parkinsonism and psychosis, the best outcome is invariably a compromise between a relatively mobile but psychotic patient and a non-psychotic but immobile individual. The patient and his carers may only be able to decide which is the lesser of these evils after experiencing both states.

Non-pharmacological interventions

Non-pharmacological interventions have the potential to ameliorate many of the symptoms and functional impairments associated with DLB, but none have yet been systematically evaluated. Cognitive dysfunction and associated symptoms such as VH can for example, be exacerbated by low levels of arousal and attention and strategies to increase these by social interaction and environmental novelty may reduce their presence and impact.

Pharmacological treatments

Pharmacological treatment strategies are based upon our knowledge of the neurochemical deficits underlying specific symptoms in DLB. The most clearly established is a correlation between

Explain hallucinations, delusions, and other neuropsychiatric symptoms of DLB to patient and carer as an intrinsic part of the disorder. **EDUCATION ↓**

Establish the nature and extent of cognitive impairment, neuropsychiatric features, sleep disorder, autonomic dysfunction and motor parkinsonism by interview with patient and carer.
Determine the extent to which of these impairs quality of life and requires active treatment.
Record pretreatment scores for each using appropriate rating scales.
BASELINE ASSESSMENT ↓

Gradually reduce, and if possible stop, antiparkinsonian medications in the following order:
- anticholinergics
- L-deprenyl
- amantadine
- direct dopamine agonists
- COMT inhibitors
- levodopa

Gradually reduce, and if possible stop other anticholinergics, sedatives and drugs which may exacerbate confusional symptoms. **REDUCE ANTI-PARKINSONIAN MEDICATIONS ↓**

After each drug change, repeat measures in all relevant symptom domains to monitor for beneficial or adverse responses. **MONITOR RESPONSE ↓**

If psychotic symptoms persist despite the reduction or withdrawal of unwanted medications a cholinesterase inhibitor should be gradually introduced in standard dose. Improvements are generally seen within the first few weeks and rapid relapse may occur if the drug is stopped abruptly. **CHOLINESTERASE INHIBITOR TRIAL ↓**

If psychotic symptoms still persist a cautious trial of an antipsychotic may be justified as long as the patient and carer understand the potential risks of severe adverse side-effects. Since most neuroleptic sensitivity reactions occur during the first 2 weeks of treatment, it may be wise to admit patients into hospital during initiation of neuroleptic therapy. Since side-effects are dose related, treatment should start with the lowest possible dose and slow titration. Approximately 50% of patients are expected to tolerate medication.
NEUROLEPTIC TRIAL—VERY LOW DOSE AND SLOW TITRATION ↓

Neuroleptics should be discontinued if parkinsonism appears for the first time in a DLB patient, and the dose substantially reduced or stopped in patients whose pre-existing parkinsonism worsens. Severe neuroleptic sensitivity reactions should be treated as a medical emergency similar to neuroleptic malignant syndrome.
STOP IF NEUROLEPTIC SENSITIVITY IS SUSPECTED ↓

Other psychotropic medications may offer short-term benefits. Clonazepam can help nocturnal hallucinations and behavioural disturbance. SSRIs and SRNIs may improve mood disorders.
CONSIDER ALTERNATIVE MEDICATIONS

Fig. 4.1.5.1 Management of the neuropsychiatric symptoms of DLB.

substantia nigra neurone loss and severity of parkinsonism. Levodopa responsiveness is less predictable in DLB than in Parkinson's disease. Activity of the cholinergic enzyme choline acetyltransferase is lower in DLB than AD, particularly in temporal and parietal cortex.[42] Clouding of consciousness, confusion, and visual hallucinations are recognized effects of anticholinergic drug toxicity, and the summative effects of subcortical and cortical cholinergic dysfunction probably play a major role in the spontaneous generation of similar fluctuating symptoms in DLB. Reductions in choline acetyltransferase activity are correlated with the severity of cognitive impairment, and hallucinations may be related to hypocholinergic and (relatively) hypermonoaminergic neocortical neurotransmitter function.

Levodopa can be used for the motor disorder of both DLB and PDD.[43,44] Medication should generally be introduced at low doses and increased slowly to the minimum required to minimize disability without exacerbating psychiatric symptoms. Anticholinergics should be avoided. Visual hallucinations are the most commonly experienced psychiatric symptom and are often accompanied by delusions, anxiety and behavioural disturbance. When pharmacological intervention is required the options include cholinesterase inhibitors or atypical antipsychotic medications. Open label studies have demonstrated the effectiveness of all three generally available cholinesterase inhibitors in DLB but placebo controlled trial data is only available to date for rivastigmine.[45] The reported reduction in symptom frequency and intensity of VH appears to be mediated at least in part by improved attentional function and the presence of VH is associated with greater cognitive improvement. Cholinesterase inhibitors also improve cognitive impairment with an effect size that is generally larger than seen with the same drugs when used in AD.[46] There is a risk of symptom of rebound on sudden withdrawal,[47] limited data on long term effects,[48] and none about possible disease modifying effects.

Side effects of cholinesterase inhibitors in DLB include hypersalivation, lacrimation, and urinary frequency, in addition to the usual gastro-intestinal symptoms and a dose dependent exacerbation of extrapyramidal motor features may occur in a minority.[49] There is no evidence that any one cholinesterase inhibitor is better than others.[50] If they are ineffective or if more acute symptom

control of behaviour is required, it may be difficult to avoid a cautious trial of an atypical antipsychotic. The clinician should warn both the carer and patient of the possibility of a severe sensitivity reaction.[5] Second generation atypicals with potentially more favourable pharmacological properties, such as quetiapine, clozapine, and aripiprazole may have theoretical advantages over traditional agents in LB disease but controlled clinical trial data is lacking and clinicians should remain vigilant to the possibility of adverse side effects.

Depression is common in DLB and there have been no systematic studies of its management. At the present time SSRI and SNRIs are probably preferred pharmacological treatment. Tricylic antidepressants and those with anticholinergic properties should generally be avoided. Apathy is also common and may improve with cholinesterase inhibition. Sleep disorders are frequently seen in LB disease and may be an early feature. Rapid eye movement sleep behaviour disorder can be treated with clonazepam 0.25 mg at bedtime, melatonin 3 mg at bedtime, or quetiapine 12.5 mg at bedtime and titrated slowly monitoring for both efficacy and side effects.[7]

Further information

http://www.lewybodydementia.org/—US-based carer organization

http://lewybody.org/—UK-based carer organization

http://www.nlm.nih.gov/medlineplus/lewybodydisease.html—MEDLINE PLUS information site

John O'Brien, Ian McKeith, David Ames, Edmond Chiu. (eds.) (2006). *Dementia with Lewy bodies and Parkinson's disease dementia*. Taylor and Francis, London—the first multidisciplinary textbook covering DLB and related disorders.

References

1. Okazaki, H., Lipkin, L.E., and Aronson, S.M. (1961). Diffuse intracytoplasmic ganglionic inclusions (Lewy type) associated with progressive dementia and quadriparesis in flexion. *Journal of Neuropathology and Experimental Neurology*, **20**, 237–44.
2. Hansen, L., Salmon, D., Galasko, D., *et al.* (1990). The Lewy body variant of Alzheimer's disease: a clinical and pathologic entity. *Neurology*, **40**, 1–8.
3. Perry, R.H., Irving, D., Blessed, G., *et al.* (1990). Senile dementia of Lewy body type. A clinically and neuropathologically distinct form of Lewy body dementia in the elderly. *Journal of the Neurological Sciences*, **95**, 119–39.
4. Burn, D.J., Rowan, E.N., Minett, T., *et al.* (2003). Extrapyramidal features in Parkinson's disease with and without dementia and dementia with Lewy bodies: a cross-sectional comparative study. *Movement Disorders*, **18**, 884–9.
5. McKeith, I., Fairbairn, A., Perry, R., *et al.* (1992). Neuroleptic sensitivity in patients with senile dementia of Lewy body type. *British Medical Journal*, **305**, 673–8.
6. Aarsland, D., Ballard, C., Larsen, J.P., *et al.* (2005). Marked neuroleptic sensitivity in dementia with Lewy bodies and Parkinson's disease. *The Journal of Clinical Psychiatry*, **66**(5), 633–7.
7. Boeve, B.F., Silber, M.H., and Ferman, T.J. (2004). REM sleep behavior disorder in Parkinson's disease and dementia with Lewy bodies. *Journal of Geriatric Psychiatry and Neurology*, **17**, 146–57.
8. Kramer, M.L. and Schulz-Schaeffer, W.J. (2007). Presynaptic alpha-synuclein aggregates, not Lewy bodies, cause neurodegeneration in dementia with Lewy bodies. *The Journal of Neuroscience*, **27**, 1405–10.
9. McKeith, I., Dickson, D., Emre, M., *et al.* (2005). Dementia with Lewy bodies: diagnosis and management: third report of the DLB consortium. *Neurology*, **65**, 1863–72.
10. McKeith, I.G., Ballard, C.G., Perry, R.H., *et al.* (2000). Prospective validation of consensus criteria for the diagnosis of dementia with Lewy bodies. *Neurology*, **54**, 1050–8.
11. Lippa, C.F. and McKeith, I. (2003). Dementia with Lewy bodies improving diagnostic criteria. *Neurology*, **60**, 1571–2.
12. Lippa, C., Duda, J.E., Grossman, M., *et al.* (2007). DLB and PDD boundary issues: diagnosis, treatment, molecular pathology, and biomarkers. *Neurology*, **68**, 812–19.
13. Cummings, J.L. (2004). Fluctuations in cognitive function in dementia with Lewy bodies. *Lancet Neurology*, **3**, 266.
14. Bradshaw, J., Saling, M., Hopwood, M., *et al.* (2004). Fluctuating cognition in dementia with Lewy bodies and Alzheimer's disease is qualitatively distinct. *Journal of Neurology, Neurosurgery, and Psychiatry*, **75**, 382–7.
15. Ferman, T.J., Smith, G.E., Boeve, B.F., *et al.* (2004). DLB fluctuations: specific features that reliably differentiate from AD and normal aging. *Neurology*, **62**, 181–7.
16. Walker, M.P., Ayre, G.A., Cummings, J.L., *et al.* (2000). The clinician assessment of fluctuation and the one day fluctuation assessment scale. Two methods to assess fluctuating confusion in dementia. *The British Journal of Psychiatry*, **177**, 252–6.
17. Walker, M.P., Ayre, G.A., Cummings, J.L., *et al.* (2000). Quantifying fluctuation in dementia with Lewy bodies, Alzheimer's disease and vascular dementia. *Neurology*, **54**, 1616–24.
18. Ballard, C., McKeith, I., Burn, D., *et al.* (1997). The UPDRS scale as a means of identifying extrapyramidal signs in patients suffering from dementia with Lewy bodies. *Acta Neurologica Scandinavica*, **96**, 366–71.
19. Mosimann, U.P., Rowan, E.N., Partington, C.E., *et al.* (2006). Characteristics of visual hallucinations in Parkinson disease dementia and dementia with Lewy bodies. *The American Journal of Geriatric Psychiatry*, **14**, 153–60.
20. McKeith, I.G., Fairbairn, A.F., Perry, R.H., *et al.* (1994). The clinical diagnosis and misdiagnosis of senile dementia of Lewy body type (SDLT). *The British Journal of Psychiatry*, **165**, 324–32.
21. McKeith, I.G., Perry, R.H., Fairbairn, A.F., *et al.* (1992). Operational criteria for senile dementia of Lewy body type (SDLT). *Psychological Medicine*, **22**, 911–22.
22. Boeve, B., Silber, M., Ferman, T., *et al.* (2001). Association of REM sleep behavior disorder and neurodegenerative disease may reflect an underlying synucleinopathy. *Movement Disorders*, **16**, 622–30.
23. Briel, R.C.G., McKeith, I.G., Barker, W.A., *et al.* (1999). EEG findings in dementia with Lewy bodies and Alzheimer's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, **66**, 401–3.
24. McKeith, I., Mintzer, J., Aarsland, D., *et al.* (2004). Dementia with Lewy bodies. *Lancet Neurology*, **3**, 19–28.
25. Barber, R., Ballard, C., McKeith, I.G., *et al.* (2000). MRI volumetric study of dementia with Lewy bodies. A comparison with AD and vascular dementia. *Neurology*, **54**, 1304–9.
26. Barber, R., Gholkar, A., Scheltens, P., *et al.* (1999). Medial temporal lobe atrophy on MRI in dementia with Lewy bodies. *Neurology*, **52**, 1153–8.
27. Lobotesis, K., Fenwick, J.D., Phipps, A., *et al.* (2001). Occipital hypoperfusion on SPECT in dementia with Lewy bodies but not AD. *Neurology*, **56**, 643–9.
28. Barber, R., Gholkar, A., Scheltens, P., *et al.* (2000). MRI volumetric correlates of white matter lesions in dementia with Lewy bodies and Alzheimer's disease. *International Journal of Geriatric Psychiatry*, **15**, 911–16.
29. O'Brien, J.T., Paling, S., Barber, R., *et al.* (2001). Progressive brain atrophy on serial MRI in dementia with Lewy bodies, AD, and vascular dementia. *Neurology*, **56**, 1386–8.
30. O'Brien, J.T., Colloby, S., Fenwick, J., *et al.* (2004). Dopamine transporter loss visualized with FP-CIT SPECT in the differential diagnosis of dementia with Lewy bodies. *Archives of Neurology*, **61**, 919–25.

31. McKeith, I., O'Brien, J., Walker, Z., *et al.* (2007). Sensitivity and specificity of dopamine transporter imaging with (123)I-FP-CIT SPECT in dementia with Lewy bodies: a phase III, multicentre study. *Lancet Neurology*, **6**, 305–13.
32. Rahkonen, T., Eloniemi-Sulkava, U., Rissanen, S., *et al.* (2003). Dementia with Lewy bodies according to the consensus criteria in a general population aged 75 years or older. *Journal of Neurology, Neurosurgery, and Psychiatry*, **74**, 720–4.
33. Holmes, C., Cairns, N., Lantos, P., *et al.* (1999). Validity of current clinical criteria for Alzheimer's disease, vascular dementia and dementia with Lewy bodies. *The British Journal of Psychiatry*, **175**, 45–50.
34. The Neuropathology Group MRC CFAS. (2001). Pathological correlates of late-onset dementia in a multicentre, community-based population in England and Wales. *Lancet*, **357**, 169–75.
35. Tsuang, D.W., DiGiacomo, L., and Bird, T.D. (2004). Familial occurrence of dementia with Lewy bodies. *The American Journal of Geriatric Psychiatry*, **12**, 179–88.
36. Gwinn-Hardy, K. and Singleton, A.A. (2002). Familial Lewy body diseases. *Journal of Geriatric Psychiatry and Neurology*, **15**, 217–23.
37. Trembath, Y., Rosenberg, C., Ervin, J.F., *et al.* (2003). Lewy body pathology is a frequent co-pathology in familial Alzheimer's disease. *Acta Neuropathologica*, **105**, 484–8.
38. Singleton, A. and Gwinn-Hardy, K. (2004). Parkinson's disease and dementia with Lewy bodies: a difference in dose? *Lancet*, **364**, 1105–7.
39. Ballard, C., O'Brien, J., Morris, C.M., *et al.* (2001). The progression of cognitive impairment in dementia with Lewy bodies, vascular dementia and Alzheimer's disease. *International Journal of Geriatric Psychiatry*, **16**, 499–503.
40. Williams, M.M., Xiong, C.J., Morris, J.C., *et al.* (2006). Survival and mortality differences between dementia with Lewy bodies vs Alzheimer disease. *Neurology*, **67**, 1935–41.
41. McKeith, I.G., Rowan, E., Askew, K., *et al.* (2006). More severe functional impairment in dementia with Lewy bodies than Alzheimer disease is related to extrapyramidal motor dysfunction. *The American Journal of Geriatric Psychiatry*, **14**, 582–8.
42. Perry, E.K., Haroutunian, V., Davis, K.L., *et al.* (1994). Neocortical cholinergic activities differentiate Lewy body dementia from classical Alzheimer's disease. *Neuroreport*, **5**, 747–9.
43. Molloy, S., McKeith, I.G., O'Brien, J.T., *et al.* (2005). The role of levodopa in the management of dementia with Lewy bodies. *Journal of Neurology, Neurosurgery, and Psychiatry*, **76**, 1200–3.
44. Bonelli, S.B., Ransmayr, G., Steffelbauer, M., *et al.* (2004). L-dopa responsiveness in dementia with Lewy bodies, Parkinson disease with and without dementia. *Neurology*, **63**, 376–8.
45. McKeith, I., Del-Ser, T., Spano, P.F., *et al.* (2000). Efficacy of rivastigmine in dementia with Lewy bodies: a randomised, double-blind, placebo-controlled international study. *Lancet*, **356**, 2031–6.
46. Samuel, W., Caligiuri, M., Galasko, D., *et al.* (2000). Better cognitive and psychopathologic response to donepezil in patients prospectively diagnosed as dementia with Lewy bodies: a preliminary study. *International Journal of Geriatric Psychiatry*, **15**, 794–802.
47. Minett, T.S.C., Thomas, A., Wilkinson, L.M., *et al.* (2003). What happens when donepezil is suddenly withdrawn? An open label trial in dementia with Lewy bodies and Parkinson's disease with dementia. *International Journal of Geriatric Psychiatry*, **18**, 988–93.
48. Grace, J., Daniel, S., Stevens, T., *et al.* (2001). Long-term use of rivastigmine in patients with dementia with Lewy bodies: an open-label trial. *International Psychogeriatrics*, **13**, 199–205.
49. Thomas, A.J., Burn, D.J., Rowan, E.N., *et al.* (2005). A comparison of the efficacy of donepezil in Parkinson's disease with dementia and dementia with Lewy bodies. *International Journal of Geriatric Psychiatry*, **20**, 938–44.
50. Bhasin, M., Rowan, E., Edwards, K., *et al.* (2007). Cholinesterase inhibitors in dementia with Lewy bodies-a comparative analysis. *International Journal of Geriatric Psychiatry,* **22**, 890–5.

4.1.6 Dementia in Parkinson's disease

R. H. S. Mindham and T. A. Hughes

Introduction

Parkinson's disease has been regarded as a neurological condition mainly affecting motor function and arising from specific lesions in the brain stem. The recognition of dementia in Parkinson's disease is of importance in management but the possibility that motor and cognitive functions may be located in the same region of the brain is of theoretical importance.

The nature of dementia in Parkinson's disease

There have been numerous reports of the impairment of specific cognitive functions in patients with Parkinson's disease. Mortimer and colleagues reported cognitive impairment in 93 per cent of a substantial group of patients with Parkinson's disease.(1) Their data showed neither a clear distinction between impaired groups nor the presence of subtypes of Parkinson's disease in which cognitive impairment is a more frequent occurrence. They proposed that cognitive impairment in Parkinson's disease lay on a continuum of severity, rather than as a feature of particular subgroups. The impairments identified included deficits in memory, language, visuospatial functioning, abstract reasoning, slowness in intellectual tasks, and difficulty in shifting from task to task. These deficits are widespread among patients with Parkinson's disease and can occur at an early stage of the disorder.

A proportion of patients with Parkinson's disease show impairment of a range of cognitive functions akin to the global impairment seen in Alzheimer's disease.(2) However, the pattern of impairment is frequently less severe than in Alzheimer's disease where the pathological changes in the brain are known to be widespread. Cognitive impairment in a range of disorders of movement where the main neuropathological changes reside in the subcortical region of the brain led to the concept of 'subcortical dementia', a form of intellectual impairment of lesser degree than in Alzheimer's disease, but affecting several cognitive functions. Albert described a syndrome of which the main features were: emotional or personality changes, impaired memory, defective ability to manipulate acquired knowledge, and a striking slowness in the rate of information processing.(3)

Many issues arose as to the nature of 'subcortical dementia'. Was it a clinical or a pathological concept? Was the difference between this and other forms of dementia simply one of degree? Did the pathological changes occur in the subcortical region of the brain alone? Was the syndrome of cognitive impairment distinctly different from other dementias or did the presence of motor features of the disorder simply give the intellectual impairment a distinct character? Was subcortical dementia a stable condition or a transitional state leading eventually to global dementia? Opinion has ranged from full acceptance of the concept to scepticism.(4,5)

McHugh[6] suggested that the subcortical region subserves functions not only in motor control and cognitive function but also in the control and display of mood. He suggested that these form a 'subcortical triad' of symptoms most convincingly seen in Huntington's disease. A notable difference between this concept and that of subcortical dementia was that the pathological disturbance of mood is only intermittently present, whereas the motor and cognitive changes are persistent.

Cummings[7] suggested that cognitive impairment in Parkinson's disease takes three forms: one which is relatively mild and meets the criteria for subcortical dementia, a more severe form showing wider impairment of cognitive function but neuropathologically distinct from Alzheimer disease and a severe form which shows neuropathological changes in both the subcortical region of the brain, and in the cortex, the latter of Alzheimer type. This proposal provides a basis for viewing cognitive changes in Parkinson's disease, albeit provisional.

Many reports have suggested that global dementia occurs in Parkinson's disease. Whether such a severe change in cognitive function can be regarded as an intrinsic feature of this disease, whether it implies an extension of a neuropathological process more widely in the brain, or whether it suggests a different neuropathology from the outset is, as yet, uncertain. More recently the debate has shifted to whether dementia in Parkinson's disease, dementia with Lewy bodies and Alzheimer's disease should be viewed as a spectrum, or as separate conditions with varying degrees of clinical and pathological overlap.

The methodology of studies of dementia in Parkinson's disease

Research to establish the status of dementia in Parkinson's disease has confronted a range of methodological issues.[8] A major problem is in the diagnosis of Parkinson's disease itself. The original description of paralysis agitans by Parkinson was the identification of a syndrome rather than of a disease. The part played by such agents as heavy metals, infections, and vascular disease was subsequently recognized. In spite of the use of standardized methods, a substantial proportion of patients diagnosed as suffering from Parkinson's disease in life do not show the expected findings in the brain postmortem. In a follow-up study, 80 per cent of cases were shown to have neuropathological changes of Parkinson's disease after death but over 20 per cent were diagnosed as having suffered from progressive supranuclear palsy, multiple system atrophy, or Alzheimer's disease.[9] Furthermore, some dementing illnesses may show disorder of movement as a clinical feature.

Studies of dementia in Parkinson's disease

Cases of dementia in Parkinson's disease have been reported for over a hundred years. The frequency of dementia reported in cross-sectional or prevalence studies ranges from 0 to over 80 per cent. A recent review found a prevalence of between 28 and 44 per cent in community studies but in older samples of subjects the prevalence was much higher.[10]

Follow-up studies have great advantages in studying the frequency of dementia in Parkinson's disease as they allow the diagnosis of Parkinson's disease to be checked, repeated assessment reduces errors in the recognition of dementia, the pattern of evolution of dementia may be followed, the underestimation of dementia by selective loss through death is avoided, and they reveal the incidence rather than the prevalence of the condition. The choice of methods of diagnosis and assessment that will remain appropriate throughout the period of the follow-up remains a problem.

A prospective, controlled study in the United Kingdom reported an incidence of dementia of 19 per cent after 4.5 years observation. A later report on the same cohort of subjects showed an incidence of dementia of 38 per cent after 10 years of observation. The control group showed cases of cognitive impairment but none amounting to dementia.[11,12] A community based, prospective, controlled study, in Norway, showed the risk of dementia was 5.9 times greater than in the control group.[13] A prospective, controlled study in the United States showed that dementia was 3.7 times greater in the Parkinson's disease group with severely affected, elderly patients especially at risk (Table 4.1.6.1).[14,15]

Prediction of dementia in Parkinson's disease

Those most likely to develop dementia are: older people, patients with Parkinson's disease of longer duration, subjects who have a greater severity of motor symptoms and signs of Parkinson's disease, and those who show greater physical disability.[11–15] Some studies have shown that male sex or late onset are associated with dementia. The apparent association between Parkinson's disease treated with levodopa and dementia is probably due to improved survival.

Neuropathology

The basic lesions are the degeneration of the pigmented neurones in the pars compacta of the substantia nigra in the brain stem; the presence of **Lewy bodies** which are intracytoplasmic neuronal inclusions composed of abnormally phosphorylated neural filament proteins aggregated with ubiquitin and alpha-synuclein; gliosis; and the formation of **Lewy neurites** which are degenerating neurites containing ubiquitin and alpha-synuclein. Alpha-synuclein may play an important role in the pathological process leading to the formation of Lewy bodies, but conclusive proof is lacking. Clinical Parkinson's disease is not apparent until about 80 per cent of the nigro-striatal dopaminergic neurones have died. Lewy bodies had come to be regarded as pathognomonic of Parkinson's disease, but are now known to be present in other diseases. An agreed though arbitrary difference between dementia in Parkinson's disease and dementia with Lewy bodies is that in the latter, parkinsonian symptoms must not precede the occurrence of dementia by more than 12 months.

The degenerative changes in the substantia nigra are known to be closely linked with decreased dopaminergic neurotransmission in the brain, and this deficiency leads to the main motor features of the disease, although other neurotransmitters are also deficient. Some of these deficiencies, which have been associated with cognitive impairment in other disorders, include a deficiency in acetylcholinesterase in the cortex, a deficiency of noradrenaline in the cortex, and a deficiency of serotonin in both striatum and cortex. The concentrations of a range of neuropeptides may also be altered.

The neuropathology of cases of Parkinson's disease showing dementia is inconsistent; some show neuropathological changes

Table 4.1.6.1 Some prospective studies of dementia in Parkinson's disease, using control subjects and employing standardized methods of diagnosis and assessment

Study	PD subjects (N)	Control subjects (N)	Length of follow-up (years)	Diagnostic criteria for dementia	% of PD demented	Number demented per 1000 person years	Dementia in PD v controls: relative risk (95% CI), †odds ratio, ‡hazard ratio
1. Biggins *et al.* 1992 (UK)	87	50	4.5	DSM-III-R	19	47.6	–
2. Hughes *et al.* 2000 (UK)	83	50	10	DSM-III-R	38	42.6	–
3. Aarsland *et al.* 2001 (Norway)	171	3062	4.2	DSM-III-R	33	95.3	†5.9 (3.9–9.1)
4. Levy *et al.* 2002 (USA)	180	180	3.6	DSM-III-R	28.9	79.9	‡3.7 (2.1–6.3)
5. Aarsland *et al.* 2003 (Norway)	224	3295	8	DSM-III-R	78.2	–	2.8
6. Hobson & Meara 2004 (UK)	86	102	4	DSM-IV	35.3	107.1	5.1 (2.1–12.5)
7. de Lau *et al.* 2005 (Holland)	139	6512	9	DSM-III-R	15.1	–	‡2.8 (1.8–4.4)

Notes on studies:

1. Prevalent cases of dementia at entry to study excluded from analysis; PD & controls assessed concurrently; DSM-III-R criteria assessed blind; PD sample drawn from neurological clinics.
2. Same PD cohort and controls, same method of assessment as study 1.
3. Community sample in Norway; controls from community sample in Demark assessed by different instruments and at different intervals. Prevalent cases of dementia excluded from entry to study. DSM-III-R criteria (dementia diagnosis) not assessed blind.
4. Community sample. Prevalent cases of dementia at entry to study excluded from analysis. Concurrent examination of PD and control groups using same instruments; DSM-III-R criteria not assessed blind.
5. Same PD cohort and controls, similar method of assessment, same method of diagnosis of dementia as study3. % PD demented includes 26% cases prevalent at entry to study. Relative risk is for dementia in PD subjects at 4 years versus control subjects at 5 years.
6. Prevalent cases of dementia excluded from entry to study. Elderly sample. DSM-IV criteria not assessed blind.
7. Community sample. Prevalent cases of dementia at entry to study excluded from analysis. DSM-III-R criteria not assessed blind.

extending beyond the subcortical region, whereas in others neuropathological changes are restricted to the subcortical region. Some studies have shown a correlation between the extent of Lewy body pathology and the severity of dementia, but others have not. Lewy bodies may be found in individuals without cognitive impairment and dementia may develop with only minimal cortical Lewy bodies.[16] In some patients with dementia the neuropathological diagnosis is of Alzheimer's disease or of other recognized degenerative conditions of the brain.

Just as there are difficulties in isolating Parkinson's disease from other conditions, there are problems in understanding the interrelationships of dementing disorders. Several distinct neurodegenerative diseases share some aetiological factors, which may represent an interaction between environmental factors and the ageing process but with differing end results arising from factors specific to the process.[17–19] Problems in the diagnosis of Parkinson's disease, the shrinking category of idiopathic Parkinson's disease, and the difficulties occasionally encountered in explaining the development of dementia in the disease, suggests that the interrelationship between causative agents, the clinical features of disorders of movement, the occurrence of cognitive impairment, and the neuropathology of this group of disorders requires substantial further work before it is understood.

The influence of dementia on mortality

Dementia is associated with increased mortality. In Parkinson's disease increased mortality is associated with age, late age of onset of the disease, cognitive impairment, dementia, and, in some studies, male sex. Many of the studies that have been carried out have been methodologically faulty, making comparisons between studies and the identification of the effect of particular factors, including dementia, problematic. One study showed a hazard ratio for Parkinson's disease compared with controls of 1.64, in general, and of 1.94 for to Parkinson's disease with dementia.[20] In another study of mortality almost 50 per cent of those who died were demented compared with a quarter of those who survived.[21]

Clinical aspects of Parkinson's disease with dementia

The most important step in the recognition of dementia in Parkinson's disease is to suspect its presence. The typical blank facial expression seen in Parkinson's disease may obscure a decline in intellectual activity, slowness in movement may conceal intellectual slowness, and sadness may suggest that morbid depression of mood is the reason for a reduction in liveliness. The clinical picture can usually be clarified by careful examination of the mental state. Standardized psychological tests may be useful in some cases.

The clinical importance of dementia in Parkinson's disease is that there is a marked increase in disability, with problems arising in areas of functioning not previously affected by motor impairment alone. Dementia may be accompanied by an increased liability to confusional episodes from the toxic effects of drugs and other causes.[22]

Management of dementia is similar to that for patients suffering from other dementing disorders (Chapter 4.1.3). Controlled trials suggest rivastigmine and donepezil have a moderate effect on cognitive function, but tolerability can be a problem, with worsening

of parkinsonism and gastrointestinal upsets.[23] Rivastigmine is started at 1.5 mg twice daily, and increased by 3 mg per day at intervals of 4 weeks upto a maximum of 12 mg daily. Donepezil should be started at 5 mg in the evenings, increasing to 10 mg after 6 weeks, if tolerated.

Further information

http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD004747/frame.html

References

1. Pirozzolo, F.J., Hansch, E.C., Mortimer, J.A., *et al.* (1982). Dementia in Parkinson's disease: a neuropsychological analysis. *Brain and Cognition*, **1**, 71–83.
2. Pollack, M. and Hornabrook, R.W. (1966). The prevalence, natural history and dementia of Parkinson's disease. *Brain*, **89**, 429–48.
3. Albert, M.L., Feldman, R.G., and Willis, A.L. (1974). The 'subcortical dementia' of progressive supranuclear palsy. *Journal of Neurology, Neurosurgery, and Psychiatry*, **37**, 121–30.
4. Mayeux, R., Stern, Y., Rosen, J., *et al.* (1983). Is 'subcortical dementia' a recognisable clinical entity? *Annals of Neurology*, **14**, 278–83.
5. Brown, R.G. and Marsden, C.D. (1988). Subcortical dementia: the neuropsychological evidence. *Neuroscience*, **25**, 363–87.
6. McHugh, P.R. (1990). The basal ganglia: the region, the integration of its systems and implications for psychiatry and neurology. In *Function and dysfunction in the basal ganglia* (ed. A.J. Franks, J.W. Ironside, R.H.S. Mindham, R.J. Smith, E.J.S. Spokes, and W. Winlow), pp. 259–69. Manchester University Press, Manchester.
7. Cummings, J.L. (1988). The dementias of Parkinson's disease: prevalence, characteristics, neurobiology, and comparison with dementia of the Alzheimer type. *European Neurology*, **28**(Suppl 1), 15–23.
8. Mindham, R.H.S. (1999). The place of dementia in Parkinson's disease: a methodologic saga. In *Advances in neurology*, Vol. 80 (ed. G.M. Stern), pp. 403–8. Lippincott, Williams & Wilkins, Philadelphia.
9. Hughes, A.J., Daniel, S.E., Kilford, L., *et al.* (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico–pathlolgical study of 100 cases. *Journal of Neurology, Neurosurgery, and Psychiatry*, **55**, 181–4.
10. Emre, M. (2003). Dementia associated with Parkinson's disease. *Lancet Neurology*, **2**, 229–37.
11. Biggins, C.A., Boyd, J.L., Harrop, F.M., *et al.* (1992). A controlled longitudinal study of dementia in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, **55**, 566–71.
12. Hughes, T.A., Ross, H.F., Musa, S., *et al.* (2000). A 10-year study of the incidence of and factors predicting dementia in Parkinson's disease. *Neurology*, **54**, 1596–602.
13. Aarsland,D., Andersen, K., Larsen, J.P., *et al.* (2001). Risk of dementia in Parkinson's disease: a community based, prospective study. *Neurology*, **56**, 730–6.
14. Levy, G., Schupf, N., Tang, M.X., *et al.* (2002). Combined effect of age and severity on the risk of dementia in Parkinson's disease. *Annals of Neurology*, **51**, 722–9.
15. Hobson, P. and Meara, J. (2004). Risk and incidence of dementia in a cohort of older subjects with Parkinson's disease in the United Kingdom. *Movement Disorders*, **19**, 1043–9.
16. Braak, H., Rüb, U., Jansen Steur, E.N.H., *et al.* (2005). Cognitive status correlates with neuropathologic stage in Parkinson's disease. *Neurology*, **64**, 1404–10.
17. Appel, S.H. (1981). A unifying hypothesis for the cause of amyotrophic lateral sclerosis, parkinsonism and Alzheimer's disease. *Annals of Neurology*, **10**, 499–505.
18. Calne, D.B., McGeer, E., Eisen, E., *et al.* (1986). Alzheimer's disease, Parkinson's disease and motor neurone disease: abiotrophic interactions between aging and environment. *Lancet*, **328**, 1067–70.
19. Ben-Shlomo, Y., Whitehead, A.S., and Smith, G.D. (1996). Parkinson's, Alzheimer's, and motor neurone disease. Clinical and pathological overlap may suggest common genetic and environmental factors. *Lancet*, **312**, 728.
20. Hughes, T.A., Ross, H.F., Mindham, R.H.S., *et al.* (2004). Mortality in Parkinson's disease and its association with dementia and depression. *Acta Neurologica Scandinavica*, **110**, 118–23.
21. de Lau, L.M.L., Schipper, C.M.A., Hofman, A., *et al.* (2005). Prognosis of Parkinson's disease. Risk of dementia and mortality: the Rotterdam study. *Archives of Neurology*, **62**, 1265–9.
22. Aarsland, D., Brønnick, K., Ehrt, U., *et al.* (2007). Neuropsychiatric symptoms in patients with Parkinson's disease and dementia: frequency, profile and associated care-giver stress. *Journal of Neurology, Neurosurgery, and Psychiatry*, **78**, 36–42.
23. Maidment, I.D., Fox, C., and Boustani, M. (2005). A review of studies describing the use of acetylcholinesterase inhibitors in Parkinson's disease dementia. *Acta Psychiatrica Scandinavica*, **111**, 403–9.

4.1.7 Dementia due to Huntington's disease

Susan Folstein and Russell L. Margolis

Introduction

Huntington's disease (HD) was first described in 1872 by an American physician living on Long Island, New York. His father and grandfather practised medicine in the same community, so he had access to case notes from several generations of families who lived there. This long period of record keeping allowed him to document a hereditary form of chorea, similar to 'common (Sydenham's) chorea', but progressing over many years to death. Its sufferers had a tendency to insanity and suicide. Huntington's brief essay, which also included a clear description of autosomal dominant inheritance, remains one of the classical descriptions of a medical disorder.[1]

Clinical features and course of illness

Huntington's disease is an inherited neuropsychiatric disorder mainly affecting the striatum and its direct connections. It is characterized by a triad of clinical features that are common to diseases of this region: a non-aphasic *dementia*, *depression* and other disorders of mood, and a variety of *dyskinesias*, most typically chorea.[2,3] Chorea, from the Greek word for 'dance', describes involuntary non-stereotyped jerky movements. The illness, insidious in onset, may begin with all or any one of these three features. Patients who present initially to psychiatrists usually have dementia, loss of temper, or depression, often with suicidal thoughts or attempts. Symptoms may appear at any time from early childhood to old age, most frequently between 35 and 45 years of age. Once the illness begins, sufferers gradually deteriorate over many years in their cognitive and motor functioning and end in a persistent vegetative state with almost complete loss of voluntary motor

function. Death occurs after about 16 years and is usually caused by inanition or aspiration pneumonia. Some patients die earlier from suicide or subdural haematoma caused by a fall. Patients with early onset progress more rapidly than those whose symptoms begin later in life.

Pathology and genetics

The earliest visible neuropathology is in the striosomes of the caudate/putamen,[4] followed by a dorsal-to-ventral progressive loss of almost all striatal output neurones. The deep layers of multiple cortical regions are also prominently affected, and there can also be milder neuronal loss in some brainstem nuclei. Protein aggregates, most easily detectable in neuronal nuclei, are prominent. Neuroimaging studies have shown that neuropathological changes typically begin before the onset of clinically detectable disease. In particular, the extent of striatal loss in presymptomatic individuals, as measured by MRI, correlates with the predicted time until disease onset.[5] Cortical thinning[6] and white matter loss and disorganization[7,8] have also been detected in presymptomatic gene carriers. Subtle changes that may be related to abnormal brain development have also been reported.

The disorder, with a point prevalence of about 6/100 000,[9] is caused by the expansion of an unstable triplet repeat sequence (CAG) in the first exon of a gene near the telomere of chromosome 4p.[10] It is transmitted as an autosomal dominant trait; if one parent is affected, each offspring (regardless of sex) has an independent 50 per cent chance of inheriting the abnormal gene. Normal repeat lengths range from about 7 to 28 triplets. Individuals with 29–35 triplets will not develop HD but may pass an expanded allele to an offspring, while individuals with 40 or more triplets will develop HD. Repeat lengths of 36–39 triplets may or may not cause disease. The rate of mutation from a normal-length allele to an expanded one is low, so that most patients have an affected parent. Family history, however, can be obscured by multiple factors, including misdiagnosis of the parent, death of the parent before disease onset, adoption, and incorrect paternity. The repeat length does not remain stable at meiosis. In HD, the number of CAG triplets is more likely to increase when the gene is transmitted by fathers. As the number of repeats increases, the age at onset is earlier. Thus, paternal transmission is often associated with 'anticipation', earlier onset in the subsequent generation; most individuals with childhood onset have affected fathers.[11]

The pathogenesis of HD is not well understood but appears to be multifaceted. The gene, *huntingtin*, with the expanded repeat is expressed as a protein, huntingtin. The CAG repeat expansion is translated as an expanded polyglutamine tract, which appears to have neurotoxic properties. The region of the huntingtin protein with the polyglutamine tract may be cleaved from the rest of the protein and adopt an abnormal configuration. This in turn is thought to lead to disruption of cellular functions, including transcriptional machinery, protein degradation processes, and cellular transport. It is also possible that the expansion mutation results in a partial loss of the normal functions of huntingtin, one of which is the stimulation of the neurotrophic factor BDNF; decreased BDNF may contribute to neurotoxicity.[12]

Diagnosis

The most difficult aspect of diagnosis is to think of HD in the differential. Diagnosis remains dependent on a thorough psychiatric history, including a detailed family history and history of changes in social adjustment, mental state examination, including a cognitive examination, and neurological examination. The features vary, depending on how long the patient has been ill.[13] Once the disease is suspected, genetic testing, available through many commercial laboratories, provides the definitive diagnosis.

Diagnosis of patients with early symptoms

Patients with HD who initially consult psychiatrists present with a variety of psychiatric syndromes, including depression, bipolar disorder, obsessive–compulsive disorder, schizophrenia, or excessive anxiety. Irritability is common with any of these or may appear outside the context of another syndrome. These psychiatric syndromes are clinically indistinguishable from idiopathic disorders and may be the only manifestation of HD for several years. It is during this prodromal phase that patients often commit suicide; this may occur even if the patient is unaware of his risk for HD.[14] Presenting symptoms and problems with functioning at work or at home must often be elicited from an informant; the patient may minimize them, be embarrassed, or even unaware of them. These include declines in work speed or accuracy, which may have resulted in demotion or warnings from superiors; a tendency to become irritated or physically aggressive in response to annoying stimuli that would not have elicited such a response in the past; and a decreased interest in activities. Most of these symptoms and behaviours are common in psychiatric disorders, but the cognitive inefficiency and irritability may seem to be out of proportion, relative to the patient's other symptoms. On cognitive examination, the patient may have difficulty recalling dates of important life events and more difficulty than expected with 'serial sevens'. Usually, the cognitive changes are easier to notice after the psychiatric disorder is treated, which can usually be accomplished using standard medications. However, unlike idiopathic disorders, cognitive inefficiency and difficulties at work, apathy (if present), and sometimes irritability remain even after the patient's mood, energy, and sleep patterns have improved. When this happens in the course of treatment of depression, a dementia work-up should be considered and the family history further scrutinized through hospital records and other family informants.

On neurological examination, motor restlessness is usually present but is easily misinterpreted as a manifestation of anxiety. Motor signs may be subtle: slightly slow saccadic eye movements,[15] writhing movements of the protruded tongue or of the fingertips when the arms are held at 90°, or mild disdiadochokinesis.

Diagnosis can be further complicated by the apparent lack of a family history of HD. The family may not have been informed about the affected parent's diagnosis, or may know only that a parent died in a psychiatric institution or committed suicide. In other cases the paternity is uncertain. If the family history is actually negative (this is quite uncommon) or unobtainable (often the case for adopted individuals who frequently present in childhood), the diagnosis may be confirmed by testing for the HD gene expansion.

(a) Diagnosis in childhood and adolescence

When HD starts in childhood or early adolescence,[16] motor signs include parkinsonian-like motor slowness of voluntary movement, with lead pipe or cogwheel rigidity and very slow saccades. Occasional children have a coarse tremor; later myoclonus is seen. Cognitively, the rate of learning in school slows, handwriting

deteriorates, and interest in school and social activities declines.[17] Of the patients who present with a schizophrenic syndrome, most are adolescents. Psychosis and loss of cognitive capacity may be the only clinical features for several years before motor impairment begins. Children with HD often have seizures, which are usually grand mal. Sometimes myoclonus is mistaken for seizures.

(b) The importance of early diagnosis

Even though it can be difficult, it is important to make the diagnosis of HD as early as possible, particularly in employed persons. Poor function at work (or in schoolwork or household duties) occurs early, and patients can lose their jobs, often on suspicion of drug or alcohol abuse. This is usually avoided if the diagnosis is made known to the family and employer, allowing modification of the work environment or retirement on the basis of disability. Prompt diagnosis does not mean that the patient needs to be informed of the diagnosis at that same time. Some patients are too depressed to do this safely; others indicate that they do not wish to be told. Treatment can usually proceed despite the patient's reluctance to label the disorder.

Diagnosis of patients with well-established signs and symptoms

After a few years of illness, diagnosis is easier. The signs and symptoms will have worsened, and usually the motor disorder is obvious. A typical patient who has been ill for about 5 to 7 years is unable to work or manage finances, but lives at home and is able to manage personal needs. Some patients remain active and energetic, continuing to participate as fully in life as their cognitive and motor disabilities allow; others are apathetic most of the time, but irritable when disturbed; still others have severe depression with delusions, obsessions, or compulsions, and most are anxious and easily upset by changes of routine. An uncommon, but very troublesome, feature of HD is sexual abnormality. While most patients become impotent or uninterested in sex, a few are hypersexual and may develop paraphilias.[18] It is important to inquire about these specifically because neither the patient nor spouse will likely mention it.

Cognitively, patients complain of forgetfulness and becoming easily distracted. Thinking is slow; patients have difficulty following a conversation and cannot complete a multistaged task. On cognitive examination, Mini Mental State Examination scores[19] may still be above the 23 cut-off score, but serial sevens will be very poor, and one or two items will be missed on recalling words after a distraction. On neuropsychological testing, IQ will be lower than expected for education, and there will be difficulty learning word lists and performing tests that require changing sets.

Most patients will have obvious involuntary choreic movements, as well as difficulty with control of voluntary motor movements, as seen by clumsiness, slowness, dysarthria, and an unsteady gait. The involuntary movements will wax and wane with the level of arousal; it can be worsened by performing serial 7 s or by fine motor tasks. Speech will have an irregular staccato, often laboured, quality. Saccadic eye movements will be slow or irregular, and the patient will be obviously clumsy on diadochokinesis and finger-thumb tapping, although finger-to-nose testing is normal. Gait will be wide based and irregular, with difficulty with tandem walking. Reflexes are usually brisk, and a history of falls can be elicited.

Diagnosis of patients with advanced disease

After 10 years of illness, dementia is more severe, with poor performance on all aspects of the cognitive examination except naming. Speech is dysfluent with long lapses between the examiner's question and the patient's reply, rather like Brocca's aphasia. Some patients will be almost unable to speak, although language comprehension is relatively preserved. Patients (if they are cooperative) can carry out simple commands and will recognize relatives and nursing staff. Patients may be irritable, particularly when their verbal requests cannot be understood or when routines are altered. Psychiatric syndromes are more difficult to discern, but most can be diagnosed by observing behaviour such as hoarding, sleeplessness, or diurnal variation in mood. Physical disabilities are much worse. Patients often need to be fed, toileted, and helped with most daily needs. They have difficulty walking and may fall, causing further disability through broken limbs or subdural haematomas. Chorea often stabilizes or subsides,[13] but the ability to carry out voluntary movements becomes seriously handicapping. If they survive long enough, patients become unable to initiate speech, swallow with great difficulty, are unable to walk, and have such severely rigid muscle tone that they may be nearly unable to move their bodies. Clonus and positive Babinski signs are present. Patients in this sort of 'persistent vegetative state'[20] are difficult to distinguish from individuals with other advanced movement disorders or dementias; as in early disease, diagnosis will depend on eliciting a family history or genetic testing.

Differential diagnosis

The differential diagnosis of HD is extensive,[3] but only a few of the disorders for which it can be mistaken are common.[2] These include other dementias, other movement disorders, and other psychiatric disorders. The most frequent subcortical dementia is **Parkinson's disease**, which has a similar motor slowness, but a pill-rolling tremor and festinating gait are rare in HD. The dementia associated with **late-life depression** can look very similar to HD, including motor slowness. **Alzheimer's disease** is easily distinguished by the lack of motor signs during the first several years of illness and more prominent difficulty with memory and language, as opposed to attention and calculation. Perhaps most difficult to distinguish clinically are the **frontotemporal dementias**, which present with prominent behavioural disturbances and a positive family history. The clinical presentation may be insufficient to distinguish these various dementias in patients with advanced disease, since they all progress to a persistent vegetative state. The family history and the duration of illness (which is longer for HD than for Alzheimer's disease or frontotemporal dementia) can be helpful.

Several other diseases classified as **movement disorders** include all the features of the subcortical triad. They often have an autosomal dominant inheritance pattern and expansion of unstable trimeric repeat sequences. These include Fahr's syndrome (calcification of the striatum), some forms of spinocerebellar degeneration, chorea acanthocytosis, Huntington's disease-like 2 (HDL2), and dentatorubropallidolusian atrophy (DRPLA). The latter three disorders, while much rarer than HD, can look so similar that they can only be distinguished by genetic testing and by a blood smear for acanthocytes. The most common movement disorder that resembles HD is **tardive dyskinesia**. Patients with HD occasionally have several years of hallucinations and delusions before the

movement disorder begins. If they have been treated with neuroleptics, the subsequent onset of involuntary movements can be mistaken for tardive dyskinesia. On the other hand, the choreoathetotic involuntary movements of severe tardive dyskinesia may involve the trunk and extremities as well as the face and can be mistaken for HD. Usually, it is possible to distinguish the patients with tardive dyskinesia by their normal saccadic eye movements, normal tandem gait, and fluid and fluent speech.[21] However, genetic testing may be necessary in some cases. Wilson's disease also presents with the subcortical triad and should be considered when neither of the parent is affected. It is recessively inherited, so that the only affected relatives are siblings. Very late-onset HD may be diagnosed as 'senile chorea' because the family history appears to be negative. Family members will also present symptoms only late in life and may have died before their manifestation.

The differential diagnosis of nearly all **psychiatric disorders** includes HD, as described above.

Treatment and management

Currently, no treatment influences the course of illness of HD, but based on research on likely genetic mechanisms, clinical trials of agents protective against oxidants, excitotoxicity, and metabolic stress are underway. No agent yet tested has had a dramatic benefit. The development of biochemical assays and cell and animal models that mimic various aspects of HD can be used to screen for effective therapies.

It is possible to alleviate some of the symptoms of HD. Small doses of neuroleptics can be helpful in decreasing involuntary movements in the first stages of the illness, as can tetrabenazine and occasionally benzodiazepines. Doses of more than 5 mg. of haloperidol do not further decrease chorea and may worsen cognition and cause motor stiffness and slowing.[22] Persons with the advanced form of the disease are often unresponsive to neuroleptics. Treatment of psychiatric manifestations significantly improves quality of life for the patients and their families. Clinical experience suggests that depression, anxiety, and obsessive–compulsive disorder associated with HD usually respond to the pharmacological treatments used for the similar idiopathic disorders. SSRIs can be particularly helpful. Because some patients seem unaware of their depressed mood (just as they can be unaware of their involuntary movements) an informant is often needed to elicit the symptoms. It is also important to distinguish depression (from which the patient is miserable and sleepless) and apathy, which does not cause distress. Occasionally, mood and anxiety disorders are chronic and unresponsive to treatment. Severe, unresponsive depression can be treated successfully with electroconvulsive therapy.[23] Bipolar disorder in patients with HD does not usually respond to lithium, but may improve with carbamazepine or valproic acid. Valproic acid, serotonin specific reuptake inhibitors, and low dose antipsychotic agents may also be helpful in the treatment of irritability. Lithium is difficult to administer because patients require high fluid intake and easily become lithium toxic if fluid intake is insufficient. In one case report, high doses of sertraline were effective for intractable aggression.[24] Schizophrenic symptoms can be difficult to treat. Sometimes a combination of an antipsychotic, including clozaril, with an antidepressant will prove helpful. Muscle rigidity and consequent contractions occur in late HD, causing pain and difficulty in positioning the patient to avoid pressure sores. Amantadine (which also has a positive effect on mood) can somewhat decrease the rigidity; chairs and beds must be padded and tailored to each patient's specific needs.

Family and environment

As with most dementias, psychopathology influences, and is influenced by, the patient's environment. Patients do best in a calm, highly predictable environment where cognitive expectations are not too complicated. When the environment is too taxing, patients become irritable, especially towards their family. HD seriously damages family relationships, which in turn affects the patient. The well spouse becomes responsible for supporting the family, caring for children and the patient, and making family and financial decisions. Spouses' lives are further complicated by patients' unwillingness to relinquish financial and family decision-making; patients usually make poor decisions that damage family relationships and finances. Some patients neglect their children or treat them badly. If the other parent cannot prevent this, it is wisest to remove the patient from the home. There is no research on the treatment of sexual aggression, which occasionally occurs in males, but the author has successfully treated a few males with depot progesterone. Supportive psychotherapy for the patient should focus on minimizing demoralization at lost abilities. Spouses can be helped to reorganize family life in ways that maximize the predictability of the patient's environment, diplomatically decrease patients' domestic responsibilities. It is crucial that the spouse has time away from the patient.

Helping persons at risk for HD

People at risk for HD vary in their abilities to deal with the burden of uncertainty, depending on their personal attributes and their experience with the illness in a relative. A few consult physicians for reassurance, but most avoid doctors unless they become ill, and even then many resist medical attention, claiming against all evidence that they are perfectly well. Currently, a minority of asymptomatic persons at risk for HD decide to have genetic testing, but these individuals, skewed towards those whose anxiety is lessened by planning for the future, have usually handled the test results well, regardless of its outcome.[25] When clinical trials are launched for individuals with the HD mutation who are without detectable symptoms, the incentive for presymptomatic testing will likely increase, with a concomitant change in the nature of individuals seeking testing. Foetal and pre-implantation genetic testing are now available in some centers, each with its own set of potentially complicated ethical and practical issues to be sorted out prior to testing.

Presymptomatic genetic testing of any sort should always be preceded by genetic counselling, provided either by a genetic counselor or by a clinician familiar with HD genetics and the potential practical and psychological consequences of both positive and negative test results. Counselling should include a discussion of the motivations for seeking testing, which may include decisions about childbearing, education, employment, finances, participation in clinical trials, or the potential at-risk status of offspring. Many individuals who come for testing have not seriously considered the possibility that they will test positive for the mutation, so that role playing about various outcome scenarios is important. Occasionally, persons request testing who have learned only recently that they are

at risk for HD. Others apply who are depressed or under unusual stress for other reasons. Such persons should be encouraged to delay testing until their situation becomes more settled. Finally, some people who request testing already have symptoms of HD, yet do not wish to have a diagnosis. Considerable care is required to decide how best to support such individuals, and family members or close friends of the person should be consulted.[26]

Further information

Wells, R. and Ashizawa, T. (2006). *Genetic Instabilities and Neurologic Diseases* (2nd edn). Academic Press, San Diego.

Bates, G. Harper, P. and Jones, L. (2002). *Huntington's disease* (3rd edn). Oxford University Press, Oxford.

Walker, F.O. (2007). Huntington's disease. *Lancet*, **369**, 218–28.

References

1. Huntington, G. (1872). On chorea. Reprinted in *Advances in Neurology*, **1**, 33–5 (1973).
2. Folstein, S. (1989). *Huntington's disease. A disorder of families*. Johns Hopkins University Press, Baltimore, MD.
3. Harper, P. (ed.) (1991). *Huntington's disease*, Vol. 22. W.B. Saunders, London.
4. Hedreen, J.C. and Folstein, S.E. (1995). Early loss of neostriatal striosome neurons in Huntington's disease. *Journal of Neuropathology and Experimental Neurology*, **54**, 105–20.
5. Aylward, E.H, Codori, A.M., Rosenblatt, A., *et al.* (2000). Rate of caudate atrophy in presymptomatic and symptomatic stages of Huntington's disease. *Movement Disorders*, **15**, 552–60.
6. Rosas, H.D., Hevelone, N.D., Zaleta, A.K., *et al.* (2005). Regional cortical thinning in preclinical Huntington disease and its relationship to cognition. *Neurology*, **65**, 745–7.
7. Paulsen, J.S., Magnotta, V.A., Mikos, A.E., *et al.* (2006). Brain structure in preclinical Huntington's disease. *Biological Psychiatry*, **59**, 57–63.
8. Reading, S.A., Yassa, M.A., Bakker, A., *et al.* (2005). Regional white matter change in presymptomatic Huntington's disease: a diffusion tensor imaging study. *Psychiatry Research*, **140**, 55–62.
9. Folstein, S.E, Chase, G.A., Wahl, W.E., *et al.* (1987). Huntington disease in Maryland: clinical aspects of racial variation. *American Journal of Human Genetics*, **41**, 168–79.
10. Huntington's Disease Collaborative Research Group. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell*, **72**, 971–83.
11. Ranen, N.G., Stine, O.C., Abbott, M.H., *et al.* (1995). Anticipation and instability of (CAG)*n* repeats in IT-15 in parent–offspring pairs with Huntington's disease. *American Journal of Human Genetics*, **57**, 593–602.
12. Ross, C.A., Russell, L.M., Becher, M.W., *et al.* (1998). Pathogenesis of neurodegenerative diseases associated with expanded glutamine repeats: new answers, new questions. *Progress in Brain Research*, **117**, 398–419.
13. Folstein, S.E., Leigh, R., Parhad, I.M., *et al.* (1986). The diagnosis of Huntington's disease. *Neurology*, **36**, 1279–83.
14. Folstein, S.E., Abbott, M.H., Franz, M.L., *et al.* (1983). The association of affective disorder with Huntington's disease in a case series and in families. *Psychological Medicine*, **13**, 537–42.
15. Lasker, A.D. and Zee, D.S. (1994). Ocular motor abnormalities in Huntington's disease. *Vision Research*, **34**, 3639–45.
16. Kosky, R. (1981). Children and Huntington's disease: some clinical observations of children at risk. *The Medical Journal of Australia*, **1**, 405–7.
17. Nance, M. (1997). Genetic testing of children at risk for Huntington's disease. *Neurology*, **49**, 1048–53.
18. Federoff, J.O., Peyser, C.E., Franz, M.L., *et al.* (1994). Sexual disorders in Huntington's disease. *The Journal of Neuropsychiatry and Clinical Neuroscience*, **6**, 147–53.
19. Folstein, M.F., Folstein, S.E., and McHugh, P.R. (1975). 'Mini-Mental State': a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **2**, 189–98.
20. Walshe, T.M. and Leonard, C. (1985). Persistent vegetative state: extension of the syndrome to include chronic disorders. *Archives of Neurology*, **42**, 1045–7.
21. David, A.S., Jeste, D.V., Folstein, M.F., *et al.* (1987). Voluntary movement dysfunction in Huntington's disease and tardive dyskinesia. *Acta Neurologica Scandinavia*, **75**, 130–9.
22. Barr, A.N., Fischer, J.H., Koller, W.C., *et al.* (1988). Serum haloperidol concentration and choreiform movements in Huntington's disease. *Neurology*, **38**, 84–8.
23. Ranen, N.G., Peyser, C.E., and Folstein, S.E. (1994). ECT as a treatment for depression in Huntington's disease. *Journal of Neuropsychiatry and Clinical Neurosciences*, **6**, 154–9.
24. Ranen, N.G., Lipsey, J.R., Treisman, G., *et al.* (1996). Sertraline in the treatment of severe aggressiveness in Huntington's disease. *The Journal of Neuropsychiatry and Clinical Neuroscience*, **8**, 338–40.
25. Almqvist, E.W., Brinkman, R.R., Wiggins, S., *et al.* (2003). Psychological consequences and predictors of adverse events in the first 5 years after predictive testing for Huntington's disease. *Clinical Genetics*, **64**, 300–309.
26. Scourfield, J., Soldan, J., Gray, J., *et al.* (1997). Huntington's disease: psychiatric practice in molecular genetic prediction and diagnosis. *The British Journal of Psychiatry*, **170**, 146–9.

4.1.8 Vascular dementia

Timo Erkinjuntti

Introduction

Vascular dementia is the second most frequent cause of dementia.[1,2] Because vascular causes of cognitive impairment are common, may be preventable, and the patients could benefit from therapy, early detection, and accurate diagnosis of vascular dementia is desirable.[3]

Vascular dementia is not only multi-infarct dementia, but is related to other vascular mechanisms and pathological changes in the brain, and has other causes and clinical manifestations. Vascular dementia is not a disease, but a syndrome. The origin of this syndrome reflects complex interactions between vascular aetiologies (cerebrovascular disorders and vascular risk factors), changes in the brain (infarcts, white-matter lesions, atrophy), host factors (age, education), and cognition.[4–8]

Conceptual issues related to of vascular dementia include the definition of the cognitive syndrome (type, extent, and combination of impairments in different cognitive domains), and the vascular causes (vascular aetiologies and changes in the brain). Variations in these definitions has led to different estimates of point prevalence, to different groups of patients, and to reports of different types and distribution of brain lesions.[9–11] The cognitive syndrome of vascular dementia is characterized by predominate executive dysfunction rather than deficits in memory and language function.[12] Although the course of cognitive decline may be

stepwise, it is often slowly progressive, and may include periods of stability or even some improvement.

The relationship between vascular lesions in the brain and cognitive impairment is important, but which type, extent, side, site, and tempo of vascular lesions in the brain relates to different types of vascular dementia is not established in detail.[4–6,13]

Current criteria for vascular dementia are based on the concept of cerebral infarcts. For example the widely used NINDS-AIREN criteria include dementia, cerebrovascular disease, and a relationship between these two disorders. The main tools for the diagnosis include detailed history, neurological examination, mental state examination, relevant laboratory examinations, and preferably magnetic resonance imaging of the brain.

Vascular dementia research, until recently overshadowed by that into Alzheimer's disease, is now developing rapidly. There is great promise for intervention. Developments in classification, diagnosis, and treatment are likely.

Aetiology and pathophysiology

Aetiology

The main causes of vascular dementia are cerebrovascular disorders and their risk factors. The prevalent cerebrovascular disorders include large artery disease (artery-to-artery embolism, occlusion of an extra- or intracranial artery), cardiac embolic events, small-vessel disease (lacunar infarcts, ischaemic white-matter lesions) and haemodynamic mechanisms.[13–15] Less frequent causes include specific arteriopathies including cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADACIL) and cerebral amyloid angiopathy (CAA), haemorrhage (intracranial haemorrhage, subarachnoidal haemorrhage), haematological factors, venous disease, and hereditary disorders. There may be as yet undiscovered causes.

In most patients diagnosed with vascular dementia, several aetiological factors are involved. However, the roles these factors play have not been identified in detail, and it is not certain which of these mechanisms distinguish vascular dementia from cerebrovascular disease without dementia.[4,5,7,16,17]

Risk factors for vascular dementia can be divided into vascular factors (e.g. arterial hypertension, atrial fibrillation, myocardial infarction, coronary heart disease, diabetes, generalized atherosclerosis, lipid abnormalities, smoking), demographic factors (e.g. age, education), genetic factors (e.g. family history, individual genetic features), and stroke-related factors (e.g. type of cerebrovascular disease, site and size of stroke).[18,19] Hypoxic ischaemic events (cardiac arrhythmias, congestive heart failure, myocardial infarction, seizures, pneumonia) may be an important risk factor for incident dementia in patients with stroke.[20]

Changes in the brain

Vascular dementia is related to both ischaemic and non-ischaemic changes in the brain.[4,5,13,14] The ischaemic lesions include arterial territorial infarct, distal field (watershed) infarct, lacunar infarct, ischaemic white-matter lesions, and incomplete ischaemic injury. Incomplete ischaemic injury incorporates laminar necrosis, focal gliosis, granular atrophy, and incomplete white-matter infarction.[21,22] In addition, both focal (around the ischaemic lesion) and remote (disconnection, diaschisis) functional ischaemic changes relate to vascular dementia, and the volume of functionally inactive tissue exceeds that of focal ischaemic lesions in vascular dementia.[23] Limitation in current clinical methods have hampered the detection of both incomplete ischaemic injury and functional ischaemic changes related to vascular dementia. Atrophy is the non-ischaemic factor related to vascular dementia. However, there are no methods to distinguish between ischaemic and degenerative causes of atrophy clinically.

Brain imaging findings

Work on the relationship between brain lesions and cognition in vascular dementia has used varying definitions and measures of cognitive impairment, varying techniques to reveal brain changes, and varying criteria for the selection of patients.[17]

CT and magnetic resonance imaging (**MRI**) studies on vascular dementia have shown that bilateral ischaemic lesions are important.[4,5,7,17] Some studies emphasize deep infarcts in the frontal and limbic areas, while others report cortical lesions especially in the temporal and parietal areas. There is disagreement about the number and volume of the infarcts, as well as the extent and location of atrophy. Diffuse and extensive white-matter lesions have been suggested as an important factor leading to functional disconnection of cortical brain areas. Some general conclusions on brain lesions in vascular dementia may be drawn.

1 There is no single pathological feature, but a combination of infarcts, ischaemic white-matter lesions of varying size and type, and atrophy of varying degree and site.

2 Infarcts associated with vascular dementia tend to be bilateral, multiple (more than two), and located in the dominant hemisphere and in the limbic structures (frontolimbic or prefrontal–subcortical and medial–limbic or medial–hippocampal circuits).

3 White-matter lesions on CT or magnetic resonance imaging (MRI) associated with vascular dementia are extensive, extending in periventricular white matter, and confluent to extending in the deep white matter.

4 It is doubtful whether a single small lesion on imaging can be accepted as evidence for vascular dementia.

5 Absence of cerebrovascular lesions on CT or MRI is contrary to a diagnosis of vascular dementia.

Pathophysiology

The extent to which pathological changes in the brain cause, compound, or only coexist with the vascular dementia syndrome is still uncertain. The vascular changes in the brain can be the main cause of cognitive impairment (as assumed in vascular dementia[24,25]), they can contribute to the clinical picture of other dementia syndromes including Alzheimer's disease (AD),[7,26] or they may be coincidental. The occurrence of infarcts may cause an earlier presentation of clinical symptoms in a brain in which there is existing and progressive Alzheimer's disease pathology.[26]

It is not certain which are the critical changes in the brain leading to the clinical picture of vascular dementia. The syndrome has been related to the volume of brain infarcts (with a critical threshold), the number of infarcts, the site of infarcts (bilateral, in strategic cortical or subcortical, or affecting white matter), to other ischaemic factors (incomplete ischaemic injury, delayed neuronal death, functional changes), to the atrophic changes (origin, location, extent), and finally to the additive effects of other pathologies

(Alzheimer's disease, Lewy body dementia, frontal lobe dementias). But it is uncertain which type, extent, side, site, and tempo of vascular lesions in the brain, and which combination with other pathologies, relate to vascular dementia.[4–6,13]

Classification and clinical criteria

Classification

Vascular dementia has been divided into subtypes on the basis of clinical, radiological, and neuropathological features. It is uncertain whether these subtypes are distinct disorders, with separate pathological and clinical features, and responses to therapy.[27] If homogenous subtypes could be identified the comparability of research studies would be greater and multicentre studies easier.[28]

The subtypes of vascular dementia included in most classifications include multi-infarct dementia (cortical lesions), small-vessel dementia or subcortical ischaemic vascular disease and dementia (SIVD) (subcortical deep lesions), and strategic infarct dementia.[12,14,27,29–33] Many include also hypoperfusion dementia.[12,14,30,34] Further suggested subtypes include haemorrhagic dementia, hereditary vascular dementia, and combined or mixed dementia (Alzheimer's disease with cerebrovascular disease).

DSM-IV[35] does not specify subtypes. ICD-10[36] includes six subtypes (acute onset, multi-infarct, subcortical, mixed cortical and subcortical, other, and unspecified). The NINDS-AIREN criteria[30] include, without detailed description, cortical vascular dementia, subcortical vascular dementia, Binswanger's disease, and thalamic dementia. In addition separate research criteria for subcortical vascular dementia, the SIVD, have been proposed.[37]

Main subtypes

Multi-infarct dementia or cortical vascular dementia, and small-vessel dementia or subcortical vascular dementia are the two common subtypes, although their frequencies vary in different series.[12,14,31]

Cortical vascular dementia relates to large-vessel disease, cardiac embolic events, and hypoperfusion. Infarcts are predominantly in the cortical and corticosubcortical arterial territories, and their distal fields (watershed). Typical clinical features are lateralized sensorimotor changes and the abrupt onset of cognitive impairment and aphasia.[31] A combination of different cortical neuropsychological syndromes has been suggested to occur in cortical vascular dementia.[38]

Subcortical vascular dementia, small-vessel dementia, the SIVD[33,37] incorporates the entities 'lacunar state' and 'Binswanger's disease'. It relates to small-vessel disease and hypoperfusion, with predominately lacunar infarcts, focal and diffuse ischaemic white-matter lesions, and incomplete ischaemic injury.[31,33,37–39] Clinically, small-vessel dementia is characterized by pure motor hemiparesis, bulbar signs, dysarthria, depression, and emotional lability, and especially deficits in executive functioning.[38–41]

Clinical criteria

Since the 1970s several clinical criteria for vascular dementia have been published.[11,42,43] The most widely used include those in DSM-IV,[35] ICD-10,[36] and NINDS-AIREN.[30]

The two cardinal elements of any clinical criteria for vascular dementia are the definition of the cognitive syndrome[44] and the definition of the cause.[11,43,45] All clinical criteria are consensus criteria, derived neither from prospective community-based studies on vascular factors affecting the cognition, nor on detailed natural histories.[28,30,42,43,46] All these criteria are based on the concept of ischaemic infarcts. They are designed to have high specificity, but have been poorly validated.[42,46] An important consequence of the different definitions of the dementia syndrome,[9,44] and the vascular cause,[10,11] is that the different diagnostic criteria identify different populations.

Table 4.1.8.1 The DSM-IV definition of vascular dementia

Focal neurological signs and symptoms (e.g. exaggeration of deep tendon reflexes, extensor plantar response, pseudobulbar palsy, gait abnormalities, weakness of an extremity, etc.)
or
Laboratory evidence of focal neurological damage (e.g. multiple infarctions involving cortex and underlying white matter)
The cognitive deficits cause significant impairment in social or occupational functioning and represent a significant decline from a previously higher level of functioning
The focal neurological signs, symptoms, and laboratory evidence are judged to be aetiologically related to the disturbance
The deficits do not occur exclusively during the course of delirium
Course characterized by sustained periods of clinical stability punctuated by sudden significant cognitive and functional losses

The DSM-IV definition of vascular dementia (Table 4.1.8.1) requires focal neurological signs and symptoms or laboratory evidence of focal neurological damage clinically judged to be related to the disturbance.[35] The course is specified by sudden cognitive and functional losses. The DSM-IV criteria do not detail brain imaging requirements. The DSM-IV definition of vascular dementia is reasonably broad and lacks detailed clinical and radiological guidelines.

The ICD-10 criteria[36] (Table 4.1.8.2) require unequal distribution of cognitive deficits, focal signs as evidence of focal brain damage, and significant cerebrovascular disease judged to be aetiologically related to the dementia. The criteria do not detail brain imaging requirements. The ICD-10 criteria specify six subtypes of vascular dementia (Table 4.1.8.3). The ICD-10 criteria for vascular dementia have been shown to be highly selective and only some of those fulfilling the general criteria for ICD-10 vascular dementia

Table 4.1.8.2 The ICD-10 criteria for vascular dementia

Unequal distribution of deficits in higher cognitive functions with some affected and others relatively spared. Thus, memory may be quite markedly affected while thinking, reasoning, and information processing may show only mild decline
There is evidence for focal brain damage, manifest as at least one of the following: unilateral spastic weakness of the limbs, unilaterally increased tendon reflexes, an extensor plantar response, pseudobulbar palsy
There is evidence from the history, examination, or test of significant cerebrovascular disease, which may reasonably be judged to be aetiologically related to the dementia (history of stroke, evidence of cerebral infarction)

Table 4.1.8.3 Characteristics of the vascular dementia subtypes in ICD-10

Acute onset (F01.0)
The dementia develops rapidly (i.e. usually within 1 month but within no longer than 3 months) after a succession of strokes, or (rarely) after a single large infarction
Multi-infarct (F01.1)
The onset of the dementia is more gradual (i.e. within 3-6 months) following a number of minor ischaemic episodes. Comments: it is presumed that there is an accumulation of infarcts in the cerebral parenchyma. Between the ischaemic episodes there may be periods of actual clinical improvement
Subcortical (F01.2)
A history of hypertension, and evidence from clinical examination and special investigations of vascular disease located in the deep white matter of the cerebral hemispheres, with preservation of the cerebral cortex.
Mixed cortical and subcortical (F01.3)
Mixed cortical and subcortical components of vascular dementia may be suspected from the clinical features, the results of investigation, or both
Other (F01.8)
Unspecified (F01.9)
In the ICD-10 criteria no specific diagnostic guidelines are given for these two vascular dementia sybtypes

can be classified into one of the subtypes.[11,45] The shortcoming of these criteria include lack of detailed guidelines (e.g. of unequal cognitive deficits and changes on neuroimaging), lack of aetiological criteria, and heterogeneity.[11,45]

The NINDS-AIREN research criteria for vascular dementia[30] include a dementia syndrome, cerebrovascular disease, and a relationship between these (Table 4.1.8.4). Cerebrovascular disease is defined by the presence of focal neurological lesions and brain imaging evidence of ischaemic changes in the brain. A relationship between dementia and cerebrovascular disorder is inferred from the onset of dementia within 3 months following a recognized stroke, or on abrupt deterioration in cognitive functions, or fluctuating stepwise progression of cognitive deficits. The criteria include a list of features consistent with the diagnosis, as well as a list of features that make the diagnosis uncertain or unlikely. Also, different levels of certainty of the clinical diagnosis (probable, possible, definite) are included. The NINDS-AIREN criteria recognize heterogeneity[47] of the syndrome and variability of the clinical course in vascular dementia, and highlight detection of ischaemic lesions and a relationship between lesion and cognition, as well as stroke and dementia onset.

The NINDS-AIREN criteria are currently most widely used in clinical drug trials on vascular dementia. In a neuropathological series, sensitivity of the NINDS-AIREN criteria was 58 per cent and specificity 80 per cent.[48] The criteria successfully excluded Alzheimer's disease in 91 per cent of cases, and the proportion of combined cases misclassified as probable vascular dementia was 29 per cent.[48] The inter-rater reliability of the NINDS-AIREN criteria is moderate to substantial ($\kappa = 0.46–0.72$).[49]

These three sets of criteria for vascular dementia are not interchangeable; they identify different numbers and clusters of patients. The DSM-IV criteria are less restrictive than the ICD-10 and NINDS-AIREN criteria.[11,50]

Table 4.1.8.4 The NINDS-AIREN criteria for probable vascular dementia

(I) The criteria for the clinical diagnosis of PROBABLE vascular dementia include *all* of the following
1 *Dementia*
2 *Cerebrovascular disease*, defined by the presence of focal signs on neurological examination, such as hemiparesis, lower facial weakness, Babinski sign, sensory deficit, hemianopia, dysarthria, etc. consistent with stroke (with or without history of stroke), and evidence of relevant CVD by brain imaging (CT or MRI) including multiple large-vessel strokes or a single strategically placed infarct (angular gyrus, thalamus, basal forebrain, PCA or ACA territories), as well as multiple basal ganglia and white-matter lacunes or extensive periventricular white-matter lesions, or combinations thereof
3 *A relationship between the above two disorders*, manifested or inferred *by* the presence of one or more of the following
(a) Onset of dementia within 3 months following a recognized stroke
(b) Abrupt deterioration in cognitive functions, or fluctuating stepwise progression of cognitive deficits
(II) Clinical features consistent with the diagnosis of PROBABLE vascular dementia include the following
(a) Early presence of a gait disturbance (small-step gait or *marche a petits-pas*, apraxic–ataxic or parkinsonian gait)
(b) History of unsteadiness and frequent unprovoked falls
(c) Early urinary frequency, urgency, and other urinary symptoms not explained by urological disease
(d) Personality and mood changes, abulia, depression, emotional incontinence, other subcortical deficits including psychomotor retardation and abnormal executive function
(III) Features that make the diagnosis of vascular dementia uncertain or unlikely include the following
(a) Early onset of memory deficit and progressive worsening of memory and other cognitive functions such as language (transcortical sensory aphasia), motor skills (apraxia), and perception (agnosia), in the absence of corresponding focal lesions on brain imaging
(b) Absence of focal neurological signs, other than cognitive disturbance
(c) Absence of cerebrovascular lesions on brain CT or MRI

CVD, cerebrovascular disease; PCA, posterior cerebral artery; ACA, anterior cerebral artery.

Vascular cognitive impairment

Vascular cognitive impairment (VCI) is currently considered the most recent modification of the terminology to reflect the all-encompassing effects of vascular disease or lesions on cognition and incorporates the complex interactions between vascular aetiologies, risk factors, and cellular changes within the brain and cognition.[51,52]

VCI refers to all aetiologies of CVD including vascular risks which can result in brain damage leading to cognitive impairment. The impairment encompasses all levels of cognitive decline, from the earliest deficits to a severe and broad dementia-like cognitive syndrome.[51,53] VCI cases that do not meet the criteria for dementia can also be labelled as VCI with no dementia, vascular CIND.[54]

VCI may include cases with cognitive impairment related to hypertension, diabetes, or atherosclerosis, transient ischaemic attacks, multiple corticosubcortical infarcts, silent infarct, strategic

infarcts, small-vessel disease with white-matter lesions an lacunae, as well as AS pathology with coexisting CVD.[55] The concept and definitions of VCI and vascular CIND are still evolving, but it seems clear that the diagnosis should not be confined to a single aetiology comparable to the traditional 'pure AD' concept.[51,52]

Clinical features

Cognitive syndrome

The cognitive syndrome of vascular dementia is characterized by memory deficit, dysexecutive syndrome, slowed information processing, and mood and personality changes. These features are found especially among patients with subcortical lesions. Patients with cortical lesions often have additional cortical neuropsychological syndromes.[38]

The memory deficit in vascular dementia is often less severe than in Alzheimer's disease. It is characterized by impaired recall, relatively intact recognition, and more benefit from cues.[56] The dysexecutive syndrome in vascular dementia includes impairment in goal formulation, initiation, planning, organizing, sequencing, executing, set-sifting and set-maintenance, as well as in abstracting.[12,38,56] The dysexecutive syndrome in vascular dementia relates to lesions affecting the prefrontal subcortical circuit including prefrontal cortex, caudate, pallidum, thalamus, and the thalamocortical circuit (capsular genu, anterior capsule, anterior centrum semiovale, and anterior corona radiata).[57] Typically, personality and insight are relatively preserved in mild and moderate cases of vascular dementia.

Features that make the diagnosis of vascular dementia disease uncertain or unlikely include early and progressive worsening of episodic memory, and other cognitive cortical deficits in the absence of corresponding focal lesions on brain imaging.[30]

Neurological findings

Frequent neurological findings indicating focal brain lesion early in the course of vascular dementia include mild motor or sensory deficits, decreased co-ordination, brisk tendon reflexes, Babinski's sign, visual field loss, bulbar signs including dysarthria and dysphagia, extrapyramidal signs (mainly rigidity and akinesia), disordered gait (hemiplegic, apraxic–ataxic, or small-stepped), unsteadiness, unprovoked falls, and urinary frequency and urgency.[30,31,39–41] Features that make the diagnosis of vascular dementia uncertain or unlikely include absence of focal neurological signs, other than cognitive disturbance.[30]

In cortical vascular dementia, typical clinical features are lateralized sensorimotor changes and abrupt onset of cognitive impairment and aphasia, and in subcortical vascular dementia disease pure motor hemiparesis, bulbar signs, dysarthria, disordered gait and unsteadiness.[31]

Behavioural and psychological symptoms of dementia

Depression, anxiety, emotional lability and incontinence, and other psychiatric symptoms are frequent in vascular dementia. Depression, abulia, emotional incontinence, and psychomotor retardation are especially frequent in subcortical vascular dementia disease.[12,38]

(a) Ischaemic scores

Cardinal features of vascular dementia disease are incorporated in the Hachinski Ischaemia Score[58] (Table 4.1.8.5). In a neuropathological series, stepwise deterioration (odds ratio, 6.0), fluctuating course (odds ratio, 7.6), history of hypertension (odds ratio, 4.3), history of stroke (odds ratio, 4.3), and focal neurological symptoms (odds ratio, 4.4) differentiated patients with definite vascular dementia from those with definite Alzheimer's disease.[59] Nocturnal confusion and depression did not discriminate. However, the ischaemia score was unable to differentiate the Alzheimer's disease patients with cerebrovascular disease from those with vascular dementia.

Table 4.1.8.5 Hachinski ischaemia score

Item	Score value
Abrupt onset	2
Stepwise deterioration	1
Fluctuating course	2
Nocturnal confusion	1
Relative preservation of personality	1
Depression	1
Somatic complaints	1
Emotional incontinence	1
History of hypertension	1
History of strokes	2
Evidence of associated atherosclerosis	1
Focal neurological symptoms	2
Focal neurological signs	2

Course and prognosis

Traditionally, vascular dementia has been characterized by a relative abrupt onset (days to weeks), a stepwise deterioration (some recovery after worsening), and fluctuating course (e.g. differences between days) of cognitive functions. These features are seen in patients with repeated lesions affecting cortical and corticosubcortical brain structures, i.e. large-vessel multi-infarct vascular dementia, and with watershed infarcts related to haemodynamic problems. However, in patients with small-vessel dementia, i.e. subcortical vascular dementia, the onset is more insidious and course more slowly progressive.[28,30,39,60]

The mean duration of vascular dementia is around 5 years.[2] In most studies survival is less than for the general population or those with Alzheimer's disease.[61,62] Surprisingly little is known about the rate and pattern of cognitive decline, either overall or among different subgroups of vascular dementia.[63] This underlines the lack of studies detailing the natural history of vascular dementia.

Diagnosis and differential diagnosis

The clinical evaluation of patients with memory impairment has two stages, the symptomatic diagnosis, i.e. evaluation of the type and extent of cognitive impairment, and the aetiological diagnosis, i.e. evaluation of vascular cause(s) and related factors. The symptomatic categories other than dementia include the more mild cognitive stages, i.e. VCI or vascular CIND, delirium, circumscribed

neuropsychological syndromes (e.g. aphasia) and functional psychiatric disorders (e.g. depression).[46] Stages of aetiological diagnosis include diagnosis of the specific causes, especially the potentially treatable conditions, evaluation of secondary factors able to affect the cognitive functioning, and more detailed differentiation between specific causes, especially that between vascular dementia disease and Alzheimer's disease.

Clinical evaluation

The cornerstone in the evaluation of a patient with suspected vascular dementia is detailed clinical and neurological history and examination, including interview of a close informant. Assessment of social functions, activities of daily living, as well as psychiatric and behavioural symptoms, is part of the basic evaluation. These patients are challenging and enough time should be allocated time for the consultation, often 40 to 60 min.

Mental status examination

Bedside mental status examination includes the Mini-Mental State Examination.[64] However, this has limitations as it emphasizes language, does not include timed elements and the recognition portion of the memory tests, is insensitive to mild deficits, and is influenced by education and age. Other proposed screening instruments for vascular dementia include a 10-word memory test with delayed recall, cube drawing test for copy, verbal fluency test (number of animals named in 1 min), Luria's alternating hand sequence, or finger rings and letter cancellation test (neglect).[30] Other test include the Clox and Exit designed to screen the dysexecutive features.[65]

Frequently a more detailed neuropsychological test is needed. It should cover the main cognitive domains including memory functions (working memory, episodic memory, semantic memory), abstract thinking, judgement, aphasia, apraxia, agnosia, orientation, attention, executive functions, and speed of information processing.[44,66]

Brain imaging

Brain imaging should be performed at least once during the initial diagnostic workout. MRI is preferred because it has high sensitivity and the ability to demonstrate medial temporal lobe and basal forebrain areas. Depending on the criteria of vascular dementia used, focal brain infarcts have been revealed in 70 to 100 per cent, and more extensive white-matter lesions in 70 to 100 per cent of cases.[13,25,30,67,68]

Single-photon emission CT and positron-emission tomography may reveal patchy reduction of regional blood flow and metabolism, as well as decreased white-matter flow and metabolism.[69]

Other investigations

Chest X-ray, electrocardiography, and screening laboratory tests are part of the basic evaluation.[15,70,71] In selected cases extended laboratory investigations, analysis of the cerebrospinal fluid, and EEG are performed, as well as examinations of the extra- and intracranial arteries and detailed cardiological investigations.[15,70,71]

In vascular dementia EEG is more often normal than in Alzheimer's disease, and if abnormal more frequently suggests a focal abnormality. Abnormalities increase with more severe intellectual decline both in vascular dementia disease and Alzheimer's disease.[60]

At present there is no specific laboratory test for vascular dementia. Tests may reveal risk factors and concomitant disorders such as hyperlipidaemia, diabetes, and cardiac abnormality.[60] Apolipoprotein E_4 is a established risk factor for Alzheimer's disease, but its relationship to vascular dementia has not been consistent.[72] Determination of apolipoprotein E status is currently not part of clinical evaluation in vascular dementia.

Differential diagnosis of vascular dementia disease

(a) Alzheimer's disease

Typical Alzheimer's disease is characterized by insidious onset and slowly progressive intellectual deterioration, absence of symptoms and signs indicating focal brain damage, and absence of any other specific disease affecting the brain.[73] Alzheimer's disease has typical clinical stages ranging from early changes to profound dementia.[74,75]

When patients with vascular dementia have a clinical history, neurological examination, and brain imaging findings compatible with ischaemic changes of the brain, the differentiation from Alzheimer's disease can be made clinically.[25]

Diagnostic problems arise when Alzheimer's disease is combined with cerebrovascular disease. Difficult clinical problems include stroke unmasking Alzheimer's disease in patients with post-stroke dementia, insidious onset, and/or slow progressive course in vascular dementia patients, and cases where it is difficult to assess the role of white-matter lesions or of infarcts found on neuroimaging. A solution to recognize patients with Alzheimer's disease and cerebrovascular disease would be to discover reliable biological markers for clinical AD. Other potential markers include early prominent episodic memory impairment, early and significant medial temporal lobe atrophy on magnetic resonance imaging, bilateral parietal hypoperfusion on single-photon emission computed tomography and low concentrations of cerebrospinal fluid amyloid peptides with high tau protein concentrations. The distinction would be also less difficult if there were more detailed knowledge of the sites, type, and extent of ischaemic brain changes critical for vascular dementia, and the extent and type of medial temporal lobe atrophy critical for Alzheimer's disease.

Other important conditions to be differentiated from vascular dementia include normal pressure hydrocephalus,[76] frontal lobe tumours and other intracranial masses,[15] Lewy body disease,[77] frontotemporal degenerations,[78] Parkinson's disease and dementia,[79] progressive supranuclear palsy,[80] and multisystem atrophy.[81]

Epidemiology

Vascular dementia is the second most common cause of dementia accounting for 10 to 50 per cent of cases, depending on the geographic location, patient population, and clinical methods used.[1,2] The prevalence of vascular dementia is from 1.2 to 4.2 per cent of persons aged 65 years and older, and the incidence is 6 to 12 cases per 1000 persons aged over 70 years per year.[2] The prevalence and the incidence of vascular dementia disease increases with increasing age, and men seem to have a higher prevalence of vascular dementia than women. Epidemiology of vascular dementia has been affected by variations in the definition of the disorder, the clinical criteria used, and the clinical methods applied.[18,82,83]

The frequency of vascular dementia disease has been higher than previously reported in recent series comprising older subjects.[10] Stroke and cerebrovascular disorders relate also to a high risk of cognitive impairment and dementia.[24,84] Finally, vascular factors such as stroke and white-matter lesions have a clinical effect on Alzheimer's disease.[26] Thus, vascular factors may even be the leading cause of cognitive impairment worldwide especially when cognitive impairment as opposed to dementia is considered.[51,85]

Treatment

The objectives of targeted treatment of vascular dementia include symptomatic improvement of core symptoms (e.g. cognitive, behavioural), slowing progression of the disorder, and treatment of secondary factors affecting cognition (e.g. depression, anxiety, agitation).

A number of drugs have been studied in the symptomatic treatment of vascular dementia including cerebro- and vasoactive drugs, nootropics, and some calcium antagonists, but largely these studies have shown negative results.[86] Studies on symptomatic improvement in vascular dementia have mostly had small numbers, short treatment periods, variations in diagnostic criteria and tools, mixed populations, and have had variation in clinical endpoints applied.

First nimodipine,[87] memantine,[88] and propentofylline[89] raised expectations for a symptomatic treatment of vascular dementia. However, all the studies failed to fulfil the current requirement by the regulators for and treatment indication in vascular dementia.[90,91]

More recently cholinesterase inhibitors (donepezil, galantamine, rivastigmine) have been tested in large randomized controlled trials in patients with probable vascular dementia. All showed significant cognitive improvement compared to placebo, but failed to show significantly better global outcome with the Alzheimer-type measures. Accordingly, none of the acetylcholinesterase inhibitors have received marketing authorization for the treatment of vascular dementia.[91] Patients with Alzheimer's disease with coexisting cerebrovascular disease show good benefit from galantamine.[92]

Possibilities for prevention

For primary prevention the target is the brain at risk of cerebrovascular disease and cognitive impairment. The methods relate to the treatment of putative risk factors of vascular dementia, and the promotion of potential protective factors. Risk factors include those related to cerebrovascular disorders and stroke, to vascular dementia, to post-stroke dementia, to white-matter lesions, and to cognitive impairment or dementia, and also those related to Alzheimer's disease.[8] The vascular risk factors include arterial hypertension, atrial fibrillation, myocardial infarction, coronary heart disease, diabetes, generalized atherosclerosis, lipid abnormalities, and smoking. The demographic factors include age and education. One putative protective factor is oestrogen.[93]

Knowledge of effects of primary prevention on these risk factors in populations free of cognitive impairment is still scant.[8,94] In a European study, treatment of mild systolic hypertension decreased the incidence of dementia.[95] Positive effects in primary prevention of stroke support the idea that action on vascular risk factors could reduce the numbers of patients with vascular dementia.

For secondary prevention the target is the brain already affected by cerebrovascular disease and at risk of vascular dementia. Actions include diagnosis and treatment of acute stroke in order to limit the extent of ischaemic brain changes, prevention of recurrence of stroke, and treatment of risk factors. Treatment is guided by the aetiology of cerebrovascular disorder such as large artery disease (e.g. aspirin, dipyridamole, carotid endarterectomy), cardiac embolic events (e.g. anticoagulation, aspirin), small-vessel disease (e.g. antiplatelet therapy), and haemodynamic mechanisms (e.g. control of hypotension and cardiac arrhythmias).[15,29,46] A recent large study showed some benefit from perindopril and indapamide in the prevention of post-stroke dementia.[96] Hypoxic ischaemic events (cardiac arrhythmias, congestive heart failure, myocardial infarction, seizures, pneumonia) are an important risk factor for incident dementia in patients with stroke and should be taken into account in the secondary prevention of vascular dementia.[20]

Detailed knowledge of the effects of secondary prevention of vascular dementia is lacking. In a small series of patients with established vascular dementia, control of high arterial blood pressure,[97] cessation of smoking,[97] and use of aspirin[98] improved or stabilized cognition. It has been suggested that lowering of plasma viscosity could also have an effect in vascular dementia.[99] The absence of progressive cognitive decline in patients receiving placebo in treatment trials of vascular dementia may also reflect an effect of intensified risk factor control.[89]

Further information

Erkinjuntti, T., and Gauthier, S. (eds). (2002). *Vascular cognitive impairment*. Martin Duniz Publishers, London, England.

Vascular burden of the brain. *International Psychogeriatrics* (Suppl 1) 2003.

Bowler, J. and Hachinski, V. (eds.) (2003). *Vascular cognitive impairment preventable dementia*. pp. 176–92, Oxford University Press, Oxford, UK.

Mohr, J.P., Choi, D.W., Grotta, J.C., *et al.* (2004). *Stroke: Pathophysiology, Diagnosis, and Management*, (4th edn), Philadelphia, Elsevier.

Burns, A., O'Brien, J. and Ames, D. (2005). *Dementia* (3rd edn). Edward Arnold (Publishers), London, UK.

Growdon, J.H. and Rossor, M.N. (eds.) (2007). *Blue books of practical neurology, The Dementias 2*. Butterworth Heineman Elsevier, Philadephia, PA.

Ritchie, C.W., Ames, D., Masters, C.L., *et al.* (2007). *Therapeutic strategies dementia*. Clinical Publishing, Atlas Medical Publishing Ltd, Oxford, UK.

References

1. Rocca, W.A., Hofman, A., Brayne, C., *et al.* (1991). The prevalence of vascular dementia in Europe: facts and fragments from 1980–1990 studies. EURODEM-prevalence research group. *Annals of Neurology*, **30**, 817–24.
2. Hebert, R. and Brayne, C. (1995). Epidemiology of vascular dementia. *Neuroepidemiology*, **14**, 240–57.
3. Bowler, J.V. and Hachinski, V. (1995). Vascular cognitive impairment: a new approach to vascular dementia. *Baillières Clinical Neurology*, **4**, 357–76.
4. Tatemichi, T.K. (1990). How acute brain failure becomes chronic. A view of the mechanisms and syndromes of dementia related to stroke. *Neurology*, **40**, 1652–9.
5. Chui, H.C. (1989). Dementia: a review emphasizing clinicopathologic correlation and brain–behavior relationships. *Archives of Neurology*, **46**, 806–14.

6. Desmond, D.W. (1996). Vascular dementia: a construct in evolution. *Cerebrovascular and Brain Metabolism Reviews*, **8**, 296–325.
7. Pasquier, F. and Leys, D. (1997). Why are stroke patients prone to develop dementia? *Journal of Neurology*, **244**, 135–42.
8. Skoog, I. (1998). Status of risk factors for vascular dementia. *Neuroepidemiology*, **17**, 2–9.
9. Pohjasvaara, T., Erkinjuntti, T., Vataja, R., *et al.* (1997). Dementia three months after stroke. Baseline frequency and effect of different definitions of dementia in the Helsinki Stroke Aging Memory Study (SAM) cohort. *Stroke*, **28**, 785–92.
10. Skoog, I., Nilsson, L., Palmertz, B., *et al.* (1993). A population-based study of dementia in 85-year-olds. *New England Journal of Medicine*, **328**, 153–8.
11. Wetterling, T., Kanitz, R.D., and Borgis, K.J. (1996). Comparison of different diagnostic criteria for vascular dementia (ADDTC, DSM-IV, ICD-10, NINDS-AIREN). *Stroke*, **27**, 30–6.
12. Cummings, J.L. (1994). Vascular subcortical dementias: clinical aspects. *Dementia*, **5**, 177–80.
13. Erkinjuntti, T. (1996). Clinicopathological study of vascular dementia. In *Vascular dementia. Current concepts* (eds I. Prohovnik, J. Wade, S. Knezevic, T.K. Tatemichi, and T. Erkinjuntii), pp. 73–112. Wiley, Chichester.
14. Brun, A. (1994). Pathology and pathophysiology of cerebrovascular dementia: pure subgroups of obstructive and hypoperfusive etiology. *Dementia*, **5**, 145–7.
15. Amar, K. and Wilcock, G. (1996). Vascular dementia. *British Medical Journal*, **312**, 227–31.
16. Pantoni, L. and Garcia, J.H. (1995). The significance of cerebral white matter abnormalities 100 years after Binswanger's report. A review. *Stroke*, **26**, 1293–301.
17. Erkinjuntti, T. and Hachinski, V.C. (1993). Rethinking vascular dementia. *Cerebrovascular Disease*, **3**, 3–23.
18. Skoog, I. (1994). Risk factors for vascular dementia: a review. *Dementia*, **5**, 137–44.
19. Gorelick, P.B. (1997). Status of risk factors for dementia associated with stroke. *Stroke*, **28**, 459–63.
20. Moroney, J.T., Bagiella, E., Desmond, D.W., *et al.* (1996). Risk factors for incident dementia after stroke. Role of hypoxic and ischemic disorders. *Stroke*, **27**, 1283–9.
21. Pantoni, L. and Garcia, J.H. (1997). Pathogenesis of leukoaraiosis: a review. *Stroke*, **28**, 652–9.
22. Englund, E., Brun, A., and Alling, C. (1988). White matter changes in dementia of Alzheimer's type. Biochemical and neuropathological correlates. *Brain*, **111**, 1425–39.
23. Mielke, R., Herholz, K., Grond, M., *et al.* (1992). Severity of vascular dementia is related to volume of metabolically impaired tissue. *Archives of Neurology*, **49**, 909–13.
24. Tatemichi, T.K., Paik, M., Bagiella, E., *et al.* (1994). Risk of dementia after stroke in a hospitalized cohort: results of a longitudinal study. *Neurology*, **44**, 1885–91.
25. Erkinjuntti, T., Haltia, M., Palo, J., *et al.* (1988). Accuracy of the clinical diagnosis of vascular dementia: a prospective clinical and post-mortem neuropathological study. *Journal of Neurology, Neurosurgery, and Psychiatry*, **51**, 1037–44.
26. Snowdon, D.A., Greiner, L.H., Mortimer, J.A., *et al.* (1997). Brain infarction and the clinical expression of Alzheimer disease. The Nun Study. *Journal of the American Medical Association*, **277**, 813–17.
27. Wallin, A. and Blennow, K. (1994). The clinical diagnosis of vascular dementia. *Dementia*, **5**, 181–4.
28. Chui, H.C., Victoroff, J.I., Margolin, D., *et al.* (1992). Criteria for the diagnosis of ischemic vascular dementia proposed by the State of California Alzheimer's disease diagnostic and treatment centers. *Neurology*, **42**, 473–80.
29. Konno, S., Meyer, J.S., Terayama, Y., *et al.* (1997). Classification, diagnosis and treatment of vascular dementia. *Drugs & Aging*, **11**, 361–73.
30. Roman, G.C., Tatemichi, T.K., Erkinjuntti, T., *et al.* (1993). Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Work Group. *Neurology*, **43**, 250–60.
31. Erkinjuntti, T. (1987). Types of multi-infarct dementia. *Acta Neurologica Scandinavica*, **75**, 391–9.
32. Loeb, C. and Meyer, J.S. (1996). Vascular dementia: still a debatable entity? *Journal of the Neurological Sciences*, **143**, 31–40.
33. Roman, G.C., Erkinjuntti, T., Wallin, A., *et al.*(2002). Subcorticalischaemic vascular dementia. *Lancet Neurology,* **1,** 426–36.
34. Sulkava, R. and Erkinjuntti, T. (1987). Vascular dementia due to cardiac arrhythmias and systemic hypotension. *Acta Neurologica Scandinavica*, **76**, 123–8.
35. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
36. World Health Organization. (1993). *ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*. World Health Organization, Geneva.
37. Erkinjuntti, T., Inzitari, D., Pantoni, L., *et al.* (2000). Research criteria for subcortical vascular dementia in clinical trials. *Journal of Neural Transam,* **59**, 23–30.
38. Mahler, M.E. and Cummings, J.L. (1991). The behavioural neurology of multi-infarct dementia. *Alzheimer's Disease and Associated Disorders*, **5**, 122–30.
39. Roman, G.C. (1987). Senile dementia of the Binswanger type. A vascular form of dementia in the elderly. *Journal of the American Medical Association*, **258**, 1782–8.
40. Babikian, V. and Ropper, A.H. (1987). Binswanger's disease: a review. *Stroke*, **18**, 2–12.
41. Ishii, N., Nishihara, Y., and Imamura, T. (1986). Why do frontal lobe symptoms predominate in vascular dementia with lacunes? *Neurology*, **36**, 340–5.
42. Rockwood, K., Parhad, I., Hachinski, V., *et al.* (1994). Diagnosis of vascular dementia: consortium of Canadian centres for clinical cognitive research consensus statement. *Canadian Journal of Neurological Sciences*, **21**, 358–64.
43. Erkinjuntti, T. (1994). Clinical criteria for vascular dementia: the NINDS-AIREN criteria. *Dementia*, **5**, 189–92.
44. Erkinjuntti, T., Ostbye, T., Steenhuis, R., *et al.* (1997). The effect of different diagnostic criteria on the prevalence of dementia. *New England Journal of Medicine*, **337**, 1667–74.
45. Wetterling, T., Kanitz, R.D., and Borgis, K.J. (1994). The ICD-10 criteria for vascular dementia. *Dementia*, **5**, 185–8.
46. Erkinjuntti, T. (1997). Vascular dementia: challenge of clinical diagnosis. *International Psychogeriatrics*, **9**(Suppl. 1), 51–8.
47. Erkinjuntti, T. (1994). Clinical criteria for vascular dementia: the NINDS-AIREN criteria. *Dementia*, **5**, 189–92.
48. Gold, G., Giannakopoulos, P., Montes-Paixao, J.C., *et al.* (1997). Sensitivity and specificity of newly proposed clinical criteria for possible vascular dementia. *Neurology*, **49**, 690–4.
49. Lopez, O.L., Larumbe, M.R., Becker, J.T., *et al.* (1994). Reliability of NINDS-AIREN clinical criteria for the diagnosis of vascular dementia. *Neurology*, **44**, 1240–5.
50. Verhey, F.R., Lodder, J., Rozendaal, N., *et al.* (1996). Comparison of seven sets of criteria used for the diagnosis of vascular dementia. *Neuroepidemiology*, **15**, 166–72.
51. O'Brien, J.T. (2003). Vascular cognitive impairment. *Lancet Neurology*, **2**, 89–98.
52. Roman, G.C., Sachdev, P., Royall, D.R., *et al.* (2004). Vascular cognitive disorder: a new diagnostic category updating vascular cognitive impairment and vascular dementia. *Journal of Neurol Sc,* **226**, 81–7.
53. Bowler, J.V., Steenhuis, R., and Hachinski, V. (1999). Conceptual background of vascular cognitive impairment. *Alzheimer Disease and Associated Disorders*, **13**, S30–7.

54. Rockwood, K., Wenzel, C., Hachinski, V., *et al.* (2000). Prevalence and outcomes of vascular cognitive impairment. *Neurology,* **54**, 447–51.
55. Kalaria, R.N., Kenny, R.A., Ballard, C.G., *et al.* (2004). Towards defining the neuropathological substrates of vascular dementia. *Journal of Neurol Sci*; **226**, 75–80.
56. Desmond, D.W., Erkinjuntti, T., Sano, M., *et al.* (1999). *The cognitive syndrome of vascular dementia: implications for clinical trials. Alzh Dis Assoc Disord*, **13** (Suppl 3), S21-S9. (Review).
57. Cummings, J.L. (1993). Fronto-subcortical circuits and human behavior. *Archives of Neurology*, **50**, 873–80.
58. Hachinski, V.C., Iliff, L.D., Zilhka, E., *et al.* (1975). Cerebral blood flow in dementia. *Archives of Neurology*, **32**, 632–7.
59. Moroney, J.T., Bagiella, E., Desmond, D.W., *et al.* (1997). Meta-analysis of the Hachinski ischemic score in pathologically verified dementias. *Neurology*, **49**, 1096–105.
60. Erkinjuntti, T. (1987). Differential diagnosis between Alzheimer's disease and vascular dementia: evaluation of common clinical methods. *Acta Neurologica Scandinavica*, **76**, 433–42.
61. Mölsä, P.K., Marttila, R.J., and Rinne, U.K. (1995). Long-term survival and predictors of mortality in Alzheimer's disease and multi-infarct dementia. *Acta Neurologica Scandinavica*, **91**, 159–64.
62. Skoog, I., Nilsson, L., Palmertz, B., *et al.* (1993). A population-based study on dementia in 85-year-olds. *New England Journal of Medicine*, **328**, 153–8.
63. Chui, H.C., and Gonthier, R. (1999). Natural history of vascular dementia. *Alzheimer Dis Assoc Disord,* **13** (Suppl 23), S124–S30.
64. Folstein, M.F., Folstein, S.E., and McHugh, P.R. (1975). 'Mini-Mental State': a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **12**, 189–98.
65. Royall, D. and Roman, G. (1999). Executive control function: a rational basis for the diagnosis of vascular dementia. *Alzheimer Disease and Associated Disorders*, **13**(Suppl. 3), 69–80.
66. Pohjasvaara, T., Erkinjuntti, T., Ylikoski, R., *et al.* (1998). Clinical determinants of poststroke dementia. *Stroke*, **29**, 75–81.
67. Erkinjuntti, T., Ketonen, L., Sulkava, R., *et al.* (1987). Do white matter changes on MRI and CT differentiate vascular dementia from Alzheimer's disease? *Journal of Neurology, Neurosurgery, and Psychiatry*, **50**, 37–42.
68. Erkinjuntti, T., Ketonen, L., Sulkava, R., *et al.* (1987). CT in the differential diagnosis between Alzheimer's disease and vascular dementia. *Acta Neurologica Scandinavica*, **75**, 262–70.
69. Launes, J., Sulkava, R., Erkinjuntti, T., *et al.* (1991). 99mTC-HM-PAO SPECT in suspected dementia. *Nuclear Medicine Communications*, **12**, 757–65.
70. Orrell, R.W. and Wade, J.P.H. (1996). Clinical diagnosis: how good is it and how should it be done? In *Vascular dementia. Current concepts* (eds I. Prohovnik, J. Wade, S. Knezevic, T.K. Tatemichi, and T. Erkinjuntii), pp. 143–63. Wiley, Chichester.
71. Erkinjuntti, T. and Sulkava, R. (1991). Diagnosis of multi-infarct dementia. *Alzheimer Disease and Associated Disorders*, **5**, 112–21.
72. Slooter, A.J., Tang, M.X., van Duijn, C.M., *et al.* (1997). Apolipoprotein E epsilon4 and the risk of dementia with stroke. A population-based investigation. *Journal of the American Medical Association*, **277**, 818–21.
73. McKhann, G., Drachman, D., Folstein, M., *et al.* (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology*, **34**, 939–44.
74. Petersen, R.C. (1995). Normal aging, mild cognitive impairment, and early Alzheimer's disease. *Neurologist*, **1**, 326–44.
75. Braak, H. and Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*, **82**, 239–59.
76. Roman, G.C. (1991). White matter lesions and normal-pressure hydrocephalus: Binswanger disease or Hakim syndrome? *American Journal of Neuroradiology*, **12**, 40–1.
77. McKeith, I.G., Galasko, D., Kosaka, K., *et al.* (1996). Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop. *Neurology*, **47**, 1113–24.
78. Anonymous (1994). Clinical and neuropathological criteria for frontotemporal dementia. The Lund and Manchester groups. *Journal of Neurology, Neurosurgery, and Psychiatry*, **57**, 416–18.
79. Marder, K., Tang, M.X., Cote, L., *et al.* (1995). The frequency and associated risk factors for dementia in patients with Parkinson's disease. *Archives of Neurology*, **52**, 695–701.
80. Erkinjunti, T., Roman, G.C., Gauthier, S., *et al.* (2004). Emerging therapies for vascular dementia and vascular cognitive impairment. *Stroke,* **35**, 1010–17.
81. Wenning, G.K., Be-Sholomon, Y., Magalhaes, M., *et al.* (1995). Clinicopathological study of 35 cases of multiple system atrophy. *Journal of Neurology, Neurosurgery, and Psychiatry*, **58**, 160–6.
82. Jorm, A.F., Korten, A.E., and Henderson, A.S. (1987). The prevalence of dementia: a quantitative integration of the literature. *Acta Psychiatrica Scandinavica*, **76**, 465–79.
83. Rocca, W.A., Hofman, A., Brayne, C., *et al.* (1991). The prevalence of vascular dementia in Europe: facts and fragments from 1980–1990 studies. *Annals of Neurology*, **30**, 817–24.
84. Tatemichi, T.K., Desmond, D.W., Mayeux, R., *et al.* (1992). Dementia after stroke: baseline frequency, risks, and clinical features in a hospitalized cohort. *Neurology*, **42**, 1185–93.
85. Hachinski, V. (1992). Preventable senility: a call for action against the vascular dementias. *Journal of the American Geriatrics Society*, **340**, 645–8.
86. Knezevic, S., Labs, K.H., Kittner, B., *et al.* (1996). The treatment of vascular dementia: problems and prospects. In *Vascular dementia. Current concepts* (eds I. Prohovnik, J. Wade, S. Knezevic, T.K. Tatemichi, and T. Erkinjuntii), pp. 301–12. Wiley, Chichester.
87. Pantoni, L., Carosi, M., Amigoni, S., *et al.* (1996). A preliminary open trial with nimodipine in patients with cognitive impairment and leukoaraiosis. *Clinical Neuropharmacology*, **19**, 497–506.
88. Görtelmeyer, R. and Erbler, H. (1992). Memantine in treatment of mild to moderate dementia syndrome. *Drug Research*, **42**, 904–12.
89. Rother, M., Erkinjuntti, T., Roessner, M., *et al.* (1998). Propentofylline in the treatment of Alzheimer's disease and vascular dementia. *Dementia*, **9**(Suppl. 1), 36–43.
90. Erkinjuntti, T. (1999). Cerebrovascular dementia: a guide to diagnosis and treatment. *CNS Drugs*, **12**, 35–48.
91. Erkinjunti, T., *et al.* (2004). Emerging therapies for vascular dementia and vascular cognitive impairment. *Stroke*, **35**, 1010–17.
92. Erkinjuntti, T., *et al.* (2002). Efficacy of galantamine in probable vascular dementia and Alzheimer's disease combined with cerebrovascular disease: a randomised trial. *Lancet*, **359**, 1283–90.
93. Mortel, K.F. and Meyer, J.S. (1995). Lack of postmenopausal estrogen replacement therapy and the risk of dementia. *Journal of Neuropsychiatry and Clinical Neurosciences*, **7**, 334–7.
94. Skoog, I. (1997). The relationship between blood pressure and dementia: a review. *Biomedicine and Pharmacotherapy*, **51**, 367–75.
95. Forette, F., Seux, M.L., Staessen, J.A., *et al.* (1998). Prevention of dementia in randomised double-blind placebo-controlled Systolic Hypertension in Europe (Syst-Eur) trial. *Journal of the American Geriatrics Society*, **352**, 1347–51.
96. Tzourio, C., *et al.* (2003). Effects of blood pressure lowering with perindopril and indapamide therapy on dementia and cognitive decline in patients with cerebrovascular disease. *Archives of Internal Medicine*, **163**, 1069–75.

97. Meyer, J.S., Judd, B.W., Tawaklna, T., *et al.* (1986). Improved cognition after control of risk factors for multi-infarct dementia. *Journal of the American Medical Association*, **256**, 2203–9.
98. Meyer, J.S., Rogers, R.L., McClintic, K., *et al.* (1989). Randomized clinical trial of daily aspirin therapy in multi-infarct dementia. A pilot study. *Journal of the American Geriatrics Society*, **37**, 549–55.
99. Lechner, H. (1998). Status of treatment of vascular dementia. *Neuroepidemiology*, **17**, 10–13.

4.1.9 Dementia due to HIV disease

Mario Maj

Introduction

The first description of a syndrome consisting of cognitive, motor, and behavioural disturbances in patients with AIDS was published in 1986.[1] The syndrome was named 'AIDS dementia complex'. In 1990, the World Health Organization (WHO) introduced the term 'HIV-associated dementia',[2] pointing out that subclinical or mild cognitive and/or motor dysfunctions without impairment of performance in daily living activities cannot be subsumed under the term 'dementia'. The expression 'mild cognitive/motor disorder' was proposed for those conditions. The same distinction was made in 1991 by the American Academy of Neurology,[3] which identified an 'HIV-associated dementia complex' and an 'HIV-associated minor cognitive/motor disorder'. The present chapter focuses on the dementia syndrome associated with HIV infection.

Clinical features

The onset of HIV-associated dementia is usually insidious. Early cognitive symptoms include forgetfulness, loss of concentration, mental slowing, and reduced performance on sequential mental activities of some complexity (the subject misses appointments, or needs lists to recall ordinary duties; loses track of conversations or his or her own train of thought; needs additional time and effort to organize thoughts and to complete daily tasks). Early behavioural symptoms include apathy, reduced spontaneity and emotional responsivity, and social withdrawal (the subject becomes indifferent to his or her personal and professional responsibilities; his or her work production decreases, as well as the frequency of social interactions; the subject complains of early fatiguability, malaise, and loss of sexual drive). Depression, irritability or emotional lability, agitation, and psychotic symptoms may also occur. Early motor symptoms include loss of balance and coordination, clumsiness, and leg weakness; the subject is less precise in normal hand activities, such as writing and eating, drops things more often than usual, trips and falls more frequently, and perceives the need to exercise more care in walking.[1,4]

Routine mental status tests, in this early stage, may be normal or show only slowing in verbal or motor responses and/or difficulty in recalling a series of objects after 5 min or more. Neurological examination may show tremor (best seen when the patient sustains a posture, such as holding the arms and fingers outstretched), hyperreflexia (particularly of the lower extremities), ataxia (usually seen only on rapid turns or tandem gait), slowing of rapid alternating movements (of the fingers, wrists, or feet), frontal release signs (snout reflex, palmar grasp), dysarthria. Tests of ocular motility may show interruption of smooth pursuits, and slowing or inaccuracy of saccades.

In the late stages of the disease, there is usually a global deterioration of cognitive functions and a severe psychomotor retardation. Speech is slow and monotonous, with word-finding difficulties and possible progression to mutism. Patients become unable to walk, due to paraparesis, and usually lie in bed indifferent to their illness and their surroundings. Bladder and bowel incontinence are common. Myoclonus and seizures may occur. Pedal paraesthesias and hypersensitivity may appear, due to concurrent sensory neuropathy. The level of consciousness is usually preserved, except for occasional hypersomnolence.

Classification

The WHO criteria for HIV-associated dementia[2] are as follows:

1 The research criteria for dementia of the ICD-10 are met, with some modifications:
   (a) decline in memory may not be severe enough to impair activities of daily living;
   (b) decline in motor function may be present, and is verified by clinical examination and, when possible, formal neuropsychological testing;
   (c) the minimum requested duration of symptoms is 1 month;
   (d) aphasia, agnosia, and apraxia are unusual.
2 Laboratory evidence for systemic HIV infection is present.
3 No evidence of another aetiology from history, physical examination, or laboratory tests should be present (specifically, cerebrospinal fluid analysis and either computed tomography (CT) or magnetic resonance imaging (MRI) should be done to exclude active central nervous system opportunistic processes).

The American Academy of Neurology criteria[3] require the following:

1 Laboratory evidence for systemic HIV infection.
2 Acquired abnormality in at least two of the following cognitive abilities (present for at least 1 month): attention/concentration, speed of processing of information, abstraction/reasoning, visuospatial skills, memory/learning, and speech/language.
3 At least one of the following:
   (a) acquired abnormality in motor function or performance;
   (b) decline in motivation or emotional control or change in social behaviour.
4 Absence of clouding of consciousness during a period long enough to establish the presence of 2.
5 Absence of evidence of another aetiology.

Both the WHO and the American Academy of Neurology criteria distinguish three levels of severity of the dementia syndrome (mild, moderate, and severe), on the basis of the degree of the impairment in activities of daily living.

Diagnosis and differential diagnosis

Neuropsychological tests

Neuropsychological examination supports the clinical diagnosis of HIV-associated dementia, by providing evidence of cognitive and

motor dysfunction. Moreover, it may be useful in the differential diagnosis with a depressive syndrome.

The most prominent impairment is observed on tests of fine motor control (finger tapping, grooved pegboard), rapid sequential problem solving (trail-making A and B, digit symbol), visuospatial problem solving (block design), spontaneity (verbal fluency), and visual memory (visual reproduction). In contrast, naming and vocabulary skills are largely preserved even in the most advanced cases.

The signs that should alert to the possible presence of a depressive 'pseudodementia' are as follows:[5]

1 the intratest variability of performance (i.e. missing easy items and then correctly answering more difficult questions);
2 mood-congruent complaints, which are at odds with objective performance (i.e. the subject complains of having difficulties with a test, whereas his or her performance is near perfect);
3 responses of 'I don't know' or giving up, which are followed by the correct answer, when the subject is further urged to respond.

It should be considered, however, that dementia and depression may coexist in HIV-seropositive subjects.

Brain imaging

Brain imaging provides additional support to the diagnosis of HIV-associated dementia, especially by excluding central nervous system opportunistic processes, in particular cerebral toxoplasmosis and primary central nervous system lymphoma.

The predominant finding in HIV-associated dementia is cerebral atrophy: both CT and MRI demonstrate widened cortical sulci and, less commonly, enlarged ventricles. Furthermore, MRI frequently shows high-intensity signal abnormalities on the T_2-weighted image (diffuse widespread involvement, patchy localized involvement, focal distinct areas of involvement, or punctuate white-matter hyperdensities). These lesions are without mass effect and are most commonly located in the periventricular white matter and the centrum semiovale (less frequently, in the basal ganglia or in the thalamus).

As to differential diagnosis, both CT and MRI are able to demonstrate the multiple bilateral ring-enhancing lesions that are characteristic of cerebral toxoplasmosis, and the contrast-enhancing mass lesions of primary central nervous system lymphoma.

Cerebrospinal fluid analysis

Cerebrospinal fluid analysis can support the clinical diagnosis of HIV-associated dementia, especially by excluding several central nervous system opportunistic infections, in particular cryptococcal meningitis.

The most frequent cerebrospinal fluid findings in HIV-associated dementia are the increase of total proteins and of the IgG fraction and index. A mononuclear pleocytosis may occur. The presence of the HIV core antigen p24 can be detected, although this finding is possible also in neurologically normal subjects. HIV RNA can be demonstrated in the cerebrospinal fluid by using the polymerase chain reaction. Increased cerebrospinal fluid levels of neopterin, β_2-microglobulin, and quinolinic acid (non-specific markers of immune activation), soluble Fas and Fas ligand (associated with apoptosis), as well as several cytokines (interleukin 1β, interleukin 6, tumour necrosis factor-α), have been reported, but may be detected also during central nervous system opportunistic infections.

As to differential diagnosis, Indian ink staining, cryptococcal antigen titres, and fungal culture can be decisive for the identification of cryptococcal meningitis. Other central nervous system opportunistic infections that can be identified by cerebrospinal fluid analysis include central nervous system tuberculosis, cytomegalovirus encephalitis, and neurosyphilis.

Epidemiology

There has been a decrease in the incidence of HIV-associated dementia after the introduction of highly active antiretroviral therapy (HAART): while between 1990 and 1992 the mean incidence was 21.1 cases per 1000 person-years, between 1996 and 1998 it decreased to 10.5 cases per 1000 person-years.[6] However, the incidence seems to have increased again in 2003.[7] A post-mortem neuropathologic study reported that, while severe HIV encephalopathy was not detected anymore in the HAART era, the prevalence of mild and moderate encephalopathy increased, probably reflecting the longer survival time after initial HIV infection.[8]

Pathogenesis

HIV crosses the blood-brain barrier by a Trojan-horse-type mechanism, using the macrophages it infects.[9] Once in the brain, it infects glial cells. Infected and activated macrophages and microglia release neurotoxins which lead to neuronal damage and apoptosis.[10] It is possible that direct effects of viral proteins on neurones also contribute to neurodegeneration. Post-mortem studies have revealed the presence of HIV in frontal lobes, subcortical white matter and the basal ganglia.[11]

Course and prognosis

In the pre-HAART era, HIV-associated dementia often progressed rapidly to severe deterioration and death, especially in patients with advanced systemic disease. Today, many patients present an attenuated form which is slowly progressive or static. The mean survival, which was 5 months in 1993–1995, increased to 38.5 months in 1996–2000.[12] Prominent psychomotor slowing, a history of intravenous drug use and low CD4 T-lymphocyte count seem to predict a more rapid progression.[13]

Available treatments

Antiretrovirals are not always successful in crossing the blood-brain barrier, but, as mentioned above, have been able to reduce the incidence and modify the course of HIV-associated dementia. There is evidence that they can improve specific aspects of cognitive functioning, such as psychomotor speed performance, in people with HIV-associated dementia.[14] The optimal HAART regimen for the treatment of HIV-associated dementia has not been established.

Neuroprotective drugs whose beneficial effect on cognitive performance in patients with HIV infection has been preliminarily documented include the monoamine oxidase inhibitor deprenyl (a putative antioxidant and antiapoptotic agent) and peptide T (which blocks the HIV gp120 envelope protein). Other investigational drugs include memantine and nitroglycerin (which are *N*-methyl-D-aspartate receptor antagonists), nimodipine (a calcium-channel blocker), pentoxifylline (an inhibitor of the production

and activity of tumour necrosis factor-α), and lexipafant (an antagonist of platelet-activating factor).

The psychostimulant methylphenidate has been found to be useful in treating apathy and cognitive slowing in patients with HIV-associated dementia, with relatively mild side effects. Only anecdotal evidence is available concerning the usefulness of cholinesterase inhibitors such as donepezil.

Patients with AIDS, when treated with typical antipsychotic drugs for the presence of psychotic symptoms or behavioural dyscontrol, are particularly prone to develop extrapyramidal side effects and neuroleptic malignant syndrome. According to preliminary research evidence, some atypical antipsychotics are well tolerated even by patients who had to stop standard neuroleptics due to extrapyramidal side effects.

AIDS patients with depressed mood have been found to respond to tricyclic antidepressants and selective serotonin reuptake inhibitors (SSRIs) as well as HIV-seronegative subjects. There is a preliminary evidence that SSRIs (or at least some of them) are better tolerated than tricyclic antidepressants, except in patients with diarrhoea.

Management

Patients with HIV-associated dementia often have additional disease processes which may aggravate the cognitive impairment, including secondary infections and metabolic disturbances. These conditions should be adequately diagnosed and managed.

An appropriate HAART regimen should be implemented and constantly monitored (taking into account that cognitive dysfunction may have a negative impact on adherence to treatment). If psychotic symptoms, behavioural dyscontrol, or mood disturbances are present, the same strategies which are used for other people with these problems should be implemented, taking into account that HIV-infected patients have an increased sensitivity to the side effects of antipsychotics and antidepressants, and that adverse interactions may occur between psychotropic drugs and antiretrovirals (for instance, the administration of St. John's Wort induces the metabolism of the protease inhibitor indinavir, thus decreasing its serum concentration to levels which may cause treatment failure).[15] Methylphenidate (5–20 mg/day) may be used to reduce apathy and psychomotor slowing.

Psychosocial interventions in HIV-associated dementia should include maintenance of a structured daily schedule, titration of external stimuli, restriction to familiar environments, frequent orienting interactions with significant others, and monitoring of personal and financial affairs. Psychoeducational intervention with families and significant others is also essential.

The care of patients with HIV-associated dementia will make increasing demands on health services, as well as on volunteer and community support systems. It is uncertain, at present, whether such care is best provided in specialized units (e.g. inpatient AIDS units), or within general psychiatric or medical services. Special management problems may arise when the behavioural disturbance (e.g. poor impulse control, sexual acting-out behaviour) is such as to constitute a risk for other patients or staff members. Placement of patients in the terminal stage of the disease may also represent a problem: the lack of appropriate options in the community may obstruct their timely and humane discharge from the hospital.

Further information

Dougherty, R.H., Skolasky, R.L., and McArthur, J.C. (2002). Progression of HIV-associated dementia treated with HAART. *The AIDS Reader*, **12**, 69–74.

Kaul, M. and Lipton, S.A. (2006). Mechanisms of neuronal injury and death in HIV-1 associated dementia. *Current HIV Research*, **4**, 307–18.

Maj, M., Starace, F., and Sartorius, N. (1993). *Mental disorders in HIV-1 infection and AIDS*. Hogrefe and Huber, Bern.

References

1. Navia, B.A., Jordan, B.D., and Price, R.W. (1986). The AIDS dementia complex: I. clinical picture. *Annals of Neurology*, **19**, 517–24.
2. World Health Organization. (1990). Report of the Second Consultation on the Neuropsychiatric Aspects of HIV-1 Infection, Geneva, 11–13 January 1990. World Health Organization, Geneva.
3. American Academy of Neurology AIDS Task Force. (1991). Nomenclature and research case definitions for the neurological manifestations of human immunodeficiency virus type-1 infection. *Neurology*, **41**, 778–85.
4. Maj, M. (1990). Organic mental disorders in HIV-1 infection. *AIDS*, **4**, 831–40.
5. Van Gorp, W.G., Satz, P., Hinkin, C., *et al.* (1989). The neuropsychological aspects of HIV-1 spectrum disease. *Psychiatric Medicine*, **7**, 59–78.
6. Sacktor, N., Lyles, R.H., Skolasky, R., *et al.* (2001). HIV-associated neurologic disease incidence changes: multicenter AIDS cohort study, 1990–1998. *Neurology*, **56**, 257–60.
7. McArthur, J.C. (2004). HIV dementia: an evolving disease. *Journal of Neuroimmunology*, **157**, 3–10.
8. Neuenburg, J.K., Brodt, H.R., Herndier, B.G., *et al.* (2002). HIV-related neuropathology, 1985 to 1999: rising prevalence of HIV encephalopathy in the era of highly active antiretroviral therapy. *Journal of Acquired Immune Deficiency Syndromes*, **31**, 171–7.
9. Lawrence, D.M. and Major, E.O. (2002). HIV-1 and the brain: connections between HIV-1-associated dementia, neuropathology and neuroimmunology. *Microbes and Infection*, **4**, 301–8.
10. Nath, A. (2002). Human immunodeficiency virus (HIV) proteins in neuropathogenesis of HIV dementia. *Journal of Infectious Diseases*, **186**(Suppl. 2), S193–8.
11. Kaul, M., Garden, G.A., and Lipton, S.A. (2001). Pathways to neuronal injury and apoptosis in HIV-associated dementia. *Nature*, **410**, 988–94.
12. Dore, G.J., McDonald, A., Li, Y., *et al.* (2003). Marked improvement in survival following AIDS dementia complex in the era of highly active antiretroviral therapy. *AIDS*, **17**, 1539–45.
13. Bouwman, F.H., Skolasky, R., Hes, D., *et al.* (1998). Variable progression of HIV-associated dementia. *Neurology*, **50**, 1814–20.
14. Sacktor, N.C., Lyles, R.H., Skolasky, R., *et al.* (1999). Combination antiretroviral therapy improves psychomotor speed performance in HIV-seropositive homosexual men. *Neurology*, **52**, 1640–6.
15. Piscitelli, S.C., Burstein, A.H., Chaitt, D., *et al.* (2000). Indinavir concentrations and St. John's Wort. *Lancet*, **355**, 547–8.

4.1.10 The neuropsychiatry of head injury

Simon Fleminger

Head injury 'imparts at a blow both physical and psychological trauma',[1] and the consequences are often devastating and enduring.[2] Not infrequently head injury leads to a psychiatric consultation, which will need to take into account the interplay between the brain and its injuries as well as the psychodynamic processes that follow from the injury.

In the immediate aftermath of the head injury, the management rests with the acute surgical and medical team.[3] The psychiatrist is usually not involved at this stage. Nevertheless, to understand the later neuropsychiatric effects of head injury it is first necessary to know what happens to the brain when it is injured.

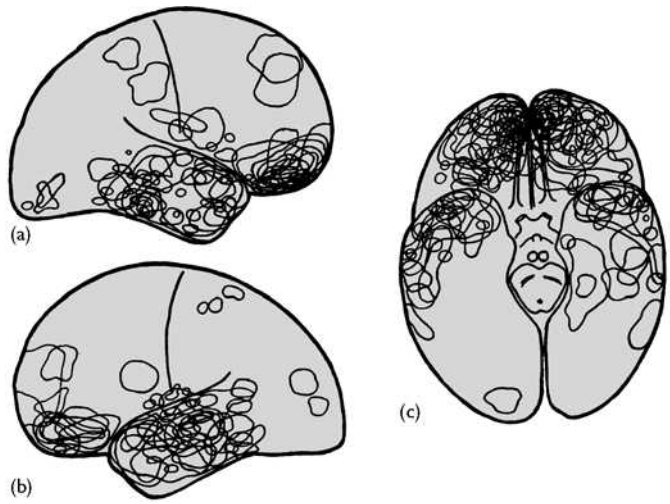


Fig. 4.1.10.1 A composite of the contusions found in 50 cases of people dying from head injury. (Reproduced with permission from Courville, C.B. (1937). *Pathology of the central nervous system*, Part IV. Pacific Press, Mountain View, CA)

Neuropathology

Open head injuries

In open head injuries there is penetration of the skull often with considerable destruction of brain tissue local to the trauma, but relatively less at a distance—particularly for lower velocity injuries such as stabbing. Open head injuries may therefore be associated with little, if any, loss of consciousness, which is generally a marker of diffuse brain injury.

Closed head injuries

(a) Contusions

In closed head injuries acceleration/deceleration forces and shearing forces damage the brain. The soft brain moves within its hard bony box and is damaged. Contusion of the brain occurs, ranging from slight localized small vessel bleeding into surrounding tissue to almost complete local destruction of the brain.

The medial orbital frontal cortex and the tips and undersurface of the temporal lobes are particularly vulnerable to contusions (Fig. 4.1.10.1). The brain becomes traumatized on adjacent bone of the floor of the skull. Contrecoup localization of contusions is sometimes evident.

(b) Intracerebral haemorrhage

Localized haemorrhage into the brain occurs at the site of a contusion. Scattered intracerebral haemorrhages found at the interface between grey and white matter are thought to be associated with diffuse axonal injury (see below). A large isolated haematoma suggests that a blood vessel has ruptured.

In very severe injury haemorrhages are also found round the aqueduct in the brainstem, perhaps caused by distortion of the brainstem as a result of cerebral herniation into the posterior fossa due to raised intracranial pressure. They are associated with prolonged coma or death.

(c) Extradural and subdural haemorrhage

Haemorrhage into the extradural or subdural space will act as a space-occupying lesion and contribute to raised intracranial pressure. The extradural haemorrhage, being under high pressure, can rapidly cause coma and death. The patient may 'talk and die', regaining consciousness after the head injury, only to lapse a few hours later into severe coma. Without acute neurosurgical intervention to drain the blood these patients will die.

Subdural haematomas tend to run a subacute course and as such are of more interest to the psychiatrist. They may present with a failure to improve, or fluctuating drowsiness, weeks or months after the head injury. They may regress spontaneously or may require surgical drainage, but they do have a propensity to recur.

(d) Diffuse axonal injury

Diffuse axonal injury occurs in the white matter tracts of the cerebral hemispheres, including the corpus callosum, and the brainstem, particularly the cerebellar peduncles. Axons break up over the course of the first 24 to 48 h following brain trauma with the formation of 'retraction balls'-globular structures at the end of transected axons.[4]

(e) Oedema and ischaemia

Oedema of damaged brain occurs over the first few hours following brain injury. The resulting raised intracranial pressure compromises the cerebral circulation and results in ischaemia, which may further contribute to brain injury. Cerebral oedema tends to resolve over the course of a few days or weeks.

(f) Neuronal death

Two fairly distinct processes result in neuronal death from traumatic brain injury.[4] Necrotic cell death occurs when there is massive cell disruption either from the direct effects of the trauma on the cell membrane or from anoxia. It is a relatively passive process though may involve toxic effects of high levels of intracellular calcium, excitatory transmitters, and free radicals. On the other hand apoptotic cell death, 'cell suicide', is a more active process triggered by various routes including ligand binding to cell death receptors. These activate, for example, endonucleases which attack cellular DNA. Markers of apoptotic cell death are elevated in the days and weeks after injury.

(g) Late effects

Ventriculomegaly may develop over the weeks and months following injury. Often it is the result of atrophy of the white matter of the cerebral hemispheres, usually attributed to diffuse axonal injury, and is associated with atrophy of the corpus callosum. More localized atrophy is observed when contusions resolve to leave a loss of brain tissue.

Of greater importance is hydrocephalus resulting from the residual effects of subarachnoid blood interfering with the normal cerebrospinal fluid flow and preventing it from escaping into the venous system. This may require insertion of a ventriculo-peritoneal shunt to prevent deterioration in cognitive function.

Fractures to the floor of the skull, particularly if they are associated with cerebrospinal fluid leaks, may allow infection into the subarachnoid space, causing meningitis sometimes years after injury. Cerebral abscesses may take months before they become clinically evident.

Loss of consciousness following head injury

The mechanism of loss of consciousness after mild blows to the head is poorly understood. Based on animal work some researchers suggest it is produced by activation of cholinergic nuclei in the pons.[5]

Loss of consciousness lasting for more than a few minutes is likely to damage either cortical areas necessary for consciousness or the subcortical arousal systems. Raised intracranial pressure, partly as a result of compromising cerebral circulation, causes coma. Large or multiple haematomas are likely to be associated with a period of coma, particularly if they are associated with cerebral oedema.

Some patients, however, show prolonged coma with little to be found on brain scan apart from some evidence of generalized cerebral oedema. In these patients, diffuse axonal injury may be the cause of their coma, possibly by damaging the white matter tracts that carry arousal signals from the brainstem to the cortex.

Remember that the head injury may have been caused by an accident triggered by a loss of consciousness, for example, due to hypoglycaemia, alcohol intoxication, or an epileptic fit. Systemic effects (e.g. hypoxaemia or fat emboli) may exacerbate unconsciousness due to head trauma, as may drug intoxication.

Head injury severity

It is surprisingly difficult to predict the degree of brain injury from the size of the blow to the head. Some patients after a severe blow to the head sustain little injury to the brain. Others will suffer severe brain injury associated with prolonged unconsciousness, merely as a result of hitting their head on the ground by falling over from the standing position. Perhaps in the very occasional case significant brain injury occurs when there is no, or only momentary, loss of consciousness (see post-concussion syndrome below). The presence of a skull fracture says little about the severity of the brain injury incurred.

There are several clinical indicators of head injury severity (Box 4.1.10.1). Of these the duration of retrograde amnesia is probably the least valuable: it correlates very poorly with head injury severity.

Box 4.1.10.1 Clinical indicators of head injury severity

- The duration of retrograde amnesia—the period leading up to the injury for which memories have been lost. Tends to shrink as the patient recovers.
- The depth of unconsciousness as assessed by the worst score on the Glasgow Coma Scale—a score of 3 indicates absent responses with severe coma, 15 is normal consciousness.
- The duration of coma—this may be difficult to ascertain because of routine sedation and ventilation following severe head injuries.
- Neurological evidence of cerebral injury—abnormality on neuroimaging or EEG.
- The duration of post-traumatic amnesia—interval between injury and the return of normal day-to-day memories.

The duration of post-traumatic amnesia is probably the best marker of outcome,[6] and is particularly useful because it can be assessed retrospectively. Most patients with a post-traumatic amnesia of less than 1 week will be left with little if any disability, while a duration of more than 1 month indicates that there is likely to be enduring and significant disability.

Predictors of a worse outcome after head injury are a previous head injury, older age, *APOE e4* positive status, and alcohol dependence.

There is no universally accepted classification of head injury severity. However, the most widely used grading system is based on the lowest rating of the Glasgow Coma Scale (**GCS**)[7] following injury.

- Mild: GCS score 13 to 15. Likely to be associated with a loss of consciousness of less than 30 min and a post-traumatic amnesia of less than 24 h. There must be clinical evidence of concussion.
- Moderate: GCS score 9 to 12. Likely to be associated with a loss of consciousness of more than a few minutes but less than 6 h and a post-traumatic amnesia of more than 1 day but less than 2 weeks.
- Severe: GCS score 3 to 8. Likely to be associated with a loss of consciousness of more than 6 h or a post-traumatic amnesia of more than 2 weeks.

Epidemiology

On average 200–300/100 000 population attend hospital with a head injury every year.[8] About one-sixth of those attending hospital will be admitted. This reflects the fact that about 80 per cent of head injuries are mild, 10 per cent moderate, and 10 per cent severe.

At greatest risk are 15- to 25-year-olds. The sex ratio is about two to three males to one female. Risk factors include alcohol misuse as well as lower socio-economic class. Road traffic accidents are the largest single cause of head injury in most civilian cohorts, followed by assaults and falls. A significant proportion will sustain their head injury as a result of deliberate self-harm.

The prevalence rate for those experiencing considerable disability as a result of head injury is in the order of 100 per 100 000.

Investigations

Neuroimaging

In the emergency room or on the trauma unit CT brain scanning is generally the preferred investigation, with its faster acquisition time and good visualization of subdural and extradural haematomas.

For later neuropsychiatric assessment magnetic resonance imaging (MRI) is preferred.[9] Cerebral contusions are often found near the bone–brain interface (see above) where the image quality of CT is reduced because of imaging artefacts from the adjacent bone. MRI has no such limitation and generally has better sensitivity and anatomical definition. MRI is able to detect, on T_2-weighted images, changes in signal associated with a diffuse axonal injury when the white matter would have appeared normal on CT brain imaging. Gradient echo MRI sequences should be performed to detect haemosiderin deposits from old small traumatic haemorrhages.

Despite its greater sensitivity a normal MRI does not rule out significant brain injury. On the other hand, particularly in the elderly, MRI may detect abnormalities unrelated to the head injury. It may not be possible to perform an MRI scan if there is magnetic material present in the body (e.g. a pacemaker).

The MRI scan can be normal and yet functional imaging of cerebral metabolism using single-photon emission computed tomography or positron-emission tomography will detect abnormalities. In general, changes on functional imaging correlate better with neuropsychological test performance than do lesions found on structural imaging.[10] However, abnormalities on functional imaging are not necessarily due to brain injury. Hypometabolism may be seen in mental illness without brain injury, for example in depression. Marked hypometabolism on positron-emission tomography imaging has been observed in a man with cognitive impairment occurring immediately after a psychological trauma.[11] He had sustained no head injury.

Electroencephalography

Electroencephalography (EEG) may be useful in the investigation of a deteriorating conscious level or unexpectedly prolonged unconsciousness, and in the investigation of unusual behavioural disturbances that may be attributable to epilepsy. However, EEG is not a good predictor of post-traumatic epilepsy and is generally not useful as a guide to prognosis.

Neuropsychological assessment

A neuropsychological assessment is an invaluable accompaniment to the psychiatric history and examination, and good liaison with the neuropsychologist is essential. Areas of impaired performance can be documented and quantified. This is often useful as a baseline for future assessments and to guide rehabilitation.

The National Adult Reading Test, for people whose first language is English, gives a good estimate of pre-injury IQ.[12] This can then be compared with the present performance on cognitive testing, to estimate the impairment produced by the head injury.

Subtle neuropsychological impairments, which are often not obvious clinically, suggest that the patient may have more problems when they return to work than would otherwise have been expected. On the other hand, if there is clinical evidence of underperformance, and yet standard neuropsychological test results are normal, then it is particularly important that executive function is tested.[13]

Function and health

Psychological symptoms far outstrip neurophysical symptoms (e.g. hemiparesis or dysarthria) as determinants of chronic disability and suffering, both of the patient and their carer, following brain injury.

The ideas encapsulated in the International Classification of Functioning, Disability and Health (ICF) (http://www3.who.int/icf/icftemplate.cfm) are important for understanding recovery from brain injury. ICF is so named because it wishes to emphasize health and functioning by moving away from a dichotomous distinction between those who are healthy and those who are disabled. ICF is a development of the earlier classification ICIDH based on:

- Impairments—abnormalities of structure, or physiological or psychological function.
- Disability—concerned with performance of activities.
- Handicap—reflects limitations fulfilling the person's normal social role and participation in society. It is very responsive to external, e.g. environmental and societal, factors.

In ICF an individual's position on the spectrum between health and disability is considered according to (i) functioning and disability and (ii) contextual factors. ICF details the environmental impacts on functioning, e.g. the consequences of living in an area prone to flooding in somebody who is wheelchair dependent. What a person can do in a standard environment (their capacity) is distinguished from what they actually do in their usual environment (their performance).

Recovery and long-term outcome from head injury

Most recovery takes place in the first year. As a general rule, the milder the head injury the sooner the patient achieves the asymptote of their recovery curve. After a mild head injury most patients will have fully recovered within 6 months. After very severe injury significant further improvements in impairment may be seen after the first year post-injury. Neuropsychological impairments tend to continue improving after neurophysical impairments are static. Nevertheless most of the recovery of cognitive function occurs within the first year.[14] Psychiatric symptoms, with their multifactorial aetiology, generally show no simple pattern of recovery.

Improvement in functioning and participation may continue long after the recovery of the underlying impairment has stopped. These further improvements often reflect improved coping strategies and environmental measures to facilitate independence. This will be the focus of the community rehabilitation team as they attempt to help minimize handicap, for example by improving access to local shops. Memory aids may enable the person to return to work. Continuing improvements in participation in social life and work can take place 5 to 10 years after head injury.[15]

But sometimes early gains are made, for example as a result of being in a return-to-work rehabilitation programme, which are subsequently lost over the longer term. In one study,[16] 25 per cent

had deteriorated at 5 years follow-up, with a similar proportion improving compared with how they were at 6 months. Those who deteriorated were more depressed and anxious, had lower self-esteem and had more problems with alcohol than those who improved.

In the longer term, decades after injury, it has been suggested that the reduced reserve of the injured brain makes it particularly vulnerable to the effects of ageing. Some studies have found an accelerated cognitive decline compared with age matched controls, for example in head-injured soldiers 25 years later.[17] Head injury may be a risk factor for the development of Alzheimer's disease, particularly in men.[18] However the evidence, both for an accelerated cognitive decline and for an increased risk of Alzheimer's disease, is inconsistent.

Aetiology of psychological sequelae

To understand the mental symptoms that follow head injury it is necessary to know about the person who has been injured, what brain injuries they sustained, and the consequences. However, the interaction between these is complex and poorly understood.

Pre-traumatic factors

People who take risks or get into fights are more likely to sustain a head injury; therefore these personality traits, present before injury, are over-represented in head-injury survivors. Young men are at high risk, as are those who have already had a head injury or have cognitive dysfunction.[19]

The poor social adjustment of many patients before the head injury partly explains why so many run into behavioural problems afterwards. But premorbid characteristics do not strongly predict who will develop emotional and behavioural problems. Nevertheless, traumatic brain injury probably has the potential to turn pre-injury personality traits into post-injury personality disorders.

The trauma

The extent of brain injury probably explains less than 10 per cent of the variance in the amount of psychiatric morbidity that follows brain injury.[20] In general, early psychiatric symptoms, within weeks and months of the injury, correlate better with the extent and location of brain injury than do late psychiatric symptoms. Left hemisphere damage seems to be associated with greater psychiatric morbidity. Specific relationships between the location of brain injury and the psychiatric symptoms are discussed below.

But the head injury is also a psychological trauma. Amnesia for the event, as a result of the head injury, protects against post-traumatic stress disorder. However, it is a mistake to believe that amnesia for the event prevents a psychological stress reaction to the event itself.

- The meaning of the event may be distressing to the patient.[21] In the case of assaults, the head injury may signal the potential for further assaults. An accident may have been life-threatening and a shocking reminder to the patient that they are mortal. They may feel aggrieved by an employer's negligent action that caused the accident.
- The patient may be amnesic for the event, lacking explicit memories of what happened, but retain implicit memory of what happened. The consequences of these implicit memories may be akin to that observed in one of Claperède's amnesic patients.[22] The doctor shook the patient's hand, pricking it while doing so with a concealed drawing pin. The next day the patient could not remember having met the doctor, but flinched from shaking his hand when it was offered.
- They may have islets of intact memories that may be extremely frightening.[23]

Post-traumatic factors

Post-traumatic factors deserve special attention because they are most likely to be amenable to intervention. The psychiatrist needs to consider the patient's reaction to any disability, as well as the consequences of the disability on the role of the patient in the family and society. There may be reinforcing cycles of maladaptive behaviour, and compensation claims may complicate the picture.

Cognitive impairment

Cognitive impairment correlates with measures of head injury severity better than any of the other mental sequelae. For example, there is a strong correlation between the duration of post-traumatic amnesia and the severity of cognitive impairment.

Attention and concentration

Non-specific cognitive impairments include slowness and reduced concentration. The severely injured patient is likely to be stimulus bound, i.e. responding to each and every stimulus they are exposed to in a rather concrete way. At the same time they may show perseveration, with previous responses inappropriately interfering with the answers to subsequent questions, or when the topic of a conversation has been changed.

Dysexecutive syndrome

More specific impairments, generally referred to as the dysexecutive syndrome, result from a disturbance of the executive system responsible for organizing, planning, scheduling, prioritizing, and monitoring cognitive activities.[24] In some patients with isolated medial orbito-frontal lesions or dorso-lateral prefrontal lesions, the dysexecutive syndrome may stand alone. Disturbance of the executive system also results in difficulties in attending to two things at once, and distractibility.

Patients with the dysexecutive syndrome may be much more impaired in everyday life than is predicted by their performance on standard neuropsychological tests. They can manage with the clear instructions of the well-structured and constrained test situation. But in the real world these are absent; priorities have to be set, a strategy planned, decisions taken, and the unexpected dealt with, all without guidance. In the real world, impairment of the executive system may be catastrophic. Tests of the dysexecutive syndrome have been developed in order to be better predictors of these real-life problems. [13]

Memory impairment

Memory impairment is perhaps the most common cognitive impairment that follows head injury, and can be very disabling.[25] People will have problems remembering where they put things, what to do next, how to get home from the shops, or what they did

yesterday. Anterograde amnesia refers to these enduring problems laying down new memories, and must be distinguished from retrograde and post-traumatic amnesia (see Box 4.1.10.1).

No consistent pattern of brain injury is associated with anterograde amnesia and it seems likely that it is the combined damage to several areas which causes the amnesia. Frontal injury may be particularly implicated perhaps by interfering with the executive processes required for normal memory, for example in memory retrieval. As with most amnesic states the amnesia following brain injury is for explicit memories, namely those which are consciously remembered. Implicit memory, for example remembering and learning a motor skill, is relatively well preserved.

Anterograde amnesia is often characterized by distortions and inaccurate recall with poor monitoring and insight. Confabulations are often seen.

Communication

Dysphasia is quite common after head injury, and may be rather different from that seen after stroke. The more diffuse and widespread injury of traumatic brain injury results in additional cognitive impairments which colour the picture. Monitoring of language errors is often particularly poor and the patient may demonstrate a jargon aphasia such that they are apparently unaware that their speech is completely incomprehensible. Dysphasia often continues to improve even many years post-injury.[26]

Dysprosody, in which the normal rhythms and intonations of speech are lost, is also seen, more so after right hemisphere damage. This interferes with social communication because the voice sounds flat and fails to convey emotion. Social communication is disrupted for other reasons, for example the patient fails in the turn-taking necessary for normal conversation. Word-finding difficulties are common.

Visuospatial impairments

Visuospatial impairments may contribute to spatial disorientation. Visual agnosia is easy to miss in someone with quite widespread cognitive impairments. Hemi-neglect can be troublesome.

Personality change

Personality change after head injury results in more suffering than any other single sequel.[27,28] In general, personality change goes hand in hand with cognitive impairment. However, a severe personality change is occasionally found in somebody with almost no impairment of cognitive function. Normal test scores for memory and intellect do not rule out brain injury as a cause of personality change after head injury.

Aetiology

It is not easy to predict who will develop a change in personality after head injury. Sometimes a personality trait present before the injury becomes much more troublesome, but often there is no obvious predisposition. The site of the brain injury may play a role.[29] Lesions of the medial and lateral surfaces of the frontal lobe can produce impairments of drive. Whereas orbito-frontal lesions, on the undersurface of the frontal lobe, may cause a more troublesome personality change with impairments in social behaviour.

Post-traumatic factors also need to be considered. Some patients seem to learn maladaptive patterns of behaviour; for example the response of the carers may unwittingly reinforce unwanted behaviours. Chronic mental illness, aggravated by chronic psychosocial stressors, may be manifest as personality change. Dependence on drugs, particularly alcohol, frequently confounds the picture.

Characteristics of the personality change

Changes in personality[27] include apathy and impairment of motivation and ambition. Patients are often described as childish; this covers a range of traits including impulsivity, poor tolerance of frustration, being demanding and self-centred, and generally lacking the ability to take on the adult role in terms of independent decision-making. Patients may be fatuous and facetious. Antisocial behaviours (see below) and disinhibition are severe handicaps that make integration back into the community difficult. Sexual disinhibition of any type is particularly worrisome. A spectrum of severity is seen, ranging from being inappropriately flirtatious through to indiscriminate sexual assaults. Head injury is probably a risk factor for borderline personality disorder.[30]

In acquired antisocial personality disorder the person is often self-centred and relatively oblivious to the needs of others. They are likely to be tactless and, on occasion, offensively rude. Irritability and aggression and impulsive behaviour are seen. They may show a lack of remorse for violent behaviour. These personality traits often are accompanied by the dysexecutive syndrome. Thus not only does the person show disturbed social decision-making, resulting in antisocial behaviours, but also disruption of the planning and organizational skills needed for cognitive tasks. For example, helpful and supportive friends may be alienated in favour of disreputable acquaintances, at the same time as money is impulsively spent and lost, on risky projects without any attempt to weigh up the options.

Effects on family and carers

Families find personality change particularly difficult to cope with.[31] Children may be ignored and the partner's needs, particularly emotional needs, forgotten. The healthy balance of the relationship with the partner may be destroyed, with the head-injured person now unable to take an effective part in the household. The partner becomes a carer and the change in roles may have a serious impact on the sexual relationship. Divorce not infrequently follows. However, parents may find the childish personality of the brain-injured person easier to cope with; they revert to taking on the parental role.

Personality change may deteriorate. Supportive social networks are lost and social isolation and financial problems may contribute to depression or alcohol abuse, which then cause a deterioration in the behavioural problems associated with the personality change. Follow-up studies lend some support to this argument. Some behavioural problems are found to have deteriorated at 5 years after head injury,[32] and family burden increases over this period.

Early mental symptoms following brain injury

On recovery of consciousness many patients after a severe head injury go through a period of delirium with clouding of consciousness. The clouding of consciousness may resolve, leaving a confusional state in clear consciousness with disorientation and thought

disorder consisting of muddled thinking, rambling talk, and perseverations. This state is often dominated by misperceptions and misrecollections as the patient flits from one false observation to another.[33] Fear is common.

Distortions of memory

Confabulations, brief-lived false memories, emerge at about this time. Confabulations occur particularly in association with memory disturbance associated with frontal injury. The patient almost invariably shows poor insight into their memory problems, and is likely to be disorientated.

Occasionally after a severe head injury there are islets of memory in the dense amnesic period immediately around the time of the injury. These may be recollections of something that was consciously experienced at the time. On the other hand, the memories may have been fabricated from information subsequently given to the patient about what happened, or the memories may have no basis in reality and be properly described as a delusional memory.

Alterations of mood and perception

In the early recovery period oneroid states may be seen. The patient may be perplexed. He or she may feel that the trauma never occurred and that the whole event, including being in hospital, is a fabrication. Derealization/depersonalization may be associated with prominent anxiety, with the patient constantly asking for reassurance. Agitation occurs in about 10 per cent of patients with severe brain injury.[34]

Hallucinations, particularly visual, are occasionally observed, whereas illusions of familiarity are quite common after brain injury. The patient may have a sense of *déja vu*, or that he or she has met clinical staff or patients before. Distortions of the sense of familiarity seem to be implicated in many of the delusions observed early after brain injury.

Apathetic states

In many patients the recovery period lacks the positive features described above and is dominated by an apathetic withdrawn state.

Psychosis after brain injury

Early psychotic symptoms

The vast majority of the delusions and hallucinations occurring during the recovery period will themselves remit spontaneously and not relapse. However, it has been shown that, in some patients who have recovered from these early delusions, amylobarbitone can produce a return of symptoms.[35] This suggests that generalized disturbance of brain function plays an important part in the development of early delusions.

(a) Delusional misidentification

Delusional misidentifications of place, persons, objects, and events may be observed early in the course of recovery. Of these the one that is most pathognomonic of brain injury, and which is also associated with other causes of organic mental disorder, is reduplicative paramnesia. The term reduplicative paramnesia covers a range of phenomena which involve duplication of events or places. Pick,[36] who introduced the term, used it to describe a patient who believed she had visited a duplicate hospital.

Delusional disorientation for place may involve the belief that the current location is a duplicate of the true location or in some way displaced, for example that the hospital is in a different country. The patient may have two incompatible attitudes to orientation; this is sometimes referred to as a double orientation. For example, a patient who lives in Edinburgh acknowledges that he is in a hospital in London, but says that his home is just a few yards down the road. A common delusional disorientation is the patient's belief that they are still at work, despite the fact that they remain in hospital recovering from their injuries. Such patients lack insight into their injuries and report, for example, that they have been sent to complete some work assignment and that staff on the ward are colleagues from work.

Whereas isolated delusional misidentifications of place are rare in the absence of manifest organic brain disease, most cases of delusional misidentification of person (e.g. Capgras syndrome) are to be found in schizophrenia. Delusional misidentifications of person may also be observed following brain injury, often alongside a reduplicative paramnesia (see also Chapter 4.4).

Delusional misidentification syndromes can best be understood as the result of an interaction between organic brain disease and psychological disorder.[37] Lesions of the right hemisphere, often in combination with frontal injury or more diffuse evidence of brain disease, are particularly associated with delusional misidentification.

Late psychosis

(a) Schizophrenia-like psychosis

A psychotic illness may develop long after the acute confusional state has resolved. The patient may develop a typical schizophrenia indistinguishable from idiopathic schizophrenia. Would he or she have developed schizophrenia regardless of having had a head injury?

Davison and Bagley, almost 40 years ago,[38] estimated that patients after a head injury had a two- to three-fold increased risk of developing a schizophrenia-like psychosis compared with the general population. But there were large variations in the different studies they examined, and most were cohorts of war veterans who will have suffered open head injuries.

Any apparent association between head injury and schizophrenia might be explained by the fact that the period from late teens to early 20s is both the period of greatest risk of head injuries and the time when schizophrenia tends to start. In addition people at risk of schizophrenia may also be at increased risk of suffering a head injury ('reverse causality'[39]). Two large studies from Denmark[40] and Sweden,[41] based on linkage of nation wide hospital case registers, have shown that there appears to be no elevated risk of being admitted to hospital with a diagnosis of schizophrenia in those who have previously suffered a head injury. However, the second study did suggest that other non-affective psychoses, not diagnosed as schizophrenia, might be more common after a head injury. This fits with clinical experience; the patients whose psychosis seems most convincingly related to their head injury are those with more severe injuries. They would be diagnosed as suffering an organic psychosis, not schizophrenia.

(b) Paranoid psychosis

Paranoid psychoses may emerge after brain injury. Not infrequently this occurs relatively early and in a patient with severe cognitive impairment and personality change. Memory impairment will facilitate the development of persecutory ideas; for example, the

patient believes that belongings have been stolen. Persecutory ideas or delusions of reference are a fairly common cause of aggression and may be hidden by communication difficulties.

Mood disorders, including anxiety disorders

Depression

The study of depression after head injury raises two fundamental questions about the nosological status of depression.[42] First, with a severe disability should the belief that life is not worth living be regarded as a symptom of depression or a 'rational' reaction to an intolerable predicament? Second, what is one to make of symptoms of apathy or anhedonia when the brain pathways involved in generating spontaneous behaviour or the experience of pleasure have been damaged? Most of the biological symptoms of depression can be produced by brain injury.

The diagnosis of depression therefore relies heavily on identifying a depressive mood. Symptoms like self-deprecation or guilt are also particularly helpful in diagnosis. Estimates of the prevalence of depression after head injury vary, partly because of the lack of uniformity in defining depression. Perhaps 25 per cent of patients meet DSM-IIIR criteria for major depression 1 month after injury.[43] A similar rate of depression at 1 year is described in several studies, though perhaps the more conservative figure of 14 per cent[44] is more realistic. Over the first year many who are initially depressed recover, to be replaced by those previously not depressed who become depressed.

Aetiological factors include a personal history of depression, which is twice as common in those who become depressed, and lack of social support. Depression after head injury interferes with rehabilitation, and is associated with aggression. It may exacerbate cognitive impairment and in some cases produce a pseudodementia.

Emotional lability may occur, particularly after severe head injury, and is frequently associated with the presence of depression.

Mania

Manic illness after head injury is much less common than depression. It needs to be distinguished from the neurobehavioural symptoms of, for example, disinhibition and fatuous behaviour that may follow frontal injury. Mania is particularly associated with aggressive and assaultative behaviour following brain injury.

Anxiety disorder

Symptoms of anxiety are common after head injury,[45] particularly in those who have suffered mild injury. Generalized anxiety disorder occurs in perhaps 10–15 per cent of cases.[46]

Early symptoms may be observed in relation to derealization/depersonalization symptoms, or perplexity. Early on, the amnesic period surrounding the injury may cause great distress. In the catastrophic reaction, which is observed in patients with moderate to severe cognitive impairment, sudden distress occurs when they fail to perform a task, or because of their inability to communicate.

Anxiety symptoms, particularly in those with a mild head injury, may develop over the weeks and months following a head injury. It is then more likely to be associated with depression, post-concussion syndrome, and with post-traumatic stress disorder. Phobic avoidance is seen, for example when there is travel anxiety following a road traffic accident. Apprehension is a common complaint, perhaps reflecting problems caused by cognitive impairments, and the person may be indecisive. Therefore anxiety symptoms may emerge on return to work. Anxiety symptoms will be inflated in the presence of financial or family stress.

Obsessive–compulsive disorder is recognized sequelae of head injury. This may partly reflect the inflexibility and rigidity of the brain-injured person, or a response to doubt resulting from memory disorder.

Suicide

The risk of suicide is increased following head injury occurring in about 1 per cent of cases over the first 15 years or so after injury.[47] This represents about a three-fold increase in suicide rate compared with the age matched population rates. There is no evidence of a specific at risk period. At least some of the increased risk is probably because those at increased risk of head injury also have a greater risk of suicide. Rates of attempted suicide are increased after head injury.

Agitation and aggression

The psychiatrist is more likely to be asked to advise about the management of agitation and aggression following head injury than any other mental symptom.

Agitation in the early recovery period after severe brain injury will generally spontaneously improve over the course of days or weeks.[34] Early agitation may be followed by more intractable aggressive behaviour.[48] A major predictor of aggression is antisocial behaviour before the head injury.

If the aggressive behaviour emerges early, namely during the confusional state or shortly after it resolves, then this suggest that the aetiology is largely organic. A pattern of aggression that is highly stereotyped, or erupts over seconds with no or trivial triggers, or is bizarre, and is against a background of calm behaviour, suggests the possibility of epilepsy.

Other causes need to be considered. It is important to rule out any medical or surgical complications of the head injury, for example a subdural haematoma. Likewise pain and sources of infection, for example a UTI. The patient's worries and fears need to be explored, and phobic anxiety disorder considered. Drugs may make agitation worse and paradoxical effects of sedative medication occur if the medication increases confusion or disinhibition, or results in akathisia. The patient may be in a withdrawal state having stopped a drug they were taking regularly before the head injury. Drug and alcohol dependence may be especially problematic. Symptoms of mental illness may not be immediately obvious because of communication difficulties. It is therefore necessary to search for evidence of persecutory delusions, mania, depression, and anxiety. Any mental illness should be treated before considering medication specifically to treat agitation (see below).

Alcohol and head injury

Alcohol dependence complicates the management of the head-injured person several-fold. The person may have suffered several previous head injuries, as well as the effects of alcoholic brain damage before the head injury. A blow to the head may result in much greater brain injury for reasons that are poorly understood.[49] Poor physical health is likely to prejudice immediate management after the head injury. Subdural haematomas may be problematic.

Alcohol craving may interfere with medical care and rehabilitation.[50] Social networks are often poor, thus complicating discharge from hospital.

Very occasionally a head injury seems to cure the alcohol dependence. Unfortunately alcohol dependence often gets worse, perhaps because the head injury has weakened impulse control. Indeed some patients develop alcohol dependence when they find that alcohol relieves their anxiety symptoms.

Post-concussion syndrome

The post-concussion syndrome is poorly defined.[51] The term is perhaps most usefully reserved to describe a constellation of symptoms that may result in surprisingly severe disability after mild head injury. These symptoms may be observed after moderate and severe head injuries, in which case they are likely to be in the company of other symptoms more readily understood as resulting from brain injury. There is no consistent relationship between the prevalence of post-concussion symptoms and injury severity.

Phenomenology

Early symptoms tend to have a more neurological flavour and include headache, dizziness, and for example diplopia. Mild head injury fairly consistently results, in the immediate aftermath, in impairment of speed of information processing and concentration. Fatigue is also evident from early on, along with symptoms of noise sensitivity. Anxiety, depression, and irritability are common and may appear after a latent period. The symptoms of post-concussion syndrome overlap with those of post-traumatic stress disorder, and chronic fatigue. Other symptoms occasionally reported include tinnitus, unsteadiness, and muscle pain.

In general after a mild to moderate injury symptoms will have recovered by 2 to 6 months. But a few patients, sometimes after a latent period, develop persistent symptoms that last for years. Psychological factors are likely to be important in such patients, particularly if their injury is mild.

Aetiology

(a) Brain injury

Several observations support the contribution of brain injury, even in those with mild injury. Microscopic lesions in the brain have been described at post-mortem, following mild head injury.[52] Imaging, particularly functional imaging with single-photon emission CT or positron-emission tomography, may show abnormalities. Early after mild head injury there is evidence of cerebral dysfunction. One month after a mild head injury patients undertaking a working memory task showed more widespread activation of cerebral cortex compared to controls, even though their actual performance on the task was normal.[53]

(b) Psychological factors

Psychosocial factors have also been found to play a part in post-concussion syndrome, more so in those with symptoms lasting longer than 1 year. If the accident occurs at work, particularly if the person blames their employer, symptoms are more likely. A meta-analysis of the effects of compensation on symptoms, any symptom, after head injury concluded that, on average, being involved in compensation claims increases symptoms by about 25 per cent.[54] This effect was larger in those with milder injuries.

(c) Model of interaction

Lishman has proposed a model in which early disturbance of brain function after mild head injury results in the early symptoms of post-concussion syndrome.[1] In most patients these gradually resolve and a good recovery is made. However, the post-concussion syndrome may develop if psychological effects interfere with the normal process of recovery. Anxiety is thought to play a large part in impeding recovery; the patient worries about the symptoms and focuses on them. These may be aggravated if the patient is vulnerable to somatization, or there are compensation issues at stake. The symptoms may cause secondary disability provoking yet more anxiety, which will be made worse if there are additional psychosocial stressors. The role of psychological factors is greatest in those with very mild head injuries and very chronic symptoms.

Post-traumatic epilepsy

Early fits, within the first week, are relatively benign, sensitive to prophylactic anti-convulsants, and are only weak predictors of later epilepsy.

Only about 5 per cent of closed head injuries go on to develop late seizures, compared with 30 per cent after an open head injury. The majority of these late seizures start in the few years following injury. By the time 5 to 10 years have elapsed without seizures any subsequent seizure development may be unrelated to the head injury.[55]

The likelihood of developing seizures in patients with a closed head injury is increased by the presence of a depressed skull fracture, intracranial haematoma, and early seizure, as well as by the severity of the injury. Mild head injuries result in only a small increased risk of epilepsy above population norms. The EEG is generally not a good predictor of post-traumatic epilepsy.

Post-traumatic epilepsy increases psychiatric morbidity, particularly mood disorders, and may increase the risk of late dementia.

Prophylactic anti-convulsants have no effect on reducing the incidence of late post-traumatic epilepsy.[56] Carbamazepine, rather than phenytoin, is the drug of choice if an anti-convulsant is needed because it has less effect on cognition.[57] Half of all patients with post-traumatic epilepsy from open head injuries are found to be in remission by 5 to 10 years.

Head injury in children

Children, compared with adults, are more likely to suffer cerebral oedema and early post-traumatic epilepsy. They tend to develop a relatively stereotyped pattern of changes in personality with emotional lability, overactivity, reduced attention span, and irritability with outbursts of temper and rage. Apart from personality changes the commonest psychiatric disorders that follow childhood head injuries are attention-deficit hyperactivity disorder, and obsessive–compulsive disorder.[58] Children who develop attention-deficit hyperactivity disorder after head injury tend to demonstrate less hyperactivity than is seen in the idiopathic form.

It is sometimes said that the greater potential for plasticity which may be present in the younger person's brain, results in a better outcome compared to adults. There is some evidence for this for mild and moderate head injuries. However, children with severe head injuries are likely to be left with persistent cognitive deficits

and behavioural problems. Very young children with severe head injury suffer a double hazard[59] with both loss of acquired skills and interference with further development. This is often complicated by the fact that many of these children will have demonstrated pre-injury behavioural problems. The quality of parenting has a powerful effect on outcome; those with poor parenting are much more likely to develop behavioural problems.[60]

Boxing

In the past, when the average number of career bouts was about 300, 10 to 20 per cent of professional boxers went on to develop a chronic traumatic encephalopathy,[61] the punch-drunk syndrome. However now, with an average boxing career of 13 bouts, cases are much less frequently seen.[62]

Patients with chronic traumatic encephalopathy suffer damage to the extrapyramidal system as well as cerebellar and pyramidal pathways. They are slow and ataxic. Cognitive impairment, in particular memory impairment, is a frequent accompaniment and about 50 per cent have dementia. Upper brainstem lesions may explain the neurological symptoms, while cerebral atrophy, white matter changes, and damage to diencephalic structures may account for cognitive changes. Perforation of the septum pellucidum, which separates the two lateral ventricles, is a characteristic finding. *APOE e4* status increases vulnerability to the punch-drunk syndrome,[63] and this is consistent with the finding that amyloid is often present.

Professional footballers may show evidence of subtle impairments of thinking. This raises the possibility that repeated blows to the head from heading the ball may be sufficient to cause slight brain injury; but a more likely explanation for any injury is head-to-head contact.[64]

Management of early neurobehavioural problems

Interventions aimed at reducing the risk of enduring post-concussional symptoms after milder injuries, using brief educational, and supportive therapy in the early days post-injury, have been shown to be effective.[65]

But there is little evidence to guide the management of behavioural problems and mental symptoms arising in the days and weeks following a more severe brain injury. Such symptoms should be regarded as a flag to indicate the need to check on the progress of recovery. The history needs to be reviewed, paying attention to the period leading up to injury. The patient will need to be examined physically including a thorough neurological examination and checking for fever. It is essential to document the conscious level and orientation. Routine blood tests should be performed, and blood gases and a chest radiograph considered. Medication needs to be scrutinized. A neurological or neurosurgical opinion may be needed with a view to considering neuroimaging or an EEG. A lumbar puncture, for example looking for meningitis, should probably not be done without specialist advice.

Causes of deterioration after head injury are listed in Box 4.1.10.2. Once these have been excluded then the principle of care should be to allow recovery to take place in a safe environment, paying attention to the general principles of the care of the delirious or demented patient as indicated. Explanation to the patient and his or her family as to what is happening, is required.

Box 4.1.10.2 Causes of late deterioration in cognitive function after brain injury

Specific

- Subdural haematoma
- Hydrocephalus
- Epilepsy, particularly complex partial status
- Late intracranial infection, including cerebral abscess

Non-specific

- Systemic illness, including fat emboli and pain
- Drug intoxication
- Severe mental illness, in particular depression
- The patient 'gives up' as he or she gains insight
- Independent dementing process

Management of late mental sequelae

The reader is referred to Chapter 4.1.14 for many of the management principles relevant to patients with severe cognitive impairment after a head injury.

Psychological interventions

(a) Evidence of effectiveness

A small RCT has shown that cognitive behavioural treatment may be useful in those with persistent post-concussion syndrome.[66] Inpatient cognitive rehabilitation is probably needed only in those with more severe injuries.[46] Community therapy can improve handicap.[67] The evidence that behavioural strategies can improve behavioural problems, particularly aggression, in those with brain injury rests very largely on single case studies[68] or case series showing marked improvement in patients with very long-standing symptoms.[69]

(b) Principles of management

The first step, after medical issues have been excluded (Box 4.1.10.2), is to ensure that the patient has received adequate rehabilitation. In those with severe cognitive impairment once the patient is medically stable appropriate inpatient rehabilitation, perhaps on a locked unit, will probably be required. Problems arise if patients have to be sedated to ensure their safety, and the safety of other patients, while they remain inappropriately in an acute hospital bed. Timely access to rehabilitation is likely to reduce the risk of mental sequelae. Not infrequently psychological problems arise if any part of this process has not gone smoothly, or is perceived to have failed. Education, and access to information, is an important part of the care plan. Good advice on strategies for return to work can be invaluable, and some will require formal vocational rehabilitation. A social worker should be asked to undertake a community care assessment, which may identify the need for respite care or modifications to the home.

The management of any mental sequelae rests on a good understanding of the severity of the brain injury in order to estimate the likely contribution of brain damage to the mental sequelae.

A neuropsychological assessment, to determine injury severity and the pattern of impairments, may be needed. The severity of brain injury will suggest whether a particular symptom is mainly due to brain injury or to psychological processes. The degree of cognitive impairment may indicate whether or not the patient is capable of benefiting from certain psychological therapies. Those with less severe impairments should be offered CBT as appropriate, for example to treat travel anxiety or depression. In addition, the individual and their family should have access to support and guidance as they try to adjust to the changes forced on them by the head injury. Sometimes carers or family will need advice on how to manage challenging behaviours, particularly if their responses seem to be reinforcing the behaviour.

Pharmacological management

By and large patients should be given psychotropics only if absolutely necessary,[70] attending to the principles described in Box 4.1.10.3 and avoiding multiple drugs given concurrently. There is an emerging literature on drugs which may enhance cognition.[71]

(a) Agitation and aggression

Evidence—There are no good trials of medication for agitation and aggression after brain injury.[72] Only β-blockers have been exposed to randomized controlled trials. These studies showed a slight effect in favour of medication. But despite this β-blockers are rarely used in the management of agitation and aggression. In the RCTs very large doses of β-blockers were used that will almost inevitably cause worrying side effects in most patients.

Management—Because of the lack of controlled trials, prescribing for agitation and aggression after brain injury is very much trial and error, requiring good monitoring and documentation of the behaviour. If there is no evidence of benefit then the drug should be withdrawn and another drug tried. Be wary of responding to every episode of aggression by increasing the dose or adding a new drug.

For many psychiatrists, based on little more than anecdotal evidence, valproate, or carbamazepine are the drugs of first choice for aggression after brain injury. They have the advantage of anticonvulsant as well as mood-stabilizing effects. Perhaps a third of patients will respond.

Antipsychotics should be used if delusions or persecutory ideas of reference or fear are also present. But akathisia may perpetuate agitated behaviour which would otherwise have resolved spontaneously. Atypical antipsychotics, having less risk of motor side effects, are to be recommended.

Antidepressants may be helpful particularly if symptoms of anxiety or depression are present. Selective serotonin-reuptake inhibitors should be given in preference to tricyclics. Trazodone given at night may be useful if there is sleep disturbance.

Benzodiazepines should be considered for agitation and aggression during the early recovery from severe head injury. But be wary of increasing the confusion, and paradoxical violence due to disinhibition. Because of the potential for addiction, benzodiazepines should not be given to a patient with a chronic aggressive disorder.

Box 4.1.10.3 Prescribing psychotropics in brain injury

No knee jerk reaction—if possible wait to see if the problem goes away spontaneously
Small doses—start low, go slow
Only continue treatment if good evidence of effect
Drug profile—choose drugs with less potential for:

- Lowering seizure threshold—avoid clozapine
- Anticholinergic activity—to minimize potential for increasing confusion
- Extrapyramidal side-effects—especially akathisia, parkinsonism, and neuroleptic malignant syndrome
- Enzyme induction or other pharmacodynamic interaction with other drugs

Regular medication with long-acting anxiolytics, compared with short-acting drugs as required, is less likely to produce:

- Withdrawal syndrome
- Development of addiction
- Reinforcement of unwanted behaviour but may produce raised blood concentrations

(b) Mood disorders and psychosis

Evidence—Good studies evaluating the efficacy of medication for depression or psychosis after head injury are lacking.[71]

Management—Depression after a head injury is probably more difficult to treat than in those without brain injury.[73] However, some studies in head-injured depressed patients have found good response rates. The selection of an antidepressant is no different from that used to treat depression in the absence of brain injury, provided that the principles given in Box 4.1.10.3 are taken into account.

Confabulations and delusions early after brain injury should be allowed to resolve spontaneously where possible. For established psychotic symptoms atypical antipsychotics are probably to be recommended.

(c) Apathy

Bromocriptine and methylphenidate may be useful for treating apathetic states but controlled trials are lacking. Methylphenidate, with its risk of addiction and troublesome side effects, should only be prescribed if bromocriptine has not been successful, and under consultant supervision.

(d) Drugs which enhance memory and concentration

Evidence—Preliminary evidence supports the use of methylphenidate for deficits in attention and speed of information processing,[74] donepezil for attention and memory problems[75] and bromocriptine for executive problems.[76]

Management—There is now a case for considering medication to enhance cognitive function after brain injury. Drugs are available that may result in small improvements in attention, memory, and executive function. However in almost every case the evidence relies heavily on a single fairly small randomized controlled trial. Longer-term adverse consequences are uncertain. These drugs should only be considered:

- for patients with definite moderate to severe brain injury to account for their cognitive complaints,

- after discussion with the patient and their family about the uncertainties of treatment,
- if there is good reason to believe that a small increase in cognitive function can result in significant improvement in handicap,
- with close monitoring of response and side effects.

Insight, capacity, and detention in hospital

Insight and capacity to consent to treatment should be assessed in all patients. Lack of awareness of deficits is a common problem for the head-injured person[77] and affects compliance with, and capacity to consent to treatment. The Mental Capacity Act, 2005 (http://www.opsi.gov.uk/acts/acts2005/20050009.htm or http://www.dca.gov.uk/menincap/legis.htm) governs decision-making on behalf of adults in England and Wales who lack capacity to consent to treatment and manage their affairs. In these patients if there is no family or friend to act on their behalf, an independent advocate may be needed. The clinical team will need to ensure that all reasonable measures have been taken to enable the patient to take part in any decision-making. Any decision to act in the patient's 'best interests' should take into account what is known about their previous views and opinions. Very occasionally advance decisions may be in place which dictate how the patient wishes to be treated.

The psychiatrist may be called when the patient demands to leave hospital against medical advice. Only the very exceptional patient who is demanding to leave hospital following a head injury, and who as a result would be putting his or her health severely at risk, will be found to be competent. If they are not, it may be necessary to consider detention under the Mental Health Act, 1983 (England and Wales) or equivalent.

Patients will also need to be assessed to see if they are capable of managing their finances and affairs. If they are not, appropriate legal arrangements should be made; in the United Kingdom a receiver may need to be appointed to protect their interests. The prospect of compensation should be considered and, if appropriate, they should be enabled to pursue a personal injury claim.

Further information

http://www.ninds.nih.gov/disorders/tbi/tbi.htm National Institute of Neurological Disorders and Stroke—Traumatic brain injury information page.

http://www.dh.gov.uk/en/Healthcare/NationalServiceFrameworks/LongtermNeurologicalConditionsNSF/index.htm The National Service Framework on Long term (neurological) conditions. This provides guidance for health and social services on therapy and support for people with long term neurological conditions. The guidance is very relevant to patients with traumatic brain injury.

Silver, J.M., McAllister, T.W., and Yudofsky, S.C. (2005). *Textbook of traumatic brain injury.* American Psychiatric Publishing Inc., Washington, DC. A large multi-author textbook on all aspects of traumatic brain injury.

Damasio, A.R. (1994). *Descartes' error: emotion, reason and the human brain.* Grosset/Putnam, New York.

Lishman, W.A. (1998). *Organic psychiatry: the psychological consequences of cerebral disorder* (3rd edn). Blackwell Science, Oxford.

References

1. Lishman, W.A. (1988). Physiogenesis and psychogenesis in the 'post-concussional syndrome'. *The British Journal of Psychiatry*, **153**, 460–9.
2. Kapur, N. (1997). *Injured brains of medical minds: views from within.* Oxford University Press, Oxford.
3. National Collaborating Centre for Acute Care. (2003). *Head injury: triage, assessment, investigation and early management of head injury in infants, children and adults.* Guideline commissioned by the National Institute for Clinical Excellence. http://www.nice.org.uk/guidance/CG4
4. Povlishock, J.T. and Katz, D.I. (2005). Update of neuropathology and neurological recovery after traumatic brain injury. *The Journal of Head Trauma Rehabilitation*, **20**, 76–94.
5. Hayes, R.L., Lyeth, B.G., and Jenkins, L.W. (1989). Neurochemical mechanisms of mild and moderate head injury; implications for treatment. In *Mild head injury* (eds. H. S. Levin, H. M. Eisenberg, and A. L. Benton), pp. 54–79. Oxford University Press, Oxford.
6. Bishara, S., Partridge, F., Godfrey, H., *et al.* (1992). Post-traumatic amnesia and Glasgow Coma Scale related to outcome in survivors in a consecutive series of patients with severe closed head injury. *Brain Injury*, **6**, 373–80.
7. Teasdale, G. and Jennett, B. (1974). Assessment of coma and impaired consciousness. A practical scale. *Lancet*, **2**, 81–4.
8. Kraus, J.F. and Chu, L.D. (2005). Epidemiology. In *Textbook of traumatic brain injury* (eds. J.M. Silver, T.W. McAllister, and S.C. Yudofsky), pp. 3–26. American Psychiatric Publishing, Washington, DC.
9. Parizel, P.M., Van Goethem, J.W., Ozsarlak, O., *et al.* (2005). New developments in the neuroradiological diagnosis of craniocerebral trauma. *European Radiology*, **15**, 569–81.
10. Goldenberg, G., Oder, W., Spatt, J., *et al.* (1992). Cerebral correlates of disturbed executive function and memory in survivors of severe closed head injury: a SPECT study. *Journal of Neurology, Neurosurgery, and Psychiatry*, **55**, 362–8.
11. Markowitsch, H.J., Kessler, J., Van Der Van, C., *et al.* (1998). Psychic trauma causing grossly reduced brain metabolism and cognitive deterioration. *Neuropsychologia*, **36**, 77–82.
12. Nelson, H.E. and Willison, J.R. (1991). *National adult reading test* (2nd edn). NFER–Nelson, Windsor.
13. Wilson, B.A., Alderman, N., Burgess, P.W., *et al.* (1996). *Behavioural assessment of the dysexecutive syndrome.* Thames Valley Test Co., Bury St Edmunds, Suffolk.
14. Ruff, R.M., Young, D., Gautille, T., *et al.* (1991). Verbal learning deficits following severe head injury: heterogeneity in recovery over 1 year. *Journal of Neurosurgery*, **75**, S50–8.
15. Wilson, B. (1992). Recovery and compensation strategies in head injured memory impaired people several years after insult. *Journal of Neurology, Neurosurgery, and Psychiatry*, **55**, 177–80.
16. Whitnall, L., McMillan, T.M., Murray, G.D., *et al.* (2006). Disability in young people and adults after head injury: 5–7 year follow up of a prospective cohort study. *Journal of Neurology, Neurosurgery, and Psychiatry*, **77**, 640–5.
17. Corkin, S., Rosen, J., Sullivan, E.V., *et al.* (1989). Penetrating head injury in young adulthood exacerbates cognitive decline in later years. *Journal of Neuroscience*, **9**, 3876–83.
18. Fleminger, S., Oliver, D.L., Lovestone, S., *et al.* (2003). Head injury as a risk factor for Alzheimer's disease: the evidence 10 years on; a partial replication. *Journal of Neurology, Neurosurgery, and Psychiatry*, **74**, 857–62.
19. Teasdale, T.W. and Engberg, A. (1997). Duration of cognitive dysfunction after concussion, and cognitive dysfunction as a risk factor: a population study of young men. *British Medical Journal*, **315**, 569–72.
20. Lishman, W.A. (1968). Brain damage in relation to psychiatric disability after head injury. *The British Journal of Psychiatry*, **114**, 373–410.
21. Pilowsky, I. (1985). Cryptotrauma and 'accident neurosis'. *The British Journal of Psychiatry*, **147**, 310–11.

22. Claparède, E. (1911). Récognition et moïïté. *Archives of Psychology* (*Genève*), **11**, 79–90.
23. Whitty, C.W.M. and Zangwill, O.L. (1966). Traumatic amnesia. In *Amnesia* (eds. C.W.M. Whitty and O.L. Zangwill), pp. 92–108. Butterworths, London.
24. Shallice, T. and Burgess, P.W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, **114**, 727–41.
25. Brooks, D.N. (1975). Long and short term memory in head injured patients. *Cortex*, **11**, 329–40.
26. Thomsen, I.V. (1984). Late outcome of very severe blunt head trauma: a 10–15 year second follow-up. *Journal of Neurology, Neurosurgery, and Psychiatry*, **47**, 260–8.
27. Lezak, M.D. (1978). Living with the characterologically altered brain injured patient. *The Journal of Clinical Psychiatry*, **39**, 592–8.
28. Parker, R.S. (1996). The spectrum of emotional distress and personality changes after minor head injury incurred in a motor vehicle accident. *Brain Injury*, **10**, 287–302.
29. Blumer, D. and Benson, D.F. (1978). Personality changes with frontal and temporal lobe lesions. In *Psychiatric aspects of neurologic disease* (eds. D.F. Benson and D. Blumer), pp. 151–70. Grune and Stratton, New York.
30. Hibbard, M.R., Bogdany, J., Uysal, S., *et al.* (2000). Axis II psychopathology in individuals with traumatic brain injury. *Brain Injury*, **14**, 45–61.
31. Thomsen, I.V. (1974). The patient with severe head injury and his family: a follow-up study of 50 patients. *Scandinavian Journal of Rehabilitation Medicine*, **6**, 180–3.
32. Olver, J.H., Ponsford, J.L., and Curran, C.A. (1996). Outcome following traumatic brain injury: a comparison between 2 and 5 years after injury. *Brain Injury*, **10**, 841–8.
33. Symonds, C.P. (1937). Mental disorder following head injury. *Proceedings of the Royal Society of Medicine*, **30**, 1081–94.
34. Brooke, M.M., Questad, K.A., Patterson, D.R., *et al.* (1992). Agitation and restlessness after closed head injury: a prospective study of 100 consecutive admissions. *Archives of Physical Medicine & Rehabilitation*, **73**, 320–3.
35. Weinstein, E.A. and Kahn, R.L. (1951). Patterns of disorientation in organic brain disease. *Journal of Neuropathology and Clinical Neurology*, **1**, 214.
36. Pick, A. (1903). Clinical studies. III: on reduplicative paramnesia. *Brain*, **26**, 260–7.
37. Fleminger, S. and Burns, A. (1993). The delusional misidentification syndromes in patients with and without evidence of organic cerebral disorder: a structured review of case reports. *Biological Psychiatry*, **33**, 22–32.
38. Davison, K. and Bagley, C.R. (1969). Schizophrenia-like psychoses associated with organic disorders of the central nervous system: a review of the literature. In *Current problems in neuropsychiatry: schizophrenia, epilepsy, the temporal lobe* (ed. R.N. Herrington), *The British Journal of Psychiatry*, Special Publication No. 4, pp. 113–84. Headley Brothers, Ashford.
39. David, A.S. and Prince, M. (2005). Psychosis following head injury: a critical review. *Journal of Neurology, Neurosurgery, and Psychiatry*, **76**(Suppl. 1), i53–60.
40. Nielsen, A.S., Mortensen, P.B., O'Callaghan, E., *et al.* (2002). Is head injury a risk factor for schizophrenia? *Schizophrenia Research*, **55**, 93–8.
41. Harrison, G., Whitley, E., Rasmussen, F., *et al.* (2006). Risk of schizophrenia and other non-affective psychosis among individuals exposed to head injury: case control study. *Schizophrenia Research*, **88**, 119–26.
42. Fleminger, S., Oliver, D.L., Williams, W.H., *et al.* (2003). The neuropsychiatry of depression after brain injury. *Neuropsychological Rehabilitation*, **13**, 65–87.
43. Jorge, R.E., Robinson, R.G., Arndt, S.V., *et al.* (1993). Depression following traumatic brain injury: a 1 year longitudinal study. *Journal of Affective Disorder*, **27**, 233–43.
44. Deb, S., Lyons, I., Koutzoukis, C., *et al.* (1999). Rate of psychiatric illness 1 year after traumatic brain injury. *The American Journal of Psychiatry*, **156**, 374–8.
45. Warden, D.L. and Labbate, L.A. (2005). Posttraumatic stress disorder and other anxiety disorders. In *Textbook of traumatic brain injury* (eds. J.M. Silver, T.W. McAllister, and S.C. Yudofsky), pp. 231–44. American Psychiatric Publishing, Washington, DC.
46. Salazar, A.M., Warden, D.L., Schwab, K., *et al.* (2000). Cognitive rehabilitation for traumatic brain injury: a randomized trial. Defense and Veterans Head Injury Program (DVHIP) study group. *The Journal of the American Medical Association*, **283**, 3075–81.
47. Teasdale, T.W. and Engberg, A.W. (2001). Suicide after traumatic brain injury: a population study. *Journal of Neurology, Neurosurgery, and Psychiatry*, **71**, 436–40.
48. Mysiw, W.J. and Sandel, M.E. (1997). The agitated brain injured patient. Part 2: Pathophysiology and treatment. *Archives of Physical Medicine and Rehabilitation*, **78**, 213–20.
49. Rönty, H., Ahonen, A., Tolonen, U., *et al.* (1993). Cerebral trauma and alcohol abuse. *European Journal of Clinical Investigation*, **23**, 182–7.
50. Sparadeo, F.R. and Gill, D. (1989). Effects of prior alcohol use on head injury recovery. *The Journal of Head Trauma Rehabilitation*, **4**, 75–82.
51. Ryan, L.M. and Warden, D.L. (2003). Post concussion syndrome. *International Review of Psychiatry*, **15**, 310–6.
52. Povlishock, J.T. and Coburn, T.H. (1989). Morphological change associated with mild head injury. In *Mild head injury* (eds. H.S. Levin, H.M. Eisenberg, and A.L. Benton), pp. 37–53. Oxford University Press, New York.
53. McAllister, T.W., Saykin, A.J., Flashman, L.A., *et al.* (1999). Brain activation during working memory 1 month after mild traumatic brain injury: a functional MRI study. *Neurology*, **53**, 1300–8.
54. Binder, L.M. and Rohling, M.L. (1996). Money matters: a meta-analytic review of the effects of financial incentives on recovery after closed-head injury. *The American Journal of Psychiatry*, **153**, 7–10.
55. Annegers, J.F., Grabow, J.D., Groover, R.V., *et al.* (1980). Seizures after head trauma: a population study. *Neurology*, **30**, 683–9.
56. Schierhout, G. and Roberts, I. (1998). Prophylactic antiepileptic agents after head injury: a systematic review. *Journal of Neurology, Neurosurgery, and Psychiatry*, **64**, 108–12.
57. Wroblewski, B.A., Glenn, M.B., Whyte, J., *et al.* (1989). Carbamazepine replacement of phenytoin, phenobarbital and primidone in a rehabilitation setting: effects on seizure control. *Brain Injury*, **3**, 149–56.
58. Max, J.E. (2005). Children and adolescents. In *Textbook of traumatic brain injury* (eds. J.M. Silver, T.W. McAllister, and S.C. Yudofsky), pp. 477–94. American Psychiatric Publishing, Washington, DC.
59. Anderson, V., Catroppa, C., Morse, S., *et al.* (2005). Functional plasticity or vulnerability after early brain injury? *Pediatrics*, **116**, 1374–82.
60. Max, J.E., Robin, D.A., Lindgren, S.D., *et al.* (1997). Traumatic brain injury in children and adolescents: psychiatric disorders at two years. *Journal of the American Academy of Child and Adolescent Psychiatry*, **36**, 1278–85.
61. Roberts, A.H. (1969). *Brain damage in boxers*. Pitman, London.
62. Clausen, H., McCrory, P., and Anderson, V. (2005). The risk of chronic traumatic brain injury in professional boxing: change in exposure variables over the past century. *British Journal of Sports Medicine*, **39**, 661–4.
63. Jordan, B.D., Relkin, N.R., Ravdin, L.D., *et al.* (1997). Apolipoprotein E e4 associated with chronic traumatic brain injury boxing. *The Journal of the American Medical Association*, **278**, 136–40.
64. McCrory, P.R. (2003). Brain injury and heading in soccer. *British Medical Journal*, **327**, 351–2.
65. Wade, D.T., King, N.S., Wenden, F.J., *et al.* (1998). Routine follow up after head injury: a second randomised controlled trial. *Journal of Neurology Neurosurgery, and Psychiatry*, **65**, 177–83.

66. Tiersky, L.A., Anselmi, V., Johnston, M.V., *et al.* (2005). A trial of neuropsychologic rehabilitation in mild-spectrum traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, **86**, 1565–74.
67. Powell, J., Heslin, J., and Greenwood, R. (2002). Community based rehabilitation after severe traumatic brain injury: a randomised controlled trial. *Journal of Neurology, Neurosurgery, and Psychiatry*, **72**, 193–202.
68. Alderman, N., Davies, J.A., Jones, C., *et al.* (1999). Reduction of severe aggressive behaviour in acquired brain injury: case studies illustrating clinical use of the OAS-MNR in the management of challenging behaviours. *Brain Injury*, **13**, 669–704.
69. Eames, P. and Wood, R. (1985). Rehabilitation after severe brain injury: a follow-up study of a behaviour modification approach. *Journal of Neurology, Neurosurgery, and Psychiatry*, **48**, 613–9.
70. Goldstein, L.B. (1995). Prescribing of potentially harmful drugs to patients admitted to hospital after head injury. *Journal of Neurology, Neurosurgery, and Psychiatry*, **58**, 753–5.
71. Neurobehavioral Guidelines Working Group. Warden, D.L., Gordon, B., McAllister, T.W., *et al.* (2006). Guidelines for the pharmacologic treatment of neurobehavioral sequelae of traumatic brain injury. *Journal of Neurotrauma*, **23**, 1468–501.
72. Fleminger, S., Greenwood, R.J., and Oliver, D.L. (2006). Pharmacological management for agitation and aggression in people with acquired brain injury. *Cochrane Database of Systematic Reviews*, **4**, CD003.
73. Dinan, T.G. and Mobayed, M. (1992). Treatment resistance of depression after head injury: a preliminary study of amitriptyline response. *Acta Psychiatrica Scandinavica*, **85**, 292–4.
74. Whyte, J., Hart, T., Vaccaro, M., *et al.* (2004). Effects of methylphenidate on attention deficits after traumatic brain injury: a multidimensional, randomized, controlled trial. *American Journal of Physical Medicine & Rehabilitation*, **83**, 401–20.
75. Zhang, L., Plotkin, R.C., Wang, G., *et al.* (2004). Cholinergic augmentation with donepezil enhances recovery in short-term memory and sustained attention after traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, **85**, 1050–5.
76. McDowell, S., Whyte, J., and D'Esposito, M. (1998). Differential effect of a dopaminergic agonist on prefrontal function in traumatic brain injury patients. *Brain*, **121**, 1155–64.
77. Prigatano, G.P. and Schacter, D.L. (1991). *Awareness of deficit after brain injury: clinical and theoretical issues*. Oxford University Press, New York.

4.1.11 Alcohol-related dementia (alcohol-induced dementia; alcohol-related brain damage)

Jane Marshall

Introduction

Long-term heavy alcohol consumption causes significant brain abnormalities and impairs cognitive functioning. A number of terms have been used to describe these effects, including: 'alcohol-related dementia', 'alcohol-induced dementia', and 'alcoholic dementia'.[1] The more pragmatic umbrella term 'alcohol-related brain damage' (ARBD) is also used. The literature is beset with limitations, in particular the lack of a diagnostic gold standard, and the difficulty in making a clinical diagnosis. Many individuals labelled as having an alcohol-related dementia are, in fact, suffering from the Wernicke–Korsakoff syndrome (WKS).[2] (This is a specific neuropathological disease caused by thiamine deficiency, which can occur secondary to alcohol misuse. It is considered in Chapter 4.1.12.) When considering the topic of 'alcohol-related dementia' it is probably sensible to take a broad clinically-based diagnostic view that includes both WKS and other cases of 'dementia' that appear to be alcohol-related.[3]

Diagnostic criteria

Diagnostic criteria for 'substance-induced persisting dementia' are included in DSM-IV[4] (Table 4.1.11.1), which also states that there must be evidence from the history, physical examination, or laboratory findings that the deficits are aetiologically related to the persisting effects of substance use (in this case alcohol). No specific inclusion criteria are offered to distinguish alcohol-related dementia from other dementias. In ICD-10,[5] the Korsakoff syndrome is listed separately under the amnesic syndrome heading (F10.6) whereas alcohol-induced 'dementia' and 'other persisting cognitive impairment' are included under the 'residual and late-onset psychotic disorder' category (F10.73 and F10.74 respectively), where diagnostic guidelines can be found.

Diagnostic criteria for establishing a diagnosis of 'alcohol-related dementia' have been proposed, conceiving it as a spectrum of alcohol-related intellectual and neurological syndromes, ranging from moderate deficits to the more severe Wernicke–Korsakoff syndrome.[3] 'Alcohol-related dementia' is thus defined as a syndrome that results from several aetiological mechanisms including the direct neurotoxic effects of alcohol, metabolic dysfunction during intoxication and withdrawal, trauma, vascular injury and thiamine or other nutritional deficiencies.

Table 4.1.11.1 DSM-IV diagnostic criteria for substance-induced persisting dementia

A.	The development of multiple cognitive deficits manifested by both (1) memory impairment (impaired ability to learn new information or to recall previously learned information) (2) one (or more) of the following cognitive disturbances: (a) aphasia (language disturbance) (b) apraxia (impaired ability to carry out motor activities despite intact motor function) (c) agnosia (failure to recognize or identify objects despite intact motor sensory function) (d) disturbance in executive functioning (i.e. planning, organization, sequencing, abstracting)
B.	The cognitive deficits in criteria A1 and A2 each cause significant impairment in social or occupational functioning and represent a significant decline from a previous level of functioning.
C.	The deficits do not occur exclusively during the course of a delirium and persist beyond the usual duration of substance intoxication or withdrawal
D.	There is evidence from the history, physical examination, or laboratory findings that the deficits are aetiologically related to the persisting effects of substance use (e.g. a drug of abuse, a medication)

Prevalence

Adequate epidemiological studies to determine the size of the problem have not been carried out. It has been estimated that 'alcohol-related dementia' accounts for 10 per cent of the dementia population.[1] Indeed alcohol misuse may contribute to as many as 21–24 per cent of all cases of cognitive impairment in mid-adulthood.[6] The prevalence is likely to be higher in areas of socio-economic deprivation, with most cases presenting between the ages of 50 and 60 years.[7] Early onset has been associated with poorer prognosis and potential for recovery. Recent evidence suggests that the prevalence of the Wernicke–Korsakoff syndrome, caused by thiamine deficiency, may be increasing.[8,9] Early identification and intervention can help to maximize optimum recovery.

Causal mechanisms

There is no single cause of 'alcohol-related dementia'. Individual susceptibility may be influenced by age; age of onset of drinking and the drinking history; gender; genetic background; family history of alcohol dependence; nutrition; alcohol exposure before birth; and general health status.[10] Causal mechanisms include: the neurotoxic effect of alcohol and its metabolite acetaldehyde; repeated episodes of intoxication and withdrawal; dietary neglect and vitamin deficiencies; repeated episodes of head trauma; cerebrovascular events; and liver damage. In particular, thiamine depletion, and metabolic factors, such as hypoxia, electrolyte imbalance, and hypoglycaemia, all of which result from acute or chronic intoxication and withdrawal, are important and interrelated. It is difficult to determine the relative contributions of these mechanisms. A number of theories have been advanced by Lishman and others to explain the mechanisms by which chronic alcohol use might lead to dementia.[1,6]

- The brain might be vulnerable to both thiamine depletion and alcohol neurotoxicity, the former affecting the basal brain regions and the latter both the basal brain and the frontal cortex.[1] Individual genetic vulnerability is likely to have a role in influencing these processes.
- Wernicke–Korsakoff pathological processes in the basal brain have the potential to damage nearby cholinergic fibres projecting to the cerebral cortex: the so-called cholinergic hypothesis.[1,6]
- Alcohol-induced brain pathology couples with other processes including 'ageing, trauma, vascular changes, and hepatic dysfunction' leading to cognitive decline: the coupling hypothesis.[1,6]
- Ethanol stimulates pituitary corticotrophin leading to elevated corticosteroid levels and possible injury to the hippocampus.
- Recurrent alcohol withdrawal has been hypothesized to have a kindling effect.[11] During alcohol withdrawal there is increased *N*-methyl-d-aspartate (NMDA) function which is postulated to lead to increased neuronal excitability and to glutamate-induced neurotoxicity.[12] The way in which alcohol interferes with glutamatergic neurotransmission, especially through the NMDA receptor, is probably central to an understanding of its long-term effects on the brain.
- Alcohol might lead to an accelerated ageing process.

Areas of the brain affected

There is evidence that the frontal lobes and sub-cortical areas such as the limbic system, the thalamus and the basal forebrain are particularly vulnerable to alcohol-related damage. The cerebellum is also vulnerable. Alcohol-related brain changes in the frontal lobes become more prominent with age.[13] Emotional processing is affected by long-standing heavy alcohol use and dependence, and probably reflects abnormalities in the limbic system and the frontal lobes.[14] This is manifested as difficulty with interpreting non-verbal emotional cues and recognizing facial expressions of emotion.

Alcohol-related brain damage has been studied using a variety of methods, ranging from the neuropathology of the post-mortem alcoholic brain to neuro-imaging techniques focusing on structural, functional and biochemical changes. There is also a considerable neuropsychological literature.

Neuropathology

Early neuropathological studies of the alcoholic brain described fairly uniform cerebral atrophy, mainly over the dorso-lateral frontal regions, widened sulci, a narrowed cortical ribbon, and enlargement particularly of the anterior horns of the lateral ventricles.[1]

The reduction in cerebral volume seen in the alcoholic brain is due mainly to the loss of white matter in the cerebral hemispheres.[15] The reduced white matter is not related to changes in hydration or changes in the chemical structure of the myelin. Selective neuronal loss in the superior frontal cortex was reported in one study[15] but not confirmed in another.[16] However, there is evidence that individual neurones are shrunken in regions where neuronal numbers are normal, such as the superior frontal, cingulate, and motor cortices.[15,16]

Animal research suggests that alcohol has a direct neurotoxic effect on the brain. Chronic ingestion of ethanol by well-nourished rats has been shown to be toxic to cholinergic projection neurones[17] and to reduce the complexity of dendritic arborization in hippocampal pyramidal neurones.[18] In the former study, transplantation of cholinergic neurones into the hippocampus and neocortex corrected the cholinergic deficits and memory abnormalities. In the latter, abstinence led to an increase in dendritic arborization.

Structural neuroimaging

Neuroimaging studies (CT and magnetic resonance imaging (MRI)) have compared recently detoxified alcoholics without obvious cognitive impairment with age-matched controls. CT studies confirmed diffuse atrophy of brain tissue, with the frontal lobes showing most extensive shrinkage.[19] Follow-up studies showed that abstinence was associated with reversibility of brain shrinkage,[19] particularly in younger individuals and in women.[20]

Structural MRI studies have reported reduced volume of both grey and white matter in the cerebral cortex, especially the frontal lobes, which are used for reasoning, judgement, and problem solving,[13] particularly in older age groups. Changes have also been shown in other structures involved in memory, such as the hippocampus (in adolescents and adults), mammillary bodies, thalamus and cerebellar cortex.[21–24] Other abnormalities include thinning of the corpus callosum and reduced volume in the pons.[25] Reduced white-matter volume is also seen in the temporal lobes (in alcohol dependent subjects with seizures) and in the cerebellar

vermis where the loss is associated with deficits in postural stability.[26] More recent MRI studies have not supported the idea of increased vulnerability among women. [27] Abstinence is associated with recovery of tissue volume.

Functional neuroimaging

Functional neuroimaging studies have reported hypometabolism in the frontal and parietal cortices of chronic alcoholics without major neurological impairment, when compared with normal controls.[28–31] These abnormalities improve following abstinence,[31,32] mainly during the 16 to 30 days after the last use of alcohol. Metabolic recovery is most marked in the frontal area.[31]

Proton magnetic resonance spectroscopy can be combined with MRI, allowing ***in vivo*** insight into brain metabolism.[33–35] The metabolic changes observed in the few magnetic resonance spectroscopy studies that have been carried out suggest neuronal loss and compensatory gliosis.

Neuropsychology

Many individuals with a history of chronic excessive alcohol consumption show evidence of moderate impairment in short- and long-term memory, learning, visuoperceptual abstraction, visuospatial organization, the maintenance of cognitive set, and impulse control.[36] This tendency for alcoholics to show proportionally greater visuospatial than language-related impairments suggests that alcohol might have a selective effect on the right hemisphere: the so-called 'right hemisphere hypothesis'.[37] However, right hemisphere functions also decline with ageing and the current view is that the functional lateralities of 'alcoholics' and ageing individuals are similar to normal controls.[37]

Neuropsychological performance improves with abstinence. However, impairments can be detected in apparently healthy, abstinent alcohol dependent individuals[38] and are still detectable even after 5 years of abstinence.[39] Performance on neuropsychological tests has generally been poorly correlated with structural imaging changes,[19,40] particularly with changes in grey-matter volume. However, one MRI study reported significant correlations between cortical (sulcal) and subcortical (ventricular) fluid volumes and some cognitive measures.[22] Another study, using a combination of structural (CT or MRI) and functional imaging (positron emission tomography) together with neuropsychological tests in older alcohol-dependent patients who were abstinent, found a significant correlation between degree of atrophy/metabolic functioning in the cingulate gyrus, and performance on the Wisconsin Card Sort Test.[41]

Neuropsychological test scores do not predict outcome in alcohol-dependent patients.[42, 43]

Management

Difficulties in establishing a diagnosis of alcohol-related dementia/brain damage mean that it remains an 'invisible disability'[7], usually goes unrecognized, and is often masked by other problems such as continuing alcohol consumption and withdrawal, physical ill-health, depression and associated traumatic brain damage.

All dementia work-ups should include a history of past and present alcohol use, confirmed with a collateral history.[6] Appropriate treatment of alcohol withdrawal syndromes, assessment and re-assessment should be carried out over a two-year period. Ongoing assessment and care planning are important as these patients have the capacity to improve with abstinence. The possibility of Wernicke–Korsakoff pathology in cognitively impaired patients with an alcohol use disorder should prompt swift and appropriate treatment with parenteral thiamine.[44] Oral B vitamins should be continued long-term.

The mainstay of long-term treatment in alcohol-related dementia is abstinence. This can be facilitated by a supportive non-drinking social network, and cognitive behavioural methods to teach recognition of factors that predispose to relapse and alternative coping strategies.[6] Families and care-givers facilitate success and must be actively educated and supported. A rehabilitation approach to activities of daily living and occupation is also a key factor.

Patients with alcohol-related dementia are younger and more physically active than the usual dementia population. They do not fit neatly into any category of care and are at risk of falling 'through the net'. Services lack the capacity to manage this population so they are passed between services and find it difficult to access specialist assessment or care.

Conclusions

Alcohol-related dementia should be recognized as a preventable condition. However, identification is hampered by a lack of clarity in terminology, and a lack of standardized and specialized screening instruments and assessment procedures.[45] These individuals make repeated use of Accident and Emergency Departments, general medical, and long stay wards. Early identification would reduce their need for these services. Abstinence is the key to recovery. Treatment services should be integrated and flexible.

Further information

Baddeley, A.D., Kopelman, M/D., Wilson, B.A. (2003). *The Handbook of Memory Disorders*, (2nd edn.) Chichester: John Wiley and Sons Ltd.

Edwards, G., Marshall, E.J., Cook, C.C.H. (2003). *The Treatment of Drinking Problems: a Guide for the Helping Professions*, (4th edn.) Cambridge: Cambridge University Press

Lishman, W.A. (1998). *Organic Psychiatry. The Psychological Consequences of Cerebral Disorder*, (3rd edn). Oxford: Blackwell Science.

The National Institute of Alcohol Abuse and Alcoholism website has a portal which supports researchers and practitioners searching for information relating to alcohol research. It has a number of links to other databases: http://etoh.niaaa.nih.gov/

References

1. Lishman, W.A. (1990). Alcohol and the brain. *British Journal of Psychiatry*, **156**, 635–44.
2. Victor, M. (1993) Persistent altered mentation due to ethanol. *Neurologic Clinics,* **11**, 639–61.
3. Oslin, D., Atkinson, R.M, Smith, D.M., *et al* (1998). Alcohol related dementia: proposed clinical criteria. *International Journal of Geriatric Psychiatry,* **13**, 203–12.
4. American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.

5. World Health Organization (1992). *International statistical classification of diseases and related health problems, 10th revision*. WHO, Geneva.
6. Smith, D. and Atkinson, R. (1995). Alcohol and dementia. *International Journal of Addictions,* **30**, 1843–69.
7. MacRae, S. and Cox, S. (2003). *Meeting the needs of people with alcohol-related brain damage: a literature review on the existing and recommended service provision and models of care*. University of Stirling, Scotland.
8. Smith, I. and Flanigan, C. (2000). Korsakoff's psychosis in Scotland. Evidence for increased prevalence and regional variation. *Alcohol and Alcoholism,* **35**, Suppl 1, 8–10.
9. Ramayya, A. and Jauher, P. 1997). Increasing incidence of Korsakoff's Psychosis in the east end of Glasgow. *Alcohol and Alcoholism*, **32**, 281–85.
10. Oscar-Berman M., Marinkovic, K. (2003) Alcoholism and the brain: an overview. *Alcohol Research and Health,* **27**, 125–33.
11. Ballenger, J.C. and Post, R.M. (1978). Kindling as a model for alcohol withdrawal syndromes. *British Journal of Psychiatry*, **33**, 1–14.
12. Tsai, G., Gastfriend, D.R., and Coyle, J.T. (1995). The glutamatergic basis of human alcoholism. *American Journal of Psychiatry*, **152**, 332–40.
13. Pfefferbaum, A., Sullivan, E.V., Mathalon, D.H., *et al.* (1997). Frontal lobe volume loss observed with magnetic resonance imaging in older chronic alcoholics. *Alcoholism: Clinical and Experimental Research*, **21**, 521–9.
14. Kornreich, C., Philippot, P., Foisy, M.L. (2002) Impaired emotional facial expression recognition is associated with interpersonal problems in alcoholism. *Alcohol and Alcoholism,* **37**, 394–400.
15. Kril, J.J. and Harper, C.G. (1989). Neuronal counts from four cortical regions in alcoholic brains. *Acta Neuropathologica*, **79**, 200–4.
16. Jensen, G.B. and Pakkenburg, B. (1993). Do alcoholics drink their neurones away? *Lancet*, **342**, 1201–4.
17. Arendt, T.A., Allen, Y., Sinden, J., *et al.* (1988). Cholinergic–rich brain transplants reverse alcohol–induced memory deficits. *Nature*, **332**, 448–50.
18. McMullan, P.A., Saint–Cyr, J.A., and Carlen, P.L. (1984). Morphological alterations in rat CA1 hippocampal pyramidal cell dendrites resulting from chronic ethanol consumption and withdrawal. *Journal of Comparative Neurology*, **225**, 111–18.
19. Ron, M.A. (1983). *The alcoholic brain: CT scan and psychological findings*. Psychological Medicine Monograph 3. Cambridge University Press.
20. Carlen, P.L. and Wilkinson, D.A. (1987). Reversibility of alcohol–related brain damage: clinical and experimental observations. *Acta Medica Scandinavica*, **222** (Suppl 717), 19–26.
21. Sullivan, E.V., Marsh, L., Mathalon, D.H., *et al* (1995). Anterior hippocampal volume deficits in nonamnesic, ageing chronic alcoholics. *Alcoholism: Clinical and Experimental Research*, **19**, 110–22.
22. Davila, M.D., Shear, P.K., Lane, B., *et al.* (1994). Mammillary body and cerebellar shrinkage in chronic alcoholics: an MRI and neuropsychological study. *Neuropsychology*, **8**, 433–44.
23. Shear, P.K., Sullivan, E.V., Lane, B.J., *et al.* (1996). Mammillary body and cerebellar shrinkage in chronic alcoholics with and without amnesia. *Journal of the International Neuropsychological Society*, **2**, 34–5.
24. De Bellis, M.D., Clark, D. B., Beers, S.R. *et al.* (2000). Hippocampal volume in adolescent-onset alcohol use disorders. *American Journal of Psychiatry,* **157**, 737–44.
25. Pfefferbaum, A., Lim, K.O., Desmond, J.E., *et al.* (1996). Thinning of the corpus callosum in older alcoholic men: a magnetic resonance imaging study. *Alcoholism: Clinical and Experimental Research*, **20**, 752–7.
26. Rosenbloom, M., Sullivan, E.V., Pfefferbaum A. (2003). Using magnetic resonance and diffusion tensor imaging to assess brain damage in alcoholics. *Alcohol Research and Health,* **27**, 146–52.
27. Pfefferbaum, A., Rosenbloom, M.J., Deshmukh, A., *et al.* (2002). Sex differences in the effects of alcohol on brain structure. *American Journal of Psychiatry,* **158**, 188–97.
28. Sachs, H., Russell, J.A.G., Christman, D.R., and Cook, B. (1987). Alteration of regional cerebral glucose metabolic rate in non–Korsakoff chronic alcoholism. *Archives of Neurology*, **44**, 1242–51.
29. Volkow, N.D., Hitzemann, R., Wang, G.–J., *et al.* (1992). Decreased brain metabolism in neurologically intact healthy alcoholics. *American Journal of Psychiatry*, **149**, 1016–22.
30. Volkow, N.D., Wang, G.–J., Hitzemann, R., *et al.* (1994). Recovery of brain glucose metabolism in detoxified alcoholics. *Americal Journal of Psychiatry*, **151**, 178–83.
31. Nicolas, J.M., Catafau, A.M., Estruch, R., *et al.* (1993). Regional cerebral blood flow—SPECT in chronic alcoholism: relation to neuropsychological testing. *Journal of Nuclear Medicine*, **34**, 1452–9.
32. Fein, G., Meyerhoff, D.J., Di Sclafani, V., *et al.* (1994). 1H magnetic resonance spectroscopic imaging separates neuronal from glial changes in alcohol–related brain atrophy. In *Alcohol and glial cells*, pp. 227–41. Research Monograph 27. National Institutes of Health, Bethesda, MD.
33. Martin, P.R., Gibbs, S.J., Nimmerrichter, A.A., *et al.* (1995). Brain proton magnetic resonance spectroscopy studies in recently abstinent alcoholics. *Alcoholism: Clinical and Experimental Research*, **4**, 1078–82.
34. Seitz, D., Widmann, U., Seeger, U., *et al.* (1999). Localised proton magnetic resonance spectroscopy of the cerebellum in detoxifying alcoholics. *Alcoholism: Clinical and Experimental Research*, **23**, 158–63.
35. Bloomer, C.W., Langleben, D.D., Meyerhoff, D.J. (2004) Magnetic resonance detects brainstem changes in chronic, active heavy drinkers. *Psychiatry Research,* **30**, 209-18.
36. Oscar–Berman, M. (2000). Neuropsychological vulnerabilities in chronic alcoholism. In: Noronha, A., Eckardt, M.J., and Warren, K., eds. *Review of NIAAA's Neuroscience and Behavioural Research Portfolio*. National Institute on Alcohol Abuse and Alcoholism (NIAAA) Research Monograph No 34. Bethesda, MD: NIAAA, pp. 437–71.
37. Ellis, R.J. and Oscar-Berman, M. (1989). Alcoholism aging, and functional cerebral asymmetries. *Psychological Bulletin*, **106**, 128–47.
38. Davies, S.J.C., Pandit, S.A., Feeney, B., *et al.* (2005). Is there impairment in clinically 'healthy' abstinent alcohol dependence? *Alcohol and Dependence,* **40**, 498–503.
39. Brandt, J., Butters, N., Ryan, C., *et al.* (1983). Cognitive loss and recovery in long term alcohol abusers. *Archives of General Psychiatry*, **40**, 435–42.
40. Carlen, P.L., Wilkinson, D.A., and Wortzman, G., *et al.* (1981). Cerebral atrophy and functional deficits in alcoholics without clinically apparent liver disease. *Neurology*, **31**, 377–85.
41. Adams, K.M., Gilman, S., Koeppe, R.A., *et al.* (1993). Neuropsychological deficits are correlated with frontal hypometabolism in positron emission tomography studies of older alcoholic patients. *Alcoholism: Clinical and Expeimental Research,* **17,** 205–10.
42. Alterman, A.I., Kushner, H., and Holahan, J.M. (1990). Cognitive functioning and treatment outcome in alcoholics. *Journal of Nervous and Mental Disorders*, **178**, 494–9.
43. Leenane, K.J. (1988). Patients with alcohol–related brain damage: therapy and outcome. *Australian Drug and Alcohol Review*, **7**, 89–92.
44. Thomson A. D. and Marshall, E.J (2006) The treatment of patients at risk of developing Wernicke's Encephalopathy in the community. *Alcohol and Alcoholism*, **41**, 159–67.
45. Smith I., Hillman A (1999). Management of the alcoholic Korsakoff syndrome. *Advances in Psychiatric Treatment,* **5**, 271–75.

4.1.12 Amnesic syndromes

Michael D. Kopelman

Introduction

Amnesic disorders can be broadly classified across two orthogonal dimensions. Along the first dimension, there can be transient or discrete episodes of amnesia as opposed to persistent memory impairment. On the second dimension, memory loss can result from either neurological damage or psychological causation, although admixtures of these factors are, of course, very common. The notion of confabulation has traditionally been associated with amnesic syndromes, particularly the Korsakoff syndrome, although it may have a separate basis, and false memories are now known to arise in a number of different situations. With the advent of drugs purporting to influence memory, there is increasing interest in the psychopharmacology of memory disorders. This chapter will consider findings from investigations of patients with memory disorders and a few selected psychopharmacological studies of relevance. It will not review the extensive literature on functional imaging in normal subjects.

Transient amnesias

Transient global amnesia

Transient global amnesia (TGA) most commonly occurs in the middle-aged or elderly, more frequently in men, and it results in a period of amnesia lasting several hours. It is characterized by repetitive questioning, and there may be some confusion, but patients do not report any loss of personal identity (they know who they are). It is sometimes preceded by headache or nausea, a stressful life event, a medical procedure, or vigorous exercise. Hodges and Ward[1] found that the mean duration of amnesia was 4 h and the maximum was 12 h. In 25 per cent of their sample, there was a past history of migraine, which was considered to have a possible aetiological role. In a further 7 per cent of the sample, the patients subsequently developed unequivocal features of epilepsy (there had been no focal signs or features of epilepsy during the original attack) and the memory loss was therefore attributed, in retrospect, to previously undiagnosed epilepsy. There was no association with either a past history of vascular disease, clinical signs suggestive of vascular pathology, or known risk factors for vascular disease. In particular, there was no association with transient ischaemic attacks. In 60 to 70 per cent of the sample, the underlying aetiology was unclear.

More recently, Quinette *et al.*[2] reviewed the findings in 1353 patients reported in the clinical literature since 1956 and their own data from 142 patients, seen between 1994 and 2004. In general, the findings were consistent across the two sources. There was no sex bias, and the vast majority of attacks occurred between the ages of 50 and 80 (mean = 60.3 ± 9.6). Most patients had a single attack, but the annual rate of recurrence ranged from 2.9 per cent to 26.3 per cent (6.3 per cent in their own study). In the literature, the duration of attacks ranged from 15 min to 24 h and, in their own investigation, the range was 30 min to 16 h (mean = 5.6 h). These authors investigated putative predisposing and precipitating factors in great detail, concluding that TGA may encompass at least three different groups of patients: (i) younger patients with a history of migraine, in whom spreading neurochemical depression may be implicated, (ii) women who have experienced acute emotional or physical stress, and often have a history of anxiety or depression, and (iii) men who, following physical exertion, develop venous congestion in the context of insufficient jugular vein valves and a precipitating Valsalva manoeuvre.

In instances of TGA where neuropsychological tests have been administered to patients during the acute episode of memory loss,[1,3] the patients show a profound anterograde amnesia, as expected, on tests of both verbal and non-verbal memory. However, performance on tests of retrograde memory is variable. Follow-up studies show either complete or almost complete recovery of memories, several weeks to months after the acute attack. In general, retrograde amnesia recovers before anterograde amnesia; the degree of shrinkage of retrograde amnesia is heterogeneous; and anterograde memory (new learning) recovers gradually.

The general consensus is that the amnesic disorder results from transient dysfunction in limbic-hippocampal circuits, crucial to memory formation. Medial temporal abnormalities have been reported bilaterally in terms of single-photon emission CT (SPECT) measures of perfusion, positron emission tomography (PET) measures of metabolism, diffusion weighted imaging (DWI), and small hippocampal cavities on T2 reversed magnetic resonance imaging (MRI) images.[4,5] In addition, venous duplex sonography has shown jugular vein valve insufficiency in a proportion of cases.[5]

Transient epileptic amnesia

This term was coined by Kapur,[6] and it refers to the minority of patients with transient global amnesia in whom epilepsy appears to be the underlying cause of the syndrome.[1] Where epilepsy has not previously been diagnosed, the main predictive factors for an epileptic aetiology are brief episodes of memory loss (an hour or less) with multiple attacks.[1] It is important to note that standard electroencephalography (EEG) and CT findings are often normal. However, an epileptic basis to the disorder may be revealed on sleep EEG recordings.[1,7]

Patients with transient epileptic amnesia may show residual deficits in between their attacks, associated with their underlying neuropathology. Kopelman *et al.*[7] found a moderate degree of residual anterograde memory impairment in their patient, related to subsequent (unpublished) findings of small foci of MRI signal alteration and PET hypometabolism bilaterally in the medial temporal lobes. Several authors have reported patients who describe 'gaps' in past personal memories, and Manes *et al.*[8] have reported disproportionate inter-ictal retrograde amnesia. The latter group also reported abnormal long-term forgetting of verbal material, but whether these gaps in memory result from faulty encoding (because of subclinical ictal activity), impaired consolidation (giving rise to accelerated forgetting), or deficits in retrieval remains controversial.

Epilepsy may, of course, give rise to automatisms or post-ictal confusional states. Where there is an automatism in such circumstances, there is always bilateral involvement of the limbic structures involved in memory formation, including the hippocampal and parahippocampal structures bilaterally as well as the mesial diencephalon. Consequently, amnesia for the period of automatic behaviour is always present and is usually complete.

Head injury

In head injury, it is important to distinguish between a brief period of retrograde amnesia, which may last only a few seconds or minutes but can be weeks or months, a longer period of post-traumatic amnesia, and islands of preserved memory within the amnesic gap.[9] Occasionally post-traumatic amnesia may exist without any retrograde amnesia, although this is more common in cases of penetrating lesions. Sometimes there is a particularly vivid memory for images or sounds occurring immediately before the injury, on regaining consciousness, or during a lucid interval between the injury and the onset of post-traumatic amnesia. These vivid ºmemories may become the intrusive flashbacks of a post-traumatic stress disorder (PTSD) syndrome.

Post-traumatic amnesia (PTA) is generally assumed to reflect the degree of underlying diffuse brain pathology, in particular rotational forces giving rise to axonal tearing and generalized cognitive impairment. The length of PTA is predictive of eventual cognitive outcome, psychiatric outcome, and social outcome.[10] However, the duration of PTA is often not well documented in medical records, and these relationships are often weaker than is generally assumed. In addition, contusion to the frontal and anterior temporal lobes is a common consequence of head injury. The clinical features and underlying pathophysiology of head injury have recently been well described elsewhere.[11]

Post-traumatic amnesia needs to be distinguished from the persisting anterograde memory impairment, which may be detected on clinical assessment or cognitive testing long after the period of PTA has ended. Moreover, forgetfulness is a common complaint within the context of a post-traumatic syndrome, which may include anxiety, irritability, poor concentration, and various somatic complaints. Commonly, these complaints persist long after the settlement of any compensation issues.[11]

Alcoholic blackouts

Alcoholic blackouts are discrete episodes of memory loss for significant events, which should not be confused with withdrawal seizures or other ictal phenomena. Alcoholic blackouts are associated with severe intoxication, usually in the context of a history of prolonged alcohol abuse. Goodwin *et al.*[12] described two types of blackout—the *fragmentary* and the *en bloc*. However, alcohol-induced state-dependent experiences can be viewed as related phenomena, and it has been suggested that the three represent gradations of alcohol-induced memory impairment. In state-dependent effects, subjects when sober cannot remember events or facts from an episode of intoxication, which they recall easily when they again become intoxicated. In fragmentary blackouts, the subjects are aware of their memory loss on being told later of an event; there are islands of preserved memory; and the amnesia tends to recover partially through time by shrinkage of the amnesic gap. In *en bloc* blackouts there is an abrupt beginning and end to the period of memory loss, and the lost memories are very seldom recovered. Blackouts may be more common in binge drinkers, because they are related to a high blood alcohol level. Hypoglycaemia may also be a contributory factor, and blackouts are more common where there is a history of previous head injuries.

After electroconvulsive therapy

This is an iatrogenic form of transient amnesia. Benzodiazepines and anticholinergic agents can also give rise to transient memory loss in more moderate form.[13]

Subjects tested within a few hours of electroconvulsive therapy (ECT) show a retrograde impairment for information from the preceding 1 to 3 years, a pronounced anterograde memory impairment on both recall and recognition memory tasks, and an accelerated rate of forgetting.[14] When retested approximately 6 to 9 months after completion of a course of ECT, memory generally returns to normal on objective tests. However, complaints of memory impairment can persist, and they may be evident three or more years after a course of ECT has been completed.[15] It seems that patients with persistent complaints of memory loss tend to be those who have recovered least well from their depression,[14,15] although their complaints tend to focus upon the period for which there was an initial retrograde and anterograde amnesia. A recent American study suggested that sine wave stimulation induces cognitive slowing in terms of reaction time, and that multiple bilateral ECT administrations can produce impairments in autobiographical memory retrieval 6 months following treatment.[16]

Verbal memory appears to be particularly sensitive to disruption. Unilateral electroconvulsive therapy to the non-dominant hemisphere produces considerably less memory impairment than bilateral ECT, although it is important to identify the non-dominant hemisphere by a valid procedure. Attempts to minimize memory disruption by either making changes in premedication or the concomitant administration of other substances—such as glycopyrrolate, physostigmine, thyroxine, dexamethasone, or acetylcholine—have produced limited or no benefit.

Post-traumatic stress disorder

This clinically important syndrome is described in Chapter 4.6.2. Post-traumatic stress disorder (PTSD) is characterized by vivid, intrusive thoughts and memories ('flashbacks'), avoidance and anxiety phenomena, and hyper-arousal and hyper-vigilance symptoms. However, there may be instances of brief memory loss, distortions, or even frank confabulations. For example, a victim of the *Herald of Free Enterprise* disaster at Zeebruge described trying to rescue a close friend still on board the ship, when other witnesses reported that the close friend had, in fact, not been seen by the victim from the moment the ship turned over. Cases of PTSD may, of course, be confounded by other factors, such as head injury. Nevertheless, it is of interest that PTSD symptoms can occur even when a subject is completely amnesic for an episode.[17] PTSD victims can show deficits in anterograde memory on formal tasks many years after the original trauma, and there is also evidence that they may show loss of hippocampal volume on magnetic resonance imaging (MRI) brain scan, which has been attributed by some to a surge in glucocorticoid secretion. Brewin[18] has recently reviewed four controversies in autobiographical memory for trauma. He found that qualitative and quantitative differences do exist between trauma and non-trauma memories in PTSD victims, and that memories for trauma can be either better or worse than non-trauma memories. In other words, some incidents may be recalled particularly vividly, and others may be forgotten.

Psychogenic fugue

A fugue state is a syndrome consisting of a sudden loss of all autobiographical memories and knowledge of personal identity, usually associated with a period of wandering, for which there is a subsequent amnesic gap on recovery. Characteristically, fugue states last a few hours or days, up to about 3 weeks. There are also

descriptions in the literature of persisting autobiographical memory loss, in which personal identity has been 're-learned', and these are better known as 'psychogenic focal retrograde amnesia'.[19] However, whenever such complaints persist, the suspicion of simulation must arise. Fugue states differ from transient global amnesia or transient epileptic amnesia in that the subject does not know who he or she is, and repetitive questioning is not a characteristic feature in fugues.

As discussed elsewhere,[20] fugue states are always preceded by a severe precipitating stress. Second, depressed mood is also an extremely common antecedent for a psychogenic fugue state, and may be associated with manifest suicidal ideas just before or following recovery from the fugue. Third, various authors have noted that there is often a past history of a previous transient neurological amnesia, such as epilepsy or head injury. In brief, it appears that patients who have experienced a previous transient organic amnesia, and who become depressed and/or suicidal, are particularly likely to go into a fugue in the face of a severe, precipitating stress. That stress may consist of marital or emotional discord, bereavement, financial problems, a criminal charge, or stress during wartime. Fugues have been described as a 'flight from suicide'. Recent neuro-imaging investigations have examined people purportedly in a fugue state with very inconsistent results, probably because the delay until imaging, the imaging techniques employed, and the clinical situations themselves have varied considerably across studies.

Amnesia for offences

This is a phenomenon commonly brought to the attention of psychiatrists, particularly forensic psychiatrists, although the empirical literature on this disorder is scanty. Amnesia is claimed by 25 to 45 per cent of offenders in cases of homicide, approximately 8 per cent of perpetrators of other violent crimes, and a small percentage of non-violent offenders.[21] It is necessary to exclude underlying neurological or endocrine factors such as an epileptic automatism, post-ictal confusional state, head injury, sleepwalking, or hypoglycaemia. Underlying medical disorder can be grounds for a so-called 'insane' automatism in English law (if the result of an internal brain disease) or a 'sane' automatism (if the consequence of an external agent), but otherwise amnesia per se does not constitute grounds for alleviation of responsibility for an offence.

Amnesia for an offence is most commonly associated with the following:

1 States of either extreme emotional arousal or peri-traumatic dissociation, in which the offence is unpremeditated, and the victim usually a lover, wife, or family member. This is most commonly seen in homicide cases ('crimes of passion').

2 Alcohol intoxication (sometimes in association with other substances), usually involving very high peak levels ('alcoholic blackout'), and often a long history of alcohol abuse. The victim is not necessarily related to the offender, and the offence may vary from criminal damage, through assault, to homicide.

3 Florid psychotic states or depressed mood. Occasionally offenders describe a delusional account of what has happened, quite at odds with what was seen by other observers, and sometimes resulting in confessions to crimes that the person could not actually have committed (a paramnesia or delusional memory). In many other cases, depressed mood is associated with amnesia for an offence, just as it is a common associate of psychogenic fugue.

Pyszora *et al.*[22] examined the psychiatric reports of all offenders given a life sentence in England and Wales in 1994, 29 per cent of whom claimed amnesia. Detailed, follow-up reports at 3 years were also examined, and these suggested that approximately one-third of those who had claimed amnesia at trial reported complete recovery, one-third showed partial recovery, and one-third reported no change in their amnesias. Only about 2 per cent were thought to have been malingering.

Persistent memory disorder

The amnesic syndrome can be defined as follows:

> An abnormal mental state in which memory and learning are affected out of all proportion to other cognitive functions in an otherwise alert and responsive patient.[23]

The Korsakoff syndrome can be defined in the same way but with the addition of the following phrase:

> . . . resulting from nutritional depletion, notably thiamine deficiency.

In fact, Victor *et al.*[23] used the first description as a definition of the Korsakoff syndrome, but it is important to distinguish between amnesic syndromes in general (for which the Victor *et al.* definition suffices) and the particular clinical condition described by Korsakoff,[24] whose cases can all be viewed (with hindsight) as having suffered nutritional depletion, whether of alcoholic or non-alcoholic causation. Various disorders can give rise to an amnesic syndrome.

The Korsakoff syndrome

As mentioned, this is the result of nutritional depletion, namely a thiamine deficiency. Korsakoff[24] described this condition as resulting from alcohol abuse or from a number of other causes, but by far the most common nowadays is alcohol abuse.

(a) Clinical

There are frequent misunderstandings about the nature of this disorder. 'Short-term memory', in the sense that psychologists employ it, is intact but learning over more prolonged periods is severely impaired, and there is usually a retrograde memory loss which characteristically extends back many years or decades.[20] Korsakoff himself noted that his patients 'reason about everything perfectly well, draw correct deductions from given premises, make witty remarks, play chess or a game of cards, in a word comport themselves as mentally sound persons'.[24] However, he also noted repetitive questioning, the extensive nature of the retrograde memory loss, and a particular problem in remembering the temporal sequence of events, associated with severe disorientation in time. As will be discussed below, he gave examples of confabulation reflecting the problem with the temporal sequence memory, such that real memories were jumbled up and retrieved inappropriately, out of temporal context.

Many cases of the Korsakoff syndrome are diagnosed following an acute Wernicke encephalopathy, involving confusion, ataxia, nystagmus, and opthalmoplegia. Usually, not all these features are present, and the opthalmoplegia in particular responds rapidly to treatment with high-dose vitamins. These features are often associated with a peripheral neuropathy. However, the disorder can also

have an insidious onset, and such cases are more likely to come to the attention of psychiatrists; in these cases, there may be either no known history of or only a transient history of Wernicke features. There are also reports that the characteristic Wernicke–Korsakoff neuropathology is found much more commonly at autopsy in alcoholics than the diagnosis is made in life, implying that many cases are being missed.

Victor *et al.*[23] reported that 25 per cent of patients with the Korsakoff syndrome 'recover', 50 per cent show improvement through time, and 25 per cent remain unchanged. Whilst it is unlikely that any established patient shows complete recovery, the present author's experience is that substantial improvement does occur over a matter of years if the patient remains abstinent. It is probably correct to say that 75 per cent of these patients show a variable degree of improvement, whilst 25 per cent show no change.[20]

(b) Pathology

The characteristic neuropathology in what is often known as the Wernicke–Korsakoff syndrome consists of neuronal loss, microhaemorrhages, and gliosis in the paraventricular and periaqueductal grey matter.[23] However, there has been a debate as to which particular lesions are critical for the manifestation of chronic memory disorder. Victor *et al.*[23] pointed out that all 24 of their cases in whom the medial dorsal nucleus of the thalamus was affected had a clinical history of persistent memory impairment (Korsakoff syndrome), whereas five cases in whom this nucleus was unaffected had a history of Wernicke features without any recorded clinical history of subsequent memory disorder. By contrast, the mammillary bodies were implicated in all the Wernicke cases, whether or not there was subsequent memory impairment. However, Mair *et al.*[25] provided a careful pathological and neuropsychological description of two patients with the Korsakoff syndrome, whose autopsies showed lesions in the mammillary bodies, the midline, and anterior portion of the thalamus, but not in the medial dorsal nuclei. Mayes *et al.*[26] obtained very similar findings in two further patients with the Korsakoff syndrome, who had also been very carefully described both neuropsychologically and at autopsy. Harding *et al.*[27] reported that pathology in the anterior principal thalamic nuclei was the critical difference between eight patients who suffered a persistent Korsakoff syndrome, and five others who experienced only a transient Wernicke episode. Taken together, these findings suggest that the mammillary bodies, the mammillothalamic tract, and the anterior thalamus may be more important to memory dysfunction than the medial dorsal nucleus of the thalamus.

There is also evidence of general cortical atrophy particularly involving the frontal lobes in patients with the Korsakoff syndrome, and this is associated with neuropsychological evidence of 'frontal' or 'executive' test dysfunction in these patients.[20]

There have been a number of neuro-imaging studies of the Korsakoff syndrome. CT scan studies indicated a general degree of cortical atrophy, particularly involving the frontal lobes.[28] MRI studies have indicated more specific atrophy in diencephalic structures.[29] PET investigations show variable findings, but hypometabolism has been reported in thalamic, orbito-medial frontal, and retrosplenial regions.[30]

Herpes encephalitis

This can give rise to a particularly severe form of amnesic syndrome.[31] Many cases are said to be primary infections, although others may involve a reactivation of the virus. Characteristically, there is a fairly abrupt onset of acute fever, headache, and nausea. There may be behavioural changes. Seizures can occur. The fully developed clinical picture with neck rigidity, vomiting, and motor and sensory deficits seldom occurs during the first week. Moreover, some cases commence more insidiously with behavioural change or psychiatric phenomena, the confusion and neurological features becoming evident only later. Diagnosis is by the PCR test or a raised titre of antibodies to the virus in the cerebrospinal fluid. A presumptive diagnosis is sometimes made on the basis of the clinical picture as well as severe signal alteration, haemorrhaging, and atrophy in the temporal lobes on MRI brain imaging.

Neuropathological and neuro-imaging studies usually show extensive bilateral temporal lobe damage,[29,32] although occasionally the changes are surprisingly unilateral. There may be frontal changes, often in the orbito-frontal regions, and there may be focal changes elsewhere as well as a variable degree of general cortical atrophy. The medial temporal lobe structures are usually particularly severely affected, including the hippocampi, amygdalae, entorhinal, and perirhinal cortices, and other parahippocampal structures. Encephalitis, like head injury, can also implicate basal forebrain structures which give cholinergic outputs to the hippocampi; this may further exacerbate the damage.

The chronic memory disorder in herpes encephalitis is often very severe,[31] but it shows many resemblances to that seen in the Korsakoff syndrome, consistent with the fact that there are many neural connections between the thalami, mammillary bodies, and the hippocampi. Patients with herpes appear to have better 'insight' into the nature of their disorder, and a 'flatter' temporal gradient to their retrograde memory loss (i.e. less sparing of early memories), and they may have a particularly severe deficit in spatial memory when the right hippocampus is involved.[20] However, the similarities in the episodic memory disorder tend to outweigh the differences.

On the other hand, a more extensive involvement of semantic memory is characteristic in herpes encephalitis, and this results from the widespread involvement of the lateral, inferior, and posterior regions of the temporal lobes. Semantic memory refers to a knowledge of facts, concepts, and language (see Chapter 2.5.3). Left temporal lobe pathology in herpes encephalitis commonly gives rise to an impairment in naming, reading (a so-called 'surface dyslexia'), and other aspects of lexico-semantic memory. Right temporal lobe damage may lead to a particularly severe impairment in face recognition memory or knowledge of people.

Severe hypoxia

Severe hypoxia can give rise to an amnesic syndrome following carbon monoxide poisoning, cardiac and respiratory arrests, or suicide attempts by hanging or poisoning with the exhaust gases from a car. Drug overdoses may precipitate prolonged unconsciousness and cerebral hypoxia, and this quite commonly occurs in heroin abusers. Zola–Morgan *et al.*[33] described a patient with repeated episodes of hypoxia and/or cardiovascular problems who developed a moderately severe anterograde amnesia. At autopsy 6 years later, this patient was shown to have a severe loss of pyramidal cells in the CA1 region of the hippocampi bilaterally, with the rest of the brain appearing relatively normal. Hippocampal atrophy on MRI has found in hypoxic, amnesic patients,[29]

and also thalamic hypometabolism on FDG–PET scanning.[34] In brief, the memory disorder is likely to result from a combination of hippocampal and thalamic changes, related to the many common neural pathways between these two structures. However, Caine and Watson[35] in an important review reported that less than 20 per cent of hypoxic patients described in the literature show either a specific amnesic syndrome (in the absence of other cognitive deficits) or damage solely confined to the hippocampi.

Vascular disorders

Two types of specific vascular lesions can particularly affect memory, as opposed to general cognitive functioning, namely thalamic infarction and subarachnoid haemorrhage. However, memory disorder may be the first manifestation of the vascular form of 'mild cognitive impairment'.

In an elegant CT scan study, von Cramon *et al.*[36] showed that damage to the anterior thalamus was critical in producing an amnesic syndrome. When the pathology was confined to the more posterior regions of the thalamus, memory function was relatively unaffected. The anterior region of the thalamus is variably supplied by the polar or paramedian arteries in different individuals, both of which are, ultimately, branches of the posterior cerebral artery that also supplies the posterior region of the hippocampi. When there is a relatively pure lesion of the anterior thalamus, anterograde amnesia without an extensive retrograde memory loss commonly results. However, cases in whom there is also retrograde memory loss, or even a generalized dementia, have been described following thalamic infarction, and this presumably relates to the extent to which thalamic projections are also implicated in the infarction.

Subarachnoid haemorrhage following rupture of a berry aneurysm can result in memory impairment, whether the anterior cerebral or posterior cerebral circulation from the Circle of Willis is involved. Most commonly described in the neuropsychological literature have been ruptured aneurysms from the anterior communicating arteries, because these affect ventro-medial frontal structures and the basal forebrain. Gade[37] has argued that it is whether or not the septal nuclei of the basal forebrain are implicated in the ischaemia which determines whether a persistent amnesic syndrome occurs in such patients. Others have attributed the florid confabulation, which these patients often exhibit, to concomitant orbito-frontal damage.[38]

Head injury

As discussed above, severe head injury can produce a persistent amnesia which may or may not be associated with generalized cognitive impairment. There may be direct trauma to the frontal and anterior temporal lobes, resulting in contusion and haemorrhaging, contrecoup damage, intracranial haemorrhage, and axonal tearing and gliosis following acceleration-deceleration or rotational forces. Memory function is commonly the last cognitive function to improve following an acute trauma, and patients can show the characteristic features of an amnesic syndrome. The phenomenon of 'isolated retrograde amnesia' has been described: in such cases, it seems likely that a mild head injury has precipitated a more purely psychiatric phenomenon. Traumatic head injury is considered in more detail in Chapter 4.1.10 and by Fleminger.[11]

Other causes of an amnesic syndrome

Deep midline cerebral tumours can give rise to an amnesic syndrome, and this may be exacerbated by surgical or irradiation treatment for pituitary tumours. Other infections, such as tuberculous meningitis or HIV, may, on occasion, give rise to an amnesic syndrome. Mild cognitive impairment and the very early stages of Alzheimer dementia may manifest themselves as a focal amnesic syndrome. Surgical treatment to the temporal lobes for epilepsy can result in profound amnesia, if there is bilateral involvement. There is increasing evidence that focal lesions in the frontal lobes can also produce severe memory impairment on aspects of anterograde and retrograde memory.[20] This can occur even in the absence of basal forebrain involvement, but it probably results from particular aspects of memory being implicated, including planning and organization, source and context monitoring, and particular aspects of retrieval processes.[39]

Neuropsychological aspects

The terms 'short-term' and 'long-term' memory should be abolished from psychiatric discourse, as they cause confusion across disciplines. It is more useful to consider current or recent memory versus remote (or autobiographical) memory. In addition, 'prospective memory' refers to remembering to do something.

Concepts of memory are considered in Chapter 2.5.3. As described in that chapter, a distinction is generally drawn between so-called 'working memory', which holds information for brief periods (a matter of several seconds) and allocates resources, and secondary memory, which holds different types of information on a permanent or semi-permanent basis. Secondary memory, in turn, can be subdivided into an episodic (or 'explicit') component, semantic memory, and implicit memory. Episodic memory refers to incidents or events from a person's past, such that he/she can 'travel back mentally in time'; this is characteristically severely affected in the amnesic syndrome. As mentioned previously, semantic memory refers to knowledge of facts, concepts, and language. The learning of new semantic memories is variably affected in the amnesic syndrome, although there is now some evidence that new facts can be learned even in the presence of severe, bilateral medial temporal lobe damage. Other aspects of semantic memory, including naming, reading, and comprehension, are affected in disorders where there is concomitant widespread temporal lobe pathology, such as herpes encephalitis, Alzheimer dementia, or semantic dementia (a form of frontotemporal dementia). Implicit memory refers to procedural or perceptuomotor skills, and to the facilitation of responses in the absence of explicit memory, known as 'priming'. Both these aspects are characteristically spared in the amnesic syndrome,[40] although the precise extent of sparing does depend on particular features in the experimental design.[41]

Over the years, there has been extensive debate concerning whether the primary deficit in the amnesic syndrome lies in the initial encoding of information, or some kind of physiological 'consolidation' into secondary memory, or accelerated forgetting of that information, or in retrieval processes.[20] There is still very little agreement about this debate, but, if anything, the consensus is that retrieval problems are secondary to initial acquisition and consolidation impairments, at least in anterograde amnesia. Retrieval deficits may be more important where there is an

extensive retrograde memory loss,[20] which might account for why there is generally a poor correlation between scores on anterograde memory measures and retrograde memory measures.

Much recent research has focused on the specific function of the hippocampi, and how there their role is distinct from other structures within the medial temporal lobes, more lateral temporal lobe regions, and the frontal lobes. Suggestions include a particular contribution to the binding of complex associations, relational memory, the binding of the distributed features of an episode into a coherent trace, novel or incremental learning, and a contribution to retrieval processes.[20,42] Aggleton and Brown[43] have suggested that the hippocampi are critical to the recall of contextual richness and detail, involved in 'remembering' or 'recollection', whereas the perirhinal cortex is particularly implicated in the familiarity judgements essential in recognition memory. The frontal lobes are generally thought to contribute to planning and organization in memory, aspects of context and source memory, awareness of memory performance (metamemory), prospective memory, and to particular aspects of retrieval processes.[20,39]

There are also many controversies concerning the nature of the extensive retrograde memory loss found in many of the above disorders. Modern neuropsychological studies have confirmed that this retrograde memory loss can extend back many years or decades, but that it often shows a 'temporal gradient' with relative sparing of early memories. The gradient is characteristically steeper in the amnesic syndrome than in dementing disorders such as Alzheimer dementia or Huntington's disease. Differing patterns of retrograde memory loss can occur; left temporal lobe damage seems to affect memory for facts and for the more linguistic components of remote memory, whereas right temporal lobe damage may affect memory for the incidents in a person's life.[20] One theory of retrograde amnesia and of the temporal gradient is that, as memories become 'consolidated' through time, they become independent of the medial temporal lobes and are relatively protected against brain injury to these structures. A second theory is that, through time, episodic memories adopt a less vivid, more 'semantic' form, and this protects earlier memories from the effects of brain injury. A third theory suggests that the hippocampi are always involved in the retrieval and reactivation of memories, and that every time a memory is retrieved, a new trace is laid down, resulting in 'multiple traces' protecting against the effects of brain injury.[44] These three theories make differing predictions and, at present, the underlying basis of retrograde amnesia remains hugely controversial.

Confabulation disorders

Confabulation can be subdivided into 'spontaneous' confabulation, in which there is a persistent, unprovoked outpouring of erroneous memories, and 'momentary' or 'provoked' confabulation, in which fleeting intrusion errors or distortions are seen in response to a challenge to memory, such as a memory test.[20,45]

Confabulation is widely believed to be particularly associated with the Korsakoff syndrome, but this is incorrect. Spontaneous confabulation arises in confusional states and in frontal lobe disease.[45] The link with frontal lobe pathology, particularly in the ventro-medial region, has been established in many investigations.[20,38] Spontaneous confabulation is often seen in the confusional state of a Wernicke encephalopathy, but it is rare in the more chronic phases of the Korsakoff syndrome. On the other hand, fleeting intrusion errors or distortions ('momentary confabulation') do occur in the chronic phase of a Korsakoff syndrome, when memory is challenged. However, such intrusion errors are also seen in healthy subjects when memory is 'weak' for any reason, such as a prolonged delay until recall.[45] They are also seen in Alzheimer dementia and other clinical amnesic syndromes, and they are certainly not specific to the Korsakoff syndrome.

There has been considerable interest of late in the nature of spontaneous confabulations. Confabulation can extend across episodic, personal semantic, and more general semantic memories.[46] A theory put forward by Korsakoff himself, as well as other authorities,[24] emphasizes problems in the temporal ordering of memories. In a particularly elegant study, Schnider *et al.*[47] found that a group of 'spontaneous confabulators' could be differentiated from other amnesic patients and controls on the basis of their errors on a temporal context memory task, but not on other memory or executive tests. More recently Schnider[48] has interpreted these findings in terms of a failure in 'reality monitoring'. Somewhat similarly, Johnson *et al.*[49] has argued that confabulation may reflect an interaction between a vivid imagination, an inability to retrieve autobiographical memories systematically, and source or context monitoring deficits. By contrast, Gilboa *et al.*[38] found that a failure to make fine-grained distinctions within memory could account for Schnider's observations. They argued that a failure in strategic retrieval and post-retrieval monitoring, related to ventromedial and orbito-frontal pathology, is critical for spontaneous confabulation to arise. Somewhat similar hypotheses have been put forward by Burgess and Shallice[50] It has also been argued that the content of confabulations may be heavily influenced by motivational factors.[51]

The notion of 'confabulation' or 'false memory' has now been extended to a variety of other disorders, including delusional memory, confabulation in schizophrenia, false confessions, apparently false memories for child sexual abuse, pseudologia fantastica, and dissociative identity disorder. Whilst each of these can potentially be accounted for in terms of a general model of memory and executive function, provided that the social context and some notion of 'self' is incorporated, there are likely to be differing mechanisms which give rise to these different types of false memory.[20]

Neurochemistry and neuropharmacology of memory disorders

The Korsakoff syndrome is relatively unusual among memory disorders in that there is a distinct neurochemical pathology with important implications for treatment. Since animal studies in the 1930s and 1940s, and the important observations of De Wardener and Lennox[52] and others in malnourished prisoners of war, it has been known that **thiamine depletion** is the mechanism which gives rise to the acute Wernicke episode, followed by a Korsakoff memory impairment. However, the genetic factor that predisposes a minority of heavy drinkers to develop this syndrome before they develop hepatic or gastrointestinal complications of alcohol abuse remains unclear. Transketolase is the enzyme which requires thiamine pyrophosphate (TPP) as a cofactor. Thiamine depletion affects six neurotransmitter systems (including acetylcholine, glutamate, aspartate, and GABA), either by reduction of TPP-dependant enzyme activity or by direct structural damage. Direct genomic PCR sequences of a high-affinity thiamine transporter gene

(SLC19A2) have identified three genetic variants in the Wernicke–Korsakoff syndrome.[53] Whatever the underlying genetic mechanism, treatment as soon as possible with high doses of parenterally administered multivitamins is essential in patients with the Wernicke–Korsakoff syndrome. The Wernicke features respond well to high-doses of vitamins, and such treatment can prevent the occurrence of a severe, chronic Korsakoff state.[20,23] The small risk of anaphylaxis is completely outweighed by the high risk of severe brain damage and the appreciable risk of litigation if such treatment is not administered.

There has been an extensive literature on the effects of **cholinergic antagonists** (such as scopolamine) upon memory. Kopelman and Corn[54] found a pattern of impairment in anterograde memory that closely resembled that seen in the amnesic syndrome. It has been argued that cholinergic blockade produces an effect upon the 'central executive' component of working memory, but Rusted[55] has concluded that this is not sufficient to account for the drug effect upon memory processes. Although some have argued that the predominant effect of scopolamine is on attention, it has been found that covarying for the sedation or psychomotor effects of the drug did not eliminate the strong drug effects on episodic memory tests.[13,54] The anticholinesterases, donepezil, rivastigmine, and galantamine are now widely used in the management of Alzheimer dementia.

Despite their very different pharmacological actions, the effects of the **benzodiazepines** upon memory and attention are remarkably similar to those of scopolamine. When recall or recognition is tested after a delay, benzodiazepines produce a marked anterograde impairment in explicit or episodic memory, similar to scopolamine.[13] As with scopolamine, however, once learning has been accomplished, the rate of forgetting is normal, and benzodiazepines do not produce any retrograde deficits.[13] Procedural learning tasks after both benzodiazepine and scopolamine administration show similar effects, with learning curves on the active drug generally paralleling those for placebo.[13,54] Benzodiazepine effects can be attenuated by coadministration of the benzodiazepine antagonist, flumazenil.

The effects of **catecholamines** upon memory have been studied for many years, but the general consensus is that they act upon 'tonic attentional processes' rather than directly upon the storage or retrieval of memories. In an elegant study, Cahill *et al.*[56] examined the effects of the β-adrenergic receptor antagonist propranolol on memory for an emotionally arousing story, compared with a carefully matched neutral story. As expected, subjects given a placebo recalled more of the emotional than the neutral story, when tested 1 week later. Subjects given propranolol recalled the neutral story as well as the placebo subjects, but were impaired on the emotional story, whereas stimulation of noradrenaline (with yohimbine) produced some enhancement of the emotional elements, and benzodiazepines impaired memory equally for both the neutral and emotional elements of the story.[13]

Some years ago, there was interest in the **serotonergic system** and alcohol-induced memory impairment. Early reports suggested that zimelidine, a serotonin reuptake inhibitor, reversed the memory impairment in healthy volunteers after the administration of ethanol. Later, it was claimed that fluvoxamine improved memory performance in five patients with the Korsakoff syndrome, and that the improvements correlated significantly with reductions in a cerebrospinal fluid breakdown product. The samples were small, and the benefits were minor. Nevertheless, 3,4-methylenedioxymethamphetamine (ecstasy) has been reported to produce memory impairments either by direct or indirect effects.[57] Some of the apparent cognitive effects of serotonergic agents may be the by-product of their effects on mood.[13]

Of forensic psychiatric importance are agents which produce transient but profound amnesia, and may be implicated in offences such as 'date rape'. These include flunitrazepam (Rohypnol) and gammahydroxybutyrate (GHB), and this topic has recently been reviewed by Curran.[58]

Conclusions

Systematic clinical descriptions of amnesic disorders and their underlying pathology have become more detailed and rigorous over the years. In particular, recent advances in neuro-imaging (structural, metabolic, and activation) have provided the opportunity to relate particular cognitive abnormalities to specific changes in brain function. The use of pharmacological agents, in parallel with such imaging techniques, may promote the development of pharmacological agents more potent than the meagre array that we have at present for the treatment of severe memory disorder.

Further information

Baddeley, A.D., Kopelman, M.D., and Wilson, B.A. (eds.) (2002). *The handbook of memory disorders* (2nd edn). John Wiley & Sons Ltd, Chichester.

Kopelman, M.D. (2002). Disorders of memory. *Brain*, **125**, 2152–90.

Squire, L.R. and Schacter, D.L. (eds.) (2002). *Neuropsychology of memory* (3rd edn). The Guilford Press, New York.

References

1. Hodges, J.R. and Ward, C.D. (1989). Observations during transient global amnesia: a behavioural and neuropsychological study of five cases. *Brain*, **112**, 595–620.
2. Quinette, P., Guillery-Girard, B., Dayan, V., *et al.* (2006). What does transient global amnesia really mean? Review of the literature and thorough study of 142 cases. *Brain*, **129**, 1640–58.
3. Kritchevsky, M., Squire, L.R., and Zouzounis, J.A. (1988). Transient global amnesia: characterization of anterograde and retrograde amnesia. *Neurology*, **38**, 213–9.
4. Nakada, T., Kwee, I.L., Fujii, Y., *et al.* (2005). High-field, T2 reversed MRI of the hippocampus in transient global amnesia. *Neurology*, **64**, 1170–4.
5. Sander, K. and Sander, D. (2005). New insights into transient global amnesia: recent imaging and clinical findings. *Lancet Neurology*, **4**, 437–44.
6. Kapur, N. (1993). Transient epileptic amnesia—a clinical update and a reformulation. *Journal of Neurology, Neurosurgery, and Psychiatry*, **56**, 1184–90.
7. Kopelman, M.D., Panayiotopoulos, C.P., and Lewis, P. (1994). Transient epileptic amnesia differentiated from psychogenic 'fugue': neuropsychological, EEG and PET findings. *Journal of Neurology, Neurosurgery, and Psychiatry*, **57**, 1002–4.
8. Manes, F., Graham, K.S., Zeman, A., *et al.* (2005). Autobiographical amnesia and accelerated forgetting in transient epileptic amnesia. *Journal of Neurology, Neurosurgery, and Psychiatry*, **76**, 1387–91.
9. Russell, W.R. and Nathan, P.W. (1946). Traumatic amnesia. *Brain*, **69**, 280–300.

10. Brooks, N. (1984). Cognitive deficits after head injury. In *Closed head injury: psychological, social and family consequences* (ed. N. Brooks), pp. 44–73. Oxford University Press, Oxford.
11. Fleminger, S. (in press). Head injury. In *Lishman's organic psychiatry* (4th edn) (eds. A. David, S. Fleminger, M.D. Kopelman, S. Lovestone, and J. Mellers). Blackwell, Oxford,.
12. Goodwin, D.W., Crane, J.B., and Guze, S.E. (1969). Phenomenological aspects of the alcoholic 'blackout'. *The British Journal of Psychiatry*, **115**, 1033–8.
13. Curran, H.V. and Weingartner, H. (2002). Psychopharmacology of human memory. In *The handbook of memory disorders* (2nd edn) (eds. A.D. Baddeley, M.D. Kopelman, and B.A. Wilson), pp. 123–41. John Wiley & Sons Ltd., Chichester.
14. Frith, C.D., Stevens, M., Johnstone, E.C., *et al.* (1983). Effects of ECT and depression on various aspects of memory. *The British Journal of Psychiatry*, **142**, 610–7.
15. Squire, L.R. and Slater, P.C. (1983). ECT and complaints of memory dysfunction: a prospective three-year follow–up study. *The British Journal of Psychiatry*, **142**, 1–8.
16. Sackeim, H.A., Prudic, J., Fuller, R., *et al.* (2007). The cognitive effects of electroconvulsive therapy in community settings. *Neuropharmacology*, **32**, 244–54.
17. Harvey, A.G., Brewin, C.R., Jones, C., *et al.* (2003). Coexistence of posttraumatic stress disorder and traumatic brain injury: towards a resolution of the paradox. *Journal of the International Neuropsychological Society*, **9**, 663–76.
18. Brewin, C.R. (2007). Autobiographical memory for trauma: update on four controversies. *Memory*, **15**, 227–48.
19. Kopelman, M.D. (2000). Focal retrograde amnesia and the attribution of causality: an exceptionally critical review. *Cognitive Neuropsychology*, **17**, 585–621.
20. Kopelman, M.D. (2002). Disorders of memory. *Brain*, **125**, 2152–90.
21. Taylor, P.J. and Kopelman, M.D. (1984). Amnesia for criminal offences. *Psychological Medicine*, **14**, 581–8.
22. Pyszora, N.M., Barker, A.F., and Kopelman, M.D. (2003). Amnesia for criminal offences: a study of life sentence prisoners. *Journal of Forensic Psychiatry and Psychology*, **14**, 475–90.
23. Victor, M., Adams, R.D., and Collins, G.H. (1971). *The Wernicke–Korsakoff syndrome*. F.A. Davis, Philadelphia, PA.
24. Korsakoff, S.S. (1889). Psychic disorder in conjunction with peripheral neuritis. Translated and republished by M. Victor and P.I. Yakovlev (1955). *Neurology*, **5**, 394–406.
25. Mair, W.G.P., Warrington, E.K., and Weiskrantz, L. (1979). Memory disorder in Korsakoff's psychosis; a neuropathological and neuropsychological investigation of two cases. *Brain*, **102**, 749–83.
26. Mayes, A.R., Meudell, P.R., Mann, D., *et al.* (1988). Location of lesions in Korsakoff's syndrome: neuropsychological and neuropathological data on two patients. *Cortex*, **24**, 367–88.
27. Harding, A., Halliday, G., Caine, D., *et al.* (2000). Degeneration of anterior thalamic nuclei differentiates alcoholics with amnesia. *Brain*, **123**, 141–54.
28. Jacobson, R.R. and Lishman, W.A. (1990). Cortical and diencephalic lesions in Korsakoff's syndrome: a clinical and CT scan study. *Psychological Medicine*, **20**, 63–75.
29. Colchester, A., Kingsley, D., Lasserson, D., *et al.* (2001). Structural MRI volumetric analysis in patients with organic amnesia, 1: methods and findings, comparative findings across diagnostic groups. *Journal of Neurology, Neurosurgery, and Psychiatry*, **71**, 13–22.
30. Reed, L.J., Lasserson, D., Marsden, P., *et al.* (2003). ^{18}FDG–PET findings in the Wernicke–Korsakoff syndrome. *Cortex*, **39**, 1027–45.
31. Hokkanen, L. and Launes, J. (2007). Neuropsychological sequelae of acute-onset sporadic viral encephalitis. *Neuropsychological Rehabilitation*. **17**, 429–49.
32. Reed, L.J., Lasserson, D., Marsden, P., *et al.* (2005). Correlations of regional cerebral metabolism with memory performance and executive function in patients with herpes encephalitis or frontal lobe lesions. *Neuropsychology*, **19**, 555–65.
33. Zola–Morgan, S., Squire, L.R., and Amaral, D.G. (1986). Human amnesia and the medial temporal region: enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus. *The Journal of Neuroscience*, **6**, 2950–67.
34. Reed, L.J., Lasserson, D., Marsden, P., *et al.* (1999). FDG-PET analysis and findings in amnesia resulting from hypoxia. *Memory*, **7**, 599–612.
35. Caine, D. and Watson, J.D. (2000). Neuropsychological and neuropathological sequelae of cerebral anoxia: a critical review. *Journal of International Neuropsychological Society*, **6**, 86–99.
36. von Cramon, D.Y., Hebel, N., and Schuri, U. (1985). A contribution to the anatomical basis of thalamic amnesia. *Brain*, **108**, 997–1008.
37. Gade, A. (1982). Amnesia after operations on aneurysms of the anterior communicating artery. *Surgical Neurology*, **18**, 46–9.
38. Gilboa, A., Alain, C., Stuss, D.T., *et al.* (2006). Mechanisms of spontaneous confabulations: a strategic retrieval account. *Brain*, **129**, 1399–1414.
39. Baldo, J.V. and Shimamura, A.P. (2002). Frontal lobes and memory. In *The handbook of memory disorders* (2nd edn) (eds. A.D. Baddeley, M.D. Kopelman, and B.A. Wilson), pp. 363–79. John Wiley & Sons Ltd., Chichester.
40. Schacter, D.L. (1987). Implicit memory: history and current status. *Journal of Experimental Psychology. Learning, Memory and Cognition*, **13**, 501–18.
41. Ostergaard, A.L. (1994). Dissociations between word priming effects in normal subjects and patients with memory disorders: multiple memory systems or retrieval? *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, **47**, 331–64.
42. Cohen, N.J., Poldrack, R.A., and Eichenbaum, H. (1998). Memory for items and memory for relations in the procedural/declarative memory framework. *Memory*, **6**, 6689–99.
43. Aggleton, J. P. and Brown, M.W. (1999). Episodic memory, amnesia and the hippocampal-anterior thalamic axis. *The Behavioural and Brain Sciences*, **22**, 425–44.
44. Moscovitch, M., Nadel, L., Winocur, G., *et al.* (2006). The cognitive neuroscience of remote episodic, semantic and spatial memory. *Current Opinion in Neurobiology*, **16**, 179–90
45. Kopelman, M.D. (1987). Two types of confabulation. *Journal of Neurology, Neurosurgery, and Psychiatry*, **50**, 1482–7.
46. Kopelman, M.D., Ng, N., and Van den Brouke, O. (1997). Confabulation extending across episodic memory, personal and general semantic memory. *Cognitive Neuropsychology*, **14**, 683–712.
47. Schnider, A., von Däniken, C., and Gutbrod, K. (1996). The mechanisms of spontaneous and provoked confabulations. *Brain*, **119**, 1365–1375.
48. Schnider, A. (2003). Spontaneous confabulation and the adaptation of thought to ongoing reality. *Nature Reviews. Neuroscience*, **4**, 662–71.
49. Johnson, M.K., O'Connor, M., and Cantor, J. (1997). Confabulation, memory deficits, and frontal dysfunction. *Brain and Cognition*, **34**, 189–206.
50. Burgess, P.W. and Shallice, T. (1996). Confabulation and the control of recollection. *Memory*, **4**, 359–411.
51. Fotopoulou, A., Solms, M., and Turnbull, O. (2004). Wishful reality distortions in confabulations: a case study. *Neuropsychologia*, **42**, 727–44.
52. De Wardener, H.E. and Lennox, B. (1947). Cerebral beriberi (Wernicke's encephalopathy): review of 52 cases in Singapore prisoner-of-war hospital. *Lancet*, **1**, 11–7.
53. Guerrini, I., Thomson, A.D., Cook, C., *et al.* (2005). Direct genomic PCR sequencing of the high affinity thiamine transporter (SLC19A2) gene identifies three genetic variants in Wernicke–Korsakoff syndrome (WKS). *American Journal of Medical Genetics*, **137B**, 17–19.

54. Kopelman, M.D. and Corn, T.H. (1988). Cholinergic 'blockade' as a model for cholinergic depletion: a comparison of the memory deficits with those of Alzheimer-type dementia and the alcoholic Korsakoff syndrome. *Brain*, **111**, 1079–110.
55. Rusted, J. (1994). Cholinergic blockade and human information processing: are we asking the right questions? *Journal of Psychopharmacology*, **8**, 54–9.
56. Cahill, L., Prins, B., Weber, M., *et al.* (1994). Beta–adrenergic activation and memory for emotional events. *Nature*, **3H**, 702–4.
57. Kopelman, M.D., Reed, L.J., Marsden, P., *et al.* (2001). Amnesic syndrome and severe ataxia following the recreational use of MDMA ('ecstasy') and other substances, *Neurocase*, **7**, 423–32.
58. Curran, H.V. (2006). Effects of drugs on witness memory. In *Witness testimony—psychological, investigative and evidential perspectives* (eds. A. Heaton-Armstrong, E. Shepherd, G. Gudjonsson, and D. Wolchover), pp 77–87. Oxford University Press, New York.

4.1.13 The management of dementia

John-Paul Taylor and Simon Fleminger

Introduction

The term dementia is used in two different ways. First there are the **dementias**. These are **diseases** that cause progressive and diffuse cerebral damage, of which Alzheimer's disease is the most common. Second, dementia can be used to refer to a **clinical syndrome**. Thus dementia is 'an acquired global impairment of intellect, memory, and personality, but without impairment of consciousness'.[1] For clinicians this is the preferred usage, and the one adopted in this chapter. It demands that the cause of the dementia is explored, and makes no comment on the likely prognosis.

This chapter will focus on the management of dementia regardless of the cause; however given the burden of dementia in older age, the discussion will be invariably, but not exclusively, slanted towards the management of dementia in this age group. Aspects of management specific to individual diseases which produce dementia will be avoided. In addition, a discourse on the management of cognitive and memory problems is excluded as these are described elsewhere (see Chapters 2.5.4 and 6.2.7). Patients who suffer the dementia before 18 years of age will, by and large, not be included; their needs are often best met by services provided for people with intellectual disability.

The newly diagnosed patient with dementia

Given that it is now possible to diagnose dementia early in the course of the disease it is important to consider when and how to disclose the diagnosis. This is often seen by clinicians as a difficult task and one to be avoided until the diagnosis is absolutely certain. Stigma is associated with the diagnosis of dementia; it is perceived as a chronic debilitating illness, with a progressive deterioration in mental faculties that ultimately leads to a loss of self-identity and an unpleasant death. The clinician may believe there is not much to offer until later in disease, and so there is not much point in disclosing the diagnosis at an early stage. Furthermore, they may find it difficult to break 'bad news', particularly when an individual with dementia may not understand or retain information.

Nevertheless, leaving these discussions until the diagnosis is certain may be too late; the patient's ability to take part in decisions about their future treatment, and their family's future, may by then be jeopardized by cognitive decline. Only early in the course of the illness will they be able to make a power of attorney, settle their will, and discuss with their doctors how they wish to be treated once the disease is well advanced.

The way in which the diagnosis is given will affect how patients and their families cope and deal with the diagnosis in long-term. Although there are no specific strategies for disclosure of a dementia diagnosis, techniques developed for breaking bad news in disclosure of cancer diagnosis are probably applicable; for example, the excellent protocol devised by Baile *et al.*[2]

Formal psychotherapy and counselling may help patients and their families come to terms with the diagnosis.[3] Clear simple pamphlets or information sheets should be available so that patients and their families can assimilate the diagnosis and its consequences outside of the interview. Referrals can also be made to dementia support groups and local dementia societies; these can provide psychoeducation, befriending services, and networking groups for patients and their families.

After initial meeting and disclosure, it is important that a follow-up meeting is arranged; this will allow patients and families to take on board the diagnosis and formulate any questions they might have. Detailed management strategies are probably best discussed during follow-up appointments, as patients and their families might be overwhelmed at the initial appointment.

Genetic counselling and testing for dementia

Many patients and their families are concerned about the heritability of the condition and will ask if any genetic tests can be performed. But such a request needs to be considered carefully.

In only about 5–10 per cent of cases is the dementia directly due to a high penetrance genetic mutation (for example, early-onset Alzheimer's, frontotemporal dementia, and dementia associated with Huntington's disease). Low penetrance gene variants, such as the apolipoprotein E (APOE) genotype, while modulating, for example, the risk of development of Alzheimer's dementia, do not adequately predict disease development. Consensus groups have therefore advised against using APOE predictive testing.[4–6]

There are also significant social ramifications of genetic testing for the relatives of patients with dementia; positive tests could have serious implications for employment, family planning, and insurance. Therefore access to appropriate pre-test counselling is important.

Box 4.1.13.1 shows a current modus operandi, based on United Kingdom guidelines produced by the National Institute of Clinical Excellence in dealing with this difficult subject.[4] However, the clinician is advised to keep abreast of current best practice given the likely rapid advances in this area.

The younger patient with dementia

The incidence of dementia under the age of 65 is rare. However dementia in younger people has significant additional consequences. Often the younger person with dementia has dependents and considerable financial commitments. Spouses may have to give up

Box 4.1.13.1 Guidelines for genetic testing for dementia

- People likely to have a genetic cause for their dementia (for example, familial autosomal dominant Alzheimer's disease or frontotemporal dementia, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy [CADASIL], or Huntington's disease) and their unaffected relatives should be offered referral for genetic counselling and testing.
- All patients referred for testing should have appropriate counselling in helping deal with psychological and social consequences.
- If a specific genetic cause for dementia is not suspected, as is the case in late-onset dementia, genotyping should not be undertaken for clinical purposes.

work to care for their partners and there are very high rates of caregiver burn-out.

Specialist service provision for the younger patient with dementia is often lacking. These patients have different life expectations than their elderly counterparts. Many will be physically fit and often do not fit easily into the service models provided for their elderly counterparts. Specialist multidisciplinary teams allied to traditional dementia services have been advocated[4] although actual implementation is still required.

Driving

Decisions about whether or not a person with dementia should be allowed to drive are often difficult. The patient's right to autonomy needs to be balanced against their social and legal responsibilities. The clinician has a duty to consider the safety of other people on the road, as well as the patient themselves. But there is no clear consensus on the best way of making the decision, although a number of regulatory authorities have issued guidance. As a rule of thumb, patients with moderate or severe dementia should not be driving; patients with mild dementia need a careful assessment.

Advice: Begin with a history from family and relatives; this may need to be done while the patient is not present. Have there been any accidents or near accidents? Do they feel the patient is unsafe and shouldn't be driving? A cognitive assessment (especially of executive and visuospatial function, and psychomotor speed) and physical examination of the patient is of some value although not definitive. The gold standard is a driving assessment on the road; a driving simulator test is an alternative.

Often, as in the United Kingdom, patients are legally obliged to inform their driving licensing authority about their diagnosis. The clinician should advise the patient and their relative of this, and document the discussion. Difficulties arise when a patient who is not fit to drive fails to inform the authority and continues to drive. A written warning to stop driving is often sufficient, particularly if the patient and their relatives are reminded that their car insurance policy is no longer valid. In some cases where the patient presents a real risk the clinician may need to break confidentiality and inform the authorities.

If the patient is deemed fit to continue driving, then they should be advised about risk reduction, for example keep to well-known routes and avoid busy roads, driving in bad weather conditions, or at night. They should be regularly reassessed with regard to their fitness to drive. Often this has to happen in any case because they will only be issued a short-term license (e.g. 1 year).

Behavioural and neuropsychiatric symptoms in dementia

Background

Behavioural and Psychological Symptoms in Dementia (BPSDs) have been defined by the International Psychogeriatric Association (1996) as 'signs and symptoms of disturbed perception, thought content, mood, or behaviour that frequently occur in patients with dementia'.[7]

Identification, assessment, and management of BPSDs are central to good dementia care. These heterogeneous symptoms are highly prevalent in dementia; one study[8] found that 61 per cent of 329 patients with dementia exhibited BPSDs, with the most common symptoms being apathy (27 per cent), depression (24 per cent), and agitation/aggression (24 per cent). The presence of BPSDs is cited by carers and relatives as being the most significant determinant in generating carer stress,[9] carer burden,[10] and increasing the likelihood of subsequent institutionalization.[11]

There appears to be only a weak correlation between the level of cognitive impairment and the occurrence and severity of BPSDs. Stronger associations have been noted between the presence of BPSDs and the degree of impairment in activities of daily living.[12]

A complex interplay of factors can give rise to these symptoms and include intrinsic host attributes and extrinsic environmental influences (Fig. 4.1.13.1). Therefore the same symptom in different individuals may be due to different causes. For example, aggression may be the response to a delusion in one individual, and the reaction to a change in caregiver in another. Often several different problem behaviours are seen in the same patient, such as wandering and sleep disturbance. There may be causal links between different BPSDs, for example the presence of distressing auditory hallucinations and persecutory delusions is strongly associated with consequent aggression.[13] Particular constellations of BPSDs are often associated with specific dementia syndromes (Table 4.1.13.1). BPSDs will change over time; for example aggression and psychosis tend to occur in the early to middle stages of Alzheimer's dementia whereas incontinence is invariably a feature of late disease.

Of all the symptoms that patients with dementia suffer, it is the problems caused by BPSDs that are most likely to trigger a pharmacological intervention or institutionalization. But whether or not a BPSD is reported as being a problem depends heavily on the informant and the situation. For example, night-time wandering may be tolerated by the spouse with the patient in their own home, but not by nursing staff in an acute medical ward.

Assessment

Assessment of a BPSD begins with a carefully taken informant history to assess the nature, history, and severity of the BPSD, and to garner the background medical, psychiatric, and social history. For example, there may be a history of phobic disorder, which is now manifest as agitation, or a lifelong tendency to aggression. Alcohol or other drug abuse must be addressed. The effect of recently prescribed, and recently stopped, medications needs to be

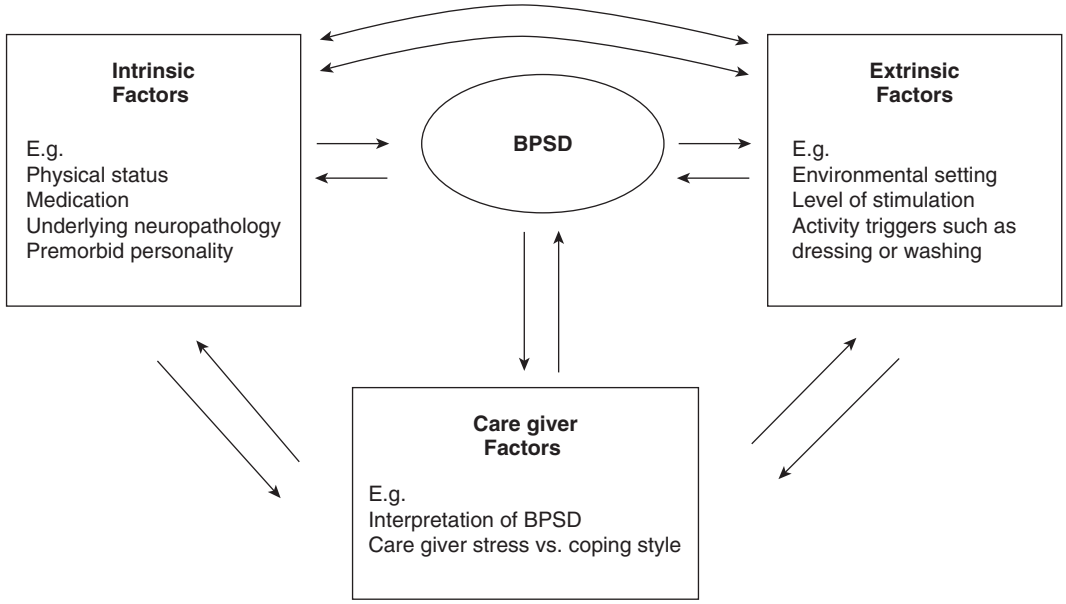


Fig. 4.1.13.1 Interaction between intrinsic host factors, extrinsic factors, and caregiver factors in the aetiology of BPSDs.

considered. The impact of the BPSD on the patient and the carer should be determined. Consider environmental influences; ask when the BPSD first occurred and whether it now occurs at any specific time, and whether it is related to any prior activity or antecedents. This will aid in formulation of specific behavioural management strategies (see below).

The mental state examination will look for evidence of anxiety, depression, or psychosis and persecutory delusions, and ascertain the patient's insight and understanding of their behavioural/neuropsychiatric symptoms. With specific problem behaviours direct observation of the behaviour can be very useful. A thorough physical examination will be needed to exclude physical illnesses; if suspected then appropriate medical investigations should be considered such as a midstream urine sample or chest X-ray. In agitated behaviours, sources of pain and fear should be considered, as well as the possibility of sleep loss or constipation, cold, or hunger. The presence of any sensory deprivation (e.g. hearing loss) should be looked for. Conversely the patient may be over-stimulated, as commonly occurs on general medical or surgical wards because of the noise and hustle and bustle.

Table 4.1.13.1 Common BPSDs in major dementia syndromes

Dementia	Common BPSDs
Alzheimer's dementia	Agitation, apathy, depression, anxiety, delusions
Vascular dementia	Depression, apathy
Dementia with Lewy bodies	Visual hallucinations, delusions, depression
Frontotemporal dementia	Disinhibition, repetitive behaviours, altered eating patterns, apathy

Assessment scales

A range of objective instruments for the measurement of BPSDs now exist.[14] Some measure multiple domains, for example the Behavioural Pathology in Alzheimer's Disease rating scale.[15] Other instruments are specific for one symptom, such as the Cohen–Mansfield Agitation Inventory[16] or Cornell scale for depression in dementia.[17] There are several caveats to the use of such scales, including the large intra-individual variations in scale scores which reflects the dynamic nature of BPSDs and variable reporting patterns of different observers. Some scales take a long time to complete or require training to administer. Nevertheless in clinical practice, the judicious use of such scales can allow for more reliable assessment of the response to a specific behavioural or pharmacological intervention.

Management of agitation and challenging behaviour

Agitation includes behaviour that is aggressive or abusive and occurs at an inappropriate frequency or is socially inappropriate.[18] Challenging behaviour has been used as a 'catch all' for a number of different behaviours including aggression, combativeness and disruptive vocalizations and may or may not be associated with agitation.

(a) Behavioural interventions

Evidence: Research into the effectiveness of behavioural techniques for agitation and challenging behaviour is based largely on studies using A–B–A–B or single-case design, and case series. Recent systematic reviews suggest that individually tailored behavioural interventions are effective.[19–22]

Advice: An intervention programme would start with a situational, or ABC, analysis:

- Antecedents—what was happening before the behaviour started?
- Behaviour—a clear description of the behaviour.

- Consequences—what happened as a result of the behaviour, particularly looking for possible reinforcers of the behaviour?

The frequency and severity of the behaviour then need to be charted as a baseline before introducing the specific intervention. Many programmes rely on the differential reinforcement of other behaviour (**DRO**); this involves positive reinforcement of other, appropriate behaviours, with the hope that these will then replace the challenging behaviour. A useful technique to be used alongside DRO is 'time out on the spot' (**TOOTS**), in which the unwanted behaviour is met with immediate withdrawal of social contact; appropriate behaviours receive warm social contact.

Unfortunately the limiting factor in use of behavioural interventions in dementia is the lack of trained individuals; often behavioural programmes can only be implemented in specialist units. In addition, the whole nursing/multidisciplinary team must be aware of the principles of reinforcement and extinction of behaviour, because behavioural programmes are unlikely to be effective unless consistently applied across the team. Evidence for the effectiveness of educating carers/family members in behaviour management techniques is currently inconclusive.[22] An alternative view which is gaining prominence is that while some challenging behaviours may not be amenable to interventions, it is possible to change the caregiver perception of the problem; this can lead to a reduction in caregiver distress and often by improving an aberrant interaction between caregiver and patient, there is a reduction in the challenging behaviour itself.

(b) Psychosocial and environmental interventions

Evidence: Despite numerous studies there is still a paucity of high quality evidence for the effectiveness of psychosocial and environmental interventions.[19,20,22] Some interventions have shown some promise helping ameliorate aggressive or disruptive behaviours in dementia. A non-exhaustive selection of the major psychosocial interventions that have been used is shown in Table 4.1.13.2.

Advice: The interventions should be tailored to the individual person taking account of their level of function and response to the approach. Psychosocial interventions are best applied when there is no clear cause for the disruptive behaviour; they should be only considered when a thorough assessment of the behaviour has be carried out to exclude treatable causes, for example, pain or psychosis. Even though certain psychosocial approaches may only have modest efficacy in decreasing disruptive behaviour, they may still be useful care adjuncts in improving general patient well-being.

(c) Antipsychotics

Evidence: Short-term treatment with atypical antipsychotics is of benefit in treating aggression, agitation, and psychosis, although the effect size is modest.[4,23,24] With regard to typical antipsychotics, meta-analyses have suggested that haloperidol might improve symptoms of aggression.[25,26] There is no evidence for benefit of longer term treatment (i.e. greater than 3 months).

Adverse effects of antipsychotics may be troublesome, particularly in the elderly; indeed outcomes from the Clinical Antipsychotic Trials of Intervention Effectiveness for Alzheimer's disease suggested that the adverse effects of atypical antipsychotics offset treatment

Table 4.1.13.2 Psychosocial interventions in dementia

Intervention	Description	Evidence for effectiveness
Psychoeducation to staff	Educating staff about dementia, neuropsychiatric symptoms, and reduced use of restraint	Possibly effective—might have sustained benefits
Reminiscence therapy	Uses materials related to patient and their era for example, old photographs and news articles, to stimulate memories and allow sharing of experiences	Possibly effective
Cognitive stimulation	Similar to reality orientation therapy, based on information processing rather than orientation knowledge	Possibly effective—might have sustained benefits
Music therapy	Can consistent of playing music as part of activity sessions or at specific times of the day. Music often of patient's era	Possibly effective—but no evidence of prolonged benefit
Snoezelen/multi-sensory therapy	Combined relaxation and use of sensory stimuli e.g. sounds, lights, touch	Possibly effective—but benefit wears off quickly. Also time/staff intensive
Aromatherapy	Mostly using lemon balm oil and lavender oil either inhaled or applied by massage	Possibly effective for agitation and restlessness
Bright light therapy	Sustained exposure to high levels of light (up to 10 000 lux)	Possibly effective (may be more benefit in sleep disturbance than behavioral disturbance). Time/staff intensive
Pet therapy	Contact with animals	Inconclusive
Exercise	Walking or light exercise sessions	Inconclusive
Simulated presence	Audiotape recorded by caregiver/family member played to patient where positive autobiographical memories are reiterated	Not effective
Reality orientation	Regular provision of orientating information e.g. time, date, etc.	Not effective
Validation therapy	Rogerian-based therapy; allowing resolution of unfinished conflicts, the acceptance of the reality and the expression of feelings	Not effective

advantages in patients with Alzheimer's disease.[24] Antipsychotic use is also an independent risk factor for falls in people with dementia.[27] Confusion may deteriorate, particularly with drugs with anticholinergic effects, and sedation may be problematic. In addition, the risk of emergent extra-pyramidal symptoms appears to be significantly increased with the use of haloperidol and risperidone;[28] these side-effects manifest even at low doses (for example, 1–2 mg of risperidone). Neuroleptic sensitivity is particularly evident in dementia with Lewy bodies; severe neuroleptic reactions can occur in up to 50 per cent of these patients.[29]

There is some evidence for an increased mortality risk for olanzapine and risperidone. There may be an increased risk of cerebrovascular adverse events in people with dementia taking these medications.[30] It is unclear whether this is drug specific or a class effect. Meta-analyses have indicated that there is a 1.5- to 1.7-fold increase in mortality risk for people with Alzheimer's disease treated with atypical neuroleptics. In 2005, the FDA asked the manufacturers of olanzapine, risperidone, aripiprazole, quetiapine, clozapine, and ziprasidone to include warning labels indicating increased risk of death on their products.[31] In the United Kingdom, the Committee for the Safety of Medicines (2004) advised that risperidone and olanzapine should not be used for the treatment of behavioural symptoms of dementia.[30] More recently, Wang *et al.* (2005), reported increased mortality rates in patients over the age of 65 treated with typical neuroleptics compared with atypicals; the risk was greatest shortly after initiation of the treatment and when higher doses were used.[32] Neuroleptic use has also been suggested to hasten cognitive decline although a more recent study refutes this.[33]

Advice: A high level of caution needs to applied to the use of antipsychotics in dementia. A careful weighing of benefits versus the risks is required. General principles for the use of antipsychotics (and other psychotropics) in dementia are similar to psychotropic prescribing for people with head injury (Box 4.1.10.3 in Chapter 4.1.10). Additionally, in dementia:

- Use drugs only if psychosocial or behavioural strategies have failed, and only if absolutely necessary.
- Review prescriptions regularly looking for side-effects. Patients with dementia often have physical co-morbidities and as a consequence take multiple medications; do not allow cocktails to build up.
- Do not prescribe for prolonged periods. Regularly reassess the need for the drug; can alternative interventions be applied now the situation is containable?
- For mild to moderate behavioural disturbances, consider the use of cholinesterase inhibitors or memantine (see below).

(d) Benzodiazepines

Benzodiazepines can be effective in agitation, particularly if it is associated with anxiety, sleep disturbance, or restlessness. However side effects are frequently associated with benzodiazepine use in people with dementia including sedation, worsening of cognitive function, paradoxical increased agitation, and increased risk of falls. The use of benzodiazepines should therefore be judicious and on a needs only basis. Short-acting benzodiazepines, for example oxazepam or lorazepam, are recommended by some, particularly in the elderly, because they are less likely to result in steadily accumulating blood levels.

(e) Mood stabilizers

There is limited evidence that carbamazepine may be beneficial in the treatment of agitation, although there are concerns about its safety in elderly patients given its propensity to induce haematological abnormalities.[23] High dose valproate does appear to reduce agitation, but, again, there is a significant risk of serious side effects.[34]

(f) Cholinesterase inhibitors and memantine

Evidence: In addition to providing benefits for cognition in dementia (see Chapter 6.2.7), cholinesterase inhibitors may help reduce BPSDs in people with dementia (see Ballard and Howard, 2006, for discussion).[28] However in most studies behaviour improvements have been secondary outcomes and a recent multi-centre 12-week trial of donepezil in patients with Alzheimer's disease found that donepezil was no more effective than placebo in treating agitation.[35] One group who do appear to gain clear benefit from cholinesterase inhibitor use is people with Lewy body dementia. Rivastigmine at a dose of up to 12 mg/day for 20 weeks appeared to significantly reduce psychotic symptoms in this group.[36] There is conflicting evidence for the use of memantine for the treatment of behavioural symptoms.

Advice: In terms of prescribing, it is probably worth considering the use of cholinesterase inhibitors and memantine in BPSDs, given their low propensity for serious adverse side effects.

(g) Other agents

Trazodone, a sedative medication, in preliminary findings appeared to be helpful in the treatment of behavioural disturbances. However more recent randomized control trials (RCTs) have refuted the benefits of this medication in dementia.[28]

Citalopram, aside from its antidepressant properties may have some beneficial effects on irritability and restlessness.[37] Propranolol has been considered; however, most trials for its effectiveness are from old, open label trials. If used, blood pressure and the ECG need to be closely monitored.

(h) Wandering

Many people with dementia will wander, and others will abscond or demand to leave. A risk assessment may be needed to determine their safety outside, for example assessing road safety and their ability to find their way back home.

Evidence: The use of two-dimensional grid patterns by the door of the ward, environmental sign-posting, or concealing the exit by use of a mirror may possibly reduce inappropriate exiting behaviour; however the evidence for these strategies is relatively weak.[22] There is probably better evidence for behavioural interventions (see above).[21]

Advice: Number entry locks which the patient with dementia cannot use can be helpful although it may frustrate the patient. The use of identification bracelets or tagging systems which sound an alarm if the patient leaves the unit is controversial and the subject of ethical debate. An inpatient or residential unit will need both an 'absent without leave' policy, which will include the protocol for informing family and police, and a locked door policy which must take into account what happens if there is a fire. Detention under a mental health or mental capacity act may need to be considered.

Mood disturbance

(a) Depression and apathy

Depression in people with dementia is quite common. Therefore it is important to consider depression as a cause for almost any change in function or behaviour, and to look for risk factors for depression, for example a recent bereavement, in the history. A screening test to detect depression may be appropriate (see above). Apathy, another common symptom in dementia, may be both a symptom of depression and a consequence of organic brain disease affecting those brain systems involved in motivation.

If the patient is depressed then review the general medical state, including any drugs that may produce depression. Make sure that all general psychosocial issues have been addressed, for example appropriate support services, leisure activities, and housing. Specific psychological therapy, for instance cognitive therapy, for depression in people with dementia is generally unavailable. However there is some limited evidence that a cognitive behavioural approach may help in treating depressive symptoms in people with dementia and be of benefit to carers.[38]

(i) Evidence for pharmacological treatment

There is some suggestion that antidepressant treatment of depression in patients with dementia is effective, although the evidence is limited.[39] The treatment of apathy in the absence of depression is less clear. The is some evidence that cholinesterase inhibitors improve apathy[40] and case series suggest that bromocriptine and methylphenidate are effective, though clinical experience indicates that the effects may be short-lived.

(ii) Advice on pharmacological treatment

The choice of which antidepressant drug to use will depend heavily on their side-effect profile. Newer antidepressants such as the serotonin reuptake inhibitors, having less anticholinergic activity and less cardiotoxicity, are generally preferred. Some method to evaluate the effectiveness of the treatment needs to be in place, preferably before treatment is started to get a baseline measure. For example, a measure of activities of daily living may be the target outcome to see whether it improves with antidepressant treatment.

Mania

Mania is rare in dementia, though there is possibly a specific association with Huntington's chorea. There is no evidence to suggest that mania treatment in a person with dementia be any different from normal protocols.

Psychotic symptoms

The phenomenology of psychotic symptoms influences treatment choice. For example, the occurrence of auditory hallucinations with secondary persecutory delusions may be more responsive to antipsychotic therapy whereas delusions of theft, founded on memory impairment, may respond better to psychosocial interventions such as strategies to help the person keep tags on where they put things.

Evidence: There are few hard data on which to base decisions about pharmacological treatment for relieving psychotic symptoms in dementia although there is reasonable evidence that cholinesterase inhibitors are successful in treating visual hallucinations in dementia with Lewy bodies.[36]

Advice: The choice of which antipsychotic to use is likely to be determined by its profile of side-effects. The same cautions and advice given for antipsychotic use in agitation in dementia (above) need to be applied for their use in treating psychotic symptoms.

Disorders of sexual behaviour

(a) Impotence or reduced libido

Reduced sexual activity and interest is the most common disorder of sexual behaviour associated with dementia, though it is the least likely to come to the attention of the clinician. It probably plays a part in the high rates of divorce seen, for example, in young couples after one partner has sustained a brain injury. Psychological effects, in particular the change in the patient's role in the partnership as a result of dementia, as well as the physiological effects of brain dysfunction on erectile function, contribute to impotence and reduced libido.

The first and most important step is to recognize the problem and talk about it. The couple may wish to be referred to a sexual disorders clinic. If reduced libido is part of a more generalized apathy or depression then it may respond when these features are appropriately treated (see above). Erectile dysfunction may respond to oral phosphodiesterase inhibitors such sildenafil.

(b) Sexual disinhibition and overactivity

Any display of sexual disinhibition, although uncommon in dementia, is likely to become a major management issue and needs a thorough behaviour assessment. Occasionally it may be part of a Klüver–Bucy-like syndrome with hyperorality and excessive eating.

Sexual disinhibition may respond to behavioural/psychosocial strategies. It may, for example, be necessary to ensure that only men nurse the patient if all the sexual disinhibition is directed towards female staff. A full behavioural programme to try to extinguish the behaviour may be effective, but if the behaviour involves touching and groping then it is essential to discuss and monitor the programme with those involved in the hands-on care of the patient. Staff often find such behaviour particularly upsetting.

Antipsychotics may reduce sexually disinhibited behaviour. The antiandrogens, cyproterone acetate, and medroxyprogesterone (Depo-Provera), may need to be tried if all else fails.

Sleep disturbance

Patients with dementia often have a disturbed sleep pattern and this is most troublesome when the sleep–wake cycle is inverted, with the patient asleep during the day but awake at night.

(a) Assessment

It is worth considering restless legs or rapid eye movement (REM-sleep) behaviour disorder as a cause of sleep disturbance. REM-sleep behaviour disorder, often a feature of dementia with Lewy bodies can be successfully treated with low dose clonazepam. Does the patient have to get up at night to empty his or her bladder because of prostatism or bladder dysfunction? Are there other medical reasons why the patient may be waking at night, for example because of pain from a duodenal ulcer? Sleep apnoea, more common in the elderly, produces sleep disturbance and is a contraindication for benzodiazepines and other drugs that may suppress respiration. Is the sleep disturbance due to depression? Has there

been any recent change to the sleeping arrangements? If so any sleep disturbance may be self-limiting.

(b) Management

Hypnotics are likely to have deleterious effects on cognition and functional abilities, and increase the risk of falls; these drugs should only be considered after techniques to improve sleep hygiene have been tried.

If sleep hygiene techniques fail, there is little definitive evidence to guide the clinician as to which hypnotic to select in patients with dementia. Benzodiazepines should, if possible, not be given indefinitely; particular caution is needed if the patient already shows disinhibition. Trazodone has been tried, particularly if there is co-morbid depression, although the evidence for its efficacy is questionable (see above). It has been suggested that bright light therapy (Table 4.1.13.2) can resynchronize aberrant circadian rhythms, but there is no definite evidence that it is effective.

Incontinence

Dementia in the elderly roughly doubles the risk of urinary incontinence. To minimize incontinence, toilets should be easily identifiable and readily accessible. Clothing may need attention to ensure that it is easy to remove. For urinary incontinence reversible causes, such as urinary tract infection, constipation, and medication (such as diuretics or drugs with anticholinergic side-effects causing urinary retention and overflow) should be excluded.

A diary recording frequency of voiding on the toilet and frequency of incontinence should be kept to see if toileting times can be adjusted to minimize incontinence. Prompted voiding (asking the person hourly if they want to go to the toilet and giving praise for successful toileting) is effective for some individuals. A behavioural programme may be needed for the patient who urinates or defecates in inappropriate places.

If incontinence persists get the advice of a continence advisor before considering drug treatment.

Risk management in dementia

Risk assessment is an important part of the management of patients with dementia and Table 4.1.13.3 suggests various areas of risk that need to be considered. A good history from carers and others involved in the patient's care is essential for a full risk assessment, which is rarely complete without an assessment by an occupational therapist.

It is important to use the outcome of risk assessment to facilitate independence. This is done by introducing appropriate strategies to minimize risk. In addition, a risk–benefit analysis may demonstrate that it is appropriate to run a risk of some adverse event happening if there are clear benefits of doing so. For example, a patient may be at risk of wandering and getting lost from his or her home; however, if the strategy to prevent this involves moving the patient to new accommodation away from family and familiar surroundings, then this may itself be regarded as a sufficiently adverse event to make transfer inappropriate. But before implementing such a strategy discuss it with other clinicians involved in the case, and with the carers, family and, if possible, the patient. Document the outcome of these discussions as well as the rationale for the management plan.

Table 4.1.13.3 Risk assessment in dementia

Consider the following areas:	
Anti-social and other behaviour	
◆ *Violence/aggression*	—towards others —towards self
◆ *Sexual disinhibition/assault*	
◆ *Other antisocial behaviours*	—which may provoke assaults e.g. argumentative, spitting, etc.
◆ *Wandering/agitation*	
Safety associated with impaired memory and cognition and poor judgement	
◆ *Home safety*	—leaving kettles, fires, etc., on. —leaving doors/windows, etc. open —cigarettes
◆ *Out and about*	—getting lost —road safety —pedestrian —driving
◆ *Financial*	—not able to handle money, loses money —inapproriately spends money, poor judgement
◆ *Work*	—fails to monitor and check for errors —unsafe with dangerous machinery
◆ *Supervising others*	—especially children (and consider aggression/sexual behaviour)
Vulnerability to	
◆ *Abuse by others*	—physical, sexual, emotional, and financial
◆ *Self neglect*	—including not eating, squalor
Physical health	
◆ *Falls*	
◆ *Managing their illness*	—e.g. diabetes, diet —taking their medication—note risk of abruptly stopping anticonvulsants, or steroids —drug dependence
◆ *Epilepsy*	

Caregivers

The impact and burden of dementia on family and caregivers is profound. In addition to the burden of caring, caregivers may experience adverse financial consequences, loss of independence, and social isolation. As a result, caregivers often exhibit high levels of psychological and physical morbidity.

The role of caregivers needs be acknowledged in maintaining individuals with dementia in the community. Significant caregiver burden and stress drastically increases the likelihood of care home admission. Therefore support of the caregiver is intrinsic to good dementia care.

Evidence: There have been a large number of studies examining the effectiveness of interventions to support the caregiver. Even though these various studies rarely used the same outcome measures a recent meta-analysis was able to show that caregiver interventions can be beneficial in reducing caregiver psychological morbidity and,

importantly, might also delay nursing home admission.[41] Intensive interventions that focus on psychoeducation, stress management, and include the person with dementia, seem particularly effective.

Advice: Support for the family and carers should consist of several components.[4] Start with education about the cause of the dementia, the possible prognosis, and the symptoms—both current and those that may develop. Family and friends need to understand that cognitive and behavioural symptoms arise from damage to the brain and are part of the illness. Caregivers need advice on the principles of care and skills training, for example ensuring that communication is simple and direct, avoiding changes to routine, not arguing with the patient, but on the other hand not endorsing false beliefs. They will need guidance on when and how to call on professional advice. Caregivers may also need help in obtaining social services input, additional care at home (including cleaning, nursing, and meals-on-wheels), as well as legal and financial advice. Voluntary organizations such as Alzheimer's Society UK (http://www.alzheimers.org.uk) and local self-help groups are often excellent sources of information and support for caregivers.

The burden of caring for someone with a dementia may result in depression and other signs of stress. The carer should have the opportunity of talking about any problems they have, if necessary getting their own psychiatric care.

Caregivers who are under stress are probably more likely to abuse, either physically or emotionally, the person with dementia. Try to ensure that any physical and emotional abuse of the demented person is picked up early. It helps if everybody involved in the person's care knows how to report any concerns they may have about what is happening.

Management of end-stage dementia

Previously, little attention was given to the end stages of the neurodegenerative dementias; most patients would die in hospital or long-term care facilities. However palliative care strategies are increasingly being used in dementia care. These emphasize physical, psychological, social, and spiritual aspects of care, with non-curative interventions aimed at maximizing quality of life. An important principle of treatment in palliative care is proportionality; any treatment should only be implemented if the balance of clinical benefit outweighs the burden such a treatment imposes.

For patients with dementia good palliative care includes management of BPSDs. In addition there are a number of end of life issues relevant to dementia:

Swallowing difficulties and aspiration pneumonia: These are common in end-stage dementia. Ethical consensus has indicated that the use of nasogastric and percutaneous endoscopic gastrostomy (PEG) tubes is seldom warranted in end-stage dementia although there may be individual cases in which the use of feeding tubes is not futile.[42] The use of antibiotics is more controversial. Certainly in other branches of palliative care there is evidence that even in the terminal stages of illness, antibiotics can relieve the distress caused by infected bronchial secretions.

Pain: Patients with dementia are often unable to communicate their distress. Be alert to the possibilities of pain; indeed always consider if a behavioural symptom is manifestation of pain. The management of pain in dementia is similar to pain management in other conditions. The aetiology of the pain should direct the choice of treatment. Adequate doses of analgesics to achieve good pain relief should be prescribed.

Further information

National Institute of Clinical Excellence. (2006). Dementia: supporting people with dementia and their carers in health and social care. National Clinical Practice Guideline Number 42. London. Available online: http://guidance.nice.org.uk/cg42.

Burns, A.B., O'Brien, J.T., and Ames, D. (2005). *Dementia* (3rd edn). Hodder Arnold, London.

Lovestone, S. and Gauthier, S. (2001). *Management of dementia* (1st edn). Martin Dunitz, London.

References

1. Lishman, W.A. (1998). *Organic psychiatry: the psychological consequences of cerebral disorder* (3rd edn). Blackwell Science, Oxford.
2. Baile, W.F., Buckman, R., Lenzi, R., *et al.* (2000). SPIKES-a six-step protocol for delivering bad news: application to the patient with cancer. *The Oncologist*, **5**, 302–11.
3. Cheston, R., Jones, K., and Gilliard, J. (2003). Group psychotherapy and people with dementia. *Aging & Mental Health*, **7**, 452–61.
4. National Institute of Clinical Excellence. (2006). Dementia: supporting people with dementia and their carers in health and social care. National Clinical Practice Guideline Number 42. London.
5. Brodaty, H., Conneally, M., Gauthier, S., *et al.* (1995). Consensus statement on predictive testing for Alzheimer disease. *Alzheimer Disease and Associated Disorders*, **9**, 182–7.
6. Post, S.G., Whitehouse, P.J., Binstock, R.H., *et al.* (1997). The clinical introduction of genetic testing for Alzheimer disease. An ethical perspective. *Journal of the American Medical Association*, **277**, 832–6.
7. Finkel, S.I., Costa, E., Silva, J., *et al.* (1996). Behavioral and psychological signs and symptoms of dementia: a consensus statement on current knowledge and implications for research and treatment. *International Psychogeriatrics*, **8** (**Suppl. 3**), 497–500.
8. Lyketsos, C.G., Steinberg, M., Tschanz, J.T., *et al.* (2000). Mental and behavioral disturbances in dementia: findings from the cache county study on memory in aging. *The American Journal of Psychiatry*, **157**, 708–14.
9. Victoroff, J., Mack, W.J., and Nielson, K.A. (1998). Psychiatric complications of dementia: impact on caregivers. *Dementia and Geriatric Cognitive Disorders*, **9**, 50–5.
10. Robert, F. Coen. (1997). Behaviour disturbance and other predictors of care burden in Alzheimer's disease. *International Journal of Geriatric Psychiatry*, **12**, 331–6.
11. Cohen, C.A., Gold, D.P., Shulman, K.I., *et al.* (1993). Factors determining the decision to institutionalize dementing individuals: a prospective study. *The Gerontologist*, **33**, 714–20.
12. Tekin, S., Fairbanks, L.A., O'Connor, S., *et al.* (2001). Activities of daily living in Alzheimer's disease: neuropsychiatric, cognitive, and medical illness influences. *The American Journal of Geriatric Psychiatry*, **9**, 81–6.
13. Aarsland, D., Cummings, J.L., Yenner, G., *et al.* (1996). Relationship of aggressive behavior to other neuropsychiatric symptoms in patients with Alzheimer's disease. *The American Journal of Psychiatry*, **153**, 243–7.
14. Burns, A., Lawlor, B., and Craig, S. (2003). *Assessment scales in old age psychiatry* (2nd edn). Martin Dunitz Ltd., London.
15. Reisberg, B., Borenstein, J., Salob, S.P., *et al.* (1987). Behavioral symptoms in Alzheimer's disease: phenomenology and treatment. *Journal of Clinical Psychiatry*, **48** (Suppl.), 9–15.

16. Cohen-Mansfield, J., Billig. (1986). Agitated behaviours in the elderly. A conceptual review. *Journal of the American Geriatric Society*, **34**, 711–21.
17. Alexopoulos, G.S., Abrams, R.C., Young, R.C., *et al.* (1988). Cornell scale for depression in dementia. *Biological Psychiatry*, **23**, 271–84.
18. Cohen-Mansfield, J., Werner, P., Watson, V., *et al.* (1995). Agitation among elderly persons at adult day-care centers: the experiences of relatives and staff members. *International Psychogeriatrics*, **7**, 447–58.
19. Ayalon, L., Gum, A.M., Feliciano, L., *et al.* (2006). Effectiveness of nonpharmacological interventions for the management of neuropsychiatric symptoms in patients with dementia: a systematic review. *Archives of Internal Medicine*, **166**, 2182–8.
20. Doody, R.S., Stevens, J.C., Beck, C., *et al.* (2001). Practice parameter: management of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*, **56**, 1154–66.
21. Spira, A.P. and Edelstein, B.A. (2006). Behavioral interventions for agitation in older adults with dementia: an evaluative review. *International Psychogeriatrics*, **18**, 195–225.
22. Livingston, G., Johnston, K., Katona, C., *et al.* (2005). Systematic review of psychological approaches to the management of neuropsychiatric symptoms of dementia. *The American Journal of Psychiatry*, **162**, 1996–2021.
23. Sink, K.M., Holden, K.F., and Yaffe, K. (2005). Pharmacological treatment of neuropsychiatric symptoms of dementia: d review of the evidence. *Journal of the American Medical Association*, **293**, 596–608.
24. Schneider, L.S., Tariot, P.N., Dagerman, K.S., *et al.* (2006). Effectiveness of atypical antipsychotic drugs in patients with Alzheimer's disease. *The New England Journal of Medicine*, **355**, 1525–38.
25. Lanctot, K.L., Best, T.S., Mittmann, N., *et al.* (1998). Efficacy and safety of neuroleptics in behavioral disorders associated with dementia. *Journal of Clinical Psychiatry*, **59**, 550–61.
26. Schneider, L.S., Pollock, V.E., and Lyness, S.A. (1990). A metaanalysis of controlled trials of neuroleptic treatment in dementia. *Journal of the American Geriatric Society*, **38**, 553–63.
27. Kallin, K., Gustafson, Y., Sandman, P.O., *et al.* (2004). Drugs and falls in older people in geriatric care settings. *Aging Clinical and Experimental Research*, **16**, 270–6.
28. Ballard, C. and Howard, R. (2006). Neuroleptic drugs in dementia: benefits and harm. *Nature Reviews. Neuroscience*, **7**, 492–500.
29. McKeith, I., Fairbairn, A., Perry, R., *et al.* (1992). Neuroleptic sensitivity in patients with senile dementia of Lewy body type. *British Medical Journal*, **305**, 673–8.
30. Committee of the Safety of Medicines. (2004). Atypical antipsychotic drugs and stroke. [online] accessed: 13/03/07, http://www.mhra.gov.uk.
31. United States Food and Drug Agency: Center for Drug Evaluation and Research. (2005). Deaths with antipsychotics in elderly patients with behavioral disturbances. [online] accessed: 13/03/07, http://www.fda.gov/cder/drug/advisory/antipsychotics.html.
32. Wang, P.S., Schneeweiss, S., Avorn, J., *et al.* (2005). Risk of death in elderly users of conventional vs. atypical antipsychotic medications. *The New England Journal of Medicine*, **353**, 2335–41.
33. Livingston, G., Walker, A.E., Katona, C.L.E., *et al.* (2007). Antipsychotics and cognitive decline in Alzheimer's disease: the LASER-Alzheimer's disease longitudinal study. *Journal of Neurology Neurosurgery, and Psychiatry*, **78**, 25–9.
34. Lonergan, E.T., Cameron, M., and Luxenberg, J. (2004). Valproic acid for agitation in dementia. *Cochrane Database Systematic Reviews*, (2), CD003945.
35. Howard, R.J., Juszczak, E., Ballard, C.G., *et al.* (2007). Donepezil for the treatment of agitation in Alzheimer's disease. *New England Journal of Medicine*, **357**, 1382–92.
36. McKeith, I., Del Ser, T., Spano, P., *et al.* (2000). Efficacy of rivastigmine in dementia with Lewy bodies: a randomised, double-blind, placebo-controlled international study. *Lancet*, **356**, 2031–6.
37. Nyth, A.L. and Gottfries, C.G. (1990). The clinical efficacy of citalopram in treatment of emotional disturbances in dementia disorders. A Nordic multicentre study. *The British Journal of Psychiatry*, **157**, 894–901.
38. Teri, L., Logsdon, R.G., Uomoto, J., *et al.* (1997). Behavioral treatment of depression in dementia patients: a controlled clinical trial. *Journal of Gerontology: Psychological Sciences and Social Sciences*, **52**, P159–66.
39. Bains, J., Birks, J.S., and Dening, T.R. (2002). The efficacy of antidepressants in the treatment of depression in dementia. *Cochrane Database Systematic Reviews*, (4), CD003944.
40. Cummings, J.L. (2003). Use of cholinesterase inhibitors in clinical practice: evidence-based recommendations. *The American Journal of Geriatric Psychiatry*, **11**,131–145.
41. Brodaty, H., Green, A., and Koschera, A. (2003). Meta-analysis of psychosocial interventions for caregivers of people with dementia. *Journal of the American Geriatrics Society*, **51**, 657–64.
42. Gillick, M.R. (2000). Rethinking the role of tube feeding in patients with advanced dementia. *The New England Journal of Medicine*, **342**, 206–10.

4.1.14 Remediation of memory disorders

Jonathan J. Evans

Introduction

Memory problems are a feature of the majority of psychiatric and neurological conditions. Any condition that affects the physical or functional integrity of the brain is likely to have an impact on some aspect of a person's ability to remember, as successful remembering involves many different interacting cognitive systems (see Chapter 2.5.3 on the neuropsychology of memory and Chapter 4.1.12 on the amnesic syndromes). Furthermore, mood disorders such as anxiety or depression, which impair concentration, also reduce the efficiency of memory.

Remembering difficulties disrupt the ability to participate effectively in activities of daily living, as well as social, leisure, and vocational activities. For some, memory problems will be mild and cause only minor inconvenience in everyday life. Others, such as those with the amnesic syndrome that accompanies dysfunction in limbic system structures, may be severely disabled by their memory impairment. People forget to do things (e.g. take medication, turn-off the cooker, pay bills, attend appointments, pass on messages), forget what they have been told, forget people's names, forget where they left things (e.g. keys, the car in the car park), find it difficult to remember routes or learn new procedures, have difficulty recollecting personal experiences, and so on. Such problems lead to frustration, lowered self-confidence, and dependence on others. As such they represent an important therapeutic target.

Assessment of the nature of the memory disorder and the functional consequences for the individual should precede remediation intervention planning. As far as remediation of memory is concerned, although the future in terms of biological treatments is promising,[1] for the present time pharmacological options remain limited (see Chapter 6.2.7). The most effective treatments are cognitive rehabilitation techniques. These include use of memory aids, which function as cognitive prostheses, and methods of learning that promote more effective acquisition of knowledge or skills.

Planning memory remediation—assessment

The World Health Organization International Classification of Functioning, Disability and Health[2] provides a helpful framework for the assessment and remediation of cognitive deficits including memory impairment.[3] ICF, which complements the diagnostic approach of ICD, emphasizes that health (or illness) and functioning can be considered at the level of body structure (pathology), body function (impairment), activities, and participation. Application of this framework in relation to assessment of memory is illustrated in Box 4.1.14.1.

An assessment of memory should therefore address both the impairment *and* the functional consequences for the individual patient. This is important because treatment interventions will differ depending on the form and severity of the memory impairment and the nature of the everyday problems. Such an assessment will of course typically be just one part of a broader assessment of cognition—memory impairment is the focus here, but the same principles apply to all other cognitive impairments.

Box 4.1.14.1 The WHO ICF model and its relationship to assessment of memory

ICF classification	Example in relation to assessment of memory	Approaches to assessment/ investigation
Body structure (*Pathology*)	Loss of cholinergic neurones in basal forebrain region affecting functioning of medial temporal lobe limbic system (Alzheimer's disease)	Physical investigations (e.g. routine medical and brain imaging investigations)
Body function (*Impairment*)	Episodic memory deficit	Standardized neuropsychological assessment tools (e.g. Wechsler Memory Scales III; Rivermead Behavioural Memory Test)
Activities (*Disability*)	Failure to remember to do important tasks, failure to remember events that have happened, or things previously told	Clinical interview with patient and significant other, questionnaires, and observation
Participation (*Handicap*)	Inability to work; increased dependence on others; inability to participate in leisure activities	Clinical interview; quality of life measures

Assessment of memory impairment

Memory impairment is assessed through the use of standardized neuropsychological assessment tools (see Chapter 1.8.3).

Assessment of functional consequences of memory impairment

Activity limitations arising from memory disorders can be assessed through clinical interview with the patient and proxy, but it can also be helpful to use a standardized questionnaire to aid information gathering. Several questionnaires exist for this purpose. The Prospective and Retrospective Memory Questionnaire[4] is one example of a useful, brief questionnaire with self-rating and proxy-rating forms that address both prospective remembering (e.g. Do you fail to do something you were supposed to do a few minutes later even though it's there in front of you, like take a pill or turn off the kettle?) and retrospective remembering (e.g. Do you fail to recall things that have happened to you in the last few days?). This questionnaire also has normative data for self-rating and proxy-ratings.[5,6]

Awareness of the functional consequences of memory impairment may be limited on the part of the patient and the carer (see section on assessment of awareness below). It is possible that functional consequences will also be minimized (again by patient and under some circumstances the carer). In some cases there may be significant impairment of memory and associated limitations of activity, but a spouse/family may take on most or all of the remembering responsibility and hence the significant disability on the part of the patient may not represent a problem for patient or spouse/family. In this circumstance it is important to investigate whether there is adequate awareness of rehabilitation options.

Assessment of use of memory aids and strategies

Pre-morbid and current use of memory aids and strategies should also be discussed as part of the clinical interview. Given that the most effective approaches to memory rehabilitation are those that enable people with memory dysfunction to compensate for their impairment, it is important to understand what past experience of use of memory aids the patient has, and which aids/strategies are used currently. Some people will have made extensive use of memory aids and strategies throughout their life, and continue to do so in response to onset of memory problems. Others may have used aids and strategies in the past, but then do not use them despite the onset of memory problems. Others have little previous experience. Some examples of aids and strategies to investigate, drawn from a survey of use of memory aids by people with memory impairment,[7] are shown in Box 4.1.14.2.

Assessment of awareness of memory deficits

Awareness of impairment should also be examined as this will impact on the approach to remediation that will follow. To what extent is the patient aware of his or her memory (and any other cognitive) problems? Insight and awareness is a complex issue. Clare[8] presents a biopsychosocial model of the construction of awareness in Alzheimer's disease, though the principles of the

Box 4.1.14.2 Assessment of pre-morbid and current use of memory aids and strategies

Use of memory aids and strategies is a central component of memory rehabilitation. Investigation of prior experience, and current use, of aids and strategies is an important part of the assessment process. Below are examples of aids and strategies that are most commonly used by people with memory impairment.[3] This list is not exhaustive and it is important to ask whether any other aids/strategies are used.

- Wall calendar/chart/memo board
- Notebook
- Lists (e.g. things to do/shopping)
- Checklists (e.g. instructions for how to operate washing machine)
- Appointment diary/personal organizer (e.g. Filofax)
- Asking others to remind you to do something or of things that have happened
- Mental retracing (e.g. of steps when lost an object)
- Placing objects in unusual places (e.g. by door if need to take it when leaving or as reminder to do something)
- Leaving notes in special places
- Dosett box or other pill reminder
- Repetitive practice (learning something new by frequent repetition)
- Making associations
- Watch with date/alarm
- Having a daily routine
- Journal (a daily diary record of personal experiences)
- Daily timetable
- Alarm clock/timer
- Mobile phone (e.g. with alarm reminder/GPS navigation function)
- Electronic organizer (Personal Digital Assistant—PDA)
- NeuroPage (paging-based reminding system)

model apply to most neurological and indeed many psychiatric conditions. Another simple model, but one that is useful in clinical practice, is the hierarchical model of Crosson and colleagues[9] which suggests that awareness may be *intellectual*, *emergent*, or *anticipatory*. Intellectual awareness refers to knowing that you have an impairment, but not necessarily recognizing the occurrence of problems as they occur. Emergent awareness refers to 'online' awareness of problems as they occur, whilst anticipatory awareness refers to using knowledge of deficits to anticipate problems and taking steps to prevent problems occurring. This tripartite model of awareness can be helpful in formulating a patient's level of awareness of memory problems. The extent to which the patient's reporting of problems is discrepant from their relative's account (in interview or on questionnaires), or from what might be expected on the basis of standardized test results will give some indication of level of awareness. In addition it is useful to establish the extent to which the patient is aware of the type of memory problems that arise and the extent to which s/he makes attempt to compensate for the problems. Bear in mind that severe memory impairment may itself impact on awareness—patients may have difficulty remembering that, or what, they forget.

Planning memory remediation—treatment approaches

Memory remediation interventions must take account of several factors, including the form and severity of memory impairment, the presence/absence of additional cognitive impairment, and awareness of the deficit. With regard to form of impairment, the major distinction that is drawn is between primary/working memory and secondary/-term memory.

Remediation of primary/working memory

Primary/working memory refers to the process of briefly holding verbal or visuospatial information in mind, and in the case of working memory, manipulating information in the mental workspace. Primary/working memory processes are reflected in tasks such as digit span, with backward digit span seen as taxing working memory. Working memory is considered to be crucial for effective mental control and executive functions such as problem solving. There is some evidence, primarily from studies of patients with a diagnosis of schizophrenia that cognitive training (involving extensive practice on tasks, sometimes computerized, which make demands on working memory) improves working memory performance over and above control conditions.[10,11] Intervention for working memory problems as part of a more comprehensive cognitive remediation programme should therefore be given serious consideration, at least for patients with schizophrenia.[12] Perhaps ironically, given the acknowledgement that cognitive rehabilitation for schizophrenia has its roots in neurorehabilitation for traumatic brain injury,[10] there is a much more limited evidence base relating to the effectiveness of this type of intervention with other neurological conditions.[13]

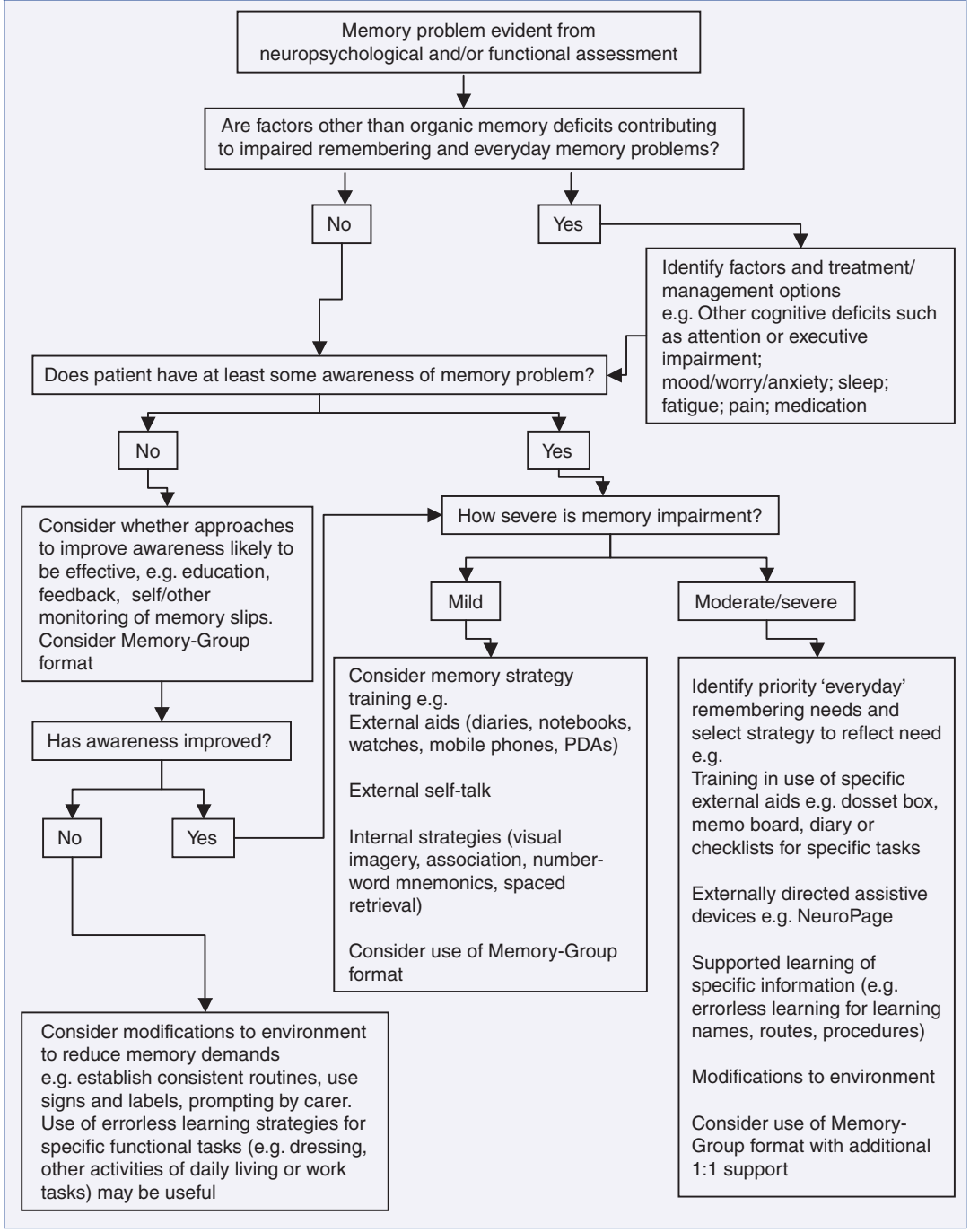
Remediation of secondary/long-term memory problems

Secondary or long-term memory refers to the process of encoding, storage, and retrieval of memories after a delay, where 'delay' means anything from a few minutes to a lifetime. Secondary memory is what is used to recall episodes, to acquire knowledge and to remember to do things. It is secondary memory that is impaired in amnesia. Box 4.1.14.3 provides a decision tree that reflects some of the processes involved in identifying memory remediation interventions.

(a) Are there contributory factors?

As part of the assessment and formulation of memory disorders and their functional consequences, one must consider whether a range of other factors are contributing to the functional disability. If so, then one should include intervention for these factors in the treatment plan. This includes treatment of mood disorders, sleep disorders, management of fatigue and pain, or adjustment of medication where possible. In some patients, memory problems will be secondary to

Box 4.1.14.3 Guidelines for genetic testing for dementia



impaired attention and concentration and if this is the case, then interventions to address these problems should be considered.

(b) Is the person aware of the problem?

The question of the patient's awareness of memory problems should be addressed. If there is no awareness then it is important to consider means of improving awareness before pressing on with specific remediation strategies. Improving awareness is sometimes straightforward and a question of providing basic information and feedback, but often it is more complex. Use of education and self/other monitoring of memory slips in conjunction with feedback can help. This must be done sensitively as minimizing of problems may be a psychological

coping mechanism and insensitive confrontation may be threatening. Working with patients in a group format can contribute to improving awareness—patients can be supported to provide feedback to each other. Patients may feel more able to acknowledge problems if others are also doing this in the group context.[14] *If awareness cannot be improved* (which may be the case if memory problems are occurring in the context of more global and severe cognitive impairment) then the strategy of modifying the environment to reduce memory demands on the patient must be considered. Environmental modifications include the use of prominent signs/labelling (e.g. of toilets, cupboards, draws, rooms) to support orientation in the physical environment. Establishing very fixed daily routines can also help develop behavioural habits. It may be the case that the patient requires prompting from carers. If this is the case then the minimal level of prompting required should be established and regularly monitored, and if possible prompting can be gradually reduced as behavioural sequences are learned.

(c) Strategies for mild problems

If the patient demonstrates at least some awareness of memory problems then the severity of the memory problem should be considered. Systematic reviews of cognitive rehabilitation[15,16] have recommended that different approaches are required for different levels of severity of memory disorder. There is no evidence that 'drill and repetitive practice' without additional strategy training is effective in improving memory.[17,18] However, in the context of mild memory impairment then memory strategy training is recommended.[15] Memory strategy training can be carried out on a 1:1 basis or in a group. The aim is to provide information and training in the use of a range of memory strategies which the patient learns to apply independently in specific situations in their own life, via homework tasks, over the course of the training programme. The patient is provided with a range of strategy options that s/he can select according to personal needs. This includes training in the use of external memory aids, such as many of those listed in Box 4.1.14.2. Internal memory strategies are also trained. These include strategies to aid deeper encoding of information. It is a well-established principle that deeper (more meaningful, personally relevant, emotionally salient), multi-modal (i.e. visual and verbal) processing of information results in more effective learning and recollection of that information.[19] For those with severe amnesia whether or not the information is processed deeply will make little difference, but for those with more mild impairment, strategies to enhance processing are more relevant. Strategies include visual imagery, categorization, association with established knowledge, motor movement (e.g. rehearsing in mind an action that has to be carried out at some later time) and spaced retrieval/expanding rehearsal (gradually increasing the time between successive trials of testing recollection of material to be learned). Craik and colleagues[20] and Evans[1] discuss internal strategies further.

(d) Strategies for more severe problems

For those with more severe memory impairment general strategy training is unlikely to be effective as the demands of learning a range of strategies and applying them when required are too great.[18] The approach recommended in this context is to try to map specific everyday remembering priorities to specific strategies. In other words, rather than providing a tool box of strategies and relying on the patient independently selecting the right tools for the right task when needed, the clinician establishes with the patient and carer what is essential to be learned/remembered and then considers how can this be achieved. For some this will be just one task for which one remembering strategy will be established. For others a more complex 'memory system' can be constructed to allow several remembering tasks to be achieved. Some of the commonly used external aids (memory notebooks, diaries, memo boards) will be used. For people with more severe impairment, formal training in learning how to use these aids consistently is required. A number of studies have shown that comprehensive training approaches can lead to effective use of memory journals, even in people with severe amnesia.[21–23] Kime[24] provides instruction on devising needs-led practical approaches to compensating for memory deficits.

(e) Electronic memory aids

These aids offer the major advantage of having the facility to prompt an action using alarms and so are particularly valuable in relation to prospective memory (remembering to do things). They also provide a means of combining a number of different memory aid functions (e.g. alarmed reminders, schedule, contact information, to-do list) into one portable tool. The most extensively evaluated electronic reminding system is NeuroPage.[25] Reminder messages are sent to standard alpha-numeric pagers worn by people with memory and/or planning problems, according to a pre-arranged schedule. This system has now been evaluated in randomized clinical trials and single case studies and shown to be very effective.[25,26] In recent years there has been a massive worldwide increase in use of mobile phones such that the vast majority of people acquiring cognitive impairment now will have had exposure to this technology before the onset of their memory deficit. This opens up the possibility of much greater use of portable reminding technology delivered via mobile phones, including the use of SMS text messaging.[27,28]

(f) Errorless learning

Another approach to remediation can be applied when there is a need to learn specific information or a procedure. Errorless learning is based on the principle that for those with severe memory impairment, learning will be most effective if errors can be avoided during the learning process. This is because memory-impaired patients may be more likely to repeat errors (as a result of intact *implicit* memory processes), but are unable to recollect that a response was an error. Thus errors become reinforced. Errorless learning techniques have been used for many years to teach new skills to people with learning disabilities and more recently this technique has been used with people with acquired neurological impairment and with schizophrenia. Baddeley and Wilson[29] published the first study demonstrating that people with amnesia learn better when prevented from making mistakes during the learning process. This finding has been replicated with people with a diagnosis of schizophrenia.[30] These were theoretical studies of word list learning. However, several single case and group studies have shown the benefit of errorless learning methods in teaching more practical, everyday information including learning names of people important in a person's life,[31] work tasks.[32] Kessels and de Haan's[33] meta-analysis concluded that errorless learning was more effective than standard, 'trial and error' conditions.

Many of the treatment interventions described here are relatively labour-intensive requiring a significant amount of clinician/therapist time for them to be successfully implemented. Occupational Therapists and Clinical Psychologists specializing in neuropsychology have relevant training in the assessment and treatment of memory disorders and hence memory remediation should be considered within the interdisciplinary context.

Summary and conclusions

Memory disorders are frequently encountered in clinical practice and cause significant disability. Memory should therefore be carefully assessed as part of routine clinical assessment. Restoration of normal functioning is not typically possible and remediation is therefore usually concerned with compensating for impaired memory. A range of treatment approaches is available, and the treatment of choice will depend on the form and severity of memory disorder and the functional problems faced by the patient.

Further information

Clare, L. and Wilson, B.A. (1997) *Coping with memory problems: a practical guide for people with memory impairments, their relatives, friends and carers*. Harcourt, London.

Evans, J.J. (2004). Disorders of memory. In *Clinical neuropsychology: a practical guide to assessment and management* (eds. L.H. Goldstein and J. McNeil), pp. 143–63. John Wiley & Sons, Chichester.

Kapur, N., Glisky, E., and Wilson, B.A. (2004). External memory aids and computers in memory rehabilitation. In *The essential handbook of memory disorders for clinicians* (eds. A.D. Baddeley, M. Kopelman, and B.A. Wilson). John Wiley & Sons, Chichester.

Information on managing memory impairment is also available in booklets from Headway, the brain injuries association (www.headway.org.uk/).

References

1. Evans, J.J. (2006). Memory dysfunction. In *Textbook of neural repair and rehabilitation* (eds. M.E. Selzer, M.D. Clarke, L.G. Cohen, P.W. Duncan, and F.H. Gage), pp. 461–74. Cambridge University Press, Cambridge.
2. World Health Organization. (2001). International classification of functioning, disability and health (ICF). WHO, Geneva.
3. Wade, D. (2003). Applying the WHO ICF framework to the rehabilitation of patients with cognitive deficits. In *The effectiveness of rehabilitation for cognitive deficits* (eds. P. Halligan and D. Wade). Oxford University Press, Oxford.
4. Smith, G., Della Sala, S., Logie, R.H., *et al.* (2000). Prospective and retrospective memory in normal ageing and dementia: a questionnaire study. *Memory*, **8**, 311–21.
5. Crawford, J.R., Smith, G., Maylor, E.A., *et al.* (2003). The Prospective and Retrospective Memory Questionnaire (PRMQ): normative data and latent structure in a large non-clinical sample. *Memory*, **11**, 261–75.
6. Crawford, J.R., Henry, J.D., Ward, A.L., *et al.* (2006). The Prospective and Retrospective Memory Questionnaire (PRMQ): latent structure, normative data and discrepancy analyses for proxy-ratings. *The British Journal of Clinical Psychology*, **45**, 83–104.
7. Evans, J.J., Needham, P., Wilson, B.A., *et al.* (2003). Which memory impaired people make good use of memory aids? Results of a survey of people with acquired brain injury. *Journal of the International Neuropsychological Society*, **9**, 925–35.
8. Clare, L. (2004). The construction of awareness in early stage Alzheimer's disease: a review of concepts and models. *The British Journal of Clinical Psychology*, **43**, 155–75.
9. Crosson, B., Barco, P.P., Velozo, C.A., *et al.* (1989). Awareness and compensation in post-acute head injury rehabilitation. *The Journal of Head Trauma Rehabilitation*, **4**, 46–54.
10. Twamley, E.W., Jeste, D.V., *et al.* (2003). A review of cognitive training in schizophrenia. *Schizophrenia Bulletin*, **29**, 359–82.
11. Kurtz, M.M., Seltzer, J.C., Shagan, D.S., *et al.* (2007). Computer-assisted cognitive remediation in schizophrenia: what is the active ingredient? *Schizophrenia Research*, **89**, 251–60.
12. Wykes, T. and Reeder, C. (2005). *Cognitive remediation therapy for schizophrenia*. Routledge, London.
13. Cicerone, K., Levin, H., Malec, J., *et al.* (2006). Cognitive rehabilitation interventions for executive function: moving from bench to bedside in patients with traumatic brain injury. *Journal of Cognitive Neuroscience*, **18**, 1212–22.
14. Evans, J.J. and Wilson, B.A. (1992). A memory group for individuals with brain injury. *Clinical Rehabilitation*, **6**, 75–81.
15. Cicerone, K.D., Dahlberg, C., Malec, J., *et al.* (2005). Evidence-based cognitive rehabilitation: updated review of the literature from 1998 through 2002. *Archives of Physical Medicine and Rehabilitation*, **86**, 1681–92.
16. Cappa, S., Benke, T., Clark, S., *et al.* (2005). EFNS guidelines on cognitive rehabilitation: report of an EFNS task force. *European Journal of Neurology*, **12**, 665–80.
17. Cicerone, K.D., Dahlberg, C., Kalmar, K., *et al.* (2000). Evidence-based cognitive rehabilitation: recommendations for clinical practice. *Archives of Physical Medicine and Rehabilitation*, **81**, 1596–615.
18. Clare, L., Woods, R.T., Moniz Cook, E.D., *et al.* (2003). Cognitive rehabilitation and cognitive training for early stage Alzheimer's disease and vascular dementia (Cochrane Review). In *The cochrane library*, (4). John Wiley & Sons, Chichester, UK.
19. Craik, F.I.M. and Lockhart, R.S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, **11**, 671–84.
20. Craik, F.I.M., Winocur, G., Palmer, H., *et al.* (2007). Cognitive rehabilitation in the elderly: effects on memory. *Journal of the International Neuropsychological Society*, **13**, 132–42.
21. Sohlberg, M.M. and Mateer, K. (1989). *Introduction to cognitive rehabilitation: theory and practice*. The Guilford Press, New York.
22. Donaghy, S. and Williams, W. (1998). A new protocol for training severely impaired patients in the usage of memory journals. *Brain Injury*, **12**, 1061–77.
23. Kime, S.K., Lamb, D.G., and Wilson, B.A. (1996). Use of a comprehensive programme of external cueing to enhance procedural memory in a patient with dense amnesia. *Brain Injury*, **10**, 17–25.
24. Kime, S.K. (2006). *Compensating for memory deficits using a systematic approach*. AOTA Press, Bethesda MD. (www.aota.org)
25. Wilson, B.A., Emslie, H., Quirk, K., *et al.* (2001). Reducing everyday memory and planning problems by means of a paging system: a randomised control crossover study. *Journal of Neurology, Neurosurgery, and Psychiatry*, **70**, 477–82.
26. Evans, J.J., Emslie, H., and Wilson, B.A. (1998). External cueing systems in the rehabilitation of executive impairments of action. *Journal of the International Neuropsychological Society*, **4**, 399–408.
27. Pijnenborg, G.H.M., Withaar, F.K., Evans, J.J., *et al.* (2007). SMS text messages as a prosthetic aid in the cognitive rehabilitation of schizophrenia. *Rehabilitation Psychology*, **52**(2), 236–40.
28. Fish, J., Evans, J.J., Nimmo, M., *et al.* (2007). Rehabilitation of executive dysfunction following brain injury: Content-free cueing improves everyday prospective memory performance. *Neuropsychologia*, **45**, 1318–30.

29. Baddeley, A.D. and Wilson, B.A. (1994). When implicit learning fails: amnesia and the problem of error elimination. *Neuropsychologia*, **32**, 53–68.
30. Pope, J.W. and Kern, R.S. (2006). An 'errorful' learning deficit in schizophrenia? *Journal of Clinical and Experimental Neuropsychology*, **28**, 101–10.
31. Clare, L., Wilson, B.A., Breen, K., *et al.* (1999). Errorless learning of face-name associations in early Alzheimer's disease. *Neurocase*, **5**, 37–46.
32. Kern, R.S., Green, M.F., Mintz, J., *et al.* (2003). Does 'errorless learning' compensate for neurocognitive impairments in the work rehabilitation of persons with schizophrenia? *Psychological Medicine*, **33**, 433–42.
33. Kessels, R.P.C. and DeHaan, E.H.F. (2003). Implicit learning in memory rehabilitation: a meta analysis on errorless learning and vanishing cues methods. *Journal of Clinical and Experimental Neuropsychology*, **25**, 805–14.

4.2

Substance use disorders

Contents

4.2.1 Pharmacological and psychological aspects of drugs abuse

David J. Nutt and Fergus D. Law

Drug abuse, misuse, and addiction are major issues in society because of their enormous personal, social, and economic costs and their important psychiatric components.[1] Many drug treatment programmes are run by psychiatrists, and the evidence strongly supports the notion that a significant proportion of severe drug abusers are psychiatrically ill. Moreover, drug misuse appears to be becoming more frequent in patients with other psychiatric disorders, where it can lead to problems in treatment and poorer outcomes. It is therefore essential for all psychiatrists and related health professionals to have a good understanding of the basis of drug misuse.

Why do people take drugs?

A very common misconception is that drug misuse is simply a **search for fun**. In fact, people take drugs for many reasons other than to get the buzz or high. Indeed, studies have shown that straightforward pleasure seeking is the primary reason for initiation of drug use in fewer than 20 per cent of individuals. Whilst the high or buzz is the most obvious pleasurable effect, many people also describe using drugs to feel comfortably numb, pleasantly drowsy, or full of energy and confidence. Many others will be chasing the high or buzz that they first experienced, always seeking the intensity of their first experience. Still others will be self-medicating for anxiety, anger, pain, boredom, lack of motivation,

lack of self-confidence, and many other aversive states including drug withdrawal.

The main reason to try to ascertain the reasons for drug use is that in many cases identification of the cause can lead to effective interventions. For example, many **alcoholics** will point to **anxiety** as their reason for drinking;[2] indeed, social anxiety is one of the most common causes of alcoholism in young men.[3] If this can be treated (e.g. by selective serotonin reuptake inhibitors) then they are frequently able to become abstinent or even drink normally. **Social anxiety** and **attention-deficit disorder** are common reasons for the use of stimulants. **Depression**, is particularly likely to lead to excess alcohol intake, and a vicious cycle can develop because both alcohol and its withdrawal are depressogenic. Alcohol is also one of the most serious risk factors for suicide. There is increasing use of **stimulants** and **cannabis** by **schizophrenic** patients. In part this reflects the behaviour of their peer group but in part is because they can offset some of the more negative aspects of the illness and medication side effects. As both these types of drugs can worsen psychotic illness, dealing with drug misuse in this group is a priority.

Other factors affecting drug use may be less amenable to psychiatric intervention, such as **pressure from peers** or others. For instance, female opiate addicts often have a male partner who also uses drugs or even deals drugs. Should she stop use, relapse is almost certain to occur if she continues to live with this partner. Another reason for drug use is to reduce **pain** or **boredom**, the latter being a common reason given by disadvantaged youth in areas of high unemployment and poor environmental quality such as inner cities or out-of-town housing estates. Other reasons for drug use, especially with the psychedelics, include the **search for meaning** or for **mystical experiences**. Whilst not directly relevant to psychiatry, this use can precipitate psychotic episodes in susceptible individuals and may trigger schizophrenia.

Finally, it is important to remember that the reasons for use of a specific drug are not static. An opiate addict may use the same dose of **heroin** to get going in the morning, to 'top off' a pleasant experience later in the day, to deal with angry feelings when they occur, and to promote sleep at night. Similarly during a **drug-using career** different motivations may become dominant. This has been well characterized in opiate users where for many the initial use was for pleasure or escape. Over months, as physical dependence becomes increasingly apparent, use becomes driven by the need to avoid withdrawal and to feel normal at almost any cost.

Drug use and misuse

It is possible to view the issue of drug abuse from different perspectives, which range from the molecular and genetic through the pharmacological to the psychological and social. Each view has its merits and is important, but there is little doubt that an integrated view is necessary, because for most drugs and for most societies no one perspective can explain all the known features of drug abuse. However, for the purpose of this chapter we have concentrated on the psychological and pharmacological.[1,4–6]

Problem use, addiction, dependence, and craving

These are some of the most commonly used terms regarding drug misuse but at the same time they are also the most problematic. The use of drugs in any circumstance, therapeutic, or otherwise, can be associated with problems, although the nature and scale of this varies (see Table 4.2.1.1). The terms problem use and misuse usually refer to use of drugs (prescription or otherwise) for pleasure but with disregard for the personal or social dangers. For example, alcohol misuse can lead to irresponsible behaviour whilst intoxicated and, if prolonged, to liver, gastric, and nervous system damage without the individual necessarily being addicted or dependent.

Table 4.2.1.1 Potential problems with drug use

Type of drug use and associated issues	Examples/effects
Therapeutic use	
Adverse effects	Sedation, poor driving
Drug interactions	Increased drug levels
Withdrawal	Convulsions, delirium tremens
Drug use with pain	Difficulty reducing opiate dose
Misuse/problem use	
Illegality	Criminal records; social stigmatization
Intoxication	Physical/social damage
Excessive regular use	Physical/social damage
Injecting drug use	Infections, thromboses, Hepatitis C
Dependence	
Tolerance	Dose escalation
Withdrawal	Physical dependence
Urge to use/cannot abstain	Psychological dependence
Craving/drug seeking	Drug dominates life
Drug becomes dominant life goal	Personal/social decline
Reinstatement on relapse	Cycles of dependency

Addiction is a term that had become so misused in general parlance and had acquired such a pejorative edge, that in the past two decades attempts have been made to remove it from the psychiatric lexicon. Unfortunately, the replacement terminology of **dependence**, or the dependence syndrome, has been similarly devalued by popular usage. In fact there exists a spectrum of dependence ranging from physiological supplementation (as with insulin in diabetes mellitus) through to life-altering dependence on illicit drugs such as heroin (see Table 4.2.1.1). Addiction is still a useful construct if it is reserved for the collection of phenomena that occur at the extreme end of the dependence spectrum, and includes the concept of social and personal decline associated with drug use, as well as **craving, tolerance**, and **withdrawal** symptoms (cf. DSM-IV and ICD-10).

Another area of some confusion is the distinction between **physical and psychological dependence**. When originally conceived, this distinction was helpful in that it emphasized that drug dependence was more than just physical adaptation to drugs as manifest by withdrawal symptoms, and that psychological processes, especially drug liking, were also important. However, drugs without obvious physical withdrawal syndromes (e.g. stimulants) also result in measurable physiological withdrawal changes in sleep and activity as well as measurable psychological changes such as those in mood. In addition, new neuroimaging techniques such as **PET, SPECT,** and **functional MRI** are beginning to reveal the brain circuits underlying the pleasurable effects of drugs, and this has

resulted in a blurring of the distinction between physical and psychological processes. For example, the plate shows a PET scan of heroin addicts in which the brain regions showing increased blood flow activated by craving for heroin are illuminated using the radiotracer oxygen-15 (Plate 4.2.1.1). Similar studies have revealed the brain regions involved in the pleasurable effects of opiates and stimulants.[7] Thus there is a clear convergence in terms of mechanisms, but in terms of treatment regimens the distinction between physical and psychological remains.

Craving is also a term that is widely used yet ill-defined. Craving is a desire, which most commonly is taken to mean a strong and sometimes irresistible desire to use a drug. The emotional valence of craving is not necessarily pleasurable. Craving can reliably be elicited in situations of negative valence. It is commonly found in withdrawal, when it can lead to relapse. Craving can also be present as an urge or desire to use a drug although the sufferer may be actively denying or resisting its presence. The complex interplay of physical and psychological processes is well exemplified by the physical responses that craving can produce. For example when opiate-dependent subjects are shown drug-related paraphernalia they may experience emotions that range from pleasurable anticipation to early withdrawal (shaking, tearing of the eyes, pupil dilatation, etc.). Each one of these experiences can lead to a desire to use the drug, that is craving.

Studies in both animals and humans have demonstrated that **conditioning** occurs to both the positive and negative aspects of craving.[8] **Tolerance** is to a large extent a conditioned response, particularly related to the environmental context in which a drug has been taken.[9] Thus an environmental context which is drug familiar results in physiological changes in the brain in preparation for the drug effect, and thus less actual drug effect occurs (i.e. tolerance). However, in a novel context, such preparatory changes do not occur so that a standard drug dose will result in a larger drug effect and a potentially fatal outcome. Thus the lethality of a drug is largely dependent on the environment in which it is taken.

Attempts have been made to dissect out the subcomponents of craving using questionnaires. The best known of these are the set designed by Tiffany *et al.*[10] who independently rate the five main subcomponents of craving—urges and desires to use, intention and planning to use, anticipation of positive outcome, anticipation of relief from withdrawal or negative outcome, and loss of control over use. Ongoing neuroimaging studies are beginning to support this multiprocess view of craving by revealing activation or inhibition of different brain regions to be correlated with individual symptom clusters.

There is also increasing evidence that the particular **cognitions** of patients may be important for treatment, especially during withdrawal. Just as panic disorder patients have catastrophic cognitions, addicted patients may have a high fear of craving and other withdrawal symptoms in association with related catastrophic cognitions. This detoxification fear has been measured in opiate addicts, and shown to predict outcome.[11] Withdrawal expectations also play a significant role,[12] and a 15 to 30 min explanation of what the opiate detoxification involves may reduce the measured withdrawal distress by over one-third. Indeed, such is the strength of psychological factors in addiction treatment, there is little doubt that drug treatments should always be combined with the appropriate psychological interventions.

Psychological processes and treatment implications

One of the most influential models in addiction treatment is known as the **stages of change model**.[13] The stage of change that a person can be identified as being at determines the therapeutic approach and type of treatment offered. Thus at the precontemplation stage where there is no recognition of a need for treatment, there is no point in offering intensive treatment interventions. Similarly, at the contemplation stage when treatment is being considered, the appropriate intervention is to help the person clarify their views and build their motivation to change rather than offering active treatment. Indeed, it is only in the decision and action stages that treatment should be actively offered and facilitated.

The brief counselling technique of **motivational interviewing**[14–16] has been proved to improve outcome effectively, and ties in well with the stages of change model. In the early stages the therapy is focused on encouraging the patient to reduce or resolve their ambivalence, which acts as their psychological barrier to treatment. The patient in this client-centred but focused therapy is facilitated to discover the solutions to their own problems themselves. This approach of accepting the client's current level of thinking (rather than offering ready-made solutions, or confronting them, or trying to argue them into the solution) has been shown to be surprisingly effective in the clinical trials.[16] The effectiveness of this technique has resulted in a new understanding of motivation, which is seen as a dynamic state rather than as a fixed state, and one which can be influenced by the therapeutic stance.

Other cognitive therapies also make significant contributions to treatment. **Relapse prevention** involves the teaching of cognitive and behavioural strategies for dealing with high-risk situations and mental states.[17,18] **Other cognitive behavioural therapies**, including extinction of conditioning, contingency management, community reinforcement techniques,[19] and indeed Beck's cognitive therapy,[20] have been effectively applied to substance misuse. The very large Project MATCH (matching alcoholism treatments to client heterogeneity) study of alcohol treatments compared three types of treatment and found that motivational enhancement, 12-step facilitation, and cognitive behavioural therapy were equally effective overall, although each therapy excelled in certain subgroups.[21,22] Based on these results it seems likely that specific therapies targeted at specific issues of importance in patients with addiction are roughly equally effective overall, but that we do not yet know enough to confidently match specific patient subtypes to specific therapies.

A number of **other therapies** have also been shown to be effective, particularly in the alcohol field, including self-control training, self-help groups, marital and family therapy, coping and social skills training, anxiety and stress management, aversion therapies, and brief intervention strategies.[23,24] The Cochrane reviews found that there was insufficient evidence to prove the effectiveness of psychosocial interventions used alone, but that there was added benefit from combining such interventions with pharmacological treatments in both maintenance and detoxification.[4–6]

Personality variables and the genetics of addiction

The role of personality in addiction is a major issue, with some believing in an 'addictive personality' and others suggesting

different personality types might predispose to different aspects or forms of drug misuse.[1] In this highly controversial field a few facts are generally agreed. Predisposition to experiment with both licit and illicit drugs is more likely in those with sensation-seeking or impulsive behaviour traits, and in extroverts rather than introverts. However, once drug dependence is established, those with obsessional, dependent, or anxious characteristics find it hardest to stop.[1]

The genetics of drug abuse are beginning to be unravelled and already these studies have thrown up some important insights in relation to personality. The best studied dependence is that on alcohol, where the Scandinavian adoption studies have found the risk of alcoholism in male children of male alcoholics is the same regardless of whether the child is reared with the alcoholic father or by a non-drinking adoptive family. Building on these data, Cloninger[25] has identified two main forms of alcoholism. Type I is the late-onset form that has low inheritance and is associated with anxiety and stress which drinking is used to relieve, often in binges. In contrast, Type II alcoholism starts at a younger age with a heavy regular intake and is associated with antisocial personality traits and criminality. This form is male limited, is associated with impulsivity, and may be related to underfunctioning of brain 5-hydroxytryptamine systems, as genetic polymorphisms of 5-hydroxytryptamine receptors and enzymes have been found in these subjects.[26]

How abused substances affect the brain

The brain works by transmitting information between neurones using the primary neurotransmitters. The **primary neurotransmitters** are glutamate, which is stimulatory (i.e. it turns neurones on), and the closely related amino acid γ-aminobutyric acid (GABA), which is inhibitory (i.e. it turns neurones off). The appropriate balance between these neurotransmitters leads to the brain processes underlying action, sensation, learning, and memory. **Secondary transmitters** are the monoamines and peptides such as dopamine, 5-hydroxytryptamine, noradrenaline (norepinephrine), acetylcholine, and endogenous opiates. These add the tone, valence, and emotion to the primary processes, and some such as noradrenaline are important in memory formation. All 'drugs' (probably even solvents through indirect effects) act by interfering with these neurotransmitters in ways summarized in Table 4.2.1.2. However, it is important to realize that the brain has its own **endogenous 'addictive' neurotransmitters**. The best known are the endogenous opioid peptides such as the endorphins and enkephalins, but there are also endogenous cannabinoids (anandamide) and probably others. It is not yet known whether these endogenous substances are mediators of addiction to cannabis or other drugs, although this would certainly seem possible.[27–29]

What is certain is that some of the most addictive agents (especially the full agonist opiates such as heroin/morphine) act on the endogenous **opioid neurotransmitter pathway**, but with a much greater effect than the natural transmitter. The profound ability of opiates such as heroin to produce addiction is because these drugs highjack the natural transmitter system leaving normal levels of stimulation seeming tame by comparison. Treatment with partial agonist opiates such as **buprenorphine** offer a compromise in that they are less addicting than heroin yet restore some of the

Table 4.2.1.2 Drugs and transmitters

Drug class	Endogenous transmitter	Treatment implications
Mimic natural transmitters		
Opiates (alcohol)	Endorphins/enkephalins	Antagonists (naltrexone) Partial agonists (buprenorphine)
Cannabis	Anandamide/others	Antagonists
Alcohol	GABA	GABA modulator (? acamprosate)
Benzodiazepines/barbiturates	GABA	Partial benzodiazepine agonists Antagonists (flumazenil)
Nicotine	Acetylcholine	Antagonists (mecamylamine) Partial agonist (varenicline)
Release transmitters		
Cocaine (buproprion)	DA	Other uptake site blockers
Amphetamines	DA	D2-receptor antagonists/partial agonists As cocaine
Nicotine	DA	As cocaine
Ecstasy	5-HT/DA	5-HT uptake blockers/antagonists
Alcohol, solvents?	NA/DA	NA/DA uptake blockers
Block transmitters		
Alcohol	Glutamate	Glutamate modulators (acamprosate)
Barbiturates	Glutamate	Glutamate modulators
LSD/other psychedelics	5-HT	Antagonists

DA, dopamine; GABA, γ-aminobutyric acid; 5-HT, 5-hydroxytryptamine; NA, noradrenaline.

brain's deficiency of opiate tone. They also have the advantage of being much safer than full agonists in overdose and rarely cause death from respiratory depression.[30] Other drugs, in particular alcohol, seem to act in part by indirectly stimulating the endogenous opioid system, which is why opioid antagonists such as naltrexone can be useful treatments.[31]

Other drugs act on the natural stimulant transmitter **dopamine**. Dopamine deficiency (for instance in Parkinson's disease) has long been known to limit motor behaviour. Stimulant drugs increase energy and stamina by increasing the synaptic levels of dopamine, either by increasing the release or by blocking its reuptake in the basal ganglia. Many drugs of addiction can also increase dopamine availability in other brain regions, the two most important being the nucleus accumbens and the prefrontal cortex.[32] A huge body of evidence points to the nucleus accumbens as being a critical gateway in drug misuse. Most abused substances (with possible exceptions of opiates and benzodiazepines) act to increase dopamine release in this region. How they do this varies—cocaine and nicotine act at the level of the dopamine terminals, whilst cannabis and alcohol activates the cell bodies in the brain stem. The net effect is to increase dopamine transmission out of the nucleus accumbens into the basal ganglia and thalamus, frontal cortex, amygdala, and hypothalamus.[1]

This circuit is the one that was shown by Olds in the 1950s to sustain electrical self-stimulation in rats and is the brain's own **reward circuit**. It is normally activated by positive reinforcers, such as food, water, and sex that are critical to survival. Because drugs of abuse produce greater effects than the natural reinforcers, the resultant effect is that the brain directs its normal drives away from the natural reinforcers and towards the more pleasurable drugs. In severe addiction, which frequently occurs with the most powerful reinforcers (such as heroin and cocaine), all natural drives may be subsumed to an overwhelming search for and use of the drug. Thus addicts may give up sex, grooming, hygiene, relationships, work, hardly eat or drink, and ignore health problems.

The routes and risks of addiction

In addition to its impact on the social aspects of life, drug misuse can lead to significant medical problems. The dangers of drug abuse relate to two main factors; the route of use of the drug and the effects the drug has outside of the reinforcement circuit of the brain.

For most drugs of abuse the faster the drugs reach their target site in the brain the better they are liked and the more psychologically reinforcing they are. Indeed, the 'pharmaceutical' history of most abused drugs illustrates the progressive refinement of their preparation, in order to accelerate their rate of entry into the brain. A good example is **cocaine**. The Andean Indians originally experienced its effects from chewing coca leaves, which released low levels of cocaine slowly. An increase in vigour and a resistance to fatigue is produced, but little pleasure. Over the centuries cocaine has become more refined, first to paste and then to cocaine hydrochloride powder (snow) which when taken nasally produces high levels in the brain within 5 to 10 min and a clear 'high'. Further refinement to the free base produces a more lipophilic form (crack) that can be smoked, resulting in entry into the brain in seconds. Intravenous drug use also serves the same purpose of getting the drug to the active site very fast.

A similar process of pharmaceutical refinement to accelerate brain entry has taken place with the **opiates**. Smoking opium is a method of delivering morphine and related substances reasonably quickly but in low amounts. Refining opium into its active constituents (e.g. morphine) means that higher doses are more easily ingested. However, morphine crosses the blood–brain barrier relatively slowly and has therefore been largely supplanted by opiates such as heroin that cross more rapidly. Heroin is a diacetylated synthetic derivative of morphine that is more lipophilic, meaning that it is able to enter the brain more rapidly and give a better rush. Interestingly, the active form of heroin is morphine; heroin has to be deacetylated before it can act, which proves that pharmacokinetic differences are the critical variable with opiate preference. Similarly, codeine is also inactive until metabolized to morphine, but because this happens very slowly codeine has less abuse potential than morphine.

The **benzodiazepines** were abused relatively rarely until the advent of gel-filled capsules of temazepam. These provided experienced intravenous opiate users with a convenient source of a concentrated drug, which they began to experiment with in the late 1980s. In an attempt to stop this, the drug was reformulated in wax, which led to users heating up the caplets until they melted and then injecting the hot solution into their veins (hot lining). Unfortunately at body temperature the wax solidified, blocking the veins and arteries into which it was administered, leading to severe ischaemia that often lead to gangrene and the loss of the limb. Since there are no therapeutic advantages of temazepam over other benzodiazepines that are much less abusable, this drug has recently been put under a higher degree of regulatory control in the United Kingdom in order to deter its prescription and misuse.

As well as affecting the relative reinforcing actions of abused drugs, **the rate of brain entry** also contributes to risk. A very rapid drug entry makes dose adjustments difficult or impossible and so predisposes to overdose. This is most obvious for intravenous use of opiates where respiratory depression is the main cause of death, but is less common with smoked opiates as intake can more easily be titrated to the desired effect.

The **route of use** also affects risk, most notably with the risk of infection from intravenous use, especially when needles are not cleaned or are shared. The majority of current intravenous users are Hepatitis C positive and we can therefore expect cirrhosis to become a major cause of their death in the next decade or so. This also raises ethical and economic issues; interferon treatment significantly reduces the progression of the disease but is costly and its routine use in addicts would be massively expensive and likely to cause public disquiet. The other main infections are hepatitis B and AIDS. The frightening rise of AIDS in drug abusers, where it occurred faster than in any other group, was the main impetus to the harm-reduction approach becoming the treatment style of the 1990s. Needle-exchange programmes and increased methadone availability were both proven to reduce the spread of AIDS and have become the cornerstone of treatment in many countries.

Relative risks of abused drugs

This is a critical issue in relation to directing legal as well as medical inputs into drug abuse. There are four main factors, which have to be taken into account in determining relative risk:[33]

- risk due to the route of use
- risk of the drug itself

- extent to which the drug controls behaviour (addictiveness)
- ease of stopping

The risks due to the route have been covered above. The risks of the drugs themselves are determined by standard tests and clinical experience and can be encapsulated in concepts such as the **therapeutic index**. This is the ratio of toxic dose to therapeutic (or usual) dose. The ratio is very low for heroin and similar opiates, for cocaine especially crack, and for intravenous temazepam and oral ecstasy. It is quite high for psychedelics, cannabis, benzodiazepines, and orally used stimulants such as amphetamines. Another important consideration is the **health complications** of long-term use, which by and large reflects the therapeutic index. An exception to this is the opiates, which, provided sterile administration is used, are thought to have little detrimental effect, even when used chronically and intravenously. Chronic cocaine can lead to cardiac damage, and heavy cannabis smoking causes precancerous change in the same way as tobacco smoking, as well as causing greater levels of chronic bronchitis.

The degree of **control over behaviour** that the drug elicits is a major factor in drug dependence, and is the closest concept to addictiveness. Although the route of administration is another critical variable, we can make some reasonable generalizations. Strong opiates and cocaine are the most addictive, being overall as addictive as nicotine. The benzodiazepines, ecstasy, and psychedelics are the least addictive, and are significantly less addictive than alcohol.

There are three main factors contributing to drugs gaining control over behaviour, all of which affect the ease with which a drug may be stopped. The first is the pleasure a drug produces—the positive drive for use (pleasure giving and seeking). The others both involve the pain of abstinence—withdrawal in both physical and psychological terms—which leads to drug use to relieve it (discomfort escape). The pattern of drug use during an addiction career generally begins with the quest for pleasure and progressively evolves into the escape from withdrawal pain as neuroadaptive processes develop. In this context it may be thought that withdrawal discomfort is best limited to symptoms with a clear physical symptomatology, that is the autonomic symptoms indicative of physical dependence. But in terms of addictiveness, psychological withdrawal may in fact be more important than physical withdrawal. This is illustrated by the finding that those dependent on opiates for medical reasons, although physically dependent, experience little craving and risk of relapse once detoxified, provided the reason for being on the opiate resolves. The ease of stopping the drug thus depends on both the physical and psychological withdrawal symptoms, as well as the ability of the drug to provide positive reinforcement.

It is possible to provide rough guides for these three processes for each drug, so that the overall addictiveness potential can be gauged (Table 4.2.1.3). For completeness, the main licit drugs are also shown as well as another highly motivated behaviour which can produce a state of addiction/dependence, that is gambling.

Table 4.2.1.3 Addictiveness of various agents and activities

	Pleasure giving	Physical withdrawal problems	Psychological withdrawal problems
Opiates	+++	+++	+++
Amphetamines	++	+	+
Crack/cocaine	+++	++	+++
Cannabinoids	+	+	+
Barbiturates	++	+++	++
Benzodiazepines	+	++	++
Ecstasy	++	+	0
Psychedelics	++	+	0
Cigarettes	+++	+	+++
Alcohol	++	+++	++
Caffeine	+	+	+
Gambling	++	0	++

0 none; + slight; ++ moderate; +++ strong.

Further information

Nutt, D.J., Robbins, T.W., Stimson, G.V., *et al.* (2006). *Drugs and the future: brain science, addiction and society*. Elsevier, Burlington, MA.

NIDA (National Institute on Drug Abuse) www.nida.nih.gov

SAMHSA (Substance Abuse and Mental Health Services Administration) www.samhsa.gov

References

1. Nutt, D.J., Robbins, T.W., Stimson, G.V., *et al.* (2006). *Drugs and the future: brain science, addiction and society*. Elsevier, Burlington, MA.
2. George, D.T., Nutt, D.J., Dwyer, B.A., *et al.* (1990). Alcoholism and panic disorder: is the comorbidity more than coincidence? *Acta Psychiatrica Scandinavica*, **81**, 97–107.
3. Marshall, J.R. (1994). The diagnosis and treatment of social phobia and alcohol abuse. *Bulletin of the Menninger Clinic*, **58**, 58–66.
4. Amato, L., Minozzi, S., Davoli, M., *et al.* (2004). Psychosocial and pharmacological treatments versus pharmacological treatments for opioid detoxification. *Cochrane Database of Systematic Reviews*, (4), CD005031.
5. Amato, L., Minozzi, S., Davoli, M., *et al.* (2004). Psychosocial combined with agonist maintenance treatments versus agonist maintenance treatments alone for treatment of opioid dependence. *Cochrane Database of Systematic Reviews*, (4), CD004147.
6. Mayet, S., Farrell, M., Ferri, M., *et al.* (2004). Psychosocial treatment for opiate abuse and dependence. *Cochrane Database of Systematic Reviews*, (4), CD004330.
7. Schlaepfer, T.E., Strain, E.C., Greenberg, B.D., *et al.* (1998). Site of opioid action in the human brain: mu and kappa agonists' subjective and cerebral blood flow effects. *The American Journal of Psychiatry*, **155**, 470–3.
8. O'Brien, C.P., Testa, T., O'Brien, T.J., *et al.* (1997). Conditioned narcotic withdrawal in humans. *Science*, **195**, 1000–2.
9. Siegel, S., Hinson, R.E., Krank, M.D., *et al.* (1982). Heroin overdose death: contribution of drug–associated environmental cues. *Science*, **216**, 436–7.
10. Tiffany, S.T., Singleton, E., Haertzen, C.A., *et al.* (1993). The development of a cocaine craving questionnaire. *Drug and Alcohol Dependence*, **34**, 19–28 (erratium at 2004, **16**, 326).
11. Schumacher, J.E., Milby, J.B., Fishman, B.E., *et al.* (1992). Relation of detoxification fear to methadone maintenance outcome: 5-year follow-up. *Psychology of Addictive Behaviors*, **6**, 41–6.

12. Phillips, G.T., Gossop, M., and Bradley, B. (1986). The influence of psychological factors on the opiate withdrawal syndrome. *The British Journal of Psychiatry*, **149**, 235–8.
13. Prochaska, J.O. and DiClemente, C. (1983). Stages and processes of self-change of smoking: towards a more integrative model of change. *Journal of Consulting and Clinical Psychology*, **51**, 390–5.
14. Rollnick, S. and Miller, W.R. (1995). What is motivational interviewing? *Behavioral and Cognitive Psychotherapy*, **23**, 325–34.
15. Miller, W.R. and Rollnick, S. (2002). *Motivational interviewing: preparing people for change* (2nd edn). Guilford Press, New York.
16. Noonan, W.C. and Moyers, T.B.(1997). Motivational interviewing. *Journal of Substance Misuse*, **2**, 8–16.
17. Marlatt, G.A. and Gordon, J.R. (eds.) (1985). *Relapse prevention: maintenance strategies in the treatment of addictive behaviors*. Guilford Press, New York.
18. Wanigaratne, S., Wallace, W., Pullin, J., *et al.* (eds.) (1990). *Relapse prevention for addictive behaviours: a manual for therapists*. Blackwell, London.
19. Stitzer, M.L. and Higgins, S.T. (1995). Behavioral treatment of drug and alcohol abuse. In *Psychopharmacology: the fourth generation of progress* (eds., F.E. Bloom and D.J. Kupfer), pp. 1807–19. Raven Press, New York.
20. Beck, A.T., Wright, F.D., Newman, C.F., *et al.* (eds.) (1993). *Cognitive therapy of substance abuse*. Guilford Press, New York.
21. Project MATCH Research Group. (1997). Matching alcoholism treatments to client heterogeneity: project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol*, **58**, 7–29.
22. Project MATCH Research Group. (1997). Project MATCH secondary *a priori* hypotheses. *Addiction*, **92**, 1671–98.
23. Miller, W.R., Brown, J.M., Simpson, T.L., *et al.* (1995). What works? A methodological analysis of the alcohol treatment outcome literature. In *Handbook of alcoholism treatment approaches: effective alternatives* (2nd edn) (eds. R.K. Hester and W.R. Miller), pp. 12–44. Allyn and Bacon, Boston, MA.
24. Edwards, E. and Dare, C. (eds.) (1996). *Psychotherapy, psychological treatments and the addictions*. Cambridge University Press, Cambridge.
25. Cloninger, C.R. (1987). Neurogenetic adaptive mechanisms in alcoholism. *Science*, **236**, 410–16.
26. Nielsen, D.A., Goldman, D., Virkkunen, M., *et al.* (1994). Suicidality and 5–hydroxyindoleacetic acid concentration associated with a tryptophan hydroxylase polymorphism. *Archives of General Psychiatry*, **51**, 34–8.
27. Nutt, D.J. (1997). Neuropharmacological basis for tolerance and dependence. In *Drug addiction and its treatment: nexus of neuroscience and behavior* (eds. B.A. Johnson and J.D. Roache), pp. 171–86. Raven Press, New York.
28. Nutt, D.J. (1999). Alcohol and the brain: pharmacological insights for psychiatrists. *The British Journal of Psychiatry*, **174**, 114–19.
29. Lingford-Hughes, A. and Nutt, D. (2003). Neurobiology of addiction and implications for treatment. *The British Journal of Psychiatry*, **182**, 97–100.
30. Law, F., Daglish, M.R.C., Myles, J.S., *et al.* (2004). The clinical use of buprenorphine in opiate addiction: evidence and practice. *Acta Neuropsychiatrica*, **16**, 246–74. (Erratum (2004), **16**, 326.)
31. Volpicelli, J.R., Alterman, A.I., Hayashida, M., *et al.* (1993). Naltrexone in the treatment of alcohol dependence. *Archives of General Psychiatry*, **49**, 876–80.
32. Di Chiara, G. (1995). The role of dopamine in drug abuse viewed from the perspective of its role in motivation. *Drug and Alcohol Dependence*, **38**, 95–137.
33. Nutt, D.J., King, L.A., Saulsbury, W., *et al.* (2007). Developing a rational scale for assessing the risks of drugs of potential misuse. *Lancet*, **369**, 1047–53.

4.2.2 Alcohol use disorders

Contents

4.2.2.1 Aetiology of alcohol problems

Juan C. Negrete and Kathryn J. Gill

Introduction

Approximately 8 out of every 10 persons living in Europe and the Americas would report consuming alcoholic beverages in their lifetime,[1] and the norm is for drinking to start in adolescence: in 2003 the average age of first drink in the United States was 14 years old.[2] Also in the year 2003, 79.3 per cent of persons aged 15 years or more in Canada reported to be current users of alcohol, and 22.6 per cent admitted to having exceeded the country's safe drinking guidelines (i.e. no more than 14 units/week for males and 12 units/week for females). The same survey elicited a rate of 'hazardous drinkers' of 13.6 per cent, defined as all respondents who scored 8+ on the AUDIT screening questionnaire.[3]

Epidemiological data in the United States indicates that roughly one in seven persons who start drinking will develop an alcohol dependence disorder in the course of their lives.[4] The figure is higher among men when compared to women. Of course it is also higher if other clinical forms of alcohol misuse (i.e. alcohol abuse/harmful drinking) are included in the rates in addition to dependence.

A moderate level of alcohol use appears to be relatively harmless; and there exist public health guidelines on 'safe' drinking practices. The recommendations vary considerably from country to country, but they all assume a greater vulnerability to alcohol effects in the female gender. In the United Kingdom, for instance, hazardous drinking is thought to start at 21 units/week for men and 16 units/week for women;[5] and in the United States the equivalent guidelines are 14 and 7 drinks per week.[6] It is among alcohol users who exceed such guidelines that the prevalence of dependence is the highest; up to 40 per cent of the more frequent violators.[7]

The expression 'alcohol problems' encompasses a wide range of untoward occurrences, from maladaptive, impaired, or harmful behaviour, to health complications and the condition of alcohol dependence. Alcohol problems are not incurred just by chronic excessive drinkers, but also by persons who drink heavily on isolated occasions (e.g. accidents, violence, poisoning, etc.). Given their

high frequency and social costs, these consequences of acute inebriation represent the most significant public health burden of drinking.[8] This section focuses rather on the causes of problems of a clinical nature, the ones presented by individuals who engage in patterns of repeated excessive drinking, i.e. 'alcohol dependence' and 'alcohol abuse' (DSM-IV nomenclature) or 'harmful drinking' (ICD-10 nomenclature).

The causality of alcohol misuse

Alcoholism is a bio-psychosocial phenomenon par excellence; it results from the contribution of multiple individual and environmental risk factors. The complex dynamics influencing its development have been well acknowledged in the literature. Theories have taken many disparate facts into consideration, from the effects of alcohol policy to the influence of familial and sociocultural environments across cultures and over time. Some ethnic groups, for instance, have traditionally had low rates of alcoholism (Asians, Jews, and some North American Aboriginals) and the prevalence is generally higher in males across both age cohorts and ethnicities. Another layer of complexity lies in the fact that alcoholism is a clinically heterogeneous disorder with variable age of onset, drinking patterns, severity, and comorbidity with other mental disorders. In general, alcoholics have one or more clinical diagnoses in addition to alcohol dependence, including drug abuse, antisocial personality disorder, anxiety, and depression. The course of the disorder is variable with high rates of remission and relapse; its manifestation changing in pattern and severity in response to life events (stressors) and other aspects of the environment. A summary of the etiological factors that have been shown to influence the development of alcoholism is shown in Box 4.2.2.1.1.

Box 4.2.2.1.1 A bio-psychosocial model of the aetiology of alcoholism

The comunity/sociocultural environment

- policies affecting availability and price (temperance, prohibition, taxation)
- cultural patterns of consumption, social acceptability of drinking/drunkenness
- availability of other reinforcers (sources of pleasure/recreation)
- employment and/or educational opportunities (anomie/marginalization)
- peer influences/role models (affiliation with deviant subculture)

The family environment

- marital breakdown (lower socio-economic status, poor parental monitoring)
- family attitudes (availability of drugs/alcohol, modelling of siblings/parents)
- intrauterine exposure to alcohol/drugs (potential effects of alcohol, nicotine and other drugs on behaviour and cognition)
- familial substance abuse (poverty, violence, and increased rates of early life trauma including neglect and physical/sexual abuse)

Individual/host factors

- heritable genetic factors (genetic loading for alcohol/drug dependence, depression, anxiety disorders in first-degree relatives)
- differences in response to alcohol (low sensitivity in terms of physiological responses and subjective effects)
- metabolic differences (thiamine deficiency, alcohol metabolizing enzymes)
- high risk taking behaviours (male sex)
- childhood psychopathology (conduct disorder, untreated ADHD)
- psychiatric disorders (bipolar, depression, anxiety, schizophrenia)

Sociocultural factors

Macrocultural influences such as values, beliefs, and mores; social role functions; local economy; customs and dietary habits; rapid social change; and cultural stress do shape and dictate the way alcohol is used in human societies. But even within a single society, there is variance in the alcohol problems profile of different subgroups. For instance, drinking, heavy drinking, alcohol use disorders, and treatment for alcoholism are more frequently recorded in men than women,[1] the risk of hospital admission for alcoholic psychosis, acute intoxication, and liver cirrhosis is elevated in unskilled and blue-collar workers when compared with higher occupational categories; alcoholics are over-represented in occupations with flexible work schedules, in those less supervised, and in the ones which facilitate access to alcohol,[9,10] and although there are a larger proportion of regular alcohol users among the older, the wealthier, and the better educated, frequency of heavy drinking (i.e. episodes of intoxication, 5+ drinks at a time) is inversely correlated with age, income, and level of education.[3]

Cultural beliefs about drinking and related social norms largely determine the manner in which alcohol is used. Disorderly conduct and drunken violence are more likely to occur in societies which, while allowing drinking, do view alcohol as an evil substance. Similar consequences can be expected if drunkenness is culturally considered as a 'time out', when socially unacceptable behaviours are tolerated or excused.[11] In fact, the social condoning of drunkenness is considered as an epidemiological risk factor. The availability of alcohol and the social promotion of frequent or heavy drinking are examples of social risk. But environmental facilitation *per se* does not explain the genesis of an alcohol dependence disorder in specific individuals. This disorder is best understood as the result of social prompting and individual vulnerabilities.

Psychological factors

Alcoholics do not present a homogeneous premorbid personality profile. However, some distinctive trait clusters have been identified which seem to characterize different types of alcoholics.[12] One such group (type 1) tend to score low in novelty seeking and high in harm avoidance and reward dependence. Another group (type 2) is formed by the natural thrill seekers, who appear to ignore harmful consequences and punitive responses. This latter cluster, which prevails mostly in males with early-onset alcoholism,

is also typical of antisocial personalities. Of all personality features, conduct disorder and antisocial behaviour are the strongest predictors of alcohol misuse.[13]

(a) Psychodynamic processes

Early psychodynamic writings portrayed alcoholism and other addictions as regressive behaviours caused by unconscious conflicts over libidinal pleasures, homosexuality, and aggression. More recent formulations emphasize ego and self-developmental problems, and consider psychoactive substance abuse as a response to psychological suffering; an attempt at re-establishing homeostasis. This is known as the *self-medication hypothesis* of addictions,[14] according to which, persons with self-regulatory deficiencies in the areas of self-care, self-esteem, self-object relations, and affect tolerance, would drink to palliate their distress.

(b) Learning

Alcohol abuse as seen by some as a behavioural pattern which has been learned through mechanisms of classical (i.e. Pavlovian) and operant conditioning. According to this interpretation, the perpetuation of heavy drinking results from its association with conditioned stimuli (cues), and from the action of positive (pleasant effects) or negative (stress reduction) behavioural reinforcement. Additional components of this equation are the so-called alcohol 'expectancies'. Alcohol abusers tend to overemphasize the pleasant aspects of drinking and to ignore the negative ones; the learning theory of alcoholism assumes that such a cognitive set is also acquired through social exposure. The Social Learning Theory posits that the positive expectancy of relaxation following a drink can facilitate more frequent alcohol use and thus contribute to the development of dependence.[15]

(c) Psychiatric comorbidity

Community and clinical epidemiology findings point to the presence of other psychiatric disorders as one of the most significant psychological risk factors in alcoholism.[16] The co-occurrence is sometimes sequential, with the psychiatric disorder preceding alcoholism; in which case a causal role in the development of heavy/frequent drinking is attributed to the former. While this is often observed in cases of conduct disorder, social phobia, attention deficit-hyperactivity (untreated) and depression, there are other psychiatric disturbances such as panic disorder, generalized anxiety and dysthymia that often become clinically significant only after the person has been abusing alcohol for sometime. These *alcohol-induced* mood and anxiety disorders represent a sizeable proportion of the comorbidity rates.[17] Whether or not it is 'primary', psychological stress is a widely recognized factor in alcoholism treatment failure and relapse.

Yet the comorbidity of some psychiatric disorders (e.g. bipolar disorder, schizophrenia) and alcoholism appears to develop in no predictable sequence, so that if not random, their co-occurrence could be assumed to result from common genetic influences (see below) and pathophysiological mechanisms. One such interpretation is the 'reward deficiency syndrome' hypothesis; it purports that both psychiatric disturbances (e.g. negative symptoms of schizophrenia) and the tendency to abuse addictive substances arise from a basic dysfunction of the dopamine mesocorticolimbic reward system.[18] The 'primary addiction' theory is another such explanation for comorbidity; it contends that a single neurobiological deficiency—primary abnormalities in the hippocampus and the frontal cortex—facilitate the development of schizophrenic symptoms and the person's toxicophilia in a parallel manner.[19]

Genetic factors in the development of alcoholism

In recent decades the biological perspective on the aetiology of alcoholism has gained considerable ground. Findings from family, twin and adoption studies demonstrate that there is significantly higher risk for alcoholism among individuals with an alcoholic biological parent or first-degree relative.[20–23] Meta-analysis has been used to jointly analyze data from twin and adoption studies grouped by country of origin (Scandinavian versus United States of America). Based on all available data, genetic factors accounted for between 40 and 60 per cent of the variance in alcoholism risk, with the effects of environment (shared and non-shared) estimated between 15 and 33 per cent. In a methodologically rigorous study, Prescott and Kendler[24] examined the concordance for alcoholism among a population-based sample from the Virginia Twin Registry. Monozygotic (MZ, $n = 861$) and dizygotic (DZ, $n = 653$) male twins were diagnosed using structured interviews and DSM criteria. Concordance rates for alcohol dependence were significantly higher for MZ (48 per cent) compared to DZ (32 per cent) twins, and analyses indicated that 48–58 per cent of the variation in alcoholism liability could be attributed to additive genetic factors.

Alcoholic males with family history of alcoholism (FHP) have been reported to have greater severity of alcoholic symptoms and poorer outcomes than alcoholics that are family history negative (FHN). Box 4.2.2.1.2 describes some characteristics of familial alcoholism. Onset of drinking prior to age 15 is associated higher rates of alcoholism,[25] ADHD, conduct and anxiety disorders, as well as a host of other negative events including unintentional injuries, physical fights, nicotine/drug dependence, and poor school performance. Children of alcoholics are significantly more likely to be exposed to high-risk environments that include poor prenatal care (alcohol/nicotine exposure, nutritional deficiencies), as well as homes in which there is more poverty and violence. Overall, it

Box 4.2.2.1.2 Characteristics of familial alcoholism

Family history positive (FHP) alcoholism is associated with:

- Earlier onset (<15 at age first drink is associated with increased rates of alcoholism, nicotine dependence, drug use, and conduct disorder. Early age of alcohol use is familial, heritable and may be related to transmission of disinhibitory psychopathology in males)
- Behavioural disturbances during childhood (conduct disorder, emotional lability, aggressivity, low attention span, low soothability)
- More severe alcohol dependence (higher levels of physical dependence, negative consequences)
- Lower educational and occupational achievement
- Deficits in executive cognitive functioning (poor problem solving, abstraction, and perceptual-motor skills)

appears likely that there are common genetic and environmental influences on a host of externalizing disorders—as well as gene–environment interactions.

Linkage studies to identify the genes underlying the heritability of alcoholism

A number of large-scale international linkage studies are currently underway that are aimed at mapping genes for alcoholism including the Irish Affected Sib Pair Study,[26] and the Collaborative Study on the Genetics of Alcoholism (COGA).[27] COGA is a multi-center program designed to detect and map susceptibility genes for alcoholism that is currently underway in the United States. Using a family-based linkage strategy, the study is examining a number of quantitative intermediary phenotypes (endophenotypes) including P300 evoked potentials, alcohol sensitivity, and personality traits (harm avoidance, novelty seeking, and reward dependence) in relation to both alcohol consumption, and alcohol dependence. In addition, the study is examining the association between polymorphisms in specific candidate genes such as alcohol dehydrogenase (ADH), monoamine oxidase (MAO_B), and the serotonin transporter and alcohol-related phenotypes. In early work, COGA-reported associations between alcoholism and regions on chromosomes 4 and 15 that encode genes for the inhibitory neurotransmitter, gamma-aminobutyric acid ($GABA_A$). Most recently, linkage and association genome scans for a broader 'addiction' vulnerability phenotype provided strong evidence for linkage to chromosome 4. Further assessment of single nucleotide polymorphism (SNP) genotypes within the chromosome 4 region provided strongest support for the involvement of the $GABA_A$ receptor $\alpha 2$ subunit (GABRA2 gene).[28] $GABA_A$ has been implicated in mediating some of the psychopharmacological effects of alcohol,[29] and the genetic studies provide convergent evidence suggesting that the predisposition to alcoholism may be inherited as a general state of CNS disinhibition/hyperexcitability that results from an altered responsiveness to GABA. However, this remains to be confirmed by additional genetic and experimental studies.

Other candidate genes and processes

It has been shown that genetic factors may influence a number of important processes such as initial sensitivity to the effects of alcohol, as well as the development of tolerance, sensitization, and physical dependence (including withdrawal complications such as seizures and delirium tremens). Several lines of research have suggested that sensitivity to alcohol may influence the propensity to abuse. Sensitivity refers to drug effects such as intoxication, physiological reactivity, and activation (tendency towards stimulation versus depression following ingestion). For example, Schuckit[30,31] found that individuals with low sensitivity to alcohol as measured by lower psychomotor responses and less subjective intoxication following alcohol dosing were *more* likely to be alcohol dependent at follow-up 10 years later.

Peripheral and central levels of alcohol metabolizing enzymes may be important modulators of the psychopharmacological response to alcohol. Ethyl alcohol (ethanol) is converted to acetaldehyde via the actions of alcohol dehydrogenase (ADH). There is evidence for linkage of gene(s) located on chromosome 4 (as discussed above) and two ADH genes closely linked on chromosome 4 (ADH1B and ADH1C) that encode for isozymes that differ in their kinetic properties. The allele ADH1B*2 (found largely in individuals of East Asian and Jewish descent) encodes a more active isozyme that has been associated with protection from alcohol dependence. Most recently Edenberg *et al.*[32] genotyped 110 SNPs across seven ADH genes in a COGA sample. There was strong evidence that variations in ADH4 were associated with alcoholism, and among African-Americans there was evidence that the ADH1B*3 allele was protective. Acetaldehyde produced by ethanol oxidation is rapidly metabolized by the enzyme aldehyde dehydrogenase (ALDH). A single base pair substitution in mitochondrial ALDH, termed the 'oriental' mutation (ALDH2*2 allele), is present in a large percentage of the Asian population. This mutation renders the enzyme inactive and produces a flushing response (warm-flushed face, tachycardia, nausea) following ingestion of small quantities of alcohol due to the buildup of acetaldehyde, particularly among ALDH2*2 homozygotes. Due to the aversive nature of the flushing response, the ALDH2*2 mutation is a significant protective factor against alcoholism.

In addition to the examination of metabolic factors that may account for some of the genetic variance in the development of alcohol dependence, there is an intense search for other neurogenetic factors related to the effects of alcohol in the brain. As shown in Box 4.2.2.1.3, a large number of candidate markers and putative genetic loci that have been investigated to date. For example, abstinent alcoholics and approximately 35 per cent of sons of alcoholics have been shown to have lower amplitude of a P300 event-related brain potential. Analyses of COGA data indicate that P300 amplitude reduction (P3-AR) is heritable, but more recent analyses have demonstrated that the P3-AR is associated with risk for substance dependence generally (e.g. frequent use of cannabis).[33]

The analysis of various neurotransmitters (including synthesis, release, receptor density, second messengers, polymorphisms) in relation to alcoholism and other alcohol-related endophenotypes is a well-developed area of research. The high degree of comorbidity between alcoholism and other mental disorders suggests that there may be common neurobiological pathways, including those that

Box 4.2.2.1.3 Potential candidate genes and markers for alcoholism

- Brain waves (P300 event-related brain potential)
- Brain enzymes (e.g. monoamine oxidase, adenylate cyclase)
- Alcohol and aldehyde metabolizing enzymes (ADH, catalase, ALDH, cytochrome P450IIE1)
- Opioids (e.g. kappa OPRK1receptor and prodynorphin ligand)
- Serotonin (e.g. polymorphisms of the 5-HT transporter and receptors (e.g. 5-HT1B, 5-HT2A, 5-HT2C), tryptophan hydroxylase TPH (218AC))
- Dopamine (polymorphisms of D_2, D_3, D_4 receptors and the dopamine transporter (DAT))
- GABA (polymorphisms in receptor subunits, variants in glutamate decarboxylase-2 (*GAD2*))

Note that this list is not exhaustive. For a more complete review consult references 32–5.

modulate reward, compulsive behaviour, anxiety, depression, and stress responses.[34] In this context, dysregulation of the serotonin (5HT) system has been implicated in the aetiology of a number of psychiatric disorders (depression, OCD, eating disorders) and alcoholism. In particular, polymorphisms in the promoter region of the 5HT transporter (5HTTLPR) producing the short ('S' allele) or long ('L' allele) variants differentially modulate transcriptional activity of the promoter, yielding differences in 5HT uptake activity in human platelets and brain. Most recently analyses conducted in the COGA sample have failed to find an association between the 5HTTLPR polymorphism and alcohol dependence. In a family-based association analyses (n = 1913 Caucasians) there was evidence for association of the S allele with depression, but not with alcohol dependence.[27]

Numerous studies have examined the association between alcohol dependence and the A1 allele of the dopamine D2 receptor (DRD2), however results have been debated for more than a decade. In general, the A1 allele is not consistently associated with alcoholism, and it does not consistently co-segregate in families with alcoholism. The effect size of this allele is likely to be very small. The human genes for the dopamine D3 and D4 receptors are polymorphic and studies are currently underway examining the potential relationship between various alleles of these receptors and substance dependence.

As noted above, synaptic actions of GABA have been implicated in the psychopharmacological effects of alcohol. Associations between variants in glutamate decarboxylase-2 (*GAD2*), a gene encoding for a major enzyme in the synthesis of GABA have been reported. In particular a functional promoter *GAD2* -243 A > G variant may influence risk for alcohol dependence in populations exhibiting severe alcoholism.[35]

In summary, plausible candidate genes for alcoholism include loci associated with alcohol and aldehyde metabolism, as well as variants within the GABA, opiate, and serotonin systems. The strongest candidate to date is for the involvement of the $GABA_A$ receptor $\alpha 2$ subunit (GABRA2 gene) on chromosome 4. Notably, associations between various loci and alcoholism reported in the literature have not been consistently replicated. Discrepancies in the literature have beesn attributed to variations in sampling (ethnicity, diagnostic criteria, severity of alcoholism, sample sizes), as well as to the clinical and genetic heterogeneity of alcoholism. Thus in this context, it is important to note that possible mechanisms for indirect transmission of an alcoholism phenotype include personality traits, and comorbid psychopathology including anxiety, depression, and conduct disorder.

Further information

Sartor, C.E., Lynskey, M.T., Heath, A.C., *et al.* (2006). The role of childhood risk factors in initiation of alcohol use and progression to alcohol dependence. *Addiction*, **102**, 216–25.

References

1. Edwards, G., Anderson, P., Babor, T.F., *et al.* (1994). *Alcohol policy and the public good.* Oxford University Press, Oxford, UK.
2. Newes-Adeyi, G., Chen, C.M., Williams, G.D., *et al.* (2003). *Trends in underage drinking in the United States, 1991–2003. Surveillance Report No. 74.* National Institute on Alcohol Abuse and Alcoholism, Bethesda, MD.
3. Health Canada and the Canadian Executive Council on Addictions. (2004). *Canadian addiction survey*. Canadian Centre on Substance Abuse, Ottawa.
4. Anthony, J.C., Chen, C.Y., and Storr, C.L. (2005). Drug dependence epidemiology. *Clinical Neuroscience Research*, **5**, 55–68.
5. Royal College of General Practitioners. (1986). *Alcohol-a balanced view*. Royal College of General Practitioners, London.
6. National Institute on Alcohol Abuse and Alcoholism. (2004). *Helping patients with alcohol problems: a health practitioner's guide*. NIAAA, Bethesda, MD.
7. Dawson, D.A., Grant, B.F., and Li, T.K. (2005). Quantifying the risks associated with exceeding recommended drinking limits. *Alcoholism: Clinical and Experimental Research*, **29**, 902–8.
8. Rhem, J., Baliunas, D., Brochu, S., *et al.* (2006). *The costs of substance abuse in Canada 2002*. Canadian Centre on Substance Abuse, Ottawa.
9. Hemmingsson, T., Lundberg, I., Romelsjo, A., *et al.* (1997). Alcoholism in social classes and occupations in Sweden. *International Journal of Epidemiology*, **26**, 584–91.
10. Mandell, W., Eaton, W.W., Anthony, J.C., *et al.* (1992). Alcoholism and occupations: a review of 104 occupations. *Alcohol: Clinical and Experimental Research*, **16**, 734–46.
11. Heath, D.B. (1993). Recent developments in alcoholism: anthropology. *Recent Developments in Alcoholism*, **11**, 29–43.
12. Cloninger, C.R. (1987). Neurogenetic adaptive mechanisms in alcoholism. *Science*, **236**, 410–16.
13. Driessen, M., Veltrup, C., Wetterling, T., *et al.* (1998). Axis I and Axis II comorbidity in alcohol dependence and the two types of alcoholism. *Alcoholism: Clinical and Experimental Research*, **22**, 77–86.
14. Khantzian, E.J. (1985). The self-medication hypothesis of addictive disorders. *The American Journal of Psychiatry*, **142**, 1259–64.
15. Rotgers, F., Keller, D.S., and Morgenstern, J. (eds.) (1996). *Treating substance abuse: theory and technique*. Guilford Press, New York.
16. Kessler, R.C., McConagle, K.A., Zhao, S., *et al.* (1994). Lifetime and 12 month prevalence of DSM-III-R psychiatric disorders in the United States. *Archives of General Psychiatry*, **51**, 8–19.
17. Schuckit, M.A. (2006). Comorbidity between substance use disorders and psychiatric conditions. *Addiction*, **101**(Suppl. 1), 76–88.
18. Bowirrat, A. and Oscar-Berman, M. (2005). Relationship between dopaminergic neurotransmission, alcoholism and reward deficiency syndrome. *American Journal of Medical Genetics B Neuropsychiatric Genetics*, **132**, 29–37.
19. Chambers, R.A., Krystal, J.H., and Self, D.W. (2001). A neurobiological basis for substance abuse comorbidity in schizophrenia. *Biological Psychiatry*, **50**, 71–83.
20. Cotton, N.S. (1979). The familial incidence of alcoholism: a review. *Journal of Studies on Alcoholism*, **40**, 89–116.
21. Cloninger, C.R., Bohman, M., and Sigvardsson, S. (1981). Inheritance of alcohol abuse: cross-fostering analysis of adopted men. *Archives General Psychiatry*, **38**, 861–8.
22. Van den Bree, M.B.M., Johnson, E.O., Neale, M.C., *et al.* (1998). Genetic and environmental influences on drug use and abuse/dependence in male and female twins. *Drug and Alcohol Dependence*, **52**, 231–41.
23. Edenberg, H.J. and Foroud, T. (2006). The genetics of alcoholism: identifying specific genes through family studies. *Addiction Biology*, **11**, 386–96.
24. Prescott, C.A. and Kendler, K.S. (1999). Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. *The American Journal of Psychiatry*, **156**, 34–40.
25. Grant, B.F. and Dawson, D. (1997). Age at onset of alcohol use and its association with DSM-IV alcohol abuse and dependence: results from the national longitudinal alcohol epidemiologic survey. *Journal of Substance Abuse*, **9**, 103–10.

26. Kendler, K.S., Kuo, P.H., Todd Webb, B., *et al.* (2006). A joint genomewide linkage analysis of symptoms of alcohol dependence and conduct disorder. *Alcoholism: Clinical and Experimental Research*, **30**, 1972–7.
27. Dick, D.M., Plunkett, J., Hamlin, D., *et al.* (2007). Association analyses of the serotonin transporter gene with lifetime depression and alcohol dependence in the Collaborative Study on the Genetics of Alcoholism (COGA) sample. *Psychiatric Genetics*, **17**, 35–8.
28. Johnson, C., Drgon, T., Liu, Q.R., *et al.* (2006). Pooled association genome scanning for alcohol dependence using 104,268 SNPs: validation and use to identify alcoholism vulnerability loci in unrelated individuals from the collaborative study on the genetics of alcoholism. *American Journal of Medical Genetics: Part B Neuropsychiatric Genetics*, **141**, 844–53.
29. Krystal, J.H., Staley, J., Mason, G., *et al.* (2006). γ-Aminobutyric acid type A receptors and alcoholism. *Archives General Psychiatry*, **63**, 957–68.
30. Schuckit, M.A. and Smith, T.L. (1997). Assessing the risk for alcoholism among sons of alcoholics. *Journal of Studies on Alcohol*, **58**, 141–5.
31. Schuckit, M.A. (1994). Low level of response to alcohol as a predictor of future alcoholism. *The American Journal of Psychiatry*, **151**, 184–9.
32. Edenberg, H.J., Xuei, X., Chen, H.J., *et al.* (2006). Association of alcohol dehydrogenase genes with alcohol dependence: a comprehensive analysis. *Human Molecular Genetics*, **15**, 1539–49.
33. Yoon, H.H., Iacono, W.G., Malone, S.M., *et al.* (2006). Using the brain P300 response to identify novel phenotypes reflecting genetic vulnerability for adolescent substance misuse. *Addictive Behaviors*, **31**, 1067–87.
34. Lesch, K.P. (2005). Alcohol dependence and gene x environment interaction in emotion regulation: is serotonin the link? *European Journal of Pharmacology*, **526**, 113–24.
35. Lappalainen, J., Krupitsky, E., Kranzler, H.R., *et al.* (2006). Mutation screen of the *GAD2* gene and association study of alcoholism in three populations. *American Journal of Medical Genetics: Part B Neuropsychiatric Genetics*, **144**, 183–92.

4.2.2.2 Alcohol dependence and alcohol problems

Jane Marshall

Introduction

The problem of excessive alcohol consumption is a major cause of public health concern in most countries of the world today. Heavy consumption, which involves far more than 'dependence', can cause untold misery to the individual, who is usually affected by other physical, psychological, and social disabilities as well.

As early as 1950, the World Health Organization (WHO) viewed the lack of a commonly accepted terminology as a serious obstacle to international action in the alcohol field.[1]

Definitions of 'alcoholism' have been proposed by a range of professional and other bodies, from biomedical scientists, medical doctors and psychiatrists, psychologists, sociologists, patients in treatment, to the general public.[2] Terms such as 'alcoholism', 'addiction', and 'chemical dependence', have passed into everyday speech, becoming 'popularly enriched' and 'technically impoverished'.[2] These terms mean different things to different people and often have pejorative connotations. The lack of a precise definition of 'drinking problems' has hampered interdisciplinary communication.

In this section, the evolution of the term 'alcohol dependence' will be traced and put into context as but one aspect of a wider spectrum of alcohol-related problems. The concept of the alcohol dependence syndrome (ADS)[3] will be introduced and its influence on the 10th revision of the *International Classification of Diseases* (ICD-10)[4] and the fourth edition of the *Diagnostic and Statistical Manual of Diseases* (DSM-IV) will be reviewed.[5] The terms 'harmful use' (ICD-10) and 'alcohol abuse' (DSM-IV) will also be discussed. Finally 'alcohol-related problems' will be considered.

The development of classification systems for alcohol use disorders

From the time of its inception in 1948, WHO played a major role in formulating public health definitions of 'alcoholism', 'addiction', and 'dependence' through a series of expert committees. Early definitions stressed the sociological rather than the physical aspects of dependence, and thus had limited utility for biological research and psychiatric classification.

'Alcoholism' was classified under 'Other non-psychotic mental disorders' in ICD-8.[6] This definition of 'alcoholism' was generic, and included the subcategories of episodic excessive drinking, habitual excessive drinking, and alcohol addiction. Alcohol addiction was defined as:[6]

> a state of physical and emotional dependence on regular or periodic, heavy, and uncontrolled alcohol consumption, during which the person experiences a compulsion to drink. On cessation of alcohol intake there are withdrawal symptoms, which may be severe.

In ICD-9 the term 'alcoholism' was dropped in favour of the 'alcohol dependence syndrome'.[7] It was, however, still classified under the category 'Other non-psychotic mental disorders'.

At the same time as WHO was formulating public health definitions of 'alcoholism', 'addiction', and 'dependence', a trend towards formal diagnostic criteria was emerging in the United States. This was driven by practical considerations such as the need for better communication between clinicians, researchers, and the general public. Other influential factors included the growing need to categorize persons in an objective fashion for legal, medical, or psychiatric reasons, to collect and communicate accurate public health information, and to standardize practice nationally and internationally. The first two editions of the *Diagnostic and Statistical Manual (DSM-I and DSM-II)*, published in 1952 and 1968 respectively, classified 'alcoholism' as a category of personality disorder. In DSM-III,[8] it was included under a new and separate category of 'Substance use disorders'. The terms 'alcoholism' and 'addiction' were dropped and the terms 'dependence' and 'alcohol abuse' were used instead. Dependence was distinguished from abuse by the presence of tolerance or withdrawal symptoms.

By the mid-1980s, DSM-III and ICD-9 were undergoing reviews for the purposes of revision. The diagnostic criteria for dependence were broadened in DSM-IIIR[9] to incorporate the elements of the alcohol dependence syndrome as hypothesized by Edwards and Gross.[3] The essential feature of the DSM-IIIR dependence category was defined in the text as a 'cluster of cognitive, behavioural, and physiological symptoms, indicating that the person has impaired control over drinking and continues to drink despite adverse consequences'.

The alcohol dependence syndrome

Clinical description

In 1976, Edwards and Gross proposed the existence of alcohol dependence within a syndrome model.[3] Their description was based on the clinical observation that certain heavy drinkers manifested an interrelated clustering of signs and symptoms. They hypothesized that dependence was not an all-or-nothing phenomenon but existed in degrees of severity. The elements of the syndrome, as originally formulated, are summarized in Table 4.2.2.2.1. Not all the elements need always be present, nor always present with the same intensity. Edwards and Gross[3] also acknowledged the fact that not everyone who drinks too much is necessarily dependent on alcohol. They hypothesized that alcohol dependence should be conceptually distinguished from alcohol-related problems.

By drawing a clear distinction between the alcohol dependence syndrome and alcohol-related problems, Edwards and Gross introduced the concept of a bi-axial model. This was described further in the report of a WHO scientific group published in 1977.[10] Alcohol-related problems are defined as comprising those physical, psychological, and social problems that are a consequence of excessive drinking and dependence. Consumption may be viewed on a third axis.

The alcohol dependence syndrome was proposed in the first instance as an empirical formulation that would require research to confirm its assumptions. Unlike previous models of 'alcoholism' that had observational elements but no theoretical input, the alcohol dependence syndrome was influenced by psychological theory and proposed as a synthesis of both general learning theory and specific conditioning models of dependence.[11,12]

Establishment of the validity of the alcohol dependence syndrome

A considerable amount of scientific research evaluating the ADS has been carried out over the past 30 years, much of it supporting its validity.[13] Studies have focused on the degree to which the elements of the syndrome co-occur.[14,15] Other areas of research have included construct validity,[16] concurrent validity,[15,17,18] and predictive validity.[19,20] Field trials conducted as background to the preparation of ICD-10, DSM-IIIR, and DSM-IV, have all contributed to the body of research evidence.[5, 21–25] Difficulties have been encountered in operationalizing elements such as narrowing of repertoire, subjective change, and reinstatement.[25]

These studies have shown a remarkable similarity in terms of the coherence and dimensionality of the syndrome, and are particularly impressive because of the diversity of methods and populations used.[11]

Table 4.2.2.2.1 Key elements of the alcohol dependence syndrome

Narrowing of repertoire
Salience of drinking
Increased tolerance to alcohol
Withdrawal symptoms
Relief or avoidance of withdrawal symptoms by further drinking
Subjective awareness of compulsion to drink
Reinstatement after abstinence

Individual elements of the alcohol dependence syndrome

(a) Narrowing of the drinking repertoire

Most drinkers vary their alcohol consumption from day to day and week to week. The pattern of their drinking is influenced by a range of internal cues and external circumstances. Heavy drinkers may initially widen their drinking repertoire. As dependence advances, so a diminished variability in drinking behaviour emerges. The dependent person begins to drink in the same manner every day. The daily pattern established ensures that a relatively high blood-alcohol level is maintained and that symptoms of alcohol withdrawal are avoided. As drinking becomes stereotyped with advanced dependence, dependent drinkers are able to describe their drinking day in minute detail.

(b) Salience of drinking-seeking behaviour

With advancing dependence, individuals give priority to maintaining their alcohol intake. Alcohol consumption is maintained despite painful direct consequences such as physical illness, rejection by family, and lack of money. They will 'beg, borrow, or steal' to obtain money for alcohol.[3]

(c) Increased tolerance to alcohol

Regular drinkers become tolerant to the central nervous system effects of alcohol and can sustain blood alcohol levels that would incapacitate the non-tolerant drinker. In short, they can 'drink others under the table'. Tolerance may decrease in the later stages of dependence, with individuals becoming intoxicated on much less alcohol than would previously have affected them. Cross-tolerance extends to other drugs, notably barbiturates and benzodiazepines.

(d) Withdrawal symptoms

The term 'alcohol withdrawal' describes a broad range of symptoms and signs, from the relatively trivial to the life-threatening. At first the symptoms are intermittent and mild, but as the degree of dependence increases, so do the frequency and intensity of withdrawal symptoms. Symptoms vary from person to person and do not require abstinence to appear; they can occur when blood-alcohol concentrations are falling. When the picture is fully developed, the dependent drinker typically has severe multiple symptoms every morning on waking; these symptoms may wake him in the middle of the night. Those who are severely dependent usually experience mild withdrawal symptoms during the day whenever their alcohol levels fall.

Withdrawal symptoms cannot occur without a high degree of central nervous system tolerance, but tolerance can occur without clinically manifest withdrawal symptoms.[3]

The spectrum of symptoms is wide, but the four key symptoms are tremor, nausea, sweating, and mood disturbance. A range of other symptoms can also occur, including sensitivity to sound (hyperacusis), ringing in the ears (tinnitus), itching, muscle cramps, sleep disturbance, perceptual distortion, hallucinations, generalized (grand mal) seizures, and delirium tremens.

The four key symptoms will be described in further detail.

(i) Tremor

The first experience of alcohol withdrawal tremor may be recalled vividly: 'One afternoon I went to cut the grass at a friend's house. She gave me a cup of tea and my hands kept shaking. I kept rattling

the cup on the saucer and couldn't put the cup to my mouth. I had to put them down and pretend that I had finished.' Men often find it difficult to shave first thing in the morning and merely getting the first drink of the day to the mouth may be an ordeal in itself.

(ii) Nausea

Dependent drinkers commonly say that their bodies want to vomit first thing in the morning, but that they have nothing to bring up. This may be described as 'dry retching' or 'the dry heaves'. Typically they find it difficult to eat breakfast and to brush their teeth. The first drink of the day is often vomited back.

(iii) Sweating

Dependent drinkers commonly describe waking up in the early morning (3 a.m. or 4 a.m.) to find the bed sheets 'drenched'. In the earlier stages of dependence they may report feeling clammy.

(iv) Mood disturbance

This is an important feature of the withdrawal syndrome. Mildly dependent individuals may feel 'a bit edgy'. Severely dependent individuals may present with clinically significant symptoms of anxiety and depression.

(e) Relief or avoidance of withdrawal symptoms by further drinking

In the early stages of dependence, individuals may find that they need a lunchtime drink to alleviate discomfort. As dependence progresses there emerges the need for an early morning drink to relieve the symptoms of alcohol withdrawal coming on after a night's abstinence. Later, individuals may wake in the middle of the night for a drink, and alcohol is often kept by the bed. If they have to go for 3 or 4 hr without a drink during the day, they value the next drink for its relief effect.

Clues to the degree of dependence can be obtained by taking a detailed history of the first drink of the day. The person drinking from a bottle kept by the side of the bed before they get up is more dependent than the person who has breakfast and reads the paper first. The woman who pours whisky into her first cup of tea is more dependent than the librarian who slips out to the lavatory at midday to drink from a quarter bottle of vodka hidden in her handbag.

(f) Subjective awareness of compulsion to drink

This describes an altered subjective experience of an inability to limit drinking to an acceptable level. Although the familiar term 'loss of control' has been used to denote this element, it is more likely that control has been 'impaired' rather than lost.

Another complex experience is that of 'craving', the subjective experience of which is greatly influenced by environment. Individuals can experience craving of very different intensities on different occasions. Cues for craving include the experience of intoxication, the withdrawal syndrome, mood (anger, depression, elation), or situational cues (being in a pub or (bar), passing an off-licence (liquor store).

Here the key experience may best be described as a compulsion to drink. The desire for a further drink is seen as irrational, and is resisted, but despite this a further drink is taken.

(g) Reinstatement after abstinence

Alcohol dependent individuals who begin to drink again after a period of abstinence invariably relapse back into the previous stage of the dependence syndrome. This process occurs over a variable time course, with moderately dependent individuals perhaps taking weeks or months and severely dependent individuals taking a couple of days.

Influence of the alcohol dependence syndrome on ICD-10 and DSM-IV

Both DSM-IV and ICD-10 diagnostic approaches have drawn on the original concept of the alcohol dependence syndrome.[26,27] Although they have undoubtedly contributed to the standardization of psychiatric practice nationally and internationally, they picture dependence as an all-or-nothing phenomenon rather than as a dimensional state.[28]

ICD-10[4]

ICD-10 includes six items under dependence, most of which are similar to DSM-IV. For a diagnosis of dependence, three or more items should have occurred in the past year. The 'strong desire or sense of compulsion to take the substance' is viewed as a central descriptive characteristic of dependence in ICD-10. This compulsive-use indicator is not included in the DSM-IV concept of dependence (Table 4.2.2.2.2).

DSM-IV[5]

In view of the major changes in criteria that had occurred between 1980 and 1987, the DSM-IV Substance Use Disorders Work Group was reluctant to make any additional major changes to DSM-IIIR. The repetitive nature of the problem was highlighted in that three or more of the items should have occurred during the same 12-month period and the associated difficulties must have led to clinically significant impairment or distress. DSM-IV also uniquely allows for the subtyping of dependence with and without physiological dependence (Table 4.2.2.2.2).

Alcohol abuse and harmful use

Alcohol abuse

DSM-III; DSM-IIIR; DSM-IV

The term 'alcohol abuse' appeared infrequently in the American literature before 1970, when the United States National Institute on Alcohol Abuse and Alcoholism was formed. It was adopted as a formal diagnostic category by DSM-III,[8] which defined abuse as a behavioural concept: 'A pattern of pathological use for at least a month that causes impairment in social or occupational functioning'. Although enshrined in DSM-IIIR and DSM-IV, the term 'abuse' has been variously regarded as 'unscientific and pejorative'[29] and 'oppobrious' and 'vindictive'.[30]

The DSM-IV Substance Use Disorders Workgroup carried out extensive analysis in an effort to define abuse more precisely. Accordingly, in DSM-IV, four separate items, not included in dependence, are listed for the diagnosis of abuse, focusing on social, physical, legal, and interpersonal problems associated with alcohol use. These problems must have occurred repeatedly over a 12-month period, and caused 'clinically significant impairment or distress' (Table 4.2.2.2.3). In practice, the DSM-IV alcohol abuse definition includes a mixture of dependence and harm criteria which could be scaled along a single continuum of severity of alcohol dependence.

Table 4.2.2.2.2 Comparison of ICD-10 and DSM-IV criteria for substance dependence

ICD-10	DSM-IV
A diagnosis of dependence should usually be made only if three or more of the following have been experienced or exhibited at some time during the previous year:	A maladaptive pattern of substance use, leading to clinically significant impairment or distress, as manifested by three (or more) of the following at any time in the same 12-month period
Evidence of tolerance such that increased doses of the psychoactive substance are required in order to achieve effects originally produced by lower doses	Tolerance as defined by either of the following: ◆ need for markedly increased amounts of the substance to achieve intoxication or desired effect ◆ markedly diminished effect with continued use of the same amount of substance
A physiological withdrawal state when substance use has ceased or been reduced, as evidenced by: ◆ the characteristic withdrawal syndrome for the substance or ◆ use of the same (or a closely related) substance with the intention of relieving or avoiding withdrawal symptoms	Withdrawal as manifested by either of the following: ◆ the characteristic withdrawal syndrome for the substance ◆ the same (or closely related) substance is taken to relieve or avoid withdrawal symptoms
A strong desire or sense of compulsion to take the substance	No equivalent criterion
No equivalent criterion	There is a persistent desire or unsuccessful efforts to cut down or control substance use
Difficulties in controlling substance-taking behaviour in terms of its onset, termination, or levels of use	The substance is often taken in larger amounts or over a longer period than was intended
Progressive neglect of alternative pleasures or interests because of psychoactive substance use	Important social, occupation or recreational activities are given up or reduced because of substance use
Increased amount of time necessary to obtain or take the substance or recover from its effects	A great deal of time is spent in activities necessary to obtain the substance, use the substance, or recover from its effects
Persisting with substance use despite clear evidence of overtly harmful consequences. Efforts should be made to determine that the user was actually, or could be expected to be, aware of the nature and extent of the harm	The substance use is continued despite knowledge of having a persistent or recurrent physical and psychological problem likely to have been caused or exacerbated by the substance. *Specify if:* *With Physiological Dependence:* evidence of tolerance or withdrawal (either item is present) *Without Physiological Dependence:* no evidence of tolerance or withdrawal

Reproduced from ICD-10 and DSM-IV:[4, 5]

Table 4.2.2.2.3 Comparison of criteria for abuse or harmful use of substances

ICD-10 criteria for harmful use

A pattern of psychoactive substance use that is causing damage to health, either physical or mental. The diagnosis requires that actual damage should have been caused to the mental or physical health of the user. Socially negative consequences, or the disapproval of others are not in themselves evidence of harmful use.
Harmful use should not be diagnosed if dependence syndrome, a psychotic disorder or another specific form of alcohol-related disorder is present.

DSM-IV criteria for substance abuse

A.
A maladaptive pattern of substance use leading to clinically significant impairment or distress by one (or more) of the following occurring within a 12-month period:
- Recurrent substance use resulting in a failure to fulfill major role obligations at work, school or home
- Recurrent substance use in situations in which it is physically hazardous
- Recurrent substance-related legal problems
- Continued substance use despite having persistent or recurrent social or interpersonal problems caused or exacerbated by the effects of the substance

B.
The symptoms have never met criteria for substance dependence

Reproduced from ICD-10 and DSM-IV:[4, 5]

Harmful use

(a) ICD-10

The ICD-10 criteria for harmful use of alcohol differ significantly from the DSM-IV abuse classification. An ICD-10 diagnosis of harmful drinking requires a pattern of drinking that has caused actual physical or psychological harm to the user. This definition excludes social harms such as marital problems and does not overlap with the DSM-IV definition of alcohol abuse.

The future

Revision of the DSM and ICD classification systems must address the fact that the current systems do not address the continuum of severity of alcohol use disorders (AUDs). Research is needed to explore the relationship between AUDs and the quantity, frequency, and pattern of drinking.[31] Further refinements of the alcohol dependence diagnosis should focus on the essential or core features of the disorder.

Alcohol-related problems

Not everyone experiencing an alcohol problem or alcohol-related disability will be suffering from alcohol dependence. Both dependent and non-dependent drinkers, particularly binge drinkers, are at risk of problems related to heavy alcohol consumption. Indeed, epidemiological evidence supports the view that most alcohol-related

harm in the general population occurs in heavy non-dependent drinkers.

Alcohol-related problems are extremely diverse. They have been defined as 'those problems that may arise in individuals around their use of beverage alcohol, and that may require an appropriate treatment response for their optimum management'.[32] The phrases 'alcohol problems' or 'alcohol-related problem' contain an assumption of causality.[33] This issue is a complex one, involving individual differences and the social context of drinking as well as the pattern, duration, and intensity of alcohol use.

Alcohol-related problems can be related to the acute or chronic consumption of alcohol. A fractured ankle sustained by falling over while acutely intoxicated is an example of the former category. Cirrhosis of the liver is an example of a chronic problem. An individual who drinks in binges will experience different problems compared with someone who drinks the same amount of alcohol spread out over a week or a month or a year. The way in which a person behaves while intoxicated is another important factor determining the nature of alcohol-related problems. The social consequences of drinking such as job loss, imprisonment, marital and family break-up, and drunk-driving have profound effects on the well being of the drinker, their family, and society.[33]

Types of alcohol-related problems

Although somewhat artificial, it is helpful to classify alcohol-related problems in individuals into physical, psychological, and social categories. There is often considerable overlap between these three areas. The more severe the dependence, the greater the likelihood of problems of all three kinds.[18]

Alcohol-related physical and psychological problems are discussed in the next section. Some of the social problems can be included here, for example the acute adverse consequences of drinking such as trauma resulting from road traffic accidents, injuries from fights, and death from overdose.[33]

The social problems that can result from drinking are legion. Alcohol is involved in all types of accidents and contributes to traffic deaths, home, and leisure injuries.[33] It is associated with domestic violence, child abuse, crime, homicide, and suicide and is also related to poor work performance, dismissal, unemployment, debt and housing problems, and crimes of violence.

There is a continuity between moderate and excessive drinking and between harmless drinking and drinking that results in harm or in problems. Such problem-clustering may reflect alcohol dependence, certainly amongst a proportion of these drinkers. Given this heterogeneity, no one form of treatment is likely to be effective for all individuals with alcohol problems.[32] A range of treatments is required and it should be possible for non-specialists to offer brief interventions (see Chapter 4.2.2.4).

The study of alcohol-related problems remains underdeveloped, compared with the study of alcohol dependence.[34] There may be several reasons for this, not least the difficulties inherent in measuring alcohol-related problems. Another important issue, central to these difficulties, is the extent to which alcohol is causally related to the problem.

Several questionnaires, measuring a variety of alcohol-related problems, have been developed. The Alcohol Problems Questionnaire (APQ)[34] is a standardized inventory, which includes 46 items covering eight problem domains: physical, psychological, friends, finances, police, marital, children, and work. All questions apply to the 6-month period prior to the completion of the questionnaire. The shorter or core version includes the first five domains (23 items). This questionnaire can make a useful contribution to the overall assessment, and is of potential value in outcome research.

Conclusions

An understanding of the concepts of alcohol dependence and alcohol-related problems is central to the therapeutic process with individual patients.

The development of diagnostic criteria has helped to standardize practice nationally and internationally, and aided interdisciplinary communication. The diagnostic criteria for dependence are imperfect because they view the syndrome as an all-or-nothing phenomenon rather than as a dimensional state. The concepts of abuse and harmful use need further refinement. The totality of alcohol problems is a vast area with major implications for the general population, not just dependent drinkers.

Further information

Reports and reviews

Raistrick D., Heather N., Godfrey, C (2006). *Review of the effectiveness of treatment for drinking problems*. London: National Treatment Agency for Substance Misuse. This review is eminently readable summary of the published international research literature on alcohol interventions and treatment. Chapter 2 describes the categories of alcohol misuse, and sets them in the context of treatment and interventions to reduce alcohol-related harm.

Room, R., Babor, T., Rehm, J. (2005). Alcohol and public health. *Lancet,* **365**, 510-530

Compilation of articles

Saunders, J. B.and Schuckit, M. A. (2006). Diagnostic issues in substance use disorders: refining the research agenda. *Addiction* (Suppl 1), **101**, 1–173. A series of 18 articles commissioned by the Substance Use Disorders Workgroup to inform the research agenda for the development of the Vth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V).

Marshall E. J, Farrell, M. Chapter editors. Addiction Psychiatry. *Psychiatry,* **5**, 421–63. Straightforward account of the classification and epidemiology of substance use disorders.

Book

Edwards, G. (2006). *Alcohol: the ambiguous molecule*. Harmondsworth, Penguin.

References

1. World Health Organization (1951). *WHO technical report series*, No 42. WHO, Geneva.
2. Babor, T.F. (1990). Social, scientific and medical issues in the definition of alcohol and drug dependence. In *The nature of drug dependence* (ed. G. Edwards and M. Lader), pp. 19–36. Oxford Medical Publications, Oxford.
3. Edwards, G. and Gross, M.M. (1976). Alcohol dependence: provisional description of a clinical syndrome. *British Medical Journal*, **1**, 1058–61.
4. World Health Organization (1992). *International statistical classification of diseases and related health problems, 10th revision*. WHO, Geneva.

5. American Psychiatric Association (1994). *Diagnostic and statistical classification of diseases and related health problems* (4th edn). American Psychiatric Association, Washington, DC.
6. World Health Organization (1974). *Glossary of mental disorders and guide to their classification: for use in conjunction with the International Classification of Diseases* (8th revision). WHO, Geneva.
7. World Health Organization (1978). *Mental disorders: glossary and guide to their classification in accordance with the Ninth Revision of the International Classification of Diseases*. WHO, Geneva.
8. American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders* (3rd edn). American Psychiatric Association, Washington, DC.
9. American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders* (3rd edn, revised). American Psychiatric Association, Washington, DC.
10. Edwards, G., Gross, M.M., Keller, M., Moser, J., and Room, R. (1977). *Alcohol–related disabilities*. WHO Offset Publication No. 32. WHO, Geneva.
11. Edwards, G. (1986). The alcohol dependence syndrome: a concept as stimulus to enquiry. *British Journal of Addiction*, **81**, 171–83.
12. Babor, T.F., Cooney, N.L., and Lauerman, R.J. (1987). The dependence syndrome concept as a psychological theory of relapse behaviour: an empirical evaluation of alcoholic and opiate addicts. *British Journal of Addiction*, **82**, 393–405.
13. Li, T.K., Hewitt, B.G., Grant, B.F (2007). The alcohol dependence syndrome, 30 years later: a commentary. *Addiction*, **102**, 1522–30.
14. Chick, J. (1980). Alcohol dependence: methodological issues in its measurement, reliability of the criteria. *British Journal of Addiction*, **75**, 175–86.
15. Stockwell, T., Murphy, D., and Hodgson, R. (1983). The severity of alcohol dependence questionnaire: its use, reliability and validity. *British Journal of Addiction*, **78**, 145–55.
16. Feingold, A. and Rounsaville, B. (1995). Construct validity of the dependence syndrome as measured by DSM–IV for different psycho–active substances. *Addiction*, **90**, 1661–9.
17. Kivlahan, D., Sher, K.J., and Donovan, D.M. (1989) The Alcohol Dependence Scale: a validation study among inpatient alcoholics. *Journal of Studies on Alcohol*, **50**, 170–5.
18. Caetano, R. (1993). The association between severity of DSM–III–R alcohol dependence and medical and social consequences. *Addiction*, **88**, 631–42.
19. Hodgson, R., Rankin, H.J., and Stockwell, T. (1979). Alcohol dependence and the priming effect. *Behaviour Research and Therapy*, **17**, 379–87.
20. Rankin, H., Stockwell, T., and Hodgson, R. (1982). Cues for drinking and degrees of alcohol dependence. *British Journal of Addiction*, **77**, 287–96.
21. Grant, B.F., Harfold, T.C., Chou, P., *et al.* (1992). DSM–III–R and the proposed DSM–IV alcohol use disorders. United States 1988. A methodological comparison. *Alcoholism, Clinical and Experimental Research*, **16**, 215–21.
22. Cottler, L.B. (1993). Comparing DSM–III–R and ICD–10 substance use disorders. *Addiction*, **88**, 689–96.
23. Rapaport, M.H., Tipp, J.E., and Schuckit, M.A. (1993). A comparison of ICD–10 and DSM–III–R criteria for substance abuse and dependence. *American Journal of Drug and Alcohol Abuse*, **19**, 143–51.
24. Rounsaville, B.J., Bryant, K., Babor, T., *et al.* (1993). Cross system agreement for substance use disorders: DSM–III–R, DSM–IV and ICD–10. *Addiction*, **88**, 337–48.
25. Cottler, L.B., Phelps, D.L., and Compton, W.M. (1995) Narrowing of the drinking repertoire criterion: should it have been dropped from ICD–10? *Journal of Studies on Alcohol*, **56**, 173–6.
26. Saunders, J.B. (2006). Substance dependence and non-dependence in the Diagnostic and Statistical Manual of Mental Disorders (DSM) and the International Classification of Diseases (ICD): can an identical conceptualization be achieved? *Addiction*,**101**, 48–58.
27. Grant, B. (1996). DSM–IV, DSM–III–R, ICD–10 alcohol and drug abuse/harmful use and dependence, United States, 1992: a nosological comparison. *Alcoholism, Clinical and Experimental Research*, **8**, 1481–8.
28. Edwards, G., Marshall, E.J., and Cook, C.C.H. (2003). *The treatment of drinking problems* (4th edn). Cambridge University Press.
29. Edwards, G., Arif, A., and Hodgson, R. (1981). Nomenclature and classification of drug– and alcohol–related problems: a WHO memorandum. *Bulletin of the World Health Organization*, **50**, 225–42.
30. Keller, M. (1982). On defining alcoholism: with comment on some other relevant words. In *Alcohol, science and society revisited* (ed. E.L. Gomberg, H.R. White, and J.A. Carpenter), pp. 119–33. University of Michigan Press, Ann Arbor, MI.
31. Hasin, D.S., Liu, X., Alderson, D., *et al.* (2006) DSM-IV alcohol dependence: a categorical or dimensional phenotype? *Psychological Medicine*, **36**, 1695–1705.
32. Institute of Medicine (1990). *Broadening the base of treatment for alcohol problems: report of a study by a Committee of the Institute of Medicine, Division of Mental Health and Behavioural Medicine*. National Academy Press, Washington, DC.
33. Babor, T.F., Caetano, R., Casswell, S., *et al.* ((2003). *Alcohol: No ordinary commodity. Research and Public Policy*. Oxford University Press.
34. Drummond, D.C. (1990). The relationship between alcohol dependence and alcohol–related problems in a clinical population. *British Journal of Addiction*, **85**, 357–66.

4.2.2.3 Alcohol and psychiatric and physical disorders

Karl F. Mann and Falk Kiefer

Intoxication

Clinical symptoms of alcohol intoxication are associated with both, blood alcohol concentration (**BAC**), and the individual's level of tolerance. Whereas in healthy persons without alcohol tolerance mild intoxication (BAC ≤ 100 mg per cent), medium intoxication (BAC 100–200 mg per cent), and severe intoxication (BAC >200 mg per cent) differ clinically, this schema does not work in patients suffering from alcoholism. In these people, different levels of tolerance can lead to completely different clinical pictures despite their having similar blood alcohol concentrations. Thus, psychopathology is more important than blood alcohol concentrations for estimating the severity of an acute intoxication state. With increasing BAC we observe elated mood, disinhibition, impaired judgement, belligerence, impaired social and occupational functioning, mood lability, cognitive impairment, reduced attention span, slurred speech, incoordination, unsteady gait, nystagmus, and stupor or coma.

The term 'pathological intoxication' can still be found in the older literature (reviewed by Lishman[1]). It was described as an outburst of aggression and uncontrollable rage, which might have led to serious destructions. As a rule, this behaviour, which was not typical for the individual, ended in terminal sleep and subsequent amnesia. However, since there is not enough empirical evidence for the existence of this syndrome, it was no longer considered in DSM-IV.[2]

Alcohol-induced amnesias ('blackouts')

This term refers to a transient state of amnesia after drinking excess. Usually patients' behaviour is no different from their behaviour during other periods of intoxication without blackouts. Nevertheless, the memory gap usually lasts for hours, but may be as long as a day or more. In extreme cases, patients find themselves in strange places with no recollection of how they got there.

Withdrawal

Withdrawal without complications

When alcohol is used regularly and withdrawn rapidly, a characteristic withdrawal syndrome can develop. It includes autonomic hyperactivity like hand tremor, insomnia, sweating, tachycardia, hypertension, and anxiety. The symptoms generally occur between 6 and 12 h after the last alcohol consumption. Depending on their severity they may last for up to 4 or 5 days. The neurobiological basis for withdrawal is a gradual upregulation of *N*-methyl-D-aspartate receptors under the influence of chronic alcohol use. As soon as the alcohol, which acts as a central nervous system depressant, is withdrawn, an overwhelming excitatory action in the brain mediated by the glutamatergic system is observed.

Withdrawal with perceptual disturbances

The individual usually experiences more discomfort and anxiety if transient visual, tactile, or auditory hallucinations or illusions are present. In this state, reality testing is still intact: the person still knows that the hallucinations are induced by the substance. If this is no longer true, a substance-induced psychotic disorder or a delirium tremens is likely.

Withdrawal with grand mal seizures (alcoholic convulsions, 'rum fits')

In about 30 per cent of the cases the typical grand mal seizures are followed by a delirium tremens. The electroencephalograph picture is only abnormal at the time of the fits, hence, alcohol convulsions differ pathophysiologically from latent epilepsy.

Alcohol-induced psychosis (delirium tremens)

In delirium tremens the symptoms of alcohol withdrawal described earlier are accompanied by a reduced level of consciousness, disorientation in time and place, impairment of recent memory, insomnia, and perceptual disturbances. The latter include misinterpretation of sensory stimuli and hallucinations; most are visual, but auditory and haptic hallucinations also occur. The hallucinations may be Lilliputian or of normal size, and may be of complex, frightening, and extremely realistic scenes. The patient is restless and fearful, and may become severely agitated. There is marked tremor, and ataxia when standing. Some patients experience vestibular disturbance. Autonomic disturbance includes sweating, tachycardia, raised blood pressure, and dilated pupils. There may be a mild pyrexia. Patients are usually dehydrated, often with abnormal electrolytes, leucocytosis, and impaired liver function. As in other forms of delirium, symptoms are worse at night.

Delirium tremens is the most severe of the states following withdrawal of alcohol, with a reported mortality of up to 5 per cent. In its fully developed form it is uncommon; the more frequent states are acute tremulousness, transient hallucinations with tremor, and uncomplicated fits. Delirium tremens usually begins after 3 to 4 days of abstinence from alcohol, although occasionally it starts while drinking continues. In the latter cases it is assumed that alcohol levels have fallen below a critical level. It is not known by what mechanism alcohol withdrawal leads to the clinical syndrome. Delirium tremens often appears to start suddenly, although close enquiry may reveal a prodromal stage of restlessness, anxiety, and insomnia. It usually lasts for 2 to 3 days, often ending with deep and prolonged sleep from which thepatient wakes symptom free and with little memory of the period of delirium. Rarely, the patient is left with an amnesic syndrome, perhaps the consequence of previous undetected Wernicke's encephalopathy.

Treatment is by sedation, usually with a benzodiazepine, together with fluid replacement under close observation. The possibility of accompanying head injury or infection should be investigated. Sedation should be adequate to prevent withdrawal seizures, with frequent monitoring of the response. High-potency vitamins are usually given to prevent Wernicke's encephalopathy. An anticonvulsant is given when there have been withdrawal seizures in the past. Cardiovascular collapse and hyperthermia occur occasionally and require urgent medical treatment.

Hallucinosis

Alcoholic hallucinosis is a rare condition in which auditory hallucinations are present in clear consciousness and without autonomic overactivity, usually in a person who has been drinking excessively for many years. The hallucinations often begin as simple noises, but are gradually replaced by voices, which may threaten, abuse, or reproach the person. Usually the voices speak to the person, but sometimes they discuss him or her in the third person. The voices may be occasional or relentlessly persistent. They may command the patient, who may respond with unrestrained or suicidal behaviour. Delusions are secondary interpretations of the hallucinations. Autochthonous hallucinations suggest schizophrenia, as do thought disorder or incongruity of affect. The patient is usually distressed, anxious, and restless.

In both ICD-10 and DSM-IV, the disorder is classified as a substance-induced psychotic disorder and not, as has been suggested in the past, a form of schizophrenia (released by heavy drinking). The differential diagnosis includes transient auditory hallucinations occurring during withdrawal from a period of heavy drinking, and delirium tremens in which auditory hallucinations may accompany the more prominent visual ones. In both conditions the auditory hallucinations are transient and disorganized, and in the latter consciousness is impaired. In contrast, the auditory hallucinations of an alcoholic hallucinosis are persistent and organized, and occur in clear consciousness. Other differential diagnoses are depressive disorder with psychotic symptoms and schizophrenia, both of which can be accompanied by heavy drinking.

The hallucinations usually respond rapidly to antipsychotic medication. The prognosis is good; usually the condition improves within days or a couple of weeks provided that the person remains abstinent. Symptoms that last for 6 months generally continue for years.[3]

Psychiatric disorders

Alcohol-dependent patients often present with symptoms of anxiety or depression. These states are generally referred to as

comorbid disorders or dual diagnosis. Alcoholism can be a consequence of anxiety and mood disorders ('secondary alcoholism'). It can develop independently after anxiety and depression, or it can precede anxiety and depressive symptoms ('primary'). As the former are discussed elsewhere in this textbook, here we concentrate on the latter.

Alcohol-induced mood disorders

Alcohol is a central nervous system depressant. Taken regularly in high doses it may provoke feelings of sadness. Episodes of withdrawal or relative withdrawal can lead to excitability and nervousness, including anxiety. The more a person drinks, the more likely it is that these symptoms will occur. Finally in the stage of alcohol dependence, up to 80 per cent of people report depressive symptoms at some time in their life. About one-third of male patients and up to 50 per cent of female patients have experienced longer periods of severe depression.[4] These high prevalence rates are noteworthy, since more than 20 per cent of alcoholics have attempted suicide once or more and about 15 per cent die in their attempt. Besides depressive features, alcohol-induced mood disorders may also comprise manic symptoms or mixed features. However, the diagnosis should only be used when the symptoms cause clinically significant impairment or distress in social, occupational, or other areas of functioning.

Concerning treatment, it is interesting to note that despite the vast majority of patients who present with depressive symptoms at the beginning of treatment for alcoholism, only very few need specific antidepressant medication or specific psychotherapy. In most other cases depressive symptoms disappear within weeks of controlled abstinence.[5]

Alcohol-induced anxiety disorders

This diagnosis should only be used when anxiety symptoms are thought to be related to the direct physiological effects of alcohol. The symptomatology may involve anxiety, panic attacks, and phobias. Both alcohol-induced anxiety disorders and mood disorders can develop during intoxication, withdrawal, or up to 4 weeks after cessation of alcohol consumption. During intoxication or withdrawal, the diagnosis should only be given when the symptomatology clearly exceeds what would be expected from anxiety or depressive symptoms during a regular intoxication or withdrawal episode.

Anxiety disorders are among the most common groups of psychiatric disorders in the general population, with prevalence rates of up to 25 per cent.[6] In clinical studies between 20 and 70 per cent of patients with alcoholism also suffer from anxiety disorders.[7] On the other hand, between 20 and 45 per cent of patients with anxiety disorders also have histories of alcoholism.[8] However, it has been argued that the comorbidity figures are overestimated, because in some of the studies the focus was on drinking patterns rather than on alcohol dependence or they describe anxiety symptoms rather than disorders according to diagnostic criteria.[9] Family studies analysing the comorbidity of alcoholism and anxiety disorders might be a means of clarifying this controversy. For instance, in the Yale study the presence of anxiety disorders in the probands slightly increased the risk for alcohol dependence in their relatives, whereas alcohol dependence in the proband did not increase their relative's risk for anxiety disorders.[10] Similarly, Maier *et al.*[11] demonstrated an increased risk of alcoholism in probands with panic disorders, but not the reverse. Kendler *et al.*[12] in a study of female twins, found evidence that common genetic factors may underlie both alcoholism and panic disorder.

Effects on the brain

Cerebral cortex

Chronic alcohol consumption leads to structural and functional changes in the brain. Alcoholic dementia is dealt with in Chapter 4.1.11. Most of the tissue loss from the cerebral hemispheres in alcoholics is accounted for by a reduction in the volume of the cerebral white matter, additionally there is a slight reduction in the volume of the cerebral cortex. This has been demonstrated both pathologically[13] and using magnetic resonance imaging with quantitative morphometry.[14]

Harper *et al.*[15] documented neuronal loss in alcoholics. There was a 22 per cent reduction in the number of neurones in the superior frontal cortex (Brodmann's area 8), while surviving neurones showed shrinkage in the superior frontal, motor, and frontal cingulate cortices.[16] This finding of cortical damage in alcoholics is consistent with neuroradiological studies.[14]

Ferrer *et al.*[17] examined the dendritic tree of cortical neurones in alcoholic subjects using Golgi-apparatus impregnation techniques. They described a significant reduction in the basal dendritic tree of layer III pyramidal neurones in both the superior frontal and motor cortices. These studies suggest that, even though there is no significant reduction in the numbers of cortical neurones in the motor cortex, there are cellular structural abnormalities that could have important functional implications.

Wernicke's encephalopathy

The best-known features of heavy alcohol consumption in adults are Wernicke's encephalopathy and Korsakoff's syndrome. Wernicke's encephalopathy is directly caused by thiamine deficiency, which results from a combination of inadequate dietary intake, reduced gastrointestinal absorption, decreased hepatic storage, and impaired utilization. Only a subset of thiamine-deficient alcoholics develop Wernicke's encephalopathy, perhaps because they have inherited or acquired abnormalities of the thiamine-dependent enzyme transketolase, which reduces its affinity for thiamine. Wernicke's encephalopathy is characterized by degenerative changes, including gliosis and small haemorrhages in structures surrounding the third ventricle and aqueduct: namely, the mammillary bodies, hypothalamus, mediodorsal thalamic nucleus, colliculi, and midbrain tegmentum. Clinical features associated with the Wernicke–Korsakoff syndrome include memory deficits, ocular signs, ataxia, and global confusional states. Most can be related to damaged functional systems in the hypothalamus, midbrain, and cerebellum. In a large Scandinavian neuropathological study, 12.5 per cent of all alcoholics exhibited signs of Wernicke's encephalopathy.[18]

Korsakoff's syndrome

About 80 per cent of alcoholic patients recovering from Wernicke's encephalopathy develop Korsakoff's amnesic syndrome. It is characterized by marked deficits in anterograde and retrograde memory, apathy, an intact sensorium, and relative preservation of other intellectual abilities. Korsakoff's amnesic syndrome may also appear without an antecedent episode of Wernicke's

encephalopathy. Acute lesions may be superimposed on chronic lesions, suggesting that subclinical episodes of Wernicke's encephalopathy may culminate in Korsakoff's amnesic syndrome. The memory disorder correlates best with the presence of histopathological lesions in the dorsomedial thalamus. (Amnesic syndrome is considered further in Chapter 4.1.12.)

Cerebellar degeneration

Many alcoholic patients develop a chronic cerebellar syndrome related to the degeneration of Purkinje cells in the cerebellar cortex. Quantitative studies revealed a significant loss of cerebellar Purkinje cells (by 10–35 per cent) and shrinkage of the cerebellar vermal, molecular, and granular cell layers.[19] Evidence for a direct toxic effect caused by ethanol is provided by animal models.[20] In neuroimaging studies, however, cerebellar ataxia in alcoholics does not correlate with the daily, annual, or lifetime consumption of ethanol. As in Wernicke's encephalopathy, thiamine deficiency due to poor nutrition has also been implicated. Cerebellar atrophy has been reported to occur in about 40 per cent of chronic alcoholics.[19] In a clinical study of alcoholic inpatients, 49 per cent had at least discrete clinical signs of cerebellar atrophy.[21]

The diagnosis of alcoholic cerebellar ataxia is based on the clinical history and neurological examination. The ataxia affects the gait most severely. Limb ataxia and dysarthria occur more often than in Wernicke's encephalopathy, whereas nystagmus is rare. Computed tomography or magnetic resonance imaging scans may show cerebellar cortical atrophy, but a considerable number of alcoholic patients with this finding are not ataxic on examination. Whether these represent subclinical cases in which symptoms will develop subsequently is unclear. It is interesting to note that impaired cerebellar function improves significantly when abstinence is maintained.[22]

Hepatocerebral degeneration

Hepatic encephalopathy develops in many alcoholics with liver disease, and is characterized by altered sensorium, frontal release signs, 'metabolic' flapping tremor, hyperreflexia, extensor plantar responses, and occasional seizures. Whereas some patients progress from stupor to coma and then death, others recover and suffer recurrent episodes. The brains of patients with hepatic encephalopathy show enlargement and proliferation of protoplasmic astrocytes in the basal ganglia, thalamus, red nucleus, pons, and cerebellum, in the absence of neuronal loss or other glial changes.[23]

Patients who do not recover fully after an episode of hepatic encephalopathy go on to develop a progressive syndrome of tremor, choreoathetosis, dysarthria, gait ataxia, and dementia. Hepatocerebral degeneration may progress in a stepwise fashion, with incomplete recovery after each episode of hepatic encephalopathy, or slowly and inexorably, without a discrete episode of encephalopathy.

Rare disorders

The **Marchiafava–Bignami syndrome** is a disorder of demyelination or necrosis of the corpus callosum and adjacent subcortical white matter. The course may be acute, subacute, or chronic, and is marked by dementia, spasticity, dysarthria, and an inability to walk. Patients may lapse into coma and die, survive for many years in a demential condition, or occasionally recover.

Central pontine myelinolysis is a disorder of the cerebral white matter that usually affects alcoholics, but it also occurs in non-alcoholics with liver disease including Wilson's disease, malnutrition, anorexia, burns, cancer, Addison's disease, and severe electrolyte disorders such as thiazide-induced hyponatraemia; however, the majority of cases occur in alcoholics, suggesting that alcoholism may contribute to the genesis of central pontine myelinolysis in, as yet, undefined ways.[23] Myelinolytic lesions can be reduced experimentally by rapid correction of chronic hyponatraemia. Symptoms include loss of pain sensation in the limbs, bulbar palsy, quadriplegia, disordered eye movements, vomiting, confusion, and coma.

Reversibility of brain damage

Alcohol-related neuroanatomical brain changes have been shown to be partially reversible. These findings created an ongoing debate on possible mechanisms and clinical correlates.[22]

Foetal alcohol syndrome

The first description of the foetal alcohol syndrome was given by French scientists in 1968.[24] As a research paradigm, it has a major impact on our understanding of alcohol's effects on the brain. Clinically the syndrome is characterized by: growth retardation involving height, weight, and head circumference; deficient intellectual and social performance and muscular coordination; minor structural anomalies of the face, together with more variable involvement of the limbs and the heart.

The basis of this pathology is a cascade of effects exerted by alcohol on the developing cell. Under normal conditions growth factors enhance the growth of cells and their differentiation, but alcohol can diminish these effects.[25] A second way of damaging the developing nerve cell is through the production of free radicals that allow calcium to accumulate in the cells.[26] The induction of a free-radical formation is induced by alcohol. The result of both pathogenic processes is a decrease in the overall size of the brain and a diminution in the thickness of the outer layers of the cortex, due to decreases in the total numbers of cells. Impaired nerve cell migration might also play a role in the development of the foetal alcohol syndrome.[27]

The effects of alcohol on the developing brain are clinically measured by assessing the head circumference, with a clear dose-dependent effect.

The foetal alcohol syndrome is considered further in Chapter 9.2.7.

Effects on the body

Malnutrition and vitamin deficiency

Malnutrition can be a consequence of deficient food intake. More important in alcoholics seem to be maldigestion and malabsorbtion ('secondary malnutrition'). Apart from the direct toxic effect of alcohol on most body tissues, malnutrition is an important contributor to organ damage in alcoholics.[28] Vitamin metabolism may be profoundly affected by chronic alcohol consumption. As a consequence, many alcoholics have deficiencies in vitamins B1 (thiamine), A, D, B6, and E, and folate. This can lead to a variety of physical consequences, including damage to different organs.

Peripheral neuropathy

Besides its effect on the central nervous system, alcohol also damages motor, sensory, and autonomic nerves that control muscles and internal organs. Symptoms are weakness, numbness, pain, and a prickly feeling or burning of the skin, especially the feet. Usually on neurological examination, the tendon reflexes are diminished or have completely disappeared and skin sensibility is reduced, especially in the feet and in the lower limbs. When patients abstain from alcohol, the progression of the symptoms can be stopped and even partial recovery is possible.

Muscle

Alcohol is toxic to skeletal muscles in a dose-dependent way. Alcoholics often suffer from malnutrition, which adds to the chronic changes in muscles. Chronic myopathy can be found in 40 to 60 per cent of alcohol-dependent patients.[29] Pathophysiological mechanisms of muscle damage include alterations in membrane fluidity, ion channels, and pumps, as well as protein synthesis and hormonal dysfunction. Patients complain of pain and weakness. Swelling of the muscle can be easily detected. In chronic states, muscle atrophy is evident. There is no acute treatment for alcoholic myopathy other than abstinence, when acute myopathy can rapidly disappear; chronic myopathy usually only improves, leaving persistent weaknesses.

Liver

The effects of ethanol on the liver are among the first and best-known symptoms of alcoholism. The first manifestation of alcoholic liver diseases is the fatty liver. It is followed by early fibrosis, which can be associated with alcoholic hepatitis. If the process continues, irreversible damage leading to severe fibrosis and to cirrhosis is observed.[30] These effects occur through heavy alcohol consumption even in the absence of dietary deficiencies.

Mortality from liver cirrhosis has long been an important correlate of the per capita consumption in a given population. Liver damage is also important because it produces an increase in liver enzymes such as aspartate transaminase, alanine transaminase, and γ-glutamyl transferase, which again are of great practical value as diagnostic markers of severe alcohol consumption. Alcohol accounts for more than 80 per cent of all cirrhosis deaths, a consequence that seems to be even more pronounced in women.[31]

Pancreas

About 5 per cent of alcoholics develop chronic pancreatitis. Ethanol seems to damage the pancreas slowly. In general, it takes between 10 and 15 years of heavy drinking before pancreatitis becomes clinically apparent. In the presomatic phase certain changes such as fibrosis, calcium deposits, and especially loss of functioning in enzyme- and hormone-producing cells can be demonstrated. The acute symptoms are abdominal pain and vomiting. Chronic complications include weight loss, steatorrhoea, and diabetes mellitus.[32]

Skin

Originally it was believed that skin alterations in alcoholics are due to alcoholic liver disease. However, more recent research has revealed that the skin may be affected much earlier by alcohol misuse.[33] Whereas the palmar erythema and spider naevi are well-known consequences of alcoholic liver disease, which also serve as diagnostic markers for alcoholism, psoriasis and facial erythema have less often been linked with high alcohol consumption. Alcohol clearly has to be on the list of agents known to exacerbate psoriasis. One possible mechanism of the action of alcohol on the skin could be a defect in the immune system.

Heart

Cardiac myopathy is one of the oldest known physical consequences of high alcohol consumption. Similar to ethanol's effects on skeletal muscles, the cells of the heart muscle are damaged by ethanol's influence on ion channels and pumps etc. Atrophy leads to a dilatation of the heart as a whole.

Recently, the effect of alcohol on coronary heart disease has been widely discussed. Indeed, it seems that there is a beneficial effect of moderate alcohol consumption. Although the reasons are currently under discussion, recent data suggest, that the combination of several actions including changes in lipid metabolism, antioxidant effects, changes in haemostasis and platelet aggregation, arterial vasodilatation mediated by NO release and the expression of cardioprotective proteins contribute to these 'French Paradox'.[34] It seems that an alcohol-induced increase in high-density lipoproteins and a decrease in low density lipoproteins may play a role in this process—an alteration in platelet aggregation could be one possible mechanism of action. Besides cardiomyopathy, cardiac arrhythmias are prominent consequences of alcohol consumption. Close to one-third of all cardiomyopathies can be attributed to alcohol consumption.

Hypertension

A dose-response relationship between drinking and diastolic and systolic blood pressure has been shown consistently.[31] In alcohol consuming population, the amount of alcohol consumption is significantly associated with hypertension and cardiovascular as well as all cause mortality. It is not clear, however, whether this relationship can only be seen above a threshold level of consumption.

Cancer

There is very clear evidence that alcohol increases the risk of cancer at the upper bronchodigestive tract. This includes cancer of the mouth, pharynx, larynx, and oesophagus. Additionally, alcohol consumption correlates with primary liver cancer. A possible link between alcohol and breast cancer is still a matter of debate: the association is not strong and not necessarily causative, at least for moderate consumption.[35] The same seems to be true for the correlation between beer drinking and cancer of the rectum.

Further information

Schuckit, M.A. (2006). *Drug and alcohol abuse: a clinical guide to diagnosis and treatment*. Springer, Berlin.

Erickson, C. (2007). *The science of addiction. From neurobiology to treatment*. W.W. Norton & Company, New York.

Icon Health Publications. (2004). *Alcohol addiction—a medical dictionary, bibliography, and annotated research guide to internet references*. Icon Group International, San Diego.

Spanagel, R. and Mann, K. (2005). *Drugs for relapse prevention of alcoholism*. Birkhäuser Verlag, Berlin.

References

1. Lishman, W.A. (1990). *Organic psychiatry. The psychological consequences of cerebral disorder* (2nd edn). Blackwell Science, Oxford.
2. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
3. Glass, I.B. (1989). Alcohol hallucinosis: a psychiatric enigma-2. Follow-up studies. *British Journal of Addiction*, **84**, 151–64.
4. Brown, S.A. and Schuckit, M.A. (1988). Changes in depression among abstinent alcoholics. *Journal of Studies on Alcohol*, **49**, 412–17.
5. Stetter, F., Rein, W., and Mann, K. (1991). How depressive are male alcoholic inpatients? Psychometric results from the Tübinger Alkoholismusprojekt. *European Psychiatry*, **6**, 243–9.
6. Kessler, R.C., McGonagle, K.A., Zhao, S., *et al.* (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. *Archives of General Psychiatry*, **51**, 8–19.
7. Merikangas, K.R. and Angst, J. (1995). Comorbidity and social phobia: evidence from clinical, epidemiologic, and genetic studies. *European Archives of Psychiatry and Clinical Neuroscience*, **244**, 297–303.
8. Kushner, M.G., Sher, K.J., and Beitman, B.D. (1990). The relation between alcohol problems and the anxiety disorders. *American Journal of Psychiatry*, **147**, 685–95.
9. Schuckit, M.A. and Hesselbrock, V. (1994). Alcohol dependence and anxiety disorders: what is the relationship? *American Journal of Psychiatry*, **151**, 1723–34.
10. Merikangas, K.R., Stevens, D., Fenton, B., *et al.* (1996). Comorbidity and co-transmission of anxiety disorders and alcoholism: results of the Yale Family Study. In *Proceedings of the American Psychiatric Association*, May 1996. American Psychiatric Association, Washington, DC.
11. Maier, W., Minges, J., and Lichtermann, D. (1993). Alcoholism and panic disorder: co-occurrence and co-transmission in families. *European Archives of Psychiatry and Clinical Neuroscience*, **243**, 205–11.
12. Kendler, K.S., Walters, E.E., Neale, M.C., *et al.* (1995). The structure of genetic and environmental risk factors for six major psychiatric disorders in women: phobia, generalized anxiety disorder, panic disorder, bulimia, major depression, and alcoholism. *Archives of General Psychiatry*, **52**, 374–83.
13. de la Monte, S.M. (1988). Disproportionate atrophy of cerebral white matter in chronic alcoholics. *Archives of Neurology*, **45**, 990–2.
14. Harper, C. (2007). The neurotoxicity of alcohol. *Human & Experimental Toxicology*, **26**, 251–7.
15. Harper, C., Kril, J., and Daly, J. (1987). Are we drinking our neurons away? *British Medical Journal*, **294**, 534–6.
16. Harper, C.G. and Kril, J.J. (1989). Patterns of neuronal loss in the cerebral cortex in chronic alcoholic patients. *Journal of Neurology Sciences*, **92**, 81–9.
17. Ferrer, I., Fabregues, I., Rairiz, J., *et al.* (1986). Decreased numbers of dendritic spines on cortical pyramidal neurons in human chronic alcoholism. *Neuroscience Letters*, **69**, 115–19.
18. Torvik, A., Lindbö, C.F., and Rodge, S. (1982). Brain lesions in alcoholics. *Journal of Neurological Sciences*, **56**, 233–48.
19. Torvik, A. and Torp, S. (1986). The prevalence of alcoholic cerebellar atrophy. A morphometric and histological study of an autopsy material. *Journal of Neurological Sciences*, **75**, 43–51.
20. Riley, J.N. and Walker, D.W. (1978). Morphological alterations in hippocampus after long term alcohol consumption in mice. *Science*, **201**, 646–8.
21. Mann, K. (1992). *Alkohol und Gehirn—über strukturelle und funktionelle veränderungen nach erfolgreicher therapie*. Springer, Berlin.
22. Mann, K., Mundle, G., Strayle, M., *et al.* (1995). Neuroimaging in alcoholism: CT and MRI results and clinical correlates. *Journal of Neural Transmission (General Section)*, **99**, 145–55.
23. Charness, M. (1993). Brain lesions in alcoholics. *Alcoholism, Clinical and Experimental Research*, **17**, 2–11.
24. Lemoine, P., Harousseau, H., Borteyru, J.-P., *et al.* (1968). Les enfants de parents alcooliques: anomalies observées à propos de 127 cas. *Ouest Médical*, **25**, 477–82.
25. Dow, K.E. and Riopelle, R.J. (1985). Ethanol neurotoxicity: effects on neurite formation and neurotrophic factor production *in vitro*. *Science*, **228**, 591–3.
26. Manning, M.A. and Eugene Hoyme, H. (2007). Fetal alcohol spectrum disorders: a practical clinical approach to diagnosis. *Neuroscience and Biobehavioral Reviews*, **31**, 230–8.
27. Kumada, T., Jiang, Y., Cameron, D.B., *et al.* (2007). How does alcohol impair neuronal migration? *Journal of Neuroscience Research*, **85**, 465–70.
28. Estruch, R. (1996). Alcohol and nutrition. In *Alcohol misuse: a European perspective* (ed. T.J. Peters), pp. 41–61. Harwood Academic, London.
29. Urbano-Márquez, A. and Fernández-Solà, J. (1996). Musculo-skeletal problems in alcohol abuse. In *Alcohol misuse: a European perspective* (ed. T.J. Peters), pp. 123–44. Harwood Academic, London.
30. Lieber, C.S. (1998). Hepatic and other medical disorders of alcoholism: from pathogenesis to treatment. *Journal of Studies on Alcohol*, **59**, 9–25.
31. Huntgeburth, M., Ten Freyhaus, H., and Rosenkrank, S. (2005). Alcohol consumption and hypertension. *Current Hypertension Reports*, **7**, 180–5.
32. Niebergall-Roth, E., Harder, H., and Singer, M.V. (1998). A review: acute and chronic effects of ethanol and alcoholic beverages on the pancreatic exocrine secretion *in vivo* and *in vitro*. *Alcoholism, Clinical and Experimental Research*, **22**, 1570–83.
33. Higgins, E.M. (1996). Alcohol misuse and the skin. In *Alcohol misuse. A European perspective* (ed. T.J. Peters), pp. 77–87. Harwood Academic, London.
34. Providencia, R. (2006). Cardiovascular protection from alcoholic drinks: scientific basis of the French Paradox. *Revista Portuguesa de Cardiologia*, **25**, 1043–58.
35. Poschl, G. and Seitz, H.K. (2004). Alcohol and cancer. *Alcohol & Alcoholism*, **39**, 155–65.

4.2.2.4 Treatment of alcohol dependence

Jonathan Chick

A chronic relapsing disorder

Some people repeatedly put themselves or others at risk by drinking. One view is that such people could drink sensibly if they were more considerate and used more will power. Another increasingly accepted view is that many such individuals are in a state, existing in degrees of severity, in which the freedom to decide whether to change their drinking, and to adhere to that decision, is reduced compared with other drinkers. This state partly depends on perceived pay-offs for changing, and on acquired dispositions, which are less accessible to conscious control. Such persons become aware of a wish, or urge, to drink, which overcomes rational thought. They may then make up an explanation, for example, 'No wonder I feel like a drink, I've had a hard day'.

Such individuals benefit from help to unlearn those patterns, and to learn different approaches to problems. Discussion, care, and encouragement from others can bolster their will to do so. Assistance to set-up controls within or from outside themselves may help. Some people can do this without external help, and others with the help of Alcoholics Anonymous (**AA**) alone.[1]

This approach argues that dependence on alcohol should be managed like other relapsing disorders, such as diabetes and asthma,[2] by using long-term monitoring coupled with intermittent or continuous treatment.

Starting treatment

The initial interview

Assessment is the first step of intervention; clumsy interviewing alienates an ambivalent patient. The key to success is accepting that the patient is probably in two minds about the interview and about changing his or her drinking habits. Avoid confrontation. The drinking has probably already shown its resistance to deterrence by fear or pain. Gently nudge the matrix of conflicting motivations in the direction of action.

Patients may or may not have been referred for help with alcohol problems. Even if they have, the interview should begin with enquiry into the patient's current concerns. Reflective listening[3] helps the patient to clarify these concerns, conveys empathy, and avoids premature closure. A spirit of collaborative enquiry helps patients to reach their own conclusions about the role of alcohol in their troubles. This will be more convincing than a recitation of medical advice. People are more likely to believe what they hear themselves say than what others tell them. The interview is less likely to slip into confrontation if the doctor conveys recognition that, for the patient, drinking alcohol has been pleasurable. Therefore the assessment should not proceed in a series of closed questions, such as: 'Do you drink more than you intend to?' 'Does alcohol make you depressed?' Instead, ask open-ended questions: 'Tell me about your pattern of drinking. What are the good aspects . . . and what are the disadvantages?' 'How does alcohol fit in with these periods of hopelessness you describe?' The patient may want it understood that at times alcohol has dulled pain. Only then will there be a concession that the cumulative effect has been to worsen mood.

A comment such as 'I'm just a heavy social drinker, not an alcoholic' is not a gauntlet to be seized—an argument about definitions will distract from the work of clarifying and planning how to deal with the current problems. Instead, a response such as 'I gather you don't like labels' may reveal pertinent fears and prejudices (e.g. that alcoholics are failures, who get locked up in hospital).

Denial permits dismissal of unpleasant or unwanted facts and feelings. It hurts to admit that you have lost your family's respect, or that you will have to give up alcohol, which you enjoy. Alcohol problems still carry disgrace. In Islamic cultures, where alcohol is forbidden, denial from shame may be deepened by fear of punishment from the authorities.

Explain symptoms

Help the patient to understand withdrawal symptoms and how they can abort attempts to reduce consumption. Patients frequently attribute withdrawal symptoms to other causes; for example, waking at 4 a.m. with sweats and anxiety may be attributed to worry, and trembling hands in the morning to stress.

Informant

If the partner, a close friend, or a relative is present from the start, the salient points usually emerge more rapidly. However, the patient should also be seen alone because matters to do with the police, an employer, the bank, or a lover may still be unknown to the partner. Relatives should hear the exchange between doctor and patient, otherwise the version they hear later from the patient may be diluted: 'The doctor says I'm not an alcoholic'. This can leave relatives even angrier than before, convinced that no one understands their distress and that the drinker has once again deceived the doctor.

Assessments

The use of a breathalyser or saliva test to measure blood alcohol concentration puts alcohol consumption firmly into the objective arena. Use the test before the individual starts to detail recent drinking, there is nothing to be gained from showing that the patient sometimes minimizes the drinking.

Physical signs may be helpful. Heavy drinking may cause excessive capillarization in the conjunctivae or in the skin of the nose and cheeks. The liver may be enlarged. Look for tremor in the outstretched tongue, which is less commonly concealed (or exaggerated) than tremor in the fingers. Tachycardia is another useful sign of withdrawal. In a hyperaroused fearful patient, who has already been without a drink for 24 h, a pulse of over 110 beats/min may presage delirium tremens.

Clinicians vary in how structured an assessment they prefer, but at some point in the first one or two interviews the following should be noted: drinking patterns, history of withdrawal symptoms, previous attempts to stop drinking, use of drugs (prescribed and not prescribed), physical complications including head injuries, police, or Court involvement (past and current), dwelling arrangements, problems at home, trouble at work (specifying whether the employer has commented on drinking alcohol and/or started disciplinary action), psychiatric illness, family history, previous treatments, and experience of AA.

Medical assistance for withdrawal

Medical assistance to reduce the short-term discomfort of withdrawal can be the beginning of restructuring of thoughts and lifestyle towards long-term abstinence.

If dependence is severe, especially in an unplanned situation where a very heavy drinker is suddenly deprived of alcohol because of an accident, illness, or police arrest, care must be taken to prevent the life-threatening complications of convulsions or delirium. Anticipation is the key.

When dependence is less marked, withdrawal symptoms are mild and the person can stop drinking by gradual reduction, encouraged by the physician or a friend.

When the patient's aim is 'controlled drinking'(see below), this may also entail an initial stage of withdrawal, as the final goal is more likely to be achieved after abstinence for 2 or 3 months.

The setting

Controlled studies have shown that outpatient withdrawal is safe and effective for mild and moderately dependent alcoholics.[4,5] Advice for patients withdrawing at home is given in Box 4.2.2.4.1. Hayashida *et al.*[4] randomly allocated 164 mild to moderately affected patients to either inpatient or outpatient detoxification. Completion was successful in 95 per cent of the former and 72 per cent of the latter; inpatient care cost eight times more than outpatient care.

Box 4.2.2.4.1 Advice to patient on withdrawing from alcohol at home

If you have been chemically dependent on alcohol, stopping drinking causes you to become tense, edgy, perhaps shaky or sweaty, and unable to sleep. There can be vomiting or diarrhoea. This 'rebound' of the nervous system can be severe. Medication controls the symptoms while the body adjusts to being without alcohol. This usually takes 3 to 7 days from the time of your last alcoholic drink. If you did not take medication, the symptoms would be worst in the first 48 h, and then gradually disappear. This is why the dose starts high and then reduces.

You have agreed not to drink alcohol. You may become thirsty. Drink fruit juices and water but do not overdo it. You do not have to 'flush' alcohol out of the body. More than 3 l of fluid could be too much. Do not drink more than three cups of coffee or five cups of tea. These contain caffeine, which disturbs sleep and causes nervousness.

Aim to avoid stress. The important task is not to give in to the urge to take alcohol. Help yourself relax by going for a walk, listening to music, or taking a bath.

Sleep. You may find that even people with capsules, or as they are reduced, your sleep is disturbed. You need not worry about this lack of sleep as it does not seriously harm you, but starting to drink again does. Your sleep pattern will return to normal in a month or so. It is better not to take sleeping pills so that your natural sleep rhythm returns. Try going to bed later. Take a bedtime snack or milky drink. The capsules may make you drowsy so you must not drive or operate machinery. If you become drowsy, miss out a dose.

Meals. Even when you are not hungry, try to eat something. Your appetite will return.

Admission to a hospital is indicated when the home social milieu is inimical to abstinence, or when there is a history of withdrawal convulsions or delirium; it is urgent when there are any signs of Wernicke's encephalopathy.

Medication

A benzodiazepine[6] is prescribed for two reasons: first, to reduce the risk of severe withdrawal symptoms with delirium or convulsions (indicated if recent consumption has been more than 15 units/day for more than 10 days); second, to assist the individual whose wish to abstain or reduce drinking is overcome by longing for alcohol (craving), shaking, anxiety, insomnia, or nausea and vomiting.

A typical outpatient regimen would be chlordiazepoxide 20 to 30 mg four times daily, reducing to zero over 5 days, with the larger doses given at night (Table 4.2.2.4.1). Medication is issued on the understanding that the patient does not also take alcohol. If there is any doubt that this instruction will be followed, medication is issued daily and a check made (ideally by breath or saliva tests) that drinking has not been resumed. Chlordiazepoxide is preferred to diazepam for outpatient use because it has a lower street value and is therefore less likely to be sold on. When managing severe withdrawal symptoms with marked agitation and tremor, or incipient delirium, diazepam (starting at 10 mg four times daily) is preferred because it has a more rapid action and can be given parenterally. A benzodiazepine with one metabolite only and a shorter half-life (e.g. oxazepam, lorazepam) is preferred if liver function is significantly impaired (i.e. there is jaundice, ascites, oedema, low serum albumin, or raised serum bilirubin).

Table 4.2.2.4.1 Example of a fixed-dose regime for outpatient alcohol withdrawal using capsules of chlordiazepoxide 10 mg

	First thing	12 noon	6 p.m.	Bedtime
Day 1		3	3	3
Day 2	2	2	2	3
Day 3	2	1	1	2
Day 4	1	1		2
Day 5		1		1

For inpatients, a benzodiazepine such as diazepam 10 mg may be given every hour until symptoms are controlled (symptom-triggered dosing). This procedure leads to lower total prescription of benzodiazepine, less oversedation, and quicker discharge from hospital.[7]

If the patient is vomiting, give metoclopropamide 10 mg intramuscularly 30 min before the first benzodiazepine tablet and/or perhaps choose a benzodiazepine that can be administered parenterally; lorazepam 1 mg is absorbed adequately from the intramuscular site, or diazepam 10 mg can be given intravenously (or rectally).

Treating convulsions

With the aim of preventing further convulsions, the patient who has just had a fit or is in a fit is given 10 mg diazepam. Consider giving 15 to 20 mg in a patient who has been taking benzodiazepines regularly prior to this event, or is much above average weight. It is illogical to commence an anticonvulsant, which may take 2 to 3 days to reach a therapeutic serum level. Rather, increase the dose of the benzodiazepines. A convulsion may presage delirium.

Preventing convulsions

Deaths have occurred in hospital, prison, and police cells from repeated alcohol withdrawal fits. When withdrawal is planned in patients with a history of fits of any cause the risk can be reduced by commencing phenytoin (300 mg daily) 4 days before the cessation of drinking. In an acute situation, larger than normal doses of long-acting benzodiazepines are given in the first 36 h without waiting until the blood alcohol level has fallen to zero. The benzodiazepine should be started as soon as the blood alcohol level can be presumed to be falling, even though the patient still smells of alcohol or has a positive breath test, provided that he or she is sober enough to understand and cooperate with the procedure.

Treating delirium tremens

Increasing the dose of the benzodiazepine may be sufficient to control the agitation. If not, the slight epileptogenic effect of antipsychotic drugs should not deter their use, especially if delusions and hallucinations have developed, provided that anticonvulsant protection by a benzodiazepine is in place. Parenteral haloperidol plus parenteral lorazepam is usually effective. When a patient's

behaviour is uncontrolled or dangerous, transfer to a secure unit may be needed. Authoritative calm nursing reduces the risk of aggression. Hospitals should have an emergency team of sufficient personnel to manage disturbed patients.

Preventing delirium tremens

If confusion and hallucinations develop, this usually occurs 48 to 72 h after the last drink. Sufficient benzodiazepine, given early in the withdrawal, reduces the risk, as does sensitive nursing in a quiet evenly-lit environment. Explaining symptoms and orientating the patient reduces anxiety, paranoia, and confusion.

Vitamin therapy

It is reasonable to prescribe thiamine 50 mg orally three times a day for 2 to 3 weeks, as thiamine stores may be depleted because of poor diet and alcohol-impaired gut absorption. Wernicke–Korsakoff syndrome is life-threatening and steps must be taken to avoid it developing.[8] The malnourished patient, or the patient who shows any sign of Wernicke's encephalopathy (confusion, ataxia, ophthalmoplegia, nystagmus—do not wait for the 'triad' of symptoms), must be given immediate parenteral B vitamins. Anaphylactic shock was a very rare complication of some older preparations. It is less likely with intramuscular than intravenous injection; infusion saline drip, when practicable, is probably preferable to slow bolus injection.

Interventions to reduce relapse

The evidence

With appropriate help, withdrawing from alcohol is not the dependent drinker's main difficulty. The main difficulty is avoiding relapse into further problematic drinking or dependence.

Before the era of randomized controlled trials, psychiatrists typically would explore with patients possible personality or psychological causes of their excessive drinking—trying to find out 'why?'. However, evidence that this reduced relapse was lacking. Indeed, it may have sometimes had an adverse effect by reinforcing the drinkers' perception of having a need to drink and by creating transference problems which might later trigger drinking.[9] Non-directive counselling may also sometimes have had negative effects, acting as a confessional, with a sense of absolution allowing further drinking.

In recent years, several systematic reviews and meta-analyses have been conducted of treatments to prevent relapse in alcohol dependence. Drawing on data from high quality trials a consensus has emerged.[10–12] Effective treatment are social skills training based on behavioural cognitive therapy principles,[13] motivational enhancement[14] albeit tested sometimes in less severe groups of patients, the community reinforcement approach,[15] behaviour contracting, and behavioural marital therapy[16] and the pharmacotherapies described below.

Abstinence or 'controlled drinking'?

Harmful or hazardous use of alcohol without severe dependence can sometimes revert to risk-free drinking. Patients with social supports (family and job) and without impulsive personalities and many social problems are most likely to succeed. For others, including most of those dependent on alcohol, the goal of abstinence is better.[29] Among patients attending specialized clinics, the proportion who can sustain problem-free drinking for at least 1 year is small—5 per cent is a typical finding.[17–19] A randomized trial comparing the goals of controlled drinking and abstinence did not favour controlled drinking.[20] However, for patients without established dependence, reduction programmes (whether or not towards abstinence) using FRAMES (Table 4.2.2.4.2) proved to be more effective than no intervention.[21–23] Interventions in primary care are discussed in Chapter 4.2.2.5.

Table 4.2.2.4.2 FRAMES: ingredients of a brief intervention

Feedback about personal risk or impairment
Responsibility: emphasis on personal responsibility for change
Advice to cut down or, if indicated because of severe dependence or harm, to abstain
Menu of alternative options for changing drinking pattern
Empathic interviewing
Self-efficacy: an interviewing style which enhances this

Reproduced with permission from T.H. Bein, *et al.* (1993), Motivational interviewing with alcoholic outpatients, *Behavioural and Cognitive Psychotherapy*, **21**, 347–56, copyright 1993, with permission from Cambridge University Press.

If controlled drinking is the agreed goal, the patient and physician collaborate to monitor the amount and pattern of the drinking as follows:

1 Limit number of days of drinking and number of drinks on any occasion.
2 Slow the rate of drinking, and/or reduce alcoholic strength of drinks.
3 Develop assertiveness skills for refusing drinks.
4 Design reward system when goals are achieved.
5 Develop awareness of triggers to overdrinking.
6 Practise other ways of coping with triggers.
7 Record pattern and amount of drinking, for example in a diary.
8 Physician and patient monitor γ-glutamyl transferase blood test results.

Maintaining motivation and compliance

Enhancing motivation has a place not only at onset, but throughout the clinical contact. Treatment aimed only at enhancing motivation was for most outcome measures equal to cognitive behavioural therapy, and intensive intervention aimed at linking patients with AA.[24] Randomized controlled studies have shown the advantage of motivational interviewing over traditional supportive therapy.[10–12,25] The style of the opening interview using motivational interviewing techniques has already been discussed. The patient is encouraged not to forget the harm that drinking caused and the benefits of abstinence, but the losses and problems of being sober are not denied. Strategies for maintaining abstinence emerge from collaborative dialogue, and are owned by patients rather than offered as advice from the clinician. If medication is part of the treatment plan, unwanted effects are actively enquired into, and are recognized and not dismissed, and remedies are sought. Any discrepancies that patients reveal between their present view of themselves and how they would like to be, or between what patients say they believe and how they actually behave, are

used as a fulcrum for shifting attitudes and testing alternative strategies. These techniques were elaborated by Miller and Rollnick[3] and enshrined as motivational enhancement therapy[26] by Project MATCH (see below).

Helping motivation: the social matrix

It is said that the only successful way to change your drinking is 'to do it for yourself'. Nevertheless, many of those dependent on alcohol start on the road to recovery because of pressure from outside. For example, if the person finds himself in Court, or has lost his driving licence, authorities may seek evidence that the offender has taken steps to alter harmful drinking patterns. Perhaps the partner is now being firmer, even demand a separation or divorce; or the employer has given a warning.

Friends, partners, colleagues at work, and even employers sometimes adopt an approach that they believe to be motivating but which has the opposite effect and enables the drinker to continue drinking. For example, they may cover-up, gloss over, make excuses, or even blame themselves for what is going wrong. This cushions drinkers from experiencing the harmful consequences of their drinking or allows them to believe that alcohol is not the chief problem, despite evidence that alcohol is in fact the critical common factor in their downward spiral.

A physician can help the parties improve communication so that important messages are not lost. If the message from the employer or partner, or even the children, is clear and positive, it can have a powerful motivating effect: 'We value our relationship with you. But the way you are drinking is harming that relationship and we will not tolerate it'.

Some physicians are overcautious about confidentiality in this situation. If a doctor is asked by a partner or an employer to comment on the patient's condition, he or she may or may not have permission, or feel it appropriate, to do so. But doctors can usefully help partners or employers clarify for themselves what they want, and then encourage a clear and firm, but positive message.

Sometimes doctors unwittingly collude in a cover-up. The smokescreen that can be set-up by a drinker who is severely dependent and ambivalent about change can be hard to penetrate: 'It's depression, doctor'; 'It's stress at work'; 'If only my wife was more understanding/my sleep was not so disturbed/I didn't get these memory blanks which I think are some kind of stroke'. The doctor may need to wait for that medical moment, perhaps a crisis, to help such an individual. Or, if the doctor has patience, the drip, drip of non-judgemental evidence, and perhaps some social pressure, may bring about the necessary change in the patient's understanding and thus the perceived motivational pay-offs. Understanding may lead to action. However, as Fig. 4.2.2.4.1 shows, that action may not be sustained and the process of helping understanding may need to be repeated many times.[27]

There are few randomized controlled studies allocating patients to different intensities of external motivation. However, alcoholics coerced into treatment have medium-term outcomes similar to those who attend voluntarily.[28]

Coping skills therapies

When incentives are powerful, many newly abstinent patients are able to abstain for short periods. Others lack the skills to cope with the triggers to drinking even when their motivation to abstain has been strong. Cognitive behavioural therapies seem to improve the coping skills of these patients. If the triggers are in social situations, assertiveness, or conversation skills training can help. If the trigger is related to relationship of work problems, checking beliefs and attitudes, and reframing stressors may reduce the urge to use alcohol as a sedative.[13] Some patients are helped by learning to handle frustration and criticism without harbouring anger and resentments. Treatment can be in groups, where the opportunity to discuss these topics with others who have similar problems is appreciated. Groups also enable learning through role playing and by modelling on others. 'Relapse prevention therapy' as originally formulated has not stood up to meta-analytic critique.[11]

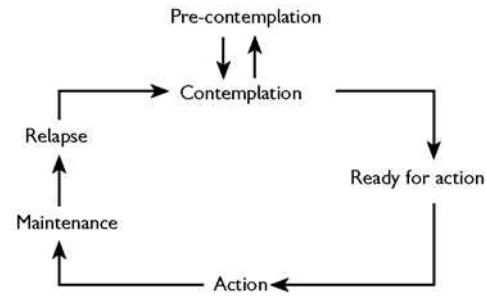


Fig. 4.2.2.4.1 Wheel of change. (Reproduced from J. Prochaska and C. DiClementi. Stages of change in the modification of problem behaviours. In *Progress in behavior modification*, Vol. 28 (eds. M. Hersen, R. Eisler, and P. Miller), pp. 183–218, copyright 1984, Sage Publishing, Sycamore, IL.

Cue exposure

The smell or sight of alcoholic drinks can be a powerful stimulus to drinking. Initial studies[30] found that 'deconditioning' by exposing inpatients to the sight and smell of their preferred drinks in a laboratory setting, without drinking, was associated in the coming 6 months with a longer period without a relapse. However, this is not a stand-alone treatment. Patients should not court danger by going into pubs and bars—these are places where people go to drink alcohol.

Couples should decide together whether or not to have alcohol in the house, but patients should not be encouraged to 'test themselves'.

Alcoholics Anonymous

There are many ingredients in the healing process of AA. Newcomers are helped to identify with others as members tell their stories. They see that it is possible to be frank about past errors and the hurt caused to others through the drinking. Telling their own story helps the members not to forget the harm that accrued from drinking. This reduces complacency, which is one of the most common precursors of relapse.

Alcoholism is viewed by AA as a physical, psychological, and spiritual illness, which can be arrested (by avoiding another drink) but cannot be cured. The meetings offer a new social network. Emotional openness is encouraged. Members learn to express warmth, and to accept that they and others have failings. The AA advice on coping with emotions and relationship difficulties has much in common with cognitive behavioural therapy and relapse prevention therapy. The method has some attractively simple concepts ('Just don't pick up that first drink'; 'HALT'—being alert to four of the most common triggers to relapse, i.e. hunger, anger,

Box 4.2.2.4.2 The 12 steps of Alcoholics Anonymous

Step 1 We admitted we were powerless over alcohol-that our lives had become unmanageable.
Step 2 Came to believe that a power greater than ourselves could restore us to sanity.
Step 3 Made a decision to turn our will and our lives over to the care of God *as we understood him*.
Step 4 Made a searching and fearless moral inventory of ourselves.
Step 5 Admitted to God, to ourselves, and to another human being the exact nature of our wrongs.
Step 6 Were entirely ready to have God remove all these defects of character.
Step 7 Humbly asked him to remove our shortcomings.
Step 8 Made a list of all persons we had harmed, and became willing to make amends to them all.
Step 9 Made direct amends to such people wherever possible, except when to do so would injure them or others.
Step 10 Continued to take personal inventory and when we were wrong promptly to admit it.
Step 11 Sought through prayer and meditation to improve our conscious contact with God *as we understood him*, praying only for knowledge of his will for us and the power to carry that out.
Step 12 Having had a spiritual awakening as a result of these steps, we tried to carry this message to practice these principles in our affairs.

loneliness, tiredness). There is a deeper aspect, which is to replace preoccupation with self by handing over to the group process, or to a 'Higher Power'.

Accepting that you are 'powerless' to control your drinking is the 'first step' in AA. This entails ceasing the struggle and letting the 'Higher Power' take over. Members vary in their interpretation of the 'Higher Power', and avowed atheists should not be deterred from sampling AA. Residential, outpatient, and day programmes, which teach the AA approach are sometimes called 12-step programmes (Box 4.2.2.4.2). One of their strengths is linking patients to the AA network. In Project MATCH,[24]

A psychiatrist can introduce patients to AA through a contact member who will tell the patient how AA works, will not ask personal details, and will extend an invitation to a meeting. Doctors are welcome to attend 'open' AA meetings to see how it works. A contact number is given in local telephone directories. AA does not work for everyone, but since it is difficult to predict who will be helped, it is good practice to offer contact to all patients with impaired control of their drinking.

A warning, often based on personal experience, may be given at AA meetings about transferring dependence from alcohol to other drugs. This usually refers to use of barbiturates or benzodiazepines, or to the danger of relying on a medication instead of adjusting one's way of living. The use of prescribed medication is not formally disapproved of by AA.

Evidence of efficacy

Naturalistic non-randomized studies have shown that treatment programmes using the AA approach are associated with outcomes in drinking and overall functioning similar to those of programmes using the cognitive behavioural approach. Patients in 12-step programmes improve on self-efficacy and coping skills scores much as patients treated by cognitive behavioural therapy.[31] Following the steps of AA is associated with improving drinking and psychosocial outcomes.[32]

Only two randomized controlled studies of 12-step programmes have been conducted. One compared inpatient treatment (with fewer hours of psychotherapy than many such programmes) with a 12-step inpatient programme (with slightly more hours of therapy). There was a non-significant trend towards a greater total abstinence programme and less relapse in the 12-step programme.[33] In Project MATCH, patients were randomly allocated to cognitive behavioural therapy, motivational enhancement therapy, or '12-step facilitation', which instructed patients in the tenets of AA, and assisted and encouraged them to attend AA meetings. The three treatments resulted in similar outcomes after 1 and 3 years. However, for those who had been relatively free of psychiatric problems at entry to the study, 12-step facilitation was associated with slightly better outcomes after 1 year. After 3 years the 12-step facilitation led to better outcome for patients who, at entry to the study, had family, social, or work environments bringing them into frequent contact with drinking.[34]

Help for the family

The family of someone with a drinking problem may suffer for years without recognition and can benefit from advice and understanding. They are a vital monitor of the patient's progress. Good family cohesion and low expressed emotion predict better outcome, even after controlling for the predictors of demographic variables and severity of alcohol dependence.

Life in the family becomes increasingly restricted. Finances dwindle. The children fear that the parent may be drunk, and so stop inviting friends to visit. They dread that arguments between mother and father will become violent. The drinker's behaviour becomes slovenly. He or she may wet the bed. Despite these hurts, the drinker may still make the family believe that they are the reason he or she drinks.

The invitation to a family member or partner to attend with the patient may be rejected if the drinker is the messenger, and the message is distorted to: 'The doctor says you're part of the problem'. A direct letter or telephone call from the clinician requesting 'your views on how I can assist' reduces the partner's fear of being burdened with extra guilt or responsibility.

The clinician can help reduce family behaviours, such as hostility or cover-up, that are damaging to the family and counterproductive for the drinker's recovery. Communication between the drinker and the spouse or children has often broken down. In many countries family groups, such as Al-Anon, provide help to families.

Behavioural marital therapy

When the patient is in a relationship, its quality can be motivating or demotivating. Reciprocal contracts are aimed at making the relationship more rewarding for each partner. Although abstinence is a prerequisite, specific agreements should not be contingent on the drinking[35]; otherwise, a relapse means that the partner ceases to work on the relationship. Another prerequisite might be that physical violence is excluded. Contracting could start thus: 'Although you

are responsible for not drinking, is there anything that your partner could do more of, or less of, that would help you stick to the plan?' Check that the requests are reasonable and available before the partner is asked to agree. The partner makes reciprocal requests and negotiation follows. Even requests for small changes can start the process.

The partners should give clear messages, owning their statements: 'This is what I would like', 'It makes me feel good if you . . . '. They will need to be reminded to state the positives and to practise being good listeners, giving non-verbal signals that they are listening, and not butting in with unsolicited good advice.

Violence in the partnership may require specific attention. If the drinker is intoxicated, the partner is advised to back off and avoid argument. Sometimes each partner is asked to sign an agreement that neither will threaten nor hit the other. If they do, time-out in another room is agreed in advance to permit slow-breathing to aid calming down, or one of them will leave the house and go to a designated place for 36 to 72 h.

When the partner 'brings up the past', this can be a major irritant to the drinker. But this can be reframed as the partner 'helping the couple not repeat their past': the partner who feels heard and understood is more ready to look at changes that he or she might also make.

Efficacy

Behavioural marital therapy produces better outcomes of drinking and marital relations than individual counselling or similar control conditions. The superior effects last for 24 months after treatment. Outcome at 1 year is better if sessions of behavioural marital therapy continue after the end of treatment to reinforce what has been learnt and rehearse relapse prevention plans.[36]

Deterrent medication

Disulfiram

If taken in a sufficient dose for at least the preceding 3 to 4 days, disulfiram causes an unpleasant reaction to develop 15 to 20 min after alcohol enters the body. The reaction is due to accumulation of acetaldehyde, an intermediate metabolite of ethanol. The reaction includes flushing, headache, pounding in the chest or head, tightness in breathing, nausea, and sometimes vomiting. Hypotension can occur and is potentially dangerous. (In some countries, calcium carbimide, which has the same action is also available.) The disulfiram–ethanol reaction varies in intensity. It is recognized practice to increase the dose of disulfiram up to 400 mg daily if the patient has tested the alcohol reaction and it has not been severe enough to act as a deterrent.

Disulfiram is an aid, not a cure. The individual can become used to life without alcohol. This allows time for confidence to recover—in the family, at work, or in the social services if there have been concerns about the safety of children. Patients may object that it is weakness to take a deterrent, and they prefer to show that they can use will power. Explain to the patient that will power is not always there when most needed. With disulfiram a decision to drink or not still has to be made, but only once a day.

Unwanted effects which occur even when no alcohol is taken include drowsiness, bad breath, and headache. These make the drug unacceptable to some patients. Concerns that disulfiram can harm the liver are based on a few case reports (the risk is about 1 in 25 000 patient-years). It appears to be a hypersensitivity reaction, and if it is to occur it is likely to be in the first month. Overall, disulfiram is associated with improved liver function tests compared with control groups, presumably owing to reduction of drinking.[37] Peripheral neuropathy (almost always reversible) has been reported following several months at doses of over 250 mg: the risk maybe greater when the patient takes other drugs such as antidepressants which are metabolized in the liver. There are a few reports of psychosis induced by disulfiram, and psychotic illness has been a formal contraindication in the licensing in some countries. The risk is so low and the need to help schizophrenic patients with alcohol problems is sometimes so great that in other countries this contraindication has been changed to a 'caution'. There are many documented cases where improvement has occurred in psychotic patients while taking disulfiram, and in a dose of up to 250 mg daily there are no problems from unwanted actions or interactions with medication for the psychiatric illness.[38]

(a) Efficacy

Disulfiram will only aid recovery if it is taken regularly in a sufficient dose to deter. Earlier studies without attempts to increase adherence to the medication did not show efficacy unlike studies in which enhanced compliance was enhanced by arranging supervision. In some of these studies there was a degree of coercion; for example, if the patient ceased taking the disulfiram the partner might withdraw from some agreed item, or disciplinary action at work might be reinstated.

The disulfiram effect depends on the patient knowing that they have ingested the disulfiram, and so only single blind studies are appropriate to test its efficacy. Single blind studies over 1 year have shown, in patients with a family member to supervise the medication, that disulfiram is associated with less relapse than acamprosate[39] and naltrexone.[40]

(b) Suggested mode of use

Before prescribing, a physical examination and baseline liver function tests are performed. The patient is encouraged to ask the partner, a nurse, or welfare officer at work or at the health centre, or a pharmacist to see that the disulfiram is taken. This can be daily, or three times a week, provided that the total weekly dose is sufficient, i.e. at least 7 × 200 mg. Some specialist clinics have follow-up clinics thrice weekly to supervise disulfiram. A programme commencing for the first months with frequent clinic attendance, and thereafter encouragement to continue using disulfiram, reported abstinence rates of over 50 per cent in patients followed for up to 7 years.[41] The product is available in a dispersible form to be taken in water so that it can be seen to be swallowed.

There should be medical follow-up, but there is no consensus as to whether monitoring of liver function tests should be carried out beyond the first month. However, monthly follow-up is appropriate to check for signs of drinking and of any unwanted effects.

It is common to prescribe disulfiram for 6 months, but many patients ask to continue for longer and there may be slips when disulfiram is withdrawn, even after long periods of abstinence. Some patients keep a supply to use when they feel an increased risk of drinking, for example on a business trip or at a social event.

Specific neurotransmitter antagonists

(a) Acamprosate (calcium acetyl homotaurinate)

Acamprosate enhances γ-aminobutyric acid (**GABA**) transmission and antagonizes glutamate transmission, probably by antagonizing *N*-methyl-D-aspartate receptors (see Chapter 6.2.8). It reduces drinking in alcohol-dependent animals, and reduces the reinstatement of drinking behaviour in animals re-exposed to alcohol after a period of abstinence. Animals do not seek out acamprosate as they do addictive substances, and it does not have mood-altering or drug-abuse potential in humans.[42] It has no deterrent or disulfiram-like effect.

Acamprosate is excreted unchanged in the kidney. It has few unwanted effects; diarrhoea, and abdominal discomfort are the only ones reported in more than 10 per cent of patients (up to 20 per cent) and these are mild and transient. It does not exacerbate psychomotor impairment caused by alcohol. There are no known drug interactions. Systematic follow-up after the end of acamprosate treatment shows no sudden relapse and no discontinuation symptoms in patients who have received the medication for up to 1 year.

(i) Efficacy

Acamprosate has a dose-related effect of improving abstinence rates in recently detoxified patients. There are no studies comparing the advantages of differing lengths of treatment. Meta-analysis of published studies finds that acamprosate is associated with improvement in abstinence rate compared to placebo with an odds ratio of 1:88 and greater cumulative days of abstinence.[43]

Acamprosate has only been tested in patients who intend to abstain from alcohol. It has not been tested formally in patients aiming for controlled drinking. However, in literature, patients who resume drinking, consume less alcohol in subsequent days[44] if they had been allocated to acamprosate than to placebo.

(ii) Suggested mode of use

Acamprosate is indicated for patients who have withdrawal symptoms and relief drinking typical of severe alcohol dependence and requiring medical assistance to withdraw. It is started 2 to 7 days after the last drink (steady state pharmacokinetics are reached after 5 days). Patients who relapse while on acamprosate are advised to continue taking the medication and exert effort to limit the lapse. However, acamprosate is not normally continued in patients who relapse more than once despite regularly taking the drug. Those who appear to be benefiting from it should continue the drug for at least 6 months, and up to 1 year if there has been a history of repeated relapsing while in treatment.

Several studies have shown that acamprosate reduces self-reported craving for alcohol. Some newly abstinent patients experience strong craving, but others experience very little and there is no evidence that this should be a criteria for deciding to whom this medication should be offered.

Acamprosate may sometimes help prolong abstinence among patients who choose to take disulfiram.[45]

(b) Opiate antagonists

Endorphins are released in one of ethanol's many acute actions on the limbic system. It has been suggested that this effect contributes to loss of control.[46] Naltrexone (and nalmefene) antagonize the neurotransmitter action of endogenous endorphins. Naltrexone has been shown to reduce ethanol-seeking in alcohol-dependent animals. It does not exacerbate the psychomotor impairment caused by alcohol.

Some patients who drink alcohol while taking naltrexone report that they feel less of the ethanol 'high'. This could lead to less impulse to carry on drinking.[47,48] However, some studies have reported an increase in total abstinence as well as a reduction of drinking overall.[48,49] It is possible that the reduced craving for alcohol and the reduced likelihood of picking up the first drink occur because the strength of the previous triggers—emotional, cognitive, or environmental—is attenuated.

Nausea following the first few doses is the commonest unwanted effect, occurring in about 10 per cent of patients. Concerns in the 1970s that naltrexone might cause dysphoria seemed to be supported by statements from heroin addicts given naltrexone to help them abstain from opiates. However, laboratory studies and randomized controlled trials in subjects who have not been opiate dependent have not found evidence of dysphoria or loss of feelings of pleasure.[50]

(i) Efficacy

Short-term administration of naltrexone reduced the rate of relapse to heavy drinking (odds ratio 0:62 in the meta-analysis of Bouza *et al.*[43] but although individual studies have reported an advantage in rates of total abstinence, this is not upheld in meta-analysis.[43]

Even though the dose is once daily, adherence has been low in some studies, and a beneficial effect only demonstrable in compliant patients.[51] Developed partly to improve compliance, a long-acting injection given monthly has become available and found to be acceptable to patients. It was more effective in reducing relapse to heavy drinking than a monthly injection of the vehicle without active naltrexone.[49]

When supervised oral naltrexone was compared to supervised disulfiram it was found to be less effective in preventing relapse to heavy drinking.[40] However, when oral naltrexone has been compared to acamprosate it was more effective.[52–54]

(ii) Suggested mode of use

Opiate antagonists have a particular role in reducing relapse to heavy drinking in patients who will not or cannot attain abstinence. As well as prescribed as a daily dose, their targeted used has also been supported in patients trying to limit the amount consumed per session, when the patient takes a dose only on days when at risk of drinking or planning to drink.[55,56] Several studies have found that patient with a positive family history of alcohol dependence are more likely to benefit from an opiate antagonist than those without.

Interactions. Opiate antagonists such as naltrexone will precipitate an immediate opiate withdrawal syndrome if given to patients who are actively dependent on opiates, and will prevent pain relief of opiate analgesics.

Helping women with alcohol problems

It has been said that when a woman has an alcohol problem, there is a man in her life with a similar problem—usually her partner or her father. When the partner also drinks heavily, he should be invited to some joint therapy meetings. Some partners have adopted a controlling role, especially if the spouse has been unreliable in

managing the children or the money, or has driven while intoxicated. The patient may allow her resentment at this to fuel her drinking, and it may need months to help her to see how this has come about.

Low self-esteem is very common in such women, even in those who were confident before the drinking became problematic. The partner, while remaining firm about the unacceptability of her drinking, may need help to be more caring and positive, to show interest in what concerns her, and to show appreciation.

When helping women with alcohol dependence to abstain some of the following may be relevant:

- Help her to stop feeling taken for granted, and to know that she has a right to set limits on what others expect of her.
- Although guilt may be proportional to what she has put her family through by her drinking, it may not help. It may prevent her from asking for the conditions at home or work that would make it easier for her to stop drinking.
- Help her let go of resentments.
- Help her find ways of recharging her batteries by, for example, taking up new interests or exercise.
- Talk with the partner, both alone and with her present. He may want to know that she acknowledges the strain on him. While still accepting complete responsibility for her drinking, she can let him know what he can do to help her.
- Self-help literature is available in many languages to help women improve self-confidence and self-assertion.[57]

Treatment of coexisting disorders

Affective disorder

Depression is common in patients who are dependent on alcohol. The drinking may have alienated friends, family, or employer, with resulting feelings of hopelessness, guilt, and lack of direction. Alcohol can reduce appetite, energy, and sexual drive. The drinker wakes in the small hours of the night feeling anxious owing to the rebound wakefulness of alcohol withdrawal. Those signs and symptoms suggesting depressive illness commonly clear with abstinence and help in tackling or tolerating personal problems and improving relationships.

Sometimes (more often in women than in men) a depressive episode precedes the alcohol dependence, the patient begins to use alcohol as self-medication. Sometimes depressive symptoms continue despite abstinence. In these cases, antidepressants should be offered in the usual way.[58,59] Relapsing alcoholism, secondary to depressive illness, is an indication for long-term antidepressants. Lithium is not a treatment for alcohol dependence itself, but is effective if alcohol dependence is secondary to manic–depressive disorder.

General practitioners and general psychiatrists often prescribe antidepressants to patients with alcohol dependence who are still drinking, because the patient has complained of low mood, insomnia, or anxiety. There is no evidence that this will improve the drinking problem, and the period of alcohol withdrawal under benzodiazepine cover can be an occasion to withdraw the antidepressant. Most depressive symptoms experienced while alcohol-dependent patients are drinking are alleviated with abstinence. Early-onset alcohol dependence, marked by novelty seeking and impulsivity, can be exacerbated by SSRIs.[60,61]

Anxiety and panic disorder

Some patients have had panic attacks for years before discovering that alcohol can end or prevent them. Others have a first panic attack during alcohol withdrawal, but the attacks continue independently even during sustained abstinence. In this case, cognitive behavioural therapy and/or medication are indicated. Anxiety symptoms, which persist are predictive of relapse in the coming year.[62] However, the majority of anxiety symptoms reported by alcohol-dependent patients resolve with abstinence[63,64] and the weight of evidence is that adding specific psychological therapy aimed at the anxiety symptoms does not improve the drinking or the anxiety outcomes beyond that achieved by the treatment for the alcohol dependence.[65,66] In Project MATCH, male patients with social phobia allocated to 12-step facilitation (i.e. encouragement to attend AA) improved their drinking as much or even slightly more than those patients allocated to cognitive behaviour therapy (CBT) who would have received specific treatment for their phobia, though an advantage to CBT showed in female socially phobic patients.[67]

One explanation for these findings could be that attending to the anxiety might, for some patients, distract attention from the drinking, or could even seem to 'justify' their continuing to drink. It is also the case that some phobic patients report that attending AA helped them to overcome their social phobia.

Three studies suggest that the serotonin agonist buspirone can help reduce both drinking and anxiety.[68] Tricyclic antidepressants and selective serotonin-reuptake inhibitors (**SSRIs**) are prescribed to patients whose anxiety disorder persists despite abstinence. Some patients with long histories of alcohol dependence and severe panic disorder fail to respond to these medications or to CBT. For these patients the risk of complications from a prescription for a long-acting benzodiazepine such as chlordiazepoxide may be less than the harm that might accrue if bouts of excessive drinking persisted. If prescribed (and to do so is controversial), the benzodiazepine should be dispensed in limited amounts. The prescription should be conditional on abstinence from alcohol, perhaps aided by disulfiram, if necessary. 'As-required' use (e.g. for travelling on public transport) helps to limit the development of tolerance, even though in theory it may perpetuate phobic beliefs. This method probably commits the patient to long-term use and an enduring risk of escalation.

Treating alcohol-dependent patients with antipsychotic medication when there is no psychotic illness may increase their drinking and should be avoided.[69,70]

Residential and inpatient treatment

It is debatable whether a period of inpatient treatment can improve the eventual outcome. Some studies have compared outcomes after patients have been randomly allocated to either inpatient or outpatient treatment. Usually no difference has been found. However, the interpretation of these results and their extrapolation to clinical reality has been debated. Finney *et al.*[71] concluded that the studies often lacked statistical power. Furthermore, the more seriously affected patients had sometimes been excluded before randomization.[72,73] While evidence that it is inpatient treatment rather than intensity of treatment which improves outcome is lacking,[74,75] admission to hospital can provide valuable respite for the drinker and the family when life is severely disorganized because

drinking is out of control. Perhaps such respite need not be offered in a relatively expensive medical environment. However, if the patient has become suicidal as difficulties increase or has developed serious medical complications, then hospital admission may be indicated, ideally to specialized facilities. Longer stays in hospital are not supported by research. For example, Trent[76] found no evidence of worse outcome when the United States Navy reduced the length of its inpatient alcoholism treatment programme from 6 to 4 weeks. The role of inpatient treatment is considered further in Chapter 4.2.2.5.

Matching patients to treatments

It is recognized that people with alcohol dependence present a range of problems, come from various backgrounds, and have different personality characteristics. Some have no accompanying emotional disturbance; others have a psychiatric disorder. The poor outcomes of treatment for alcohol dependence have been attributed to their use with unsuitable patient, and better matching of patients to treatments has been sought. A North American study of 1726 outpatients (Project MATCH) set out to test hypotheses about matching treatments to patients. Three treatments were studied, each established in previous randomized controlled trials as more effective than 'supportive therapy': motivational enhancement therapy, cognitive behavioural therapy, and instruction in the AA approach with encouragement to take part in AA meetings ('12-step facilitation').

Few matching effects reached statistical significance. In patients recruited from outpatient clinics, those who scored high on anger at initial assessment averaged 85 per cent of abstinent days if they had been allocated to motivational enhancement therapy compared with 75 per cent if they had been allocated to 12-step facilitation or cognitive behavioural therapy.[77] In the first year of follow-up, patients with initially less severe psychiatric symptoms had more abstinent days after the 12-step facilitation than after cognitive behavioural therapy. Patients with critically high psychiatric severity did no better with cognitive behavioural therapy.[77]

Another marker of who benefits most from AA emerged in the 3-year Project MATCH data. Patients who came from a social milieu where they mixed a lot with other drinkers owing to family, neighbourhood, or work influences did better if they had received 12-step facilitation than with either cognitive behavioural therapy or motivational enhancement therapy.[34]

There are several reasons for the absence of evidence of other powerful predictors of treatment outcome in the Project MATCH data. Perhaps the key behaviour—not taking the first drink—can be arrived at in different ways.

Some clinical situations

Morbid jealousy

This is discussed in Chapter 4.4.

The homeless alcohol-dependent person

It is difficult to conduct randomized controlled studies with adequate follow-up to test the efficacy of interventions to reduce drinking and improve social conditions for the homeless, and few answers have been found. A brief hospital admission to 'dry out' and assessment for transfer to residential care may result in transient improvement in physical health and is more humane than prison. However, supporting evidence is lacking. The structured intensive outpatient intervention, 'community reinforcement approach', has been shown in a North American study to reduce drinking (corroborated by improvement in serum γ-glutamyl transferase) and increase the number of clients at work and in satisfactory housing.[78] The community reinforcement approach combined an offer of free housing, a place at a 'job club' to assist with finding employment, training in problem-solving skills, communication, goal-setting, refusal of drinks, and independent living. Patients had access to an alcohol-free social club. The housing offer was contingent on sobriety and some evidence of saving money. Continuation in the housing was contingent on sobriety checked by breathalyser. Disulfiram had been shown to improve the effects of the community reinforcement approach.[15,79]

Young people

There is a dearth of evaluation of programmes to help young people with alcohol problems. AA groups may have teenage members. When education or employment is in jeopardy, young people may accept disulfiram, supervised perhaps by the family. However, without the support of a non-drinking peer group (which they would have in AA), most young people will try again and again to resume 'social drinking'. Job or marriage commitments sometimes alter the pay-off matrix sufficiently for recovery to be sustained. Otherwise, it may not be until age 30 that the young person is sufficiently convinced that he or she cannot control drinking and takes serious steps to seek help.

Employment referrals

It is common for individuals to seek help when their drinking has put their job in jeopardy. Having a job helps recovery, and for the person to lose employment while paying only lip-service to treatment is common and disheartening for all. The psychiatrist should find out whether disciplinary procedures are in motion or threatened. It can be helpful if the psychiatrist and the patient are told this directly by the employer. If the consultation is part of an undertaking under a company 'alcohol and drugs policy', the patient may have given permission for the psychiatrist to answer the employer's request to know whether he or she is attending and following advice.

Patients who are on the point of dismissal may offer to take disulfiram supervised in the company's occupational health or welfare department. This can bring about recovery and employment for as long as the threat of dismissal remains, and sometimes afterwards.[80]

The liver transplant candidate

Some transplant centres require a demonstration of months of abstinence, to show commitment, before offering transplant to a patient with alcoholic liver disorder. Other centres have no such restrictions. From 6 to 80 per cent of transplant recipients, varying between centres, have recommenced drinking and exceeded safe limits by the end of the first year. Their eventual outcome in terms of quality of life and psychiatric health is no worse than for other transplant patients, and there is no evidence to support demanding lengthy preoperative abstinence. However, patients who relapse to problematic drinking are more likely to have had a history of definite alcohol dependence, and/or depressive illness.[81,82]

Physicians as patients

Alcohol dependence is commoner among doctors than among most other occupational groups, other than those groups who are employed in the alcohol beverage manufacturing or retailing. Doctors' outcome, once in treatment, tends to be good if they can return to their practice. This is probably partly due to the requirement by the licensing body that 'impaired physicians' accept monitoring by an independent specialist to corroborate that they are following advice and continuing to progress.[83,84]

Doctors' reluctance to accept help for their illnesses, and their tendency to treat themselves, is well known and true for substance misuse. Initial denial often means that problems escalate until there are disciplinary or Court proceedings and attempts to treat their own alcohol dependence may result in dependence on other substances. In some instances, where there is any risk to safety of the doctor's patients, the professional licensing body should be informed if not already involved.

The alcoholic doctor should be treated in the same way as a lay person. The partner should be invited to the interview. Ideally, information should be obtained from the employer or from a colleague about the nature of any problems at work or any disciplinary action, actual or threatened.

In some countries there are support groups for recovering doctors and dentists who meet together and are ready to offer advice and encouragement to individuals and their families.

Follow-up

Systematic follow-up has been shown to improve outcomes.[17,85] Early detection of relapse is important, and is aided by regular contact with the family or the workplace, a breathalyser test at interview, and tests for blood markers of drinking (γ-glutamyl transferase or carbohydrate-deficient transferrin).[86] Objective markers are required when a patient requests a report for a Court, the driving licence authority, or an employer.

Some guiding principles

Research into alcoholism spanning 50 years has shown that the attitudes of the agency and the therapist influence patients' outcome, as they may do for many illnesses. The therapeutic alliance is a strong predictor of outcome in the treatment of alcohol dependence.[87] However, agencies must also be prepared to set limits on drunken behaviour at the clinic and telephone calls when intoxicated. And for patients who repeatedly relapse, resumption of treatment should sometimes be made conditional on complying with a new treatment plan, such as supervision of medication.[88]

Showing respect, enhancing dignity, conveying accurate empathy, adopting objective and not moral criteria, involving the family, and reducing hurdles to seeking help have been shown to improve compliance, and often outcome, for alcohol dependence.

Further information

http://www.niaaa.nih.gov/Publications/AlcoholResearch/

References

1. Vaillant, G. (1997). *The natural history of alcoholism revisited*. Harvard University Press, Boston, MA.
2. O'Brien, C.P. and McLellan, A.T. (1996). Myths about the treatment of addiction. *Lancet*, **347**, 237–40.
3. Miller, W.R. and Rollnick, N. (2002). *Motivational interviewing* (2nd edn). Guilford Press, New York.
4. Hayashida, M., Alterman, A.I., McLellan, T., *et al.* (1989). Comparative effectiveness and costs of inpatient and outpatient detoxification with mild-moderate alcohol withdrawal syndrome. *The New England Journal of Medicine*, **320**, 358–65.
5. Bennie, C. (1998). A comparison of home detoxification and minimal intervention strategies for problem drinkers. *Alcohol and Alcoholism*, **33**, 157–63.
6. Mayo–Smith, M.F., for the American Society for Addiction Medicine Working Group. (1997). Pharmacological management of alcohol withdrawal: a meta-analysis and evidence-based practice guideline. *The Journal of the American Medical Association*, **278**, 144–61.
7. Foy, A., March, S., and Drinkwater, V. (1988). Use of an objective clinical scale in the assessment and management of alcohol withdrawal in a large general hospital. *Alcoholism, Clinical and Experimental Research*, **12**, 360–4.
8. Thomson, A.D. and Marshall, J. (2006). The treatment of patients at risk of developing Wernicke's encephalopathy in the community. *Alcohol and Alcoholism*, **41**, 159–67.
9. Vaillant, G.E. (1991). An alternative to psychotherapy. In *The international handbook of addiction behaviour* (ed. I.B. Glass), pp. 236–9. Routledge, London.
10. Raistrick, D., Heather, N., and Godfrey, C. (2006). *Review of the effectiveness of treatment for alcohol problems* (*full report for the* (*English*)) *NHS: National Treatment Agency for Substance Misuse*, http://www.nta.nhs.uk/publications/documents/nta_review_of_the_effectiveness_of_treatment_for_alcohol_problems_fullreport
11. Slattery, J., Chick, J., Craig, J., *et al.* (2003). *Prevention of relapse in alcohol dependence*, www.docs.scottishmedicines.org/docs/pdf/Alcohol%20Report.pdf
12. Berglund, M., Thelander, S., Salaspuro, M., *et al.* (2003). Treatment of alcohol abuse: an evidence-based review. *Alcoholism, Clinical and Experimental Research*, **27**, 1645–56.
13. Monti, P.M., Kadden, R.M., Rohsenow, D.J., *et al.* (2002). *Treating alcohol dependence: a coping skills training guide* (2nd edn). Guilford Press, New York.
14. Miller, W.R., Benfield, R.G., and Tonnegan, J.S. (1993). Enhancing motivation for change in problem drinkers: a comparative outcome study of three controlled drinking therapies. *Journal of Consulting and Clinical Psychology*, **61**, 455–61.
15. Azrin, N.H. (1976). Improvements in the community reinforcement approach to alcoholism. *Behavioural Research and Therapy*,**14**, 339–48.
16. O'Farrell, T.J., Choquette, K.A., and Cutter, H.S.G. (1998). Couples relapse prevention sessions after behavioural marital therapy for male alcoholics: outcomes during the three years after starting treatment. *Journal of Studies on Alcohol*, **59**, 357–70.
17. Chick, J., Connaughton, J., Ritson, B., *et al.* (1988). Advice versus extended treatment for alcoholism: a controlled study. *British Journal of Addiction*, **83**, 159–70.
18. Helzer, J.E., Robins, L.N., Taylor, J.R., *et al.* (1985). The extent of long-term moderate drinking among alcoholics discharged from medical and psychiatric treatment facilities. *The New England Journal of Medicine*, **312**, 1678–82.
19. Vaillant, G.E. (1996). A long-term follow-up of male alcohol abuse. *Archives of General Psychiatry*, **53**, 243–9.
20. Rychtarik, R.G., Foy, D.W., Scott, T., *et al.* (1987). Five year follow-up of broad spectrum behavioral treatment for alcoholism: effects of training controlled drinking skills. *Journal of Consulting and Clinical Psychology*, **55**, 106–8.
21. Sanchez-Craig, M., Leigh, G., Spivak, K., *et al.* (1989). Superior outcome of females over males after brief intervention for the reduction of heavy drinking. *British Journal of Addiction*, **84**, 395–404.

22. WHO Brief Intervention Group. (1996). A cross-national trial of brief intervention with heavy drinkers. *American Journal of Public Health*, **86**, 948–55.
23. Bien, T.H., Miller, W.R., and Tonigan, J.S. (1993). Brief interventions for alcohol problems: a review. *Addiction*, **88**, 315–36.
24. Project MATCH Research Group. (1997). Matching alcoholism treatments to client heterogeneity: project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol*, **58**, 7–29.
25. Bien, T.H., Miller, W.R., and Boroughs, J.M. (1993). Motivational interviewing with alcoholic out-patients. *Behavioural and Cognitive Psychotherapy*, **21**, 347–56.
26. Miller, W.R., Zweben, A., DiClementi, C.C., *et al.* (1992). *Motivational enhancement therapy manual: a clinical research guide for therapists treating individuals with alcohol abuse and dependence.* NIAAA Project MATCH Monograph Series, Vol. 2. DHHS Publication No. (ADM) 92-1894. US Government Printing Office, Washington, DC.
27. Prochaska, J. and Di Clemente, C. (1992). Stages of change in the modification of problem behaviors. In *Progress in behavior modification*, Vol. 28 (eds. M. Hersen, R. Eisler, and P. Miller), pp. 183–218. Sycamore Publishing, Sycamore, IL.
28. Chick, J. (1998). Treatment of alcoholic violent offenders—ethics and efficacy. *Alcohol and Alcoholism*, **33**, 20–5.
29. Hall, S.M., Havassy, B.E., and Wasserman, D.A. (1990). Commitment to abstinence and acute stress in relapse to alcohol, opiates and nicotine. *Journal of Consulting and Clinical Psychology*, **58**, 175–81.
30. Drummond, C. and Glautier, S. (1994). A controlled trial of cue exposure treatment in alcohol dependence. *Journal of Consulting and Clinical Psychology*, **62**, 809–17.
31. Oiumette, P.C., Finney, J.W., and Moos, R.H. (1997). Twelve-step and cognitive–behavioural treatment for substance abuse: a comparison of treatment effectiveness. *Journal of Consulting and Clinical Psychology*, **65**, 230–40.
32. Suire, J.G. and Bothwell, R.K. (2006). The psychosocial benefits of alcoholics anonymous. *American Journal of Addiction*, **15**, 252–5.
33. Keso, L. and Salaspuro, M. (1990). In-patient treatment of employed alcoholics: a randomised clinical trial of Hazelden-type and traditional treatment. *Alcoholism, Clinical and Experimental Research*, **14**, 584–9.
34. Longabough, R., Wirtz, P.W., Zweben, A., *et al.* (1998). Network support for drinking: alcoholics anonymous and long-term matching effects. *Addiction*, **93**, 1313–34.
35. O'Farrell, T.J. (1995). Marital and family therapy. In *Handbook of alcoholism treatment approaches* (eds. R.K. Hester and W.R. Miller), pp. 195–220. Allyn and Bacon, Boston, MA.
36. O'Farrell, T.J., Choquette, K.A., and Cutter, H.S.G. (1998). Couples relapse prevention sessions after behavioural marital therapy for male alcoholics: outcomes during the three years after starting treatment. *Journal of Studies on Alcohol*, **59**, 357–70.
37. Chick, J. (1998). Safety aspects of disulfiram in the treatment of alcohol dependence. *Drug Safety*, **20**, 427–35.
38. Larson, E.W., Olincy, A., Rummans, T.A., *et al.* (1992). Disulfiram treatment of patients with both alcohol dependence and other psychiatric disorders: a review. *Alcoholism, Clinical and Experimental Research*, **16**, 125–30.
39. De Sousa, A. and De Sousa, A. (2005). An open randomized study comparing disulfiram and acamprosate in the treatment of alcohol dependence. *Alcohol and Alcoholism*, **40**, 545–8.
40. De Sousa, A. and De Sousa, A. (2004). A one-year pragmatic trial of naltrexone versus disulfiram in the treatment of alcohol dependence. *Alcohol and Alcoholism*, **39**, 528–31.
41. Krampe, H., Stawicki, S., Wagner, T.Y., *et al.* (2006). Follow-up of 180 alcoholic patients for up to 7 years after outpatient treatment: impact of alcohol deterrents on outcome. *Alcoholism, Clinical and Experimental Research*, **30**, 86–95.
42. Littleton, J. (1995). Acamprosate in alcohol dependence: how does it work? *Addiction*, **90**, 1179–88.
43. Bouza, C., Magro, A., Munoz, A., *et al.* (2004). Efficacy and safety of naltrexone and acamprosate in the treatment of alcohol dependence: a systematic review. *Addiction*, **99**, 811–28.
44. Chick, J., Lehert, P., and Landron, F. (2003). Does acamprosate improve reduction of drinking as well as aiding abstinence? *Journal of Psychopharmacology*, **17**, 387–92.
45. Besson, J., Aeby, F., Kasas, A., *et al.* (1998). Combined efficacy of acamprosate and disulfiram in the treatment of alcoholism: a controlled study. *Alcoholism, Clinical and Experimental Research*, **22**, 573–9.
46. Gianoulakis, C., Krishnan, B., and Thavundayil, J. (1996). Enhanced sensitivity of pituitary β-endorphin to ethanol in subjects at high risk of alcoholism. *Archives of General Psychiatry*, **53**, 250–7.
47. Volpicelli, J.R., Alterman, A.I., Hayashida, M., *et al.* (1992). Naltrexone in the treatment of alcohol dependence. *Archives of General Psychiatry*, **49**, 876–80.
48. O'Malley, S., Jaffe, A.J., Chang, G., *et al.* (1992). Naltrexone and coping skills therapy for alcohol dependence. *Archives of General Psychiatry*, **49**, 881–7.
49. Garbutt, J.C., Kranzler, H.R., O'Malley, S.S., *et al.* (2005). Efficacy and tolerability of long-acting injectable naltrexone for alcohol dependence: a randomized controlled trial. *The Journal of the American Medical Association*, **293**, 1617–25. [Erratum in *JAMA* 2005;293:1978; *JAMA* 2005;293:2864.]
50. Doty, P. and de Wit, H. (1995). Effects of naltrexone pretreatment on the subjective and performance effects of ethanol in social drinkers. *Behavioural Pharmacology*, **6**, 386–94.
51. Baros, A.M., Latham, P.K., Moak, D.H., *et al.* (2007). What role does measuring medication compliance play in evaluating the efficacy of naltrexone? *Alcoholism, Clinical and Experimental Research*, **31**, 596–603.
52. Anton, R.F., O'Malley, S.S., Ciraulo, D.A., *et al.* (2006). Combined pharmacotherapies and behavioral interventions for alcohol dependence: the COMBINE Study: a randomized controlled trial. *The Journal of the American Medical Association*, **295**, 2003–17.
53. Morley, K.C., Teesson, M., Reid, S.C., *et al.* (2006). Naltrexone versus acamprosate in the treatment of alcohol dependence: a multi-centre, randomized, double-blind, placebo-controlled trial. *Addiction*, **101**, 1451–62.
54. Kiefer, F., Jahn, H., Tarnaske, T., *et al.* (2003). Comparing and combining naltrexone and acamprosate in relapse prevention of alcoholism. *Archives of General Psychiatry*, **60**, 92–99.
55. Heinala, P., Alho, H., Kiianmaa, K., *et al.* (2001). Targeted use of naltrexone without prior detoxification in the treatment of alcohol dependence: a factorial double-blind, placebo-controlled trial. *Journal of Clinical Psychopharmacology*, **21**, 287–92.
56. Karhuvaara, S., Simojoki, K., Virta, A., *et al.* (2007). Targeted nalmefene with simple medical management in the treatment of heavy drinkers: a randomized double-blind placebo-controlled multicenter study. *Alcoholism, Clinical and Experimental Research*, **31**, 1179–87.
57. Jeffers, S. (1987). *Feel the fear and do it anyway*. Century Hutchinson, London.
58. McGrath, P.J., Nunes, E.V., Stewart, J.W., *et al.* (1996). Imipramine treatment of alcoholics with primary depression: a placebo controlled clinical trial. *Archives of General Psychiatry*, **53**, 232–40.
59. Cornelius, J.R., Salloun, I.M., Ehler, J.G., *et al.* (1997). Fluoxetine reduced depressive symptoms and alcohol consumption in patients with comorbid major depression and alcohol dependence. *Archives of General Psychiatry*, **54**, 700–5.
60. Kranzler, H.R., Burleson, J.A., Brown, J., *et al.* (1996). Fluoxetine treatment seems to reduce the beneficial effects of cognitive-behavioral therapy in type B alcoholics. *Alcoholism, Clinical and Experimental Research*, **20**, 1534–41.

61. Chick, J., Aschauer, H., and Hornik, K. (2004). Efficacy of fluvoxamine in preventing relapse in alcohol dependence: a one-year, double blind, placebo-controlled multicentre study with analysis by typology. *Drug and Alcohol Dependence*, **74**, 61–70.
62. Willinger, U., Lenzinger, E., Hornik, K., *et al.* (2002). Anxiety as a predictor of relapse in detoxified alcohol-dependent patients. *Alcohol and Alcoholism*, **37**, 609–12.
63. Driessen, M., Meier, S., Hill, A., *et al.* (2001). The course of anxiety, depression and drinking behaviours after completed detoxification in alcoholics with and without comorbid anxiety and depressive. *Alcohol and Alcoholism*, **36**, 249–55.
64. Allan, C.A., Smith, I., and Mellin, M. (2002). Changes in psychological symptoms during ambulant detoxification. *Alcohol and Alcoholism*, **37**, 241–4.
65. Bowen, R.C., D'Arcy, C., Keegan, D., *et al.* (2000). A controlled trial of cognitive behavioral treatment of panic in alcoholic inpatients with comorbid panic disorder. *Addictive Behaviors*, **25**, 593–7.
66. Randall, C.L., Thomas, S., and Thevos, A.K. (2001). Concurrent alcoholism and social anxiety disorder: a first step toward developing effective treatments. *Alcoholism, Clinical and Experimental Research*, **25**, 210–20.
67. Thevos, A.K., Roberts, J.S., Thomas, S.E., *et al.* (2000). Cognitive behavioral therapy delays relapse in female socially phobic alcoholics. *Addictive Behaviors*, **25**, 333–45.
68. Kranzler, H.R., Burleson, J.A., Boca, F.K., *et al.* (1994). Buspirone treatment of anxious alcoholics. *Archives of General Psychiatry*, **51**, 720–31.
69. Guardia, J., Segura, L., Gonzalvo, B., *et al.* (2004). A double-blind, placebo-controlled study of olanzapine in the treatment of alcohol-dependence disorder. *Alcoholism, Clinical and Experimental Research*, **28**, 736–45.
70. Wiesbeck, G.A., Weijers, H.-G., Lesch, O.M., *et al.* (2001). Flupenthixol decanoate and relapse prevention in alcoholics: results from a placebo-controlled study. *Alcohol and Alcoholism*, **36**, 329–34.
71. Finney, J., Hahn, A.C., and Moos, R.H. (1996). The effectiveness of inpatient and outpatient treatment for alcohol abuse: the need to focus on mediators and moderators of setting effect. *Addiction*, **91**, 1773–96.
72. Schuckit, M. (1998). Penny-wise, ton-foolish? The recent movement to abolish inpatient alcohol and drug treatment. *Journal of Studies on Alcohol*, **59**, 5–6.
73. Mattick, R.P. and Jarvis, T. (1994). Inpatient setting and long duration for the treatment of alcohol dependence? Outpatient care is as good. *Drug and Alcohol Review*, **13**, 127–35.
74. Annis, H.M. (1996). Inpatient versus outpatient setting effects in alcoholism treatment: revisiting the evidence. *Addiction*, **91**, 1804–7.
75. Rychtarik, R.G., Connors, G.J., Whitney, R.B., *et al.* (2000). Treatment settings for persons with alcoholism: evidence for matching clients to inpatient versus outpatient care. *Journal of Consulting and Clinical Psychology*, **68**, 277–89.
76. Trent, L.K. (1998). Evaluation of a four—versus six-week length of stay in the Navy's alcohol treatment program. *Journal of Studies on Alcohol*, **59**, 270–9.
77. Project MATCH Research Group. (1997). Project MATCH secondary a priori hypotheses. *Addiction*, **92**, 1671–98.
78. Smith, J.E., Meyers, R.I., and Delaney, H.D. (1998). The community reinforcement approach with homeless alcohol-dependent individuals. *Journal of Consulting and Clinical Psychology*, **66**, 541–8.
79. Azrin, N.H., Sisson, R.W., Meyers, R., *et al.* (1982). Alcoholism treatment by disulfiram and community reinforcement therapy. *Journal of Behavior Therapy and Experimental Psychiatry*, **13**, 105–12.
80. Robichaud, C., Strickland, D., Bigelow, G., *et al.* (1979). Disulfiram maintenance employee alcoholism treatment: a three phase treatment. *Behaviour Research and Therapy*, **17**, 618–21.
81. Kelly, M., Chick, J., Gribble, R., *et al.* (2006). Predictors of relapse to harmful alcohol consumption after orthotopic liver transplantation. *Alcohol and Alcoholism*, **41**, 278–83.
82. McCallum, S. and Masterton, G. (2006). Liver transplantation for alcoholic liver disease: a systematic review of psychosocial selection criteria. *Alcohol and Alcoholism*, **41**, 358–63.
83. Shore, J.H. (1987). The Oregon experience with impaired physicians: an eight year follow-up. *The Journal of the American Medical Association*, **257**, 2931–4.
84. Fowlie, D.G. (1999). The misuse of alcohol and other drugs by doctors: a UK report and one region's response. *Alcohol and Alcoholism*, **34**, 666–71.
85. Ahles, T.A., Schlundt, D.G., Prue, D.M., *et al.* (1983). Impact of aftercare arrangements on the maintenance of treatment success in abusive drinkers. *Addictive Behaviours*, **8**, 53–8.
86. Reynauld, M., Hourcade, F., Planche, F., *et al.* (1998). Usefulness of carbohydrate–deficient transferrin in alcoholic patients with normal gamma-glutamyl transferase. *Alcoholism, Clinical and Experimental Research*, **22**, 615–18.
87. Connors, G.J., Carroll, K.M., DiClemente, C.C., *et al.* (1997). The therapeutic alliance and its relationship to alcoholism treatment participation and outcome. *Journal of Consulting and Clinical Psychology*, **65**, 588–98.
88. Sereny, G., Sharma, V., Holt, S., *et al.* (1986). Mandatory supervised antabuse therapy in an out-patient alcoholism program: a pilot study. *Alcoholism, Clinical and Experimental Research*, **10**, 290–2.

4.2.2.5 Services for alcohol use disorders

D. Colin Drummond

A spectrum of disorders needing a range of services

The provision of services for alcohol use disorders has been driven by the prevailing view of their nature and prevalence. Following the Second World War, the disease concept of alcoholism gained increasing support in both the United States (US) and the United Kingdom (UK).[1] According to this concept, alcoholism is an all-or-nothing phenomenon affecting a relatively small subgroup of the population, and requires intensive specialist treatment. In the UK this led to the development of specialist alcohol treatment centres with an emphasis on intensive inpatient treatment involving group therapy, often with close affiliation to the Alcoholics Anonymous (**AA**) fellowship. Such programmes tended to be targeted at relatively socially stable men, and catering for the more severely alcohol dependent.[2]

In the 1970s and 1980s came a recognition that there existed a much wider range of alcohol-related problems in the population than would meet the narrow criteria of alcoholism or alcohol dependence, but which might nevertheless benefit from intervention. Research began to show that alcohol problems existed on a continuum of severity and thus might not necessarily require intensive specialist treatment with a lifelong goal of complete abstinence from alcohol. Screening and brief intervention with presymptomatic heavy drinkers in the primary care or general hospital medical ward setting could be effective in reducing excessive alcohol consumption and alcohol-related harm.[3,4] This led to the proposal that greater benefit could be accrued from less intensive

approaches aimed at the large number of hazardous drinkers, than more intensive and expensive interventions catering for the minority of very heavy drinkers: the 'preventive paradox'.[5]

In a ground-breaking report, the US Institute of Medicine advocated 'broadening the base of treatment for alcohol problems'.[6] Recognizing the potential for increased prevention and treatment activity in health care personnel without specialist addiction training (e.g. general practitioners, physicians, social workers), and the limitations of expanding specialist treatment given the high prevalence of alcohol misuse, the report emphasized the need for an expanded range of locations and methods of intervention, across the spectrum of alcohol use disorders (see Fig. 4.2.2.5.1). Importantly however, the report also recognized that alcohol use disorders are heterogeneous, and different types of disorders are likely to require different types or intensities of treatment, that is, the need to match treatments to the nature of the presenting problem.

Since this report there has been some progress made towards increasing the range and accessibility of treatment. However, in some cases this has been disappointing. This chapter describes the range and organization of treatment approaches and explores the barriers to implementation of a comprehensive system of care for alcohol use disorders. The evidence suggests that we have a long way to go to deliver an optimal level of access to alcohol treatment for those in need. The evidence on the cost-effectiveness of alcohol treatment is discussed, and consideration is given to the needs of special groups in the population who may find access to treatment more difficult. The main conclusion is that on the basis of the existing research evidence there remain considerable opportunities to expand and improve treatment services for alcohol use disorders. This will require further training and dissemination initiatives and the political will and funding to achieve this throughout the health system.

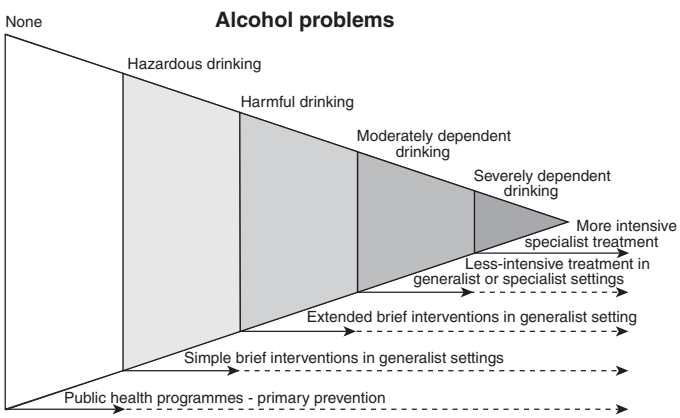


Fig. 4.2.2.5.1 A spectrum of responses to alcohol problems. (Reprinted from D. Raistrick, N. Heather, and C. Godfrey, (2006), *Review of the effectiveness of treatment for alcohol problems*, National Treatment Agency for Substance Misuse, London © 2001–2008 National Treatment Agency.) The triangle represents the general population, with the spectrum of alcohol problems experienced by the population shown along the upper side of the figure. Responses to these problems are shown along the lower side. The dotted lines suggest that primary prevention, simple brief intervention, extended brief intervention and less-intensive treatment may have effects beyond their main target area. Although the figure is not drawn to scale, the prevalence in the population of each of the categories of alcohol problem is approximated by the area of the triangle occupied; most people have no alcohol problems, a very large number show risky consumption but no current problems, many have risky consumption and less serious alcohol problems, some have moderate dependence and problems and a few have severe dependence or complicated alcohol problems.

Location and intensity of treatment

Brief interventions

There has been considerable research interest in the potential of brief interventions in primary care, and to a lesser extent in the general hospital.[7] There are several potential advantages in conducting treatment interventions in primary care. Patients with alcohol use disorders consult their general practitioner more frequently than other patients. Excessive drinkers identified by screening in primary care are largely at an earlier stage in their drinking career and are potentially more likely to benefit from brief early intervention than more severely dependent drinkers presenting to specialist treatment services. Further, primary care is often seen as less stigmatizing than a specialist clinic. Brief interventions typically involve opportunistic screening using tools such as the AUDIT questionnaire[8] or other shorter variants,[9] followed by 5–30 min. of brief intervention conducted by a practitioner who does not have training in specialist alcohol treatment.

Several studies have demonstrated the effectiveness of screening and brief intervention in hazardous drinkers in primary care. In a large randomized controlled trial, Wallace *et al.*[4] found that brief intervention was more effective than a control treatment in reducing alcohol consumption and γ-glutamyl transferase at 1-year follow-up. Similar findings were obtained in a large World Health Organization multicentre trial.[10]

Fewer screening and brief intervention studies have been conducted in the general hospital setting. A recent meta-analysis of this literature showed no difference between intervention and control in this setting.[11] However, some recent studies have shown benefits of screening and brief intervention in accident and emergency departments. One UK study showed reduced alcohol consumption and fewer reattendances in A&E attenders identified by screening and referred to an alcohol health worker.[12] Another UK study in young men with alcohol-related facial injuries found that brief intervention significantly reduced alcohol consumption and alcohol-related problems.[13]

Meta-analyses of brief interventions have mostly found advantages of brief intervention over control treatments with effect sizes of 10–20 per cent on reduced alcohol consumption at 1 year.[14–16] Some earlier reviews concluded that brief interventions are at least as effective as more intensive specialist treatments. However, the populations studied in brief intervention trials are heterogeneous.[17] Most trials have been conducted with opportunistic screening in non-treatment seeking populations in primary care. A smaller number have compared brief interventions to more intensive treatments in specialist alcohol treatment settings. A recent review found that brief interventions are effective only in less severe alcohol disorders in the context of opportunistic screening.[15]

There are barriers to implementation of brief intervention in non-specialist settings, which may limit its effectiveness. In a UK national survey,[18,19] it was found that general practitioners and

primary care practice nurses were reluctant to engage in screening and brief interventions because of a perceived lack of training and support to carry-out this work. Effective implementation of large-scale screening and brief intervention programmes will require attention to the training and support needs of non-specialist personnel. Further, screening programmes will identify more severely alcohol dependent drinkers who may not respond to brief interventions alone. Thus, effective working arrangements between generalists and specialists are needed. Drummond[20] has also questioned the generalizability of brief intervention research findings in the typical clinical setting, given the large number of exclusions in research studies and a lack of pragmatic trials.

Specialist treatment in inpatient settings

The main treatment response to alcohol use disorders continues to be delivered by specialists, although this is mostly delivered in the community rather than in inpatient settings. There has been extensive research on the location and intensity of specialist treatment. An early influential study was that of Edwards *et al.*[21] in which 100 alcohol-dependent men referred to the Maudsley Hospital in London were randomized to receive either intensive specialist treatment, including specialist inpatient care, or a single session of counselling. At 1-year follow-up there was no difference in outcome between the two treatments. It was concluded that the reliance on intensive treatments up to that time was called into question by the findings. This controversial study gave rise to considerable debate and several studies have subsequently investigated the same issues. Another British study attempted to replicate the Edwards study and found only modest differences between advice only and extended treatment in a randomized controlled trial at 2 years' follow-up.[22] There were, however, no differences between treatments in abstinence rates or alcohol consumption level during follow-up. However, a later follow-up of the Edwards cohort found that more severely dependent drinkers benefited more from intensive treatment.[23]

In a larger study in the US, employees who were identified as drinking excessively were randomized to one of the three options: compulsory inpatient treatment, compulsory AA attendance, or a choice of these two options.[24] At 2-year follow-up there were no differences between the groups in terms of work-related outcome measures. However, drinking-related measures the inpatient group had the best, and the AA group the poorest outcome, with the choice group having an intermediate outcome. The compulsory AA group was more likely than the others to require subsequent inpatient treatment. However, the length of inpatient treatment does not appear to influence outcome significantly.[25,26]

Studies comparing inpatient versus outpatient alcohol detoxification have generally found the two approaches to be equally effective. For example, Hayashida *et al.*[27] randomized male military veterans to inpatient and outpatient detoxification. At 6 months' follow-up no differences in outcome were found between the two groups. Indeed, outpatient detoxification is generally regarded as the treatment of choice for the majority of patients. It should be noted, however, that studies comparing inpatient and outpatient treatment (including detoxification) have tended to exclude patients with particularly poor prognosis (e.g. poor social circumstances, severe psychiatric or physical comorbidity, those at risk of harm to themselves or others). Hence, the clinician needs to interpret the research evidence with caution in applying it to patients in the typical clinical setting. However, it is probably safe to assume that in less complicated alcohol dependence there is no evidence of an advantage of inpatient over outpatient treatment.

A recent randomized trial by Rychartic *et al.*[28] assigned alcohol dependent patients to inpatient, intensive outpatient, or standard outpatient treatment. Following treatment inpatients had reduced jail and subsequent inpatient episodes, and those with greater alcohol dependence or impaired cognitive function had better outcomes with inpatient treatment.

Overall, the majority of studies that have compared intensive specialist treatment with less intensive treatment have not supported the use of more intensive approaches. However, most of these studies excluded patients with more complex needs. Few studies have examined the interaction between treatment setting and problem severity. The emerging evidence now is that alcohol dependent patients with more complex needs (more severe alcohol dependence, psychiatric comorbidity, cognitive impairment, poor social circumstances, or support) are more likely to benefit from inpatient treatment.[9]

Community-based specialist treatments

The growth of studies questioning the value of specialist inpatient treatment and a move towards cost containment in health care have led to a shift in resources to treating alcohol use disorders in community settings. In the US for example 87 per cent of specialist alcohol treatment is delivered on an outpatient basis.[29] A similar survey in England found that 69 per cent of specialist alcohol treatment agencies were community based.[30]

Apart from the potential advantage of lower cost, community-based treatment provides the least social disruption for the individual and offers the opportunity to mobilize existing community resources to support sustained recovery. In the UK, the past 30 years have seen the widespread development of the community alcohol team (CAT) model of treatment following the original Maudsley Alcohol Pilot Project.[31] The main principle of the CAT model is that the specialist multi-disciplinary team (typically consisting of specialist medical, nursing, social work, and psychology staff) work to train and support generic teams, mainly in primary care, to manage alcohol use disorders more effectively. In practice, CATs have tended to find difficulty in avoiding becoming involved in a more traditional specialist role, often providing direct care for alcohol use disorders in the face of reluctance on the part of primary care personnel to take on this work.[32]

There has been remarkably little research conducted to evaluate the CAT model. One study randomly allocated 40 problem drinkers referred to the specialist alcohol treatment clinic at the Maudsley Hospital to receive either routine specialist treatment or 'shared care'.[33] Following specialist assessment, the shared care group was returned to the care of their general practitioner, who was then supported by the specialist CAT. Shared care within this model included advice and training for the general practitioner, a shared treatment plan, regular phone contact between specialist and general practitioner, and the offer of further specialist care should the patient remain unchanged or deteriorate. At 6 months' follow-up the specialist and shared care groups both showed significant improvements, but there was no difference in outcome between the two groups.

Another study in Scotland evaluated the efficacy of a home detoxification service compared with minimal intervention in a

randomized controlled trial in 95 patients referred by their general practitioner.[34] At 6 months' follow-up the home detoxification group remained abstinent twice as long after treatment than the minimal intervention group.

The 'community reinforcement approach' has been demonstrated to have benefits in the treatment of alcohol dependence in the US.[35] This approach aims to provide reinforcers for abstinence from alcohol including positive family support, help in finding employment, membership of an alcohol-free social club, and alcohol counselling. The specialist treatment input aims to ensure that these supports are put in place. There is some evidence from small-scale controlled trials[35,36] that this approach is effective in reducing alcohol consumption and improving social adjustment compared to standard treatment, but it has not so far been fully evaluated, and has never been tested in the UK.

Another variant on the CAT approach has been the evaluation of community psychiatric nurse (CPN) aftercare following specialist inpatient treatment.[37] One study in the UK evaluated the effectiveness of regular CPN follow-up consisting of weekly 1 to 2 h visits to the patient's home for a period of 6 weeks post-discharge from inpatient care, followed by less frequent visits up to 1 year. The home-based sessions involved advice, support, counselling, partner involvement, and encouragement to attend AA. This was compared to routine 6-weekly hospital appointments. The study, which involved a non-randomized design, found significant improvements in abstinence and engagement in support in the CPN approach compared to the routine aftercare group.

There is growing interest in the potential application of Assertive Community Treatment (ACT) approaches for alcohol dependence, particularly for patients with more severe, complex and chronic problems who are difficult to engage in standard treatment approaches. This borrows from the experience of ACT in severe mental illness[38] and acknowledges that for some patients, alcohol dependence is a chronic disorder that in these cases may be more suited to a 'disease management' model of care, commonplace in treatment of many physical illnesses such as diabetes or hypertension. Assertive approaches appear promising in alcohol dependence,[39] but a definitive trial of ACT is needed.

In summary, the CAT model of alcohol service delivery has been widely implemented in the UK in advance of clear evidence of its effectiveness. Evidence is emerging showing at least the equivalence, and in some cases, the superiority, of outcome from community-based services compared with more traditional inpatient treatment approaches. However, the CAT approach is implemented in a range of ways in the UK, and encompasses many different models and specific interventions. More research is needed to evaluate the cost-effectiveness of community alcohol team approaches and to identify the specific elements and methods that contribute to treatment effectiveness.

Matching and stepped care

The Institute of Medicine report emphasized the need to match the level of intervention to the severity and nature of the presenting problems.[6] There is some empirical evidence of matching effects in relation to both inpatient and outpatient treatment.[40] Up until recently, however, matching effects have generally been explored in *post hoc* analyses in studies that lacked sufficient statistical power. The Project MATCH study in the US aimed to assess a wide range of matching hypotheses in a prospective design, but found no strong matching effects[42] (see Chapter 4.2.2.4). However, it should be noted that most controlled trials, including MATCH, excluded the more complex patients, including those with limited social support and those with severe psychiatric comorbidity. This tends to work against finding matching effects as the study sample lacked clinical heterogeneity.[43] Further, many of the patient and treatment programme characteristics likely to mediate treatment matching and treatment effectiveness, remain largely unresearched. The matching results of a similar trial in the UK are awaited.[43]

Stepped care is an alternative method of matching treatments to patient needs that has become accepted in the fields of smoking intervention and general medicine. Until now it has received relatively scant attention in the alcohol field. In essence, stepped care involves initially providing relatively low-intensity treatments, and only offering more intensive treatments to those who fail to respond.[44] This provides a potentially resource-efficient means of delivering treatment, and provides clinicians with clinical algorithms. A recent trial of alcohol screening in primary care compared stepped care intervention with minimal 5 min of advice delivered by a practice nurse.[45] In the stepped care group received an initial 40 min session of Behaviour Change Counselling delivered by a trained practice nurse who then followed up the patients (Step 1). Those who did not respond to Step 1 were referred to four sessions of Motivational Enhancement Therapy delivered by trained alcohol counsellors (Step 2). Finally those not responding to Step 2 were referred to more intensive treatment delivered by a CAT (Step 3). The study found no significant difference in alcohol consumption at 6 months, which may in part be due to a small sample size. However, the stepped care intervention was more cost-effective than minimal intervention, mainly through reduced health care and criminal justice costs.

Overall, few community-based studies have found significant treatment matching effects. But this may be in part due to exclusion of the most severe cases. Some studies have found advantages of inpatient compared with community treatment for patients with more severe and complex needs as described above. A more promising approach is stepped care, which is effectively pragmatic matching: that is patients not responding to less intensive treatments receive more intensive treatments. This now forms an important principle of the national framework for alcohol services in England.[46]

Cost-effectiveness of alcohol treatment

With a trend towards containment of health care costs in industrialized societies, there has been an increase in the application of health economic research in the alcohol treatment field. It has been estimated that the annual cost to society of alcohol misuse is in the region of US$184 billion in the US[47] and £20 billion in the UK.[48] In comparison, the direct treatment costs of alcohol use disorders by specialist treatment agencies amounted to approximately US$7.5 billion in the US and about £217 million in the UK.[47,30] Thus there is a need to demonstrate the cost-effectiveness of treatments for alcohol use disorders.

Until recently, research on cost-effectiveness has been largely speculative and not based on direct estimates of cost benefits. In a landmark study, Holder *et al.*[49] provided a 'first approximation' of the cost-effectiveness of treatment. In their analysis they

used a combination of findings of efficacy from clinical trials, typical costs of different treatments, and recommendations from experts and treatment providers about appropriate treatment approaches. While noting the lack of studies directly assessing the cost-effectiveness of treatments, they concluded the cost of care was inversely correlated with evidence of effectiveness. They also noted that those treatments with the highest cost and lowest evidence of effectiveness were amongst the most prevalent in the North American treatment system. While this review has been criticized on methodological grounds, it has stimulated an important debate and has contributed to an increasing number of clinical trials including a health economic component in outcome evaluation.

Cost-effectiveness analysis, which takes a societal perspective rather than a narrow intervention cost perspective, provides a better measure of the overall impact of an intervention. These wider costs include patient out-of-pocket costs, lost productivity, unplanned health care utilization (e.g. admissions with alcohol-related physical and mental illnesses, primary care utilization), criminal justice costs, accidents, premature deaths, social work involvement, childcare costs, and costs associated with illnesses in relatives and carers.[50]

Also important in cost-effectiveness analysis is the estimation of improvements in quality of life following an intervention. This is beginning to be studied in the alcohol treatment field. Quality of life can be measured in a variety of ways (e.g. Euroqol, Short Form 36). In a randomized controlled trial an estimate of the difference in Quality Adjusted Life Years (QALYs) can be compared between treatment and control groups. The National Institute of Clinical Excellence (NICE) in the UK determines which treatments should be funded by the National Health Service (NHS) based on the available evidence. NICE regards a net cost per QALY, taking into account the costs of treatment and the savings to society, of £20 000 or less to be the maximum cost acceptable to implement the treatment in the NHS. Alcohol misusers typically have a much lower quality of life than non-alcohol misusers. One study compared quality of life in alcohol dependent drinkers before treatment with controls. Mean quality of life measured by the Euroqol (EQ5D) was 0.57 in the alcohol dependent group compared to age matched controls: 0.9 (1.0 being the best possible quality of life and 0 being death).[43] However, a recent review found that while treatment significantly reduces alcohol consumption and societal costs, it has a limited impact on quality of life.[9]

A recent review of cost-effectiveness examined the available literature on alcohol interventions.[9] In terms of intensive specialist alcohol treatment it was estimated that providing evidence-based alcohol interventions would result in a saving of £5 for the public sector for every £1 spent. However, it was noted that several studies showed an initial increase in costs in people newly entering alcohol treatment. This is likely to be because people not in treatment and in an active drinking phase find it harder to access services, and specialist alcohol services assist patients to address outstanding health and social care needs. Cost savings as a result of treatment therefore need to be examined over the longer term (>1 year). Comparing individual specialist interventions including pharmacotherapies and psychosocial interventions one review found the net health cost per death averted ranged from –£3073 to £2076 for most of the interventions that also provided significant clinical improvements.[51]

There is now good evidence that brief intervention in hazardous/harmful drinkers is highly cost-effective. Fleming *et al.*[52] found that, as well as reducing excessive drinking, there was a reduced length of hospitalization during the 12-month follow-up period. In addition, while brief intervention cost more than minimal intervention, this cost was more than offset by reductions in subsequent health care, criminal justice, and road traffic accident costs (US$56 000 savings per US$10 000 intervention costs). An analysis of several brief intervention studies found the cost per life year gained was approximately £2000, well within the NICE definition.[53]

There is considerable scope for further development of health economic research in the alcohol field. This will prove important in providing health care commissioners with appropriate information to make rational decisions in the provision of cost-effective evidence-based services for alcohol use disorders. (For a further account of cost-effectiveness analysis, see Chapter 7.7.)

The availability of alcohol services

The availability of alcohol services is likely to affect the overall impact of treatment at a whole population level. There is some evidence that the availability of alcohol treatment services is related to the prevalence of alcohol use disorders at a population level. Mann *et al.*[54] found that increased treatment services in Ontario, Canada, were associated with decreased hospital discharges for liver cirrhosis. A similar study in North Carolina examining the 20-year period between 1968 and 1987 found an association between increased alcohol treatment admissions and decreased cirrhosis mortality. Further, Mann *et al.*[55] found a relationship between AA membership and alcohol-related problems including cirrhosis mortality rates in the US, Canada, and other countries. They estimated that a 1 per cent increase in AA membership was associated with a 0.06 per cent decrease in cirrhosis mortality. These studies of course demonstrate associations rather than causal links between treatment availability and prevalence, but do provide support to the hypothesis that access to treatment could have an impact at a population level.

The National Drug and Alcohol Treatment Utilization Survey (**NDATUS**), which is a national census of public and private treatment programmes in the US, provides a unique data set to study treatment availability. It has been conducted intermittently since 1979 and provides a method to study trends over time. An analysis by the Institute of Medicine found large regional variations in the availability of treatment places.[6] There was no association found between treatment place availability and prevalence of alcohol misuse across states in the US. This points to the importance of 'needs assessment' in the rational allocation of public resources to fund treatment services. This involves a variety of data sources as indicators of alcohol use disorder prevalence in a particular locality, including general population surveys, mortality statistics (e.g. deaths from alcoholic liver disease), crime statistics (e.g. public drunkenness and driving whilst intoxicated arrests), and alcohol-related hospital admissions. Such indicators provide measures of relative 'need' in different localities, gaps between need and access to treatment, and can be used to direct resource allocation.

Examining data from NDATUS surveys between 1982 and 1993, Weisner *et al.*[56] found an increase in activity over this period of 147 per cent. In the US the impact of managed care organizations, which aim to limit access to treatment on the basis of individual

need and cost, has yet to be fully established in relation to overall access to alcohol services. Such measures are likely to reduce the availability of inpatient services and to reduce the rate of readmission for those with chronic alcohol problems.

A recent national needs assessment in England examined *inter alia* the regional variation in the prevalence of alcohol use disorders and access to specialist alcohol treatment services.[30] The prevalence of alcohol dependence in adults was 3.6 per cent overall (men: 6 per cent; women: 2 per cent), varying from 1.6 per cent to 5.2 per cent across regions. The overall level of access to treatment also varied across regions. In England overall, 1 in 18 people with alcohol dependence gained access to treatment per annum (the Prevalence–Service Utilization Ratio). But this varied from 1 in 12, to 1 in 108 between the best and worst served regions. A ratio of 1 in 10 is regarded as a 'low' level of access and 1 in 5 a 'high' level in North America.[57] Factors associated with better access included a greater number of treatment agencies, greater overall spending on treatment, and shorter waiting times.

Special groups

Specialist alcohol treatment services typically attract younger, male, single patients of lower socio-economic and educational background, with more severe alcohol dependence. Relative to the prevalence of alcohol use disorders in the general population, women, older people, and people from ethnic minorities are typically under-represented, as are the homeless. Further, there are limited specialist alcohol services for young people. This is of particular concern as the prevalence of alcohol use disorders is increasing in young people in the UK.

Women

Thom and Green have identified three main factors that may account for the under-representation of women in alcohol treatment.[58] Women tend to perceive their problems differently from men, less often identifying themselves as 'alcoholic'. This may in part be related to negative public stereotypes of female drinking and negative attitudes towards female problem drinkers amongst professionals. Women have also been found to perceive the 'costs' of entering treatment differently from men. This is particularly in relation to the perceived social stigma as well as other costs, both financial, in relationships, and in terms of losing their children into the care system. Finally, women often find the services offered to be less appropriate in meeting their needs than do men. Often specialist alcohol services do not offer childcare or 'women-only' facilities. However, an increasing number of women are seeking help for alcohol use disorders, both in the US and the UK on the basis of general population surveys and surveys of treatment populations. In England we found that women with alcohol dependence were 1.6 times more likely to access treatment than men.[30] This suggests that some of the barriers to access identified by Thom and Green have been overcome. Nevertheless, more still needs to be done to provide alcohol treatment services that are sensitive to women's needs. Further, there is a need to develop services catering for pregnant women.[59]

Ethnic minority groups

The evidence concerning help-seeking in ethnic minority groups is complex (see Chapter 7.3). Harrison *et al.*[60] have provided a review of the evidence. In the US, Hispanics tend to be under-represented and African–Americans are over-represented in alcohol treatment compared with the general population prevalence. However, interpretation of the evidence is complicated by the fact that household surveys tend to under-represent socially disadvantaged individuals from ethnic minorities. Marmot *et al.*[61] found that cirrhosis mortality rates were elevated compared to the national average for men from the Asian subcontinent and from Ireland, but lower than average for African–Caribbean men. In women, cirrhosis mortality was lower than average in Asian and African–Caribbean women but higher in Irish women. However, there were few cirrhosis deaths in total in ethnic minorities, which may lead to large errors in extrapolation to the whole population alcohol misuse estimates. In terms of alcohol treatment populations, studies have tended to find higher rates of admission (per 100 000 population) in Indian-, Scottish- and Irish-born people than in those born in the Caribbean or Pakistan. Differences in culturally related health beliefs and help-seeking, as well as service factors such as the availability of interpreters or treatment personnel from appropriate ethnic minority groups, may account for some of these differences. There remain few specific services for people from ethnic minorities, although some examples of good practice exist in the UK.[60]

The homeless

There is a high prevalence of alcohol use disorders (as well as mental and physical health and social problems) amongst the homeless population, a group that is not typically well catered for by mainstream alcohol services (see Chapter 7.10.2). The prevalence of alcohol problems in the homeless has been found to be as high as 38 per cent in the UK[62] and between 2 and 86 per cent in the US; typically the prevalence is between 20 and 45 per cent in North American studies.[59] This has contributed to the development of specific alcohol services for the homeless and street drinkers, notably 'wet' hostels. In the 'wet' hostel, residents are able to continue drinking but are cared for in an environment that is designed to minimize the harm associated with heavy drinking and to tackle issues associated with homelessness.[59,62] Such facilities tend to be restricted to large urban centres and have restricted places compared to the number of street drinkers. Similarly, outreach services and 'crisis centres' have been developed to attract alcohol-misusing homeless people into treatment facilities.[63] Often those entering 'wet' hostels can subsequently be persuaded to undergo alcohol detoxification and progress to 'dry' (or alcohol-free) supported accommodation.

Young people

The prevalence of alcohol use disorders is increasing in young people, particularly young women in the UK.[48] The young are over-represented in alcohol-related road traffic accidents, and alcohol is a leading cause of accidental death in this group. Alcohol misuse is also associated with unprotected sexual activity. Nevertheless, there are few specialist alcohol services for young people. Most initiatives have been directed at prevention and health promotion in this group, but the evidence to support these is lacking. This has led to the proposal that individually targeted interventions, for example by the primary health care team or in accident and emergency departments, are more likely to be effective.[64]

Relatives and carers

Relatives and carers of people with alcohol use disorders often experience significant social and psychological problems related to the drinking of a 'significant other'. Alcohol use disorders are associated with a high level of domestic violence and child neglect and abuse. Many specialist treatment programmes provide help and support to relatives and carers, and Al-Anon (for adult carers and relatives) and Alateen (for the young), which are affiliates of AA, provide a widely available source of mutual aid for these groups.

Services for individuals with comorbidity

There is an increasing recognition of the problems associated with alcohol and other drug misuse and mental illness (see Chapter 4.2.2.3). Often alcohol misuse is complicated by multiple substance misuse. For example, in the Epidemiologic Catchment Area Study half of all patients with schizophrenia also had a substance use disorder,[65] and a recent British survey of psychiatric inpatients found that half had an alcohol use disorder.[66] However, there is currently no consensus on the most appropriate treatment services for patients with comorbidity.[67] Alcohol and substance misuse can be particularly problematic in the context of mental illness, and is associated with higher rates of violence and poor treatment outcome. Such patients are often poorly engaged with, and disruptive in mental health services, and typically have difficulty in engaging in alcohol or drug services. Assertive community outreach and integrated service models, covering both mental illness and substance misuse, have been advocated, but more research is needed to evaluate the cost-effectiveness of these approaches.

Conclusions

The alcohol treatment field has seen considerable change over the past 30 years. Some of this has been evidence based, and some has been largely politically driven, particularly in the pursuit of containing health care costs. On the positive side, a shift in policy from a limited number of treatment services catering only for the small minority of severely dependent drinkers, to more community orientated services with a greater emphasis on early identification and intervention, is to be broadly welcomed. However, in some places a move towards services catering for early stage 'at-risk' drinkers has been at the expense of losing services for those with more severe alcohol problems.[40] While the evidence in favour of matching treatments to individual needs is still at a relatively early stage of development, and clear evidence of matching effects is not yet available, clinical practice needs to be guided by pragmatic principles by which more intensive treatments are provided to more complex patients, and/or in a stepped care paradigm. It must be concluded that, despite a large research effort in evaluating intensive versus less intensive alcohol interventions, there is still a long way to go in developing pragmatic clinical trials that evaluate effectiveness and cost-effectiveness of treatment in a way that can best advise practitioners in the typical treatment setting.

On the positive side, research has begun to address fundamental health economic issues that are highly relevant to the rational funding of treatment services. Important in this is the development of health economic analysis in randomized controlled trials. The assessment of the impact of treatment availability on the prevalence of alcohol-related harm also represents a significant advance.

Health services research that does not influence clinical practice fails in its fundamental aim. For example, while there is a now considerable evidence base in support of brief intervention in the primary care setting, there is a resistance within primary care to adopt such approaches, often despite exhortations from governments and professional bodies. Part of the problem may lie in the disparity between the priorities of public health, which is directed towards population level benefits of an intervention, and the priorities of the individual practitioner, whose first duty is to the patient in his or her care.[17] If the individual practitioner remains unconvinced about the value of a particular intervention for the patient, such public health policies are likely to fail even if they are supported by research evidence.

Similarly, as Holder *et al.*[49] have pointed out, often treatment programmes continue to provide alcohol services that are not supported by research evidence. On occasions, but not exclusively, this criticism is levelled at private for-profit agencies, with the implication that their motivation is financial rather than being principally for the benefit of their patients. In many cases, however, the evidence base is lacking because the fundamental research has not yet been conducted. Or, as in the case of self-help organizations such as AA, the methodology necessary adequately to evaluate an intervention would be extremely complex, or perhaps impossible, to conduct to the standard typically expected in evidence-based medicine (i.e. a randomized controlled trial).

Nevertheless, treatment research cannot occur in a vacuum. Research needs to take account of the funding environment in which treatment takes place. Further, treatment research needs to provide answers to the key issues facing commissioners of health care. With the gradual improvement in the quality of treatment research over the past three decades[68] and the development of more advanced health economic methods to evaluate treatment, the treatment research community is in a much better position than ever before to provide evidence to guide the rational development of treatment services for alcohol use disorders.

While many differences between health care systems exist in different countries, the evidence points to the need for a wide spectrum of services to cater for different needs. The development of low-threshold community-based services should not occur at the expense of more specialized services for more severe alcohol use disorders. Similarly, a treatment system that provides only specialist services for the minority of severe cases misses a significant public health opportunity to reduce the prevalence of alcohol use disorders through early, brief interventions.

Further information

Edwards, G., Marshall, J., and Cook, C.C.H. (1997). *The treatment of drinking problems: a guide for the helping professions* (3rd edn). Cambridge University Press, Cambridge.

Institute of Medicine. (1990). *Broadening the base of treatment for alcohol problems*. National Academy Press, Washington, DC.

Raistrick, D., Heather, N., and Godfrey, C. (2006). *Review of the effectiveness of treatment for alcohol problems*. National Treatment Agency for Substance Misuse, London.

National Treatment Agency for Substance Misuse. (2006). *Models of care for alcohol misusers*. National Treatment Agency for Substance Misuse, London.

References

1. Jellinek, E.M. (1960). *The disease concept of alcoholism*. Hillhouse, New Haven, CT.
2. Edwards, G., Marshall, J., and Cook, C.C.H. (1997). *The treatment of drinking problems: a guide for the helping professions* (3rd edn). Cambridge University Press, Cambridge.
3. Chick, J., Lloyd, G., and Crombie, E. (1985). Counselling problem drinkers in medical wards: a controlled study. *British Medical Journal*, **290**, 965–7.
4. Wallace, P., Cutler, S., and Haines, A. (1988). Randomised controlled trial of general practitioner intervention in patients with excessive alcohol consumption. *British Medical Journal*, **297**, 663–8.
5. Kreitman, N. (1986). Alcohol consumption and the preventive paradox. *British Journal of Addiction*, **81**, 353–63.
6. Institute of Medicine. (1990). *Broadening the base of treatment for alcohol problems*. National Academy Press, Washington, DC.
7. Heather, N. (1998). Brief opportunities for change in medical settings. In *Treating addictive behaviors* (2nd edn) (eds. W.R. Miller and N. Heather), pp. 133–47. Plenum, New York.
8. Saunders, J.B., Aasland, O.G., Babor, T.F., *et al.* (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption—II. *Addiction*, **88**, 791–804.
9. Raistrick, D., Heather, N., and Godfrey, C. (2006). *Review of the effectiveness of treatment for alcohol problems*. National Treatment Agency for Substance Misuse, London.
10. Babor, T.F. and Grant, B. (eds.) (1992). *Programme on substance abuse: project on identification and management of alcohol related problems. Report on phase II: a randomised clinical trial of brief intervention in primary health care*. World Health Organization, Geneva.
11. Emmen, M.J., Schippers, G.M., Bleijenberg, G., *et al.* (2004). Effectiveness of opportunistic brief interventions for problem drinking in a general hospital setting: systemative review. *British Medical Journal*, **328**, 318–22.
12. Crawford, M.J., Patton, R., Touquet, R., *et al.* (2004). Screening and referral for brief intervention of alcohol misusing patients in an emergency department: a pragmatic randomized controlled trial. *Lancet*, **364**, 1334–9.
13. Smith, A.J., Hodgson, R.J., Bridgeman, K., *et al.* (2003). A randomized controlled trial of a brief intervention after alcohol-related facial injury. *Addiction*, **98**, 3–52.
14. Effective Health Care Team. (1993). *Brief interventions and alcohol use. Effective Health Care Bulletin 7*. Department of Health, London.
15. Moyer, A., Finney, J., Swearigan, C., *et al.* (2002). Brief interventions for alcohol problems: a meta-analytic review of controlled investigations in treatment seeking and non-treatment seeking populations. *Addiction*, **97**, 279–92.
16. Kaner, E.F.S., Beyer, F., Dickinson, H.O., *et al.* (2007). Effectiveness of brief alcohol interventions in primary care populations. *Cochrane Database of Systematic Reviews*, (2).
17. Heather, N. (1995). Interpreting the evidence on brief interventions for excessive drinkers: the need for caution. *Alcohol and Alcoholism*, **30**, 287–96.
18. Deehan, A., Templeton, L., Taylor, C., *et al.* (1998). Low detection rates, negative attitudes and the failure to meet 'Health of the Nation' targets: findings from a national survey of GPs in England and Wales. *Drug and Alcohol Review*, **17**, 249–58.
19. Deehan, A., Templeton, L., Taylor, C., *et al.* (1998). How do general practitioners manage alcohol misusing patients? Results from a national survey of GPs in England and Wales. *Drug and Alcohol Review*, **17**, 259–66.
20. Drummond, D.C. (1997). Alcohol interventions: do the best things come in small packages? *Addiction*, **92**, 375–9.
21. Edwards, G., Orford, J., Egert, S., *et al.* (1977). Alcoholism: a controlled trial of 'treatment' and 'advice'. *Journal of Studies on Alcohol*, **38**, 1004–31.
22. Chick, J., Ritson, B., Connaughton, J., *et al.* (1988). Advice versus extended treatment for alcoholism: a controlled study. *British Journal of Addiction*, **83**, 159–70.
23. Orford, J., Oppenheimer, E., and Edwards, G. (1976). Abstinence or control: the outcome for excessive drinkers two years after consultation. *Behaviour Research and Therapy*, **14**, 409–18.
24. Walsh, D.C., Hingson, R.W., Merrigan, D.M., *et al.* (1991). A randomised trial of treatment options for alcohol–abusing workers. *The New England Journal of Medicine*, **325**, 775–82.
25. Trent, L.K. (1998). Evaluation of a four–versus six–week length of stay in the navy's alcohol treatment program. *Journal of Studies on Alcohol*, **59**, 270–9.
26. Long, C.G., Williams, M., and Hollin, C.R. (1998). Treating alcohol problems: a study of programme effectiveness according to length and delivery of treatment. *Addiction*, **93**, 561–71.
27. Hayashida, M., Alterman, A.I., McLellan, A.T., *et al.* (1989). Comparative effectiveness and costs of inpatient and outpatient detoxification of patients with mild–to–moderate alcohol withdrawal syndrome. *The New England Journal of Medicine*, **320**, 358–65.
28. Rychartic, R., Connors, G., Whitney, R., *et al.* (2000). Treatment settings for persons with alcoholism: evidence for matching clients to inpatient versus outpatient care. *Journal of Consulting and Clinical Psychology*, **68**, 277–89.
29. Fuller, R.K. and Hiller-Sturmhofel, S. (1999). Alcoholism treatment in the US. *Alcohol Research and Health*, **23**, 69–77.
30. Drummond, C., Oyefeso, N., Phillips, T., *et al.* (2005). *Alcohol needs assessment research project: the 2004 national alcohol needs assessment for England*. Department of Health, London.
31. Shaw, S.J., Cartwright, A.K.J., Spratley, T.A., *et al.* (1978). *Responding to drinking problems*. Croom Helm, London.
32. Clement, S. and Stockwell, T. (eds.) (1987). *Helping the problem drinker: new initiatives in community care*. Croom Helm, London.
33. Drummond, D.C., Thom, B., Brown, C., *et al.* (1990). Specialist versus general practitioner treatment of problem drinkers. *Lancet*, **336**, 915–18.
34. Bennie, C. (1998). A comparison of home detoxification and minimal intervention strategies for problem drinkers. *Alcohol and Alcoholism*, **33**, 157–63.
35. Hunt, G.N. and Azrin, N.H. (1973). A community reinforcement approach to alcoholism. *Behaviour Research and Therapy*, **11**, 91–104.
36. Azrin, N.H., Sisson, R.W., Meyers, R., *et al.* (1982). Alcoholism treatment by disulfiram and community reinforcement therapy. *Journal of Behavior Therapy and Experimental Psychiatry*, **13**, 105–12.
37. Patterson, D.G., MacPherson, J., and Brady, N.M. (1997). Community psychiatric nurse aftercare for alcoholics: a five–year follow–up. *Addiction*, **92**, 459–68.
38. Marshall, M. and Lockwood, A. (1998). Assertive community treatment for people with severe mental disorders. *Cochrane Database of Systematic Reviews*, (2).
39. Hesse, M., Broekaert, E., Fridell, M., *et al.* (2006). Case management for substance use disorders. *Cochrane Database of Systematic Reviews*, (4).
40. Mattson, M.E., Allen, J.P., Longabaugh, R., *et al.* (1994). A chronological review of empirical studies matching alcoholic clients to treatment. *Journal of Studies on Alcohol*, **12**, 16–29.
41. Project MATCH Research Group. (1997). Matching alcoholism treatments to client heterogeneity: project MATCH post-treatment drinking outcomes. *Journal of Studies on Alcohol*, **58**, 7–29.
42. Drummond, D.C. (1999). Treatment research in the wake of Project MATCH. *Addiction*, **94**, 39–42.
43. UKATT Research Team. (2005). Effectiveness of treatment for alcohol problems: findings of the randomized UK Alcohol Treatment Trial (UKATT). *British Medical Journal*, **311**, 541–4.
44. Breslin, F.C., Sobell, M.B., Sobell, L.C., *et al.* (1997). Toward a stepped care approach to treating problem drinkers: the predictive utility of within treatment variables and therapist prognostic ratings. *Addiction*, **92**, 1479–89.

45. Drummond, D.C., James, D., Coulton, S., *et al.* (2003). *The effectiveness and cost effectiveness of screening and stepped care intervention for alcohol use disorders in the primary care setting*. Final Report to the Wales Office for Research and Development. St George's Hospital Medical School, London.
46. National Treatment Agency for Substance Misuse. (2006). *Models of care for alcohol misusers*. National Treatment Agency for Substance Misuse, London.
47. Harwood, H. (2000). *Updating estimates of the economic costs of alcohol abuse in the United States: estimates, update methods, and data*. National Institute of Alcohol Abuse and Alcoholism, Rockville, MD.
48. Prime Minister's Strategy Unit. (2004). *Alcohol harm reduction strategy for England*. Cabinet Office, London.
49. Holder, H.D., Longabaugh, R., Miller, W.R., *et al.* (1991). The cost effectiveness of treatment for alcohol problems: a first approximation. *Journal of Studies on Alcohol*, **52**, 517–40.
50. Drummond, M.F., O'Brien, B., Stoddard, G.L., *et al.* (1997). *Methods for the economic evaluation of health care programmes*. Oxford University Press, Oxford.
51. Slattery, J., Chick, J., Cochrane, M., *et al.* (2003). *Prevention of relapse in alcohol dependence*. Health technology assessment report 3. Health Technology Board for Scotland, Glasgow.
52. Fleming, M.F., Marlon, M.P., French, M.T., *et al.* (2000). Benefit cost analysis of brief physician advice with problem drinkers in primary care settings. *Medical Care*, **31**, 7–18.
53. Ludbrook, A., Godfrey, C., Wyness, L., *et al.* (2002). *Effective and cost effective measures to reduce alcohol misuse in Scotland*. Scottish Executive Health Department, Ediburgh.
54. Mann, R.E., Smart, R., Anglin, L., *et al.* (1988). Are decreases in liver cirrhosis rates a result of increased treatment for alcoholism? *British Journal of Addiction*, **83**, 683–8.
55. Mann, R.E., Smart, R., Anglin, L., *et al.* (1991). Reductions in cirrhosis deaths in the United States: associations with per capita consumption and AA membership. *Journal of Studies on Alcohol*, **52**, 361–5.
56. Weisner, C., Greenfield, T., and Room, R. (1995). Trends in the treatment of alcohol problems in the US general population, 1979 through 1990. *American Journal of Public Health*, **85**, 55–60.
57. Rush, B. (1990). A systems approach to estimating the required capacity of alcohol treatment services. *British Journal of Addiction*, **85**, 49–59.
58. Thom, B. and Green, A. (1996). Services for women with alcohol problems: the way forward. In *Alcohol problems in the community* (ed. L. Harrison), pp. 200–22. Routledge, London.
59. Institute of Medicine. (1988). *Homelessness, health, and human needs*. National Academy Press, Washington, DC.
60. Harrison, L., Harrison, M., and Adebowale, V. (1997). Drinking problems among black communities. In *Alcohol problems in the community* (ed. L. Harrison), pp. 223–40. Routledge, London.
61. Marmot, M., Adelstein, A., and Bulusu, L. (1984). *Immigrant mortality in England and Wales, 1970–78*. HMSO, London.
62. Harrison, L. and Luck, H. (1997). Drinking and homelessness. In *Alcohol problems in the community* (ed. L. Harrison), pp. 53–75. Routledge, London.
63. Freimanis, L. (1993). Alcohol and single homelessness: an outreach approach. In *Homelessness, health care and welfare provision* (eds. K. Fisher and J. Collins), pp. 44–51. Routledge, London.
64. May, C. (1993). Young heavy drinkers: is there a problem, is there a solution? *Health and Social Care*, **1**, 203–10.
65. Reiger, D.A., Farmer, M.E., Rae, D.S., *et al.* (1990). Co–morbidity of mental disorder with alcohol and other drug abuse: results from an Epidemiological Catchment Area (ECA) study. *The Journal of the American Medical Association*, **264**, 2511–18.
66. Barnaby, B., Drummond, D.C., McCloud, A., *et al.* (2003). Substance misuse in psychiatric inpatients: comparison of a screening questionnaire survey with case notes. *British Medical Journal*, **327**, 783–4.
67. Weaver, T., Renton, A., Stimson, G., *et al.* (1999). Severe mental illness and substance misuse. *British Medical Journal*, **318**, 137–8.
68. Moncrieff, J. and Drummond, D.C. (1998). The quality of alcohol treatment research: an examination of influential controlled trials and development of a quality rating system. *Addiction*, **93**, 811–23.

4.2.2.6 Prevention of alcohol-related problems

Robin Room

Alcohol consumption is widely distributed in the population in most parts of the world, with abstainers in a minority among adults in most developing societies but in a majority in many less developed societies.[1,2] Those qualifying to be diagnosed with an alcohol use disorder are usually a relatively small minority of drinkers.

On the other hand, alcohol is causally implicated in a wide variety of health and social problems. The WHO *Global Burden of Disease* (*GBD*) study for 2000 estimated that alcohol accounted globally for 4 per cent of the total health-related loss of disability-adjusted life years (DALYs), for 6.8 per cent in developed societies like those in Western Europe and North America, and for 12.1 per cent in Eastern Europe and Central Asia.[3] In terms of where this burden appears in the health system, while psychiatric conditions (including dependence) and chronic physical disease are both important, casualties often play a predominant role. The *GBD* 2000 study calculated that injuries accounted for 40 per cent of the DALYs lost worldwide due to alcohol.[3]

The public health importance of acute effects of a particular episode of intoxication underlies what is often described as the 'prevention paradox'. In many societies, a fairly substantial proportion of the population (particularly of males) gets intoxicated at least occasionally, and by that fact is at risk of experiencing and causing social and health harm from drinking.[4] Preventing alcohol problems thus requires looking beyond the considerably smaller segment of the population diagnosable with an alcohol use disorder, or the even smaller segment receiving treatment for such a disorder.

A complication in preventing alcohol problems is that there is also evidence of a health benefit from drinking in terms of reduced cardiovascular disease. This benefit is, however, important mainly for men over 45 and women past menopause, and can be attained with a pattern of very light regular drinking, as little as a drink every second day.[5] There is thus little potential conflict between taking alcohol as a preventive heart medication and any prevention policy short of total prohibition.

Simplifying somewhat, there are seven main strategies to minimize alcohol problems. One strategy is to educate or persuade people not to use or about ways to use so as to limit harm. A second strategy, a kind of negative persuasion, is to deter drinking-related behaviour with the threat of penalties. A third strategy, operating in the positive direction, is to provide alternatives to drinking or to

drink-connected activities. A fourth strategy is in one way or another to insulate the use from harm. A fifth strategy is to regulate availability of the drug or the conditions of its use. Prohibition of supply may be regarded as a special case of such regulation. A sixth strategy is to work with social or religious movements oriented to reducing alcohol problems. And a seventh strategy is to treat or otherwise help people who are in trouble with their drinking. We will consider in turn these strategies and the evidence on their effectiveness in reducing rates of alcohol problems in the population.

Education and persuasion

In principle, education can be offered to any segment of the population in a variety of venues, but it is usually education of youth in schools, which first comes to mind in the prevention of alcohol problems. Community-based prevention programmes, which may be also directed at adults, often also include an educational component.

Education offers new information or ways of thinking about information, and leaves it to the listener to draw conclusions concerning beliefs and behaviour. However, most alcohol education programmes go beyond this. A commonplace of the North American evaluative literature on alcohol education is that 'knowledge-only' approaches do not result in changes in behaviour.[6] School-based alcohol education has thus usually had a persuasional element, aiming to influence students in a particular direction.

Persuasion is directly concerned with changing beliefs or behaviours, and may or may not also offer information. Mass-media campaigns aimed at persuasion have been a very common component of prevention programmes for alcohol-related problems, but persuasion can be pursued also through other media and modalities.

In most societies, public health-oriented persuasion about alcohol must compete with a variety of other persuasional messages, including those intended to sell alcoholic beverages. The evidence that alcohol advertising influences teenagers and young adults towards increased drinking and problematic drinking is becoming stronger.[7,8] Even where alcohol advertising is not allowed on the mass media, these messages are conveyed to consumers and potential consumers in a variety of other ways.

Evidence on effectiveness

The literature on effectiveness of educational approaches is dominated by studies from the United States on school-based education. This means that the alcohol education has usually been in the context of drug and tobacco education, and that the emphasis has been on abstention,[9] or at least on delaying the start of drinking, in cultural circumstances where the median age of actually starting drinking is about 13, while the minimum legal drinking age is 21. In general, despite the best efforts of a generation of researchers, this literature has had difficulty showing substantial and lasting effects.[10] There is a good argument from general principles for alcohol education in the context of consumer and health education, but there is little evidence from the formal evaluation literature at this point of its effectiveness beyond the short term.

Persuasional media campaigns have also been a favourite modality in many places in recent decades for the prevention of alcohol problems. In general, evaluations of such campaigns have been able to demonstrate impacts on knowledge and awareness about substance use problems, but little effect on drinking behaviour. As with school education approaches, there are hints in the literature that success may come more from influencing the community environment around the drinker—in terms of attitudes of significant others, or popular support for alcohol policy measures—than from directly persuading the drinker him/herself. Thus, media messages can be effective as agenda-setting mechanisms in the community, increasing or sustaining public support for other preventive strategies.[11]

Deterrence

In its broadest sense, deterrence means simply the threat of negative sanctions or incentives for behaviour—a form of negative persuasion. Criminal laws deter in two ways: by general deterrence, which is the effect of the law in preventing a prohibited behaviour in the population as a whole, and specific deterrence, which is the effect of the law in discouraging those who have been caught from doing it again.[12] A law tends to have a greater preventive effect and to be cheaper to administer to the extent it has a strong general deterrence effect.

Prohibitions on driving after drinking more than a specified amount are now in effect in most nations.[13] In many societies, there have also been laws against public drunkenness (being in a public place while intoxicated), and against obnoxious behaviour while intoxicated. Other common prohibitions are concerned with producing or selling alcoholic beverages outside state-regulated channels, and with aspects of drinking under a specified minimum age.

Evidence on effectiveness

Drinking-driving legislation, such as 'per se' laws outlawing driving while at or above a defined blood-alcohol level, has been shown to be effective in changing behaviour and reducing rates of alcohol-related problems (Babor *et al.*,[14] pp. 157–72). The effect is through both general and specific deterrence. The quickness and certainty of punishment, as well as its severity, are important in the deterrent value (too much severity tends to undercut its quickness and certainty). Drinking driving is an ideal area for applying general deterrence, since the gains from breaking the law are limited, and automobile drivers typically have something to lose by being caught.

Many English-speaking and Scandinavian countries have had a tradition of criminalizing drinking in public places or public drunkenness as such, but the trend has been to decriminalize public drunkenness. Though there are few specific studies, criminalizing public drunkenness is not very effective in changing behaviour, particularly of those who have little to lose (Parliament of Victoria, Drugs and Crime Prevention Committee,[15] pp. 309–20).

Providing and encouraging alternative activities

Another strategy, in principle involving positive incentives, is to provide and seek to encourage activities, which are an alternative to drinking or to activities closely associated with drinking. This includes such initiatives as making soft drinks available as an alternative to alcoholic beverages, providing locations for sociability

as an alternative to taverns, and providing and encouraging recreational activities as an alternative to leisure activities involving drinking. Job creation and skill development programmes are other examples.

Evidence on effectiveness

'Boredom' and 'because there's nothing else to do' are certainly among the reasons that are given for drinking by some drinkers. And there are often good reasons of general social policy for providing and encouraging alternative activities. But it has been noted that the problem with alternatives to drinking is that drinking combines so well with so many of them. Soft drinks are indeed an alternative for quenching thirst, but they may also serve as a mixer in an alcoholic drink. Involvement in sports may go along with drinking as well as replace it. The few evaluation studies of providing alternative activities, again from a restricted range of societies, have generally not shown lasting effects on drinking behaviour,[16,17] though they undoubtedly often serve a general social purpose in broadening opportunities for the disadvantaged.[18]

Insulating use from harm

A major social strategy for reducing alcohol-related problems in many societies has been measures to separate the drinking, and particularly heavy drinking, from potential harm. This separation can be physical (in terms of distance or walls), it can be temporal, or it can be cultural (e.g. defining the drinking occasion as 'time-out' from normal responsibilities). These 'harm reduction' strategies, as they are called in the context of illicit drugs, are often built into cultural arrangements around drinking, but can also be the object of purposive programmes and policies (Moore and Gerstein,[19] pp. 100–11).

A variety of modifications of the driving environment affect casualties associated with drinking and driving, along with other casualties. These include mandatory use of seat belts, airbags, and improvements in the safety of road vehicles and roads. Many other practical measures tending to separate intoxication episodes from casualties and other adverse consequences have been put into practice, though usually without formal evaluation.

The main focus for self-conscious strategies of alcohol harm reduction in recent years has been on modifying the drinking environment, particularly in public drinking places, primarily by modifying the behaviour of alcohol servers through server training and enforcement of bans on serving those under age or already intoxicated (Babor *et al.*,[14] pp. 141–47).

Evidence on effectiveness

Drinking-driving countermeasures are a prime example of an approach in terms of insulating drinking behaviour from harm, since they seek to reduce alcohol-related traffic casualties without necessarily stopping or reducing alcohol use. There is substantial evidence of the success of a range of such countermeasures (Babor *et al.*,[14] pp. 159–68). Environmental measures which reduce road casualties in general—e.g. requiring wearing of seat belts in cars, providing sidewalks separated from the road—may prevent casualties associated with intoxication at least as much as other casualties. While there is also evidence for the effectiveness of some other harm reduction approaches, there are also many examples of well-meaning efforts which proved ineffective (Room *et al.*,[2] pp. 186–92).

Regulating the availability and conditions of use

In terms of the substantial harms to health and public order they can cause, alcoholic beverages are not ordinary commodities. Governments have thus often actively intervened in the markets for such beverages, far beyond usual levels of state intervention in markets for commodities.

Total prohibition can be viewed as an extreme form of regulation of the market. In this circumstance, where no one is licensed to sell alcohol, the state has no formal control over the conditions of the sales, which nevertheless occur, and there are no legal sales interests, controlled through licensing, to cooperate with the state in the market's regulation.

With a general prohibition, typically the consumption of alcohol does fall in the population, and there are declines also in the rates of the direct consequences of drinking such as cirrhosis or alcohol-related mental disorders.[19,20] But prohibition also brings with it characteristic negative consequences, including the emergence and growth of an illicit market, and the crime associated with this. Partly for this reason, prohibition is not now a live option in any developed society, although it is in some other societies.

The features of alcohol control regimes, regulating the legal market in alcohol, vary greatly. Special taxes on alcohol are very common, imposed often as much for revenue as for public health considerations. Many societies have minimum age limits forbidding sales to underage customers, and regulating forbidding sales to the already intoxicated. Often the regulations include limiting the number of sales outlets, restricting hours and days of sale, and limiting sales to special stores or drinking places. Rationing of alcohol purchases—limiting the amount individuals can buy in a given time-period—has also been used as a means of regulating availability. Regulations restricting or forbidding advertising of alcoholic beverages attempt to limit or channel efforts by private interests to increase demand for particular alcoholic beverage products. Such regulations potentially complement education and persuasion efforts. State monopolization of sales of some or all alcoholic beverages at the retail and/or wholesale level has also been commonly been used as a mechanism to minimize alcohol-related harm.[21]

The effectiveness of specific types of regulation of availability

The last 25 years have seen the development of a burgeoning literature on the effects of alcohol control measures. Specific types of regulation of the alcohol market, and the evidence on their effectiveness, are discussed below.

(a) Minimum age limits

A minimum age limit is a partial prohibition, applied to one segment of the population. There is a strong evaluation literature showing the effectiveness of establishing and enforcing minimum age limits in reducing alcohol-related problems (Babor *et al.*,[14] pp. 127–28). However, this literature has been primarily North America based, focuses mostly on youthful driving casualties, and mostly evaluates reduction from and increases to age 21 as the limit, a higher minimum age limit than in most societies. The applicability of the literature's findings in other societies and where youth cultures are less automobile-focused has been little tested.

(b) Taxes and other price increases

Generally, consumers show some response to the price of alcoholic beverages, as of all other commodities. If the price goes up, the drinker will drink less; data from developed societies suggests this is at least as true of the heavy drinker as of the occasional drinker (Babor *et al.*,[14] pp. 110–11). Studies have found that alcohol tax increases reduce the rates of traffic casualties, of cirrhosis mortality, and of incidents of violence.[22,23]

(c) Limiting sales outlets, and hours and conditions of sale

There is a substantial literature showing that levels and patterns of alcohol consumption, and rates of alcohol-related casualties and other problems, are influenced by such sales restrictions, which typically make the purchase of alcoholic beverages slightly inconvenient, or influence the setting of and after drinking (Babor *et al.*,[14] pp. 125–42). Enforcing rules influencing 'house policies' in drinking places on not serving intoxicated customers, etc., has also been shown to have some effect (Babor *et al.*,[14] pp. 142–45).

(d) Monopolizing production or sale

Studies of the effects of privatizing retail alcohol monopolies have often shown some increase in levels of alcohol consumption and problems, in part because the number of outlets and hours of sale typically increase with privatization.[24] From a public health perspective, it is the retail level which is important, while monopolization of the production or wholesale level may facilitate revenue collection and effective control of the market.

(e) Rationing sales

Rationing the amount of alcohol sold to an individual potentially directly impacts on heavy drinkers, and has been shown to reduce levels both of intoxication-related problems such as violence, and of drinking-history-related problems such as cirrhosis mortality.[25,26] But while a form of rationing—the medical prescription system—is well accepted in most societies for psychoactive medications, it has proved politically unacceptable nowadays for alcoholic beverages in developed societies.

(f) Advertising and promotion restrictions

Many societies have regulations on advertising and other promotion of sales of alcoholic beverages.[13] As noted, the evidence on the effects of advertising and promotion on overall demand has become stronger in the recent literature.[27] However, studies of the effects of advertising and promotion restrictions on alcohol consumption have so far found at best weak effects, at least in the short term (Babor *et al.*,[14] pp. 180–83).

Social and religious movements and community action

Substantial reductions in alcohol-related problems have often been the result of spontaneous social and religious movements, which put a major emphasis on quitting intoxication or drinking. In recent decades, there have also been efforts to form partnerships between state organizations and nongovernmental groups to work on alcohol problems, often at the level of the local community. There has been an active tradition of community action projects on alcohol problems, often using a range of prevention strategies.[28–31] School-based prevention efforts have also moved increasingly to try to involve the community, in line with general perceptions that such multifaceted strategies will be more effective.[32]

While some of the biggest historical reductions in alcohol problems rates have resulted from spontaneous and autonomous social or religious movements, support or collaboration from a government can easily be perceived as official cooptation or manipulation.[33] Thus there is considerable question about the extent to which such movements can or should become an instrument of government prevention policies.

Evidence on effectiveness

In the short term, movements of religious or cultural revival can be highly effective in reducing levels of drinking and of alcohol-related problems. Alcohol consumption in the United States fell by about one-half in the first flush of temperance enthusiasm in 1830–1845 (Moore and Gerstein,[19] p. 35). Rates of serious crime are reported to have fallen for a while to a fraction of their previous level in Ireland in the wake of Father Mathew's temperance crusade.[34] The enthusiasm which sustains such movements tends to decay over time, though they often leave behind new customs and institutions with much longer duration. For instance, though the days when the historic temperance movement in English-speaking societies was strong are long gone, the movement had the long-lasting effect of largely removing drinking from the workplace in these societies.

There are some good examples of well-evaluated community action projects with demonstrated effectiveness.[30,31] However, the strategies used in such projects should be guided by evidence on effectiveness; good intentions and effort in themselves offer no guarantee of effectiveness.

Treatment and other help

Providing effective treatment or other help for drinkers who find they cannot control their drinking can be regarded as an obligation of a just and humane society. The help can take several forms: a specific treatment system for alcohol problems, professional help in general health or welfare systems, or non-professional assistance in mutual-help movements. To the extent such help is effective, and it is also potentially a means of reducing rates of alcohol-related problems.

Treatments for alcohol problems need not be complex or expensive. The evaluation literature suggests that brief outpatient interventions aimed at changing cognitions and behaviour around drinking are as effective in most circumstances as longer and more intensive treatment.[35,36] Positive results from such interventions in a primary health care settings were shown in a WHO study including a number of countries.[37]

Evidence on effectiveness

In terms of the effects of treatment on those who come for it, there is good evidence of effectiveness of treatment for alcohol problem. Typically, the improvement rate from a single episode of treatment is about 20 per cent higher than the no-treatment condition. Further treatment episodes are often needed. Brief treatment interventions or mutual-help approaches usually result in net savings in social and health costs associated with the heavy drinker (at least where health care is not self-paid), as well as improving the quality of life.[38,39] However, evaluations of brief interventions

by medical general practitioners have not always found effects.[40] Getting general practitioners to use the methods on a sustained basis has not proved easy,[41] and their patients are often unreceptive[42] or recalcitrant.[43] It remains to be seen whether and in what sociocultural circumstances making brief interventions for problematic drinking a routine part of general medical practice is a feasible and effective strategy.

The effectiveness of providing treatment as a strategy for reducing rates of alcohol problems in a society is more equivocal. In a North American context, it has been argued that the steep increase in alcohol problems treatment provision and mutual-help group membership in recent decades has contributed to reducing alcohol problems rates.[44] But the strength of the evidence for this contention is disputed.[45,46] A treatment system for alcohol problems is an important part of an integrated national alcohol policy, but as an instrument of prevention—of reducing societal rates of alcohol problems—it is probably not cost-effective.

Building an integrated societal alcohol policy

Often the different strategies for preventing alcohol problems appear to be synergistic in their effects.[47] Controls of availability, for instance, are more likely to be adopted, continued, and respected when the public has been successfully persuaded of their effects and effectiveness. But strategies can also work at cross-purposes: a prohibition policy, for instance, makes it difficult to pursue measures which insulate drinking from harm.

In a society where alcohol is a regular item of consumption, in view of the resulting rates of alcohol-related social and health problems, there is a strong justification for adopting a comprehensive policy concerning alcohol, taking into account production, marketing and consumption, and the prevention and treatment of alcohol-related problems. But while the adoption of overall alcohol strategies or policies has become common,[48] governments often tend to shy away from the most effective strategies.

In terms of strategies we have reviewed for managing and reducing the rates of alcohol problems in the society, there is clear evidence for effectiveness and cost-effectiveness of measures regulating the availability and conditions of use, and some such evidence for some measures which insulate use from harm. With respect to some aspects of alcohol problems, notably drinking-driving, deterrence measures also fall in the same category. The literature has begun to move beyond the question of effectiveness of these measures, and to consider questions of the relative cost-effectiveness of different measures in different societies.[49]

Despite their perennial popularity, evidence of the effectiveness of education/persuasion and treatment strategies in reducing societal rates of problems is limited at best. Education and treatment are good things for a society and a government to be doing about alcohol problems, but they do not constitute in themselves a public health policy on alcohol. These strategies will be nevertheless be pursued in most societies, and they can best pursued with attention to using cost-effective methods, and to integrating targets and messages with other aspects of alcohol policy.

Physicians and other health workers observe the adverse effects of alcohol in their daily practice, and are well-positioned to argue for public health approaches to reducing the burden of alcohol problems. Reports by colleges of psychiatrists and other physicians have played an important role in such countries as Britain[50] and Sweden[51] in putting a public health response to alcohol problems on the societal agenda.

Further information

References 2, 13, and 14 are useful also as sources of more general information.

Chikritzhs, T., Gray, D., Lyons, Z., *et al.* (2007). *Restrictions on the sale and supply of alcohol: evidence and outcomes. National Drug Research Institute, Curtin University of Technology, Perth,* http://www.ndri.curtin.edu.au/pdfs/publications/R207.pdf

WHO Expert Committee on Problems Related to Alcohol Consumption. (2007). *Second report*. WHO Technical Report Series, No. 944, http://www.who.int/substance_abuse/activities/expert_comm_alcohol_2nd_report.pdf

References

1. Rehm, J., Room, R., Monteiro, M., *et al.* (2004). Alcohol use. In *Comparative quantification of health risks. Global and regional burden of disease attributable to selected major risk factors*, Vol. 1 (eds. M. Ezzati, A.D. Lopez, A. Rodgers, and C.J.L. Murray), pp. 959–1108. World Health Organization, Geneva.
2. Room, R., Jernigan, D., Carlini-Marlatt, B., *et al.* (2002). *Alcohol and developing societies: A public health approach*. World Health Organization, Finnish Foundation for Alcohol Studies & Geneva, Helsinki.
3. Room, R., Babor, T., and Rehm, J. (2005). Alcohol and public health: a review. *Lancet*, **365**, 519–30.
4. Stockwell, T., Hawks, D., Lang, E., *et al.* (1996). Unravelling the preventive paradox. *Drug and Alcohol Review*, **15**, 7–16.
5. Bondy, S., Rehm, J., Ashley, M.J., *et al.* (1999). Low-risk drinking guidelines: the scientific evidence. *Canadian Journal of Public Health*, **90**, 272–6.
6. Botvin, G.J. (1995). Principles of prevention. In *Handbook on drug abuse prevention: a comprehensive strategy to prevent the abuse of alcohol and other drugs* (eds. R.H. Coombs and D. Ziedonis), pp. 19–44. Allyn and Bacon, Boston.
7. Ellickson, P.L., Collins, R.L., Hambarsoomians, K., *et al.* (2005). Does alcohol advertising promote adolescent drinking? Results from a longitudinal assessment. *Addiction*, **100**, 235–46.
8. Wyllie, A., Zhang, J.F., and Casswell, S. (1998). Positive responses to televised beer advertisements associated with drinking and problems reported by 18- to 29-year-olds. *Addiction*, **93**, 749–60.
9. Beck, J. (1998). 100 years of "just say no" versus "just say know". *Evaluation Review*, **22**, 15–45.
10. Foxcroft, D.R., Ireland, D., Lister-Sharp, D.J., *et al.* (2003). Longer-term primary prevention for alcohol misuse in young people: a systematic review. *Addiction*, **98**, 397–411.
11. Casswell, S., Gilmore, L., Maguire, V., *et al.* (1989). Changes in public support for alcohol policies following a community-based campaign. *British Journal of Addiction*, **84**, 515–22.
12. Ross, H.L. (1982). Deterring the drinking driver: legal policy and social control. Lexington Books, Lexington, MA.
13. WHO. (2004). *Global status report: alcohol policy*. World Health Organization, Geneva. http://www.who.int/substance_abuse/publications/en/Alcohol%20Policy%20Report.pdf
14. Babor, T., Caetano, R., Casswell, S., *et al.* (2003). *Alcohol: no ordinary commodity—research and public: policy*. Oxford University Press, Oxford.
15. Parliament of Victoria, Drugs and Crime Prevention Committee. (2006). *Report on inquiry into strategies to reduce harmful alcohol consumption*. Government Printer, Melbourne, http://www.parliament.vic.gov.au/dcpc/

16. Moskowitz, J.M., Mailvin, J., Schaeffer, G.A., *et al.* (1983). Evaluation of a junior high school primary prevention program, *Addictive Behaviors*, **8**, 393–401.
17. Norman, E., Turner, S., Zunz, S.J., *et al.* (1997). Prevention programs reviewed: what works? In *Drug-free youth: a compendium for prevention specialists* (ed. E. Norman), pp. 22–45. Garland Publishing, New York.
18. Carmona, M. and Stewart, K. (1996). *Review of alternative activities and alternatives programs in youth-oriented prevention*, CSAP Technical Report 13. Center for Substance Abuse Prevention, Rockville, MD.
19. Moore, M.H. and Gerstein, D.R. (eds.) (1981). *Alcohol and public policy: beyond the shadow of prohibition*. National Academy Press, Washington, DC.
20. Teasley, D.L. (1992). Drug legalization and the 'lessons' of prohibition. *Contemporary Drug Problems*, **19**, 27–52.
21. Room, R. (1993). The evolution of alcohol monopolies and their relevance for public health. *Contemporary Drug Problems*, **20**, 169–87.
22. Cook, P. (1981). Effect of liquor taxes on drinking, cirrhosis, and auto accidents. In *Alcohol and public policy: beyond the shadow of prohibition* (eds. M.H. Moore and D.R. Gerstein), pp. 255–85. National Academy Press, Washington, D.C.
23. Cook, P.J. and Moore, M.H. (1993). Violence reduction through restrictions on alcohol availability. *Alcohol Health and Research World*, **17**, 151–6.
24. Her, M., Giesbrecht, N., Room, R., *et al.* (1999). Privatizing alcohol sales and alcohol consumption: evidence and implications. *Addiction*, **94**, 1125–39.
25. Schechter, E.J. (1986). Alcohol rationing and control systems in Greenland. *Contemporary Drug Problems*, **13**, 587–620.
26. Norström, T. (1987). Abolition of the Swedish alcohol rationing system: effects on consumption distribution and cirrhosis mortality. *British Journal of Addiction*, **82**, 633–41.
27. Hastings, G., Anderson, S., Cooke, E., *et al.* (2005). Alcohol marketing and young people's drinking: a review of the research. *Journal of Public Health Policy*, **26**, 296–311.
28. Giesbrecht, N., Conley, P., Denniston, R., *et al.* (eds.) (1990). *Research, action and the community: experiences in the prevention of alcohol and other drug problems*. DHHS Publication No. (ADM) 89-1651, Office of Substance Abuse Prevention, Rockville, MD.
29. Greenfield, T. and Zimmerman, R. (eds.) (1993). *Experiences with community action projects: new research in the prevention of alcohol and other drug problems*. DHHS Publication No. (ADM) 93-1976, Center for Substance Abuse Prevention, Rockville, MD.
30. Holmila, M. (ed.) (1997). *Community prevention of alcohol problems*. Macmillan, Basingstoke, UK.
31. Holder, H.D. (1998). *Alcohol and the community: a systems approach to prevention*. Cambridge University Press, Cambridge, UK.
32. WHO Expert Committee on Problems Related to Alcohol Consumption. (2007). *Second report*. Technical Report Series, No. 944, WHO, Geneva.
33. Room, R. (1997). Voluntary organizations and the state in the prevention of alcohol problems. *Drugs & Society*, **11**, 11–23.
34. Room, R. (1983). Alcohol and crime: behavioral aspects. In *Encyclopedia of crime and justice*, Vol. 1 (ed. S. Kadish), pp. 35–44. Free Press, New York.
35. Finney, J.W. and Monahan, S.C. (1998). Cost-effectiveness of treatment for alcoholism: a second approximation. *Journal of Studies on Alcohol*, **57**, 229–43.
36. Long, C.G., Williams, M., and Hollin, C.R. (1998). Treating alcohol problems: a study of program effectiveness and cost effectiveness according to length and delivery of treatment. *Addiction*, **93**, 561–71.
37. Babor, T.F. and Grant, M. (1994). Randomized clinical trial of brief interventions in primary health care: summary of a WHO project (with commentaries and a response). *Addiction*, **89**, 657–78.
38. Holder, H.D., Lennox, R.D.L., and Blose, J.O. (1992). Economic benefits of alcoholism treatment: a summary of twenty years of research. *Journal of Employee Assistance Research*, **1**, 63–82.
39. Holder, H.D. and Cunningham, D.W. (1992). Alcoholism treatment for employees and family members: its effect on health care costs. *Alcohol Health and Research World*, **16**, 149–53.
40. Richmond, R., Heather, N., Wodak, A., *et al.* (1995). Controlled evaluation of a general practice-based brief intervention for excessive drinking. *Addiction*, **90**, 119–32.
41. Roche, A.M. and Freeman, T. (2004). Brief interventions: good in theory but weak in practice? *Drug and Alcohol Review*, **23**, 11–18.
42. Conigliaro, J., McNeil, M., Kraemer, K., *et al.* (1997). Are patients diagnosed with alcohol abuse in primary care ready to change their behavior? *Journal of General Internal Medicine*, **12**(Suppl. 1), 113.
43. Edwards, A.G.K. and Rollnick, S. (1997). Outcome studies of brief alcohol intervention in general practice: the problem of lost subjects. *Addiction*, **92**, 1699–704.
44. Smart, R.G. and Mann, R.E. (1990). Are increased levels of treatment and alcoholics anonymous large enough to create the recent reduction in liver cirrhosis? *British Journal of Addiction*, **85**, 1385–7.
45. Holder, H. (1997). Can individually directed interventions reduce population-level alcohol-involved problems? *Addiction*, **92**, 5–7.
46. Smart, R.G. and Mann, R.E. (1997). Interventions into alcohol problems: what works? *Addiction*, **92**, 9–13.
47. DeJong, W. and Hingson, R. (1998). Strategies to reduce driving under the influence of alcohol. *Annual Review of Public Health*, **19**, 359–78.
48. Crombie, I.K., Irvine, L., Elliott, L., *et al.* (2007). How do public health policies tackle alcohol-related harm? A review of 12 developed countries. *Alcohol and Alcoholism*, **42**, 492–9.
49. Chisholm, D., Rehm, J., van Ommeren, M., *et al.* (2004). Reducing the global burden of hazardous alcohol use: a comparative cost-effectiveness analysis. *Journal of Studies on Alcohol*, **65**, 782–93.
50. Baggott, R. (1990). *Alcohol, politics and social policy*. Avebury, Aldershot, UK.
51. Sutton, C. (1998). *Swedish alcohol discourse: constructions of a social problem*. Uppsala University Library, Studia Sociologica Upsaliensia 45, Uppsala.

4.2.3 Other substance use disorders

Contents

4.2.3.1 Opioids: heroin, methadone, and buprenorphine

Soraya Mayet, Adam R. Winstock, and John Strang

Opium is derived from the seed of the opium poppy (Papaver somniferum) and has been used for thousands of years for its analgesic and euphoriant effects.

What are opioids?

Opioids are drugs which mimic endogenous opioid peptides (e.g. endorphins and enkephalins) and activate opioid receptors. Opioids include both naturally occurring opiates extracted from opium (morphine and codeine) and synthetic opioids (heroin and methadone). Heroin is the most common opioid to be misused partly because of its distinctive euphoriant effects.

Neurobiology of opioids

Opioid receptors are widely distributed throughout the central nervous system. Activation of opioid receptors inhibits activity of the dorsal horn. The three most important opioid receptors are μ-receptor, δ-receptor, and κ-receptor. The μ-receptor is concentrated in brain areas and is mostly involved in nocioception (pain sensation). μ-receptor and δ-receptor activation cause hyperpolarization of neurones by activating K+ channels involving G-proteins. Exogenous opioids such as heroin and methadone act at all opioid receptors but particularly at μ-receptors. Opioid receptor antagonists such as naloxone reverse the effects of opioid agonists.[1]

The precise mechanisms of tolerance and dependence to opioids are not clear. Short term use of opioids causing intoxication may not lead to neuroadaption of opioid receptors. However, continued use of opioids may lead to opioid receptor desensitization and dependency. Acutely, opioids lead to the inhibition of adenylate cyclase with reduced conversion of ATP to cAMP, resulting in reduced firing at noradrenergic neurones located on the locus coeruleus. Chronic opioid administration leads to compensatory upregulation of cAMP, returning levels towards baseline. On cessation of opioid use (or following opioid receptor antagonism), withdrawal ensues, characterized by a massive surge in unopposed noradrenergic activity (termed the 'noradrenergic storm') from the locus coeruleus. This noradrenergic hyperactivity is thought to underlie many symptoms of opioid withdrawal, and explains some of the efficacy of the presynaptic α2 agonists in the treatment of the symptoms of acute opioid withdrawal. Opioid receptors can readapt back to normal in the absence of opioids.[2]

Both glutamate and γ-aminobutyric acid (GABA) are also likely to be involved. Positive reinforcement is thought to be mediated via the dopaminergic mesolimbic system. In the ventral tegmental area, GABA inhibits dopaminergic neurones, which in turn are inhibited via μ-opioid receptor activation. Consequently, opioid use leads to increased dopaminergic activity which is thought to mediate the drive to use and its positive reinforcement.

Route of administration

Heroin may be administered by a number of routes including injecting (intravenous, intramuscular and subcutaneous 'skin popping'), snorting and 'chasing the dragon' (inhaling after heating on tin foil).[3] These different routes of administration have profound effects on bioavailability, speed of onset, severity of dependence and physical complications. Different types of heroin may be preferentially used by different routes of administration.[4] Brown heroin is poorly water soluble with a high oil content which 'runs' well on a heated foil making it better for 'chasing'. In contrast, white heroin tends to be more water soluble and better suited for intravenous use, although it may still be snorted or smoked after preparation.

Whilst smoking heroin ('chasing the dragon') is probably the most commonly used route of self-administration, it is not as effective or efficient as injecting. Consequently, as heroin users develop tolerance, many subsequently change to the intravenous route. Injecting heroin use is also associated with a greater risk of fatal overdose and hence reducing the transition from smoking to injecting may be associated with reduced harms. Although injecting in the upper limbs is the most common site for administration, as venous access becomes compromised increasingly risky sites such as the groin or neck my be used. However, non-injecting routes of administration are not without risks and may still result in dependence and similar treatment outcomes.

Opioid metabolism

The oral bioavailability of heroin (Diacetylmorphine) itself is poor due to complete first pass metabolism, which is the reason it is often administered by alternative routes. Following administration, heroin is rapidly metabolized to 6-monoacetylmorphine (6-MAM), which is the only metabolite that specifically indicates that heroin has been used. 6-MAM is metabolized to morphine, which is catalyzed in the liver. Morphine (either as a metabolite of heroin or given as a drug) is mostly metabolized by UDP-glucuronosyltransferase (UGT) to the inactive metabolite morphine-3-glucuronide (M3G) and in lesser amounts to the active morphine-6-glucuronide (M6G). Morphine is also ***N***-demethylated to normorphine by hepatic CYP3A4 and CYP2C8 enzymes (Fig. 4.2.3.1.1). Heroin and its metabolites can be monitored in the blood, hair, saliva and urine. 6-MAM is detectable up to 12 h post administration, whilst morphine can be detected in the urine for several days after use.[5] Methadone is primarily metabolized in the liver by CYP3A4 to the inactive metabolite 2-ethylidine-1,5-dimethyl-3,3-diphenylpyrrolidine (EDDP), then to 2-ethyl-5-methyl-3,3-diphenylpyraline (EMDP). Buprenorphine is metabolized to norbuprenorphine and to conjugated buprenorphine and norbuprenorphine. Codeine (3-methylmorphine) is synthesized from morphine and metabolized to codeine-6-glucuronide. Additionally codeine is

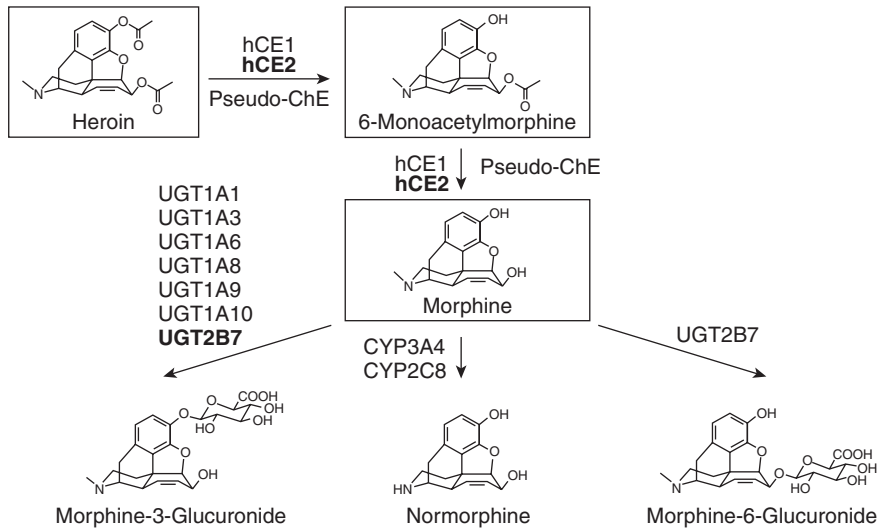


Fig. 4.2.3.1.1 Major metabolic pathways of heroin and morphine in humans.[6] Reproduced from Maurer, H.H., Sauer, C., Theobald, D.S. (2006). Toxicokinetics of drugs of abuse: current knowledge of the isoenzymes involved in the human metabolism of tetrahydrocannabinol, cocaine, heroin, morphine, and codeine. The Drug Monit. 2006 June; **28**(3), 447–53, copyright 2006, Lippincott Williams & Wilkins

metabolized to morphine by CYP2D6. Therefore a morphine positive result is possible after consuming codeine only.[5] Dihydrocodeine is metabolized to nordihydrocodeine and dihydromorphine and cannot normally be metabolized to codeine or morphine.

Epidemiology

The illicit nature of opioid use means that it is often difficult to estimate the exact prevalence. The European Monitoring Centre for Drugs and Drug Addiction reports that heroin use in the general population is less than one per cent.[6] Gender differences in opioid use have been reported with women accounting for approximately one third of all opioid users. Non-dependent recreational heroin use has been reported, but is rare. Dependence will often develop gradually over the first few years, most commonly in the late teens and early twenties. Treatment can alter the course of opioid dependence, by prolonging periods of abstinence and improving outcomes (see later sections). The risk of death from heroin dependence is approximately 12 times that of the general population, with most deaths occurring in males (> 80 per cent). Opioid use has a major effect on crime and it has been estimated that half of all crime is drug related, with estimated costs within the UK criminal justice system at £1 billion per annum in 1996.

The effects of opioids

The effects of opioids are outlined in Table 4.2.3.2.1. The effects vary depending on dose and route.

Opioid dependence and withdrawal

Opioid dependence may be classified according to the ICD-10 criteria (see appendix). Continued use of heroin (or other opioids) tends to lead to physiological dependence with the development of tolerance and withdrawal symptoms on discontinuing heroin. Tolerance occurs when the same dose gives a reduced effect or conversely, an increased opioid dose is required to have the same effect. Once physiological dependence is established, abrupt cessation or a marked reduction in dose will result in a withdrawal syndrome. During this time there is an 'undoing' of the neuroadaptation which had occurred during the development of tolerance. Withdrawal leads to 'gooseflesh' (piloerection) of the skin (which is the reason behind the term 'cold turkey'). Insomnia (with increase in Rapid Eye Movement sleep) and craving for the drug may persist for weeks. Most withdrawal symptoms appear within 4–12 h and peak at 48–72 h lasting 7–10 days. Longer acting opioids such as methadone may result in a more prolonged withdrawal. Opioid withdrawal is not usually considered to be life threatening.

Complications from opioid use

Complications may be biological, psychological, or social (Table 4.2.3.1.2).

Biological complications

Opioid overdose is the most common cause of death among opioid users, while blood-borne virus infection and other injecting related complications also contribute to increased morbidity and premature deaths.

(a) Opioid overdose

Opioid overdose occurs when the opioid dose exceeds the individual's tolerance to the respiratory depressant effect of the drug.

Table 4.2.3.1.1 Effects of opioids

Mood change (euphoria)
Analgesia
Drowsiness/Sleep
Respiratory depression
Cough reflex depression
Sensitization of the labyrinth with nausea and vomiting
Decreased sympathetic outflow (bradycardia and hypotension)
Lowering of the body temperature
Pupillary constriction
Constipation

Table 4.2.3.1.2 Complications of opioid use

Biological	Psychological	Social
Infections – hepatitis and HIV – bacterial endocarditis – septicaemia – abscesses – cellulitis Cardiorespiratory – pulmonary emboli – aspiration – cardiac arrthymias – respiratory depression – overdose death	Psychiatric complications – depression – suicide	Criminal behavior – fund dependence Loss of housing Unemployment Loss of family – family breakdown

It is more common when other Central Nervous System depressants are concurrently consumed. Respiratory depression is caused by the opioid action on the brain stem nuclei and death can follow within minutes of injecting excessive amounts. Risk factors for opioid overdose are injecting use, return to opioid use following recent abstinence (such as following detoxification and release from prison), during the early stages of dependence and starting opioid substitution treatment.[7] In addition, variability in opioid purity, increased central depressant effects following polysubstance use and high levels of psychiatric co-morbidity may increase the vulnerability for accidental or intentional overdose. Opioid overdose can be rapidly reversed if naloxone (opioid antagonist) is administered to the person who has overdosed. The management of an opioid overdose is described later in the chapter.

(b) Blood borne virus transmission

The risk of viral transmission is high among injecting drug users, and routine testing with counseling should be available to those at risk. Transmission of blood borne viruses is primarily related to sharing injecting equipment and involvement in the sex industry. Needle sharing appears to occur less frequently than sharing of spoons and filters, but any shared equipment may pose the risk of viral transmission. The prison population is particularly at risk, as injecting is more common likely to involve sharing injecting equipment.

Opioid substitution treatment is one of the most effective interventions for reducing the extent of injecting, thereby also reducing both the spread of and morbidity from blood borne viruses. Stabilization on methadone/buprenoprehine with abstinence from injecting, needle and other injecting equipment sharing, and unprotected sex should be encouraged among all dependent heroin users. Blood borne virus testing and referral to specialists for treatment is important. In addition, close liaison with medical and psychiatric services is important for improving outcomes and compliance with treatment.

(i) Human immunodeficiency virus (HIV)

Rates of HIV seropositivity amongst current injecting drug users in England and Wales although low, have recently increased, with an incidence of 1.5 per cent reported in 2004 (the highest since 1992), with higher rates reported in London.[8] The relatively low levels of HIV within the UK and elsewhere in Europe is believed to be due to the widespread availability of 'needle-exchange' services and provision of services focused towards 'harm minimization'.[9]

(ii) Hepatitis B & C

Screening of drug users in treatment has revealed prevalence rates of 20 per cent for Hepatitis B and more than 50 per cent for Hepatitis C. Intravenous drug users are likely to have higher rates estimated at 30 to 50 per cent (Hepatitis B) and up to 90 per cent (Hepatitis C).[10] Prognosis is worsened by high levels of alcohol consumption and therefore liaison with hepatitis/gastroenterology services is important. Screening for Hepatitis B and C and targeted vaccination for Hepatitis B, in addition to education and harm-reduction provision, should be provided.

Psychological complications

Numerous large epidemiological studies have identified that co-morbid psychiatric illness is common among those with opioid dependence, with prevalence rates of about 50 per cent.[11] Concurrent use of benzodiazepines, alcohol, and especially stimulant drugs increases psychiatric morbidity in addition to female gender, poor physical function, and difficulties in personal relationships.[12] Many of those with opioid dependence will have had childhood behavioural problems such as conduct disorder which may be a marker for subsequent drug use.[13] The rate of suicide among heroin users is estimated at 14 times that of age matched peers with reports of 3–35 per cent of deaths related to suicide. Risk factors include a history of depression, poly substance use in addition to generic risk factors.[14]

At entry into treatment, it may be difficult to make an accurate determination of an opioid user's psychiatric diagnosis and generally assessment should be repeated once stability on a substitution medication or detoxification has been achieved. Waiting to review a patient's mental health once they are out of crisis can prevent early misdiagnosis since much of psychiatric distress dissipates rapidly on cessation of illicit use and stabilization. In one follow up study of opioid users entering treatment, baseline levels of depression fell from 25 per cent at baseline to 11 per cent at 12 month follow up, with the observed decline being strongly related to treatment exposure.[15] Follow-up and provision of treatment for co-morbid disorders are thus essential. Enhancing compliance with prescribed medications through supervised dispensing or engagement of carers is useful. Treatment can be effective in reducing psychiatric distress observed on entry to treatment in this group. However, in a significant proportion, even on cessation of use or abstinence, major psychiatric illnesses can persist. If left untreated, co-morbid conditions can lead to a poorer prognosis particularly in respect to relapse and suicide.

Social complications

The ramifications of heroin dependence upon individual functioning and that individual's ability to relate and function within their family and community are immense. Although high rates of socioeconomic disadvantage often precede entry into heroin use, it is the other associated problems of relatively poorer premorbid functional and educational attainment that frequently compound later efforts at rehabilitation. High rates of criminal activity, homelessness, and unemployment are associated with opioid addiction, although treatment can improve socio-economic status.

Treatment of opioid dependence

Opioid maintenance treatment

Opioid maintenance treatment generally involves substituting heroin for an oral long-acting opioid, thereby reducing the plasma level variability and stopping injecting drug use. Oral methadone and buprenorphine have been licensed in several countries for use in the treatment of opioid dependence and both are approved in the UK.[16] The decision regarding whether methadone or buprenorphine is used should be based on individual factors estimating the risks and benefits and following discussion with the patient. If both drugs are equally suitable, cost-effectiveness examination by the National Institute of Clinical Excellence (NICE) concludes that methadone should be prescribed as the first choice.[16]

(a) Methadone maintenance treatment

Methadone is a synthetic orally active full opioid agonist with a half-life of 24–36 h, making it suitable for daily administration. This is an effective treatment for heroin dependence and has significantly better outcomes than non pharmacological substitution for retaining patients in treatment, decreasing heroin use, reducing crimes, reducing overdose deaths, reduced injecting and sharing of injecting equipment and consequent reduced risk behaviours leading to transmission of HIV.[17] Higher doses of methadone (60 to 120 mg/day) have been shown to be more effective than lower dosages[18] and doses greater than 80 mg daily are believed to provide a reasonable level of opioid receptor blockade such that euphoria from illicit opioids is diminished.

Methadone steady state plasma levels take approximately 4 to 5 days and so there is potential for methadone accumulation (and overdose) when initiating treatment. Deaths have been recorded during the induction phase onto methadone especially when the recipient is not as tolerant as believed or is using other opioids or substances.[19] Therefore, confirmation of the patient's dependent status is paramount as described in the chapter and it is safer to start treatment at low doses (not more than 30 mg daily). Treatment should be initiated under supervision with oral methadone liquid where consumption can be easily monitored. Doses should generally be increased slowly, and titrated against withdrawal symptoms. As methadone can accumulate, there should be increased observation over the first few weeks of treatment. Doses can then gradually be increased to within therapeutic levels (60–120 mg daily). Prescribing should follow the guidelines outlined later in the chapter.

(b) Buprenorphine maintenance treatment

Buprenorphine (Subutex) is a synthetic partial opioid agonist which is given as a sublingual tablet and has a high affinity at μ-opioid receptors. It is an effective treatment for use in maintenance treatment for heroin addiction, but is not more effective than methadone at adequate dosages.[20]

Buprenorphine undergoes extensive first metabolism. Therefore, it is administered as sublingual tablets (2 mg and 8 mg) with bioavailability of between 30–40 per cent. Optimizing sublingual absorption while minimizing diversion is a practical challenge that supervised dispensing points still need to address. Taking 5–7 minutes to dissolve, buprenorphine reaches a plateau on most physiological subjective effect at a daily dose of 4–16 mg. At higher doses, the duration of action increases, permitting less than daily dosing in about one-third of patients. Because of its high affinity for the opioid receptor, buprenorphine will precipitate withdrawal if administered to someone with an opioid agonist on board (typically within 8 h of heroin use or 24–36 h of methadone use). Therefore, patients are advised to wait until they are in mild withdrawal before commencing treatment. Finally, it is thought to be safer in overdose (with a 'ceiling' in the respiratory depression, unlike with full opioid agonists) and so induction can be quite rapid aiming for doses of between 8–16 mg by day 3.

It is an effective analgesic but may be less suitable than methadone for those with chronic pain. Although safer in overdose than methadone, fatal overdose can occur especially when taken in combination with other substances. As with methadone dose, reduction in someone stable on maintenance should be done gradually (typically 2 mg ever 2 weeks) and only supported when there is evidence of continued abstinence from illicit opioid use. Continued dose reduction in the face of return to illicit use is likely to further destabilize the patient. In some countries (e.g. many European countries since 2007), another preparation has become available (Suboxone), in which buprenorphine has been combined with naloxone in a ratio of 4:1 to reduce the desirability of injecting diverted medication. When taken as directed (sublingual), the bioavailability of naloxone is very poor, whereas when injected by a dependent opiate user, a severe withdrawal reaction may be precipitated.

(c) Other opioid maintenance treatments

Injectable methadone and injectable heroin are rarely prescribed in the United Kingdom and at present there is insufficient evidence to guide this use.[21] However, it may be considered in some patients as a 'second-line' treatment option for whom an adequate trial (e.g. at least 6 months) of optimized methadone maintenance treatment (e.g. doses > 80 mg daily, regular supervised dosing, regular appointments and appropriate management of medical or psychiatric co-morbidity) is ineffective in controlling illicit injecting heroin use. This should only be initiated by a specialist. Both long-acting morphine and dihyrocodeine have been compared to methadone in two randomized controlled trials which have revealed broadly equivalent outcomes between the groups.[23,23]

Opioid detoxification

Detoxification may be based on suppression of the 'nor-adrenergic storm' that accompanies opioid withdrawal, by prescribing either α2-receptor agonists or a gradual reduction of an opioid agonist. When a patient is stable and motivated, a gradual reduction in the dose of opioid maintenance dose can be an effective means for attaining abstinence. However, it must be noted that detoxification is generally associated with poor long term rates of abstinence and retention in treatment.

(a) Buprenorphine detoxification

Buprenorphine has been found to be more effective than clonidine (below) for the management of opioid withdrawal. In addition, there is no significant difference between buprenorphine and methadone in terms of completion of treatment, but withdrawal symptoms may resolve more quickly with buprenorphine.[24] This can be undertaken as an inpatient or outpatient. Buprenorphine should only be taken after cessation of heroin use as it can precipitate withdrawal. Withdrawal may be achieved by the burpenorphine dose being stabilized according to withdrawal over

a 24–48 h period, after which a gradual dose reduction should occur over 5–21 days in the inpatient or outpatient setting.

(b) Methadone detoxification

Methadone detoxification can be used for pharmacologically assisted opioid detoxifications. A review revealed that withdrawal programs vary widely with regard to duration, design and treatment objectives. This review confirmed that slow tapering with temporary substitution of methadone accompanied by medical supervision and ancillary medications can reduce withdrawal severity. Nevertheless, the majority of patients relapsed to heroin use.[25]

(c) α2-Agonist assisted detoxification

α2-agonists, such as lofexidine and clonidine, reduce pre-synaptic nor adrenaline release alleviating many withdrawal symptoms associated with opioid withdrawal. The dose is titrated against the symptoms and signs of withdrawal whilst also avoiding hypotensive episodes. Detoxification can occur in both outpatient and inpatient settings. Lofexidine that has a lower incidence of hypotension than clonidine is the preferred non-opioid method of assisting opioid withdrawal. A review has concluded that there is no significant difference in efficacy for treatment regimes based on clonidine and lofexidine, compared to reducing doses of methadone over a period of around 10 days, for the management of opioid withdrawal.[26]

(d) Naltrexone assisted detoxification

Some research has been published suggesting that withdrawal may also be completed more quickly by additionally administering the long-acting opioid antagonist naltrexone. However, further research is still needed to confirm effectiveness and safety of this treatment.

Naltrexone for relapse prevention

Oral naltrexone has recently been recommended in the UK for the management of opioid dependence. Naltrexone is relevant for highly motivated patients who have completed an opioid detoxification. This may be combined with arrangements for supervision and should be given as part of a programme of supportive care. Regular reviews of effectiveness of naltrexone should be undertaken by the clinician and discontinuation of treatment should be considered if there is evidence of continued opiate misuse.[27]

Psychosocial interventions for opioid dependence

There are numerous psychosocial approaches that are currently used in the management of substance misuse. Treatment based on a holistic approach encompassing biological, psychological and social aspects of care is likely to improve outcomes. Psychosocial treatments such as relapse prevention, motivational interviewing and contingency management are often used in the addiction setting. Cognitive behaviour therapy and other types of psychotherapy are also used particularly for those with a specific psychiatric disorder such as depression and can be referenced elsewhere in the book.

Relapse Prevention looks at identification of triggers for craving (e.g. people, places, or moods) and uses learning techniques (distraction, relaxation) to handle high-risk situations.[28] Motivational interviewing techniques can help patients move along a 'cycle of change' from pre-contemplation (no interest in changing behaviour) to contemplator, to determination and action without confrontation.[29] This is based on five key principles (Table 4.2.3.1.3) which are often used in both addiction and eating disorders, but may also be used in any aspect of the doctor-patient relationship where the patient is ambivalent about change.[30] Although used rarely in the UK, Contingency Management (CM) has a stronger evidence base. This behavioural treatment uses rewards or other reinforcers to promote abstinence or other selected goals. Narcotics Anonymous is based on the 12-step program (the philosophy behind Alcoholics Anonymous) where the individual accepts they have a drug problem and uses 12 steps (Table 4.2.3.1.4) within the group setting, to attain abstinence.[31]

Table 4.2.3.1.3 Principles of motivational interviewing[30]

- express empathy
- help the client to see discrepancies in their behaviours
- avoid argument
- roll with resistance
- support the patient's sense of self-efficacy.

Reproduced from Miller, W. and Rollnick, S. Motivational Interviewing, copyright 1991, Guildford Press NY.

Psychosocial interventions alone have not been shown to be more effective than no treatment.[32] However psychosocial interventions combined with pharmacological interventions have been shown to lead to better outcomes.[33]

Educating the drug user about safe practices (including safer injecting, not sharing equipment) and harm-reduction techniques is important, as is appropriate liaison with other agencies such as social services or voluntary sector supports. Therapeutic communities, residential rehabilitation and 'concept houses' based on a religious or abstinent theme offer longer-term care.

Management of opioid dependence

The treatment plan should be made jointly between the clinician and patient. The actual management plan depends on whether the person has opioid dependence, the amount of opioids used, and the outcome of the mutually agreed treatment objectives and

Table 4.2.3.1.4 12 Steps of Narcotics Anonymous[31]

1. We admitted that we were powerless over our addiction, that our lives had become unmanageable.
2. We came to believe that a Power greater than ourselves could restore us to sanity.
3. We made a decision to turn our will and our lives over to the care of God as we understood Him.
4. We made a searching and fearless moral inventory of ourselves.
5. We admitted to God, to ourselves, and to another human being the exact nature of our wrongs.
6. We were entirely ready to have God remove all these defects of character.
7. We humbly asked Him to remove our shortcomings.
8. We made a list of all persons we had harmed, and became willing to make amends to them all.
9. We made direct amends to such people wherever possible, except when to do so would injure them or others.
10. We continued to take personal inventory and when we were wrong promptly admitted it.
11. We sought through prayer and meditation to improve our conscious contact with God as we understood Him, praying only for knowledge of His will for us and the power to carry that out.
12. Having had a spiritual awakening as a result of these steps, we tried to carry this message to addicts, and to practice these principles in all our affairs.

treatment plan. Opioid dependence management is sometimes based on 'abstinence', where the person refrains from taking drugs, but also needs to be based on the principles of harm reduction. Harm reduction aims to reduce harms from opioid use often in terms of reducing deaths, spread of blood-borne viruses or improving psychosocial outcomes. This may optimally be achieved through cessation of injecting illicit opioid use by stabilizing the person on an opioid replacement.

Assessment of opioid use and dependence

A comprehensive assessment of opioid use patterns and associated risks forms the basis of any treatment plan. A suggested plan of enquiry that allows both accurate diagnosis and risk assessment is outlined below.

(a) Clinical assessment

1 **Current consumption:** How much heroin (or other opioid) is consumed on a typical day, in terms of either weight or money spent, and for how long has consumption been at this level? Are opioids taken daily? What happens when no opioids are taken? The route of use (smoking, intravenous injection), number of administrations per day, and the minimum amount of opioid required each day to avoid withdrawal symptoms should also be assessed. Enquiry should be made to determine if more than one type of opioid is being used (e.g. prescribed medication or street methadone). All other substances being used should be identified (e.g. cocaine, benzodiazepines and alcohol) and their patterns of use.

2 **Typical day:** A systematic enquiry of the person's typical day and the use of opioids and other substances is particularly important for assessing evidence of withdrawal. This can also be useful for identifying risk activities that the user engages in to support their ongoing use, including criminal activity, high risk sex, or injecting practices.

3 **Drug use history:** A careful enquiry needs to be made to determine the temporal relationship between the onset of drug use and any psychological behaviors. The age of first use and the psychosocial precipitants of use should be established, as should the development of tolerance and craving through increased frequency of use, escalating dose, and where relevant, the onset of injecting.

4 **Biopsychosocial complications:** This should assess episodes of overdose (intentional or accidental), viral screening status and injecting behaviour including use of needle exchanges, sharing equipment and use of high-risk injection sites such as the groin and neck. Effects on family relationships, employment and criminal activity, and co-morbid psychiatric conditions related to opioid use should be assessed.

5 **Past treatments and abstinent periods:** Have they ever been in contact with treatment services, maintained on substitute medication or undertaken a detoxification? What was their longest period of abstinence? What has helped in the past? When and why did relapses occur? What are high-risk situations and other triggers for use?

6 **Motivation for change:** Why seek treatment now? What support is needed? Thinking of **L**ivelihood (financial), **L**ife (physical health), **L**ove life, **L**egal problems and **L**osing it (loss of control) can be helpful as these are often precipitants for seeking treatment and can be used to encourage behavioural change.

(b) Confirmation of dependence

Although a diagnosis of opioid dependence can be made by taking a full history; dependence and tolerance to opioids should be corroborated before commencing substitute treatment. Confirmation of dependence and tolerance prior to commencing treatment is important since the greatest risk associated with prescribing methadone is the possibility of overdose following consumption of methadone by a non-tolerant individual. Corroborative information may come from urine drug screens and other health care practitioners such as the GP or criminal justice worker. Physical examination is essential and may reveal stigmata of injecting drug use such as evidence of recent intravenous injection sites or the more long-term 'track marks' (linear scarring along veins from repeated intravenous use) on the drug user's limbs. All patients should be asked where they usually inject and recent injection sites should be examined. In those with poor upper limb veins, with evidence suggestive of ongoing illicit use, it may be appropriate to examine groin or neck sites.

Urines tests are supportive of use but not confirmatory of dependence. Sequential urine testing over a few days may allow the confirmation of regular opioid (although not necessarily heroin) consumption. Measurement of withdrawal by using withdrawal scales (Table 4.2.3.1.5) in addition to examining for presence of tachycardia and hypertension, following a period of abstinence, is helpful in confirming withdrawal. The Objective Opiate Withdrawal Scale (Table 4.2.3.1.5) is useful when assessing whether a person is in withdrawal before commencing opioid substitution treatment and useful after commencing treatment to aid assessing whether the opioid substitute dose is adequate. Partial or full reversal of withdrawal following a measured dose of opioids administered on site will provide some information of the patient's level of tolerance.

Table 4.2.3.1.5 Objective opiate withdrawal scale[27]

This is to be completed by clinician. A score should be given for each observation within a 5 minute observation period.

Observations	Scoring
1. Yawning	0 = no yawns 1 = ≥ 1 yawns
2. Rhinorrhoea	0 = <3 sniffs 1 = ≥3 sniffs
3. Piloerection (observe arm) 'gooseflesh'	0 = absent 1 = present
4. Perspiration	0 = absent 1 = present
5. Lacrimation	0 = absent 1 = present
6. Tremor	0 = absent 1 = present
7. Mydriasis	0 = absent 1 = ≥ 3mm
8. Hot and cold flushes	0 = absent 1 = shivering / huddling for warmth
9. Restlessness	0 = absent 1 = frequent shifts of position
10. Vomiting	0 = absent 1 = present
11. Muscle twitches	0 = absent 1 = present
12. Abdominal cramps	0 = absent 1 = Holding stomach
13. Anxiety	0 = absent 1 = mild – severe
TOTAL SCORE	

National Institute for Health and Clinical Excellence (NICE) (2005) TA115 Drug misuse–naltrexone. London: NICE. Available from www.nice.org.uk/ TA115 Reproduced with permission.

Investigations for patients with opioid dependence

Heroin dependence is associated with high rates of physical and psychiatric morbidity. Since access to primary health care may be difficult, basic physical health checks should be a fundamental part of all drug treatment. The core assessment should include history, physical examination, routine blood tests, blood borne virus screening and vaccination where appropriate, and assessment of nutritional status, mental and dental health. Referral to appropriate specialist services should be facilitated through coordinated care planning which should form the cornerstone of structured drug treatment delivery.

Management options for opioid dependence

Following a careful and thorough assessment which allows confirmation of dependence as outlined above, there are several treatment options. In the short term, opioid substitution treatments with methadone or buprenorphine may be offered. Opioid substitution treatment should be monitored closely, especially during the initial phase of treatment. Initiation of opioid substitution treatment usually takes place in the community setting, but may occur in an inpatient setting if there is a complicated history involving medical or psychiatric morbidity. Opioid maintenance treatment is continued in the community setting and can be continued as maintenance for the long term, with some patients continuing treatment for over 50 years. Treatment for opioid dependence is likely to be improved when framed within a comprehensive treatment package including psychosocial and pharmacological interventions.[32] Relapse prevention and motivational interviewing can be carried out during regular appointments with health professionals. Community substance misuse teams often use keyworking as a model of care for opioid dependent patients. This is where one healthcare professional looks after the patients care and provides the main source of contact, often following up referrals and ensuring medical and psychosocial needs are met. Opioid maintenance treatment within the community substance team may continue in the long term or be a prelude to starting abstinence based treatments, with the aim of stopping all illicit heroin use.

Abstinence can be achieved by pharmacologically assisted withdrawal from the opioid via a detoxification with either lofexidine or an opioid substitute such as methadone or buprenorphine (as above). This usually takes 10–21 days but may be longer if withdrawal from a longer acting opioid is required. Opioid detoxification may take place in the inpatient or community setting, depending on the level of medical and psychiatric morbidity. Following opioid detoxification patients may benefit from treatment within a therapeutic community or residential rehabilitation centre where life skills for dealing with a world without opioids may be developed. For patients who do not enter rehabilitation centres, regular sessions with keyworkers using relapse prevention and motivational interviewing may also be beneficial. Highly motivated patients may also benefit from naltrexone to prevent relapse (described above).

Prescribing opioids for opioid dependence

In the United Kingdom, all doctors may prescribe methadone or buprenorphine for the treatment of dependence, although prescribing should generally be initiated by a specialist or special interest general practitioner. Prescriptions for opioid dependence should ideally be dispensed daily with supervised doses, particularly for the first three months of treatment. After about three months, if a patient is stable (based on psychosocial outcomes and illicit substance use) the number of dispensings per week and level of supervision may be reduced, although this may be varied when there are extenuating circumstances. UK guidelines advocate that no more than a week's medication is dispensed at a time.

Liaison between the pharmacist, the general practitioner and the specialist is important to prevent double scripting, reduce diversion and improve safety. Clinicians in all specialties should be aware of the potential for all opioid-containing analgesics to be diverted for abuse or develop into iatrogenic dependence. Repeat prescriptions of such analgesics should be carefully reviewed. Prescriptions for controlled drugs can only be for 28 days in total and the total prescribed amount must be written in words and figures. It is recommended that the prescription should state the name of the pharmacy where the prescription is to be dispensed, how often the prescription should be dispensed and whether the dose should be consumed under supervision. Installment prescriptions for daily dispensing are available in the UK for burpenorphine and methadone. Only doctors in possession of a Home Office license are able to prescribe heroin for the treatment of opioid dependence.[34]

Opioid overdose management

Opioid overdose management training is particularly important as early recognition (Table 4.2.3.1.6) of an opioid overdose and prompt action (Table 4.2.3.1.7) can save lives. The antidote to heroin is naloxone and this should be given if an opioid overdose is suspected. Although intravenous naloxone is quicker acting, venous access may be difficult and therefore, intramuscular injection may be preferable and also results in a more gradual reversal of the overdose which may be less likely to provoke aggression. Hospital monitoring should always be recommended, since the plasma half-life of naloxone is shorter (<1 h) than the physiological effects of heroin (4–6 h) and methadone (24–36 h). In addition many overdoses are a result of concomitant substance use—the effects of which will not be reversed by naloxone alone. Opioid overdoses involving buprenorphine (partial agonist) will not be readily be reversed by naloxone; however buprenorphine is believed to have much less respiratory depressant effects.

The supply of 'take home' emergency naloxone may help reduce opioid related overdose deaths. It has, therefore, been advocated that providing take-home naloxone in combination with opioid overdose management training to opioid using patients and their families may help reduce deaths.[35]

Table 4.2.3.1.6 Recognition of an opioid overdose

Recognition
Respiratory Arrest with a pulse
Pinpoint pupils (unreactive to light)
Snoring giving way to shallow respiration
Respiratory Depression (<8 breaths per min)
Bradycardia and hypotension
Varying degrees of unconsiousness

Table 4.2.3.1.7 Management of an opioid overdose

Check area safe, then try to rouse overdose victim
If unrousable - Call for help/ambulance
Check airway and breathing a. If not breathing, give 2 rescue breaths b. If breathing — place in recovery position
Administer 0.4 mg Naloxone Intramuscularly — Increase dose until adequate reversal achieved
Consider use of high flow oxygen
Patient to have medical monitoring after naloxone, as opioid overdose may re-emerge. Patients may need additional doses of naloxone

Table 4.2.3.1.8 Aims of managing the pregnant drug user

- Engage and maintain contact with the patient and partner.
- Aim to reduce risk-taking behaviours (sharing needles, prostitution).
- Stabilize on oral methadone maintenance treatment (or extremely slow detoxification, if required).
- Ensure that other drug and alcohol use are assessed routinely.
- Provide health and psychosocial care including blood borne virus screening.
- Close liaison with multi-agency teams including possible social work assessment.
- Social stability and provisions for motherhood.

Special groups

Young people

Young people with opioid problems often have other emotional and/or behavioural problems, and frequently fall between the adult and child psychiatric services as well as addiction services, compounding the difficulties in providing effective services to this group. Increasingly dedicated services have recently been developed and these integrated services focus as much on the family and re-integration with education as they do on substance use issues. Approaches effective for the adult population may be less effective in a group with less developed emotional and cognitive abilities. Separation of such a service from adult providers would also assist in preventing experienced drug users from influencing more naive users. Ultimately, a tiered approach would appear appropriate, since it would allow maximum utility of current services and focused development of new services. Generic services in primary health care could provide accurate screening with initial referral to youth-oriented services within existing departments. Beyond this, referral to specialist and super-specialist regional services could be employed to provide secure environments with the option of residential rehabilitation and therapeutic communities. Once engaged they may benefit from a range of possible therapies from family work and cognitive behavioural therapy to pharmacotherapy and self-help groups.

Pregnancy and breastfeeding

Maternal opioid use poses a risk to both the mother and foetus. Pregnancy can be a specific point when women try to address their opioid use problem. The management of the pregnant opioid user should follow the same guiding principles as for other opioid users; additionally, there should be close liaison between addictions and maternity services, the general practitioner and other relevant agencies (Table 4.2.3.1.8).

Women who are already on methadone maintenance treatment can remain on methadone but should be encouraged to stop illicit opioid use. For women who are not prescribed methadone, the first step is to initiate stabilization on methadone. Methadone maintenance may continue at a stable dose throughout pregnancy. During the third trimester, maternal metabolism may increase the need for methadone and so the dose may need to be increased or alternatively the daily dose could be split. Methadone increases the risk of respiratory depression in the neonate and should always be pre-considered for the delivery plan. Women prescribed opioids may also need increased pain relief during delivery. The long-term outcome in women who enter methadone treatment programmes during pregnancy is better in terms of their pregnancy, childbirth, and infant development, irrespective of continuing illicit drug use.[36] Methadone is not contraindicated for breast feeding.

Some women prefer to be abstinent from opioids during pregnancy. These women will often need a gradual pharmacologically assisted detoxification, which should be avoided in the first trimester, ideally undertaken in the second trimester and undertaken with caution in the third trimester. Detoxification is best undertaken within a dedicated inpatient facility, however, this may not always be possible and therefore community detoxification can be undertaken. During stabilization on methadone or detoxification, it is important to prevent the pregnant woman experiencing opioid withdrawal as this is dangerous for both mother and fetus.

Currently, there is insufficient evidence regarding the use of buprenorphine during pregnancy or breastfeeding to be able to define its safety profile. However, women well maintained on buprenorphine prior to pregnancy refusing alternative pharmacotherapy could be kept on buprenorphine following full informed consent.

Forensic

It has been estimated that up to two-thirds of people arrested have taken substances prior to arrest, whilst approximately 15–50 per cent of the prison population were previously dependent on drugs or alcohol. It is difficult to estimate exact opioid use and dependence, but approximately one-fifth of the prison population injects drugs. Access to illicit substances is not prevented by imprisonment; indeed some users may increase or start using drugs whilst imprisoned.[37] Improving identification of drug users before sentencing is important. In addition, identification of drug related crimes and offering court diversion schemes with drug treatment interventions can be an alternative to a custodial sentence. Prisoners will benefit from education, good primary health care, blood borne virus testing, and hepatitis vaccination. In addition, prisoners can undertake opioid detoxification or commence/continue opioid maintenance treatment whilst in prison. Release from prison is associated with extremely high risk of opioid overdose death, particularly in the first few weeks, and therefore quick access to drug services following release (or prior to release) may reduce deaths.

Accessing treatment and the range of services

Those who experience problems with opioids may present to a wide range of professionals within the health-care, social, and legal systems. The range of treatment options available from statutory and non-statutory agencies in any particular area will vary, as will the provision of either maintenance or detoxification for opioid dependents depending upon differing treatment philosophies and treatment settings.

Outcomes

Heroin dependence is a chronic relapsing condition and opioid use reduces morbidity and mortality more than any other drug use.[38] Treatment saves lives and improves psychosocial function as well as reducing risk to both the individual and the community. Outcomes are broadly comparable to those seen with other chronic medical conditions.[39] Abstinence rates following treatment vary widely, but 10 to 40 per cent of treated patients are still drug free at 6 months.[40] The majority of those who relapse following treatment do so within 3 months of discharge. Longer treatment is associated with better outcomes and greater pre-treatment severity of psychopathology is associated with worse outcomes. Long-term follow-up studies suggest that successful and lasting cessation of opioid use can be a very slow process and becomes increasingly unlikely if users continue into their late thirties.

Results from three longitudinal studies ranging from 3–5 years based in the United Kingdom, Australia and United States of America have shown that treatment leads to better outcomes against all parameters as compared to no treatment.[41,42,43] Patients who enter drug treatment are more likely to significantly reduce use of heroin and other illicit drugs and longer treatment times have been associated with better treatment outcomes. Reductions in heroin use have also been mirrored by reductions in heroin overdose rates. Therefore opioid treatment increases morbidity and mortality in opioid users.

The Australian Treatment Outcome Study (ATOS) reported that half the number of opioid overdoses occurred in the participants in treatment at 1 year as compared to the same participants prior to entering treatment. In addition this study showed that levels of psychopathology reduced, physical health improved and crime rates in this population reduced following treatment.[43] The National Treatment Outcome Research Study (NTORS) based in the UK had increased rates of abstinence, reduced heroin use and reduced injecting and sharing of injecting related equipment with both residential and community treatment.[42] Findings from the Drug Abuse Treatment Outcome Studies (DATOS) based in the USA were similar to both the UK and Australia studies.[41]

Longer term outcomes have been observed in a study following up 581 male heroin users since 1962. At the 33-year follow up, 284 subjects were dead and 242 were interviewed. Over a fifth were still using heroin and 40.5 per cent admitted to using heroin in the last year. There were high rates of health problems, mental health problems, and criminal justice system involvement. Heroin abstinence was associated with improved outcomes in all these domains. Deaths increased steadily over time with a relatively stable pattern of heroin use in the subjects.[44] A further analysis of the results assessed years of life lost through heroin use and concluded that on average, 18.3 years of potential life were lost before the age 65, which is significantly higher than that of US population.[45]

Conclusion

Opioid dependence is a chronic relapsing and remitting disorder affecting a large proportion of people throughout the world with severe physical, psychological, and social consequences. Opioid overdose and spread of blood borne viruses are major causes of morbidity and mortality. Assessment of opioid use and dependence should be systematic and confirmation of dependence is of paramount importance before initiating treatment. The prescription of substitute opioids should be managed carefully to prevent harm, diversion to others and improve safety. Management of opioid dependence can greatly improve outcomes and may be based on opioid maintenance stabilization or detoxification combined with psychosocial interventions.

Further information

Department of Health (England) and the developed administrations. Drug misuse and dependence: UK guidelines on clinical management. 1999. HMSO, London. http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4078198.pdf

National Treatment Agency for Substance Misuse website for guidelines. http://www.nta.nhs.uk/

The Cochrane Library for Cochrane Systematic Reviews on Opioids. http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME

National Institute for Clinical Excellence. www.nice.org.uk/

References

1. Stahl, S.M. (2006). *Essential Psychopharmacology*, Cambridge University Press, Cambridge.
2. Simonato, M.(1996). The neurochemistry of morphine addiction in the neocortex. *Trends in Pharmacological Science*, **17**, 410–15.
3. Strang, J., Griffiths, P., Gossop, M., *et al.* (1997). Heroin in the United Kingdom: different forms, different origins, and the relationship to different routes of administration. *Drug and Alcohol Review*,**16**, 329–37.
4. Strang, J., Griffiths, P., Powis, B., *et al.(1999).* Heroin chasers and heroin injectors: differences observed in a community sample in London. *American Journal on Addictions*, **8**(2),148–60.
5. Maurer, H.H., Sauer, C., Theobald, D.S., *et al.* (2006) Toxicokinetics of drugs of abuse: current knowledge of the isoenzymes involved in the human metabolism of tetrahydrocannabinol, cocaine, heroin, morphine, and codeine. *Therapeutic Drug Monitoring,* **28**(3), 447–53.
6. Annual report on the state of the drugs problem in the European Union 2000 European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). Accessed April 2007.
7. Gossop, M., Griffiths, P., Powis, B., *et al.* (1996). Frequency of non-fatal heroin overdose. *British Medical Journal*, **313**, 402.
8. United Kingdom drug situation: annual report to the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) 2005.
9. Stimson, G.V. (1995). AIDS and injecting drug use in the UK, 1987–1993: the policy response and the prevention of the epidemic. *Social Sciences and Medicine*, **41**, 699–716.
10. Fingerhood, M.I., Jasinski, D.R., Sullivan, J.T. *et al.*(1993). Prevalence of hepatitis C in a chemically dependent population. *Archives of International Medicine*, **153**, 2025–30.
11. Verthein, U., Degkwitz, P., Haasen, C., *et al.*(2005). Significance of comorbidity for the long-term course of opiate dependence. *European Addiction Resarch*, **11**(1),15–21.

12. Marsden, J., Gossop, M., Stewart, D., *et al.* (2000). Psychiatric symptoms among clients seeking treatment for drug dependence. Intake data from the National Treatment Outcome Research Study. *The British Journal of Psychiatry*, **176**,285–9.
13. Lloyd, C.(1998). Risk factors for problem drug use: identifying vulnerable groups. *Drugs: Education, Prevention and Policy*, **5**, 217–32.
14. Darke, S., Ross, J.(2002). Suicide among heroin users: rates, risk factors and methods. *Addiction,* **97**(11),1383–94.
15. Havard, A., Teesson, M., Darke, S., *et al.* (2006). Depression among heroin users: 12-Month outcomes from the Australian Treatment Outcome Study (ATOS). *Journal of Substance Abuse Treatment*, **30**(4), 355–62.
16. NICE technology appraisal guidance 114. (2007). Methadone and buprenorphine for the management of opioid dependence.
17. Mattick, R.P., Breen, C., Kimber, J., *et al.* (2003). Methadone maintenance therapy versus no opioid replacement therapy for opioid dependence. *Cochrane Database of Systematic Reviews*. Issue 2, Art. No.: CD002209. DOI: 10.1002/14651858.CD002209.
18. Faggiano, F., Vigna-Taglianti, F., Versino, E., *et al.* (2003). Methadone maintenance at different dosages for opioid dependence. *Cochrane Database of Systematic Reviews*. Issue 3, Art. No.: CD002208. DOI: 10.1002/14651858.CD002208.
19. Caplehorn, J. and Drummer, O.H. (1999). Mortality associated with New South Wales methadone programs in 1994: lives lost and saved. *Medical Journal of Australia*, **170**, 104–9.
20. Mattick, R.P., Kimber, J., Breen, C., *et al.* (2003). Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. *Cochrane Database of Systematic Reviews.* Issue 2, Art. No.: CD002207. DOI: 10.1002/14651858.CD002207.pub2.
21. Ferri, M., Davoli, M., Perucci, C.A., *et al.* (2006). Heroin maintenance treatment for chronic heroin-dependent individuals: a Cochrane systematic review of effectiveness. *Journal of Substance Abuse Treatment*, **30**(1), 63–72.
22. Eder, H., Jagsch, R., Kraigher, D., *et al.* (2005) Comparative study of the effectiveness of slow-release morphine and methadone for opioid maintenance therapy. *Addiction,* **100**(8), 1101–9.
23. Robertson, J.R., Raab, G., Bruce, M., *et al.* (2006). Addressing the efficacy of dihydrocodeine versus methadone as an alternative maintenance treatment for opiate dependence: a randomized controlled trial. *Addiction*, **101**, 1752–9.
24. Gowing, L., Ali, R., White, J., *et al.* (2006). Buprenorphine for the management of opioid withdrawal. *Cochrane Database of Systematic Reviews.* Issue 2, Art. No.: CD002025. DOI: 10.1002/14651858.CD002025.pub3.
25. Amato, L., Davoli, M., Minozzi, S., *et al.* (2005). Methadone at tapered doses for the management of opioid withdrawal. *Cochrane Database of Systematic Reviews*. Issue 3, Art. No.: CD003409. DOI: 10.1002/14651858.CD003409.pub3.
26. Gowing, L., Farrell, M., Ali, R., *et al.* (2004). Alpha2 adrenergic agonists for the management of opioid withdrawal. *Cochrane Database of Systematic Reviews.* Issue 4, Art. No.: CD002024. DOI: 10.1002/14651858.CD002024.pub2.
27. NICE technology appraisal guidance 115 – January 2007. Naltrexone for the management of opioid dependence.
28. Marlatt, G.A., and Gordon, J.R. (ed.) (1985). *Relapse prevention: maintenance strategies in the treatment of addictive behaviour*. Guilford Press, New York.
29. Prochaska, J.O., and Di Clemente, C.C. (1986). Towards a comprehensive model of change. In *Treating addictive behaviours: process of change* (eds. W.R. Miller and N. Heather). Plenum Press, New York.
30. Miller, W., and Rollnick, S. (1991). *Motivational interviewing*. Guilford Press, New York.
31. Narcotics Anonymous webpage http://www.na.org/accessed April 2007
32. Mayet, S., Farrell, M., Ferri, M., *et al.* (2005), Psychosocial treatment for opiate abuse and dependence. *Cochrane Database of Systematic Reviews*, **25**(1).
33. Amato, L., Minozzi, S., Davoli, M., *et al.* (2004). Psychosocial and pharmacological treatments versus pharmacological treatments for opioid detoxification. *Cochrane Database of Systematic Reviews.* Issue 4, Art. No.: CD004147. DOI: 10.1002/14651858.CD004147.pub2.
34. Department of Health. (1999). Drug misuse and dependence: guidelines on clinical management. HMSO, London.
35. Strang, J., Kelleher, M., Best, D., *et al.* (2006). Emergency naloxone for heroin overdose. *British Medical Journal*, **333**(7569), 614–5.
36. Fischer, G. (2000). Treatment of opioid dependence in pregnant women. *Addiction*, **95**(8), 1141–4.
37. Stover H. (2002). Drug and HIV/AIDS services in European Prisons Bibliotheks-und Unfomationssystems der Universitat Oldenurg BJS
38. Darke, S., Degenhardt, L., Mattick, R., *et al.* (2007). *Mortality amongst illicit drug users: epidemiology, causes and intervention*. Cambridge University Press, Cambridge.
39. McLellan, A.T., Lewis, D.C., O'Brien, C.P., *et al.* (2000). Drug dependence, a chronic medical illness: implications for treatment, insurance, and outcomes evaluation. *Journal of the American Medical Association*, **284**(13), 1689–95.
40. McLellan, A.T., Alterman, A.I., Metzger, D., *et al.* (1994). Similarity of outcome predictors across opiate, cocaine and alcohol treatments: role of treatment services. *Journal of Consulting and Clinical Psychology*, **62**, 1141–58
41. Hubbard, R.L., Craddock, S.G., Anderson, J., *et al.* (2003). Overview of 5-year follow up outcomes in the drug abuse treatment outcome studies (DATOS). *Journal of Substance Abuse Treatment*, **25**(3), 125–34.
42. Gossop, M., Marsden, J., Stewart, D., *et al.* (2003) The National Treatment Outcome Research Study (NTORS): 4-5 year follow-up results. *Addiction,* **98**(3), 291–303.
43. Darke, S., Ross, J., Teesson, M., *et al.* (2007). The Australian Treatment Outcome Study (ATOS): what have we learnt about treatment for heroin dependence? *Drug and Alcohol Review*, **26**(1), 49–54.
44. Hser, Y.I., Hoffman, V., Grella, C.E., *et al.* (2001). A 33-year follow-up of narcotics addicts. *Archives of General Psychiatry*, **58**(5), 503–8.
45. Smyth, B., Hoffman, V., Fan, J., *et al.* (2007). Years of potential life lost among heroin addicts 33 years after treatment. *Preventive Medecine.* **44**(4), 369–74. Epub 2007, Feb 8.

4.2.3.2 Disorders relating to the use of amphetamine and cocaine

Nicholas Seivewright and Robert Fung

Introduction

Amphetamine and cocaine are classed as stimulant drugs, although the distinction between stimulants and depressants can be criticized on the grounds that the same drug may have both actions in turn. This does indeed occur with amphetamine and cocaine, but the initial desired effects are increased energy and activity, and elevation in mood. These appear to be mainly due to enhanced central transmission of dopamine and noradrenaline (norepinephrine), with a similar enhancement of serotonin playing a less certain role.

Pharmaceutical preparations of amphetamine were previously widely used for treatment of depression and obesity, with some misuse of these occurring. In the period since the 1970s of increasing recreational drug use, the powder preparation of 'street'

amphetamine (commonly known as 'speed' or 'whizz') has largely displaced the pharmaceutical forms to become a common drug of misuse in many countries. The powder is typically very impure and constitutes a racemic mixture of d- and l-isomers, with the l-form being relatively inactive. A stronger street preparation is the 'base', sometimes more a moist paste, and both these forms of the drug may be swallowed (either on their own or in a drink), snorted, or injected. (For the methylated forms 'ecstasy' and 'crystal meth', see Chapter 4.2.3.5.)

The coca shrub is indigenous to countries in South America, where the leaf is chewed, and use of the derived cocaine powder has spread extensively to the United States and elsewhere. The powder may be injected, sometimes along with heroin, by polydrug users, or snorted, the image of which is sometimes linked with executive lifestyles. Cocaine has become more dangerous as usage has gradually transferred in many countries to the 'crack' form, which is made from cocaine hydrochloride powder in a simple chemical process, and is more potent in its effects and withdrawal effects. Very rapid increases in blood levels of the drug can be achieved by smoking crack, and this is the usual route, although it is injected by committed intravenous drug users.

Clinical features

The effects and withdrawal effects of amphetamine and cocaine can be considered together, as the main features are equivalent. However, amphetamine has a slower onset of action than cocaine and longer elimination half-life, while crack is the most quickly absorbed of the cocaine preparations. This is reflected not only in the generally more intense effects of cocaine than amphetamine, but in the timescales involved. Thus an amphetamine user may experience desired effects, unwanted mental effects, and withdrawal features over the course of a few days, while a crack user can report the same sequence occurring in a matter of hours or even less. The main effects and withdrawal effects of these two stimulant drugs are shown in Table 4.2.3.2.1.

The list of effects can be seen as merging from the desired to the undesired. These drugs are typically taken in situations where stimulation is the aim, with sleep and eating regarded as hindrances. Mood is elevated, but characteristically this progresses to suspicion, in which true paranoid symptoms may be experienced. This is usually recognized by the individual as indicating that the episode of use should be terminated, but if use persists symptoms may become severe, or a more confused state develop. After stopping the drugs there are typically withdrawal effects of depressed mood, hyperphagia, and hypersomnia; no consensus exists as to whether such features are best viewed as 'rebound' symptoms, a truer withdrawal syndrome, or simply users catching up on sleeping and eating after a period without either.

Table 4.2.3.2.1 Effects and withdrawal effects of amphetamine and cocaine

Effects	Withdrawal effects
Increased energy	Depression
Hyperactivity	Irritability
Euphoria	Agitation
Reduced appetite	Craving
Insomnia	Hyperphagia
Paranoid symptoms	Hypersomnia
Confusion	

Such withdrawal features have been delineated most closely in relation to cocaine. A three-stage process has long been recognized[1]: initially agitation, anorexia, and acute craving; second, excessive tiredness, depression, and hyperphagia; finally, a normalization of most features, but a return of craving when triggered by environmental cues. This description is from before the escalation in crack use, and depression, craving, and agitation especially are often much more severe with this form of the drug. While environmental cues are clearly relevant in precipitating the use of any drug, a powerful surge of craving on encountering situations associated with previous use appears particularly characteristic of cocaine and crack.[2] The three-stage description of withdrawal features suggests that this phenomenon may occur after months or even years of abstinence.

Are amphetamine and cocaine addictive?

It is commonly observed that amphetamine and cocaine are non-addictive, or cause psychological but not physical dependence. Such observations rest on a distinction in which the condition of addiction, or physical dependence, requires visible bodily withdrawal symptoms, but critics claim this is of limited meaning now that there is an understanding of the neurobiological basis of drug withdrawal states. The current classification systems do retain some distinctions between physical and psychological dependence, and the issue is largely one of definition and semantics. The credibility of the label 'non-addictive' is certainly tested by individuals who have injected amphetamine 10 or more times every day for many years, or who spend vast amounts of money using crack in a highly compulsive manner.

Classification

Table 4.2.3.2.2 shows the classification within ICD-10 and DSM-IV-TR of disorders that may relate to the use of cocaine. In both systems the same diagnoses can be applied to amphetamine, in ICD-10 within a category 'other stimulants'.

Importantly, the list of diagnoses in ICD-10 is a standard one to be used across all psychoactive substances, with the second digit of the code number simply changed according to substance, and so does not imply that all those conditions can be caused by amphetamine or cocaine. The DSM-IV-TR listing is somewhat more specific, in that the diagnoses are selected from a wider general list of conditions which can apply to the range of substances. In this way the DSM-IV-TR classification recognizes that cocaine and amphetamine can produce states of dependence and withdrawal, as well as psychosis, affective disorders, and the other conditions included.

Diagnosis

The use of amphetamine or cocaine can be detected by drug screening of a plain urine sample, in laboratory testing or with instant kits. The importance of urine testing as a relatively simple procedure to employ, in any setting, in cases where drug use is suspected must be emphasized, as it is surprisingly often neglected. The two main limitations are possible doubts about authenticity, where mouth swabs for oral mucosal transudate can be a useful

Table 4.2.3.2.2 Classification of disorders relating to cocaine in ICD-10 and DSM-IV-TR

ICD-10		DSM-IV	
F14.0	Acute intoxication	292.89	Cocaine intoxication
		292.81	Cocaine intoxication delirium
F14.1	Harmful use	305.60	Cocaine abuse
F14.2	Dependence syndrome	304.20	Cocaine dependence
F14.3	Withdrawal state	292.0	Cocaine withdrawal
F14.4	Withdrawal state with delirium		
F14.5	Psychotic disorder	292.11	Cocaine-induced psychotic disorder, with delusions
		292.12	Cocaine-induced psychotic disorder, with hallucinations
		292.84	Cocaine-induced mood disorder
		292.89	Cocaine-induced anxiety disorder
		292.89	Cocaine-induced sexual dysfunction
		292.89	Cocaine-induced sleep disorder
F14.6	Amnesic syndrome		
F14.7	Residual and late-onset psychotic disorder		

alternative, and the short time for which drugs remain detectable in urine—as little as 24 h for cocaine. By contrast drugs remain in hair from the head or other parts of the body for the whole period of growth, but this technique which gives much longer-term information is a specialized one.

Obtaining a history and compliance with sampling may be particularly problematic in psychotic states. In such conditions it is also important to recognize that detected drug use may be incidental rather than necessarily causative.

Epidemiology

In most countries, the use of illicit drugs is commonest among young males of lower socio-economic status. Stimulant use overall reflects this, although of the various drugs of misuse, cocaine powder has been exceptional in the extent of usage also by more affluent individuals.

The biggest epidemic of cocaine use outside South America has been in the United States, where it peaked in the mid-1980s.[3] Household surveys at that time estimated that approximately one-tenth of the population had used the drug; the same epidemiological method has charted the subsequent general decline in occasional use, but an increase in more dependent use of crack. Cocaine use in other countries has not generally spread as widely as was predicted from the United States experience. In Europe the population lifetime prevalence of cocaine use has remained in low single-figure percentages,[4] much of it among inner-city polydrug users, although snorting cocaine powder is seemingly increasing among young people, rather displacing the recreational use of 'ecstasy'.

Even in areas where stimulant use is common, such users tend to present relatively rarely to treatment services. Priority is generally given to opioid substitution treatment of heroin addicts, and so service statistics will nearly always underestimate stimulant problems.

Aetiology

Broadly the same familial, social, and psychological factors are relevant in the aetiology of amphetamine and cocaine misuse as in other forms of drug misuse. Approximately half of the drug misusers are deemed in studies to have an underlying personality disorder,[5] usually of the antisocial type, although the figure has sometimes been found to be lower for stimulant misusers than for those dependent on opiates. This may be partly methodological, to do with the difficulty in distinguishing true personality characteristics from behaviours inherent in the activity of highly dependent drug misuse, but is probably also a reflection of the use of stimulant drugs by a generally broader population.

Course and prognosis

Course

A far greater proportion of amphetamine and cocaine misuse than opiate misuse is recreational in nature, with few significant complications if the medical harms are avoided. It is assumed that the vast majority of those who are identified in school and teenage surveys as having used stimulants simply give them up in due course, although little systematic data is available. Complications and involvement with treatment services are more likely where there is dependent usage, and there may be psychiatric contact in episodes of psychosis. A very small proportion of amphetamine injectors progress to high-dose daily usage, while the heavy use of cocaine appears to be less sustainable and is therefore usually periodic in nature.

Other drug use

After being stimulated with amphetamine or cocaine, many individuals will use sedatives such as alcohol, benzodiazepines, or cannabis to 'come down' from their drug. Increasingly heroin is being used for this purpose, sometimes to the point of becoming dependent on the opiate and requiring substitution treatment. The use of cocaine in particular is commonly encountered as a secondary form of drug misuse in methadone patients,[6] with some individuals undoubtedly switching their preferred illicit drug from heroin to cocaine when treatment is established.

Prognosis

The drug misuse literature in general would suggest that stimulant use is more likely to progress and become problematic in individuals with associated personal or social difficulties or psychiatric disorder. Usage by individuals with severe mental illness, which often contains an element of 'self-medication' of distress even though in the long run stimulants will render symptoms worse, can be particularly entrenched.[7]

Complications

Many of the complications of amphetamine and cocaine misuse are complications of drug misuse in general, including those related to injecting. The range includes general physical decline,

weight loss, dental problems, infective complications ranging from abscesses to hepatitis and infection with the human immunodeficiency virus (HIV), reduced foetal growth in pregnancy, mood disturbances, and various social problems. Complications in the following areas are somewhat more specific to stimulant misuse:

- Cardiovascular—hypertension, arrhythmias, myocardial infarction, cerebrovascular accident
- Obstetric—premature labour, placental abruption
- Psychiatric—anxiety, depression, aggressive behaviours, psychosis
- Other—perforation of nasal septum (cocaine snorting)

The cardiovascular problems relate to increased catecholamine secretion, and represent the most serious hazard of cocaine abuse.[8] With obstetric complications, it is difficult to separate the effects of drugs from other risk factors such as poor diet, smoking, or adverse social conditions, but there appears to be a particular link between stimulants and placental abruption.

There are also various psychiatric disorders that are particularly associated with amphetamine and cocaine misuse. Anxiety as a symptom is common in relation to the agitation produced by the drugs, while depression is a classic withdrawal effect. An assessment of the true clinical significance of these features therefore requires withdrawal from drugs, while in acute presentations both can be extremely distressing. Aggressive behaviour may be due to an underlying personality disorder, but it is also characteristic of withdrawal from crack cocaine when severe craving is experienced. Paranoid psychosis, sometimes indistinguishable from an acute schizophrenic episode, is the best-known complication of stimulant misuse. The earliest descriptions were of cases where symptoms quickly subsided after withdrawal of the drugs, but it is now recognized that through mechanisms which represent a kind of sensitization, symptoms which are drug-induced can persist and recur even with avoidance of substance use.[9]

Treatment

Evidence

A very large number of medications have been investigated in cocaine misuse, mainly compounds which through actions on catecholamines or serotonin could be expected to alleviate withdrawal effects. After decades of such work the evidence is very discouraging, with no medications consistently found to reduce stimulant abuse.[10] Inpatient programmes and psychological treatments basically represent modifications of those approaches used across all forms of drug misuse, although cocaine abuse appears particularly amenable to the 'contingency management' approach of providing material incentives for abstinence.[11]

Management

Faced with the limitations in treatment for these forms of drug misuse that have high morbidity and mortality, drug services have had to consider how best to achieve some benefits in terms of very practical management.[12] The factors that appear important in such provision are:

- specific outreach programmes
- harm-reduction approaches
- rapid response where necessary
- targeted use of treatments
- admission in severe cases

To engage stimulant users at all can require outreach aimed at the subcultural groups in whom usage is common. Basic harm-reduction measures must be offered, including drug information, education about health risks, advice to reduce damaging injecting practices, and the provision of clean equipment. Counselling of a supportive or more behavioural kind may be provided by various types of agency.

The periodic nature of stimulant problems means that rapid response can be important, for instance in states of acute crack withdrawal or psychiatric disturbance. Use of tranquillizers and antipsychotic medications may be necessary for some presentations, while fluoxetine appears to be increasingly favoured over other antidepressants, due to a possible anticraving effect and good acceptability. Inpatient admission can be required in cases where no long-term measure is able to make much impact between acute crises. The possibility of any substitute prescribing in stimulant misuse is highly controversial, with some services seeing a role for oral dexamphetamine in heavily dependent amphetamine users experiencing extreme problems from injecting.[13]

Drug-induced psychosis

The two aspects of management of this complication are the treatment of psychotic symptoms and the withdrawal of the drug which is thought to be causative. The latter can be very problematic other than as an inpatient, and is not guaranteed even then. In practice, ongoing low-grade psychotic states in individuals who have not completely stopped drug use are common, and treatment may have to be attempted in such circumstances. The use of antipsychotic medications does not differ significantly from that in psychoses not produced by drugs.

Prevention

The prevention of drug misuse lies largely outside the clinical domain, in the areas of education, enforcement and social improvement. A more biological development in cocaine misuse of vaccination, whereby limited exposure produces antibodies to subsequently block the drug's effects, remains experimental.

Further information

National Institute of Drug Addiction-Research Report for Cocaine Addiction: www.nida.nih.gov/ResearchReports/cocaine.html

Amphetamine Dependence: www.mentalhealth.com/dis/p20-sb02.html

Sadock, B.J. and Sadock, V.A. (2003). Kaplan and Sadock's synopsis of psychiatry; Substance-related disorders, Chap. 12. Lippincott, Williams & Wilkins, Philadelphia.

European Monitoring Centre for Drugs and Drug Addiction. (2006). *Annual report: the state of the drugs problem in Europe*. Office for Official Publications of the European Communities, Luxembourg. Chapter 4, Amphetamines, ecstasy and other psychotropic drugs, and Chapter 5, Cocaine and crack cocaine.

References

1. Gawin, F.H. and Kleber, H.D. (1986). Abstinence symptomatology and psychiatric diagnoses in cocaine abusers: clinical observations. *Archives of General Psychiatry*, **43**, 107–13.

2. van de Laar, M.C., Licht, R., Franken, I.H.A., *et al.* (2004). Event-related potentials indicate motivational relevance of cocaine use in abstinent cocaine addicts. *Psychopharmacology*, **177**, 121–9.
3. Withers, N.W., Pulvirenti, L., Koob, G.F., *et al.* (1995). Cocaine abuse and dependence. *Journal of Clinical Psychopharmacology*, **15**, 63–73.
4. European Monitoring Centre for Drugs and Drug Addiction. (2006). Cocaine and crack cocaine. In *Annual report: the state of the drugs problem in Europe*. Office for Official Publications of the European Communities, Luxembourg.
5. Grant, B.F., Stinson, F.S., Dawson, D.A., *et al.* (2004). Co-occurrence of 12-month alcohol and drug use disorders and personality disorders in the United States. *Archives of General Psychiatry*, **61**, 361–8.
6. Williamson, A., Darke, S., Ross, J., *et al.* (2006). The association between cocaine use and short-term outcomes for the treatment of heroin dependence: findings from the Australian treatment outcome study. *Drug and Alcohol Review*, **25**, 141–8.
7. Seivewright, N., Iqbal, M.Z., and Bourne, H. (2004). Treating patients with comorbidities. In *Drug treatment: what works?* (eds. P. Bean and T. Nemitz), pp. 123–41. Routledge, London.
8. Edgred, M. and Davis, G.K. (2005). Cocaine and the heart. *Postgraduate Medical Journal*, **81**, 568–71.
9. Curran, C., Byrappa, N., and McBride, A. (2004). Stimulant psychosis: systematic review. *The British Journal of Psychiatry*, **185**, 196–204.
10. de Lima, M.S., de Oliveira Soares, B.G., Alves, A., *et al.* (2002). Pharmacological treatment of cocaine dependence: a systematic review. *Addiction*, **97**, 931–49.
11. Higgins, S.T., Heil, S.H., Dantona, R., *et al.* (2007). Effects of varying the monetary value of voucher-based incentives on abstinence achieved during and following treatment among cocaine-dependent outpatients. *Addiction*, **102**, 271–81.
12. Seivewright, N., McMahon, C., and Egleston, P. (2005). Stimulant use still going strong. *Advances in Psychiatric Treatment*, **2**, 262–9.
13. Grabowski, J., Shearer, J., Merrill, J., *et al.* (2004). Agonist-like replacement pharmacotherapy for stimulant abuse dependence. *Addictive Behaviors*, **29**, 1439–64.

4.2.3.3 Disorders relating to use of PCP and hallucinogens

Henry David Abraham

PCP

Introduction

Phencyclidine (PCP, 'angel dust') is an arylcyclohexylamine dissociative anaesthetic. It was first abused in the United States in New York and San Francisco in the 1960s, but abuse declined when a broad range of adverse complications was noted.(1)

Epidemiology

While use of the unadulterated drug occurs, PCP is more frequently mixed with LSD or cannabis. The drug may be ingested or injected, but is more commonly smoked or snorted. Data suggest use in the United States and Europe. In the United States, a stable trend of 3 per cent of high school seniors have tried PCP at least once.(2) It traffics under a long and colourful list of street names. It has been suggested that any illicit smoked drug with an unrecognized street name (dust, mist, THC, embalming fluid, *inter alia*) should be considered PCP until proven otherwise.

Acute physiological effects

The drug has a delayed onset of activity when taken orally. Unlike the major hallucinogens, PCP requires doses in milligrams to be effective, a factor facilitating toxicological identification. When smoked, the onset of its main effects occurs immediately. The drug has particular affinity for the sigma opioid receptor, and non-competitively blocks the *N*-methyl-D-aspartate-type excitatory amino acid receptor. Other effects appear to be mediated indirectly by catecholamine release, cholinergic stimulation, and serotonergic receptors.

DSM-IV lists as criteria for acute PCP intoxication the following:

- agitation
- belligerence
- impaired judgement
- nystagmus
- hyperacusis
- hypertension
- tachycardia
- numbness
- ataxia
- dysarthria
- rigidity
- salivation
- seizures
- coma

It is clear from this daunting inventory that impaired judgement is likely to be present beforehand in any person intentionally choosing to abuse this drug.

Adverse effects

PCP affects not only adults, but fetuses and nursing infants.(3) Neurological consequences in infants include poor attention, hypertonia, and depressed neonatal reflexes.(4) *In vitro* studies show that PCP causes inhibited axon outgrowth, degeneration, and death in human fetal cerebral cortical neurones.(5)

In adults signs of severe PCP toxicity include:

- hyperthermia
- opisthotonus
- cardiac arrhythmia
- stroke

PCP is capable of provoking extreme muscular agitation, rhabdomyolysis and renal failure in 2.5 per cent of users.(6) DSM-IV lists psychiatric effects of PCP including intoxication, delirium, PCP-induced psychotic, mood, and anxiety disorders, and PCP abuse and dependence. A criterion for diagnosis is the emergence of the disorder within a month of drug use.

PCP delirium

Unlike acute intoxication with other hallucinogens, PCP delirium is associated with neurological disturbances. A continuum of effects is noted depending on dose.(7) Psychiatric symptoms occur early in

drug use, with stupor and coma occurring later. Shortly after drug use, patients appear confused and ataxic. Analgesia in fingers and toes may be described. PCP can produce complex hallucinations resembling LSD intoxication. Differentiating the two drugs in emergencies is important, since high-potency neuroleptics, which are useful in PCP toxicity, may exacerbate LSD, while the use of benzodiazepines, helpful in acute LSD toxicity, may disinhibit an assaultive PCP patient. Unlike LSD, PCP is readily identified in routine toxicological screening of blood and urine, but such data may not be readily available. One rapid bedside technique to differentiate the two drugs is the palm sign. The examiner asks the patient to describe the names of all the colours seen in the examiner's outstretch palm. A typical LSD patient reports a vision of multiple colours and images. A PCP patient simply attacks the hand. Dexterity of the examiner is suggested. Unfocused aggression makes PCP delirium a particularly dangerous disorder. The spectrum of violence includes both suicide and homicide.[8] The technique of 'talking down' acutely toxic patients is contraindicated. Environmental stimuli should be minimized, and the patient provided with protective supervision. The use of physical restraints is relatively contraindicated because of the potential for rhabdomyolysis.

Specific treatments involve:

- intravenous naloxone to rule out narcotics overdose
- activated charcoal
- acidification of the patient's urine with vitamin C, ammonium chloride, cranberry juice
- diuresis with frusemide (furosemide, Lasix)
- antihypertensives
- high-potency neuroleptics or barbiturates

PCP has mixed agonist and antagonist effects at cholinergic receptors. Anticholinergic drugs may precipitate a synergistic reaction with PCP, worsening delirium. Thus, low-potency neuroleptics, tricyclic antidepressants, and the anticholinergic antiparkinsonian drugs should be avoided.

PCP-induced psychotic disorder

PCP delirium may evolve into a chronic PCP psychosis that is differentiated from schizophrenia only with difficulty. Alternatively, a PCP delirium may clear, only to be replaced by the insidious onset of a post-PCP psychotic disorder. Certain features of PCP psychosis, namely neurological abnormalities, dose-related severity of symptoms, and regularity of the length of illness, are not noted with other psychedelic drugs, leading to the suggestion that PCP psychosis is a toxic drug effect rather than a functional illness. Four classes of agents are reported to help PCP psychosis:

- benzodiazepines
- neuroleptics
- acetylcholinesterase inhibitors (physostigmine)
- catecholamine depleters (reserpine)

Otherwise, treatment considerations are those for PCP delirium. The long-term prognosis for this disorder appears to be poor, according to data from an 8-year follow-up of 10 patients.[9]

PCP abuse, dependence, and organic mental disorder

Rhesus monkeys will self-administer PCP in a dose-dependent way,[10] suggesting that repeated abuse in humans may be associated with psychophysiological dependence. This in turn is likely to be associated with a decline in social and occupational function characteristic of other forms of addiction. Because of its widespread neuropsychological effects, any intentional, informed use of PCP should be considered maladaptive. For the habituated patient, long-term treatment is indicated. Issues that should be addressed in the process are:

- emotional lability
- cognitive defects
- depression
- possible PCP withdrawal
- nutritional status

Many of the treatments applicable to patients addicted to the opiates, alcohol, and cocaine apply to this population. Several aspects of treating the PCP patient depart from the more conventional addiction treatments. A triad of confusion, decreased cognitive function, and assaultiveness mark an organic mental disorder associated with PCP use. Reduced cognition is a barrier to recovery that must be recognized and addressed in any prospective treatment plan. Neuropsychological assessment is helpful in this regard. Secondly, there is murine evidence that PCP is sequestered in fat and by melanin for at least 3 weeks following a single exposure.[11] Conditions associated with weight loss are likely to release long-held PCP into the blood and brain.

Hallucinogens

Introduction

Agents that alter perception and mood without disorientation typify hallucinogenic drugs. They have been known and used for millennia for purposes ranging from magical to medical. Anthropologists trace back the earliest use of hallucinogens to Paleolithic Europe, although 80 per cent of extant hallucinogenic plants are to be found in the New World. Galen (AD 130–200) wrote that it was customary to give dinner guests hemp seeds to promote the evening's proceedings. The ergot-bearing fungus *Claviceps purpurea* infected rye in tenth-century France and claimed 40 000 lives. Despite such calamities, ergot continued to be used by midwives in medieval Europe. In search of a benign ergot derivative for use in childbirth, Albert Hofmann synthesized lysergic acid diethylamide (LSD-25) in 1938, described in his classic monograph, *LSD: my problem child*.

In 1947, Stoll in Switzerland published the first experimental use of LSD in psychiatry. Intelligence agencies worldwide seized on the misnomer of LSD as an instrument of 'mind control'. Academicians including Sandison and Elkes in England, Cohen and Eisner in the United States, Leuner in Germany, and Grof in Czechoslovakia engaged in human studies. But within a decade the drug the genie was out of the bottle, as the drug moved from the hands of scientists to clinicians, clergy, curious professors, and a widening number of students on both sides of the Atlantic. Military investigators in the United States gave the drug surreptitiously to recruits. By the late 1960s LSD and cannabis led the way to a pandemic of drug abuse among the young.

Drug preparations

Hallucinogenic drugs comprise not so much a single class of compounds, but a multiple classes affecting different neuronal receptors. Hofmann and Schultes describe 11 classes of hallucinogenic compounds which can be isolated from botanicals.[12] Hallucinogens are readily available. Botanicals are easily grown. Indole and phenethylamines can be easily synthesized, especially with the rise of the Internet. Chemically pure hallucinogens are psychoactive in microgram quantities, and are easily concealed, transported and sold, accounting for their enduring role as abusable substances.

LSD is psychoactive in a single droplet of solvent. The drug is easily dissolved in an aqueous solution. Drops of the drug are placed on sugar cubes or blotting paper stamped with coloured cartoon figures to mark the drug's location. Sheets of the paper are then distributed, and the figures ingested. Dosages commonly range from 25 to 100 µg. A hallucinogenic trip can occur after 75 µg. Other hallucinogens, such as dimethyltryptamine, are injected. The serotonin-2A receptor has been shown to bind strongly to many hallucinogenic drugs, and these drugs appear to act as partial agonists.[13]

Common **botanical hallucinogens** include fungi and angiosperms (flower-bearing plants), of which approximately 100 are recognized with hallucinogenic properties. Ibogaine is derived from the root of the *Tabernanthe iboga* plant cultivated in Gabon and eaten as a rite of passage. Ayahuasca is a tea of dimethyltryptamine from the Amazon vine, *Banisteriopsis*, potentiated by beta-carbolines which inhibit monoamine oxidase. Mescaline is a predominant hallucinogen in the cactus *Lophophora williamsii*. Strips of cactus are cut from the plant, dried, and eaten. The Mexican mint, *Salvia divinorum*, contains a diterpene kappa-opioid agonist. *Salvia* is easily bought from the Internet from scores of Websites in the United States and Europe.[14]

Hallucinogenic mushrooms contain psilocybin and psilocin, which are phosphorylated hydroxylated congeners of dimethyltryptamine. Mushrooms are ingested for their effect, or brewed first and the broth consumed. Responses vary widely between individuals and occasions. The American psychologist William James reported ingesting several dozen hallucinogenic mushrooms and only experiencing headache. Shulgin has synthesized and tested 179 phenylethylamines for hallucinogenic properties. Their effects on the human brain are complex and largely unknown.

Epidemiology

Annual surveys in United States college students indicate that LSD use fell from 1995 to 2005 from a lifetime prevalence of 11.5 to 3.7 per cent. This was offset by an increase in the lifetime use of psilocybin mushrooms, from 6.5 to 10.6 per cent. This increase in the use of non-LSD hallucinogens is reflected across secondary grades and young adults between 19 and 28 as well,[1] and follows a period of rising LSD use in Germany, the United Kingdom, and the United States in the 1990s. Factors which may explain this decrease in LSD use include student reports of less availability, greater perceptions of risk, and the substitution of psilocybin and MDMA for LSD. A cross-sectional study of 904 women from 14 to 26 found that LSD users were more often Caucasian, victims of physical abuse, and suffering depression.[15]

Acute effects

The characteristic LSD trip comprises:

- autonomic arousal
- marked mydriasis
- sensory disturbances
- emotional lability

Progressive modulations of visual imagery appear to be generated both from external objects and distortions of eidetic sources. Ordinarily benign objects may take on new emotional meanings. Geometric imagery may rise and fall before one's eyes. A prevalent feeling one experiences is a sense of helplessness to control one's streaming images and emotions, hence the hippie advice of 'going with the flow'. The loss of cognitive, perceptual, and affective control for some users leads to panic, which in turn results in the so-called 'bad trip'. As these effects decline, they may be replaced with a sense of oceanic well-being or residual paranoia.

Adverse effects

Adverse reactions to hallucinogens include panic reactions associated with a bad trip, hallucinogen persisting perception disorder, and prolonged psychoses.

(a) Panic reaction

Panic may arise during the acute drug experience. It is characterized by a crescendo of rising anxiety accompanied by autonomic arousal in the context of streaming emotions and imagery. Mydriasis is greater than that seen in non-drug-induced panic. The use of an oral benzodiazepine such as diazepam 20 mg is utterly effective in stopping the panic within minutes.

(b) Hallucinogen persisting perception disorder (HPPD)

It became apparent within the first few years of experimentation with LSD that this class of drugs was capable of inducing visual disturbances days to weeks following drug exposure. Subsequent research found that these disturbances, dubbed 'flashbacks' because of their evanescent visual appearance, appeared to be an intermittent form of post-drug visual disorder that in its extreme form was experienced continually. Thus, HPPD patients are capable of describing a range of visual disturbances that fluctuate in intensity, but are observable from moment to moment (Abraham, 1983). Such imagery includes:

- geometric hallucinations
- false perceptions of movement
- afterimagery
- the perception of trails behind moving images
- pinpoint dots in the air (aeropsia)

Symptoms drawn by HPPD patients have been published on the Internet.[16] Patients also describe derealization and depersonalization. Symptoms are intensified by stimulation from:

- emergence into a dark environment
- marijuana
- amphetamines
- cocaine

- anxiety
- the stress of intercurrent illnesses

While recovery may occur over months and years following last drug use, approximately half of the patients so afflicted appear to develop a permanent alteration of the visual apparatus. Studies of psychophysics in HPPD patients reveal quantified prolongations in afterimagery.[17] Neurophysiological studies confirm cerebral disinhibition involving those regions of the cortex processing visual information.[18]

Management: Because the disorder is exacerbated by psychological and physiological conditions of arousal, benzodiazepines have been helpful for management of visual symptoms.[19] The results of these efforts are palliative at best, and complicated by the prospect of treating a drug abuser with an abusable substance. Recent case reports of treatment with sertraline, naltrexone, and clonidine are encouraging.

In addition to pharmacotherapy, HPPD patients often require supportive psychotherapy to deal with the issues of learning to cope with what may be a permanent alteration in perception. Therapy is also indicated to educate the patient, and prevent the development of common comorbid disorders in HPPD. These are:

- major depression
- panic disorder
- alcohol dependence from self-medication

(c) Psychosis

Evidence supporting the hypothesis that the use of potent hallucinogens can trigger prolonged psychotic episodes is found in multiple longitudinal, cross-sectional, and case studies (Abraham *et al.* 1996). Psychiatric patienthood appears to be a risk factor for the development of psychosis following LSD. The clinical picture of post-LSD psychosis resembles schizoaffective disorder more than it does schizophrenia, with the commonly added feature of chronic visual disturbances. Clinically such patients resemble those with good-prognosis schizophrenia, since they possess more affect than those with poor prognosis, have less thought disorder, are more socially related, and appear to have fewer signs of negative schizophrenia. Mystical preoccupations reminiscent of acute drug experiences can predominate. Visual hallucinations often are of the variety that are seen in HPPD, although in contradistinction to such patients, psychotic patients may describe delusions and auditory hallucinations as well.[20]

Management: Atypical pharmacotherapies appear to have an important role in treatment, and in selected cases are preferable to dopamine-blocking neuroleptics. Reports in the literature describe cases responding to electroconvulsive therapy, lithium, anticonvulsants, and the serotonin precursor, l-5-hydroxytryptophan. Long-term supportive psychotherapy is almost always indicated to help the patient and his family make painful adjustments to the patient's chronic disappointments in relationships and employment, frequently made all the more poignant by the illness' propensity to preserve the patient's insight as it progresses. This last factor may partially explain the apparently high risk for suicide.

Human experimentation with hallucinogens

The discovery of the effects of LSD in 1943 led to a flurry of experimental activity with the drug in humans, at its worst with dubious methodology and indifferent to the protection of human subjects. But the ability of this unique class of drugs to alter perception, cognition, and affect has prompted a new wave of research with selected hallucinogens with regulatory oversight. Studies have examined the safety of peyote in religious use[21]; the psychological effects of psilocybin, dimethyltryptamine, and ketamine[22]; their use as a tool in modelling the pathophysiology of psychosis,[23] and possible therapeutic uses.

Hallucinogens have been used as experimental psychoses. Vollenweider *et al.* have found increased metabolic activity in the frontal cortex of subjects on the dissociative anaesthetic ketamine,[24] and in subjects during an experimental psychosis from psilocybin, increased serotonin-2 agonist activity.[25] Finally, the study of treatment with hallucinogens for psychiatric disorders has cautiously reemerged. In a randomized trial of ketamine for depression, Zarate *et al.* found that the drug had benefits for a week following a single dose.[26]

Further information

Abraham, H.D. (1983). Visual phenomenology of the LSD flashback. *Archives of General Psychiatry*, **40**, 884–9.

Abraham, H.D., Aldridge, A.M., and Gogia, P. (1996). The psychopharmacology of hallucinogens. *Neuropsychopharmacology*, **14**, 285–98.

Hofmann, A. (1980). *LSD: my problem child*. McGraw-Hill, New York.

McCarron, M.M., Schulze, B.W., Thompson, G.A., *et al.* (1981). Acute PCP intoxication: incidence of clinical findings in 1000 cases. *Annals of Emergency Medicine*, **10**, 237–42.

References

1. McCarron, M.M., Schulze, B.W., Thompson, G.A., *et al.* (1981). Acute PCP intoxication: incidence of clinical findings in 1000 cases. *Annals of Emergency Medicine*, **10**, 237–42.
2. Johnston, L.D., O'Malley, P.M., Bachman, J.G., *et al.* (2006). *Monitoring the future national survey results on drug use, 1975–2005: secondary school students*, Vol. I (NIH Publication No. 06-5883). National Institute on Drug Abuse, Bethesda, MD.
3. Tabor, B.L., Smith-Wallace, T., and Yonekura, M.L. (1990). Perinatal outcome associated with PCP versus cocaine use. *American Journal of Drug and Alcohol Abuse*, **16**, 337–48.
4. Golden, N.L., Kuhnert, B.R., Sokol, R.J., *et al.* (1987). Neonatal manifestations of maternal PCP exposure. *Journal of Perinatal Medicine*, **15**, 185–91.
5. Mattson, M.P., Rychlik, B., and Cheng, B. (1992). Degenerative and axon outgrowth-altering effects of PCP in human fetal cerebral cortical cells. *Neuropharmacology*, **31**, 279–91.
6. Akmal, M., Valdin, J.R., McCarron, M.M., *et al.* (1981). Rhabdomyolysis with and without acute renal failure in patients with PCP intoxication. *American Journal of Nephrology*, **1**, 91–6.
7. Milhorn, H.T.J. (1991). Diagnosis and management of PCP intoxication. *American Family Physician*, **43**, 1290–302.
8. Poklis, A., Graham, M., Maginn, D., *et al.* (1990). PCP and violent deaths in St. Louis, Missouri: a survey of medical examiners' cases from 1977 through 1986. *American Journal of Drug and Alcohol Abuse*, **16**, 265–74.
9. Wright, H.H., Cole, E.A., Batey, S.R., *et al.* (1988). PCP-induced psychosis: eight-year follow-up of ten cases. *Southern Medical Journal*, **81**, 565–7.
10. Campbell, U.C., Thompson, S.S., and Carroll, M.E. (1998). Acquisition of oral PCP (PCP) self-administration in rhesus monkeys: effects of dose and an alternative non-drug reinforcer. *Psychopharmacology (Berlin)*, **137**, 132–8.
11. Misra, A.L., Pontani, R.B., and Bartolomeo, J. (1979). Persistence of PCP (PCP) and metabolites in brain and adipose tissue and implications for long-lasting behavioural effects. *Research Communications in Chemistry, Pathology, and Pharmacology*, **24**, 431–45.

12. Schultes, R. and Hofmann, A. (1980). *The botany and chemistry of hallucinogens*. Thomas, Springfield, IL.
13. Glennon, R.A., Titeler, M., and McKenney, J.D. (1984). Evidence for 5-HT2 involvement in the mechanism of action of hallucinogenic agents. *Life Sciences*, **35**, 2505–11.
14. Prisinzano, T.E. (2005). Psychopharmacology of the hallucinogenic sage *Salvia divinorum*. *Life Sciences*, **78**, 527–31.
15. Rickert, VI., Siqueira, L.M., Dale, T., *et al.* (2003). Prevalence and risk factors for LSD use among young women. *Journal of Pediatric and Adolescent Gynecology*, **16**, 67–75.
16. Abraham, H.D. (2005). at http://www.drabraham.com/html/hppd.htm
17. Abraham, H.D. and Wolf, E. (1988). Visual function in past users of LSD: psychophysical findings. *Journal of Abnormal Psychology*, **97**, 443–7.
18. Abraham, H.D. and Duffy, F.H. (1996). Stable quantitative EEG difference in post-LSD visual disorder by split-half analysis: evidence for disinhibition. *Psychiatry Research and Neuroimaging*, **67**, 173–87.
19. Lerner, A.G., Gelkopf, M., Skladman, I., *et al.* (2003). Clonazepam treatment of lysergic acid diethylamide-induced hallucinogen persisting perception disorder with anxiety features. *International Clinical Psychopharmacology*, **18**, 101–5.
20. Bowers, M.B. Jr (1977). Psychoses precipitated by psychotomimetic drugs: a follow up study. *Archives of General Psychiatry*, **34**, 832–5.
21. Halpern, J.H., Sherwood, A.R., Hudson, J.I., *et al* (2005). Psychological and cognitive effects of long-term peyote use among Native Americans. *Biological Psychiatry*, **58**, 624–31.
22. Gouzoulis-Mayfrank, E., Heekeren, K., Neukirch, A., *et al.* (2005). Psychological effects of (S)-ketamine and *N*,*N*-dimethyltryptamine (DMT): a double-blind, cross-over study in healthy volunteers. *Pharmacopsychiatry*, **38**, 301–11.
23. Javitt, D.C. and Zukin, S.R. (1991). Recent advances in the PCP model of schizophrenia. *The American Journal of Psychiatry*, **148**, 1301–8.
24. Vollenweider, F.X., Leenders, K.L., Scharfetter, C., *et al.* (1997). Metabolic hyperfrontality and psychopathology in the ketamine model of psychosis using positron emission tomography (PET) and [^{18}F]fluorodeoxyglucose (FDG). *European Neuropsychopharmacology*, **7**, 9–24.
25. Vollenweider, F.X., Vollenweider-Scherpenhuyzen, M.F., Babler, A., *et al.* (1998). Psilocybin induces schizophrenia-like psychosis in humans via a serotonin-2 agonist action. *Neuroreport*, **9**, 3897–902.
26. Zarate, C.A. Jr, Singh, J.B., Carlson, P.J., *et al.* (2006). A randomized trial of an *N*-methyl-D-aspartate antagonist in treatment-resistant major depression. *Archives of General Psychiatry*, **63**, 856–64.

4.2.3.4 Misuse of benzodiazepines

Sarah Welch and Michael Farrell

Epidemiology and patterns of use

The rise in benzodiazepine prescribing in the United Kingdom in the 1960s and 1970s was a development that followed the previous period of prescribing of barbiturates and other sedatives. Concerns about the obvious toxicity of barbiturates, and previously other sedatives, in overdose, together with knowledge of their dependence-inducing characteristics, led to their replacement with the safer benzodiazepines as the commonly prescribed anxiolytic and hypnotic drugs. Case reports of patients who escalated their dose of benzodiazepines above the recommended dose, and who experienced convulsions and confusional states on stopping them, began to appear in the 1970s.[1,2] In the mid-1970s and onward, regulatory bodies in the United Kingdom and United States began to recognize the abuse potential of benzodiazepines even in therapeutic doses. Dependence on benzodiazepines was well described in the early literature on the development of these drugs but surprisingly clinical dependence was not reported in the medical literature until the early 1980s.[3,4] Dependence on therapeutic doses of prescribed benzodiazepines is covered in Chapter 6.2.2. In this chapter we are concerned with abuse of and dependence on high doses of benzodiazepines.

The upsurge in the drug epidemic in the 1980s was associated with an increase in misuse of hypnosedatives, and in the late 1980s there was a series of reports on the intravenous use of benzodiazepines, in particular temazepam.[5,6] Because benzodiazepines are the most commonly prescribed anxiolytics and hypnotics it is not surprising to find that they are also reported to be commonly misused. However, patterns of misuse vary, from episodic use of non-prescribed medication with up to 15 per cent of young people reporting some experience with benzodiazepines, to continuous high-dose use. Since the mid-1980s there has been a substantial drop in the prescription of benzodiazepines as anxiolytic agents but use as hypnotics has remained relatively steady with the concentration of long-term use being in the elderly population. Changes in prescribing practices are likely to influence diversion of benzodiazepines to the illicit market.

Reports indicate that supra-therapeutic dose misuse and dependence is strongly associated with polydrug and alcohol abuse and dependence.[7,8] This pattern of benzodiazepine misuse and dependence is probably much less common than iatrogenic benzodiazepine dependence. However, it presents a substantial problem to many clinicians in primary care and specialist settings. In particular, high-dose misuse is likely to be associated with 'doctor shopping' and efforts to extract additional medication on top of that already prescribed. The high doses used present a particular risk because they are often used in combination with other substances such as alcohol, opiates, and stimulants. High-dose use may be intermittent in nature (in a 'binge' pattern), and not associated with dependence (in which case the initiation of a prescription may change a pattern of intermittent binge use to daily dependent use in a manner that entrenches polydrug use; see the section below on guidelines for management). Drug misusers use benzodiazepines in a non-dependent fashion for a variety of reasons. For example, benzodiazepines may be used to enhance the effect of other drugs (such as boosting the euphoria with heroin), or to alleviate unwanted effects from other drugs (to 'cushion' the 'come-down' from cocaine), or to help alleviate withdrawal symptoms when drugs such as heroin are unobtainable, or in attempts at self-detoxification from other drugs. Misuse of benzodiazepines may also arise from injudicious patterns of prescribing for the treatment of alcohol dependence, or from attempts at self-treatment for alcohol dependence. Some drug users will also develop a dependent pattern of use of benzodiazepines in their own right. Benzodiazepine dependence may be a factor contributing to poor outcome for patients who are attempting opiate detoxification.

High-dose use is associated with substantial tolerance to the sedating effects of the medication but some of the other effects may not be equivalently protected by tolerance. Thus, some individuals may consume extraordinarily high doses, and not appear sedated, but experience profound amnesia for their actions. Such effects of

amnesia may also be associated with the reported high rates of risk-taking behaviour, and amnesia may be more pronounced in injecting benzodiazepine users, although there are no good data to confirm this.

The potential of different benzodiazepines for misuse and dependence

In view of the frequency of prescribing of benzodiazepine drugs, it is an important question as to whether some benzodiazepines are more likely to be misused, or to lead to dependence, than others. The similarities between different benzodiazepines are much greater than the differences. Patients may maintain that they need a specific named benzodiazepine, but there is marked cross-tolerance, and patients changed to an equivalent dose of a different benzodiazepine under double-blind conditions show almost complete cross-dependence (i.e. no difference in withdrawal symptoms from those whose medication has been unchanged).[9] However, this cross-dependence was shown for patients who were already benzodiazepine dependent. It is possible that the properties of certain benzodiazepines lead to a stronger motivation for people to desire their effects, to escalate the dose, and to persist with their use.

Factors that have been considered to influence the liability to misuse and development of dependence include the relative potency of the drug, and its elimination half-life.[10] Triazolam, a very short-acting benzodiazepine prescribed for insomnia, was withdrawn from the British market following concerns about the severity of rebound anxiety experienced even after a single dose. Triazolam is a very potent benzodiazepine that binds very readily to benzodiazepine receptors, and experience with its use suggested that it had a more euphoriant effect than other benzodiazepines, resulting in greater potential for misuse and for the development of dependence. Other benzodiazepines that have high potency and have caused concern include flunitrazepam and lorazepam. Flunitrazepam is relatively rarely prescribed in the United Kingdom, but frequently reported as one of the most common benzodiazepines misused in many European countries and in Australia. It is not marketed in the United States. Concerns about its availability on the illicit market continue.[11,12] It has attracted media attention as a drug used to facilitate 'date rape'; it is unclear why this particular benzodiazepine should have this image, although it is a potent drug. Lorazepam, also a potent benzodiazepine, is much more widely prescribed in the United Kingdom, and alprazolam is used in the United States. Some studies suggest that lorazepam and alprazolam may be associated with an earlier and more difficult withdrawal process than diazepam.[9,13] In one European study, triazolam and lorazepam were found to feature more highly among individuals dependent on high doses of benzodiazepine drugs than among those dependent on low or 'therapeutic' doses.[14] In summary, it appears that potency is a contributory factor in the abuse and dependence-inducing potential of benzodiazepine drugs. However, this picture is somewhat complicated by the fact that drugs such as lorazepam tend to have been marketed at higher equivalent doses than some other benzodiazepines. The elimination half-life influences the nature of withdrawal; if it is short, withdrawal symptoms appear more rapidly and may appear more severe, although withdrawal of more insidious onset in longer-acting benzodiazepines may be just as problematic.

As well as properties of the drugs themselves, the abuse potential of different benzodiazepines is also associated with broader prescribing patterns which affect the potential for diversion to illicit market. Diazepam and temazepam have been the most widely prescribed benzodiazepines in the United Kingdom, and are therefore the most likely to be obtained by drug users and problem drinkers. Drugs such as clonazepam which tend to be prescribed much less widely, and generally for epilepsy rather than for anxiety or insomnia, seem infrequently to raise concerns about misuse[10] but are frequently requested as the treatment of choice for those who are both epileptic and drug dependent. In other parts of Europe where flunitrazepam is more commonly prescribed, there are reports of its high levels of misuse among the illicit drug-using population.

The potential for misuse by injecting

Over time, there have been reports of misuse of various benzodiazepines by injection.[7] A number of factors have influenced this practice. These include the availability of the drug, its short-acting nature, and also the formulation of the drug. In the 1980s temazepam was marketed as a liquid-filled capsule, which enabled easy extraction of the contents to put into a syringe for injecting. The later gel formulation was also injected, by heating to liquidize the gel, resulting in very damaging injecting complications. Some medications that come in easily soluble form can also be converted into a form for injecting, such as liquid diazepam.

The injection of benzodiazepine drugs is associated with substantially more harmful drug misuse in a number of respects,[15] with increased rates of reported sharing of injecting equipment,[9,16] increased risks of overdose, and poorer general health.[8]

Evidence-base for management of benzodiazepine misuse

The literature on management of 'ordinary dose' benzodiazepine dependence, relating mainly to patients prescribed benzodiazepines for treatment of psychiatric disorders, is far more extensive and systematic than that concerning illicit drug users. At the time of writing, there are no meta-analyses or indeed well-conducted randomized controlled trials specifically addressing this problem. In practice, the applicability of the 'ordinary dose' dependence literature is affected not only by clinical differences, but also by concerns about abuse and diversion of prescribed supplies by illicit drug users. So far, no clear guidelines for management have been produced, and the management principles covered in the section below are based on expert clinical consensus statements: these involve some extrapolation from the 'ordinary dose' benzodiazepine dependence literature, and some from established evidence-based principles for management of misusers of other drugs such as opiates.[17] For example, unlike the established evidence that supervised substitute prescribing of methadone reduces injecting in intravenous heroin users, there is no clear evidence that substitute prescribing of benzodiazepines reduces injecting behaviour in injecting misusers of these drugs. Nevertheless, prescribing of oral benzodiazepines with daily supervised consumption is sometimes instituted for this group of patients, especially if they are already established in supervised treatment for concurrent intravenous heroin use (see section on concurrent opioid dependence below).

Guidelines for management of misuse of non-prescribed benzodiazepines

These are a complex group of patients to manage and in general such patients should be referred for specialist assessment.

Assessment

Assessment should attempt to confirm evidence for benzodiazepine dependence. Assessment over a number of visits should involve obtaining urine specimens to determine objectively the regularity (or intermittent nature) of benzodiazepine consumption. Part of the initial assessment process should identify underlying psychiatric disorders that may have been the trigger for a doctor initiating a prescription or for the patient obtaining drugs for the purpose of self-medication. High levels of psychiatric morbidity have been found among samples of patients with severe benzodiazepine dependence.[18] The commonest conditions are probably anxiety disorders for which anxiolytics have been prescribed injudiciously; however, in some instances major depressive disorders may be treated or self-medicated with benzodiazepines. Thorough assessment should explore for evidence of major depression and if identified, consideration should be given to the use of antidepressant medication combined with cognitive behavioural therapy.

Treatment

In the presence of polydrug abuse or dependence, caution needs to be exercised about initiating a prescription for benzodiazepines (see section on opioid drug users). Where dependence is established, dose titration should aim to ameliorate withdrawal symptoms rather than to match the large doses of medication that the patient reports consuming. Long-acting benzodiazepines such as diazepam are preferred (Chapter 6.2.2), and doses should rarely exceed 40 to 60 mg daily. Where large doses are prescribed these should ideally be dispensed on a daily pick-up basis from the community pharmacist. The prescribing doctor should avoid succumbing to pressure from the patient to increase the dose without a thorough dose assessment and evidence of withdrawal symptoms. Prescriptions issued on a routine 'repeat' basis should be avoided, and the patient should be informed that lost medication will not be replaced. Requests for replacement medication or additional medication should encourage the doctor to review the overall care plan and to consider reducing and stopping the benzodiazepine prescription. Virtually all benzodiazepine prescriptions should be time limited and part of a detoxification plan with gradual reduction over a clearly stated and negotiated time period. Alternatively, inpatient detoxification, perhaps with longer admissions, may be the best course of action where reduction regimens fail in the community setting. Reduction regimens with benzodiazepines are very variable and many clinicians opt for a very gradual withdrawal with small dose reduction at wide intervals. There is no evidence to indicate that such a gradual approach is any more effective than 20 to 25 per cent reductions over a shorter, more clearly defined time (such as 6 weeks). Carbamazepine can also be effective in ameliorating withdrawal symptoms and in seizure prevention, and may be used as an adjunct in withdrawal management in high-dose benzodiazepine dependence,[19] though drug interactions need to be considered for patients who are also prescribed opioid substitute drugs for treatment of heroin dependence (see section on concurrent benzodiazepine and opioid dependence).

Special considerations for managing concurrent benzodiazepine and alcohol dependence

There are high reported rates of alcohol dependence among the homeless, and there are substantial reports of benzodiazepine abuse and dependence in this population also. Cross-tolerance between benzodiazepines and alcohol permits individuals who are alcohol dependent to tolerate high doses of benzodiazepines. Patients may be prescribed benzodiazepines to manage alcohol withdrawal, but injudicious prescribing that is not targeted towards the management of alcohol withdrawal symptoms may result in iatrogenic benzodiazepine dependence.

The extensive research on pharmacological interventions for management of alcohol withdrawal in alcohol-dependent patients has been examined in a meta-analysis by Mayo-Smith, who subsequently produced a systematic review with treatment guidelines.[20] This review supports the use of benzodiazepines as the treatment of choice for managing withdrawal symptoms for patients whose symptoms are of sufficient severity to warrant medication, and is consistent with a more recent Cochrane review.[21] However, carbamazepine is established as an acceptable alternative to benzodiazepine drugs in treatment of alcohol withdrawal, and is effective both in reducing withdrawal symptoms and in seizure prevention.[17] Agents such as clomethiazole and the barbiturate phenobarbital, are less well-supported by controlled trials than benzodiazepines, and carry higher risks of adverse effects.

The potential for misuse of benzodiazepines must be considered in alcohol-dependent patients. However, this is not an adequate reason for avoiding the use of benzodiazepines in the management of withdrawal in view of their superiority in effectiveness, and possibly in potential for harmful misuse, over other treatments. Benzodiazepines with a slower onset of action such as chlordiazepoxide, appear to have less potential for misuse. The prolonged use of benzodiazepines in withdrawal is rarely necessary or helpful, and evidence for benefits of 'substitute prescribing' of benzodiazepines for alcohol users in the longer term is lacking. Use of benzodiazepines to manage phobic and anxiety disorders associated with alcohol dependence should be avoided. For alcohol-dependent patients with a history of benzodiazepine misuse, especially where the patient wishes to avoid further benzodiazepine use, management of withdrawal using carbamazepine alone should be explored.

Special considerations for managing concurrent benzodiazepine and opioid drug dependence

Currently there is no evidence base for long maintenance prescribing of benzodiazepines for those who are high-dose injecting polydrug abusers. However, this intervention has not been subject to any structured evaluation and merits further study in the face of the difficulties of managing such patients. There are two particular areas of concern in managing opioid drug users who also misuse benzodiazepines. The first is the potential for toxicity and overdose when combining illicit or prescribed opioid drugs with benzodiazepines, particularly because of the sedative and respiratory

depressant effect of both. Two studies comparing the effects of benzodiazepine use in conjunction with either methadone or buprenorphine treatment have reported higher subjectively reported opioid toxicity symptoms[22] and greater peak effects on objectively assessed performance measures[23] for patients on methadone than for patients on buprenorphine. Where substitute prescribing for heroin or other opioid drug dependence is to be undertaken, the patient needs clear information about the effects of combining opioid and benzodiazepine drugs. Close supervision of prescribed medication (such as daily supervised consumption in the clinic or pharmacy) should be maintained in the early stages of treatment. As many opioid drug users use benzodiazepines somewhat erratically, for example as a substitute when heroin is not available, then stabilization on either methadone or buprenorphine as a regular dose may allow them to stop benzodiazepine use without any other intervention. Initiation of any prescribing intervention addressing benzodiazepine dependence is usually best delayed until it is clear that dependent use has continued despite well-established treatment for opioid dependence. The second area of concern is the potential poor prognosis of opioid withdrawal programmes for patients with concurrent benzodiazepine dependence. Opioid withdrawal symptoms may be more pronounced in patients with both problems, and treatment protocols may need to be adjusted to address this.[24]

Conclusion

There is a valuable role for benzodiazepines, and a need for vigilance and care in their use, as well as active recognition and management of those who are dependent. However, there is a need for greater awareness of the risks of polydrug dependence with misuse of high doses of benzodiazepines in conjunction with both alcohol dependence and opiate dependence. Caution needs to be used in assessing patients, and benzodiazepine prescribing should be restricted to those where there is clear evidence of dependence.

The risk of synergistic effects with other drugs and consequent overdose should be explained to all patients who are identified as being involved in such behaviour. Community detoxification or inpatient detoxification is the best option based on the evidence of available research and evaluation of current interventions.

Further information

Seivewright, N. (2000). *The case for other substitute drugs. Community treatment of drug misuse: more than methadone*, pp. 49–81. Cambridge University Press, Cambridge, UK.

Lingford-Hughes, A.R., Welch, S., and Nutt, D. (2004). Evidence-based guidelines for the treatment of substance misuse, addiction and comorbidity: recommendations from the British Association for Psychopharmacology. *Journal of Psychopharamacology*, **18**, 293–335.

References

1. Woody, G.E., O'Brien, C.P., and Greenstein, R. (1975). Misuse and abuse of diazepam: an increasingly common medical problem. *International Journal of Addiction*, **10**, 843–8.
2. Bliding, A. (1978). The abuse potential of benzodiazepines with special reference to oxazepam. *Acta Psychiatrica Scandinavia Supplementum*, **274**, 111–16.
3. Petursson, H. and Lader, M.H. (1981). Benzodiazepine dependence. *British Journal of Addiction*, **76**, 133–45.
4. Hallstrom, C. and Lader, M. (1981). Benzodiazepine withdrawal phenomena. *International Pharmacopsychiatry*, **16**, 235–44.
5. Farrell, M. and Strang, J. (1988). Misuse of temazepam. *British Medical Journal*, **297**, 1402.
6. Ruben, S.M. and Morrison, C.L. (1992). Temazepam misuse in a group of injecting drug users. *British Journal of Addiction*, **87**, 1387–92.
7. Strang, J., Seivewright, N., and Farrell, M. (1992). Intravenous and other novel abuses of benzodiazepines: the opening of Pandora's box? *British Journal of Addiction*, **87**, 1373–5.
8. Ross, J., Darke, S., and Hall, W. (1997). Transitions between routes of benzodiazepine administration among heroin users in Sydney. *Addiction*, **92**, 697–705.
9. Murphy, S.M. and Tyrer, P. (1991). A double-blind comparison of the effects of gradual withdrawal of lorazepam, diazepam and bromazepam in benzodiazepine dependence. *The British Journal of Psychiatry*, **158**, 511–16.
10. Tyrer, P. (1993). Pharmacological differences in the dependence potential of benzodiazepines. In *Benzodiazepine dependence* (ed. C. Hallstrom), pp. 77–90. Oxford University Press, Oxford.
11. Simmons, M.M. and Cupp, M.J. (1998). Use and abuse of flunitrazepam. *Annals of Pharmacotherapy*, **32**, 117–19.
12. Woods, J.H. and Winger, G. (1997). Abuse liability of flunitrazepam. *Journal of Clinical Psychopharmacology*, **17**(Suppl. 2), 1S–57S.
13. Rickels, K., Case, W.G., Schweizer, E.E., *et al.* (1986). Low dose dependence in chronic benzodiazepine users; a preliminary report on 119 patients. *Psychopharmacology Bulletin*, **22**, 415–17.
14. Martinez-Cano, H., Vela-Bueno, A., De Iceta, M., *et al.* (1996). Benzodiazepine types in high versus therapeutic dose dependence. *Addiction*, **91**, 1179–86.
15. Darke, S., Hall, W., Ross, M., *et al.* (1992). Benzodiazepine use and HIV risk-taking behaviour among injecting drug users. *Drug and Alcohol Dependence*, **31**, 31–6.
16. Klee, H., Faugier, J., Hayes, C., *et al.* (1990). AIDS-related risk behaviour, polydrug use and temazepam. *British Journal of Addiction*, **85**, 1125–32.
17. Lingford-Hughes, A.R., Welch, S., and Nutt, D. (2004). Evidence-based guidelines for the treatment of substance misuse, addiction and comorbidity: recommendations from the British Association for Psychopharmacology. *Journal of Psychopharamacology*, **18**, 293–335.
18. Busto, U.E., Romach, M.K., and Sellers, E.M. (1996). Multiple drug use and psychiatric comorbidity in patients admitted to the hospital with severe benzodiazepine dependence. *Journal of Clinical Psychopharmacology*, **16**, 517.
19. Denis, C., Fatseas, M., Lavie, E., *et al.* (2006). Pharmacological interventions for benzodiazepine monodependence management in outpatient settings. *Cochrane Database of Systematic Reviews*, (3), CD005194.
20. Mayo-Smith, M.F. (1997). Pharmacological management of alcohol withdrawal: a meta-analysis and evidence-based practice guideline. *The Journal of the American Medical Association*, **278**, 144–50.
21. Ntais, C., Pakos, E., Kysaz, P., *et al.* (2005). Benzodiazepines for alcohol withdrawal. *Cochrane Database of Systematic Reviews*, (3), CD005063.
22. Nielsen, S., Dietze, P., Lee, N., *et al.* (2007). Concurrent buprenorphine and benzodiazepines use and self-reported toxicity in opioid substitution treatment. *Addiction*, **102**, 616–22.
23. Lintzeris, N., Mitchell, T.B., Bond, A., *et al.* (2006). Interactions on mixing diazepam with methadone or buprenorphine in maintenance patients. *Journal of Clinical Psychopharmacology*, **26**, 274–83.
24. de Wet, C., Reed, L., Glasper, A., *et al.* (2004). Benzodiazepine co-dependence exacerbates the opiate withdrawal syndrome. *Drug and Alcohol Dependence*, **76**, 31–5.

4.2.3.5 Disorders relating to the use of ecstasy and other 'party drugs'

Adam R. Winstock and Fabrizio Schifano

Introduction

Participation in the dance music/rave scene has been associated with an ever-growing range of primarily stimulant and hallucinogenic drugs since its inception in the Balearic Islands in the mid-1980s. The last two decades have seen the globalization of the dance music scene and the gradual adoption of these and other drugs by mainstream drug using populations. Although 'Ecstasy' (**MDMA**, 3,4-methylenedioxymethamphetamine) was the archetypical dance drug inducing both stimulant and empathogenic effects, dance drug use is polydrug use with cocaine, amphetamine, nitrates, ketamine, and GHB all being common.[1] The use of this diverse group of drugs is now no longer confined to either young adults or the dance floor with use common for example, at house parties or other more intimate social gatherings.[2] Although not typically identified as dance drugs, alcohol and cannabis are of course highly prevalent among this group.

Ecstasy (3,4-methylenedioxymethamphetamine)

Background

Incorrectly termed a designer drug (a drug whose chemical structure is modified to avoid being included within a list of drugs/chemical structures prohibited by legislation), MDMA was first synthesized in 1912 by Merck Pharmaceuticals. It was never marketed and remained largely ignored until the 1950s when the United States Army explored its military potential. It was not until the late 1960s and early 1970s however that drug users on the west coast of America began to popularize its recreational use along with MDA (methylenedioxymethamphetamine).

Although MDMA may claim its place as the mother of dance drugs and possessor of the best branding in terms of name, MDMA is only one of a large number of synthetic amphetamine type drugs possessing varying degrees of stimulant, hallucinogenic, and empathogenic effects that are used within the dance scene. Characterized by its stimulant and prosocial effects the sought after experiences of disinhibition, euphoria, energy, and empathy are ideally suited to the 'dance scene' where energetic and prolonged dancing is commonplace. Indeed it may be that dancing offsets the psychomotor agitation that stimulants can induce or that MDMA-like drugs may enhance enjoyment and ability to dance to dance music. In the United Kingdom MDMA is classified as a Class B drug.

Preparations, purity, and routes of use

MDMA is most commonly taken orally though it may be snorted or injected. MDMA is most commonly sold as branded tablets ('pills'), with different tablets being identified by an imprinted logo (for example of a cartoon character, car manufacturer, or animal), but may also be found as capsules or powder. The average dose of an ecstasy tablet containing MDMA is about 70 mg (range 50–150 mg). The cost has reduced markedly over the last 20 years from £20/tablet in 1985 to as little at 50 pence/tablet when bought in bulk (typical price in the United Kingdom for a single tablet would be 2–5).[3]

Because of the illicit nature of MDMA production, variation in precursor availability, and the large number of possible synthetic pathways for its production, tablets sold as ecstasy may contain a wide range of substances other than MDMA. In the United Kingdom and Europe especially in the 1990s, ecstasy tablets were often found to contain substances (usually psychoactive) other than MDMA. These included analogues of MDMA, such as methylenedioxyamphetamine (**MDEA**), *N*-methyl-1-(1,3-benzodioxol-5-yl)-2-butanamine (**MBDB**), and methylenedioxyethylamphetamine (**MDEA**), or combinations of stimulants such as ephedrine or amphetamine and hallucinogens such as LSD or ketamine.[4] While these other psychoactive substances may be marketed as distinct substances with their own branding (e.g. 4-Bromo-2, 5-dimethoxyphenethylamine sold as 2CB and 4 MTA as 'flatliners') more commonly they and the MDMA analogues MDEA, MBDB, and MDA, which broadly share the effects of MDMA, are sold under the generic term 'ecstasy'. More recent evidence would suggest that the proportion of tablets sold as ecstasy that contains MDMA in the United Kingdom is very high, with purity levels of 80–90 per cent not being uncommon.[5]

Pill testing

Some users utilize various pill testing methods (such as the Marquis test which gives colorimetric result by mixing the substance with a reagent) and websites (www.EcstasyData.org) which provide the contents of different pills following more elaborate analytical methods. Although it may be the case that these approaches may have some role in getting users to consider risks and promote the uptake of harm reduction practices, it is only on very few occasions that such methods have identified potentially far riskier psychoactive contents such as paramethoxyamphetamine (PMA).[6] In such cases, early warning through such websites may be potentially helpful. However, even with more sophisticated analytic processes (e.g. GMS, HPLC), knowing the content of your tablet never guarantees the user a positive experience and does not guarantee that they will not experience severe adverse effects.[4] In addition, most ecstasy-related deaths have involved tablets containing MDMA. Most deaths have not been related to dose and would have been unpredictable from knowing the content of the tablet.[3]

Prevalence and patterns of ecstasy use

British population studies show that 54 per cent of 20- to 22-year-olds have been offered ecstasy at some time and 15 per cent have tried it at least once.[7] The prevalence of use appears higher in those associated with the dance music drug scene with over 90 per cent reporting ever use. Similar findings have been reported from Europe,[8] Australia,[9] and the United States.

The typical pattern of use in the United Kingdom and Europe is one to four tablets a night, though many users will often consume larger number of tablets during a binge session especially when holidaying in summer dance resort destinations. Regular users will use between once or twice a week to once every fortnight, though there has been increasing recognition of a minority of users who take very large numbers of tablets (20 or 30) over a single session,

or with the availability of very cheap 'pills' extended periods of low-level daily use. Ecstasy, especially within the context of dance clubs is rarely taken in isolation and polydrug use is the norm, with different adjunctive substances taken at different times over the course of a night.[1] For example, alcohol is taken with ecstasy at the beginning of the night to get a stronger/better high.[10] Cocaine, amphetamines, and/or additional ecstasy tablets are taken to maintain arousal and a state of alertness (the MDMA enactogenic effects fade away in 2–4 h). Finally depressants such as cannabis, alcohol, benzodiazepines, and more rarely opiates, may be taken in the last part of the night to calm down before going home, since the untoward after-effects of ecstasy (namely irritability, insomnia, and restlessness) may persist well beyond its 'pleasurable' effects. With a chronic high dosage, ecstasy users develop tolerance and experience a decrease in the desired effects over time, which could lead to exploration of use of other stimulants and hallucinogens.[1]

Physical effects and complications

Physiologically sympathomimetic properties similar to amphetamine predominate including tachycardia, anorexia, increased respiratory rate, blood pressure, increased motor activity, tremor, mydriasis, increased temperature, and sweating. Jaw tightening (bruxism), xerostomia, teeth grinding with molar erosion, may also be seen (see Table 4.2.3.5.1). Sleep architecture modification[11] and sexual activity alteration[12,13] have also been described.

After MDMA intake a number of untoward effects may commonly occur including nausea, vomiting, diarrhoea, tachycardia, and palpitations. Pathologies less commonly seen include arrhythmias, hypertension as well as potentially life-threatening, metabolic acidosis, cerebral haemorrhages, convulsions, coma, rhabdomyolysis, thrombocytopenia, disseminated intravascular coagulation, SIADH, acute kidney failure, acute liver failure, dehydration, and malignant hyperthermia.[14–18]

Hyperthermia although enhanced by exertional activity and poorly ventilated environments may be somewhat independent from the setting in which the drug is taken, with MDMA having thermal dysregulation effects in its own right. Dehydration is common and thirsty clubbers naturally tend to replace body fluids lost during sweating sensibly with fruit juices, other isotonic fluids or water, or less sensibly but quite commonly with alcohol. Very rarely excessive intake of hypotonic fluids, coupled with an increase in vasopressin levels, has led to the occurrence of lethal hyponatraemia.[19] Deaths as a result of SIADH are very rare but can be fatal in association with excessive hypotonic fluid consumption with MDMA potentially impairing judgement or stimulating repetitive compulsive behaviours. In normal subjects who take ecstasy and do not develop SIADH there does appears to be an increase in both ADH and oxytocin levels,[20] the latter perhaps responsible for the drug's prosocial effects. MDMA is a potentially damaging cardiac stimulant;[21] with reports suggesting long-term MDMA use may possibly lead to a fenfluramine-like valvular heart disease condition. All users of ecstasy develop a (mild, in most cases) serotonin syndrome after acute drug intake.[13,22] Unfortunately although being reported by almost half ecstasy users, Verheyden and Henry[23] found that concerns about physical health are not perceived as important as concerns about mental health.[23]

Table 4.2.3.5.1 Psychological and physical effects of MDMA

Physical	Psychological
Increase in physical and emotional energy	Relaxation/euphoria
Dilated pupils, dry mouth	Feelings of well-being
Tachycardia, hypertension, increased respiratory rate	Enhanced closeness and sociability
Increased sweating, dehydration	Heightened perceptual awareness
Increased motor activity, tremor	Disinhibition
Blurred/double vision	Increased response to touch/empathy
Anorexia, nausea, weight loss	Anxiety/panic/paranoia
Teeth grinding, jaw clenching	Agitation and restlessness

MDMA psychological effects and problems (see Table 4.2.3.5.1, adapted from Liester *et al.*[24])

Being structurally related to both amphetamine and mescaline, 'empathogens' or 'entactogens'[25] like MDMA possess both stimulant and hallucinogenic properties which allow them to be discriminated from other related substances. After MDMA ingestion, enhanced mood, increased energy, openness, heightened sensory perception, and mild perception alterations are reported[15,17] (Table 4.2.3.5.1).

MDMA is described as evoking 'an easily controlled altered state of consciousness with emotional and sensual overtones',[26] with the substance's appeal resting in its 'dramatic and consistent ability to induce in the user a profound feeling of attachment and connection'. With this in mind it is perhaps not surprising that the Los Angeles dealer who coined the street name 'ecstasy' for MDMA would have preferred the name 'empathy' but he did not feel that his typical customer would know what it meant. It was also these qualities that led to the enthusiasm of a small number of physicians and therapists in the United States to explore its use within a clinical psychotherapeutic setting and more recently led to its approval as a research agent in the treatment of PTSD.[27]

Acute psychological problems associated with MDMA use

There have been reports of acute episodes of anxiety, panic, paranoia, and rarely brief psychotic episodes following consumption of MDMA by some users. Many users of MDMA report 'midweek blues', with some individuals reporting clinically borderline levels of depression in the days following MDMA[28] which could reflect depletion of serotonin following the acute elevation that follows ingestion of MDMA. This could be seen as a parallel to the 'crash' reported after abstinence of cocaine use or as a hangover effect from all night dancing, excessive alcohol, and minimal sleep. Although depression, anxiety, and mood fluctuations attributed to ecstasy are reported to be strongly related to the number of occasions of MDMA use,[29,30] Morgan *et al.*[31] found that higher depression scores among current heavy ecstasy users, in comparison to drug-naïve and polydrug controls, were no longer significant after treating cannabis use as a covariate.

Other consequences of use

(a) Neurotoxicity and evidence for 5-HT disruption in humans (see Table 4.2.3.5.2)

Although, in humans, the relationship between MDMA intake, putative 5-HT neurotoxicity, and persistent functional consequences is somewhat controversial,[32] the average single dose size consumed by humans approaches levels found to be neurotoxic in animals.[10] Core ambient temperature and hydration status have been implicated as key factors in the development of neurotoxicity.[13]

Although Colado *et al.*[33] suggested that MDMAs ability to produce neurodegeneration of dopamine nerve endings is open to debate, MDMA is generally considered to be a selective 5-HT neurotoxin. After administration of MDMA, animals have reduced levels of 5-HT, 5-hydroxyindole acetic acid, and tryptophan hydroxylase. Abnormal 5-HT regrowth has been reported after MDMA-induced damage with a decrease in 5-HT terminal density, suggestive of a 'chemical axotomy'. Pathological investigations suggest that 5-HT nerve terminals arising from the dorsal raphe nucleus are specifically involved. Duration and magnitude of these neurotoxic effects are dose dependent and are followed by differential rates of recovery, with 5-HT damage persisting for up to a year in the rat, and dopaminergic damage for up to 3 years in the rhesus monkey. These changes appear to be species specific with primates being more sensitive to the neurotoxic damage than rodents.

Markers for 5-HT damage may be sought either by direct assessment of metabolite levels or indirectly by assessing those functions thought to be dependent on an intact 5-HT system (see Table 4.2.3.5.2). MDMA users may show reduced brain 5-HTT (serotonin transporter) levels; there might be an association between degree of MDMA exposure and degree of reduction in 5-HTT ligand binding.[34]

Further evidence for disruption of the 5-HT system comes from blunted neuroendocrine responses (cortisol and prolactin) to d-fenfluramine in former MDMA users;[13] from cognitive disturbances in former MDMA users and from PET studies showing a decrease in a structural component of 5-HT neurones.

(b) Neuropsychological impairment and psychiatric presentation of ecstasy users

MDMA users as a group demonstrate a range of cognitive deficits in comparison to alcohol users, non-drug controls, and MDMA-naïve polydrug controls, including cannabis users.[31,35] Although the issue is somewhat controversial the most consistent neuropsychological finding in former MDMA users is a deficit in verbal memory under both immediate- and delayed-recall conditions.[36] Deficits in other areas of cognitive function, such as verbal fluency, executive function, impulse control, reaction time, and processing speed have been reported as well.[31] Evidence for attentional deficits varies depending on the task employed.[37] MDMA use may be associated with longer visual scanning times, reaction times, or planning times. Interpretation of these data is somewhat complicated by the fact that MDMA users typically use other drugs which may exert independent or interactive effects on cognitive performance.[37]

MDMA intake may put users at significant risk for developing psychiatric problems,[23] although some have suggested that this may occur only in vulnerable individuals.[39] Studies suggest that ecstasy users may report both childhood emotional/physical abuse[40] and history of familial depression, anxiety, and panic attacks[41] more frequently than ecstasy-naïve controls.

Thus although fraught with confounders, especially other drug use and premorbid functioning, there does appear to be an association between MDMA use and increased rates of anxiety,[24] panic, major depressive disorder,[38] prolonged depersonalization, psychosis, flashbacks, and even craving for chocolate.

(c) Depression and its management in ecstasy users

In the case of a patient presenting with psychological problems who has a history of MDMA use, the crucial assessment issues are the identification of any premorbid disorders, where in the cycle of use/post use they are, and the persistence of any symptoms beyond a 2–4 week period following cessation of use. As with amphetamine use, in the days after taking MDMA there is a period characterized by symptoms attributable to monoamine depletion and subsequently repletion. A period of acute 5-HT depletion due to vesicular monoamine depletion (Tuesday blues), is likely to be the most potent cause for the relative reduction in monoamine neurotransmitters. Repeated use of MDMA over several days will be associated with markedly diminished effects. Recovery is delayed further by inhibition of the rate-limiting enzyme (tyrosine hydroxylase in the case of MDMA) and the relative absence especially in chronic users, of a good source of monoamine precursors following stimulant-induced anorexia and malnutrition. It is likely that, as with other stimulant drugs, including cocaine, a period of extended but less intense withdrawal symptoms (mood, sleep) may be seen with persistent abstinence which may take weeks or months to recede and are associated with the more gradual reversal of neuroadaptive changes in dopaminergic receptor sensitivity/expression.[42]

History taking should specifically endeavour to identify any pre-existing/persistent depressive/other disorders and to ascertain the functionality of the use of MDMA and the consequences on underlying mood and functioning in days following use. Antidepressant treatments should usually not be commenced until 2–4 weeks after cessation of MDMA use in order to allow for reassessment and confirmation of any disorder. In addition, resolution of symptoms with cessation of use may also act as powerful reinforcer of continued abstinence if an individual's mood recovers. Reassessment and treatment where necessary, is important since being depressed is associated with an enhanced initial response to stimulant drugs and higher relapse risk.

Table 4.2.3.5.2 Clinical signs of intoxication with amphetamine

Physical	Psychological	Behavioural
Elevated P, BP, temperature	Euphoria/energized	Motor hyperactivity
Increased respiratory rate	Anxious/irritable	Restless/twitching
Sweating/dehydrated	Rapid thoughts	Talkative, pressured speech
Dilated pupils	Paranoia	Aggressive
Tremor/shakiness	Perceptual disturbance	Stereotyped movements

In prescribing an antidepressant to a client with a previous history of MDMA use, confining prescribing to only abstinent users is recommended since their effectiveness during a period of current use would be expected to be poor both as a result of poor compliance and monoamine depletion. In addition there are at least theoretical causes for concern over potentially fatal interactions between MDMA and selective serotonin reuptake inhibitors (SSRIs) that have very rarely been reported, possibly because some SSRIs (i.e. citalopram) can inhibit the CYP2D6 enzyme.[22] The precise effects of combining SSRIs and MDMA appears to be related to whether use of the SSRI was before or after the MDMA and whether SSRI dosing is acute or chronic. For example, SSRIs given acutely after MDMA (taken by users to intensify the ecstasy effects) may theoretically increase the risk of precipitating a serotoninergic syndrome. It is probable that SSRIs and other classes of antidepressant can be used effectively in this group if a diagnosis of responsive affective/anxiety disorder is confirmed and abstinence is maintained. CBT may be useful in this group both to address their underlying drug as well as to address any coexisting anxiety/depressive disorders.

Dependence with the development of heavy regular use patterns is possible, though there are unlikely to be any specific signs or symptoms that differentiate diagnosis of management significantly from other forms of stimulant dependence.

Methamphetamine-'crystalline methamphetamine hydrochloride' ice, crystal, shabu, yaba, meth, tina

Background

Methamphetamine is one of number of synthetic amphetamine type stimulants that includes dex-amphetamine. Whilst its use has been problematic in SE Asian countries such as Thailand, Japan, and Korea for many years it is only in the last decade that it has become a significant problem in eastern Europe, America, Australia, and elsewhere. In the United Kingdom methamphetamine was reclassified as Class A drug in 2007.

Preparation, purity, and routes of use

Unlike illicit amphetamine sulphate powder (speed), methamphetamine is often of very high purity. Crystalline methamphetamine hydrochloride (known as ice-because it can resemble shards of glass) can be up to 80% pure. Base amphetamine (sometimes known as paste), is an oily, waxy intermediate product on the way the manufacture of the crystalline hydrochloride salt of methamphetamine and has a lower purity of about 40–50 per cent. Methamphetamine is a versatile drug and can be smoked, snorted, injected, and taken orally.

Mechanism of action and metabolism

Methamphetamine closely resembles amphetamine sulphate (commonly referred to as speed) in structure and mechanism of action but is considerably more potent in its sympathomimetic effects and has a longer duration of action (half-life about 12 h). Methamphetamine has both direct sympathomimetic effects secondary to disruption of vesicular storage of monoamines and inhibition of their breakdown by MAOIs and indirect actions through inhibition of central presynaptic reuptake of catecholamines.

Prevalence and patterns of use

In the United Kingdom reports of its use are becoming more common, particularly in association with the gay and dance music scene but compared to the use of cocaine in all its forms, the prevalence of methamphetamine use at present is still low.[43]

Physical effects and complications

Sympathetic arousal induced by methamphetamine produces rapid and sometimes irregular heartbeat, sweating, pupillary dilation, hypertension, dry mouth, tremor and blurred vision, and increased body heat. Occasionally, serious medical complications arise including coronary artery syndrome, seizures, and cerebral bleeds (see Table 4.2.3.5.2).

Psychological effects and complications

Acute sought after-effects are similar to those of amphetamine and include euphoria, enhanced stamina, confidence, disinhibition, reduced appetite, improved coordination, and heightened alertness and awareness.

Other consequences of use

(a) Dependence and withdrawal

Dependence may occur and is more common among heavy male users and in those who smoke or inject the drug. Although dependent users may use every day to avoid withdrawal, more typically users tend to consume a large amount of the drug (often several grams) over several days going out without sleep (a binge) before ceasing use through physical exhaustion or an exhaustion of funds. 'Crashing' refers to the period following a binge, which is characterized by fatigue, hypersomnia, hyperphagia, and low mood due to acute monoamine depletion. The crash and subsequent comedown period may last 2–7 days. If abstinence persists a longer term withdrawal period may be seen, characterized by craving, low mood, anergia, irritability, sleep, and appetite disturbance. Similar neurobiological mechanisms involving alterations in the function and activity of the monoamine neurotransmitters are responsible for the overlap between the symptoms of depression and those of stimulant withdrawal.[42,44] Typically the withdrawal gradually diminishes over 2–4 weeks though dysphoric symptoms may persist for up to 10 weeks.

(b) Withdrawal, depression, and management

The frequency of depressive symptoms is highest during the withdrawal period. As with alcohol, there are a far fewer number of people who present with depression symptoms outside of stimulant use or withdrawal. Management of withdrawal is largely supportive and with a safe, well-supported home environment. In patient admission is rarely required other than in those with severe mental or physical illness. The patient should be placed in quiet surroundings for several days and allowed to sleep and eat as much as is needed. Because a significant component of the withdrawal syndrome is probably related to neurotransmitter depletion, recovery may be delayed because of anorexia associated with amphetamine use. It may be useful in some to provide nutritional supplements or a well-balanced diet rich in monoamine precursors: phenylalanine, tyrosine, l-tryptophan for example, pumpkin seeds, chocolate, marmite, bananas. Benzodiazepines may be prescribed

on a short-term basis for agitation. Some patients may become markedly despondent during withdrawal and a suicide assessment may be necessary.

Since antidepressants have no specific anti-craving effects, and the efficacy of antidepressants in reducing depression is confined to those stimulant users who are depressed, it is useful to wait until after they have stopped using for 2–4 weeks and reassess them for depressive symptoms. The advantages of waiting are improved diagnostic accuracy, avoidance of potentially unnecessary medication, and probably an improvement in compliance and efficacy. However the persistence of depressive symptoms beyond 2–4 weeks after stopping amphetamine use may suggest that there is an underlying depressive illness and this should be treated[42] since left unmanaged its presence represents a high risk for relapse. Psychosocial treatments for stimulant abuse and dependence have been found to be effective in reducing levels of use,[45] but to date, no reliably effective pharmacological treatments have been identified and there are currently no widely accepted evidence-based pharmacotherapy regimes for the treatment of psychostimulant withdrawal.[46]

(c) Stimulant-induced psychosis

The use of high doses of methamphetamine may lead to the induction of a temporary psychotic state that may be clinically indistinguishable from paranoid schizophrenia. First recognized in 1938 in association with Benzedrine nasal inhalers, it was not until Connell's classic 1958 study that the syndrome was well described. Acute transient psychotic episodes (typically characterized by suspiciousness, unusual thought content, or hallucinations) occur in about 10–15 per cent of users. Psychotic episodes are more common in dependent users, men, injectors, and smokers, polydrug users, those with past history and following a binge in association with prolonged insomnia.[47]

Characterized by persecutory delusions and hallucinations which are typically auditory but may be visual or tactile (which can be associated with secondary delusion of parasitic infestation) amphetamine-induced psychosis typically remit within few days or at most a few weeks. Little has changed in the way of management since Connell's time who recommends 'removal of the drug and appropriate sedation'. Often, patients present with high levels of hostility and violence secondary to persecutory delusions or hallucinations, and safe containment and management of the disturbed individual can require enormous levels of both physical and chemical restraint. Benzodiazepines (often required in very high doses) should be the first-line medication with antipsychotics used only where additional tranquilization is required. A diagnosis of a possible underlying or persistent psychotic disorder must be deferred until a reassessment can be made in a drug-free state. These often florid psychoses usually remit within a few days and the user returns to normal functioning, although some retain a vulnerability to such episodes.[48] Only a minority (1–15 per cent) persist beyond 1 month and many of these patients will have underlying psychiatric disorders.[47]

The prognosis is variable with those who have experienced stimulant-induced psychotic episodes being more vulnerable to future episodes (possibly through behavioural sensitization) on re-exposure to the drug often at lower levels. Recent positron emission tomography (PET) imaging studies in chronic methamphetamine users have demonstrated a reduction in dopamine transporter concentration and this reduction was significantly associated with the duration of methamphetamine use and closely related to the severity of persistent psychiatric symptoms. Moreover, the severity of psychiatric symptoms was significantly correlated with the duration of methamphetamine use. Cessation is still potentially important since there does however appear to be some recovery of dopamine transporter function with abstinence.

Gamma hydroxy butyrate, GHB, GBH, fantasy, G, liquid ecstasy, and GBL

Background

GHB is an endogenous short-chain fatty acid found in the CNS and elsewhere in the body. A putative neurotransmitter, its precise role is yet to be identified although specific binding sites have been identified in hippocampus (linked to DA neurones). Trace amounts may also be found in certain fruits such a guava. In the United Kingdom GHB is classified as Class C drug.

Like ketamine, GHB was originally developed as an anaesthetic though the high incidence of tonic clinic seizures dampened enthusiasm among surgeons for its routine use. Subsequently it found clinical utility as a sedative, a treatment for narcolepsy, as a detoxification agent (it is effective in the management of alcohol withdrawal) and as a putative muscle growth enhancer for body builders (through its effect on increasing slow wave sleep). Since the 1990s however it has become best known for its place among the smorgasbord of drugs commonly used by those involved in the dance/rave scene. It has a reputation as a cheap stimulant drug of short duration with marked aphrodisiac properties but a narrow therapeutic threshold carrying a significant risk of overdose especially in combination with alcohol.[49]

Preparation (pro drugs), purity, and routes of use

Until the late 1990s GHB and its precursors including the psychoactive pro compounds GBL (gamma butyl-lactone) and 1,4 butanediol were widely available over the Internet. Because of its relative ease of manufacture attempts at home production were common, resulting in preparations of widely varying concentrations. The resultant formulations were sometimes caustic, resulting in gastrointestinal discomfort, vomiting, aspiration, and coma.

GHB is sold most commonly as a free acid (a colourless and odourless liquid in its pure form, with a slightly salty, acidic taste) or as a sodium salt (usually a white powder). Although varying widely in concentration typically doses are sold in plastic vials holding 5–10 ml.

Mechanism of action and metabolism

GHB readily cross the BBB and acutely leads to a transient decrease followed by increase in dopamine levels (accompanied by increase in endogenous opioid release). Increases in other neurotransmitters such as GABA, Ach, and 5-HT are also seen. At higher doses it exhibits some partial GABA-b activity (epileptogenic). GHB is usually taken orally often mixed in fruit juice or alcoholic beverage. It has a very rapid onset of action with noticeable effects occurring within 15 min of administration it has a relatively short duration of action ($t_{1/2}$ 27 min) with effects peaking at 30–60 min and being over within 2–4 h, being eliminated though its breakdown to CO_2 and H_2O.

Physical effects and complications

GHB exhibits a very narrow therapeutic index and as a result of wide interpersonal variation in tolerance and significantly enhanced toxicity (depressant effects) when combined with alcohol, overdose with GHB has been reported more widely than for any other dance drug,[50,51] overdose should be suspected in someone who presents with nystagmus, ataxia, nausea, vomiting, sedation, weakness, bradycardia, hypotension, and the rapid onset of unconsciousness (quite similar to severe alcohol intoxication but without alcohol on the breath). Management should include placing the person in the recovery position, airway management and pulse oximetry. GCS scores may be very low (<7). If oxygen saturation drops or they are so unconscious that they can tolerate a Guedels airways then ventilation should be considered. Overdoses are short lived and most awake somewhat aroused and disorientation after a few hours. Other clinical presentations include agitation, anxiety, coma, amnesia, and collapse. These patients when in coma may require ventilation and typically suddenly emerge from their coma with high levels of agitation, arousal, and violence.

Although there have been press reports of GHB being commonly used as 'date rape drug', such cases are very rare and it is still the case that the most common drug used for such purposes is alcohol alone.

Psychological effects and complications

Consumption of GHB results in a dose-related euphoria and stimulation which gives way to sedation at higher doses. In combination with stimulant drugs there may be an increase of precipitating a brief psychotic reaction, whilst with alcohol the risk of fatal overdose is the primary concern. There has been a single case report of Wernicke Korsakoff syndrome.

Other consequences of use

(a) Dependence and withdrawal

More recently there have been number of reports describing GHB dependence and withdrawal. The later may present as a rapid onset, prolonged alcohol withdrawal picture. Although associated with lower levels of autonomic arousal and seizure risk, patients may exhibit marked confusion, delirium, and hallucinations. Management may require doses of diazepam markedly in excess of those typically used to manage alcohol withdrawal with a waxing and waning clinical progression that may last 2 weeks.[52]

Ketamine (K, Special K, Super K, Vitamin K, Green, Mean Green, Jet)

Background

Ketamine (2-(2-chlorophenyl)-2-(methylamino)-cyclohexanone) and PCP (angel dust) are very similar drugs, the main difference being ketamine's shorter half and less problematic 'emergence phenomena'. Ketamine has a range of useful clinical applications. It is used across several areas of medicine including paediatric analgesia and anaesthesia, emergency anaesthesia, obstetrics, and battle-zones[53] and benefits from having a wide margin of safety in overdose. Ketamine is an NMDA antagonist and is almost unique as an anaesthetic in its ability to produce a 'dissociative' state, which results in higher brain structures in the brain centres being prevented from perceiving auditory, visual or painful stimuli leading to 'a lack of responsive awareness'. Overall the effect has been described as somato-aesthetic sensory blockade with amnesia and analgesia. In recent years its non-medical use as a psychedelic has become more common. In the United Kingdom ketamine is classified as Class C drug. For a recent review see Wolff and Winstock.[54]

Mechanism of action and metabolism

Ketamine has multiple actions at numerous receptor sites particularly affecting glutaminergic and monoaminergic neurotransmission. The most significant pharmacological action of ketamine is the non-competitive antagonist binding at the cation channel of the NMDA receptor and consequent interference with excitatory amino acid transmitters—glutamate and aspartate. As a research probe in the study of schizophrenia, ketamine has given increasing prominence to the role of glutamate in the aetiology of psychotic illness. Ketamine also enhances monoaminergic transmission resulting in marked sympathomimetic effects as well as modulating activity at opioid receptors, thought to be responsible for its analgesic and dysphoric effects.

Ketamine undergoes marked first-pass metabolism and is fairly ineffective when taken orally (bioavailability may be <20 per cent). However, oral consumption does result in a two-fold higher concentration of its primary metabolite norketamine compared to, for example, intramuscular dosing. Norketamine is pharmacologically active with anaesthetic potency approaching one-third that of the parent compound. Hence although the onset of effects following ingestion may be somewhat slower than by parental routes the duration of effects would almost certainly be longer. The majority of the parent drug will be eliminated from the body within 24 h, although prolonged effects due to the presence of active metabolites may occur.

Preparations, purity, and routes of use

Ketamine may be sold illicitly in a number of preparations; as crystalline powder for intranasal use (dose-100–400 mg), in liquid, tablet, powder, or capsular form (dose-350–500 mg) for ingestion. When obtained from diverted licit sources, the formulation of the drug is a solution prepared for intravenous use. This solution may be injected or swallowed, but more typically the solution is dried and taken intranasally.[54] However, this process of desiccation may reduce the purity of the crystalline residue, and is an obvious point at which, via contamination, purity may be altered. Ketamine may be adequately absorbed via the intranasal, intravenous, subcutaneous, intramuscular, and intrathecal routes, with snorting and injecting being the most common recreational routes of use. Reports of clinical use via the rectal and transdermal routes have also been described.

Prevalence and patterns of use

Far less common in the general population than MDMA, ketamine none the less has become increasingly popular among those associated with the dance scene with a prevalence of about 20 per cent being reported among clubbers. Whilst ketamine maintains a good safety record within clinical settings, the increase in its unregulated use outside such controlled environments is a cause for concern

Table 4.2.3.5.3 Ketamine, psychological and physical effects

Psychological	Physical
Rapid onset, short duration of action (1 h), wide safety margin	
Dissociative anaesthesia 'somatosensory blockade' analgesia	Dilated pupils
Perceptual distortion/hallucinations/near death	Tachycardia
Out of body experience	Hypertension
Though disorder/synaethesia	Ataxia
Emergence phenomena	Paralysis
Cognitive impairment	Sweating
Amnesia	Hypersalivation
Derealization/depersonalization	Little effect on cough reflex

with its effects being highly sensitive to age, dose, route, set, and setting (see Table 4.2.3.5.3). Outside clinical settings ketamine is most commonly snorted or injected, with typically administered doses being small fractions of a gram (an eighth). Because of its short half-life (17 min) the psychedelic effects experienced are generally short-lived with effect duration of about 1–2 h. The short duration of effect and rapid onset of action when taken by intranasal or intravenous routes often leads recreational users to administer repeated doses over the course of an evening (session) in order to maintain a desired psychoactive effect.

Physical effects and complications

Because of its fast urinary excretion (within 2 h) the ability to identify ketamine in urine screens is almost impossible and thus a level of clinical suspicion is required especially if a history of its use is not forthcoming.[55] Detection by clinical examination relies on identifying mydriasis, moderate tachycardia, elevated BP, slurred speech, blunted affect, ataxia, delirium, nystagmus (less commonly than with PCP). Tachycardia is the most common finding on physical examination. Its short half-life of 17 min (see Table 4.2.3.5.3).

Admissions to hospital are most commonly for complaints related to sympathetic over activity with chest pain, palpitations and taccycardia, nausea, vomiting, difficulty breathing, ataxia, temporary paralysis/inability to speak, blurred vision, no awareness of pain as well as derealization/depersonalization, and amnesia. Other risks associated with its use include accidents, trauma, and risky sexual behaviours. Rarely more severe complications are reported including severe agitation and rhabdomyolysis. Although relatively safe in overdose, in combination with ethanol or other CNS depressants the use of ketamine can result in death.

Clinical findings and detection

(a) Psychological effects and complications

At low doses marked elevation in mood predominate. At higher doses intense psychedelic effects commence with sensory and perceptual distortion, euphoria, and out of body and floating experiences (see Table 4.2.3.5.3). The Harvard academic, Timothy Leary, described it as 'the ultimate psychedelic journey'. Users describe entering the 'K hole' where they experience—visits to god, aliens, their birth, past lives and the 'experiences of evolution'. Some users report taking issues of set and setting into careful consideration prior to using ketamine such preparation cannot be performed if the drug is consumed unwittingly when it has been marketed under the guise of another drug such as ecstasy. Being an amnestic it may become difficult to remember the total doses consumed.

Ketamine can also produce a psychotic picture that can briefly mimic schizophrenia. Both positive and negative symptoms of schizophrenia can be transiently seen in normal users and its use can exacerbate symptoms in those with pre-existing psychotic disorders. Other adverse effects can include frightening hallucinations/out of body experiences, thought disorder, confusion, and dissociation. Such episodes tend to be short-lived, resolving in a few hours or at the most a few days. In many respects these are similar to those adverse effects seen LSD, though with ketamine they come on after a shorter period following use and recede more quickly.

Management is by supportive monitoring (cardiovascular) in a quiet low stimulation room with symptomatic treatment with benzodiazepines if needed. Unusually, the effects of benzodiazepines are inconsistent varying between compounds and dose. For instance, whilst lorazepam may reduce emotional distress, it appears to have little impact upon the psychosis or perceptual changes observed. Midazolam, on the other hand, is able to negate ketamine's effect on thought process and perceptual disorder but has little impact upon mood problems. Interestingly haloperidol also has little effect upon the psychosis associated with ketamine, suggesting a role for receptors other than D2. Chlorpromazine should be avoided (anticholinergic effects).

Other consequences of use

(a) Ketamine dependence and long-term cognitive impairment

Animal studies demonstrate the ability for intravenous ketamine to produce dependence in rat models, with disruption of operant behaviour on withdrawal. Ketamine demonstrates reinforcing efficacy in animal self-administration models and is found to be a discriminative stimuli in operant tasks. Its effects are thus readily distinguishable from other drugs and may have abuse liability. However, ketamine is somewhat unusual in its pharmacodynamics, almost acting as a partial antagonist with regard to brain reward enhancements, being stimulatory at low doses, and inhibiting brain reward centres at higher doses.

Clinically ketamine dependence has been described, with compulsive use as primary symptom. Heavy habitual use has been described by Jansen (1990), and cases of dependence have also been reported among anaesthetic staff. Although tolerance develops there are only a few case reports of a withdrawal syndrome occurring.

Longer-term follow-up studies suggest that any impairment of semantic memory may be reversible upon cessation of use but persistent deficits may be seen in episodic memory and in subjective experience with one study suggesting persistence of schizotypal and perceptual changes after cessation of use.[56]

(b) New class of drugs

Tryptamine (1H-indole-3-ethanamine) is a naturally occurring metabolite of tryptophan. It forms the parent nucleus of a wide

range of hallucinogenic drugs, some entirely synthetic (LSD, *N*,*N*-dimethyltryptamine) but many naturally occurring in plants, fungi (psilocybin—the psychoactive component of 'magic mushrooms'), and occasionally animals. It seems unlikely that many tryptamines other than LSD and psilocybin will be used unduly in dance clubs because they possess few stimulant properties, and need to be smoked or injected because most are inactive by mouth unless taken with a monoamine oxidase inhibitor.An example of the latter is the combination of *N*,*N*-dimethyltryptamine (the hallucinogen) and harmine (the activator) in the hallucinogenic drink ayahuasca or caapi used in rituals by South American Indians. Drugs such a *N*,*N*-dimethyltryptamine may have adverse effects upon both the cardiovascular system and on temperature regulation. They may induce unpleasant hallucinogenic experiences.

As described earlier the analogues MBDB, MDEA, and MDA all share similar properties with MDMA and at times have been marketed as distinct drugs. More recently 4-Bromo-2, 5-dimethoxyphenethylamine (also known as 2C-B, Nexus) a drug with similar hallucinogenic properties as psilocybin and mescaline has become available within the dance scene in Europe. Others such a 2C-T2 (2,5-dimethoxy-4-ethylthio-β-phenethylamine) suggest that new additions to an expanding street pharmacy are unlikely to become a thing of the past any time soon. Newer drugs are likely to be less easily detected by toxicologist and may have unfamiliar or atypical clinical manifestations.

Conclusion

MDMA, methamphetamine, GHB, and ketamine are all capable of producing acute adverse psychological experiences in normal users and exacerbating symptoms in those with underlying psychological disorders. They also to varying degrees pose the risk of long-term neuropsychiatric consequences. Although dependent patterns of use are not commonly seen with this group of drugs, methamphetamines certainly can result in the very rapid development of severe dependence. Most acute presentations are typically short-lived and self-limiting and are only very rarely life-threatening. The precipitation of an underlying psychiatric disorder or an exacerbation of premorbid traits may well be one of the longer term consequences of heavy use of these drugs. In those who present with acute drug-related psychological symptoms there should be an emphasis on follow-up since in some cases the symptoms will represent the onset of a persistent independent disorder which requires treatment. Users who have experienced acute psychological problems should be encouraged to make the attribution that there may be something inherent in them that makes them susceptible to experiencing the unpleasant reactions with a drug and that they are likely to remain vulnerable to those adverse experiences. This may be difficult to accept for potentially vulnerable young people who may prefer to think that the experience was not enjoyable because the drugs were not good – 'it was a bad pill'.

Further information

www.emcdda.europa.eu
www.erowid.org
www.maps.org

References

1. Winstock, A.R., Griffiths, P., and Stewart, D. (2001). Drugs and the dance music scene: a survey of current drug use patterns among a sample of dance music enthusiasts in the UK. *Drug and Alcohol Dependence*, **64**, 9–17.
2. Hansen, D., Maycock, B., and Lower, T. (2001). 'Weddings, parties, anything', a qualitative analysis of ecstasy use in Perth, Western Australia. *International Journal of Drug Policy*, **12**, 181–99.
3. Schifano, F., Corkery, J., Deluca, P., *et al.* (2006). Ecstasy (MDMA, MDA, MDEA, MBDB) consumption, seizures, related offences, prices, dosage levels and deaths in the UK (1994–2003). *Journal of Psychopharmacology*, **20**, 456–63.
4. Winstock, A.R., Wolff, K., and Ramsey, J. (2001). Ecstasy pill testing: harm minimization gone too far? *Addiction*, **96**, 1139–48.
5. Parrott, A.C. (2004). Is ecstasy MDMA? A review of the proportion of ecstasy tablets containing MDMA, their dosage levels, and the changing perceptions of purity. *Psychopharmacology*, **173**, 234–41.
6. Byard, R.W., Gilbert, J., James, R., *et al.* (1998). Amphetamine derivative fatalities in South Australia: is 'ecstasy' the culprit? *The American Journal of Forensic Medicine and Pathology*, **19**, 261–5.
7. Webb, E., Ashton, C.H., Kelly, P., *et al.* (1996). Alcohol and drug use in UK university students. *Lancet*, **348**, 922–5.
8. Condon, J. and Smith, N. (2003). *Prevalence of drug use: key findings from the 2002/2003 British Crime Survey. 229*. Home Office, London.
9. Copeland, J., Dillon, P., and Gascoigne, M. (2006). Ecstasy and the concomitant use of pharmaceuticals. *Addictive Behaviors*, **31**, 367–70.
10. Schifano, F. (2004). A bitter pill. Overview of ecstasy (MDMA, MDA) related fatalities. *Psychopharmacology*, **173**, 242–8.
11. McCann, U.D., Eligulashvili, V., and Ricaurte, G.A. (2000). (±)3,4 Methylenedioxymethamphetamine ('Ecstasy')-induced serotonin neurotoxicity in clinical studies. *Neuropsychobiology*, **42**, 11–6.
12. Topp, L., Hando, J., Dillon, P., *et al.* (1999). Ecstasy use in Australia: patterns of use and associated harm. *Drug and Alcohol Dependence*, **55**, 105–15.
13. Parrott, A.C. (2001). Human psychopharmacology of ecstasy (MDMA): a review of 15 years of empirical research. *Human Psychopharmacology*, **16**, 557–77.
14. Schifano, F., Oyefeso, A., Corkery, J., *et al.* (2003). Death rates from ecstasy (MDMA, MDA) and polydrug use in England and Wales 1996–2002. *Human Psychopharmacology Clinical Experimental*, **18**, 519–24.
15. Tancer, M. and Johanson, C.E. (2003). Reinforcing, subjective, and physiological effects of MDMA in humans: a comparison with d-amphetamine and mCPP. *Drug and Alcohol Dependence*, **72**, 33–44.
16. de la Torre, R., Farré, M., Ortuño, J., *et al.* (2000). Non-linear pharmacokinetics of MDMA ("ecstasy") in humans. *British Journal of Clinical Pharmacology*, **49**, 104–9.
17. Harris, D.S., Baggott, M., Mendelson, J.H., *et al.* (2002). Subjective and hormonal effects of 3,4-methylenedioxymethamphetamine (MDMA) in humans. *Psychopharmacology*, **162**, 396–405.
18. Hernandez-Lopez, C., Farre, M., Roset, P.N., *et al.* (2002). 3,4- Methylenedioxymethamphetamine (ecstasy) and alcohol interactions in humans: psychomotor performance, subjective effects, and pharmacokinetics. *The Journal of Pharmacology and Experimental Therapeutics*, **300**, 236–44.
19. Budisavljevic, M.N., Stewart, L., Sahn, S.A., *et al.* (2003). Hyponatremia associated with 3,4-methylenedioxymethylamphetamine ('ecstasy') abuse. *The American Journal of the Medical Sciences*, **326**, 89–93.
20. Wolff, K., Tsapakis, E.M., Winstock, A.R., *et al.* (2006). Vasopressin and oxytocin secretion in response to the consumption of ecstasy in a clubbing population. *Journal of Psychopharmacology*, **20**, 400–10.

21. Parrott, A.C. (2007). Ecstasy versus alcohol: Tolstoy and the variations of unhappiness. *Journal of Psychopharmacology*, **21**, 3–6.
22. Liechti, M.E. and Vollenweider, F.X. (2000). The serotonin uptake inhibitor citalopram reduces acute cardiovascular and vegetative effects of 3,4 methylenedioxymethamphetamine ('ecstasy') in healthy volunteers. *Journal of Psychopharmacology*, **14**, 269–74.
23. Verheyden, S.L., Henry, J.A., and Curran, H.V. (2003). Acute, sub-acute and long-term subjective consequences of 'ecstasy' (MDMA) consumption in 430 regular users. *Human Psychopharmacology Clinical Experimental*, **18**, 507–17.
24. Liester, M.B., Grob, C.S., Bravo, G.L., *et al.* (1992). Phenomenology and sequelae of 3,4-methylenedioxymethamphetamine use. *The Journal of Nervous and Mental Disease*, **180**, 345–54.
25. Nichols, D.E. (1986). Differences between the mechanism of action of MDMA, MBDB, and the classic hallucinogens. Identificationof a new therapeutic class: entactogens. *Journal of Psychoactive Drugs*, **18**, 305–13.
26. Shulgin, A.T. and Nichols, D.E. (1978). Characteristics of 3 new psychomimetics. In *The pharmacology of hallucinogens* (eds. N.C. Stillman and N.E. Willete). Pergamon Press, Oxford.
27. Parrott, C., Buchanan, T., Scholey, A.B., *et al.* (2002). Ecstasy/MDMA attributed problems reported by novice, moderate and heavy recreational users. *Human Psychopharmacology Clinical Experimental*, **17**, 309–12.
28. Curran, H.V. and Travill, R.A. (1997). Mood and cognitive effects of ±3,4-methylenedioxymethamphetamine (MDMA, 'ecstasy'): weekend 'high' followed by mid-week low. *Addiction*, **92**, 821–31.
29. Guillot, C. and Greenway, D. (2006). Recreational ecstasy use and depression. *Journal of Psychopharmacology*, **20**, 411–6.
30. Schifano, F., Di Furia, L., Forza, C., *et al.* (1998). MDMA ('ecstasy') consumption in the context of polydrug abuse: a report on 150 patients. *Drug and Alcohol Dependence*, **52**, 85–90.
31. Morgan, M.J., McFie, L., Fleetwood, H., *et al.* (2002). Ecstasy (MDMA): are the psychological problems associated with its use reversed by prolonged abstinence? *Psychopharmacology*, **159**, 294–303.
32. Verrico, C.D., Miller, G.M., and Madras, B.K. (2007). MDMA (Ecstasy) and human dopamine, norepinephrine, and serotonin transporters: implications for MDMA-induced neurotoxicity and treatment. *Psychopharmacology*, **189**, 489–503.
33. Colado, M.I, O'Shea, E., and Green, A.R. (2004). Acute and long-term effects of MDMA on cerebral dopamine biochemistry and function. *Psychopharmacology*, **173**, 249–63.
34. Cowan, R.L. (2007). Neuroimaging research in human MDMA users: a review. *Psychopharmacology*, **189**, 539–56.
35. Roiser, J.P., Cook, L.J., Cooper, J.D., *et al.* (2005). Association of a functional polymorphism in the serotonin transporter gene with abnormal emotional processing in ecstasy users. *The American Journal of Psychiatry*, **162**, 609–12.
36. Hanson, K.L. and Luciana, M. (2004). Neurocognitive function in users of MDMA: the importance of clinically significant patterns of use. *Psychological Medicine*, **34**, 229–46.
37. Gouzoulis-Mayfrank, E., Daumann, J., Tuchtenhagen, F., *et al.* (2000). Impaired cognitive performance in drug free users of recreational ecstasy (MDMA)[see comment]. *Journal of Neurology, Neurosurgery, and Psychiatry*, **68**, 719–25.
38. McCann, U.C. and Ricaurte, G.A. (1991). Lasting neuropsychiatric sequelae of (±)3,4 methylenedioxymethamphetamine ('ecstasy') in recreational users. *Journal of Clinical Psychopharmacology*, **11**, 302–5.
39. Gerra, G., Zaimovic, A., Giucastro, G., *et al.* (1998). Serotonergic function after (+/-)3,4-methylene-dioxymethamphetamine ('Ecstasy') in humans. *International Clinical Psychopharmacology*, **13**, 1–9.
40. Singer, L.T., Linares, T.J., Ntiri, S., *et al.* (2004). Psychosocial profiles of older adolescent MDMA users. *Drug and Alcohol Dependence*, **74**, 245–52.
41. Guillot, C.R. and Berman, M.E. (2007). MDMA (Ecstasy) use and psychiatric problems. *Psychopharmacology*, **189**, 575–6.
42. Kosten, T.R., Markou, A., and Koob, G.F. (1998). Depression and stimulant dependence: neurobiology and pharmacotherapy. *The Journal of Nervous and Mental Disease*, **186**, 737–45.
43. Bolding, G., Hart, G., Sherr, L., *et al.* (2006). Use of crystal methamphetamine among gay men in London. *Addiction*, **101**, 1622–30.
44. Lambert, G., Johansson, M., Agren, H., *et al.* (2000). Reduced brain norepinephrine and dopamine release in treatment-refractory depressive illness: evidence in support of the catecholamine hypothesis of mood disorders. *Archives of General Psychiatry*, **57**, 787–93.
45. Baker, A., Lee, N.K., and Jenner, L. (eds.) (2004). *Models of intervention and care for psychostimulant users* (2nd edn).
46. Shearer, J. and Gowing, L.R. (2004). Pharmacotherapies for problematic psychostimulant use: a review of current research. *Drug and Alcohol Review*, **23**, 203–11.
47. McKetin, R., McLaren, J., Lubman, D.I., *et al.* (2006). The prevalence of psychotic symptoms among methamphetamine users. *Addiction*, **101**, 1473–8.
48. Harris, D. and Batki, S.L. (2000). Stimulant psychosis: symptom profile and acute clinical course. *The American Journal on Addictions*, **9**, 28–37.
49. Rodgers, J., Ashton, C.H., Gilvarry, E., *et al.* (2004). Liquid ecstasy: a new kid on the dance floor. *The British Journal of Psychiatry*, **184**, 104–6.
50. Degenhardt, L., Darke, S., and Dillon, P. (2003). The prevalence and correlates of gamma-hydroxybutyrate (GHB) overdose among Australian users. *Addiction*, **98**, 199–204.
51. Deveaux, M., Renet, S., Renet, V., *et al.* (2002). Utilisation de l'acide gamma-hydroxybutyrique (GHB) dans les rave-parties et pour la soumission chimique en France: mythe ou realite? *Acta Clinica Belgica—Supplementum*, (1), 37–40.
52. Galloway, G.P., Frederick, S.L., and Gonzales, M. (1997). Gamma-hydroxybutyrate: an emerging drug of abuse that causes physical dependence. *Addiction*, **92**, 89–96.
53. White, P.F., Way, W.L., and Trevor, A.J. (1982). Ketamine—its pharmacology and therapeutic uses. *Anesthesiology*, **56**, 119–36.
54. Wolff, K. and Winstock, A.R. (2006). Ketamine: from medicine to misuse. *CNS Drugs*, **20**, 199–218.
55. Lim, D.K. (2003). Ketamine associated psychedelic effects and dependence. *Singapore Medical Journal*, **44**, 31–4.
56. Curran, H.V., Rees, H., Hoare, T., *et al.* (2004). Empathy and aggression: two faces of ecstasy? A study of interpretative cognitive bias and mood change in ecstasy users. *Psychopharmacology*, **173**, 425–33.

4.2.3.6 Disorders relating to the use of volatile substances

Richard Ives

Introduction

Volatile substance abuse (**VSA**)—also known as 'solvent abuse' and 'inhalant abuse'—is the deliberate inhalation of any of a range of products (see Table 4.2.3.6.1[1]), to achieve intoxication. Amyl (pentyl) and isobutyl nitrites ('poppers') have different patterns of misuse, and are not discussed here.[2]

VSA has dose-related effects similar to those of other hypnosedatives. Small doses rapidly lead to 'drunken' behaviour similar to the

Table 4.2.3.6.1 Some products which can be abused by inhalation

Product	Major volatile components
Adhesives	
Balsa wood cement	Ethyl acetate
Contact adhesives	Butanone, hexane, toluene, and esters
Cycle tyre repair cement	Toluene, and xylenes
Woodworking adhesives	Xylenes
Polyvinylchloride (PVC) cement	Acetone, butanone, cyclohexanone, trichloroethylene
Aerosols	
Air freshener	LPG, DME, and/or fluorocarbons
Deodorants, antiperspirants	LPG, DME, and/or fluorocarbons
Fly spray	LPG, DME, and/or fluorocarbons
Hair lacquer	LPG, DME, and/or fluorocarbons
Paint sprayers	LPG, DME, and/or fluorocarbons and esters
Anaesthetics/analgesics	
Inhalational	Nitrous oxide, cyclopropane Diethyl ether, halothane, enflurane, isoflurane
Topical	FC 11, FC 12, monochloroethane
Dust removers (air brushes)	DME, FC 22
Commercial dry cleaning and degreasing agents	Dichloromethane, FC 113, methanol, 1,1,1-trichloroethane, tetrachloroethylene, toluene, trichloroethylene (now rarely carbon tetrachloride, 1,2-dichloropropane)
Domestic spot removers and dry cleaners	Dichloromethane, 1,1,1-Trichloroethane, tetrachloroethylene, trichloroethylene
Fire extinguishers	Bromochlorodifluoromethane, FC 11, FC 12
Fuel gases	
Cigarette lighter refills	LPG
'Butane'	LPG
'Propane'	Propane and butanes
Nail varnish/nail varnish remover	Acetone and esters
Paints/paint thinners	Acetone, butanone, esters, hexane, toluene, trichloroethylene, xylenes
Paint stripper	Dichloromethane, methanol, toluene
'Room odorizer'	

(Reproduced from B. Flanagan and R. Ives Volatile substance abuse, *Bulletin on Narcotics*, **XLVI**, 49–78, copyright 1994, UNODC.).

effects of alcohol, and may induce delusions and hallucinations. Some heavy misusers inhale large quantities; 6 l of adhesive weekly have been reported.

Long-term effects include listlessness, anorexia, and moodiness. The hair, breath, and clothing may smell of the substance(s) used, and empty product containers (e.g. glue cans, cigarette lighter refills, and aerosol spray cans), and bags used to inhale from, may be found.

Being readily available, volatile substances are, along with alcohol and tobacco, the first intoxicating substances some children try. However, most VSA is experimental and does not lead to the use of other psychoactive substances; problematic misusers have other difficulties in their lives.[3]

History

Inhaling substances to achieve intoxication is not new. Inhaling ether and nitrous oxide ('laughing gas'), as well as commercially available volatile products, has a long history.

Public concern is more recent. In the United States during the 1950s and 1960s there was much publicity about glue sniffing; this helped to publicize the possibilities of glue as an intoxicant.[4] Only in the 1970s did public concern about VSA emerge in the United Kingdom, to reach a peak in 1983 when there were more press cuttings on the subject than on all other drugs.[5] Public anxiety has since waned, although the problem has not disappeared.

Prevalence of VSA

VSA is a worldwide problem. For an overview, see a WHO report,[6] and a National Institute on Drug Abuse (NIDA) report.[7] The European Schools survey Project on Alcohol and other Drugs (ESPAD) report provides 2003 data from 35 European countries: lifetime experience of VSA (i.e. whether *ever* tried VSA) among 15- to 16-year-olds varied from 2 per cent (in Romania) to 22 per cent (in Greenland). In the United Kingdom (and Iceland) 12 per cent reported trying VSA—nine countries had a higher prevalence. There was little difference between boys' and girls' lifetime prevalence.[8] Although young people from all socio-economic groups experiment with volatile substances, for some among the poor and the dispossessed, VSA is the drug of choice.[9] VSA is a particular problem among people living on the street. Chronic VSA is associated with poor socio-economic conditions, with delinquency and illegal drug use,[10] disrupted families, and other social and psychological problems.[11]

VSA deaths

Even for first-time experimenters, death from VSA is an ever-present risk. Death may ensue from convulsions and coma, inhalation of vomit, or direct cardiac or central nervous system toxicity. Sudden deaths of young people should be thoroughly investigated, as VSA-related deaths can be overlooked. Post-mortem examination usually reveals little, except perhaps acute lung congestion and possibly cold-induced burns to the mouth and throat. Toxicological examination of blood and tissue specimens is used to confirm a diagnosis of VSA-related death.[12]

A long-term study in the United Kingdom identified 2152 VSA-related deaths between 1971 and 2004.[13] The death rate peaked in 1990 with 152 deaths, declining since, with 47 deaths being recorded in 2004. Under-18s make up half the deaths, although the age of death is increasing. Most are male, although the overall proportion of female deaths has risen and in the first 5 years of the new millennium females comprised 22 per cent of the deaths (Table 4.2.3.6.2). Volatile-substance-related deaths occur in all social classes in the United Kingdom. However, ' . . . in deaths of

Table 4.2.3.6.2 VSA-related deaths in the United Kingdom—selected years

Year	1983	1985	1987	1989	1990	1993	1996	1998	1999	2000	2001	2002	2003	2004
N	82	117	115	113	152	79	78	80	75	66	63	65	53	47

Reproduced from Field-Smith, M.E., Butland, B.K., Ramsey, J.D. and Anderson, H.R. Trends in deaths associated with abuse of volatile substances 1971–2004, copyright 2006, St. George's University of London, London.

those under 16, there was a marked difference in mortality between social classes I and V, with nearly four times as many deaths occurring in social class V . . . compared with social class I'.[14]

Health issues

Health effects of volatile substances include the following:

- A sensitization of the heart—so that cardiac arrhythmias may occur if VSA is followed by exertion or fright.
- Cooling of the throat tissues—caused by spraying substances directly into the mouth, which may causing swelling and suffocation.
- A risk of fire—especially when combined with smoking; many products are inflammable.
- Suffocation—a particular danger if large plastic bags are used.
- Most products are mixtures of chemicals, and manufacturers do not list the constituents. Changing product formulations make the dangers unpredictable.
- Using alone in an isolated place presents special hazards.
- When combined with alcohol or other drugs, the effects can be unpredictable.
- Intoxication itself has potential dangers, for example, greater recklessness, doing bizarre things in response to hallucinations, becoming unconscious, and choking on vomit.

Apart from the real risk of death, VSA rarely causes long-term damage. However, some products contain poisonous substances, such as lead in some petrol or *n*-hexane in some glues. Chronic abuse of toluene-containing products and of chlorinated solvents such as 1,1,1-trichloroethane sometimes causes damage to the liver and kidneys. Damage to the lungs, bone marrow, and nervous system is also known, but is uncommon and generally reversible. Some people are more vulnerable (genetically or otherwise) than others to certain harmful effects. However, the long-term effects of sniffing have not been thoroughly studied, and virtually all reports of chronic toxicity are case studies, so the actual morbidity from VSA is not known.

A review article looking at the possibility of cognitive impairments concluded that: 'the possibility that permanent structural brain damage, with accompanying psychiatric manifestations, results from solvent abuse remains inconclusive'.[15]

Users develop tolerance. Although no dependence syndrome exists, a few young people develop a more compulsive and long-term habit. The UK Advisory Council on the Misuse of Drugs suggested that: 'There are . . . pharmacological reasons for suspecting that persistent exposure to volatile substances might be able to induce a dependence of the so-called depressant type'.[16]

Many volatile substance misusers are also users of other drugs, both legal and illegal. Poly-drug use may potentiate the effects of individual drugs and make it difficult to assess the risks of individual substances.[17]

VSA during pregnancy is associated with increased maternal and foetal morbidity.[18] Paternal exposure to volatile substances may also have deleterious effects on their offspring.[19] But the complexities of the chemicals involved—and the complexities of people's lives—make it difficult to identify specific causes of foetal damage difficult.

Treatment of VSA-related disorders

Emergency treatment

The immediate treatment of an intoxicated person needs a calm and firm approach. The product being misused should be removed; although not if this would lead to conflict—exertion or high emotion may raise adrenaline to dangerous levels for an over-sensitized heart. Therefore, an intoxicated person should be kept calm and never chased. It is unlikely that it will be possible to have a serious conversation with an intoxicated misuser, but calming and reassuring talk may help. After 5 to 20 min without inhalation the abuser will sober up (unless alcohol or other drugs have also been used). Subsequently, medical help might be needed; a check-up may identify particular health problems.

Cessation

No special regime is necessary when stopping misusing volatile substances. Although, being lipid-soluble, the chemicals may be detectable in the body tissues for some weeks after cessation of use, they do not have any psychoactive effect. There is no clearly defined withdrawal syndrome and special detoxification regimes are unnecessary, but rest, sleep, and good food may aid recovery.

Dealing with experimental misuse

Most teenagers never even try misusing volatile substances; those who do, do so only a few times, and even those who do so more frequently do not continue for long. Experimental or the occasional misuse of volatile substances occurs mainly from curiosity or as part of peer group activity. Appropriate intervention may simply involve a warning of the dangers, plus increased supervision. Specialist treatment is not required, and may be counterproductive, entrenching an otherwise transient activity.

Dealing with dependent misuse

Biology may predispose to dependent use, but chronic VSA is connected with other problems. As group of United Kingdom professionals put it:[20]

> Persistent misuse of volatile substances is a complex behaviour . . . frequently associated with low self-esteem, family problems, isolation and psychological difficulties. These are factors that may also be associated with the problematic use of legal and illegal drugs, and indeed,

> a large proportion of people who misuse volatile substances also misuse other drugs. Chronic VSA is thus intertwined with social and psychological problems and with the misuse of illegal drugs. Therefore, counselling services for young people should not be narrowly focused on volatile substances, but should be able to deal with VSA in the context of a range of problematic behaviours.

Often, these other problems need attention first, and until these are dealt with, the misuser—even while recognizing the harm—may not give up. Consequently, generic services, which can deal more effectively with these broader problems, should lead on care, supported where necessary by specialist agencies. Mental health services, as well as drugs' services, have an important role in giving this support, for example, in the treatment of psychiatric comorbidity (dual diagnosis). Specialist services can also help to identify areas for intervention; implementation should take account of social and cultural patterns of the misuser's life. Female volatile substance misusers may not readily present for treatment and can suffer additional stigmatization. Services for young people need to be specifically designed for their needs and cognizant of issues such as confidentiality and consent. Families who struggle unaided with problematic VSA by a young family member may also need help.

The Modified Social Stress Model, developed by the WHO Street Children Project, gives a framework for understanding substance use.[21] Potential for change can be assessed using Prochaska and DiClemente's 'revolving door' model of stages of change.[22] Jumper-Thurman and colleagues point out that the treatment of volatile substance misusers:[23]

> has presented a particularly difficult challenge . . . given the general lack of direction for effective treatment strategies. In addition to the physiological, neurological, and emotional challenges abusers face . . . [they] bring with them a multitude of other problems—academic, legal, social, and family issues.

Some groups (such as people living on the street, and indigenous peoples) have special problems with substance use that require different, more holistic, attention. Treatment should work 'with the grain' of the culture, rather than imposing inappropriate 'alien' treatment models. Indigenous peoples are beginning to insist that their cultures have useful perspectives and approaches that can be utilized in the treatment of people with drug and volatile substance problems.[24]

Follow-up

After-care, long-term rehabilitation, social reinsertion, relapse management, and follow-up of discharged patients are important aspects of the treatment process.

Relapse, which is common, should be treated non-judgementally; not as 'failure' but as an opportunity for learning. Support in maintaining improvements may be helpful; for example, events for ex-users of volatile substances to help them to maintain abstinence and to utilize group support.

Harm minimization

Because of the unpredictable dangers of VSA, harm minimization advice should not be routinely given. However, very entrenched misusers may benefit from careful individual guidance on minimizing the risks, such as avoiding spraying gases directly into the mouth.

Wider aspects

Chronic VSA is not an individual problem: it arises not only from individual pathology, but also from failures in social structures. Treatment, in the broadest sense, needs to help the healing of the family, the community, and to assist in making changes in society.

Measuring outcomes

Evaluation and monitoring need careful thought and planning. But treatment interventions have multiple aims and varied outcomes. Aims may alter and be adapted as the work develops, so that outcomes will be difficult to assess in relation to the original aims. Evaluation should handle this complexity: identifying 'success' requires measures beyond the simple calculation of reduction in, or abstention from, substance use.

Building evaluation in from the start, and using it to inform the intervention throughout, makes it part of the process of intervention; the reflection that monitoring and evaluation encourages can increase the effectiveness of the intervention.

Prevention

There are many different sniffable products, many possibilities for substitution of one product for another, many different chemicals involved, and insufficient information about the relative harm of various products and practices. Volatile substance misusers are generally young, and VSA-related deaths are sudden and unpredictable. All these factors make prevention difficult.

Tackling the supply of products

This can be approached in several ways:

- Product elimination—while it is not possible to eliminate all volatile substances, some products are particularly dangerous, and have satisfactory substitutes.
- Product modification—there are three possibilities (a review paper gives more details:[25])
  - changing the formulation of the product to remove the intoxicating substance. There has been great success in Australian indigenous communities through substituting petrol with unsniffable 'Opal', an unleaded fuel with low levels of aromatics[26,27]
  - adding a chemical to make the product unpalatable (experiments with the bittering agent, *Bitrex*, have been inconclusive)
  - and modifying the container to make misuse difficult.
- Warning labels—these may be helpful, although labelling draws attention to the potential for misuse. Many sniffable products in the United Kingdom carry the 'SACKI' warning, 'Solvent Abuse Can Kill Instantly'.[28]
- Education for suppliers—retailers need information and advice about a product's potential for misuse. This is difficult, as the various products are sold through many retail outlets.
- Legal controls on the sale and supply of misusable products—these exist in many countries but are difficult to enforce.

Tackling the demand for products

- Legal controls—in Japan, Singapore, and the Republic of Korea, VSA is an offence. However, this is not so in most countries because the criminalization of misusers of volatile substances is considered counterproductive.
- Information and education—this can be provided through public advertising, leaflets, helplines, in schools and informal education. Early education about volatile substances is essential; these products are in most people's homes and therefore (unlike illegal drugs) they can be accessed by very young children. Because many parents are unaware of the misuse potential of household products, information should also be targeted at parents.

All these strategies should be considered; as the Advisory Council on the Misuse of Drugs pointed out, 'good practice' will constitute a layered series of alternative or multiple strategies rather than any one master stroke.[29]

Further information

Advisory Council on the Misuse of Drugs. (1995). *Volatile substance abuse*. HMSO, London.

Skellington Orr, K. and Shewan, D. (2006). *Review of evidence relating to volatile substance abuse in Scotland*. Scottish Executive Substance Misuse Research Programme, Edinburgh (www.scotland.gov.uk/Resource/Doc/147377/0038818.pdf accessed 22-03-07).

The UK Government's. (2005). Strategy for tackling volatile substance (*Out of sight—Not out of mind. Children, young people and volatile substance abuse (VSA) A framework for VSA*) can be downloaded from http://www.dh.gov.uk/assetRoot/04/11/56/05/04115605.pdf.) (accessed 22-03-07).

Flanagan, R.J., Streete, P.J., and Ramsey, J.D. (1997). *Volatile substance abuse: practical guidelines for the analytical investigation of suspected cases and interpretation of results*. United Nations Drug Control Programme, Vienna.

Field-Smith, M.E., Butland, B.K., Ramsey, J.D., *et al.* (2006). *Trends in deaths associated with abuse of volatile substances 1971–2004*. St George's, University of London, London.

Ron, M. (1986). Volatile substance abuse: a review of possible long-term neurological, intellectual and psychiatric sequelae. *The British Journal of Psychiatry*, **148**, 235–46.

Reading list from the UK's DrugScope www.drugscope.org.uk/wip/7/PDFS/Inhalants%20and%20solvents.pdf (accessed 22-03-07).

Information on UK VSA-related deaths: www.vsareport.org

USA National Inhalants Prevention Coalition: www.inhalants.org

References

1. Flanagan, B. and Ives, R. (1994). Volatile substance abuse. *Bulletin on Narcotics*, **46**, 49–78.
2. Haverkos, H. and Dougherty, J. (eds.) (1988). *Health hazards of nitrite inhalants, NIDA Research Monograph 83*. National Institute on Drug Abuse, Rockville, MD.
3. Davies, B., Thorley, A., and O'Connor, D. (1985). Progression of addiction careers in young adult solvent misusers. *British Medical Journal*, **290**, 6482.
4. Brecher, E. (1972). How to launch a nation-wide drug menace. In *Licit and illicit drugs*, (eds. E.M. Brecher and the Editors of Consumer Report Magazine) pp. 321–34. Little, Brown, Boston, MA.
5. Ives, R. (1986). The rise and fall of the solvents panic. *Druglink*, **1**, 10–12.
6. WHO Programme on Substance Abuse. (1999). *Volatile substance abuse: a global overview*. WHO, Geneva.
7. Kozel, N., Sloboda, Z., and De La Rosa, M. (eds.) (1995). *Epidemiology of inhalant abuse: an international perspective (NIDA Research Monograph series No. 148)*. NIDA, Washington, DC.
8. Ives, R. (2006). Volatile Substance Abuse: a review of findings in ESPAD 2003. *Drugs: Education, Prevention and Policy*, **13**, 441–9.
9. Beauvais, F., Wayman, J. C., Jumper-Thurman, P., *et al.* (2004). Inhalant abuse among American Indian, Mexican American, and non-Latino white adolescents. *The American Journal of Drug and Alcohol Abuse*, **28**, 2171–87.
10. NSDUH. (2005). *Inhalant use and delinquent behaviors among young adolescents* (USA 'National Survey on Drug User and Health') (http://oas.samhsa.gov/2k5/inhale/inhale.htm (accessed 22-03-2007)).
11. Ives, R. (1999). *Volatile substance abuse: a report on survey evidence*. Health Education Authority, London.
12. Flanagan, R.J., Streete, P.J., and Ramsey, J.D. (1997). *Volatile substance abuse: practical guidelines for the analytical investigation of suspected cases and interpretation of results.* United Nations Drug Control Programme, Vienna.
13. Field-Smith, M.E., Butland, B.K., Ramsey, J.D., *et al.* (2006). *Trends in deaths associated with abuse of volatile substances 1971–2004.* St George's, University of London, London.
14. Esmail, A., Meyer, L., Pottier, A., *et al.* (1993). Deaths from volatile substance abuse in those under 18 years: results from a national epidemiological study. *Archives of Disease in Childhood*, **69**, 356–60 (see p. 358).
15. Ron, M. (1986). Volatile substance abuse: a review of possible long-term neurological, intellectual and psychiatric sequelae. *The British Journal of Psychiatry*, **148**, 235–46.
16. Advisory Council on the Misuse of Drugs. (1995). *Volatile substance abuse*, paragraph 3.11. HMSO, London.
17. Ives, R. and Ghelani, P. (2006). Polydrug use (the use of drugs in combination): a brief review. *Drugs: Education, Prevention and Policy*, **13**, 225–32.
18. Wilkin, H. and Gabow, P. (1991). Toluene abuse during pregnancy: obstetric complications and perinatal outcomes. *Obstetrics and Gynaecology*, **77**, 504–8.
19. Robaire, B. and Hales, B. (1993). Paternal exposure to chemicals before conception (editorial). *British Medical Journal*, **307**, 341–2.
20. Network VSA. (1997). *Network VSA statement about volatile substance misuse*. Network VSA, Blackburn, Lancs.
21. WHO Programme on Substance Abuse. (1997). *Street children, substance use and health: training for street educators*, p. 48. WHO, Geneva.
22. Prochaska, J. and DiClemente, C. (1986). Towards a comprehensive model of change. In *Treating addictive behaviours: processes of change* (eds. W. Miller and N. Heather), pp. 3–27. Plenum Press, New York.
23. Jumper-Thurman, P., Plested, B., and Beauvais, F. (1992). Treatment strategies for volatile substance abusers in the United States. In *Inhalant abuse: a volatile research agenda* (eds. C. Sharpe, F. Beauvais, and R. Spence), NIDA Research Monograph 129. National Institute on Drug Abuse, Rockville, MD.
24. Brady, M. (1992). *Heavy metal. The social meaning of petrol sniffing in Australia*. Aboriginal Studies Press, Canberra.
25. MacLean, S. and d'Abbs, P. (2006). Will modifying inhalants reduce volatile substance misuse? A review. *Drugs: Education, Prevention and Policy*, **13**, 423–39.
26. Shaw, G., Biven A., Gray, D., *et al.* (2004). *An evaluation of the Comgas scheme*. Australian Government Department of Health and Human Ageing, Canberra, ACT.
27. www.health.gov.au/petrolsniffingprevention (accessed 22-03-07).
28. www.bama.co.uk/educ_3.html (accessed 22-03-07).
29. Advisory Council on the Misuse of Drugs. (1995). *Volatile substance abuse*, paragraph 5.5. HMSO, London.

4.2.3.7 The mental health effects of cannabis use

Wayne Hall

Cannabis the drug

Cannabis products are derived from the female plant of *Cannabis sativa* and contain the psychoactive constituent delta-9-tetrahydrocannabinol (**THC**).[1] Marijuana (THC content typically 0.5–5 per cent) is prepared from the dried flowering tops and leaves of the plant. Hashish (THC content typically 2–20 per cent) consists of dried cannabis resin and compressed flowers.[1]

Cannabis is usually smoked in a 'joint', like a tobacco cigarette, or in a water pipe, often mixed with tobacco. Although marijuana and hashish may be eaten, cannabis is usually smoked because this is the most efficient way to achieve the desired effects.[2]

THC acts on a widely distributed, specific receptor in brain regions that are involved in cognition, memory, reward, pain perception, and motor coordination.[1] These receptors respond to an endogenous ligand, anandamide, which is considerably less potent and has a shorter duration of action than THC.[1]

Patterns of cannabis use

Cannabis has been tried by many young adults in Europe, the United States, and Australia.[3] Most cannabis users in these countries start in their mid to late teens and stop in their middle to late 20s.[3,4] In the United States and Australia, about 10 per cent of those who ever use cannabis become daily users, and another 20 to 30 per cent use weekly.[3] This pattern of use differs from that found in traditional cannabis-using countries, such as Egypt and India, where recreational cannabis use is uncommon and heavy cannabis use is confined to small, marginalized groups.[3]

'Heavy' cannabis use is usually defined as daily or near-daily use.[5] This pattern of use places users at the greatest risk of experiencing adverse psychological and physical consequences.[2,3] Daily cannabis users are also more likely to be regular users of alcohol and tobacco and to use amphetamines, hallucinogens, psychostimulants, sedatives, and opioids.[2,3]

Acute psychological effects of cannabis use

Cannabis produces euphoria and relaxation, perceptual alterations, impaired short-term memory and attention, and intensification of ordinary sensory experiences.[2] The most common unpleasant psychological effects are anxiety and panic reactions,[2] that are most often reported by naive users and are a common reason for discontinuing use.[2] Cannabis produces dose-related impairments in cognitive and behavioural functions that may impair ability to drive an automobile.[6]

Chronic psychological effects of cannabis use

Cannabis dependence

Animals and humans develop tolerance to the effects of THC,[1] and some heavy users experience withdrawal symptoms on the abrupt cessation of cannabis use.[7] During the 1990s there was an increase in the number of persons in the United States, Australia, and Europe seeking help to stop their cannabis use.[3]

A cannabis-dependence syndrome occurs in heavy chronic users of cannabis who report problems in controlling their cannabis use, but who continue despite experiencing adverse personal and social consequences.[8] The lifetime prevalence of cannabis abuse and dependence (as defined in DSM-IIIR) in the United States has been estimated at 4.4 per cent of adults.[9] Around 10 per cent of those who ever use cannabis will meet criteria for dependence at some point in their lives.[2,9]

It is not clear how cannabis dependence is best managed. Roffman and Stephens,[10] in summarizing the results of controlled trials of cognitive behavioural, relapse prevention and other psychological approaches to treatment, report low rates of abstinence at 12 months but substantial reductions in cannabis use and problems among those who continue to use cannabis.

Cannabis psychosis

High doses of THC have been reported to produce visual and auditory hallucinations, delusional ideas, and thought disorder in normal volunteers.[2] In traditional cannabis-using cultures, such as India, a 'cannabis psychosis' has been reported in which the symptoms are preceded by heavy cannabis use and remit after abstinence[11] but the existence of a 'cannabis psychosis' in Western cultures is still a matter for debate.[11]

Cannabis and schizophrenia

Cannabis use and schizophrenia are associated[12,13] and there is consistent evidence from a series of prospective studies in a number of different countries suggesting that cannabis use can precipitate schizophrenia in persons who are vulnerable because of a personal or family history of this disorder,[12,13] and possibly a genetic vulnerability.[14] This hypothesis is consistent with the stress-diathesis model of schizophrenia[15] and it is also biologically plausible because psychotic disorders involve disturbances in the dopamine neurotransmitter systems and cannabinoids, such as THC, increase dopamine release.[1]

Individuals with psychotic symptoms who use cannabis should be encouraged to stop or, at the very least, to reduce their frequency of use. The major challenge is in finding ways to persuade individuals with psychoses to stop doing something they enjoy and to help those who want to stop using cannabis but find it difficult to do so. Psychological interventions for cannabis dependence in individuals without psychoses produce modest rates of abstinence and many individuals with schizophrenia lack social support, may be cognitively impaired, are often unemployed, and may not comply with treatment. A recent Cochrane review[16] found no clear evidence that supported any type of substance abuse treatment in schizophrenia over standard care. The development of more effective pharmacologic and psychological methods of treatment for cannabis dependence in persons with psychoses is a research priority.

Other disorders

Cognitive impairment

Cannabis use acutely impairs cognitive functioning but long-term heavy use of cannabis does not appear to produce severe or grossly

debilitating impairment of cognitive function that is comparable to the impairments found in chronic heavy alcohol drinkers.[17] There is evidence that the long-term use of cannabis produces more subtle cognitive impairment in the higher cognitive functions of memory, attention and organization, and the integration of complex information.[16] This evidence suggests that, longer the period of heavy cannabis use, the more pronounced is the cognitive impairment.[17] But it remains to be decided whether these cognitive impairments antedate cannabis use, reflect poorer learning in non-academically oriented young people, or reflect neurotoxic effects of cannabis use that can be reversed after an extended period of abstinence.[18]

An 'amotivational syndrome'

Anecdotal reports that chronic heavy cannabis use impairs motivation and social performance have been described in societies with a long history of cannabis use, such as Egypt, the Caribbean, and elsewhere.[2] A similar pattern of behaviour among young Americans who were heavy cannabis users in the early 1970s was described as an 'amotivational syndrome'.[19] Field studies of chronic heavy cannabis users in societies with a tradition of such use, for example Costa Rica and Jamaica,[2] have produced evidence that has usually been interpreted as failing to demonstrate the existence of the amotivational syndrome. Critics have argued that these studies are unconvincing because the chronic users studied have come from socially marginal groups, so that the cognitive and motivational demands of their everyday lives were insufficient to detect any impairment caused by chronic cannabis use.[2]

The status of the amotivational syndrome remains contentious. Many clinicians find the cases of 'amotivational syndrome' compelling, while many researchers are more impressed by the largely negative findings of the field and epidemiological studies. Regular cannabis users can experience a loss of ambition and impaired school and occupational performance[2] and former cannabis users report that impaired occupational performance was their reason for stopping.[2] It may be more parsimonious to explain impaired motivation as a symptom of chronic cannabis intoxication and dependence than to invent a new syndrome.[2,3]

Flashbacks

There are case reports of users experiencing cannabis 'flashbacks', i.e. symptoms of cannabis intoxication days or weeks after the individual last used cannabis.[20] Because of their rarity, and the fact that many affected individuals have also used other drugs, it is difficult to decide whether these are rare events that are coincidental with cannabis use, the effects of other drugs that are often taken together with cannabis, rare consequences of cannabis use that only occur at much higher than usual doses, experiences that require unusual forms of personal vulnerability, or the results of interactions between cannabis and other drugs.[2]

Behavioural effects in adolescence

There has been understandable societal concern about the effects of rising rates of cannabis use among adolescents on their school performance, mental health and adjustment, and their use of other more hazardous illicit drugs.[3,21]

There is a strong cross-sectional association between heavy cannabis use in adolescence and the risk of discontinuing a high-school education and experiencing job instability in young adulthood.[20,22] However, in longitudinal studies the strength of this association is reduced but not eliminated when statistical adjustments are made for the fact that heavy cannabis users have lower academic aspirations and poorer high-school performance prior to using cannabis than their peers.[22]

There is some evidence that heavy cannabis use has adverse effects upon family formation, mental health, and involvement in drug-related crime.[3] In each case, the strong associations in cross-sectional studies are more modest in longitudinal studies after statistically controlling for associations between cannabis use and other pre-existing characteristics which independently predict these adverse outcomes.[22] It remains uncertain to what degree these modest relationships represent residual confounding[22] or real relationships.

A consistent finding in the United States[23] has been the regular sequence of initiation into drug use, in which cannabis use has typically preceded involvement with 'harder' illicit drugs such as stimulants and opioids. The interpretation of this sequence of events remains controversial.[24] There is support for two hypotheses: (i) there is a selective recruitment into cannabis use of non-conforming adolescents who have a propensity to use a variety of other illicit drugs and (ii) once recruited to cannabis use, it is the social interaction with drug-using peers and greater access to illicit drug markets that increases the likelihood of using other illicit drugs.[23]

A major public health challenge will be finding effective ways of explaining the mental health risks of cannabis use to young people. In addition to a possible increased risk of psychosis, young people also need to be informed about the risks of developing dependence on cannabis, impairing their educational attainment, and possibly increasing their risk of depression.[21]

School-based drug education programmes produce small, statistically significant reductions in cannabis use but their primary effect is on knowledge rather than behaviour change and most reductions in use occur among less frequent users rather than the heavier users who are at greater risk of adverse mental health effects.[25,26]

Summary

The major adverse acute psychological effects of cannabis use are as follows:

- Anxiety, dysphoria, panic, and paranoia, especially in naive users
- Impairment of attention, memory, and psychomotor performance while intoxicated
- An increased risk of accident if an intoxicated person attempts to drive a vehicle.

The major psychological effects of daily heavy cannabis use over many years remain contested but probably include the following:[3]

- A cannabis-dependence syndrome
- Subtle forms of cognitive impairment that affect attention and memory and which persist while the user remains chronically intoxicated
- Impaired educational achievement in adolescents with a history of poor school performance, whose achievement may be limited

by the cognitive impairments produced by chronic intoxication with cannabis

- Among those who initiate cannabis use in the early teens, a higher risk of progressing to heavy cannabis and other illicit drug use, and becoming dependent on cannabis.

Further information

Some useful books

Castle, D.J. and Murray, R.M. (2004). *Marijuana and madness*. Cambridge University Press, Cambridge.

Hall, W.D. and Pacula, R.L. (2003). *Cannabis use and dependence: public health and public policy*. Cambridge University Press, Cambridge.

Iversen, L. (2000). *The science of marijuana*. Oxford University Press, Oxford.

Kandel, D.B. (2002). *Stages and pathways of drug involvement: examining the gateway hypothesis*. Cambridge University Press, New York.

Roffman, R.A. and Stephens, R.S. (2006). *Cannabis dependence: its nature, consequences and treatment*. Cambridge University Press, Cambridge.

Solowij, N. (1998). *Cannabis and cognitive functioning*. Cambridge University Press, Cambridge.

Some useful websites

Commonwealth of Australia Department of Health and Ageing. http://www.health.gov.au/internet/wcms/publishing.nsf/Content/health-pubhlth-publicat-mono.htm (A monograph reviewing the health effects of cannabis that is periodically updated)

European Monitoring Centre for Drugs and Drug Addiction. http://www.emcdda.europa.eu/ (Regularly reports data on patterns of cannabis use and cannabis related harm in Europe including treatment seeking)

National Drug and Alcohol Research Centre (Australia). http://ndarc.med.unsw.edu.au/ (This site includes research and resources on the health effects of cannabis and the treatment of cannabis dependence)

National Institute on Drug Abuse (USA). http://www.nida.nih.gov/ (Provides regular research updates on the effects of cannabis and the treatment of cannabis dependence)

Substance Abuse and Mental Health Services Administration (USA). http://www.samhsa.gov/(Provides regular updates of USA survey data on patterns of cannabis use)

Trimbos-Instituut/Netherlands Institute of Mental health and Addiction. http://www.trimbos.nl/ (A useful source for research on cannabis use and dependence in the Netherlands)

United Nations Office on Drugs and Crime (Vienna). World drug report. http://www.unodc.org/pdf/WDR_2006/wdr2006_volume1.pdf. (A very useful overview of global trends in cannabis use and the global cannabis market)

World Health Organization. http://www.who.int/substance_abuse/publications/psychoactives/en/index.html (This contains authoritative reviews of the health effects of cannabis that are periodically updated)

References

1. Adams, I.B. and Martin, B.R. (1996). Cannabis: pharmacology and toxicology in animals and humans. *Addiction*, **91**, 1585–614.
2. Hall, W.D., Degenhardt, L., and Lynskey. M.T. (2001). *The health and psychological effects of cannabis use*. Commonwealth Department of Health and Aged Care, Canberra.
3. Hall, W.D. and Pacula, R.L. (2003). *Cannabis use and dependence: public health and public policy*. Cambridge University Press, Cambridge.
4. Chen, K. and Kandel, D.B. (1995). The natural history of drug use from adolescence to the mid-thirties in a general population sample. *American Journal of Public Health*, **85**, 41–7.
5. Kandel, D.B. and Davies, M. (1992). Progression to regular marijuana involvement: phenomenology and risk factors for near-daily use. In *Vulnerability to drug abuse* (ed. M.D. Glantz), pp. 211–53. American Psychological Association, Washington, DC.
6. Ramaekers, J.G., Berghaus, G., van Laar, M., *et al.* (2004). Dose related risk of motor vehicle crashes after cannabis use. *Drug and Alcohol Dependence*, **73**, 109–19.
7. Lichtman, A.H. and Martin, B.R. (2006). Understanding the pharmacology and physiology of cannabis dependence. In *Cannabis dependence: its nature, consequences and treatment* (eds. R.A. Roffman and R.S. Stephens), pp. 37–57. Cambridge University Press, Cambridge, New York.
8. Babor, T. (2006). The diagnosis of cannabis dependence. In *Cannabis dependence: its nature, consequences and treatment* (eds. R.A. Roffman and R.S. Stephens), pp. 21–36. Cambridge University Press, Cambridge, New York.
9. Anthony, J.C., Warner, L., and Kessler, R. (1994). Comparative epidemiology of dependence on tobacco, alcohol, controlled substances and inhalants: basic findings from the National Comorbidity Survey. *Experimental and Clinical Psychopharmacology*, **2**, 244–68.
10. Roffman, R.A. and Stephens, R.S. (2006). *Cannabis dependence: its nature, consequences and treatment*. Cambridge University Press, Cambridge, New York.
11. Hall, W.D. and Degenhardt, L. (2004). Is there a specific cannabis psychosis? In *Marijuana and madness* (eds. D.J. Castle and R.M. Murray), pp. 89–100. Cambridge University Press, Cambridge.
12. Hall, W.D. and Degenhardt, L. (2006). What are the policy implications of the evidence on cannabis and psychosis? *Canadian Journal of Psychiatry. La Revue Canadienne de Psychiatrie*, **51**, 15–23.
13. Arseneault, L., Cannon, M., Witton, J., *et al.* (2004). Causal association between cannabis and psychosis: examination of the evidence. *The British Journal of Psychiatry*, **184**, 110–7.
14. Caspi, A., Moffitt, T.E., Cannon, M., *et al.* (2005). Moderation of the effect of adolescent-onset cannabis use on adult psychosis by a functional polymorphism in the catechol-*O*-methyltransferase gene: longitudinal evidence of a gene X environment interaction. *Biological Psychiatry*, **57**, 1117–27.
15. Gottesman, I.I. (1991). *Schizophrenia genesis: the origins of madness*. W.H. Freeman, New York.
16. Jeffery, D.P., Ley, A., and McLaren, S.N.S. (2004). Psychosocial treatment programmes for people with both severe mental illness and substance misuse. *The Cochrane Database of Systematic Reviews*, 2000, CD001088.
17. Solowij, N. (1998). *Cannabis and cognitive functioning*. Cambridge University Press, Cambridge.
18. Pope, H.G. Jr, Gruber, A.J., Hudson. J.I., *et al.* (2003). Early-onset cannabis use and cognitive deficits: what is the nature of the association? *Drug and Alcohol Dependence*, **69**, 303–10.
19. Cohen, S. (1982). Cannabis effects upon adolescent motivation. *Marijuana and youth: clinical observations on motivation and learning*. National Institute on Drug Abuse, Rockville, MD.
20. Edwards, G. (1983). Psychopathology of a drug experience. *The British Journal of Psychiatry*, **143**, 139–42.
21. Hall, W.D. (2006). Cannabis use and the mental health of young people. *Australian and New Zealand Journal of Psychiatry*, **40**, 105–13.
22. Macleod, J., Oakes, R., Copello, A., *et al.* (2004). Psychological and social sequelae of cannabis and other illicit drug use by young people: a systematic review of longitudinal, general population studies. *Lancet*, **363**, 1579–88.
23. Kandel, D.B. (2002). *Stages and pathways of drug involvement: examining the gateway hypothesis*. Cambridge University Press, New York.

24. Hall, W.D. and Lynskey, M. (2005). Is cannabis a gateway drug? Testing hypotheses about the relationship between cannabis use and the use of other illicit drugs. *Drug and Alcohol Review*, **24**, 39–48.
25. White, D. and Pitts, M. (1998). Educating young people about drugs: a systematic review. *Addiction*, **93**, 1475–87.
26. Tobler, N.S., Lessard, T., Marshall, D., *et al.* (1999). Effectiveness of school-based drug prevention programs for marijuana use. *School Psychology International*, **20**, 105–37

4.2.3.8 Nicotine dependence and treatment

Mª Inés López-Ibor

Introduction

Tobacco use is the single most important preventable health risk in the developed world, and an important cause of premature death worldwide.(1) Smoking causes a wide range of diseases, including many types of cancer, chronic obstructive pulmonary disease, coronary heart disease, stroke, peripheral vascular disease, and peptic ulcer disease.(2) According to the World Health Organization (WHO) smoking prevalence is estimated at around 28.6 per cent (40 per cent among males and 18.2 per cent among females),(3) it is therefore the most prevalent form of drug dependence in the world.

Tobacco causes around 13 500 deaths per day, currently, approximately 5 million people are killed annually by tobacco use. In the last few years the standardized death rate for lung cancer among men across the European region has fallen but it has increased in women.(3) By 2030, estimates based on current trends indicate that this number will increase to 10 million, with 70 per cent of deaths occurring in low- and middle-income countries.(4)

Tobacco remains the leading contributor to the disease burden in the majority of the developed countries. According to the WHO tobacco-related health care costs between 1 per cent and 1.1 per cent of Gross Domestic Product (GDP) in many countries.(5,6)

Despite the considerable efforts made to fight smoking in the last few decades, there are still substantial number of people who, in full knowledge of the health hazards, begin smoking or continue smoking. Since 2002 many countries have implemented smoke-free policies, strengthening product regulation, restrictions on smoking in public places and in work places, which for the first time are extended to bars and restaurants.(7)

Traditionally, experimentation with and the initiation of the smoking habit were related to issues such as rebellious adolescent behaviour, a need to affirm maturity, challenging authority, imitating idols, peer group pressure (from friends or relatives who are smokers) and associating smoking with being successful from a professional, financial, or sexual point of view. More recently, other perspectives, such as the specific personality pattern typified by the search for challenges (sensation seeking behaviour) and the characteristics of neuropsychological development, also began to be considered.(8)An other important issue is that smokers are much more likely than non-smokers to use or even to abuse other psychoactive drugs. Over 90 per cent of alcoholic persons smoke, drink more coffee, or take other drugs like cannabis, cocaine, or amphetamines. The reasons that smokers have such difficulty in ceasing smoking are probably similar to those of organic dependence.(9–11)

Cigarette smoking and nicotine dependence

Approximately one-third of those individuals who experiment with cigarettes become regular smokers.(12) Once dependence develops, tobacco addiction can become a chronic relapsing disorder with direct and serious medical consequences.(13)

Nicotine dependence explains why approximately 70 per cent of smokers who want to quit smoking do not succeed. Of these, approximately one-third succeed for only 1 day and less than 10 per cent remain abstinent for 12 months.(13) The definitive cessation of smoking generally occurs only after various attempts, and the relapse rate is very high, 88 per cent.(14) The percentage of smokers in which relapse occurs is similar in almost all social classes, even when including individuals, such as health care professionals, who are more informed about tobacco-related diseases.

On the other hand, only a portion of smokers develop such dependence. Why is it that not all smokers follow the same course? This question about smoking relates to the wider one of why only some people exposed to drugs become addicted to them.(15)

Why is nicotine so addictive?

The psychoactive component of tobacco is nicotine, which has its central nervous system effects by acting as agonist at the nicotine subtype of acetylcholine receptors. About 25 per cent of the nicotine inhaled when smoking a cigarette reaches the blood, and reaches the brain in about 15 s. The half-life of nicotine is about 2 h.(13)

Nicotine is believed to have positive reinforcing and addictive properties because it activates the dopaminergic pathway projecting from the ventral tegmental area to the cerebral cortex and the limbic system, the system that is affected by cocaine and amphetamine. In addition to activating the reward system, nicotine causes an increase in the concentrations of circulating norepinephrine and epinephrine, and increased release of vasopressin, β-endorphin, adrenocorticothropine, and cortisol.(16)

The development of dependence is enhanced by strong social factors that encourage smoking in some settings.(17)

Other central effects

The stimulatory effects of nicotine result in improved attention, learning, reaction time, and problem-solving ability. Nicotine also decreases psychological tension and lessens depressive feelings.

The effects of nicotine in the cerebral blood flow (CBF) have been studied and results suggest that short-term nicotine exposure increases the CBF without changing cerebral oxygen metabolism but that long-term nicotine exposure is associated with decrease in the CBF.(18)

Genetic issues

Epidemiological studies have shown that the genetic component can play a significant role in the smoking habit, being responsible for 40 per cent to 60 per cent of the variability in the risk of addiction.(19,20) The first studies relating genetics to smoking date from 1958(21) when it was suggested that there were genes that, in

youth, predispose individuals to become smokers and, later, to present with lung cancer.[21]

Many twin concordance studies have indicated that genetic inheritance plays a role in smoking addiction.[22] Such studies have demonstrated a higher concordance rate in relation to smoking among monozygotic twins than among dizygotic twins, whether raised together or separately.[23,24] More recent studies with larger study samples, a better classification of phenotypes, and more sophisticated statistical models, point to a rather significant influence of the genome in determining the smoking phenotype.[25]

The most extensively studied genes of the dopaminergic pathway are those that regulate the flow of dopamine in the central nervous system. Five different dopamine receptors are known, and the genes that encode them have been cloned (DRD1, DRD2, DRD3, DRD4, and DRD5). Among those, the DRD2 receptor has been studied most widely, because of its association with other addictive behaviours, and because nicotine has a dopamine-releasing effect.[26–28]

Smoking in psychiatric patients

Nicotine has been said to provide anxiety relief, oral gratification, and self-medication of psychotic symptoms in psychiatric patients.[29] Patients with schizophrenia and severe mental illness smoke cigarettes at rates that well exceed the general population.[30] Little is known about the correlates and sequels of increased smoking severity on persons with severe mental illness. Greater smoking severity has however been associated with greater perceived stress, poorer overall subjective quality of life, and lower satisfaction with finances, health, leisure activities, and social relationships, results that may lend support to a self-medication hypothesis.[31]

Nicotine intoxication and withdrawal symptoms

The primary addictive substance in cigarette smoking is nicotine. Cigarette smoking is very efficient nicotine delivery system because nicotine is nebulized and subsequently absorbed through the extensive pulmonary lung vessels. Consequently, smoking produces high arterial nicotine concentrations.[12] These high arterial concentrations, higher than venous concentrations, deliver a bolus of 1–3 mg of nicotine rapidly to the brain, a few seconds after smoking. With nicotine receptor activation, neurotransmitters are released including dopamine, norepinephrine, serotonin, and endogenous opioids. The immediate positive reinforcing effects of smoking include a reduction in anxiety and increased alertness and concentration.[32]

Nicotine's half-life is 2 h; therefore, repeated administration is needed through the day to maintain its effects. Consequently smokers usually smoke at frequent intervals to maintain narrow range of nicotine concentration in the blood. Paradoxically, chronic administration of nicotine results in an increase in the number of nicotine receptors. This paradoxical effect is probably due to a chronic nicotine receptor desensitization and inactivation. An increased number of receptors may play a role in the withdrawal symptoms that many smokers experience with prolonged cigarette abstinence.[33]

Nicotine is a highly toxic chemical; doses of 60 mg in adults are fatal secondary to respiratory paralysis (an average cigarette has an average dose of 0.5 mg). At low doses, symptoms of toxicity are nausea, vomiting, salivation, pallor due to peripheral vasoconstriction, weakness, abdominal pain, diarrhoea, dizziness, headache, increased blood pressure, tachycardia, tremor, and cold sweats. Toxicity is also associated with inability to concentrate, confusion, and sensory disturbances. During pregnancy, smoking is associated with an increased incidence of low-weight-birth babies.[32]

DSM-IV does not have a diagnostic category for nicotine intoxication; however, it has a category for nicotine withdrawal.[34] ICD-10 does have a category for nicotine intoxication (F17.0, mental and behavioural disorders due to use of tobacco, acute intoxication).[35] Withdrawal symptoms can develop within 2 h after having smoked the last cigarette. Withdrawal symptoms include dysphoria and depressed mood, insomnia, irritability, anxiety, frustration, difficulty in concentration and increase in appetite, and weight gain. Withdrawal symptoms peak within 24–36 h after cessation and usually diminish after 1 week of abstinence but can last for much longer. Some individuals continue smoking to avoid the negative symptoms of withdrawal.[36,37]

Nicotine dependence

The cumulative findings of more than 2500 scientific papers were summarized in the *1988 Surgeons General's Report on the Health Consequences of Smoking: Nicotine Addiction*.[38] Nicotine has a pronounced effect on the major stress hormones, and the dose-related effects of nicotine on neuroendocrine responses appear to constitute a critical component of its pharmacological action. Hypothalamic corticotrophin-releasing factor (CRF) is stimulated by nicotine, and levels of hypophyseal hormones including ACTH and arginine-vasopressin are increased in a dose-related manner. At higher doses, growth hormone (GH) and prolactin are entrained, and corticosteroid levels are related to plasma nicotine levels.[15,39,40]

Treatment for tobacco dependence

Many of the adverse health effects of smoking are reversible, and smoking cessation treatments represent some of the most cost-effective of all health care interventions. Although the greatest benefit accrues from ceasing smoking when young, even quitting in middle age avoids much of the excess health care risk associated with smoking.[41–43]

In order to improve smoking cessation rates, effective behavioural and pharmacological treatments, coupled with professional counselling and advice, are required. Since smoking duration is the principal risk factor for smoking-related morbidity, the treatment goal should be early cessation and prevention of relapse.[44,45]

The health benefits are various: smoking cessation has a major and immediate health benefit for persons with or without smoking-related diseases; former smokers live longer than those who continue to smoke; smoking cessation decreases the risk of lung cancer, myocardial infarction, cerebrovascular diseases, and chronic lung diseases.[46]

Non-pharmacological treatment

Physician counselling and pharmacotherapeutic interventions for smoking cessation are among the most cost-effective clinical interventions. Several strategies can be used for smoking cessation, counselling differs according to the patient's readiness to quit. For smokers who do not intend to quit smoking, physicians should

inform and sensitize them about tobacco use and cessation. For smokers ready to quit, the physician should show strong support, help set a quit date, prescribe pharmaceutical therapies for nicotine dependence, if needed.[47,48]

There is insufficient information to know which elements of behavioural support are effective or whether one approach, such as motivational interviewing or cognitive behavioural therapy, is more effective than another. There is some evidence to suggest that group support may be more effective in general than one-to-one support[49] and that it should involve multiple sessions.[50] There is also evidence that such sessions can be effective even if conducted over the telephone.[51]

Of the many web-based support packages available on the Internet, only two have been evaluated in randomized controlled trials,[52,53] these trials of tailored programmes enrolled smokers who were using nicotine replacement. Both trials showed significant benefits 10–12 weeks after the quit date.

There are a few adequate studies examining complementary therapies in smoking cessation. Meta-analysis of trials of acupuncture and hypnotherapy showed no benefit but could not exclude small effects.[54,55]

Nicotine replacement therapies

Nicotine replacement therapies (NRTs) were designed in order to enhance efficacy rates during smoking cessation by replacing some of the nicotine usually delivered by smoking. All replacement therapies have shown to have high rates of efficacy. The choice of NRT depends on patient's preference, side effects, presence of concomitant medical conditions, and a history of previous success or failure.[56,57]

(a) Nicotine gum and nicotine lozenge

Nicotine gum was the first NRT marketed for smoking cessation, contains nicotine bound to an iron resin. The nicotine is slowly released into the mouth and is absorbed through the mucosa of the mouth, only 50 per cent of the nicotine in a piece of gum is systemically absorbed and concentrations reach a maximum peak in 30 min after onset of chewing.[12] The staring dose for individuals who smoke 20 cigarettes a day should be 2 mg. The gum should not be prescribed in patients with temporo-mandibular joint disease and those with dental or oral problems. The nicotine lozenge contains nicotine bound to a prolacrilex ion-exchange resin, does not require chewing and therefore is preferable to the gum for patients with dental problems.[58]

(b) Transdermal nicotine

Transdermal nicotine, commonly known as the nicotine patch, produces a constant delivery of nicotine that is very useful for patients with poor treatment adherence. Eight weeks of treatment are generally sufficient for smoking cessation.[59]

Transdermal nicotine is available in a variety of formulations and dosing schedules (i.e. 15 mg/16 h; 7, 14, 21 mg/24 h). Peak nicotine concentrations for the various systems are reached 2–6 h after application and steady state conditions occur 2–3 days after continued patch use.[59]

(c) Nicotine nasal spray and inhaler

Nicotine nasal spray delivers nicotine through the nasal mucosa. One advantage is that it relieves tobacco cravings quickly. It is available only by prescription. One spray to each nostril constitutes a dose, approximately 1 mg nicotine. Patients should use one or two doses per hour; the nasal spray delivers nicotine rapidly, with venous nicotine peaking at 5–10 min after administration.[60] Side effects include some initial irritation of the nasal mucosa should be avoided in patients with rhinitis, nasal polyps, or sinusitis.

Nicotine vapour inhaler is used by puffing through a cartridge inhaler, and may be useful for smoking cessation in some patients because its use is similar to the smoking ritual, and it delivers nicotine rapidly. It is only available by prescription. The recommended treatment period is up to 24 weeks.[45]

Pharmacological treatments

Considering that not all smokers respond well to nicotine replacement therapies and some smokers have comorbid symptoms there has been considerable interest in non-nicotine medications to treat nicotine dependence. The observations that some antidepressant-like bupropion and other selective serotonin-reuptake inhibitors (SSRIs) are useful as a treatment for smoking cessation, led to intensive research to study dopamine, serotonin, norepinephrine, glutamate, gamma-aminobutyric acid (GABA), nicotine, cannabinoid, and opiod receptors.[61,62]

Sustained-released bupropion

The sustained-released bupropion is currently considered as a first-line treatment for cigarette smokers. The mechanism of action of this antidepressant in the treatment of nicotine dependence likely involves blockade of dopamine and norepinephrine reuptake as well as antagonism of high-affinity nicotine acetylcholine receptors.[63]

The goals of bupropion therapy for nicotine dependence are (1) cessation of smoking behaviour and (2) reduction of nicotine withdrawal symptoms. In addition, bupropion SR may delay cessation-induced weight gain.[64]

A study by Hurt *et al.*[65] established the efficacy and safety of bupropion SR for treatment of nicotine dependence, and led to its approval for this indication by the FDA in 1998. Bupropion treatment also reduces weight gain associated with smoking cessation and significantly reduced nicotine withdrawal symptoms at a dose of 150–300 mg/day. Major side effects are headache, dry mouth, nausea, vomiting, and insomnia.[65]

Nortrityline

This tricycle antidepressant appears to have efficacy rates similar to bupropion in smoking cessation.[66] The mechanism of action is thought to be related to its noradrenergic and serotoninergic reuptake blockade. Side effects are those of the typical tricyclic antidepressants and include dry mouth, blurred vision, constipation, and orthostatic hypotension. Nortryptiline has been recommended as a second-line treatment.[61,67]

Clonidine

Because clonidine appears to have some efficacy for alcohol and opioid withdrawal, it has been evaluated for the treatment of nicotine withdrawal. However, it has not proved to be as effective as other therapies. Several clinical trials used oral or transdermal clonidine in doses of 0.1–0.4 mg/day for 2–6 weeks with or without behaviour therapy. Most common side effects of clonidine are dry mouth, sedation, and constipation, postural hypotension, and depression.[61]

Selective serotonin reuptake inhibitors

The available evidence provides little support for the use of SSRIs to assist in smoking cessation, either alone or in combination with other therapies. Placebo-controlled trials of fluoxetine or paroxetine failed to show an increase in smoking cessation.[61]

Varenicline

Varenicline is a selective alpha (4) beta (2) nicotinic acetylcholine receptor partial agonist and the first non-nicotine-containing medication developed with the sole purpose of treating nicotine addiction.[68]

Varenicline seems to be more efficacious than bupropion 24 weeks after randomization to a 12-week treatment course and 1 year after randomization in an identical trial. There are no contraindications except hypersensitivity and the drug is generally well tolerated. Varenicline, which is recently approved for smoking cessation, offers an option to patients who cannot tolerate the adverse effects associated with nicotine-replacement therapy and bupropion. It is also an alternative to consider for patients with contraindications to such therapies. Varenicline is completely absorbed orally and not affected by food. Steady state is reached within 4 days of administration.[69,70]

Other pharmacological treatments

Several new therapies are emerging as possible treatment options for smoking cessation. Rimonabant, a selective cannabinoid antagonist, blocks dopamine release in the nucleus accumbens, a primary reward centre for the brain. Studies have found that Rimonabant may not only be effective as a smoking cessation aid but may also assist in the maintenance of nicotine abstinence. Rimonabant has also demonstrated a weight-loss benefit, which may be attractive to smokers concerned with weight gain associated with smoking cessation.[71–73]

Three nicotine vaccines are currently in development, each acting to sequester nicotine from the bloodstream, thereby preventing its penetration of the central nervous system. Ongoing studies will evaluate their use as established therapies for smoking cessation.[74,75]

Conclusions

Despite the reality that smoking remains the most important preventable cause of death and disability, most clinicians underperform in helping smokers quit. Nearly 70 per cent of smokers want to quit, and 42.5 per cent attempt to quit each year. The most effective smoking cessation programmes involve a combination of pharmacotherapy and behavioural and/or cognitive counselling to improve abstinence rates. Ways to counter clinicians' pessimism about cessation include the knowledge that most smokers require multiple attempts before they succeed in quitting.

Further information

Fiore, M.C., Bailey, W.C., Cohen, S.J., *et al.* (2000). *Treating tobacco use and dependence. Quick reference guide for clinicians*. U.S. Department of Health and Human Services, Rockville, MD. Public Health Service. (http://www.surgeongeneral.gov/tobacco/tobaqrg.htm)

www.tobaccofreekids.org Information on Youth Smoking

www.findhelp.com Foundation for Innovations in Nicotine Dependence

www.quitnow.org Self-help Website Support for Quitting Smoking

References

1. Peto, R., López, A.D., Boreman, J., *et al.* (1996). Mortality from smoking world wide. *British Medical Bulletin*, **52**, 12–21.
2. Collihaw, N.E. and López, A.D. (1996). *The tobacco epidemic: a public health emergency*. World Health Organization (WHO) Tobacco Alert. WHO, Geneva.
3. Esson, K.M. and Leeder, S.R. (2004). *The millennium developments goals and tobacco control*. World Health Organization Publications, Geneva.
4. Thun, M.J., Day-Lacally, C.A., Calle, E.E., *et al.* (1995). Excess mortality among cigarette smokers: changes in a 20 years interval. *American Journal of Public Health*, **85**, 1223–30.
5. Mathers, C.D. and Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine*, **3**(11), e512.
6. Jha, P. and Chaloupka, F.J. (2000). *Tobacco control in developing countries*. Oxford University Press, Oxford, UK.
7. Davis, K.C., Nonnemaker, J.M., and Farrelly, M.C. (2007). Association between national smoking prevention campaigns and perceived smoking prevalence among youth in the United States. *The Journal of Adolescent Health*, **41**(5), 430–6. Epub 4 September 2007.
8. Fagerström, K. (2002). The epidemiology of smoking: health consequences and benefits of cessation. *Drugs*, **62**(Suppl. 2), 1–9.
9. Choquet, M., Morin, D., Hassler, C., *et al.* (2004). Are alcohol, tobacco, and cannabis use as well as polydrug use increasing in France? *Addictive Behaviours*, **29**, 607–14.
10. Riala, H., Hakko, M., Isohanni, M.R., *et al.* (2004). Teenage smoking and substance use as predictors of severe alcohol problems in late adolescence and in young adulthood. *The Journal of Adolescent Health*, **35**, 245–54.
11. Farrell, M., Howes, S., Bebbington, P., *et al.* (2001). Nicotine, alcohol and drug dependence and psychiatric comorbidity. Results of a national household survey. *The British Journal of Psychiatry*, **179**, 432–7.
12. Henningfield, J.E. (1995). Nicotine medications for smoking cessation. *The New England Journal of Medicine*, **333**, 1196–203.
13. Scelling, T.C. (1992). Addictive drugs: the cigarette experience. *Science*, **225**, 430.
14. Yudkin, P., Hey, K., Roberts, S., *et al.* (2003). Abstinence from smoking eight years after participation in randomized controlled trial of nicotine patch. *British Medical Journal*, **327**(7405), 28–9.
15. Volkow, N.D. (2005). What do we know about drug addiction? *The American Journal of Psychiatry*, **162**(8), 1401–2.
16. Brody, A.L., Mandelkern, M.A., London, E.D., *et al.* (2006). Cigarette smoking saturates brain alpha 4 beta 2 nicotinic acetylcholine receptors. *Archives of General Psychiatry*, **63**(8), 907–15.
17. Mamede, M., Ishizu, K., Ueda, M., *et al.* (2007). Temporal change in human nicotinic acetylcholine receptor after smoking cessation: 5IA SPECT study. *Journal of Nuclear Medicine*, EPub 17 October 2007.
18. Heath, A.C., Kirk, K.M., Meyer, J.M., *et al.* (1999). Genetic and social determinants of initiation and age at onset of smoking in Australian twins. *Behavior Genetics*, **29**(6), 395–407.
19. Sullivan, P.F. and Kendler, K.S. (1999). The genetic epidemiology of smoking. *Nicotine & Tobacco Research*, **1**(Suppl. 2), S51–7; discussion S69–70.
20. Batra, V., Patkar, A.A., Berrettini, W.H., *et al.* (2003). The genetic determinants of smoking. *Chest*, **123**(5), 1730–9.
21. Fisher, R.A. (1958). Cancer and smoking. *Nature*, **182**(4635), 596.
22. Berrettini, W.H. and Lerman, C.E. (2005). Pharmacotherapy and pharmacogenetics of nicotine dependence. *The American Journal Psychiatry*, **162**(8), 1441–51.
23. Tapper, A.R., McKinney, S.L., and Nashmi, R. (2004). Nicotine activation of alpha4 receptors: sufficient for reward, tolerance, and sensitization. *Science*, **306**, 1029–32.
24. True, W.R., Heath, A.C., Scherrer, J.F., *et al.* (1997). Genetic and environmental contributions to smoking. *Addiction*, **92**(10), 1277–87.

25. Munafo, M., Clark, T., Johnstone, E., *et al.* (2004). The genetic basis for smoking behavior: a systematic review and meta-analysis. *Nicotine & Tobacco Research*, **6**(4), 583–97.
26. Swan, G.E., Benowitz, N.L., Jacob, P. III, *et al.* (2004). Pharmacogenetics of nicotine metabolism in twins: methods and procedures. *Twin Research*, **7**(5), 435–48.
27. do Prado-Lima, P.A., Chatkin, J.M., Taufer, M., *et al.* (2004). Polymorphism of 5HT2A serotonin receptor gene is implicated in smoking addiction. *American Journal of Medical Genetics. Part B Neuropsychiatric Genetics*, **128**(1), 90–3.
28. Arinami, T., Ishiguro, H., and Onaivi, E.S. (2000). Polymorphisms in genes involved in neurotransmission in relation to smoking. *European Journal of Pharmacology*, **410**(2–3), 215–26.
29. Medoff, D.R., Wohlheiter, K., DiClemente, C., *et al.* (2007). Correlates of severity of smoking among persons with severe mental illness. *American Journal of Addiction*, **16**(2), 101–10.
30. Strand, J.E. and Nybäck, H. (2005). Tobacco use in schizophrenia: a study of cotinine concentrations in the saliva of patients and controls. *European Psychiatry*, **20**(1), 50–4.
31. Gershon, R.B., Hwang, S., Han, J., *et al.* (2007). Short-term naturalistic treatment outcomes in cigarette smokers with substance abuse and/or mental illness. *The Journal of Clinical Psychiatry*, **68**(6), 892–8; quiz 980–1.
32. Di Chiara, G. (2000). Role of dopamine in the behavioural actions of nicotine related to addiction. *European Journal of Pharmacology*, **393**, 295–314.
33. Dani, J.A. and De Biasi, M. (2001). Cellular mechanism of nicotine addiction. *Pharmacology, Biochemistry, and Behavior*, **70**, 439–66.
34. American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th edn, revised). American Psychiatric Association, Washington, DC.
35. WHO. (1992). *The ICD-10 classification of mental and behavioural disorders*, pp. 70–4. World Health Organization, Geneva.
36. Hughes, J.R. and Hatsukami, D. (1986). Signs and symptoms of tobacco withdrawal. *Archives of General Psychiatry*, **43**, 289–94.
37. Heatherton, T.F., Kozlowski, L.T., Frecker, R.C., *et al.* (1991). The Fagerström test for nicotine dependence: a revision of the Fagerström tolerance questionnaire. *British Journal of Addiction*, **86**, 1119–27.
38. Jarvis, M. (2001). Patterns of use, epidemiology, adverse effects and specific issues concerning treatment for nicotine dependence. In *Oxford textbook of psychiatry* (eds. N. Andreasen, M. Gelder, and J.J. López-Ibor), pp. 554–9.
39. Maskos, U., Molles, B.E., and Pons, S. (2005). Nicotine reinforcement and cognition restored by targeted expression of nicotinic receptors. *Nature*, **436**, 103–7.
40. Kremer, I., Bachner-Melman, R., Reshef, A., *et al.* (2005). Association of the serotonin transporter gene with smoking behavior. *The American Journal of Psychiatry*, **162**(5), 924–3028.
41. Fiore, M.C., Bailey, W.C., and Cohen, S.J. (2000). *Treating tobacco use and dependence. Quick reference guide for clinicians*. US Dept of Health and Human Services, Public Health Service, Rockville, MD.
42. Schroeder, S.A. (2005). What to do with a patient who smokes. *The Journal of the American Medical Association*, **294**(4), 482–7.
43. Law, M. and Tang, J.L. (1995). An analysis of the effectiveness of interventions intended to help people to stop smoking. *Archives of Internal Medicine*, **155**, 1993–41.
44. Lerman, C., Patterson, F., and Berrettini, W. (2005). Treating tobacco dependence: state of the science and new directions. *Journal of Clinical Oncology*, **23**(2), 311–23.
45. Ocken, C.A. and George, T.P. (2005). Tobacco. In *Clinical manual of addiction psychopharmacology*, American Psychiatric Publishing, Arlington, pp. 315–38.
46. Chatkin, J.M. (2006). The influence of genetics on nicotine dependence and the role of pharmacogenetics in treating the smoking habit. *Journal of Brasileiro de Pneumologia*, **32**(6), 573–9.
47. [No authors listed] (2000). A clinical practice guideline for treating tobacco use and dependence: a US public health service report. The tobacco use and dependence clinical practice guideline panel, staff, and consortium representatives. *The Journal of the American Medical Association*, **283**(24), 3244–54.
48. Cornuz, J. (2007). Smoking cessation interventions in clinical practice. *European Journal of Vascular Endovascular Surgery*, **34**(4), 397–404. Epub 1 August 2007.
49. McEwen, A., West, R., and McRobbie, H. (2006). Effectiveness of specialist group treatment for smoking cessation vs. one-to-one treatment in primary care. *Addictive Behaviors*, **31**, 1650–60.
50. Aveyard, P., Brown, K., Saunders, C., *et al.* (2007). Weekly versus basic smoking cessation support in primary care: a randomised controlled trial. *Thorax*, **62**(10), 892–903.
51. Stead, L.F., Perera, R., and Lancaster, T. (2006). Telephone counselling for smoking cessation. *Cochrane Database of Systematic Reviews*, (3), CD002850.
52. Strecher, V.J., Shiffman, S., and West, R. (2005). Randomized controlled trial of a web-based computer-tailored smoking cessation program as a supplement to nicotine patch therapy. *Addiction*, **100**, 682–8.
53. Swartz, L.H., Noell, J.W., Schroeder, S.W., *et al.* (2006). A randomised control study of a fully automated internet based smoking cessation programme. *Tobacco Control*, **15**, 7–12.
54. White, A.R., Rampes, H., and Campbell, J.L. (2006). Acupuncture and related interventions for smoking cessation. *Cochrane Database of Systematic Reviews*, (1), CD000009.
55. Abbot, N.C., Stead, L.F., White, A.R., *et al.* (1998). Hypnotherapy for smoking cessation. *Cochrane Database of Systematic Reviews*, (2), CD001008.
56. George, T.P. and O'Malley, S.S. (2004). Current pharmacological treatment for nicotine dependence. *Trends in Pharmacological Sciences*, **25**, 42–48.
57. Perkins, K.A., Stitzer, M., and Lerman, C. (2006). Medication screening for smoking cessation: a proposal for new methodologies. *Psychopharmacology* (*Berl*), **184**(3–4), 628–36.
58. Shiffamn, S., Dresler, C.M., and Hajek, P. (2002). Efficacy of a nicotine lozenge for smoking cessation. *Archives of Internal Medicine*, **162**, 1267–76.
59. Fiore, M.C., Smith, S.S., and Jorenby, D.E. (1994). The effectiveness of the nicotine patch for smoking cessation. *The Journal of the American Medical Association*, **271**, 1940–8.
60. Shuterland, G., Sappleton, J.A., and Russell, M.A. (1992). Randomised controlled trial of nasal nicotine spray in smoking cessation. *Lancet*, **340**, 324–9.
61. Sullivan, M.A. and Covey, L.S. (2002). Nicotine dependence: the role for antidepressants and anxiolytics. *Current Opinion in Investigational Drugs*, **3**(2), 262–71.
62. Slemmer, J.E., Martin, B.R., and Damaj, M.I. (2000). Bupropion is a nicotinic antagonist. *The Journal of Pharmacology and Experimental Therapeutics*, **295**, 321–7.
63. Lerman, C., Shields, P.G., Wileyto, E.P., *et al.* (2003). Effects of dopamine transporter and receptor polymorphisms on smoking cessation in a bupropion clinical trial. *Health Psychology*, **22**(5), 541–8.
64. Gonzales, D.H., Nides, M.A., and Ferry, L.H. (2001). Bupropion SR as an aid to smoking cessation in smokers treated previously with bupropion: a randomized placebo-controlled study. *Clinical Pharmacology and Therapeutics*, **69**, 438–44.
65. Hurt, R.D., Sachs, D.P., and Glover. E.D. (1997). A comparison of sustained released bupropion and placebo for smoking cessation. *The New England Journal of Medicine*, **337**, 1195–202.
66. Hall, S.M., Humfleet, G.L., and Reus, V.I. (2002). Psychology intervention and antidepressant treatment in smoking cessation. *Archives of General Psychiatry*, **59**, 930–6.

67. Haggstram, F.M., Chatkin, J.M., Sussenbach-Vaz, E., *et al.* (2006). A controlled trial of nortriptyline, sustained-release bupropion and placebo for smoking cessation: preliminary results. *Pulmonary Pharmacology & Therapeutics*, **19**(3), 205–9.
68. Stack, N.M. (2007). Smoking cessation: an overview of treatment options with a focus on varenicline. *Pharmacotherapy*, **27**(11), 1550–7.
69. Potts, L.A. and Garwood, C.L. (2007). Varenicline: the newest agent for smoking cessation. *American Journal of Health-System Pharmacy*, **64**(13), 1381–4.
70. Tonstad, S., Tønnesen, P., Hajek, P., *et al.* (2006). For the varenicline phase 3 study group. Effect of maintenance therapy with varenicline on smoking cessation: a randomized controlled trial. *The Journal of the American Medical Association*, **296**, 64–71.
71. Garwood, C.L. and Potts, L.A. (2007). Emerging pharmacotherapy's for smoking cessation. *American Journal of Health-System Pharmacy*, **64**(16), 1693–8.
72. Fowler, J.S., Volkow, N.D., Wang, G.J., *et al.* (1996). Brain monoamine oxidase. An inhibition in cigarette smokers. *Proceedings of the National Academy of Sciences of the United States of America*, **93**(24), 14065–9.
73. Markou, A., Paterson, N.E., and Semenova, S. (2004). Role of gamma-aminobutyric acid (GABA) and metabolic glutamate receptors in nicotine reinforcement: potential pharmacotherapy for smoking cessation. *Annals of the New York Academy of Sciences*, **1025**, 491–503.
74. Dwoskin, L.P., Joyce, B.M., Zheng, G., *et al.* (2007). Discovery of a novel nicotinic receptor antagonist for the treatment of nicotine addiction: 1-(3-Picolinium)-12-triethylammonium-dodecane dibromide (TMPD). *Biochemical Pharmacology*, **74**(8), 1271–82. Epub 21 July 2007.
75. Oliver, J.L., Pashmi, G., Barnett, P., *et al.* (2007). Development of an anti-cotinine vaccine to potentiate nicotine-based smoking cessation strategies. *Vaccine*, **25**(42), 7354–62. Epub 30 August 2007.

4.2.4 Assessing need and organizing services for drug misuse problems

John Marsden, Colin Bradbury, and John Strang*

Introduction

In the present decade, there has been substantial investment in drug misuse treatment thereby expanding the workforce, the capacity of the treatment system and leading to reduced waiting times and better integration of local services. In 2006–07, an in-treatment population of approximately 200 000 individuals were recorded by the National Drug Treatment Monitoring System (NDTMS). Capture-recapture estimates suggest that there are approximately 327 000 users of opioids and/or crack cocaine.

About two-thirds of adults entering drug misuse treatment services are dependent on illicit heroin—a clinical presentation complicated by between 20 per cent to 50 per cent of admissions by concurrent dependence on cocaine and other substances such as the misuse of pharmaceutical medications (such as benzodiazepines). Cannabis is reported as the main problem drug for younger patients under 18 years of age. Overall, treatment services for clients of all ages are able to assess and provide interventions across all illicit drugs including amphetamine-type stimulants, sedative/hypnotics, cannabis, hallucinogens and volatile substances (solvents and inhalants). Hazardous and harmful alcohol use characterizes a significant, but priority group of drug misuse treatment seekers.

In 2006, a revised national drug misuse treatment effectiveness strategy stressed the need for better local partnerships to commission and organize local services and promote reintegration of treated patients into the community. A core component of the strategy was the creation of Criminal Justice Integrated Teams (CJITS) who were given the role of treatment case coordination for individuals involved in the justice system with identified drug misuse. Nevertheless, improvements to the reach, operation, and effectiveness of treatments remains a priority—particularly tackling high-risk behaviours linked to the acquisition and transmission of blood-borne infections and ensuring that all service users receive good quality assessment and care coordination.

Local coordination of treatment

Drug Action Teams (DATs) were originally set up in 1995 under a Government white paper on drug misuse. The purpose of the DAT was to co-ordinate the activity and spend of local statutary commissioning agencies who have an interest in reducing the harm caused by illicit drug use to individuals, their families, and the community. DATs and their membership typically consists of senior commissioning representatives from local Police, Health, Local Authority and Probation services. Increasingly, the Prison Service are represented following the announcement of a new Integrated Drug Treatment System for prisoners which seeks to ensure that the same appropriate and evidenced drug treatment interventions are available to individuals regardless of whether they are in prison or in the community. Under the Police Reform Act (2002), the process of combining the activity of DATs with their equivalent bodies for crime (Crime and Disorder Reduction Partnerships (CDRPs) was started. Local areas organize their activity in differing ways, but the concept of co-ordination of action between the DAT and CDRP is now universal.

DATs are charged with consulting with and involving local communities, stakeholders, treatment providers, and crucially—users and carers—in the development of their local commissioning strategies. DATs are allocated a hypothecated fund for improving capacity and quality of drug treatment services for their residents. This Pooled Treatment Budget is typically banked by a partner agency (usually the PCT) but is intended to be commissioned jointly (along with mainstream monies that partner agencies allocate for drug treatment) via the DAT Partnership structure. DAT Partnerships typically have a sub-group know as the Joint Commissioning Group (JCG) which seeks to operationalize the DAT's agreed strategy for the locality.

The National Treatment Agency for Substance Misuse (NTA) is a Special Health Authority set up in 2001 to oversee and performance manage the commissioning of effective drug treatment.

* The views expressed in this chapter are those of the authors and do not necessarily reflect the views of the National Treatment Agency. The commissioning, performance management and planning of drug treatment varies significantly across the United Kingdom. Unless stipulated to the contrary, the following text applies specifically to England.

DAT Partnerships submit a treatment plan on an annual basis which is signed off by the NTA and other regional partners and performance monitored on a quarterly basis. Since 2005–06, the NTA has issued guidance on conducting a Needs Assessment for the local population and increasingly assessment of need is being seen as the centrepiece of DAT Partnership commissioning activity. DAT Partnerships are encouraged to set up expert groups which (in combination with available local data sources of prevalence and treatment) should be used to carefully consider available information and intelligence in order to inform the local Joint Commissioning Group of assessed levels of unmet need, therefore enabling them to set and update commissioning priorities on a cyclical basis.

Types of treatment

In the UK, treatment for substance use disorders vary on several core dimensions, as follows: (a) setting (outpatient/community or inpatient/residential), modality (pharmacological or behavioural); (b) content (e.g. cognitive behavioural therapy, motivational approaches, contingency management; couples therapy); (c) goals (harm reduction, partial or complete abstinence); (d) intensity (brief interventions or intensive therapeutic contact); (e) extent of external contingency (e.g. self-referral or criminal justice mandate); and (f) type of provider (NHS, non-governmental organiztion and private/commercial). In 2002, the NTA promulgated a national service framework for drug misuse services and updated this four years later. The framework uses a practical framework to aid rational and evidence-based commissioning of drug treatment in England with services for drug misusers grouped into four broad bands, or tiers.

Tier 1 interventions

This first tier involves the provision of information, advice, screening and referral to drug users by generic medical and social care services (e.g. Accident and Emergency Departments, community retail pharmacies). It includes liaison and partnership working with specialist drug treatment services to provide specific interventions (e.g. treatment of patients with health problems caused by Hepatitis C infection).

Tier 1 services for adults are not structured drug or alcohol treatment, but can be part of the local substance misuse treatment system. These services work with a wide range of clients including drug and alcohol misusers, but their sole purpose is not drug or alcohol treatment. Tier 1 services comprise a range of interventions which are not drug-specific, but offer a variety of generic health and social care interventions. In this context, the role of Tier 1 includes the provision of their own services plus, as a minimum, screening drug misusers and referral to local drug and alcohol treatment services in Tiers 2 and 3. Tier 1 provision for drug and alcohol misusers may also include assessment, services to reduce drug-related harm, and liaison or joint working with Tiers 2 and 3 specialist drug and alcohol treatment services. Tier 1 services are crucial to providing services in conjunction with more specialized drug and alcohol services (e.g. general medical care for drug misusers in community-based or residential substance misuse treatment, or housing support and aftercare for drug misusers leaving residential care or prison).

Tier 2 interventions

The second tier describes interventions involving specific drug-related information and advice to help drug users reduce or avoid hazardous and harmful patterns of use or attain and maintain abstain harm. Services are delivered from dedicated community locations as well as outreach and may also include brief, structured psychosocial interventions, various harm minimization interventions (including syringe and needle exchange) and aftercare support. Tier 2 services may also provide triage assessment and linked referral to structured drug treatment and in this respect may operated independently or in the same setting as a Tier 3 intervention team. Tier 2 interventions for adults provide accessible drug and alcohol specialist services for a wide range of drug and alcohol misusers referred from a variety of sources, including self-referrals. This tier is defined by its low threshold to access services, and limited requirements on drug and alcohol misusers to receive services. Often drug and alcohol misusers will access drug or alcohol services through Tier 2 and progress to higher tiers. Tier 2 interventions include advice and information, drop-in services, needle exchange and motivational interviewing.

Tier 3 interventions

In terms of the volumes of people receiving treatment, this tier is at the centre of the system. It includes specialized care-planned pharmacotherapy (opioid agonist and antagonist and adjunctive medication prescribing to treat dependence) and a broad array of psychosocial interventions delivered by combined or separate teams in the community and primary care. There is an emphasis on high-quality assessment, care planning, liaison and review and regular contact with a clinical keyworker and other team members. The frequency of scheduled contact varies widely across Tier 3 services but is particularly indented to be intensive among users attending 'structured day programmes'. Tier 3 interventions for adults are provided solely for drug and alcohol misusers in structured programmes of care. Tier 3 structured services include psychotherapeutic interventions and structured counselling (e.g. cognitive behavioural therapy, motivational interventions), methadone maintenance programmes, community detoxification, or day care provided either as a drug- and alcohol-free programme or as an adjunct to methadone treatment. Community-based aftercare programmes for drug and alcohol misusers leaving residential rehabilitation or prison are also included in Tier 3 interventions. There is interest in developing behaviour therapies to treatment drug dependence on contingency management. Psychoactive substances can exert unconditioned reinforcing effects and repeated administration produce several conditioned responses. For example, voucher-based reinforcement therapy uses vouchers of increasing value for goods and services with various bonus incentives to subjects who can provide drug-free urine tests. The National Institute for Health and Clinical Excellence (NICE) has produced guidelines for the effective delivery of various psychosocial treatment interventions tailored to the needs of drug misusers, including brief motivational interventions, contingency management and behavioural couples therapy.

Tier 4 interventions

The fourth tier of the treatment system denotes specialist inpatient (and general ward) inpatient services providing stabilization and

medically supervised withdrawal (detoxification), residential rehabilitation programmes (providing psychosocial and practical, vocational supports designed to maintain abstinence and promote long-term recovery) and a range of halfway houses and supportive accommodation. Some inpatient and residential programmes are directly linked. These services vary in duration from brief (<10 days), short-term (<3 months) and long-term (>3 months). Tier 4 services are highly structured interventions underpinned by assessments and close monitoring of clinical progress. Rehabilitation programmes have been pioneered and then sustained chiefly in the voluntary sector. Some adhere to or have adopted a therapeutic philosophy (e.g. 12-Step based on the Minnesota Model of addiction recovery developed in the USA) or therapeutic community model, while others operate as 'general houses'—which seek to foster responsible communal living and community reintegration. Tier 4 substance misuse interventions for adults are aimed at individuals with a high level of presenting need and usually require a higher level of commitment from drug and alcohol misusers than is required for services in lower tiers. Tier 4 services are rarely accessed directly by clients. Referral is usually from Tiers 2 or 3 services or via community care assessment.

NICE has produced guidelines for the delivery of psychosocial interventions in residential rehabilitation services and also for the organization and delivery of opioid detoxification services.

Commissioning treatment services

The national drugs strategy requires Crime and Drug Partnerships to commission services (or ensure access to) structured treatment (Tiers 3 and 4). The balance of l*f*ocal drug misuse treatment services and their detailed delivery mechanisms should be tailored to fit the needs of the local population; commissioners are encouraged to think systemically rather than focusing on putting in place individual services. Poorly defined care pathways between services and the lack of a joined-up care planned approach is clearly an unsatisfactory situation. Many individuals may require the provision of several different types of treatment service over time. It is quite common for an individual receiving treatment from one provider to receive additional welfare support and other social inclusion services which are provided by other agencies (e.g. housing support, legal advice). These supports are important elements in an effective package of care services that can evolve over the course of an individual's treatment. Together, the four tiers are meant to imply a continuum of care. Generic service providers and state agencies can refer an individual both up and down the four tiers to access appropriate treatment or support services.

Needs assessment

In the following section, we use an epidemiologically-based conceptualization of population treatment needs to discuss the organization of treatment services and methods for assessing need. Needs assessment occupies an importance place in the evidence-based planning process for the design and delivery of substance misuse services. It is the systematic collection of information about a geographically defined population and then applying this to make changes that will be beneficial to health. In the drug misuse field, there is a specific focus on two groups in the community: (a) those that are not in contact with services and treatment agencies and have unmet need; (b) those in contact with inefficient, ineffective or inappropriate health care services who have unmet need or for whom outcomes could be improved. Good needs assessment practice involves the application of epidemiological (and sometimes spatial geographical) techniques to estimate the number of people in the two groups above, clear understanding of the costs and benefits of interventions, a close collaboration with clinical services and the range of community stakeholders, and a planning and evaluation process to effect change. There is active encouragement for drug misuse partnerships and commissioners to undertake comprehensive needs assessments in the area of drug misuse with a specific target to assess the needs of young people. However, there have been few systematic quantitative and qualitative studies conducted in the drug misuse field in the UK. In fact, most studies in the mental health service field have been mainly or exclusively qualitative, relying on focus group discussion material. Multiple indicator methods and capture-recapture techniques have enabled estimates to be derived of the number of problem drug users in local areas.

Target groups

At the level of the individual patient, several headline factors may be influential in the assessment and treatment planning process: age, gender, race, and culture; pregnancy; familial pattern; quantity, frequency, and route of administration of psychoactive substances used; acute intoxication (overdose liability); extent of impairment and complications; social and occupational environmental supports and stressors—including acute housing need, training, and education. Complex cases will usually (but not always) be characterized by drug-related impairment, dependence, regular injecting, high tolerance levels and co-morbid problems across physical, psychological and personal/social functioning domains. At the population level, we identify six, non-independent groups. The prevalence of these groups and their case-mix at the local level will have ramifications for the assessment of health care needs and the planning, commissioning, delivery and monitoring of treatment services.

(a) Non-dependent, hazardous substance users

This group comprises individuals who are experiencing drug-related problems but they do not meet the criteria for dependence. This group may include large numbers of younger users who have begun to use drugs relatively recently. Because members of this group (both adults and particularly young people) are at risk of advancing their drug involvement to more serious levels they may be ideal clients for early intervention services.

(b) Drug injectors

This group comprises individuals who are injecting drugs and who may be at risk of acquiring and transmitting blood borne diseases. Community surveys suggest that less than 1:3 drug injectors share needles and syringes but 1:2 share injecting equipment (filters, spoons and flushing water). Research Individuals who inject drugs are much more likely to be dependent and experience drug-related harms. They constitute a priority group to be attracted to appropriate harm reduction and structured treatment programmes and retained in treatment as appropriate.

(c) Acutely intoxicated drug users

The specific needs of this group are identified because of the morbidity and mortality risks to health due to adverse reactions and drug overdose. This sub-group may overlap with sub-group B

(the IDU). There is evidence that some two-thirds of heroin users have experienced an overdose. Risk of overdose is increased for opiates users who have also consumed other central nervous system depressants—commonly other opiates, alcohol and benzodiazepines. Preventing drug overdose and overdose mortality is a specific priority area. Acute intoxication is a discrete event although an individual's needs may advance to those associated with dependence, co-morbidity and withdrawal management and support. Most services provided to the intoxicated drug user will be found outside specialist drug or mental health services (e.g. accident and emergency departments, police custody). All services, which have contact with opiate users, should have prompt access to the injectable opiate antagonist naloxone which may be administered intravenously, intramuscularly or subcutaneously and can be life-saving in the event of an opiate overdose. There is now widespread recognition of the problem of drug dependence among the prison population and evidence from database linkage studies showing that newly released prisoners are at substantial risk of fatal overdose. In a study of 48 771 male and female sentenced prisoners in England and Wales released during 1998–2000, there were 442 recorded deaths (59 per cent drug-related) in year following release. There were 342 observed male deaths (45.8 expected in the general population) and 100 observed female deaths (8.3 expected). Drug-related male deaths were relatively more likely to involve heroin and female deaths were relatively more likely to involve benzodiazepines, cocaine and tricyclic antidepressants.

(d) Dependent drug users

This group comprise individuals who have drug-related problems and meet ICD/DSM dependence criteria. Dependence ranges in severity and is characterized by substantial impairment in the ability to control the frequency and amount used and various neuro-adaptational aspects. The majority of people presenting to specialist drug misuse services are in this group. They will require carefully planned community (and often residential treatment) together with the offer of aftercare support and access to social inclusion services to assist problems with housing, employment and training.

(e) Drug users with psychiatric co-morbidity

There is widespread concern about improving services and outcomes for people who have co-morbid psychiatric and substance use disorders. There is currently no research and clinical evidence-base for the effective management and care of patients in psychiatric inpatient units with psychoactive substance misuse co-morbidity and this is an important development area. There is some evidence that people with substance use problems and co-morbid psychiatric disorders appear to have a relatively high contact with medical services and may require more intensive treatment. However, it would appear that substance use disorders amongst people admitted for psychiatric treatment are of a less severe nature than those entering treatment for primary substance use problems. It is also important to consider and plan for the possibility that people with drug misuse and severe mental illness will not respond well or comply with traditional care plans and arrangements. In terms of client attributes, the presence of psychiatric co-morbidity in drug users entering treatment has been linked to poorer outcomes. Pre-treatment psychiatric severity has been found to be predictive of outcome and this should be taken into account when selecting appropriate treatments. The importance of providing social inclusion and reintegration services, particularly in the first three months of treatment has been advocated for community-based treatment services. However, the intensity or comprehensiveness of services *per se* is not consistently associated with improved outcome. The matrix of client attributes and treatment factors and processes has important implications for referral, assessment and client treatment-placement activities.

(f) Drug users in recovery

This group denotes individuals who have achieved a state of abstinence from their main problem drug (or all drugs), usually through successful completion of a health care treatment episode. This group may require residential rehabilitation services or community based aftercare programmes and other supports.

The process and techniques of needs assessment

Service commissioners should follow a sequence of steps to inform the needs assessment for their population requirements. The overarching goal is to produce a strategic commissioning framework which can be agreed across the health, social, and criminal justice partners. This remains an evolving area with guidance available from the NTA to assist the commissioning and implementation of work in this area. The usual steps when conducting a needs assessment are as follows:

- Allocation of resources and establishment of an agreed plan/methods.
- Prevalence estimation of target population and identification and profiling of sub-groups.
- Mapping of treatment services provided in the locality and an audit of treatment commissioning purchasing from services located outside the geographical boundary of the drug misuse partnership (e.g. Social Services purchasing of residential rehabilitation) to determine the extent to which demand is being met elsewhere.
- Audit of the demand profile of treatment services (capacity; number of episodes; estimated number in need).
- Personal interviews with key informants across commissioning, provision and advocacy sectors.
- Focus group discussions with key stakeholders (commissioners, clinicians, treatment providers, service users, carers of service users) to explore what they want from services.
- A 'gaps' analysis of current and desired profile of service provision (often qualitative exercise involving estimation of desired range of services to increase coverage for specific special groups).
- Recommendations for increasing treatment coverage, purchasing efficiency, and service effectiveness based on available evidence.
- Assessment of reactions to recommendations from strategists, commissioners, purchasers, services providers, and service users.
- Development of an implementation plan based on the identification of activities, resources, and timetables.

People in the seven sub-groups summarized above are not all the same and it is necessary to characterize each group on the basis of

the severity of their problems (and extent of any complications). It is important to note that the above sub-categories are not mutually exclusive. Indeed, it is likely that an individual patient will occupy more than one category at any particular point in time (e.g. the injecting dependent heroin user with co-morbidity of HBV infection). The multiple occupation of different categories may also vary over time. In addition to these six primary groups, there is a further category which can be labelled 'at risk'. There is particular concern about segments of the younger population (see below) thought to be at risk; prevention initiatives and general educational programmes are Drug misuse services and treatment modalities.

The appraisal of the healthcare needs of the target populations and commissioning of strategic service responses should be flexible and adaptive to changing circumstances in each locality, including: variations and new trends in drugs use and consumption patterns; the geographical distribution and concentration of drug use; variations in demand for services; the changing relationship between drug use and other conditions (notably HIV infection, and blood borne viral hepatitis); changes in the organization of health services and monitoring the evidence-base for current and new treatment services.

Needs assessment activities are potentially costly activities. Intensive surveys of the resident population in most areas will be time consuming and expensive. It is quite likely that most partnerships will employ alternative (and less precise) estimation methods with which to inform the direction and success of commissioning strategies. A qualitative approach to needs assessment can be undertaken relatively quickly and can answer important questions concerning what commissioners, purchasers, service providers and service users want from treatment services and supports. Service user satisfaction surveys may be a useful means of gathering information about the extent to which a programme is perceived to have met an individual's treatment wants and needs. A range of issues has been examined including the accessibility, adequacy, content, and impact of services received. In addition to serving a simple monitoring function for treatment service providers and their commissioners, treatment satisfaction is argued to be a valuable indicator of treatment experience. Treatment satisfaction can act as a moderator of treatment outcome, since it is reasonable to assume that less satisfied clients may leave treatment prematurely or have different responses to interventions. Both users and carers should be routinely involved and consulted on service development and the setting of commissioning activities. Many DAT Partnerships (and the services they commission) now have good user/carer involvement mechanisms and guidance exists to enable the development of these mechanisms where they are still lacking.

Monitoring the impact of interventions

Drug interventions are from identical in their structure and operation, and outcome studies show that their level of effectiveness varies widely. Many service providers have been interested in monitoring their own outcomes reflecting organizational learning and quality values. The National Treatment Agency has now further developed the NDTMS as a national outcomes monitoring system. There are several benefits from collecting assessment and outcome information. Firstly, many patients perceive that a structured approach to assessment and recording outcome is a reflection of a service that is committed to providing the best care.

The feedback of information describing during-treatment changes to the patient (and his/her spouse/partner or carers) by clinical staff can be a powerful motivational influence to reinforce progress and assist in personal treatment goal setting. Secondly, clinical staff can monitor the characteristics and outcomes of their caseloads and identify areas of priority work. Thirdly, service managers can aggregate information across staff as an indicator of how well the agency is serving its patients. Fourthly, information on samples of cases of aggregate summaries across services can be provided to funding bodies and other government agencies to show the overall impact of treatment provision for a particular area.

In 2007, the NTA launched a national outcome monitoring system for substance misuse treatment services. A brief, outcome monitoring instrument—the Treatment Outcomes Profile (TOP) has been validated for this purpose. The TOP contains 30 items in four domains and is shown in Table 4.2.4.1. From these, a subset of 20 items are compiled at a national level to assess the overall effectiveness of the treatment system.

Case mix and performance analysis

Prognostic models of treatment outcome have been developed in our health care arenas with some success but the approach is in its infancy in the substance use disorders field. Essentially, a set of variables which predict a health or other outcome are identified using a statistical technique. Various methods can then be used to rank individual treatment provider's outcome performance can then be ranked against this averaged outcome. Having adjusted for

Table 4.2.4.1 The Treatment Outcomes Profile (TOP)

Domain/section	Item—in past 28 days
Substance use	Number of days used the following substances alcohol; opiates; cocaine; crack cocaine; amphetamines; cannabis; other (named) Typical quantity consumed on typical day (recorded as units, grams or amount spent) Number of days injected drugs in past 28 days Whether shared needles and syringes (direct receptive sharing); Yes/No Whether injected using a spoon, water or filter used by someone else (indirect receptive sharing; yes/no)
Crime	Days committed shop theft; days sold drugs Whether committed theft from or of vehicle (yes/no) Whether committed other property theft (yes/no) Whether committed fraud, forgery handling stolen goods (yes/no) Whether committed assault or violence (yes/no)
Health and social functioning	Had acute housing need (yes/no) Was at risk of eviction (yes/no) Number of days had paid work in past 28 days Number of days attended college or school in past 28 days Subjective rating of physical health (0 [poor] to 20 [good]) Subjective rating of psychological health (0 [poor] to 20 [good]) Subjective rating of quality of life (0 [poor] to 20 [good])

patient case mix differences, the objective is then to isolate the characteristics of very successful service providers as well as the correlates of less successful delivery of care. The results of this process can then be used to guide the development of the local treatment system.

Further information

For information on the UK drug misuse strategy and national treatment system in UK see the following websites: http://drugs.homeoffice.gov.uk/ http://www.nta.nhs.uk/

References

1. Audit Commission (2004). *Drug Misuse 2004: Reducing the Local Impact.* London:Audit Commission.
2. NTA and Department of Health (2005). *Statistics From the National Drug Treatment Monitoring System* (NDTMS), 1 April 2003, 31 March 2004. London: DH.
3.. Hay, G., Gannon, M., MacDougall, J., *et al.* (2006). Local and national estimates of the prevalence of opiate use and/or crack cocaine use (2004/05). In: *Measuring Different Aspects of Problem Drug Use: Methodological Developments* (eds. N. Singleton, R. Murray & L. Tinsley). London: Home Office. http://www.homeoffice.gov.uk/rds/pdfs06/rdsolr1606.pdf.
4. National Treatment Agency (2002). *Models of Care for Treatment of Adult Drug Misusers*. NTA, London.
5. National Treatment Agency (2006). *Models of Care for Treatment of Adult Drug Misusers:* update 2006.London: NTA.
6. Higgins, S. T., Delaney, D. D., Budney, A. J., *et al.* (1991). A behavioral approach to achieving initial cocaine abstinence. *American Journal of Psychiatry*, **148**, 1218–24.
7. Silverman, K., Chutuape, M. A., Bigelow, G. E.*et al.* (1999). Voucher-based reinforcement of cocaine abstinence in treatment-resistant methadone patients: effects of reinforcement magnitude. *Psychopharmacologia*, **146**, 128–38.
8. National Institute for Health and Clinical Excellence (2007). Drug misuse: psychosocial interventions.NICE guideline (http://guidance.nice.org.uk/CG51). Accessed on 27.07.07.
9. National Institute for Health and Clinical Excellence (2007). Drug misuse: opioid detoxification. NICE guideline (http://guidance.nice.org.uk/CG52). Accessed on 27.07.07.
10. Marsden, J., Strang, J., Lavoie, D., *et al.* (2000). Epidemiologically-based needs assessment: Drug misuse. In *Health Care NeedsAssessment: the epidemiologically based needs assessment reviews* (eds. A. Stevens, J. Raftery, and J. Mant). First Series Update. Radcliffe Medical Press, Ltd, Abington.
11. Stevens, A., and Gillam, S., (1998)/ Needs assessment: from theory to practice. *British Medical Journal*, **316**, 1448–52.
12. Hostick, T., (1995)/ Research design and methodology for a local mental health needs assessment. *Journalof Psychiatric & Mental Health Nursing*. **2**, 295–99.
13. Frischer, M., Heatlie, H., and Hickman, M., (2004). Estimating the prevalence of problematic and injecting drug use for Drug Action Team areas in England: a feasibility study using the Multiple Indicator Method. Home Office: Online Report 34/04 http://www.homeoffice.gov.uk/rds/pdfs04/rdsolr3404.pdf). Accessed 27.07.07.
14. Health Protection Agency (2005). *Shooting up: infections among injecting drug users in the United Kingdom 2004*. An update: October 2005. HPA, London.
15. Darke, S., and Zador, D., (1996). Fatal heroin 'overdose': a review. *Addiction*, **91**, 1765–72.
16. Darke, S., Ross, J., and Hall, W., (1996). Overdose among heroin users in Sydney, Australia: II. Responses to overdose. *Addiction*, **91**, 413–17.
17. Gossop, M., Griffiths, P., Powis, B., *et al.* (1996). Frequency of non-fatal heroin overdose: survey of heroin users recruited in non-clinical settings, *British Medical Journal*, 313, 402.
18. Darke, S., Ross, J., and Hall, W., (1996). Overdose amongst heroin users in Sydney: I. Prevalance and correlates of non-fatal overdose. *Addiction*, **91**, 405–11.
19. Powis, B., Strang, J., Griffiths, P., *et al.*(1999). Self-reported overdose among injecting drug users in London: Extent and nature of the problem. *Addiction*, **94**, 471–8.
20. Strang, J., Griffiths, P., Powis, B., *et al.* (1999). Which drugs cause overdose among opiate misusers? Study of personal and witnessed overdoses. *Drug and Alcohol Review*, **18**, 253–61.
21. Boys, A., Farrell, M., Bebbington, P., *et al.* (2002). Drug use and initiation in prison: results from a national prison survey in England and Wales. Addiction, **97**,1551–60.
22. Bird, S.M. and Hutchinson, S.J. (2003). Male drugs-related deaths in the fortnight after release from prison: Scotland, 1996–99. Addiction, **98**, 185–90.
23. Farrell, M., and Marsden, J. (2005). Drug-related mortality among newly released offenders 1998 to 2000. Home Office online report 40/05.
24. Johns, A., (1997). Substance misuse: a primary risk and a major problem of comorbidity. *International Review of Psychiatry*, **9**, 233–41.
25. Alterman, A.I., McLellan, T., *et al.* (1993). Do substance patients with more psychopathology receive more treatment? *Journal of Nervous and Mental Disease*, **181**, 576–82.
26. Lehman, A.F., Myers, C.P., and Corty, E. (1994). Severity of substance use disorders among psychiatric in-patients. *Journal of Nervous and Mental Disease*, **182**, 164–7.
27. McLellan, A.T., and Wisner, C. (1996). Achieving the public health potential of substance abuse treatment: implications for patient referral, treatment "matching" and outcome evaluation. In, *Drug Policy and Human Nature* (eds. W. Bickel and R. DeGrandpre), Philadelphia, PA.
28. National Treatment Agency for Substance Misuse (2007). Needs Assessment Manual. (http://www.nta.nhs.uk/publications/documents/nta_needs_assessment_manual_july_06.pdf) Accessed on 27.07.07.
29. Pascoe, G. (1983). Patient satisfaction in primary health care: a literature review and analysis. *Evaluation and Program Planning*, **6**, 185–210.
30. Marsden, J., Farrell, M., Bradbury, C., *et al.* (2007). The Treatment Outcomes Profile (TOP): A structured interview for the evaluation of substance misuse treatment. London: National Treatment Agency for Substance Misuse.

4.3

Schizophrenia and acute transient psychotic disorders

Contents

4.3.1 Schizophrenia: a conceptual history

Nancy C. Andreasen

We know that psychotic disorders have been present and publicly recognized at least since classical times because of their portrayals in literature: the madness of Medea, the frenzied behaviour in *The Bacchae*, or the paranoia of Othello. Perhaps the most 'valid' portrayal from a modern clinical perspective is the feigned madness of 'Poor Tom' in King Lear. Poor Tom is a 'bedlam beggar' who encounters Lear during the great scenes of madness, portrayed while the world itself is also in the midst of a terrible storm. Tom's speech is a classical example of schizophrenic thought disorder, but he also experiences delusions and visual hallucinations:

> Who gives anything to poor Tom? Whom the foul fiend hath led through fire and through flame, and through ford and whirlpool, o'er bog and quagmire, that hast laid knives under his pillow, and halters in his pew; set ratsbane by his porridge; made him proud of heart, to ride on a bay trotting-horse over four-inch'd bridges, to course his own shadow for a traitor. Bless thy five wits! Tom's a-cold, –O, do de, do de, do de. Bless thee from whirlwinds, star-blasting, and taking! Do poor Tom some charity, whom the foul fiend vexes. There could I have him now, –and there, –and there again, and there. (*King Lear*, III. iv. 51–60)

However, the definition and delineation of schizophrenia as a discrete disorder is a relatively recent phenomenon.

The founding fathers of the concept: Kraepelin and Bleuler

The earliest academic formulations of the concept of schizophrenia occurred in the mid-nineteenth century in the work of Bénédict-Auguste Morel and Karl Kahlbaum.[1] Morel coined the term 'démence precoce' to refer to a disorder that he observed in young people that was characterized by cognitive impairments and progressive degeneration. He did not develop the concept fully, however. Instead, under the influence of Darwinian thinking, he became preoccupied with the general concept of hereditary

degeneration, which he described in disorders ranging from intellectual disability to alcoholism. This general concept was highly influential throughout the nineteenth and early twentieth century, which led to some of the earliest studies of the familiality of mental illnesses, and laid early foundations for later efforts to examine the role of genetic factors in schizophrenia. Kahlbaum's seminal contribution was an emphasis on using course of illness (as opposed to symptoms) to define discrete disorders. He objected to the concept that there was only one form of severe mental illness ('unitary psychosis' or 'einheitspsychose') and argued that various kinds of psychotic disorders could be differentiated from one another based on changing patterns of symptoms and long-term outcome. Kahlbaum identified one type as 'hebephrenia'.

Our modern concept of schizophrenia primarily derives, however, from the interaction between two great clinicians early in the twentieth century: Emil Kraepelin and Eugen Bleuler.

Although his ideas were presaged by Morel and Kahlbaum, Emil Kraepelin was clearly the first to give a detailed description of this syndrome and a compelling justification for its delineation. Kraepelin highlighted his concept of the key features of the disorder in the name that he chose for it: It was an illness that tended to begin at an early age ('praecox') and to have a relatively chronic course characterized by significant cognitive and social impairment ('dementia'). Alois Alzheimer was a member of Kraepelin's department in Munich and used the tools of neuropathology to study a similar dementia that began at a later age; examination of the brains of these individuals at post-mortem revealed a characteristic neural signature—plaques and tangles. Kraepelin began to call this disorder Alzheimer's disease and thus gave it its current name, as well as its differentiation from dementia praecox. A similar neuropathological signature was sought for dementia praecox, but it was never found, although Kraepelin hypothesized that it must be a disease involving prefrontal and temporal regions[2]:

> If it should be demonstrated that the disease attacks by preference the frontal areas of the brain, the central convolutions and the temporal lobes, this distribution would in a certain measure agree with our present views about the site of the psychic mechanisms which are principally injured by the disease. (p. 219)

Kraepelin did not select any specific clinical feature as pathognomic, but he did stress the importance of several symptoms as characteristic:

> . . . there are apparently two principal groups of disorders which characterise the malady. On the one hand we observe a *weakening of those emotional activities which permanently form the mainsprings of volition*. In connection with this, mental activity and instinct for occupation become mute. The result of this part of the morbid process is emotional dullness, failure of mental activities, loss of mastery over volition, of endeavor, and of ability for independent action.
>
> The second group of disorders . . . consists in the *loss of the inner unity* of the activities of intellect, emotion, and volition in themselves and among one another … the near connection between thinking and feeling, between deliberation and emotional activity on the one hand, and practical work on the other is more or less lost. Emotions do not correspond to ideas. The patients laugh and weep without recognizable cause, without any relation to their circumstances and their experiences, smile while they narrate the tale of their attempts at suicide . . . (pp. 74–5).

Thus, for Kraepelin, what we now refer to as negative symptoms and fragmenting of thought were two key features of the disorder.

Bleuler was a near contemporary of Kraepelin. During their two long careers they maintained a dialogue between their native countries of Germany and Switzerland. Kraepelin was a thoroughgoing empiricist with a keen eye for detail, while Bleuler was primarily a high-level conceptualizer, although he clearly also had vast clinical experience. Bleuler chose to highlight fragmenting of thinking as the most fundamental feature of schizophrenia and designated it as the pathognomonic symptom. That is, he explicitly stated that this particular symptom ('loosening of associations') was present in all patients with schizophrenia and did not occur in other disorders. Because of the importance that he gave to this particular symptom, he renamed the illness after it (schizophrenia = fragmenting of mind). To this symptom, he added several others that he also considered to be of high importance. These included loss of volition, impairment in attention, ambivalence, autism, and affective blunting. He regarded these symptoms as basic or fundamental and the other symptoms observed in the disorder, such as delusions or hallucinations, as secondary or accessory. He pointed out that these accessory symptoms tended to occur in a variety of other conditions, such as manic-depressive illness, delirium, or dementia.

> Certain symptoms of schizophrenia are present in every case and in every period of the illness even though, as with every other disease symptom, they must have attained a certain degree of intensity before they can be recognized with any certainty . . . Besides the specific permanent or fundamental symptoms, we can find a host of other, more accessory manifestations such as delusions, hallucinations, or catatonic symptoms . . . As far as we know, the fundamental symptoms are characteristic of schizophrenia, while the accessory symptoms may also appear in other types of illness (p. 13).

Bleuler's conceptualization of the disorder captured the imagination of clinicians and investigators throughout the world, and the name he chose for the disorder eventually became the one that is now universally used. The prophecy of Kraepelin's tombstone came true: 'though his name will be forgotten, his work will live on'. During the much of the twentieth century, Bleuler's conceptualization and terminology prevailed. Although he drew on Kraepelinian concepts, very few people were aware of the magnitude of Kraepelin's contributions. Students of schizophrenia used Bleuler's name for the disorder and defined it in terms of 'the four A's' (associations, autism, affect, and ambivalence).

Schneiderian symptoms, psychosis, and the dominance of diagnostic criteria

The Bleulerian emphasis slowly began to change, however, beginning in the late 1960s and 1970s. This change in emphasis arose primarily from an interest in improving diagnostic precision and reliability. Because they are essentially 'all or none' phenomena, which are relatively easy to recognize and define, florid psychotic symptoms such as delusions and hallucinations were steadily given greater prominence and indeed even placed at the forefront of the definition of schizophrenia. Bleuler's secondary or accessory symptoms began to be treated as the pathognomonic symptoms.

The emphasis on florid psychotic symptoms arose because of the influence of Kurt Schneider and the interpretation of his thinking

by influential British psychiatrists. Schneider was greatly influenced by the work of Karl Jaspers, who explored phenomenology and created a bridge between psychiatry and philosophy. Jaspers believed that the essence of psychosis was the experience of phenomena that were 'nonunderstandable'—i.e. symptoms that a 'normal' person could not readily imagine experiencing. Schneider, like Bleuler, wished to identify symptoms that were fundamental. He concluded that one critical component was an inability to find the boundaries between self and not-self and a loss of the sense of personal autonomy. This led him to discuss various 'first-rank' symptoms that were characterized by this loss of autonomy, such as thought insertion or delusions of being controlled by outside forces.[3–5]

Schneiderian ideas were introduced to the English-speaking world by British investigators and began to exert a powerful influence on the concept of schizophrenia. An emphasis on Schneiderian first-rank symptoms satisfied the fundamental need to find an anchor in the perplexing flux of the phenomenology of schizophrenia. Schneiderian symptoms were incorporated into the first major structured interview developed for use in the International Pilot Study of Schizophrenia, the Present State Examination (PSE).[6] From this major base, they were thereafter introduced into other standard diagnostic instruments such as the Schedule for Affective Disorders and Schizophrenia (SADS),[7] Research Diagnostic Criteria (RDC),[8] and the *Diagnostic and Statistical Manual* (*DSM-III*).[9]

The emphasis on positive symptoms, and especially Schneiderian symptoms, derived from several concerns. The first was that Bleulerian symptoms were difficult to define and rate reliably. They are often continuous with normality, while positive psychotic symptoms were clearly abnormal. In addition to concerns about reliability, work with the IPSS and the US/UK study also had indicated that in the United States the concept of schizophrenia had broadened to an excessive degree, particularly in the Northeastern parts of the United States. Thus, in the United States, there was clearly a need to narrow the concept of schizophrenia. Stressing florid psychotic symptoms, particularly Schneiderian symptoms, was a useful way to achieve this end, since it appeared that schizophrenia was often being diagnosed on the basis of mild Bleulerian symptoms. When diagnostic criteria such as the RDC and later *DSM-III* were written, these placed a substantial emphasis on positive symptoms and minimized negative symptoms.

While there have been many good consequences of this progression and of the interest in Schneider's work, there have also been problems.

From a Schneiderian perspective, Schneider's work and point of view has been oversimplified and even misunderstood. As a Jasperian phenomenologist, Schneider was in fact deeply interested in the subjective experience of schizophrenia—in understanding the internal psychological processes that troubled his patients. For him, the fundamental core of the illness was not the specific first-rank symptoms themselves, but rather the internal cognitive and emotional state that they reflected. It is somewhat ironic that he has become the symbol of objective quantification and reductionism. He himself was a complex thinker who was concerned about individual patients.

The development of diagnostic criteria for schizophrenia has also had both advantages and disadvantages. When *DSM-III* was originally developed, it was intended only as a 'provisional consensus agreement' based on clinical judgement. The criteria were created by a small group of individuals who reached a decision about what to include based on a mixture of clinical experience and research data available up to that point. The criteria were chosen to serve as a gatekeeper that would include or exclude individual cases, and they were not intended to be a full description of the illness. Unfortunately, they are now sometimes treated as a textbook of psychiatry. Further, the criteria have become reified and given a power that they originally were never intended to have.

Diagnostic criteria have substantial and undeniable advantages, they improve reliability, provide a basis for cross-centre standardization both nationally and internationally, improve clinical communication, and facilitate research. However, they may also have potential disadvantages and even abuses: they provide an oversimplified and incomplete view of the clinical picture, discourage clinical sensitivity to individual patients and comprehensive history-taking, lead students and even clinicians to believe that 'knowing the criteria is enough', reify an agreement that was only intended to be provisional, and discourage creative or innovative thinking about the psychological and neural mechanisms of schizophrenia.

The concept of positive and negative symptoms

Neither Kraepelin nor Bleuler actually used the terms 'positive symptoms' or 'negative symptoms', although the concepts are embedded in their writings. While various sources for this term can be cited,[10] one of the earliest and most prominent was Hughlings-Jackson.[11] Although Jackson's work was not published until much later, in the late nineteenth century Jackson speculated about the mechanisms that might underlie psychotic symptoms:

> Disease is said to 'cause' the symptoms of insanity. I submit that disease only produces negative mental symptoms, answering to the dissolution, and that all elaborate positive mental symptoms (illusions, hallucinations, delusions, and extravagant conduct) are the outcome of activity of nervous elements untouched by any pathological process; that they arise during activity on the lower level of evolution remaining.

Thus Jackson believed that some symptoms represented a relatively pure loss of function (negative symptoms answer to the dissolution), while positive symptoms such as delusions and hallucinations represented an exaggeration of normal function and might represent release phenomena. Jackson presented these ideas at a time when Darwinian evolutionary theories were achieving ascendance, and his concepts concerning the mechanisms that produced the various symptoms were clearly shaped by a Darwinian view that the brain is organized in hierarchical evolutionary layers. Positive symptoms represent aberrations in a primitive (perhaps limbic) substrate that is for some reason no longer monitored by higher cortical functions. Thus Jackson's concept of negative versus positive symptoms rather closely resembles those which are currently discussed. Although today most investigators do not necessarily embrace the specific mechanism that he proposed, they accept his view that they must be understood in terms of brain mechanisms, as well as his basic descriptive psychopathology.

References to positive and negative symptoms occurred sporadically during the 1970s, sometimes making clear references to Jackson's ideas and sometimes simply presenting notions about the clinical meaning of the distinction. Notable examples include the descriptions of Fish,[3] a reference to the terms by Strauss and others,[12] and Iowa work on affective blunting and on thought disorder, classifying it as positive versus negative.[13, 14]

In 1980, Crow published an influential paper describing a two syndrome hypothesis of schizophrenia using the terms 'positive' and 'negative' symptoms.[15] He proposed that schizophrenia could be divided into two different syndromes, which he referred to as Type I and Type II. Type I schizophrenia was characterized by prominent positive symptoms, an acute onset, good premorbid adjustment, a good response to treatment, intact cognition, intact brain structure, and an underlying mechanism that was neurochemical (dopaminergic) and therefore reversible. Type II schizophrenia was characterized by prominent negative symptoms, an insidious onset, poor premorbid adjustment, a poor response to treatment, impaired cognition, structural brain abnormalities (i.e. ventricular enlargement as visualized by Computerized Tomography [CT]), and an underlying mechanism that was characterized by neuronal loss and therefore irreversible. This proposal was highly generative for research in schizophrenia during the 1980s, primarily because it combined speculations about clinical presentation and about underlying neural mechanisms within a single hypothesis.

Two major problems were inherent in Crow's presentation of this hypothesis, however, which initially limited its empirical testing. One problem was its failure to specify a clear method for measuring positive and negative symptoms, and the second was its failure to indicate which of the broad array of variables associated with each of the two syndromes should be considered dependent or independent. Which symptoms of schizophrenia should be considered to be positive and which negative? Which variable—or group of variables—should be used to define the separate syndromes and test whether the hypothesized relationships were present?

Solutions to these problems were proposed by the investigative team at the University of Iowa.[16] Reliable methods for defining and differentiating positive and negative thought disorder and other negative symptoms such as affective blunting had already been developed at Iowa[13, 14] and the research group there also had a long tradition of developing diagnostic criteria. Consequently, we developed structured scales for the assessment of both positive and negative symptoms, the Scale for the Assessment of Negative Symptoms (SANS) and the Scale for the Assessment of Positive Symptoms (SAPS).[16–18] These scales were intended to provide a more comprehensive, reliable, and well-anchored set of measurements for the evaluation of psychopathology in schizophrenia than had been provided by standard instruments such as the Brief Psychiatric Rating Scale (BPRS).[19] They were subjected to rigorous assessment of their psychometric properties, including internal consistency, reliability, and validity. They were quickly translated into a variety of languages and widely used throughout the world.

In addition, a solution to the second problem was implemented by using a standard strategy in the study of psychopathology: the core clinical syndrome would be treated as the independent variable, while the various associated features would be (somewhat arbitrarily) designated as dependent. To this end, criteria were developed that could be used to classify schizophrenic patients as positive, negative, or mixed. Initial work with these criteria suggested that this strategy could be quite useful.[16, 17] As Crow hypothesized, negative patients differed from positive patients in the predicted direction: larger ventricular brain ratio (VBR), poorer premorbid adjustment, lower educational achievement, and poorer cognitive performance. Many subsequent studies continued to explore the two syndrome hypothesis, with the majority confirming at least some aspects of it. However, other problems with this hypothesis also became evident. Perhaps the most vexing for the hypothesis, and for the application of criteria to categorize patients based on clinical presentation, was the large number of patients with a mixture of positive and negative symptoms. The hypothesis had difficulty in explaining how patients who were both positive and negative could have both reversible and irreversible abnormalities, good and poor premorbid adjustment, and other counterintuitive and contradictory findings. For these reasons, most investigators have regretfully abandoned this appealingly simple and heuristic hypothesis during recent years.

However, the clinical distinction between positive and negative symptoms has remained relatively robust. The tendency to distinguish between these two classes of symptoms has become widespread, and the terms have passed into standard clinical usage. The alacrity with which the terms have been adopted suggests that they fill a useful linguistic and conceptual niche. Negative symptoms are an important component of schizophrenia, and the use of the 'positive' and 'negative' terminology gives them recognition and even equal weight. As the Bleulerian symptoms received de-emphasis because of concerns about reliability, they left a void in the descriptive lexicon. Patients were designated as having 'recovered' when their delusions and hallucinations were no longer present or prominent; yet many remained unemployed, unable to return to school, or socially isolated. What might explain this outcome if their symptoms were genuinely absent? Upon reflection, it became evident that only some of their symptoms were absent, and that a group of 'no name' symptoms were the likely explanation. If these symptoms could be named 'negative', grouped together, measured objectively and reliably, and related to outcome and treatment, an important mechanism for clinical description and communication was restored. Although the oversimplified distinction between positive and negative implied in the two syndrome hypothesis might be misleading, it was useful to recognize that some symptoms tend to get patients hospitalized and to call these 'positive' and that other symptoms tend to lead to psychosocial morbidity and to call these 'negative'. Thus the distinction at the level of symptoms (as opposed to syndromes or disease categories) is helpful descriptively.

The distinction has also persisted because standardized and reliable methods have been developed for assessing these symptoms and placing them in broad general classes. Instruments such as the SANS and SAPS have facilitated the persistence of the terminology because they have provided the tools for rating and measuring. Although tools were at hand for most positive symptoms, no scale was available at all for negative symptoms prior to the SANS. The extensive and repeated documentation of its reliability has quieted concerns that negative symptoms are too 'soft' to be assessed precisely, accurately, and objectively. Furthermore, other simpler scales, targeted primarily for use in

clinical drug trials, have also been developed.[20] By the time that *DSM-IV*[21] and *ICD*-10[22] were written, the concept of positive and negative symptoms was so widely accepted that negative symptoms were included in their diagnostic criteria for the first time.

Beyond diagnostic criteria and the search for fundamental mechanisms

As the present moves towards the future, corrective readjustments are continuing to occur. Paradoxically, these often occur by returning to the past and coming back full circle to the work of Kraepelin, Bleuler, Jackson, and Schneider.

Clinically, the emphasis on negative as well as psychotic symptoms is leading to increased interest in the full range of symptoms of schizophrenia and in developing methods for treating that full range. In particular, there has been a growing interest in developing improved treatments for negative symptoms. The interest in negative symptoms has been complemented by a return to an interest in cognitive aspects of schizophrenia. Many negative symptoms are cognitive in nature—alogia (poverty of thought and speech), avolition (inability to formulate plans and pursue them), and attentional impairment. While their assessment may emphasize objective aspects of behaviour in order to achieve reliability, their underlying essence is in the domains of thought and emotion. Increasingly, therefore, investigators are returning to the original insights of Kraepelin and Bleuler that the core symptoms of schizophrenia represent a fundamental deficit in cognition and emotion.

Several prominent investigators have turned from a focus on explaining and 'localizing' the specific symptoms of schizophrenia to a search for more fundamental underlying cognitive mechanisms.[23] Examples include Frith's hypotheses concerning an inability to think in 'metarepresentations',[24] Goldman-Rakic's studies of working memory,[25] our descriptions of cognitive dysmetria Andreasen,[26] or the work of Holzman,[27] Braff,[28] Swerdlow and Geyer,[29] and Freedman[30] on information processing and attention. These cognitive models provide a general theory of the disease that is consistent with its diversity of symptoms, permit testing in human beings with a variety of convergent techniques (e.g. imaging, neurophysiology), and even permit modelling in animals. This efficient and parsimonious approach offers consider hope for the future because it facilitates the search both for improved treatments and for molecular mechanisms.

Further information

Ackerknect, E. (1959). *A short history of psychiatry*, Hafner, New York.
Zilboorg, G. (1967). *A history of medical psychology*, WW Norton, New York.

References

1. Shorter, E. (1997). *A history of psychiatry: from the era of the asylum to the age of prozac*. John Wiley & Sons, New York.
2. Kraepelin, E., Barclay, R.M., and Robertson, G.M. (1919). *Dementia praecox and paraphrenia*. E. & S. Livingstone, Edinburgh.
3. Fish, F.J. (1962). *Schizophrenia*. Williams and Wilkins, Baltimore.
4. Mellor, C.S. (1970). First rank symptoms of schizophrenia. I. The frequency in schizophrenics on admission to hospital. II. Differences between individual first rank symptoms. *The British Journal of Psychiatry*, **117**(536), 15–23.
5. Schneider, K. (1959). *Clinical psychopathology*. Grune & Stratton, New York.
6. Wing, J.K., Cooper, J.E., and Sartorius, N. (1974). *Measurement and classification of psychiatric symptoms; an instruction manual for the PSE and Catego Program*. Cambridge University Press, New York, London.
7. Endicott, J. and Spitzer, R.L. (1978). A diagnostic interview: the schedule for affective disorders and schizophrenia. *Archives of General Psychiatry*, **35**(7), 837–44.
8. Spitzer, R.L., Endicott, J., and Robins, E. (1978). Research diagnostic criteria: rationale and reliability. *Archives of General Psychiatry*, **35**(6), 773–82.
9. Association AP, Statistics CoNa. (1980). *Diagnostic and statistical manual of mental disorders* (*DSM-III*) (3rd edn). American Psychiatric Association, Washington, DC.
10. Berrios, G.E. (1985). Positive and negative symptoms and Jackson. A conceptual history. *Archives of General Psychiatry*, **42**(1), 95–7.
11. Hughlings-Jackson, J. (1931). In *Selected writings of John Hughlings Jackson* (ed. J. Taylor). Hodder & Stoughton, Ltd., London.
12. Strauss, J.S., Carpenter, W.T. Jr, and Bartko, J.J. (1974). The diagnosis and understanding of schizophrenia. Part III. Speculations on the processes that underlie schizophrenic symptoms and signs. *Schizophrenia Bulletin*, **11**, 61–9.
13. Andreasen, N.C. (1979). Thought, language, and communication disorders. I. Clinical assessment, definition of terms, and evaluation of their reliability. *Archives of General Psychiatry*, **36**(12), 1315–21.
14. Andreasen, N.C. (1979). Affective flattening and the criteria for schizophrenia. *The American Journal of Psychiatry*, **136**(7), 944–7.
15. Crow, T.J. (1980). Positive and negative schizophrenic symptoms and the role of dopamine. *The British Journal of Psychiatry*, **137**, 383–6.
16. Andreasen, N.C. and Olsen, S. (1982). Negative v positive schizophrenia. Definition and validation. *Archives of General Psychiatry*, **39**(7), 789–94.
17. Andreasen, N.C. (1982). Negative symptoms in schizophrenia. Definition and reliability. *Archives of General Psychiatry*, **39**(7), 784–8.
18. Andreasen, N.C. (1987). The diagnosis of schizophrenia. *Schizophrenia Bulletin*, **13**(1), 9–22.
19. Overall, J.E. and Gorham, D.R. (1962). The brief psychiatric rating scale. *Psychological Reports*, **10**, 799–812.
20. Kay, S.R., Fiszbein, A., and Opler, L.A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, **13**(2), 261–76.
21. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders, fourth edition* (*DSM-IV*) (4th edn). American Psychiatric Press, Inc., Washington, DC.
22. World Health Organization. (1990). Chapter V, mental and behavioral disorders (including disorders of psychological development): diagnosis criteria for research. International classification of diseases, tenth edition (ICD-10). Geneva.
23. Andreasen, N.C. (1997). Linking mind and brain in the study of mental illnesses: a project for a scientific psychopathology. *Science*, **275**(5306), 1586–93.
24. Frith, C.D. (1992). The cognitive neuropsychology of schizophrenia. Lawrence Erlbaum, East Sussix.
25. Goldman-Rakic, P.S. (1994). Working memory dysfunction in schizophrenia. *The Journal of Neuropsychiatry and Clinical Neuroscience*, **6**(4), 348–57.
26. Andreasen, N.C., Nopoulos, P., O'Leary, D.S., *et al.* (1999). Defining the phenotype of schizophrenia: cognitive dysmetria and its neural mechanisms. *Biological Psychiatry*, **46**(7), 908–20.
27. Holzman, P.S., Levy, D.L., and Proctor, L.R. (1976). Smooth pursuit eye movements, attention, and schizophrenia. *Archives of General Psychiatry*, **33**(12), 1415–20.
28. Braff, D.L. (1993). Information processing and attention dysfunctions in schizophrenia. *Schizophrenia Bulletin*, **19**(2), 233–59.

29. Swerdlow, N.R. and Geyer, M.A. (1993). Clozapine and haloperidol in an animal model of sensorimotor gating deficits in schizophrenia. *Pharmacology, Biochemistry, and Behavior*, **44**(3), 741–4.
30. Freedman, R., Waldo, M., Bickford-Wimer, P., *et al.* (1991). Elementary neuronal dysfunctions in schizophrenia. *Schizophrenia Research*, **4**(2), 233–43.

4.3.2. Descriptive clinical features of schizophrenia

Peter F. Liddle

The clinical features of schizophrenia embrace a diverse range of disturbances of perception, thought, emotion, motivation, and motor activity. It is an illness in which episodes of florid disturbance are usually set against a background of sustained disability. The level of chronic disability ranges from a mild decrease in the ability to cope with stress, to a profound difficulty in initiating and organizing activity that can render patients unable to care for themselves.

Disorders of thought and perception

Delusions

Although there are no features that provide an unambiguous distinction between the delusions of schizophrenia and those of other psychotic illnesses, the delusions most typical of schizophrenia have an enigmatic character rarely seen in other disorders. In contrast to the delusions of affective psychosis, which usually have content consistent with the prevailing emotional state, in schizophrenia delusions often appear to reflect a fragmented experience of reality. This fragmentation is manifest in several ways.

- There is a lack of logical consistency between the components of the belief, or between the belief and common understanding of what is possible. For example, a patient was very distressed by the belief that he had no head and also that there was blood all over his face.
- Behaviour bears an unpredictable relationship to the delusional belief. In some instances, the patient believes he has a special role or identity, yet for the most part, lives a life that is scarcely influenced by the belief. In the words of Bleuler[1]: 'Kings, Emperors, Popes, and Redeemers engage for the most part, in quite banal work, provided they still have any energy at all for activity'.
- In the chronic phase of the illness, patients might acknowledge that a former delusion was not justified, yet in the same interview they reiterate the delusional belief. Bleuler[1] reported: 'sometimes the patients even produce thoughts which are only understandable if it is assumed that the delusions still retain some reality for these patients even though consciously they may reject them'.

The mental mechanism of schizophrenic delusions remains to be ascertained. It is not a lack of capacity for logical thought; rather it appears that certain ideas acquire an attribute that exempts them from the normal processes of validation. This phenomenon is illustrated by the historic case of Daniel Schreber,[2] a high-ranking judge from Leipzig, who suffered a late-onset schizophrenic illness. After obtaining a court order for discharge from his second hospital admission he published his memoirs[3] in a volume that includes his own account of his beliefs, and also the report prepared by the asylum director, Dr Weber, opposing his discharge. For the purpose of understanding the nature of delusions in schizophrenia, Schreber's account is of special value because we have access to his own perceptions of his condition in addition to detailed accounts by his physician. Dr Weber reported that Schreber exhibited lively interest in his social environment, a well-informed mind, and sound judgement, while nonetheless maintaining his delusional beliefs in a manner that would accept no contrary argument. Schreber himself agreed that his beliefs were unchangeable. He believed that he had a mission to redeem the world and restore humankind to its lost state of bliss. His system of delusions included the belief that he was being transformed into a voluptuous female partner of God. He considered that his beliefs belonged to a domain that was exempt from normal logic: 'I could even say with Jesus Christ: My kingdom is not of this world; my so-called delusions are concerned solely with God and the beyond'. Furthermore, he maintained total conviction in his core beliefs despite recognizing that his experiences earlier in his illness had been unrealistic. He stated:

> Having lived for months among miracles, I was inclined to take more or less everything I saw for a miracle. Accordingly, I did not know whether to take the streets of Leipzig through which I traveled as only theatre props, perhaps in the fashion in which Prince Potemkin is said to have put them up for Empress Catherine II of Russia during her travels through the desolate country, so as to give the impression of a flourishing countryside.

Thus, in the stable phase of his illness, Schreber recognized that his earlier experiences were unrealistic and that his current beliefs defied normal logic, but appeared to regard them as exempt from the need for validation. The late onset of his illness and his high level of professional achievement are unusual for an individual with schizophrenia, and raise questions about the diagnosis. However, the fact that he eventually suffered a marked deterioration in function during his third episode of illness strongly supports the diagnosis of schizophrenia.

In many instances, the delusions of schizophrenia appear to arise from an altered experience of self. The phenomena identified by the German psychiatrist, Kurt Schneider[4] as first-rank symptoms of schizophrenia (discussed in greater detail below) include several symptoms that entail an aberrant experience of ownership of one's own thought, will, action, emotion, or bodily function, which the patient attributes to alien influence. In some cases, delusions might arise from a delusional mood, i.e. an altered sense of reality in which the current circumstances acquire an indefinable transcendental quality.

Although the delusions most characteristic of schizophrenia have an incongruous quality, it is not uncommon for schizophrenic patients to have coherent delusions that are internally consistent and produce predictable behavioural responses. In particular, coherent persecutory delusions are common, and can lead to defensive actions such barricading oneself in one's room with blinds drawn. Ideas of reference and delusions of reference are also prevalent. For example, a patient might report that television

programmes refer specifically to him or her. In the International Pilot Study of Schizophrenia[5] conducted by the World Health Organization, ideas of reference were reported in 70 per cent of cases, suspiciousness in 66 per cent, and delusions of persecution in 64 per cent.

Hallucinations

Hallucinations in any modality can occur, but auditory hallucinations are the most prevalent in schizophrenia. Hearing voices speaking in the third person is the most specific. This experience is listed among the Schneiderian first-rank symptoms. Sometimes the content is mundane, as in the instance when a patient of Bleuler[1] heard a voice saying 'Now she is combing her hair' while she was grooming in the morning. In other instances there is an implied criticism, as in the case reported by Schneider[4] of a woman who heard a voice saying 'Now she is eating; here she is munching again', whenever she wanted to eat.

Second-person auditory hallucinations are also common. In the International Pilot Study of Schizophrenia,[5] voices speaking to the patient were reported in 65 per cent of cases. Voices might issue commands that the patient obeys. In some instances, the patient engages in a dialogue with the voices.

During the acute phase of illness, auditory hallucinations usually have the same sensory quality as voices arising from sources in the external world. In some instances the voice is attributed to a radio-transmitter implanted in the body, especially in the teeth. In the chronic phase, the voices are often recognized as coming from within the person's own mind. Kraepelin[6] reports: 'at other times they do not appear to the patient as sense perceptions at all; they are 'voices of conscience'; 'voices which do not speak with words". These experiences are pseudohallucinations, but nonetheless they are a significant feature in many cases.

In schizophrenia, visual hallucinations are less common than auditory hallucinations, but do occur. Somatic hallucinations are also relatively common, and often are associated with a delusional misinterpretation. For example, a young man reported sensations in his belly that he attributed to a snake, which he believed had crawled up his anus.

Schneiderian first-rank symptoms

Kurt Schneider[4] identified a set of phenomena that he considered were strongly indicative of schizophrenia in the absence of overt brain disease. These symptoms, listed in Table 4.3.2.1, have become known as first-rank symptoms. Schneider did not consider that the diagnosis could be made simply on the presence of one such symptom; on the contrary, he warned,[4] 'a psychotic phenomenon is not like a defective stone in an otherwise perfect mosaic'. Schneider did not define the phenomena precisely, and clinicians have interpreted his writings differently. Mellor[7] formulated a precise set of definitions and found that, according to these strict criteria, 72 per cent of patients with schizophrenia exhibited at least one first-rank symptom. Applying the same criteria, O'Grady[8] found that in a series of cases assessed at admission to hospital, 73 per cent of schizophrenic patients exhibited at least one first-rank symptom, while no cases of affective psychosis did. However, applying less strict criteria, O'Grady found more broadly defined first-rank symptoms in 14 per cent of patients with affective psychosis.

Table 4.3.2.1 Schneiderian first-rank symptoms

Voices commenting—a hallucinatory voice commenting on one's actions in the third person
Voices discussing or arguing—hallucinations of two or more voices discussing or arguing about oneself
Audible thought—hearing one's thoughts aloud
Thought insertion—the insertion, by an alien sources, of thoughts that are experienced as not being one's own
Thought withdrawal—the withdrawal of thoughts from one's mind by an alien agency
Thought broadcast—the experience that one's thoughts are broadcast so as to be accessible to others
Made will—the experience of one's will being controlled by an alien influence
Made acts—the experience that acts executed by one's own body are the actions of an alien agency, rather than oneself
Made affect—the experience of emotion that is not one's own, attributed to an alien influence
Somatic passivity—bodily function is controlled by an alien influence
Delusional perception—the attribution of a totally unwarranted meaning to a normal perception

Three of the first-rank symptoms (voices commenting, voices discussing, and audible thoughts) involve auditory hallucinations, while the remainder entail delusional attributions to experiences or perceptions. Although Schneider himself avoided speculating on the theoretical implications of these phenomena, it is notable that most of them involve a disorder of the sense of ownership of one's own mental or physical activity. Thought broadcast, thought withdrawal, and thought insertion reflect the experience of loss of autonomy over thought, while made will, made acts, made affect, and somatic passivity reflect loss of autonomy over action, will, affect, and bodily function.

Mellor[7] emphasizes that there are two aspects to these phenomena: the experience of loss of autonomy and the delusional attribution to alien influence. As an illustration of made acts, Mellor reports a patient who reported that his fingers moved to pick up objects 'but I don't control them … I sit there watching them move, and they are quite independent, what they do is nothing to do with me. I am just a puppet … I am just a puppet who is manipulated by cosmic strings'. To illustrate made affect, Mellor quotes a young woman: 'I cry, tears roll down my cheeks and I look unhappy, but inside I have a cold anger because they are using me in this way, and it is not me who is unhappy, but they are projecting unhappiness into my brain'.

Delusional perception, in which an entirely unwarranted conclusion is drawn from a normal perception, illustrates the incongruity between a delusional idea and concurrent mental activity, which is characteristic of schizophrenia. However, the way in which delusional perceptions often crystallize from a delusional mood indicates that it is not merely a matter of illogical inference; the delusional idea is more like a divine revelation. Mellor[7] gives the example of an Irishman who experienced a sense of foreboding

while seated at the breakfast table in a lodging house. When another lodger innocently pushed the salt cellar towards him, he suddenly knew this meant that he must return home to greet the Pope who was visiting his family to thank them because Our Lord was to be born again to one of the women.

Disorders of the form and flow of thought

The speech of schizophrenic patients is often difficult to understand because of abnormalities of form of the underlying thought. However, the clinical assessment of thought form disorder remains a major challenge. This is due in part to the fact the essential features of formal thought disorder in schizophrenia have yet to be defined in a fully satisfactory manner. Furthermore, thought disorder is usually manifest during spontaneous speech, making it difficult to create circumstances in which the phenomena can be elicited reliably.

Bleuler[1] coined the term loosening of associations to describe the weakening of the connections between words and ideas that bind thoughts into a coherent whole. While this term is a useful label for one of the major types of disorder of the form of speech and thought, it does not encompass the entire range of such disorders. In addition to disordered connections between words and ideas, there are oddities in the use of language. One of the most comprehensive catalogues is the Thought, Language, and Communication Scale compiled by Andreasen.[9] This scale includes several items that involve different aspects of the loosening of associations:

- Derailment—wandering off the point during the free flow of conversation
- Tangentiality—answers to questions that are off the point
- Incoherence—a breakdown of the relationships between words within a sentence so that the sentence no longer makes sense
- Loss of goal—failure to reach a conclusion or achieve a point.

The Thought, Language, and Communication Scale also includes several items that refer to unusual use of language:

- Metonyms—unusual uses of words (e.g. hand-shoe instead of glove)
- Neologisms—new words invented by the patient.

The various aspects of loosening of associations and peculiarities of language use are commonly regarded as positive thought disorder. The Thought, Language, and Communication Scale also include negative thought disorders that entail impoverishment of thinking:

- Poverty of speech—a disorder of the flow of speech in which the rate of speech production is reduced
- Poverty of content—the amount of information conveyed is relatively little in proportion to the number of words uttered.

The Thought, Language, and Communication Scale has proved to be one of the most successful of recent attempts to define and quantify formal thought disorder, but it has several limitations. Most important of these is that the positive thought disorder items defined in the scale do not discriminate well between manic thought disorder and florid schizophrenic thought disorder.[10] Secondly, the scale is not sensitive to the subtle thought form disorders that occur in first-degree relatives of schizophrenic patients.

These limitations are dealt with, at least partially, in the Thought Disorder Index devised by Holzman.[11] This scale employs ratings based on thought and speech elicited by the Rorschach inkblot figures and during an assessment of IQ. Two categories of disorder, disorganization (comprising vagueness, confusion, and incoherence) and idiosyncratic verbalizations, appear to discriminate fairly well between schizophrenic and manic thought.[11] Unfortunately, this scale is too cumbersome for routine clinical use.

Positive formal thought disorder is usually a transient feature of acute episodes of illness. Nonetheless, after resolution of the acute episode there is often a subtle residual thought disorder that is manifest as vague, wandering speech, or minor idiosyncrasies of word usage or ideas. Negative formal thought disorder has a greater tendency to be persistent. Chronic poverty of speech is associated with impairment in several domains of cognition[12] including abstract reasoning. It leads to impaired social relationships,[13] although it is also influenced by the social milieu. Transient poverty of speech can occur during acute episodes of illness. At its most severe, the patient is mute.

Insight

Lack of insight is one of the defining characteristics of psychotic illness. Lack of insight entails a failure to accept that one is ill and to appreciate that symptoms are due to illness. In the International Pilot Study of Schizophrenia[5] lack of insight occurred in approximately 90 per cent of cases. Insight is often partial. In particular, even in instances in which a patient acknowledges suffering from an illness, he or she might fail to accept that psychotic symptoms such as delusions or hallucinations are a manifestation of that illness. Lack of insight is one factor that contributes to unwillingness to accept treatment. However, the clinician should be aware that other factors, including lack of appropriate education about the illness and justified fear of side-effects of treatment, can also impede the development of a therapeutic collaboration between physician and patient.

Impaired cognition

In addition to delusions and disorders of thought form, a wide range of cognitive deficits occur in schizophrenia. These are discussed in Chapter 4.3.3. This chapter focuses on the relationship between cognitive impairment and other features of the illness.

In the acute phase of the illness, attentional impairment is common and is often associated with psychomotor excitation and/or formal thought disorder. It might also reflect preoccupation with delusions and hallucinations.

During the chronic phase of illness, many schizophrenic patients exhibit persistent cognitive impairments. Longitudinal studies of individuals who subsequently develop schizophrenia reveal that the deficits are discernible during childhood, suggesting that these deficits are an aspect of the predisposition to schizophrenia. The major cognitive impairments are in the realm of executive function, working memory, and long-term memory. Executive dysfunction includes impaired ability to initiate and select self-generated mental activity. Impaired ability to form and initiate plans is associated with chronic poverty of speech, blunted affect, and lack of spontaneous activity, while impaired ability to inhibit inappropriate responses is associated with chronic formal thought disorder.[12]

Disorders of emotion

An extensive range of disorders of emotion occur in schizophrenia. Blunted affect and inappropriate affect are the most characteristic, and also tend to be the most persistent, but transient excitation, irritability, lability, and depression are also common.

Blunted affect

Blunting of affect is manifest as decreased responsiveness to emotional issues, loss of vocal inflection, and diminished facial expression. These objective signs of affective blunting are sometimes accompanied by awareness of loss of emotional tone that, paradoxically, patients find to be distressing. More commonly, there is a lack of concern and even a lack of awareness of the problem. Affective blunting is one of the hallmarks of chronic schizophrenia. Bleuler[1] remarked that when the affects disappear, the illness becomes chronic. While blunted affect is usually chronic, it can also be a feature of acute episodes of the illness that resolves as the acute episode resolves.

Inappropriate affect

Inappropriate or incongruous affect is the expression of affect that is inappropriate in the circumstances. At its most severe it takes the form of hollow laughter that is unrelated to any apparent stimulus.

Excitation and depression

During acute exacerbations of schizophrenia, excitation, manifest as irritability, sleeplessness, agitation, and motor overactivity, is common. Depression is also common around the time of an acute episode of schizophrenia,[14] and is often a feature of the prodromal phase of the illness.

Depression also occurs during the chronic phase of the illness. The cross-sectional rate is approximately 10 per cent in the chronic phase,[15] while in a longitudinal study, Johnson[16] found that 65 per cent of schizophrenic patients exhibited an episode of depression in a period of 36 months after a florid psychotic episode.

Motor disorders and catatonia

Subtle disturbance of motor co-ordination is common. Home videos of children who subsequently develop schizophrenia demonstrate that even in infancy they are noticeably more clumsy than their siblings, suggesting that disturbed motor co-ordination is an aspect of the predisposition to schizophrenia.[17]

Catatonia is a term embracing disorders of the initiation or organization of voluntary movement or posture. The most characteristic catatonic phenomena are:

- *Immobility*—absence of motor activity
- *Posturing*—adopting an unusual body posture
- *Waxy flexibility*—allowing an examiner to adjust one's posture, yielding like a warm candle
- *Negativism*—resisting manipulation by an examiner, with force proportional to that applied by the examiner
- *Stereotypy*—aimless repetitive motor behaviour
- *Mannerisms*—apparently purposeful actions that appear odd because they are exaggerated in form or occur out of the usual context
- *Echo phenomena*—repetition of an examiners utterances or movements
- *Excitement*—excessive motor activity, usually accompanied by excessive mental activity.

These disorders can occur not only in schizophrenia, but also in other psychiatric or neurological disorder such as bipolar mood disorder or encephalitis, or alone as a primary disorder of motility.

Kahlbaum[18] provided the classic description of catatonia. He emphasized not only the typical phenomena but also a characteristic time course, in which a prodromal phase dominated by melancholic symptoms evolved into a fluctuating disorder in which episodes of diminished motility typically lasting for several weeks or months, were interspersed in periods of near normal function, but with a tendency towards eventual dementia in many cases. Episodes were often accompanied by confusion and in some instances, a state of stupor. In some cases, episodes of excitation occurred as well. While Kahlbaum's emphasis on both characteristic phenomena and time course laid a foundation for the subsequent delineation of major mental illnesses at the end of the nineteenth century and in particular, for Kraepelin's delineation of schizophrenia,[7] it is probably best to regard catatonia as a cluster of clinical features that can occur within various different illnesses.

Two major forms are retarded catatonia, characterized by slowed or diminished activity; and excited catatonia in which the dominant feature is excessive motor activity. Many variants of catatonia differing in the relative prominence of the characteristic features; or in the associated features such as autonomic instability; or in time course, have been reported in the past 150 years. Fink and Taylor[19] argue on the grounds similarity of clinical features and response to treatment (with benzodiazepines and/or ECT) that the variants of catatonia share a common brain pathophysiology, though different predisposing or precipitating factors might lead to variation in clinical features and time course. The variants include malignant catatonia, in which there is a sudden onset of excitation, associated with fever and autonomic instability, leading to fatal outcome in a substantial proportion of cases. The relationship between malignant catatonia and neuroleptic malignant syndrome (NMS), which is characterized by features very similar to malignant catatonia, but is triggered by antipsychotic medication, is an issue of practical clinical importance. A careful review nine cases of NMS and 17 cases of malignant catatonia by Carroll and Taylor[20] failed to find differences in clinical features, supporting the conclusion that NMS is a form of catatonia.

Disorders of volition

Among the most disabling of the clinical phenomena of schizophrenia are disruptions of motivation and will. Voluntary activity can be disjointed or weakened. Disjointed volition is manifest in poorly organized ill-judged activities which appear to be prompted by impulse. For example, an artistic, intelligent young woman felt cold so she lit a fire on the carpet in her bedroom, even though she was able to appreciate that this was a dangerous thing to do.

Weakened volition results in prolonged periods of underactivity. The patient might lie in bed or sit in an armchair for hours.

Anxiety and somatoform disorders

Various forms of anxiety and somatic symptoms are common in schizophrenia. Huber[21] described a non-characteristic defect state which is dominated by anxiety and asthenia. Coenesthesia, in which the patient suffers unusual or debilitating bodily experiences that do not have an apparent somatic cause, occurs frequently.

Dimensions of psychopathology in schizophrenia

Schizophrenia is heterogeneous in its clinical presentation, suggesting that several different pathophysiological processes might contribute to the illness.

Positive and negative symptom dimensions

Positive symptoms are those that reflect the presence of an abnormal mental process, and include delusions, hallucinations, and formal thought disorder. Negative symptoms reflect the diminution or absence of a normal mental function. They include poverty of speech and blunted affect. In schizophrenia, positive symptoms tend to be transient, while negative symptoms tend to be chronic. In an influential hypothesis, Crow[22] proposed that positive symptoms arise from dopaminergic overactivity, while negative symptoms reflect structural brain abnormality. While this hypothesis is consistent with a substantial body of evidence, it does not account adequately for the complexity of the heterogeneity of the clinical features in schizophrenia.

Three dimensions of characteristic symptoms

The preponderance of evidence[12, 13, 23] from factor analysis of schizophrenic symptoms indicates that the characteristic symptoms of schizophrenia segregate into three syndromes, as shown in Table 4.3.2.2. These syndromes do not reflect separate illnesses, but different dimensions of illness, in the sense that a patient might exhibit more than one of the syndromes.

The three syndromes embrace only the characteristic symptoms that are given weight in making a diagnosis of schizophrenia. In addition, there are two affective syndromes, depression and psychomotor excitation, which are prevalent in schizophrenia,[12] despite being more characteristic of mood disorders. These affective syndromes are usually transient.

Table 4.3.2.2 Three syndromes of symptoms characteristic of schizophrenia

Reality distortion
Delusions
Hallucinations
Disorganization
Thought form disorder
Inappropriate affect
Bizarre behaviour
Psychomotor poverty (core negative symptoms)
Poverty of speech
Blunted affect
Decreased spontaneous movement

An accumulating body of evidence[12] from brain imaging studies indicates that the three characteristic syndromes are associated with three distinguishable patterns of cerebral malfunction involving the areas of association cortex and related subcortical nuclei, which serve higher mental functions. In an individual case, several of these neural systems might be involved.

Although many details of the relationships between the diverse clinical features of schizophrenia remain uncertain, a growing understanding of the neural pathways involved is beginning to provide the foundation for understanding the protean manifestations of this disorder.

Further information

Cutting, J. (2003). Descriptive psychopathology. In *Schizophrenia*, Chap. 2 (2nd edn) (eds. S.R. Hirsch and D. Weinberger), pp. 15–24. Blackwell, Oxford.

Fuller, R.L.M., Schultz, S.K., and Andreasen, N.C. (2003). The symptoms of schizophrenia. In *Schizophrenia*, Chap. 3 (2nd edn) (eds. S.R. Hirsch and D. Weinberger), pp. 25–33. Blackwell, Oxford.

References

1. Bleuler, E. (1950). *Dementia praecox or the group of schizophrenias* (trans. J. Zinkin). International Universities Press, New York.
2. Spitzer, R.L., Gibbon, M., Skodol, A.E., *et al.* (eds) (1989). *DSM-IIIR casebook*. American Psychiatric Press, Washington, DC.
3. Schreber, D.P. (1955). *Denkwurdigkeiten eines Nervenkranken* (*Memoirs of my nervous illness*) (trans. I. Macalpine and R. Hunter). Dawson, London.
4. Schneider, K. (1959). *Clinical psychopathology* (trans. M.W. Hamilton). Grune & Stratton, New York.
5. World Health Organization. (1973). *The international pilot study of schizophrenia*. World Health Organization, Geneva.
6. Kraepelin, E. (1919). *Dementia praecox and paraphrenia* (trans. R.M. Barclay). Facsimile edition, 1971. Kreiger, New York.
7. Mellor, C.S. (1970). First rank symptoms of schizophrenia. *The British Journal of Psychiatry*, **117**, 15–23.
8. O'Grady, J.C. (1990). The prevalence and diagnostic significance of first-rank symptoms in a random sample of acute schizophrenia in-patients. *The British Journal of Psychiatry*, **156**, 496–500.
9. Andreasen, N.C. (1979). Thought language and communication disorders. I. Clinical assessment, definition of terms and evaluation of their reliability. *Archives of General Psychiatry*, **36**, 1315–21.
10. Andreasen, N.C. (1979). Thought, language and communication disorders. II. Diagnostic significance. *Archives of General Psychiatry*, **36**, 1325–30.
11. Holzman, P.S., Shenton, M.E., and Solovay, M.R. (1986). Quality of thought disorder in differential diagnosis. *Schizophrenia Bulletin*, **12**, 360–71.
12. Liddle, P.F. (1999). The multi-dimensional phenotype of schizophrenia. In *Schizophrenia in a molecular age* (ed. C.A. Taminga), pp. 1–28. American Psychiatric Press, Washington, DC.
13. Liddle, P.F. (1987). The symptoms of chronic schizophrenia: a re-examination of the positive-negative dichotomy. *The British Journal of Psychiatry*, **151**, 145–51.
14. Siris, S.G. (1991). Diagnosis of secondary depression in schizophrenia: implications for DSMIV. *Schizophrenia Bulletin*, **17**, 75–98.
15. Barnes, T.R.E., Curson, D.A., Liddle, P.F., *et al.* (1989). The nature and prevalence of depression in chronic schizophrenic inpatients. *The British Journal of Psychiatry*, **154**, 486–91.

16. Johnson, D.A.W. (1988). The significance of depression in the prediction of relapse in chronic schizophrenia. *The British Journal of Psychiatry*, **152**, 320–3.
17. Walker, E. and Lewine, R.J. (1990). Prediction of adult-onset schizophrenia from childhood home videos of the patients. *The American Journal of Psychiatry*, **89**, 704–16.
18. Kahlbaum, K.L. (1874; translated 1973). *Die Katatonie oder das Spannungsirresein*. Verlag August Hirschwald, Berlin, 1874. Translated: Kahlbaum, K. Catatonia. Translated by Y. Levis and T. Pridon. Johns Hopkins University Press, Baltimore, 1973.
19. Fink, M. and Taylor, M.A. (2003). *Catatonia. A clinicians guide to diagnosis and treatment*. Cambridge University Press, Cambridge.
20. Carroll, B.T. and Taylor, B.E. (1997) The nondichotomy between lethal catatonia and neuroleptic malignant syndrome. *Journal of Clinical Psychopharmacology*, **17**, 235–36.
21. Huber, G., Gross, G., and Schuttler, E. (1975). A long term follow-up study of schizophrenia: psychiatric course of illness and prognosis. *Acta Psychiatrica Scandinavica*, **52**, 49–57.
22. Crow, T.J. (1980). The molecular pathology of schizophrenia: more than one disease process. *British Medical Journal*, **280**, 66–8.
23. Arndt, S., Andreasen, N.C., Flaum, M., *et al.* (1995). A longitudinal study of symptom dimensions in schizophrenia. *Archives of General Psychiatry*, **52**, 352–60.

4.3.3 The clinical neuropsychology of schizophrenia

Philip D. Harvey and Christopher R. Bowie

Introduction

Impairments in a variety of cognitive functions are found in patients with schizophrenia. These impairments affect a wide array of different cognitive abilities and are often quite severe, when compared to standards based on healthy individuals of the same age, education levels, and gender. Cognitive impairments appear to be present across the lifespan, detectable at the time of the first treatment episode, if not before, and to manifest a generally stable course over time. Although the current knowledge base regarding cognition in schizophrenia is quite broad, additional research information is constantly accruing. The main purpose of this chapter is to provide a broad overview of the domains, severity, and course of cognitive impairments in schizophrenia, with a focus on functional relevance and treatment possibilities.

History

Cognitive impairments were reported by both Emil Kraepelin and Eugen Bleuer, both of whom noted that they believed that cognitive impairments were amongst the core features of the illness. The conception of dementia praecox introduced by Kraepelin focused on the cognitive and functional deficits in the illness and likened the condition to a condition such as Alzheimer's disease with an earlier onset age. Over the first half of the twentieth century research on cognition in schizophrenia focused on a variety of different topics, including memory, attention, and language skills.[1]

Clinical neuropsychology and schizophrenia

The development of clinical neuropsychology and formalized neuropsychological (NP) tests led to a substantial increase in interest in cognition in schizophrenia. Classical NP ability domains, as well as the types of tests typically used to assess them are presented in Table 4.3.3.1. Clinical NP assessments develop an understanding of areas of relative strength and weakness, comparing current functioning following illness or injury to evidence or estimates regarding prior functioning.[2] Then a profile can be developed, contrasting better or more poorly performed ability areas. Performance across these ability areas can be converted to standard scores, considering demographic factors that influence performance such as age, education level, and sex.[3] Thus, the results of a clinical NP assessment provide a summary of relative strengths and weaknesses. Clinical NP assessment has moved away from earlier efforts to anatomically localize deficits through test performance or to distinguish 'functional' versus 'organic' impairments. The current conception of neuropsychological performance is largely based on the concept of functional neural networks, which link cortical and subcortical regions through patterns of linked activation during task performance.[4]

Table 4.3.3.1 Important cognitive ability domains and tests

Ability areas	Tests
Perceptual skills	Pattern recognition
Motor skills	Manual dexterity
Attention	
Sustained attention	Continuous performance tests
Selective attention	Resistance to distraction
Working memory	
Spatial working memory	Spatial delayed response tests
Verbal working memory	Measures of verbal memory span
Episodic memory	
Verbal memory	List learning; paragraph recall
Non-verbal memory (spatial memory)	Object learning tests
Procedural memory	Pursuit rotor; mirror writing
Long-term semantic memory	Word recognition reading
Executive functions	
Concept formation	Comprehension tests
Reasoning	Proverb interpretation
Problem-solving	Wisconsin card sort; Tower of London
Inhibition	Stroop test
Processing speed	Trail-making; digit symbol
Verbal skills	
Naming	Object naming test
Verbal fluency	Animal naming

Cognitive impairment in schizophrenia

Severity

Patients with schizophrenia demonstrate impaired performance on NP tests measuring a variety of ability areas. As shown in Table 4.3.3.2, impairments across abilities range from mild to severe.[5] Further, aspects of spared functioning are quite rare, with patients performing at levels worse than population means on nearly all domains other than reading skills, object naming, and recognition memory. These impairments are not due to poor motivation or the presence of psychosis[6]; it is well understood that patients demonstrate persistent NP impairments following recovery from acute psychotic episodes and that cognitive impairments are quite stable over time.

Profile

It is important to consider that patients with schizophrenia show considerably smaller overall decline in intelligence than in some specific ability areas.[7] The majority of these impaired domains are often seen to be those that are associated with the functions of the frontal lobe. However, the notion that the whole array of cognitive impairments seen in schizophrenia could originate from a single localized lesion is implausible, as impairments in cognitive functions that are impaired individuals with medial temporal-hippocampal lesions are also quite profound in patients with schizophrenia.[8]

Table 4.3.3.2 Level of impairment in cognitive abilities in schizophrenia

	Mild	Moderate	Severe
Perceptual skills	X		
Motor skills	X		
Attention			
Sustained attention			X
Selective attention		X	
Working memory			
Spatial working memory		X	
Verbal working memory			X
Episodic memory			
Verbal learning			X
Non-verbal memory (spatial memory)		X	
Delayed recall		X	
Delayed recognition	X		
Procedural memory		X	
Long-term factual memory	X		
Executive functions		X	
Processing speed			X
Verbal skills			
Naming	X		
Verbal fluency		X	

See Heinrichs and Zakzanis[5] for a description of the methods used to evaluate these levels of impairment.

There has been considerable debate about the structure of cognitive deficits in schizophrenia. This debate has focused on whether the profile of relative deficits is generalized, with similar severity across all components, or specific.[9] Proponents of the specific profile argument often site evidence of regional brain dysfunction detected with neuroimaging procedures[10] or more extreme deficits on certain NP tests such as episodic memory.[8] There have been recent factor analytic studies that found complex solution with up to six factors[11] and other studies that found a single factor characterized all of the cognitive data in large samples.[12] It does seem that tests requiring cognitive capacity and processing speed are amongst the most poorly performed and that tests examining the ability to use information acquired prior to the onset of schizophrenia are performed best.

Onset

At the time of the first treatment for schizophrenia, either inpatient or outpatient, people who receive the diagnosis perform in a manner that is nearly as impaired as more chronic patients, with a similar profile of impairment.[13] These data support the idea that cognitive impairment is not continuously progressive over the entire course of illness.[14] It is clear, however, that cognitive impairments may also be detectable in at least some people who are destined to develop schizophrenia. In population-based studies of apparently healthy individuals being screened for induction into compulsory military service, there are clear group differences in performance between individuals who eventually develop schizophrenia and those who do not.[15] These impairments have some level of sensitivity and specificity, but are clearly not diagnostic indicators at that stage. Findings of impairments in cognition prior to the patient's meeting formal diagnostic criteria for the illness do provide additional suggestions that cognitive deficits are central features of the illness.

Cognitive decline in schizophrenia?

While the course of cognitive impairments in schizophrenia appears generally stable over the lifespan, there is a substantial minority of patients with schizophrenia who manifest considerable cognitive impairments that worsen over time. The patients who show these changes tend to be older and with a chronic course of treatment-refractory positive symptoms, often accompanied by a lifetime of institutional care.[16] Longitudinal studies have suggested that younger institutionalized patients with similarly severe positive do not show declines during similar follow-up periods,[17] suggesting that there may be age-associated vulnerability to decline. Studies from multiple research sites have found a low prevalence of neurodegenerative changes at post-mortem in older schizophrenia patients,[18] suggesting that these abnormalities cannot be fully explained by degenerative conditions. At present, there is no information on whether these changes could be due to the experience of institutionalization alone, but studies of patients who were released from chronic psychiatric care have not shown evidence of reversal of these cognitive impairments. As patients with schizophrenia have evidence of considerable reduction in their 'cognitive reserve', based on the lower levels of premorbid functioning, it would expected that a variety of risk factors could lead to cognitive changes, including subclinical neurodegenerative pathology, vascular abnormalities, or other factors which can influence cognitive impairments in older individuals.

Presence in relatives and individuals with 'spectrum' conditions

Cognitive impairments are present in the relatives of people with schizophrenia and these impairments have evidence of heritability. Longitudinal studies have suggested that some aspects of cognitive impairment, such as attentional deficits, predict the development of psychotic symptoms in high risk children with at least one schizophrenic parent.[19] Further, individuals with schizophrenia spectrum conditions such as schizotypal personality disorder (SPD) show evidence of cognitive deficits similar in profile, yet reduced in severity compared to people with schizophrenia.[20] As patients with SPD do not have a markedly increased risk for schizophrenia, some aspects of cognitive functioning may represent a stable correlate of some aspects of the predisposition to schizophrenia. While studies have been in process to identify candidate cognitive processes as potential genetically mediated intermediate phenotypes,[21] specific gene-performance correlations are not large enough in magnitude yet to demonstrate that any cognitive impairment is clearly related to specific susceptibility genes for schizophrenia.

Functional relevance

One of the reasons for the increased interest in NP impairment in schizophrenia over the past decade is the developing understanding of the functional relevance of NP impairment. In specific, NP impairment in schizophrenia is the single strongest correlate of impairments in everyday living skills, in social outcomes, and in seeking and maintaining employment or other productive activities. This realization was spurred by several high-profile reviews of the literature[22] and a developing interest in both disability reduction and the direct measurement of disability. While the correlations between impairments in individual NP ability areas and specific aspects of everyday disability are only moderate in size, correlations between composite measures of multiple NP domains and global measures of outcome are often fairly substantial, in the range of Pearson correlations of $r = 0.7$ (reflecting 50 per cent shared variance). In contrast, in similar studies, the cross-sectional correlation between the severity of psychotic (i.e. positive) symptoms is often closer to $r = 0.1$, reflecting about 1 per cent shared variance.[12] This difference in correlations is likely accentuated by the unstable and episodic nature of positive symptoms, in contrast to both functional disability and NP performance, which are both known to be quite stable over time.

Studies of the ability to perform skilled acts (i.e. independent living and social skills) in analogue situations have found that the correlation between impairment on NP tests and deficits in 'functional capacity' is greater than the correlation between NP performance and real world functional performance.[23] This difference in correlations is probably due to the fact that there are multiple factors other than ability that determine everyday outcomes. Opportunities, disability compensation, environmental support, and familial resources are all factors that could lead to discrepancies between what a person can do (i.e. their competence) and what they actually do (i.e. their everyday performance). The fact that disability, in terms of reduced competence, can be measured directly with performance-based tests is quite important, as some of these measures could actually be used in everyday clinical practice or as outcomes in treatment studies.

Treatment of cognitive impairment in schizophrenia

Although antipsychotic medications have been shown for years to be effective in reducing psychotic symptoms in about 70 per cent of patients with schizophrenia, effects on cognitive impairments are much smaller. Although cognitive impairments are apparently not worsened by conventional antipsychotic treatments, their beneficial effects are small and limited to a subset of cognitive domains. Atypical antipsychotic treatment appears to have a somewhat greater effect, suggested by meta-analyses and large-scale studies to be about 0.25 standard deviations.[24] Given the substantial magnitude of impairments in the illness, this level of improvement does not come close to normalization for most patients.

Targeted treatments aimed at cognitive functioning have come from both pharmacological and cognitive remediation domains. Most pharmacological interventions have had quite modest effects, while the results of recent cognitive remediation interventions have been more promising. At least three different interventions, using computerized interventions in randomized trials have shown both cognitive improvements and generalization of improvement to functionally relevant aspects of everyday outcome.[25] Concurrent antipsychotic treatments may be responsible for the poor outcomes of pharmacological interventions, as some of these treatments that have shown minimal benefits in patients with schizophrenia receiving antipsychotic treatments have shown beneficial effects in healthy individuals and in persons with schizotypal personality disorder. This is an issue that will require further study.

Conclusion

Cognitive impairments in schizophrenia are related to the functional disability in the illness and may produce much of the morbidity associated with the condition. These impairments are wide ranging and are found in multiple important domains, with onset at the time of, or in many cases, prior to the first episode of illness. No single focal lesion appears responsible for the array of deficits seen. Relatives of people with schizophrenia and individuals with non-psychotic schizophrenia-spectrum conditions are also affected by these cognitive deficits. Cognitive impairments have proven difficult to treat, but multiple initiatives are underway to improve treatment success. Both pharmacological and cognitive remediation interventions are being studied in detail at this time.

Further information

Resources: National Association for Research in Schizophrenia and Affective Disorders (NARSAD). Promotes research on these topics for junior to distinguished investigators.

Websites: Schizophrenia Research Forum. The ultimate resource for new developments in schizophrenia. http://www.schizophreniaforum.org/

References

1. Chapman, L.J. and Chapman, J.M. (1973). *Disordered thought in schizophrenia*. Appleton, Century, Crofts, New York.
2. Lezak, M.D., Howieson, D.B., and Loring, D.W. (2004). *Neuropsychological assessment* (4th edn). Oxford, New York.

3. Heaton, R.K., Miller, S.W., Taylor, M.J., *et al.* (2005). *Revised comprehensive norms for an expanded Halstead-Reitan Battery (HRB): demographically adjusted neuropsychological norms for African American and Caucasian adults*. Psychological Assessment Resources, Lutz, FL.
4. Braver, T.S. and Barch, D.M. (2002). A theory of cognitive control, aging cognition, and neuromodulation. *Neuroscience and Biobehavioral Reviews*, **26**, 809–17.
5. Heinrichs, R.W. and Zakzanis, K.K. (1998). Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology*, **12**, 426–45.
6. Elevevag, A. and Goldberg, T.E. (2000). Cognitive impairment in schizophrenia is the core of the disorder. *Critical Reviews in Neurobiology*, **14**, 1–21.
7. Weickert, T.W., Goldberg, T.E., Gold, J.M., *et al.* (2000). Cognitive impairments in patients with schizophrenia displaying preserved and compromised intellect. *Archives of General Psychiatry*, **57**, 907–13.
8. Saykin, A.J., Gur, R.C., Gur, R.E., *et al.* (1991). Neuropsychological function in schizophrenia. Selective impairment in memory and learning. *Archives of General Psychiatry*, **48**, 618–24.
9. Dickinson, D., Iannone, V.N., Wilk, C.M., *et al.* (2004). General and specific cognitive deficits in schizophrenia. *Biological Psychiatry*, **55**, 826–33.
10. Hazlett, E.A., Buchsbaum, M.S., Jeu, L.A., *et al.* (2000). Hypofrontality in unmedicated schizophrenia patients studied with PET during performance of a serial verbal learning task. *Schizophrenia Research*, **43**, 33–46.
11. Gladsjo, J.A., McAdams, L.A., Palmer, B.W., *et al.* (2004). A six-factor model of cognition in schizophrenia and related psychotic disorders: relationships with clinical symptoms and functional capacity. *Schizophrenia Bulletin*, **30**, 739–54.
12. Keefe, R.S., Bilder, R.M., Harvey, P.D., *et al.* (2006). Baseline neurocognitive deficits in the CATIE schizophrenia trial. *Neuropsychopharmacology*, **31**, 2033–46.
13. Saykin, A.J., Shtasel, D.L., Gur, R.E., *et al.* (1994). Neuropsychological deficits in neuroleptic naive patients with first-episode schizophrenia. *Archives of General Psychiatry*, **51**, 124–31.
14. Heaton, R.K., Gladsjo, J.A., Palmer, B.W., *et al.* (2001). Stability and course of neuropsychological deficits in schizophrenia. *Archives of General Psychiatry*, **58**, 24–32.
15. Davidson, M., Reichenberg, A., Rabinowitz, J., *et al.* (1999). Behavioral and intellectual markers for schizophrenia in apparently healthy male adolescents. *The American Journal of Psychiatry*, **156**, 1328–35.
16. Harvey, P.D., Silverman, J.M., Mohs, R.C., *et al.* (1999). Cognitive decline in late-life schizophrenia: a longitudinal study of geriatric chronically hospitalized patients. *Biological Psychiatry*, **45**, 32–40.
17. Friedman, J.I., Harvey, P.D., Coleman, T., *et al.* (2001). Six-year follow-up study of cognitive and functional status across the lifespan in schizophrenia: a comparison with Alzheimer's disease and normal aging. *The American Journal of Psychiatry*, **158**, 1441–8.
18. Arnold, S.E., Gur, R.E., Shapiro, R.M., *et al.* (1995). Prospective clinicopathological studies of schizophrenia: accrual and assessment of patients. *The American Journal of Psychiatry*, **152**, 731–7.
19. Cornblatt, B.A. and Erlenmeyer-Kimling, L. (1985). Global attentional deviance as a marker of risk for schizophrenia: specificity and predictive validity. *Journal of Abnormal Psychology*, **94**, 470–86.
20. Mitropoulou, V., Harvey, P.D., Zegarelli, G., *et al.* (2005). Neuropsychological performance in schizotypal personality disorder: importance of working memory. *The American Journal of Psychiatry*, **162**, 1896–903.
21. Goldberg, T.E., Egan, M.F., Gscheidle, T., *et al.* (2003). Executive subprocesses in working memory: relationship to catechol-O-methyltransferase Val158Met genotype and schizophrenia. *Archives of General Psychiatry*, **60**, 889–96.
22. Green, M.F., Kern, R.S., Braff, D.L., and Mintz, J. (2000). Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the 'right stuff?' *Schizophrenia Bulletin*, **26**, 119–36.
23. Bowie, C.R., Reichenberg, A., Patterson, T.L., *et al.* (2006). Determinants of real-world functional performance in schizophrenia subjects: correlations with cognition, functional capacity, and symptoms. *The American Journal of Psychiatry*, **163**, 418–25.
24. Harvey, P.D. and Keefe, R.S. (2001). Studies of cognitive change in patients with schizophrenia following novel antipsychotic treatment. *The American Journal of Psychiatry*, **158**, 176–84.
25. McGurk, S.R., Mueser, K.T., and Pascaris, A. (2005). Cognitive training and supported employment for persons with severe mental illness: one-year results from a randomized controlled trial. *Schizophrenia Bulletin*, **31**, 898–909.

4.3.4 Diagnosis, classification, and differential diagnosis of schizophrenia

Anthony S. David

The diagnosis of schizophrenia

Until the early 1970s, the diagnosis of schizophrenia was one of the most contentious and fraught issues in the whole of psychiatry. Since then a massive international effort has been put in motion out of which explicit diagnostic criteria emerged. Some achieved widespread and even multinational agreement, allowing the painstaking process of calculating diagnostic specificity, sensitivity, reliability, and (perhaps) validity to begin. Although criticism of the diagnosis of schizophrenia continues, mostly from outside psychiatrists, the vast majority of psychiatrists look upon the major sets of diagnostic criteria with weary acceptance, seeing them as flawed but useful and possibly 'as good as it gets' given our current state of knowledge/ignorance.

Throughout the 1970s and early 1980s there was an overabundance of criteria including the St. Louis criteria[1] and the Research Diagnostic Criteria,[2] followed by the Present State Examination (PSE-CATEGO), the ICD-9, and the DSM-III. Perhaps because of the 'cookbook' explicitness of the DSM-III or the pervasive influence of American psychiatric practice, dubbed by some 'neocolonial', the DSM, in its fourth revision with a fifth due in 2010, is the mostly widely used. The ICD-10 is also used throughout the world, but seldom in North America.

Diagnostic criteria

The signs and symptoms of schizophrenia and related disorders are discussed in detail in Chapter 4.3.2. Also, the diagnostic process is described in general in Chapter 1.8.1. As noted, the signs and symptoms, weighted in terms of their typicality or specificity, combined with additional clinical factors such as onset, duration, social consequences, etc., are used to make a diagnosis of schizophrenia and subsequently to classify the disorder into subtypes. The DSM and ICD criteria are described below (Tables 4.3.4.1–4.3.4.3).

Table 4.3.4.1 Major diagnostic criteria for schizophrenia

	DSM-IV	ICD-10
Characteristic symptoms		
One or more for 1 month	1. Bizarre delusions	1. Thought echo/insertion/ withdrawal/broadcasting
	2. Commenting voice or voices conversing	2. Delusions of control
		3. Hallucinatory voices
		4. Persistent delusions
Or two or more	1. Delusions	1. Persistent hallucinations
	2. Hallucinations	2. Thought block/disorder
	3. Disorganized speech	3. Catatonia
	4. Grossly disorganized or catatonic behaviour	4. Negative symptoms
	5. Negative symptoms	5. Significant personality change
Time course	1 month ('significant proportion') for symptoms listed plus 6 months social/occupational disturbance	1 month (most of the time)
Exclusions	Schizoaffective disorder or brief mood disturbance Direct effect of drugs of abuse/ medication or general medical condition	Extensive depressive/manic symptoms or diagnosis of schizoaffective disorder Overt brain disease; drug intoxication/withdrawal

Another group of psychotic disorders which may be distinguished on the basis of formal phenomenological properties are the delusional disorders[3, 4] formally known as paranoia (see Chapter 4.4).

Basis of classification

Atheoretical: Schneider's first-rank symptoms

These are still important for the diagnosis of schizophrenia using the ICD-10 frame of reference. They are too rare to achieve high levels of sensitivity and their specificity has been challenged. Nevertheless, first-rank symptoms perform creditably on these parameters when compared to negative symptoms.[5,6] On the other hand, the lack of aetiological and prognostic significance of first-rank symptoms has undermined the prominence claimed for them.[7,8] The negative[9] or so-called deficit syndrome[10] relates more consistently to outcome/prognosis and shows more stability over time. The constituent symptoms such as social withdrawal, apathy, lack of initiative, and self-care, have rather poor diagnostic specificity in isolation and must be distinguished from depression and parkinsonism, chronic drug dependence, and organic brain damage.

Theoretical

Attempts at a theoretical classification have been made. The first in the modern era was Crow's Type I and Type II distinction,[11]

Table 4.3.4.2 Criteria for the diagnosis of schizophrenia subtypes

Schizophrenia subtypes	DSM-IV	ICD-10
Paranoid	One or more delusions plus frequent auditory hallucinations; no prominent thought disorder, catatonia, or negative symptoms	Delusions, hallucinatory voices, hallucinations in other modalities; disturbances of affect, volition, and speech 'inconspicuous'
Disorganized DSM Hebephrenic ICD	Prominent disorganized speech behaviour and flat/inappropriate affect; no catatonia	Prominent disturbances of affect, volition, and thought; 2–3 months duration; adolescents/young adults only
Catatonic	Two of motoric immobility, excessive activity, negativism, peculiar voluntary movements, echolalia/ praxia	One or more of stupor, excitement. posturing, negativism, rigidity, waxy flexibility, automatic compliance and perserveration
Undifferentiated	Meets criteria for schizophrenia but none of the above subtypes	Meets criteria for schizophrenia but none of the above subtypes plus residual
Residual	Absence of prominent characteristic symptoms (but two or more must be present in attenuated form); continuing evidence of disturbance including negative symptoms	Prominent negative symptoms; clear-cut episode(s) in past; at least 1 year history; no dementia or depression etc.
Simple	Slowly progressive negative symptoms without other psychotic symptoms	(See schizoid personality disorder)

Table 4.3.4.3 Terminology used to describe the course of schizophrenia in the DSM-IV and ICD-10 classifications

DSM-IV	ICD-10
Continuous	Continuous
Episodic with residual symptoms	Episodic with stable deficit
Episodic with no interepisode symptoms	Episodic remittent
Single episode in partial remission	Incomplete remission
Single episode in full remission	Complete remission
Other	Other Episodic with progressive deficit

[a]Course specifiers in both DSM-IV and ICD-10 require 1 year of observation.

although it echoes older notions of 'process'-chronic and deteriorating versus 'reactive' (relapsing and remitting) typologies. The innovation was to link the distinction with proposed differences in dopamine receptor hyperactivity (Type I), associated with positive symptoms and good response to dopamine antagonist drugs, and on the other hand, to neurological damage (Type II) as evidenced by ventricular enlargement on Computerized Tomography (CT) brain scans, associated with chronicity, poor premorbid functioning, and poor response to treatment.

Building on this was the 'aetiological classification' proposed by Murray *et al.*[12] which contrasted cases with a presumed genetic aetiology and those who had other putative risk factors such as early brain damage (see Chapter 4.3.6.1). Although these attempts have served as useful stimuli for research, they have not been found to aid clinical decision-making and in fact now appear to support a blurring of diagnostic boundaries rather than a sharpening or subdivision.[13] In fact the search for 'biological markers' often called 'endophenotypes', which might validate diagnostic distinctions continues. Take for example, the presence of ventricular enlargement or cortical thinning, first detected using CT and now magnetic resonance imaging (**MRI**). Meta-analyses have confirmed that indices of 'cerebral atrophy' are strongly associated with schizophrenia but the effect sizes are small.[14] Medial temporal lobe structures are the region of most grey matter volume loss. However, there is substantial overlap between normal controls and schizophrenia cases and MRI cannot be considered a useful diagnostic test. A host of genetic markers have been identified in the last 5 years, each of small effect and some showing overlap between the major schizophrenic and affective syndromes.[15]

Positive family history remains an important finding in the psychiatric history of an individual patient. Although none of the diagnostic criteria permits the influence of family history, in clinical settings, 'odd' or withdrawn behaviour takes on a very different meaning if seen in the first-degree relative of someone with a firm schizophrenia diagnosis.

Early diagnosis?

The premorbid personality in schizophrenia is typically described as emotionally and socially detached. Such people have few friends, are often cold and aloof, and engage in solitary occupations. Their behaviour may be eccentric and they are indifferent to praise or criticism. Recent studies, including United Kingdom national cohort studies[16] and a Swedish conscript cohort study[17] indicate that children who later develop schizophrenia are more likely to have lower IQs and educational achievements than other children. They are also more likely to have interpersonal and behavioural difficulties. Parents recognize 'preschizophrenic' children as being different from their other siblings. However, such characteristics are very common in the general population so have virtually no positive predictive value.

Early diagnosis is only successful when based on psychotic symptoms. Here the diagnosis of schizophreniform psychosis (DSM-IV) and the acute schizophrenia-like psychotic disorder of the ICD-10 are relevant. The former must last for more than 1 but less than 6 months (otherwise the diagnosis is brief reactive psychosis). Hence the disorder is substantial by any common-sense definition, and unsurprisingly many cases (70 per cent) go on to develop full-blown schizophrenia, affective disorder, or schizoaffective disorder.[18] The temporal stability of the diagnosis is poor, with around 30 per cent recovering over follow-up periods averaging 16 months in one study.

New services have built up around ever earlier diagnosis with the explicit aim of secondary or even primary prevention. Criteria have been developed for the diagnosis of high-, ultra-high, or so-called 'at-risk' mental states based on transient psychotic experiences—even briefer than schizophreniform or more persistent disturbances in the sense of self ('basic symptoms') which fall short of true psychosis.[19,20] One impetus to this being the discovery that most patients when first ill endure a long duration—months or years—of untreated psychosis (DUP).

Differential diagnosis

Other psychiatric disorders

(a) Other psychoses

It could be argued that distinguishing schizophrenia from schizoaffective disorder, schizophreniform disorder, delusional disorder, etc. is an academic exercise. Despite passing enthusiasms, treatment in psychiatry is largely symptom or syndrome based.[21] Thus manic symptoms respond to antimanic agents including lithium, psychotic symptoms respond to 'neuroleptics' or first and now second-generation antipsychotic drugs (SGA), and depressive symptoms respond to antidepressants. Other 'mood-stabilizing' agents are also of value especially when combined with antipsychotics. Several SGAs are licensed for bipolar affective disorder and schizoaffective disorder although it is not clear whether they have distinct advantages over older drugs in this regard. Clozapine remains the only antipsychotic medication which is proven to be effective in at least some patients who are otherwise treatment resistant. However, it is possible that with increasing clinical experience and research more specific indications for newer agents will emerge. This will depend on the preservation of skills in history taking and the mental state examination, and a careful attitude towards making a diagnosis rather than use of sloppy catch-all labels such as 'serious mental illness' favoured by healthcare planners.

The prognostic significance of a diagnosis of schizophrenia (versus schizoaffective and affective disorders) has been discussed in Chapters 4.3.7 and 4.3.9. Although predicting outcome in individual patients is notoriously difficult because of the influence

of idiosyncratic factors such as services, relationships within the family, compliance, intelligence, personality, demographics, etc., the more a disorder approaches 'typical' schizophrenia, the poorer the prognosis tends to be.

That said, schizoaffective disorder is the closest disorder, phenomenologically, to schizophrenia but combines schizophrenic symptoms with affective symptoms. The criteria are discussed in Chapter 4.3.9. Schizophreniform (DSM-IV) or acute schizophrenia-like disorders (ICD-10) differ only in terms of duration, as operationally defined (see Chapter 4.3.10). Delusional disorders (Chapter 4.4) differ from schizophrenia in being based around 'non-bizarre' delusions and few or no hallucinations. The onset and course are characteristically later and more benign respectively.

(b) Affective disorders

Typical presentations of either mania or depression usually cause few diagnostic difficulties. Overdiagnosis of schizoaffective disorder is to be resisted although the distinction from schizophrenia proper remains controversial and debatable. The guidelines given in DSM-IV attempt to exclude transient mood disturbances (<2 weeks) in people with psychosis as a basis for a schizoaffective diagnosis.

In practice reaching a diagnosis of schizophrenia in a person with evidence of one or more core symptoms of psychosis (listed under the DSM-IV and ICD-10) may be complicated for the following reasons.

(c) The presence of mood-incongruent delusions (or hallucinations)

'Congruence' is somewhat in the eye of the beholder, especially where mood may be labile or where disturbed mood is suspected but fails to follow clinical stereotypes. The clinician should try to determine if a 'grandiose' delusion is being enjoyed by the patient, and whether the content (e.g. elevated status, magical powers, material riches) is seen as justified by the patient. Similarly a delusion of depressive content (e.g. physical illness, imminent death) must be seen as undeserved or inexplicable to be deemed 'incongruent'. Auditory hallucinations may be comforting, complimentary, or, more commonly, hostile and critical. It is probably their complexity and personification which makes them 'schizophrenic' rather than their mood-incongruent content.

(d) The duration and acuteness of onset criteria

A good history may simply not be available. Symptoms may wax and wane. Partial or successful treatment may modify or curtail a potentially long episode, and onset may be complicated by the use of psychoactive drugs.

(e) Social and occupational disturbance

This is critical to the diagnosis of schizophrenia, especially the DSM-IV criteria. Here the difficulty is in distinguishing 'premorbid deficits', an illness prodrome and the illness itself. Premorbid personality factors will obscure or set in relief discontinuities in an individual's social trajectory. Objective information and informant testimony is crucial as in most of the diagnostic process. Other individual differences such as intelligence will also shape the presentation of schizophrenia. At the extreme, people with intellectual disability (learning disability) may manifest psychosis in less obvious ways (see Chapters 10.5.1 to 10.5.3). The old diagnosis of 'simple schizophrenia', retained in the ICD-10 describes 'insidious and progressive development of oddities of conduct' and the 'inability to meet the demands of society' that is, social disturbance of long duration. The progressive element distinguishes it from personality disorder although problems adjusting to changing social demands through the life cycle may give the appearance of progression in a fixed personality disorder.

Organic conditions

Differentiation of a 'primary' psychotic illness from one secondary to an organic condition may arise in essentially two situations:

- a person with a clear-cut diagnosis of a medical or neurological syndrome in which psychosis is a recognized complication (e.g. epilepsy)
- a person with a presumptive diagnosis of schizophrenia in whom significant abnormalities are detected usually following special investigation (e.g. CT brain scanning).

The list of medical conditions that could potentially give rise to psychosis is enormous. These have been the subject of extensive reviews.[22,23] While it appears that almost any disease that causes a cerebral perturbation can give rise to psychosis, abnormalities affecting the temporal lobes and diencephalon are somewhat more likely to do this.

The time course is obviously important in this context. Chronic inflammatory lesions (e.g. sarcoidosis), degenerative disorders (e.g. presenile dementias), chronic infections (e.g. neurosyphilis, AIDS), space-occupying lesions (e.g. tumour or abscesses), metabolic disorders (e.g. hyper- or hypothyroidism and vitamin deficiencies) may mimic schizophrenia by virtue of a gradual deterioration in social functioning and self-care punctuated perhaps by odd or inexplicable behaviour and rarely hallucinations and delusions. The features of the primary disease are usually evident. Rarer conditions may be misdiagnosed, for example, Wilson's disease (hepatolenticular degeneration). This usually presents with a motor disorder with bulbar features and abnormal liver function, but personality changes and psychotic symptoms are also associated. Diagnosis is made on other associated clinical features (e.g. Kayser–Fleischer rings), copper studies, and liver biopsy. Huntington's disease is characterized by chorea and cognitive decline. Affective disorder and occasionally psychotic symptoms may occur. The main differential diagnosis is with patients with chronic psychosis and tardive dyskinesia and is usually clarified by the family history, inexorable progression, and caudate atrophy on CT or MRI. Neurosyphilis is still encountered from time to time and in the 'general paralysis of the insane' form, may present with chronic delusions (often grandiose) plus dementia. Diagnosis is by appropriate serological testing of blood and cerebrospinal fluid. Finally, metachromatic leukodystrophy, a rare inherited progressive demyelinating condition, has recently been identified as a cause of a schizophrenia-like psychosis, when onset is in childhood or early adult life.[24] Arylsulphatase-A is a diagnostic marker detectable in peripheral white blood cells.

Acute disturbances following head trauma, acute infections (viral encephalitis), cerebrovascular accidents, metabolic abnormalities (e.g. electrolyte disturbances, porphyria), or drug intoxication or withdrawal (including prescribed medication) (see below) may present with a florid psychotic picture, classically dominated by visual distortions or hallucinations and fluctuating levels of

alertness, rather than the stereotyped auditory hallucinations in clear consciousness which are characteristic of schizophrenia.[25]

In practice there are few common conditions that ever give rise to real diagnostic uncertainty. The most important is **epilepsy**. It is well established that epilepsy, particularly focal (complex partial or 'temporal lobe epilepsy') can give rise to psychosis and there are inter-ictal and post-ictal patterns (see Chapter 5.3.3). A survey from a large neurology clinic showed that the incidence of schizophrenia is about nine times that of the rest of the population.[26]

Inter-ictal psychoses include the chronic schizophrenia-like psychoses described by Slater *et al.*[27] and Trimble.[28] These almost always arise in people with many years of well-established temporal lobe seizures, while the post-ictal variety occurs earlier in the life cycle but again in a person with previously diagnosed epilepsy. In post-ictal psychosis the temporal relationship to seizures, sometimes occurring in a cluster, is diagnostic, although a lucid interval is often observed. A clear history and independent description of seizures is the foundation of a diagnosis of epilepsy, with EEG confirmation. Resting EEGs show slight and subtle abnormalities in a substantial minority of patients with schizophrenia which may be accentuated by antipsychotic medication. As such, the EEG may be of limited value in differential diagnosis unless pronounced slowing or frank seizure activity is picked up (see also Chapter 5.3.3.).

(a) Symptoms

Symptoms of schizophrenic psychosis in relative isolation may give rise to diagnostic difficulties.

Auditory hallucinations may occur in alcoholic hallucinosis (see below and Chapter 4.2.2.3). Hallucinations in the context of dissociation (voices representing figures from the patient's past or embodiments of aspects of their personality) must also be distinguished from typical schizophrenic hallucinations. These are often multimodal. Pure auditory hallucinations in organic conditions including epilepsy in the absence of other psychotic features are surprisingly rare.

Certain forms of delusion suggest alternative diagnoses. Transient ill-formed but usually paranoid delusions occur in the context of confusion, memory impairment, or dementia (i.e. things going missing, strange people loitering). Delusions of misidentification are particularly associated with organic illness such as dementia or stroke.

Thought disorder may be confused with a fluent aphasia following stroke or cerebral tumours.

Personality deterioration and inappropriate or disinhibited behaviour can occur in many organic conditions in the absence of overt psychotic features. Isolated frontal lesions may cause diagnostic problems since general cognitive impairments may be absent. The widespread availability of CT and MRI in the more developed world has reduced the likelihood of such patients being misdiagnosed.

A small proportion (approximately 5 per cent) of prevalent and incident cases of schizophrenia, if investigated thoroughly, are found to have a variety of 'organic' conditions which may contribute to the illness.[29] These include metabolic abnormalities, cerebral tumours, multisystem autoimmune disease, cerebrovascular disease, etc. Some of these may be incidental; others may have precipitated the psychosis. The range of diseases counts against any specific aetiological mechanism. Similarly, the phenomenology found in such 'organic' patients is usually indistinguishable from their 'functional' counterparts.[30]

Thanks to increased application of non-invasive neuroimaging techniques to psychiatric patients, particularly those with schizophrenia, another class of organic abnormalities have been noted, namely cerebral anomalies which are often congenital. These include agenesis of the corpus callosum, cavum septum pellucidum, aqueduct stenosis, etc. Again, it is difficult to know how often such findings occur in the normal population and are asymptomatic, although the widespread use of MRI for 'minor' complaints such as mild head injury and headache is uncovering such anomalies. The examples above certainly appear to be associated with psychiatric disorders in general more than would be expected by chance. They tend to be associated with below-average IQ and other neurological problems (epilepsy in the cases of callosal agenesis).

Other factors to be taken into account in the differential diagnosis from organic conditions include the presence of a family history of schizophrenia, and abnormal premorbid personality, both of which weight aetiological judgement in favour of the functional diagnosis. This applies to the psychoses of epilepsy and those related to drug abuse especially. 'Secondary' schizophrenias also tend to have less pervasive effects on the person's personality. Treatment is again based on symptoms with the added complication that antipsychotic drugs lower the epileptic seizure threshold, and will tend to worsen extrapyramidal symptoms in patients with primary movement disorders. Treatment of the primary condition (if this has remained undiagnosed for some time) may be disappointing but should always be attempted especially in the case of chronic infections. Reversal of metabolic abnormalities, even long-standing, can lead to dramatic improvements in the mental state.

(b) Drug-induced psychoses

Many drugs of abuse and prescribed drugs can cause psychotic symptoms. The associations are also considered in Chapters 4.2.3.1 to 4.2.3.9. In the context of a differential diagnosis, drugs of abuse—in adolescents and young adults—must be considered. Chronic amphetamine psychosis may be indistinguishable from schizophrenia. The psychosis is florid and may include visual and auditory hallucinations. Phencyclidine (PCP or angel dust) is a drug of abuse in the United States and causes an acute psychosis with prominent affective symptoms as well as perceptual distortions and depersonalization. Other psychotogenic drugs include cocaine, ecstasy, and Lysergic Acid Diethylamide(LSD).

Cannabis is widely used, especially in large metropolitan areas and by certain ethnic groups (e.g. African-Caribbeans). Cannabis intoxication is more characterized by perceptual distortions and depersonalization than frank psychosis. Clinical experience suggests that cannabis has a propensity to precipitate psychotic relapse in patients with established schizophrenia and a recent meta-analysis of cohort studies concludes that cannabis use is certainly a risk factor for schizophrenia, and other psychiatric disorders.[31]

Delusions and hallucinations may occur rarely during states of alcohol intoxication but are more commonly associated with withdrawal syndromes (Chapter 4.2.2.2). Alcoholic hallucinosis is a chronic hallucinatory state of uncertain nosological status in which the patient with long-standing alcohol dependence often hears 'voices' which may be derogatory and commenting, in clear consciousness, after a lengthy withdrawal period.

(c) Prescribed medication

Again the list of agents that can cause psychotic reactions to be distinguished from schizophrenia is very long, and psychotropic drugs are particularly liable to cause psychotic reactions. Two classes of drug deserve mention because of their widespread use and relatively high incidence of major psychiatric adverse effects:

- steroids can cause a wide range of psychiatric disturbances including psychosis
- dopamine agonists used in the treatment of Parkinson's disease and some pituitary adenomas.

Frank psychosis and affective disorders may be seen. In the treatment of neurological diseases, such as Parkinson's disease, and the use of steroids for diseases of the central nervous system, there is often an interaction between the agent and the underlying condition which increases the likelihood of a drug-induced psychosis.

The diagnostic process

It used to be argued that a diagnosis of schizophrenia in itself caused disability and morbidity due to social 'labelling' and stigmatization. Evidence that this accounts for schizophrenic disability is lacking but the reality of the stigma of mental illness and negative attitudes towards 'schizophrenics' cannot be denied. This is especially delicate in the case of early intervention with people presenting without the full-blown schizophrenic syndrome since arguably, the balance of harms and benefits of diagnosis is tilted slightly away from benefit. Hence, making a diagnosis of schizophrenia should not be taken lightly. In the author's experience, very few psychiatrists spontaneously convey the diagnosis to the patient. If a patient asks whether he or she has schizophrenia, the clinician should first try to understand the motivation behind the question and the patient's knowledge and understanding of the term. Ultimately there is seldom justification in withholding the diagnosis if it is established. A schizophrenia diagnosis can be framed in a relatively positive light—this is a condition which we are now beginning to understand and for which there are effective treatments—and may lessen the burden of responsibility and blame that the patient and his or her family may carry for the disorder.

Further information

McKenna, P.J. (2007). *Schizophrenia and related syndromes* (2nd edn). Routledge, Hove, East Sussex.

The Cochrane Schizophrenia Group: http://www.update-software.com/Abstracts/SCHIZAbstractIndex.htm

References

1. Feighner, J.P., Robins, E., Guze, S., *et al.* (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, **26**, 57–62.
2. Spitzer, R.L., Endicott, J., and Robins, E. (1977). *Research diagnostic criteria for a selected group of functional disorders* (3rd edn). Biometrics Research Division, New York State Psychiatric Institution, New York.
3. World Health Organization. (1992). F20-F29 schizophrenia, schizotypal and delusional disorders. *The ICD-10 classification of mental and behavioural disorders. Clinical descriptions and diagnostic guidelines*, pp. 97–109. WHO, Geneva.
4. American Psychiatric Association. (1994). Schizophrenia and other psychotic disorders. *Diagnostic and statistical manual of mental disorders, DSM-IV* (2nd edn), pp. 284–306. American Psychiatric Association, Washington, DC.
5. David, A.S. and Appleby, L. (1992). Diagnostic criteria in schizophrenia: accentuate the positive. *Schizophrenia Bulletin*, **18**, 551–7.
6. Andreasen, N.C. and Flaum, M. (1991). Schizophrenia: the characteristic symptoms. *Schizophrenia Bulletin*, **17**, 27–39.
7. Wing, J.K., Cooper, J.E., and Sartorius, N. (1974). *Measurement and classification of psychiatric symptoms*. Cambridge University Press, Cambridge.
8. McGuffin, P., Farmer, A., Gottesman, I.I., *et al.* (1984). Twin concordance for operationally defined schizophrenia: confirmation of familiarity and heritability. *Archives of General Psychiatry*, **41**, 541–5.
9. Andreasen, N.C. and Olsen, S. (1982). Negative v. positive schizophrenia: definition and validation. *Archives of General Psychiatry*, **39**, 789–94.
10. Amador, X.F., Kirkpatrick, B., Buchanan, R.W., *et al.* (1999). Stability of the diagnosis of deficit syndrome in schizophrenia. *The American Journal of Psychiatry*, **156**, 637–9.
11. Crow, T.J. (1980). Molecular pathology of schizophrenia: more than one disease process? *British Medical Journal*, **280**, 1–9.
12. Murray, R.M., Lewis, S.W., and Reveley, A.M. (1985). Towards an aetiological classification of schizophrenia. *Lancet*, **i**, 1023–6.
13. Murray, R.M., Sham, P., Van Os, J., *et al.* (2004). A developmental model for similarities and dissimilarities between schizophrenia and bipolar disorder. *Schizophrenia Research*, **71**, 405–16.
14. Wright, I.C., Rabe-Hesketh, S., Woodruff, P.W.R., *et al.* (2000). Meta-analysis of regional brain volumes in schizophrenia. *The American Journal of Psychiatry*, **157**, 16–25.
15. Craddock, N., O'Donovan, M.C., and Owen, M.J. (2006). Genes for schizophrenia and bipolar disorder? Implications for psychiatric nosology. *Schizophrenia Bulletin*, **32**, 9–16.
16. Jones, P. (1997). The early origins of schizophrenia. *British Medical Bulletin*, **53**, 135–55.
17. Malmberg, A., Lewis, G., David, A., *et al.* (1998). Premorbid adjustment and personality in schizophrenia. *The British Journal of Psychiatry*, **172**, 308–13.
18. Strakowski, S.M. (1994). Diagnostic validity of schizophreniform disorder. *The American Journal of Psychiatry*, **151**, 815–24.
19. Yung, A.R., Stanford, C., Cosgrave, E., *et al.* (2006). Testing the ultra high risk (prodromal) criteria for the prediction of psychosis in a clinical sample of young people. *Schizophrenia Research*, **84**, 57–66.
20. Owens, D.G.C. and Johnstone, E.C. (2006). Precursors and prodromata of schizophrenia: findings from the Edinburgh High Risk Study and their literature context. *Psychological Medicine*, **36**, 1501–14.
21. Johnstone, E.C., Crow, T.J., Frith, C.D., *et al.* (1988). The Northwick Park 'functional' psychoses study: diagnosis and treatment response. *Lancet*, **ii**, 119–26.
22. Davison, K. and Bagley, C.R. (1969). Schizophrenia-like psychoses associated with organic disorders of the central nervous system. In *Current problems in neuropsychiatry* (ed. R.N. Herrington). British Journal of Psychiatry Special Publication No. 4. Headley Brothers, Ashford, Kent.
23. Lishman, W.A. (1997). *Organic psychiatry: the psychological consequences of cerebral disorder* (3rd edn). Blackwell Science, Oxford.
24. Hyde, T.M., Ziegler, J.C., and Weinberger, D.R. (1993). Psychiatric disturbances in metachromatic leukodystrophy. Insights into the neurobiology of psychosis. *Archives of Neurology*, **50**, 131.
25. Cutting, J. (1987). The phenomenology of acute organic psychosis: comparison with acute schizophrenia. *The British Journal of Psychiatry*, **151**, 324–32
26. Mendez, M.F., Grau, R., Doss, R.C., *et al.* (1993). Schizophrenia in epilepsy: seizure and psychosis variables. *Neurology*, **43**, 1073–7.

27. Slater, E., Beard, A.W., and Glithero, E. (1963). The schizophrenia-like psychoses of epilepsy. *The British Journal of Psychiatry*, **109**, 95–150.
28. Trimble, M.R. (1990). First-rank symptoms of Schneider. A new perspective? *The British Journal of Psychiatry*, **156**, 195–200.
29. Johnstone, E.C., Cooling, N., Frith, C.D., *et al.* (1988). Phenomenology of organic and functional psychoses and the overlap between them. *The British Journal of Psychiatry*, **153**, 770–6.
30. Johnstone, E.C., Macmillan, J.F., and Crow, T.J. (1987). The occurrence of organic disease of possible or probably aetiological significance in a population of 268 cases of first episode schizophrenia. *Psychological Medicine*, **17**, 371–9.
31. Moore, T.H.M., Zammit, S., Lingford-Hughes, A., *et al.* (2007). Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. *Lancet*, **370**, 319–28.

4.3.5 Epidemiology of schizophrenia

Assen Jablensky

Introduction

Epidemiological research into schizophrenia aims to answer four essential questions.

- What is the 'true' population frequency of the disorder in various populations and how is it distributed across the various groups within populations?
- Do the incidence, manifestations, and course of schizophrenia vary in relation to factors of the physical and social environment?
- Who is at risk and what forces determine or influence the risk of developing schizophrenia?
- Can the answers to the above questions help explain what causes the disorder and how to prevent it?

The hallmark of the epidemiological method (see Chapter 2.7) is the referral of a measure (numerator) of the occurrence of a disorder, or of any associated characteristics, to a population base (denominator), such as **person-years at risk**. The epidemiological study of diseases usually proceeds from a description of its frequency and associations (establishing rates of occurrence) to testing hypotheses about risk factors and causes by analysing ratios between rates.

Schizophrenia has been studied extensively from an epidemiological perspective since Kraepelin[1] introduced the concept of *dementia praecox* in 1896. In the first half of the twentieth century, epidemiological research into schizophrenia took two divergent paths. While European studies tended to focus on population distributions and genetic risks, North American researchers investigated the social ecology of the disorder. A variety of methods were explored and successfully applied by the pioneers of psychiatric epidemiology, and the contours of the epidemiological map of schizophrenia in Europe and North America were effectively laid down between the two World Wars. The early studies were carried out by dedicated researchers who often spent months or years collecting data 'door-to-door' in small communities. Close knowledge of the respondents, access to multigenerational records from the local parish registers, and the cooperation of the community resulted in studies that remain landmarks of psychiatric epidemiology (Table 4.3.5.1).

During the last several decades, the scope of epidemiological studies of schizophrenia has expanded to include populations in Asia, Africa, and South America about which little had been known previously. The World Health Organization (**WHO**) International Pilot Study of Schizophrenia and its successor, the WHO 10-country epidemiological study[9,10] were the first systematic investigations of the comparative incidence, clinical manifestations, and course

Table 4.3.5.1 Historical landmarks in the epidemiology of psychoses

Author	Method	Target population	Case-finding	Assessment
Koller (1895)[2]	The first epidemiological case-control study of psychoses	Probands with psychoses (n = 287) and non-psychiatric controls (n = 370)	Records of psychiatric hospitals and clinics	Genealogical inquiry
Luxenburger (1928)[3]	Twin concordance/discordance analysis; sampling design	Monozygotic and dizygotic twin pairs	Census of inpatients; search of birth registers for twin births	Emphasis on reliability of diagnosis: 'definite' and 'probable'
Brugger (1931)[4]	Census (door-to-door survey)	Area in Thuringia, population 37 561	Records and key informants consulted to detect 'suspected' cases	Personal examination of 'suspected' cases and of a control sample
Klemperer (1933)[5]	Birth cohort study	Random sample (n = 1000) from all births in Munich, 1881–90	Attempted tracing of all cohort members, 44% successfully traced	Personal examination or key informant interview (271 examined)
Ödegaard (1946)[6]	Cumulative national case register	Entire population of Norway	Registration of all first-admissions 1926–35 (n = 14 231)	Statistical analysis of hospital diagnoses and records
Essen-Möller *et al.* (1956);[7] Hagnell (1966)[8]	Census followed by repeated follow-up surveys	Rural community, initial population 2550 (+1013 new residents in the course of follow-up)	Complete census; tracing of migrants	Personal examination (and re-examination) of all residents

of schizophrenia in both developing and developed countries. The WHO programme was an impetus for similar studies in India, China, the Caribbean, and Australia. Two major studies of psychiatric morbidity in the United States, the Epidemiological Catchment Area project,[11] and the National Comorbidity Survey,[12] generated data on the prevalence of DSM-III/IIIR schizophrenia and related disorders in representative population samples. In the 1980s and 1990s, epidemiological studies increasingly utilized existing large databases such as cumulative case registers or birth cohorts to test hypotheses about risk factors, and began to include methods of genetic epidemiology. There is a current tendency towards integrating epidemiological approaches with other types of aetiological research in schizophrenia. This predicts an important role for epidemiology in the era of molecular biology of mental disorders.

Epidemiological methods and instruments in the study of schizophrenia

The measurement of the prevalence, incidence, and disease expectancy of schizophrenia depends critically on the sensitivity of the case-finding method (i.e. its capacity to identify all affected persons in a given population) and the availability of a diagnostic instrument or procedure that selects 'true' cases (i.e. those corresponding to an established clinical concept).

Case-finding

Case-finding designs fall into three broad groups: case detection in clinical populations, door-to-door surveys of population samples or whole communities, and birth cohort studies. Each method has its advantages and limitations.

While case-finding through the mental health services provides a relatively easy access to a substantial proportion of all persons with schizophrenia, the **cases in treatment** may not be fully representative of all individuals with the disorder. Bias related to gender, marital status, socio-economic factors, culture, or ethnicity are known to affect the probability of being in treatment at a given time in a given setting, and generalizations about schizophrenia from hospital or clinic samples are liable to error. Some of the deficiencies of case-finding through service contacts are avoided in cumulative national or regional psychiatric case registers, which cover large well-defined populations and can be linked to other population databases (e.g. birth records). This makes registers efficient research instruments in low-incidence disorders such as schizophrenia.

Surveys involve accounting for every person at risk within a defined community or a population sample in terms of either being or not being a case. Face-to-face interviews (and follow-up) of all residents in defined communities has been a feature of some high-quality research, especially in the Scandinavian countries. However, since the size of the populations surveyed in this way is limited, the number of detected cases of schizophrenia is usually too small to generate stable estimates of epidemiological parameters. Surveys of large populations involve two basic designs: a single-phase survey of a probability sample drawn from the general population, and a two-phase survey where a validated screening test is first applied to the entire population and only those scoring as screen-positive proceed to a full assessment. In the instance of schizophrenia, logistics dictates a choice between assessing large numbers less rigorously and investigating a smaller sample in greater depth. In the absence of a simple and valid screening procedure for schizophrenia, such as a biological or psychological test, the advantages of the two-phase survey may be offset by poor sensitivity or specificity of the screening device which is usually a questionnaire or checklist.

The study of **birth cohorts** at ages when their members have passed through the greater part of the period of risk for onset of schizophrenia is usually done by direct interviewing or by analysing available case register data. Well-characterized birth cohorts are among the best tools for the study of the incidence of schizophrenia and associated risk factors. However, even in settings where the population is stable and mortality and morbidity are adequately monitored, the size of birth cohorts with prospectively collected data may not be sufficient for conclusive epidemiological inferences.

All this suggests that there is no single 'gold standard' of case-finding for schizophrenia that could be applied across all possible settings, and the assets and liabilities of particular case-finding procedures need to be evaluated in the context of each study. This makes the detailed reporting of case-finding methods a mandatory prerequisite for an 'evidence-based' epidemiology of schizophrenia.

Diagnosis

Variation in diagnostic concepts and practices always explains a proportion of the variation in the results of schizophrenia studies, especially if they involve different populations or different periods. Until the 1960s, the diagnostic rules used in epidemiological research were seldom explicitly stated. In the late 1960s, the WHO International Pilot Study of Schizophrenia[10] examined diagnostic variation in schizophrenia across nine countries by comparing the diagnoses made by psychiatrists using a semi-structured clinical interview with diagnostic classification by a computer algorithm[13] utilizing the same interview data. The results demonstrated that in the majority of settings psychiatrists were using comparable diagnostic concepts in the Kraepelin–Bleuler tradition. The introduction of explicit diagnostic criteria and rules with the consecutive editions of **DSM** and the WHO's **ICD-10** improved further the reliability of diagnosis but did not resolve all diagnostic issues with implications for epidemiology. While ICD-10 and DSM-IV tend to agree well on the core cases of schizophrenia, they agree less well on the classification of atypical or milder cases. Such differences may be less important in clinical practice but they present a problem for epidemiological and genetic studies. By providing more restrictive criteria for schizophrenia, both classifications aim to identify clinically similar cases and to minimize false-positive diagnoses. This is not an unequivocal advantage for epidemiology. Applying such criteria at case-finding may result in the rejection of potential cases which fail to meet the full set of criteria at initial assessment. Therefore it is desirable to develop less restrictive screening versions of the DSM and ICD criteria for epidemiological research.

Instruments

The diagnostic instruments used in surveys which involve interviewing fall into two categories: fully structured interviews such as

the Diagnostic Interview Schedule (**DIS**)[12] and the Composite International Diagnostic Interview (**CIDI**)[14] both written to match exactly the diagnostic criteria of DSM-IIIR/IV and ICD-10, and semi-structured interview schedules such as the Present State Examination (**PSE**)[13] and the Schedules for Clinical Assessment in Neuropsychiatry (**SCAN**),[15] which cover a broad range of psychopathology and elicit data that can be processed by alternative diagnostic algorithms.

The DIS/CIDI type of instrument is reliable and capable of generating standard diagnoses of common mental disorders in a single-phase survey design. Its clinical validity in schizophrenia is less certain because symptoms may not be reported accurately or impairment may be underestimated by the respondent. In contrast, the PSE/SCAN allows a greater amount of psychopathological data to be elicited in a flexible clinical interview format, but its use in epidemiological studies presupposes availability of clinically trained interviewers. While SCAN and other similar interviews are suitable as second-stage diagnostic instruments, there is still a need for a relatively simple and effective screening procedure for case-finding of schizophrenia in field surveys.

Persons, place, time: descriptive epidemiology of schizophrenia

The epidemiological description of schizophrenia draws on extensive evidence available today on its frequency, age, and sex distribution in relatively large populations or geographical areas. Less than complete information is available on variations in its epidemiological characteristics that may be found in unusual or isolated populations, or on the temporal trends in its occurrence.

Prevalence, incidence, and disease expectancy

(a) Prevalence

Prevalence provides an estimate of the number of cases per 1000 persons at risk present in a population at a given time or over a defined period. **Point prevalence** refers to the 'active' (i.e. symptomatic) cases on a given date, or within a brief census period. Since asymptomatic cases (e.g. persons in remission) will be missed in a point prevalence survey, it is useful to supplement the assessment of the present mental state with an enquiry about past episodes of the disorder to obtain a **lifetime prevalence** index. In schizophrenia, which tends to a chronic course, estimates of point and lifetime prevalence will be closer to each other than in remitting illnesses.

An overview of selected prevalence studies of schizophrenia spanning nearly seven decades is presented in Table 4.3.5.2. The studies differ in many aspects of methodology but the majority of them feature a high intensity of case-finding. Several studies are repeat surveys in which the original population was reinvestigated following an interval of 10 or more years (the resulting consecutive prevalence figures are indicated by arrows).

The majority of studies have produced point prevalence estimates in the range 2.1 to 7.0 per 1000 population at risk and lifetime prevalence of schizophrenia in the range 15.0 to 19.0 per 1000. The figures are not uniformly standardized, and should be compared with caution because of demographic differences between populations related to factors such as age-specific mortality and migration. A **systematic review** of 188 studies in 46 countries, published between 1965 and 2002,[16] estimated the median value for point prevalence at 4.6 per 1000 persons and for lifetime prevalence at 7.2 per 1000.

Certain populations and groups deviate markedly from the central tendency. Strikingly high prevalence of schizophrenia (two to three times the national or regional average) has been found in geographically and genetically **isolated populations**, including small communities in Northern Sweden and Finland, and several Western Pacific islands (see Table 4.3.5.2). At the other extreme, a virtual absence of schizophrenia and a high rate of depression have been claimed for the Hutterites of South Dakota, a Protestant sect whose members live in close-knit endogamous communities sheltered from the outside world.[33] Negative social selection for schizoid individuals who fail to adjust to the lifestyle of the majority and eventually migrate without leaving progeny has been suggested (but not definitively proven) as an explanation. Results of two surveys in Taiwan,[21] separated by 15 years, point to a falling prevalence of schizophrenia (from 2.1 to 1.4 per 1000) in the context of major socio-economic change and an overall increase in total mental morbidity in the population.

The question about the extent of true variation in the prevalence of schizophrenia across populations has no simple answer. Methodological differences among studies, related to sampling, case-finding, and diagnostic assessment are likely to account for a good deal of the observed variation. As an example, the high mean prevalence rate of DSM-III schizophrenia reported from the Epidemiologic Catchment Area study in the United States[25] is difficult to reconcile with inconsistencies, such as a 13-fold difference in the rates for age group 18–24 years across the various sites of the survey. One possible reason is that the principal diagnostic instrument of the survey (DIS), administered by lay interviewers, may produce both false-positive and false-negative diagnoses of schizophrenia in a number of cases. Similarly, computer-generated diagnoses of 'non-affective psychosis' in the National Comorbidity Survey,[12] based on a version of the CIDI administered by lay interviewers, were found to agree poorly with clinicians' diagnoses when a subsample of the respondents were re-interviewed over the telephone.[34]

Notwithstanding such caveats in the interpretation of survey findings, the prevalence rates are fairly similar in the majority of studies, though certain specific populations clearly deviate from the modal value. Even in those instances, however, the magnitude of the deviation is modest compared with the 10- to 30-fold differences in prevalence observed in other multifactorial diseases (e.g. diabetes, ischaemic heart disease, multiple sclerosis) across populations.

(b) Incidence

The incidence rate (an estimate of the annual number of first-onset cases in a defined population per 1000 persons at risk) is of greater interest for the study of risk factors than prevalence since it represents the so-called force of morbidity (the probability of disease occurrence) in a given population, and is closer in time to the action of antecedent or precipitating factors. The estimation of incidence depends critically on the ability to determine reliably the point of **onset** of the disorder. In the case of schizophrenia, the long prodromal period and the fuzzy boundary between premorbid state and onset of psychosis make this particularly difficult. In the absence of an objective biomarker of the disease, onset is usually defined as

Table 4.3.5.2 Selected prevalence studies of schizophrenia

Author	Country	Population	Method	Prevalence per 1000 population at risk
Brugger (1931)[4]	Germany	Area in Thuringia (*n* =37 561); age 10+	Census; interview of sample	2.4
Strömgren (1938)[17]; Bøjholm and Strömgren (1989)[18]	Denmark	Island population (*n* = 50 000)	Census interviews; repeat census	3.9→3..3
Böök (1953);[19] Böök *et al.* (1978)[20]	Sweden	Genetic isolate (*n* = 9000); age 15–50	Census interviews; repeat census	9.5→17.0
Essen-Möller *et al.* (1956);[7] Hagnell (1966)[8]	Sweden	Community in Southern Sweden	Census interviews; repeat census	6.7→4.5
Lin *et al.* (1989)[21]	Taiwan	Population sample	Census interviews; repeat census	2.1→1.4
Crocetti *et al.* (1971)[22]	Croatia	Sample of 9201 households	Census based on hospital records and interviews	5.9
Dube and Kumar (1972)[23]	India	Four areas in Agra (*n* = 29 468)	Census based on hospital and clinic records	2.6
Rotstein (1977)[24]	Russia	Population sample (*n* = 35 590)	Census based on hospital and clinic records	3.8
Keith *et al.* (1991)[25]	USA	Aggregated data across five ECA sites	Sample survey; interviews	7.0 (point) 15.0 (lifetime)
Jeffreys *et al.* (1997)[26]	UK	London health district (*n* = 112 127)	Census; interview of sample (*n* = 172)	5.1
Kebede *et al.* (1999)[27]	Ethiopia	25 districts of Addis Ababa (*n* = 2 228 490)	Screening by self-report questionnaire, interviews of sample (*n* = 2042)	7.0 (point) 9.0 (lifetime)
Jablensky *et al.* (2000)[28]	Australia	Four urban areas (*n* = 1 084 978)	Census, screen for psychosis; interviews of sample (*n* = 980)	3.1–5.9 (point)[a] 3.9–6.9 (period, one year)[b]
Waldo *et al.* (1999)[29]	Micronesia	Island of Kosrae Genetic isolate	Screen of hospital records, interviews	6.8 (point)
Arajärvi *et al.* (2005)[30]	Finland	Birth cohort (*n* = 14 817) Genetic isolate	Case register data; interviews of 55% of register cases	15.0 (lifetime) 19.0[c] (lifetime)
Wu *et al.* (2006)[31]	USA (California)	Medicaid/Medicare health insurance data	20% random sample of insured subjects	5.1 (period, 1 year)
Perälä *et al.* (2007)[32]	Finland	National sample (*n* = 8028)	Screen for psychosis, interviews of sample; register and case note data also used	10.0 (lifetime) 22.9[d] (lifetime)

[a]All psychoses.

[b]Schizophrenia and other non-affective psychotic disorders.

[c]Schizophrenia spectrum disorders.

[d]Non-affective psychotic disorders.

the point in time when clinical manifestations become recognizable and diagnosable according to specified criteria. The first hospital admission, which has been used as a proxy for disease onset in many studies, is not a robust indicator because of the variable time lag between the earliest appearance of symptoms and the first-admission across treatment facilities and settings. A better approximation is provided by the first-contact, i.e. the point at which any psychiatric, general medical, or alternative 'helping' agency is accessed by symptomatic individuals for the first time. A limitation common to both first-admission and first-contact studies is that they produce rates of 'treated' incidence and miss symptomatic cases that do not present for assessment or treatment. This limitation can be overcome by periodically repeated door-to-door surveys of the same population or by longitudinal cohort studies (though both are difficult to mount for reasons of cost and logistics).

Table 4.3.5.3 summarizes the essential features of 12 selected incidence studies of schizophrenia. Studies using a 'broad' definition of schizophrenia (ICD-8 or ICD-9) estimate about three-fold difference in the variation of rates, in the range from 0.17 to 0.57 per 1000 population per year, for first-admissions or first contacts. Studies using more stringent criteria, such as the Research Diagnostic Criteria (**RDC**),[121] DSM-IV, ICD-10, or **Catego S+**,[13] have reported incidence rates two to three times lower than those based on 'broad' criteria. A **systematic review** of data from some 160 studies from 33 countries, published between 1965 and 2001,[35] yielded a median value of 0.15 and mean value of 0.24 per 1000, with a five-fold range of the rates and a tendency for more recent studies to report lower rates.

Considering the methodological differences among individual studies, generalizing about the incidence of schizophrenia from pooled data may be problematic. To date, the only investigation

Table 4.3.5.3 Selected incidence studies of schizophrenia

Author	Country	Population	Method	Rate per 1000
Ödegaard (1946)[6]	Norway	Total population	First-admissions 1926–35 (n = 14 231)	0.24 (Hospital diagnoses)
Walsh (1969)[36]	Ireland	City of Dublin (n = 720 000)	First-admissions	0.57 (males, ICD-8); 0.46 (females, ICD-8)
Murphy and Raman (1971)[37]	Mauritius	Total population (n = 257 000)	First-admissions	0.24 (Africans); 0.14 (Indian Hindus); 0.09 (Indian Moslems)
Lieberman (1974)[38]	Russia	Moscow district (n = 248 000)	Follow-back of prevalent cases	0.20 (males) 0.19 (females)
Helgason (1977)[39]	Iceland	Total population	First-admissions (case register)	0.27 (ICD-8)
Lin *et al.* (1989)[21]	Taiwan	Three communities (n = 39 024)	Door-to-door survey	0.17 ('Bleulerian' criteria)
Castle *et al.* (1991)[40]	UK	London (Camberwell)	First-admissions (case register)	0.25 (ICD-9); 0.17 (RDC); 0.08 (DSM-III)
Rajkumar *et al.* (1993)[41]	India	Area in Madras (n = 43 097)	Door-to-door survey and key informants	0.41 (ICD-9)
Wig *et al.* (1993)[42]	India	A rural area (n = 1 036 868) and an urban area (n = 348 609) in Northern India	Case-to-case finding and key informants	0.38 (urban, ICD-9); 0.09 (urban, Catego S+); 0.44 (rural, ICD-9); 0.12 (rural, Catego S+)
Brewin *et al.* (1997)[43]	UK	Nottingham	Two cohorts of first contacts (1978–80 and 1992–94)	0.25→0.29 (All psychoses, ICD-10); 0.14→0.09 (ICD-10 schizophrenia)
Mahy *et al.* (1999)[44]	Barbados	Total population (n = 262 000)	First contacts; PSE interviews; Catego	(0.32 ICD-9); (0.28 Catego S+)
Bresnahan *et al.* (2000)[45]	USA (California)	Birth cohort (n = 12 094)	Case register study; cumulative risk by age 38	0.93 (males, DSM-IV) 0.35 (females, DSM-IV)

that has applied a uniform design and common research tools to generate directly comparable incidence data for different populations is the **WHO 10-country study**.[9] Incidence counts in the WHO study were based on first-in-lifetime contacts with any 'helping agency' within defined areas (including traditional healers in the developing countries) which were monitored over a 2-year period. Potential cases and key informants were interviewed by clinicians using standardized instruments, and the timing of onset was ascertained for the majority of the patients. In 86 per cent of the 1022 patients the onset of diagnostic symptoms of schizophrenia was within the year preceding the first-contact, and therefore the first-contact incidence rate was adopted as a reasonable approximation to the 'true' onset rate. Two definitions of 'caseness', differing in the degree of specificity, were used to determine incidence: a 'broad' clinical definition comprising ICD-9 schizophrenia and paranoid psychoses, and a more restrictive definition of PSE/Catego S+[13] 'nuclear' schizophrenia manifesting with Schneiderian **first-rank symptoms**. The rates for eight of the catchment areas are shown in Table 4.3.5.4.

The differences between the area rates for 'broadly' defined schizophrenia (0.16–0.42 per 1000) were significant ($p < 0.001$) but those for 'nuclear' schizophrenia were not, suggesting that the frequency of this diagnostic subgroup varies less across different populations. No differences were found between cases meeting only 'broad' ICD-9 criteria and the Catego S+ cases with regard to age at onset, or 2-year course and outcome. Therefore it is unlikely that 'nuclear' and 'broad' schizophrenia define two different clinical illnesses.

Replications of the design of the WHO 10-country study, including its research procedures and instruments, have been carried out with very similar results in India, the Caribbean, and the United Kingdom (Table 4.3.5.3).

(c) Disease expectancy (morbid risk)

This is the probability (expressed as a percentage) that an individual born into a particular population will develop the disease if he or she survives the period of risk for that disease. In the instance of schizophrenia the **period of risk** is usually defined as 15 to 54 years. If age- and sex-specific incidence rates are known, disease expectancy can be estimated directly by a summation of the age-specific rates within the period of risk. Alternatively, disease expectancy can be estimated indirectly from prevalence data.

The estimates of disease expectancy produced by a number of studies are fairly consistent across populations and over time. Excluding outliers, such as the northern Swedish isolate,[19,20] they vary about five-fold; in the WHO study, they range from 0.59 per cent (Aarhus) to 1.8 per cent (Chandigarh, rural area) for ICD-9 schizophrenia and from 0.26 per cent (Honolulu) to 0.54 per cent (Nottingham) for Catego S+ 'nuclear' schizophrenia. The frequently cited modal estimate of lifetime disease expectancy for broadly defined schizophrenia at around 1 per cent seems to be consistent with the evidence.

(d) Associations with age and sex

Schizophrenia may have its onset at any age—in childhood as well as past middle age—although the vast majority of onsets fall within the 15 to 54 years of age interval. Onsets in men peak steeply in the

Table 4.3.5.4 Incidence rates per 1000 population, age 15–54, for a 'broad' and a 'narrow' case definition of schizophrenia (WHO 10-country study)

Country	Area	'Broad' definition (ICD-9)			'Narrow' definition (CATEGO S+)		
		Male	Female	All	Male	Female	All
Denmark	Aarhus	0.18	0.13	0.16	0.09	0.05	0.07
India	Chandigarh (rural area)	0.37	0.48	0.42	0.13	0.09	0.11
	Chandigarh (urban area)	0.34	0.35	0.35	0.08	0.11	0.09
Ireland	Dublin	0.23	0.21	0.22	0.10	0.08	0.09
Japan	Nagasaki	0.23	0.18	0.20	0.11	0.09	0.10
Russia	Moscow	0.25	0.31	0.28	0.03	0.03	0.02
United Kingdom	Nottingham	0.28	0.15	0.22	0.17	0.12	0.14
United States of America	Honolulu	0.18	0.14	0.16	0.10	0.08	0.09

(Taken from Report of the international pilot study of schizophrenia, WHO 10-country study, © World Health Organization, www.who.int)

age group 20 to 24 years; thereafter the rate of inception remains more or less constant at a lower level. In women, a less prominent peak in the age group 20 to 24 years is followed by another increase in incidence in age groups older than 35. While the age-specific incidence up to the mid-thirties is significantly higher in men, the male-to-female ratio becomes inverted with age, reaching 1:1.9 for onsets after age 40 and 1:4 or even 1:6 for onsets after age 60. There seems to be no real 'point of rarity' between the symptomatology of late-onset schizophrenia and schizophrenia of an early onset.

The sex differences in mean age at onset are unlikely to be an invariant biological characteristic of schizophrenia. For example, within families carrying high-genetic risk (two or more affected members), no significant differences in age at onset have been found between male and female siblings with schizophrenia. In some populations (e.g. India and China) the male predominance in the frequency of onsets in the younger age groups is attenuated or even inverted.[46,47]

The question of whether the total lifetime risks for men and women are about the same, or different, has not been answered definitively. In the WHO 10-country study, the cumulated risks for males and females up to the of age 54 were found to be approximately equal. Scandinavian studies which followed up population cohorts into very old age (over 80) reported a higher cumulated lifetime risk in women than in men.[48]

Male–female differences have been described in relation to the premorbid history (better premorbid functioning in women), the occurrence of brain abnormalities (more frequent in men), course (a higher percentage of remitting illness episodes and shorter hospital stay in women), and outcome (higher survival rate in the community, less disability in women). However, there is no unequivocal evidence of consistent sex differences in the symptom profiles of schizophrenia, including the frequency of positive and negative symptoms. Generally, the sex differences described in schizophrenia are more likely to result from normal sexual dimorphism in brain development, as well as from gender-related social roles, rather than from sex-specific aetiological factors.

Fertility, mortality, and comorbidity

(a) Fertility

Earlier studies reported low fertility in both men and women diagnosed with schizophrenia. The mean number of children fathered by men with schizophrenia in Sweden was 0.9, and the average number of live births over the entire reproductive period of women treated for schizophrenia in Norway between 1936 and 1975 was 1.8, compared with 2.2 for the general female population.[49] Yet this phenomenon is neither universal nor consistent over time. In the WHO 10-country study,[9] the fertility of women with schizophrenia in India did not differ from that of women in the general population within the same age groups and geographic areas. Although men with schizophrenia continue to be reproductively disadvantaged, the fertility of women with schizophrenia has increased over the last decades and this trend is likely to be sustained as a result of deinstitutionalization and the greater number of people with mental disorders being able to live in the community.

(b) Mortality

Excess mortality associated with schizophrenia has been well documented by epidemiological studies on large cohorts. National case register data for Norway, 1926–1941 and 1950–1974, indicate that, while the total mortality of psychiatric patients was decreasing, the relative mortality of patients with schizophrenia remained unchanged at a level higher than twice that of the general population.[6] Similar findings have been reported from other European countries and North America, with standardized mortality ratios of 2:6 or higher for patients with schizophrenia, which corresponds to about 20 per cent reduction in life expectancy. A **meta-analysis** of 18 studies[50] estimated a crude mortality rate of 189 deaths per 10 000 population per year and a 10-year survival rate of 81 per cent. Mortality among males was significantly higher than among females, and the difference was primarily due to an excess in suicides and accidents. Unnatural causes apart, the leading causes of death among schizophrenia patients are similar to those in the general population, with the exception of a significantly lower than expected cancer morbidity and mortality, especially for tobacco-related malignancies in males with schizophrenia.[51] This puzzling phenomenon has been replicated by several case register and record linkage studies[52,53] and does not appear to be an artifact. Its causes remain unknown, though protective effects of both genes and antipsychotic pharmacological agents have been proposed.

The single most common cause of death among schizophrenia patients at present is **suicide** (aggregated standardized mortality ratios 9.6 for males and 6.8 for females) which accounts for 28 per cent

of the excess mortality in schizophrenia.[54] The suicide rate in schizophrenia patients is at least equal to, or may indeed be higher, than the suicide rate in major depression. In China, the relative risk of suicide in individuals with schizophrenia compared to those without has been estimated at 23.8.[47] Several risk factors, relatively specific to schizophrenia, have been suggested: being young and male, experiencing multiple relapses and remissions, comorbid substance use, awareness of the deteriorating course of the condition, and loss of faith in treatment. Data from successive patient cohorts in Denmark,[55] United Kingdom,[56] and Australia[57] suggest an alarming trend of increasing mortality in first-admission patients with schizophrenia. In the Danish study,[55] the 5-year cumulated standardized mortality ratios increased from 5.30 (males) and 2.27 (females) between 1971 and 1973 to 7.79 (males) and 4.52 (females) between 1980 and 1982. Particularly striking was the standardized mortality ratio of 16.4 for males with schizophrenia in the first year after diagnosis. In the Australian study,[57] suicide risk was highest in the first 7 days after discharge from inpatient care. These trends seem to parallel the significant reductions in the number of psychiatric beds. Whether the increases in suicide mortality are associated with the shift in the management of schizophrenia from hospital to community care remains to be established.

(c) Comorbidity: physical disease

There is significant comorbidity in schizophrenia, comprising: (i) common medical problems and diseases that affect schizophrenia patients more frequently than attributable to chance; and (ii) certain rare conditions or abnormalities which tend to co-occur with the disorder.

Physical disease is common but tends to be seriously undetected and underdiagnosed. Between 46 per cent and 80 per cent of inpatients with schizophrenia, and between 20 per cent and 43 per cent of outpatients, have been found in different surveys to have concurrent medical illnesses.[58] Persons with schizophrenia, and especially those who are homeless or injection drug users, are at increased risk for potentially life-threatening **communicable diseases**, such as HIV/AIDS, hepatitis C, and tuberculosis.[59,60] Among the chronic non-communicable diseases, patients with schizophrenia have significantly higher than expected rates of epilepsy, diabetes, arteriosclerosis, and ischaemic heart disease.[61–63] Obesity and the concomitant **metabolic syndrome** involving insulin resistance are becoming increasingly common problems in schizophrenia patients.[64] Although a high incidence of **diabetes** in schizophrenia patients had been described long before the introduction of neuroleptic treatment, a contributing role for some of the second-generation antipsychotic agents has not been ruled out.

Some rare genetic or idiopathic disorders, such as metachromatic leucodystrophy, acute intermittent porphyria, and coeliac disease, as well as dysmorphic features such as high-steepled palate, malformed ears and other minor physical anomalies have also been reported to co-occur with schizophrenia.[65,66] On the other hand, several studies have found a lower than expected rate of rheumatoid arthritis in schizophrenia patients.[67]

(d) Comorbidity: substance abuse

Substance abuse is at present by far the most common associated health problem among patients with schizophrenia[68] and may involve any drug of abuse or a polydrug combination. It seems, however, that the addictive use of cannabis, stimulants, and nicotine is disproportionately high among schizophrenia patients and may be linked to the underlying neurobiology of the disorder.[69,70] In a nationwide sample of patients with psychotic disorders in Australia,[28] a lifetime diagnosis of comorbid drug abuse, or dependence was made in 36.3 per cent of males and 15.7 per cent of females with schizophrenia (compared to 3.1 per cent and 1.3 per cent respectively in the general population). In addition to poor prognosis of schizophrenia in patients with heavy cannabis use,[71] a **systematic review** of published data on **cannabis** exposure and the onset of schizophrenia[72] concluded that early use increased the risk of psychosis in a dose-related manner, especially in persons at high genetic risk of schizophrenia. Similarly, **stimulants** tend to exacerbate acute psychotic symptoms in over 50 per cent of schizophrenia patients.[73] The prevalence of cigarette **smoking** among schizophrenia patients is, on the average, two to three times higher than in the general population,[74] but the evidence regarding any adverse effects of nicotine use on the onset and course of schizophrenia is equivocal. A population cohort study[75] found that smoking at ages 18–20 was associated with a lower risk of schizophrenia in later life and could have a specific neuroprotective effect independent of its overall harmful impact on health.

Geographical and cultural variation

To date, no population or culture has been identified in which schizophrenic illnesses do not occur. Also, there is no strong evidence that the incidence of schizophrenia is either uniform, or varies widely across populations, provided that the populations being compared are large enough to minimize the effects of small-area variation. The evidence that specific **psychosocial** or cultural factors play an aetiological role in schizophrenia is also inconsistent.[76] However, there are well-replicated findings of variations in the course and outcome of schizophrenia across populations and cultures that involve, above all, a higher rate of symptomatic recovery and a lower rate of social deterioration in traditional rural communities. Data supporting this conclusion were provided by the WHO studies[9] which found a higher proportion of recovering or improving patients in developing countries such as India and Nigeria than in the developed countries. Sampling bias (e.g. a higher percentage of acute-onset schizophreniform illnesses of good prognosis among Third World patients) was not a likely explanation. A better outcome in the developing countries was found in patients with various modes of onset, and the initial symptoms of the disorder did not distinguish good-outcome from poor-outcome cases. What causes such differences in the prognosis of schizophrenia remains largely unknown. The follow-up in the WHO studies demonstrated that the outcome of paranoid psychoses and affective disorders was also better in the developing countries. Such a general effect on the outcome of psychiatric disorders may result from psychosocial factors, such as availability of social support networks, non-stigmatizing beliefs about mental illness, and positive expectations during the early stages of psychotic illness, as well as from unknown genetic or ecological (including nutritional) factors influencing brain development.

The disease and disability burden of schizophrenia

According to WHO estimates[77,78] no less than 25 per cent of the total 'burden of disease' in the established market economies is

at present attributable to neuropsychiatric conditions. Measured as proportion of the disability-adjusted life-years (**DALYs**) lost, schizophrenia, bipolar affective disorder, and major depression together account for 10.8 per cent of the total, i.e. they inflict on most communities losses that are comparable to those due to cancer (15 per cent) and higher than the losses due to ischaemic heart disease (9 per cent).

An epidemiological perspective on risk factors and antecedents

Studies on clinical samples suggest a great variety of putative risk factors in schizophrenia. As clinical samples are rarely representative and often vulnerable to bias, epidemiological evidence helps in evaluating the significance of such conjectures. Genetic and environmental risk factors are considered further in Chapter 4.3.6.1.

Genetic risk: necessary and sufficient?

Family aggregation of schizophrenia is at present the only epidemiologically well-established risk factor for the disorder, with a relative risk for first-degree relatives of persons with schizophrenia in the range from 9 to 18. Allowing for diagnostic variation, the risk estimates generated by different studies are similar and suggest a general pattern of descending risk as the proportions of shared genes between any two individuals decrease.[79,80] Although **heritability** (commonly estimated at about 80 per cent) provides the basis for the search of specific genes and gene networks involved in schizophrenia causation, the extent to which genetic vulnerability alone is necessary and sufficient to produce the disorder remains unclear. While an environmental contribution to the aetiology of schizophrenia is highly plausible, the evidence in support of it is inferential, typically proceeding from the observation that the concordance for schizophrenia in monozygotic twins (sharing 100 per cent of their genes) is only about 50 per cent. The majority of investigators now agree that genes and environments should be studied jointly and three models of conjunction have been proposed[81]:

- The effects of predisposing genes and environmental factors are additive and increase the risk of disease in a linear fashion;
- Genes modulate the sensitivity of the brain to environmental insults;
- By fostering certain personality traits and associated behaviour, genes influence the likelihood of an individual's exposure to stressful environments.

Epidemiological research into possible environmental contributions to the causation of schizophrenia focuses on three main areas: pre- and perinatal brain damage, factors affecting neurodevelopment from infancy to late adolescence, and factors of the social and urban ecology. (See also Chapter 4.3.6.1)

Factors maintaining the incidence of schizophrenia in populations

Since the first epidemiological study on the reproduction patterns of people with psychoses,[82] reduced fertility among individuals with schizophrenia has been documented by numerous investigators. Coupled with the evidence that the lifetime risk of the disorder (about 1 per cent) is similar across populations and remains stable over time, the question about factors that sustain the incidence of schizophrenia despite a reduced **reproductive fitness**. An early hypothesis was proposed in 1964 by Huxley *et al.*[83] who argued that the high frequency of schizophrenia was evidence of 'genetic morphism' (a balanced polymorphism) whereby the low fertility of affected individuals could be compensated for by a higher than average fertility of clinically unaffected 'cryptoschizophrenic carriers' who possessed some selective advantage, e.g. resistance to shock, autoimmune disease, or infection. However, attempts to demonstrate such advantage in terms of disease resistance, adaptability to extreme environments, or ability and creativity, have been unsuccessful. Importantly, the selective advantage hypothesis assumed that schizophrenia was a single-gene disorder with low penetrance, whereas the majority of investigators today agree that schizophrenia is a **complex polygenic disorder** with incomplete or variable expression of the genotype, and widespread locus and allelic heterogeneity. The polygenic model implies that loss of susceptibility alleles resulting from the lower reproductive fitness of affected individuals would have a negligible effect on the overall gene pool in the population. The more recent hypothesis, that *de novo* germ-line mutations inherited from an ageing father[84] may be responsible for a substantial proportion of incident cases of schizophrenia, is difficult to reconcile with current knowledge that mutation rates for most human genes are within the range of 10^{-6} to 10^{-5} per generation, i.e. their contribution to the maintenance of schizophrenia in the population would be insignificant. Considering that both multiple genes and multiple exogenous factors are likely to be involved in the causation of schizophrenia, neither increased fertility in asymptomatic carriers of the risk genes, nor paternal inheritance of germ-line mutations appear to be necessary or sufficient for the persistence of the disorder.

Environmental insults during early development

(a) Season of birth

A 5 per cent to 8 per cent winter–spring excess of schizophrenic births was first described in 1929[85] and since then reported by numerous studies, mostly in the northern hemisphere (southern hemisphere data are less consistent). Though some of these studies did not have the sample size or statistical design needed to definitively prove or rule out a seasonal effect, **winter–spring births** were associated with a mild but significant increase of the relative risk for schizophrenia (RR = 1.11; CI 1.06–1.18) in a large population cohort from Denmark.[86] Thus, birth seasonality appears to be a robust finding in the epidemiology of schizophrenia,[87] though few biologically plausible and testable causal hypotheses have been advanced to explain it. One of them is the seasonally increased risk of intrauterine exposure to viral infection.

(b) Prenatal exposure to infection

In utero exposure to **influenza** has been implicated as a risk factor since a report that a significant proportion of adult schizophrenia in Helsinki was associated with presumed second-trimester *in utero* exposure to the 1957 A2 influenza epidemic.[88] Numerous studies, attempting to replicate the link between maternal influenza and schizophrenia, have since reached conflicting results, with negative findings reported from an increasing number of studies based on large population samples,[89,90] as well as studies including data on schizophrenia risk in the offspring of women with prospectively

recorded influenza infection during pregnancy.[91] However, positive association between schizophrenia in the offspring and maternal infection during pregnancy has been reported for **rubella**[92] and **toxoplasmosis**[93] and the issue of prenatal infection contributing to schizophrenia risk merits further study.

(c) Pregnancy and birth complications

Maternal obstetric complications, ranging from placental abnormalities in the first trimester of pregnancy to diabetes, pre-eclampsia, perinatal hypoxia, and low birth weight, are widely regarded to be risk factors in schizophrenia. This view is supported by a number of studies of small to moderate size, typically using a case-control design and relying on maternal recall of adverse events during pregnancy.[94] Population-based studies[95,96] using prospectively recorded obstetric data tend to report conflicting or inconclusive results, with generally small effect sizes (odds ratio less than two) for any positive findings.[97] However, several birth cohort studies with long-term follow-up have found significantly increased risk of adult schizophrenia in individuals who had survived severe, mainly **hypoxic perinatal brain damage**.[98,99] Birth weight (adjusted for gestation) is another factor that may have a complex relationship with schizophrenia risk. A large cohort study in Sweden[100] found a reverse J-shaped association between **birth weight** and adult schizophrenia, with significant hazard ratios of 7.03 for males of low birth weight (<2500 g) and 3.37 for those of high birth weight (>4000 g). It remains unclear, however, if severe obstetric complications, such as perinatal hypoxia or low birth weight, are capable of raising substantially the risk of schizophrenia in the adult in the absence of increased genetic risk. Maternal schizophrenia is associated with a higher rate of pregnancy complications, including low birth weight,[101] but it is not known if the effects of genetic liability and obstetric complications on schizophrenia risk in the offspring are additive or interactive. It is also possible that genetic predisposition sensitizes the developing brain to lesions resulting from randomly occurring less severe obstetric complications. Such gaps in knowledge or inconsistencies among research findings caution against an unqualified acceptance of obstetric complication as a proven risk factor in schizophrenia. Clarification of their role remains an important priority for epidemiological research.

Further information about studies of obstetric complications and hypoxic-ischaemic damage as risk factors for schizophrenia can be found in Chapter 4.3.6.1.

Developmental antecedents of schizophrenia

(a) Brain development and neurobehavioural markers

Children at high genetic risk for schizophrenia (i.e. having parents or other first-degree relatives with the disorder) tend to show early signs of aberrant neurodevelopment, including ventricular enlargement on computerized tomography[102] and decreased activation in the prefrontal and parietal regions of the heteromodal association cortex on functional magnetic resonance imaging.[103] Such imaging studies are limited by small sample size and their results may not be generalizable. However, population-based or cohort studies, such as the National Child Development Study in the United Kingdom have demonstrated a higher incidence of abnormal **motor and speech development** before 2 years of age, and of soft neurological signs (poor motor control, coordination, and balance), non-right handedness and speech defects between ages 2–15.[104]

(b) Cognitive and neurophysiological markers

Deficits in verbal memory, sustained attention and executive functions, as well as abnormalities in event-related brain potentials and oculomotor control[105–107] are common in patients with schizophrenia and antedate the onset of clinical symptoms. They also occur in a proportion of their clinically normal biological relatives, but are rare in control subjects drawn from the general population (see Chapter 4.3.3). Their specificity to schizophrenia needs to be investigated in larger population samples. Should such **endophenotypes** be validated as biological markers of schizophrenia by epidemiological studies, the power of risk prediction at the level of the individual is likely to increase substantially.

(c) Premorbid intelligence (IQ)

In a cohort study from Sweden,[108] involving a 15-year follow-up of 109 643 men conscripted into the army at age 18 to 20, the individuals who subsequently developed schizophrenia were compared with the rest of the cohort on the performance of IQ-related tests and tasks at the time of conscription. Controlling for potential confounders, the risk of schizophrenia was found to increase linearly with the decrement of IQ. The effect was mainly attributable to poor performance on verbal tasks and tests of reasoning. Similar results have been reported from a study in Israel linking psychometric assessment data of the army draft board with the national psychiatric case register.[109]

Premorbid social impairment

Individuals who develop schizophrenia as adults are more likely to manifest difficulties in social interaction during childhood and adolescence than individuals who do not develop schizophrenia. Among children at increased genetic risk (having a parent with schizophrenia), those who develop schizophrenia as adults have been found to show poorer social competence at age 7 to 12 and more passivity and social isolation in adolescence, as compared to those who do not develop the disorder.[110] The association between early 'schizoid' traits and risk of schizophrenia is not restricted to offsprings of parents with schizophrenia. Population-based evidence of early **socialization difficulties** (school problems, social anxiety, and preference for solitary play) in children who develop schizophrenia as adults is provided by the prospective study of a national birth cohort in the United Kingdom.[104] In the Swedish conscript study,[108] poor social adjustment during childhood and adolescence was significantly more common among those who subsequently developed schizophrenia than among those who did not. It should be noted, however, that the early behavioural traits that tend to be associated with schizophrenia in adult life have low specificity and their predictive value is limited.

Further information about studies of premorbid social impairment can be found in Chapter 4.3.6.1.

The social and family environment

(a) Early rearing environment

Support for an effect of the early rearing family environment on the risk of developing schizophrenia is provided by a study of a Finnish sample of **adopted children** born to mothers with schizophrenia (a high-risk group) and a control sample of adoptees at no increased genetic risk.[111] Though the rates of adult psychosis

or severe personality disorder were significantly higher in the high-risk group compared with the control group, the difference was entirely attributable to a subset of the high-risk children who grew up in dysfunctional or otherwise disturbed adoptive families—a result consistent with a gene-environment model of genetic influence on a person's sensitivity to psychosocial adversity.

(b) The urban environment

Earlier hypotheses that urban environments increase the risk of psychosis, either by contributing to causation (the breeder effect) or by attracting vulnerable individuals (the drift effect), have been revived in the light of recent epidemiological findings suggesting that urban birth is associated with a moderate but statistically significant increase in the incidence of schizophrenia, affective psychoses, and other non-affective psychoses.[112] It remains unclear whether the effect is linked to a factor operating pre- or perinatally, or a factor influencing postnatal development (see also Chapter 4.3.6.1).

(c) Social class

Since the 1930s, numerous studies in North America and Europe have consistently found that the economically disadvantaged social groups contribute disproportionately to the first-admission rate for schizophrenia. Two explanatory hypotheses, of **social causation** ('breeder') and of **social selection** ('drift'), were originally proposed.[113] According to the social causation theory, the greater socio-economic adversity characteristic of lower-class living conditions could precipitate psychosis in genetically vulnerable individuals who have a restricted capacity to cope with complex or stressful situations. In the 1960s, this theory was considered to be refuted by a single study[114] which found that the social class distribution of the fathers of schizophrenic patients did not deviate from that of the general population, and that the excess of low socio-economic status among schizophrenic patients was mainly attributable to individuals who had drifted down the occupational and social scale prior to the onset of psychosis. As a result, aetiological research in schizophrenia in recent decades has tended to ignore such 'macrosocial' variables. However, the possibility remains that social stratification, socio-economic status, and acculturation stress are contributing factors in the causation of schizophrenia.

(d) Migrants and ethnic minorities

An exceptionally high-incidence rate of schizophrenia (about 6.0 per 1000) has been found in the African–Caribbean population in the United Kingdom.[115,116] The excess morbidity is not restricted to recent immigrants and is higher in the British-born second generation of migrants. Similar findings of nearly four-fold excess over the general population rate have been reported for the Dutch Antillean and Surinamese immigrants in Holland.[117]

The causes of the phenomenon remain obscure. Incidence studies in the Caribbean do not indicate any excess morbidity in the indigenous populations from which migrants are recruited. Explanations in terms of biological risk factors have found little support.[118,122] A finding in need of replication is the significant increase of schizophrenia among the siblings of second-generation African–Caribbean schizophrenia patients compared with the incidence of schizophrenia in the siblings of white patients.[119] Such 'horizontal' increase in the morbid risk suggests that an environmental factor may be modifying the penetrance of the genetic predisposition to schizophrenia carried by a proportion of the African–Caribbean population. Psychosocial hypotheses involving acculturation stress, demoralization due to racial discrimination, and blocked opportunities for upward social mobility have been suggested but not yet definitively tested (see also Chapter 4.3.6.1).

Epidemiological issues for the next decade

The unprecedented growth of basic knowledge about the brain and the human genome opens up novel perspectives and opportunities in the study of complex disorders such as schizophrenia, which integrate concepts and tools of genetics, neuroscience, and epidemiology. Several issues with wide implications for future research are already emerging.

Is schizophrenia a single disease or a group of aetiologically distinct disorders?

Schizophrenia is characterized by extensive phenotypic variability and likely genetic heterogeneity. These two factors may be contributing disproportionately to the multitude of research findings that are inconsistent or difficult to replicate and there is increasing concern that the categorical diagnostic concept of schizophrenia may not demarcate a biologically homogeneous entity.[120] The likely existence of different **subtypes** of the disorder (Bleuler's notion of a 'group of schizophrenias') is rarely considered in genetic and other biological research into schizophrenia. Disaggregating a **complex phenotype** by identifying intermediate (endo-) phenotypes and quantitative traits as covariates has been a successful strategy in the genetic study of disorders such as type I diabetes, asthma, and dementia. While the clinical concept of schizophrenia as a broad syndrome with some internal cohesion and a characteristic course over time is well supported by current epidemiological evidence, its dissection into modular endophenotypes with specific neurocognitive and neurophysiological underpinnings is beginning to be perceived as a promising approach in schizophrenia genetics. The study of endophenotypes cutting across the conventional diagnostic boundaries may reveal unexpected patterns of associations with symptoms, personality traits or behaviours, as well as genetic polymorphisms, providing epidemiology with rich material for hypothesis testing at population level.

Molecular epidemiology of schizophrenia

Notwithstanding the difficulties accompanying the genetic dissection of complex disorders, novel methods of genetic analysis will eventually identify genomic regions, genes, and **interacting gene networks** underlying the predisposition to schizophrenia. The majority of genes involved are believed to be of small effect, although one cannot exclude the possibility that genes of moderate effects may also be found, especially in relation to the neurophysiological abnormalities associated with schizophrenia. Clarifying the function of such genes will be a complex task. Part of the solution is likely to be found in the domain of epidemiology, since establishing their population frequency and associations with a variety of phenotypic expressions is a prerequisite for understanding their causal role. Thus the molecular epidemiology of schizophrenia is likely to be the next major chapter in the search for its causes and cures.

Can schizophrenia be prevented?

The increasing investment in early diagnosis and treatment of **first episodes** of schizophrenia has raised questions whether people likely to develop schizophrenia can be reliably recognized prior to the onset of symptoms, and whether early pharmacological, cognitive, or social intervention can prevent the development of the disorder. While early diagnosis and timely treatment of symptomatic cases may improve the short- or medium-term outcome, the detection of people at risk with a view to preventative intervention is problematic. Screening young age groups in the population by using predictors such as family history of psychosis, obstetric complications, or abnormal eye tracking is likely to result in multiple false-positive and false-negative results and a generally low positive predictive value. Other candidate risk factors have not been evaluated at all epidemiologically. Problems of reliability of measurement apart, population-based screening will pose huge practical and ethical problems of having to treat a large number of individuals who do not have the disorder and missing many others who eventually will develop the disorder. From an epidemiological point of view, pre-symptomatic detection and preventative intervention in schizophrenia do not appear to be feasible for the time being.

Summary and conclusions

After nearly a century of epidemiological research, essential questions about the nature and causes of schizophrenia still await answers. Two major conclusions stand out.

- The clinical syndrome of schizophrenia is robust and can be identified in diverse populations, regardless of wide-ranging demographic, ecological, and cultural differences among them. This suggests that a common pathophysiology is likely to underlie the characteristic symptoms of schizophrenia. On balance, the evidence suggests that schizophrenia incidence and disease risk show relatively modest variation at the level of large population aggregates. However, the study of 'atypical' populations or pockets of very high or very low frequency of schizophrenia, such as in genetic isolates or minority groups, may provide **novel clues** to the aetiology and pathogenesis of disorder.
- No single environmental risk factor of major effect on the incidence of schizophrenia has yet been discovered. Further studies using large samples are required to evaluate potential risk factors, antecedents, and predictors for which the present evidence is inconclusive. Assuming that methodological pitfalls will be avoided by risk-factor epidemiology, and that multiple environmental risk factors of small to moderate effect will eventually be identified, the results will complement those of genetic research which also implicate multiple genes and networks. All this suggests that the key to understanding schizophrenia is likely to be in the unraveling of complex **gene-environment interactions**.

Further information

Susser, E., Schwartz, S., Morabia, A., and Bromet, E. (eds.) (2006). *Psychiatric epidemiology. Searching for the causes of mental disorders*. Oxford University Press, Oxford.

Murray, R.M., Jones, P.B., Susser, E., van Os, J., and Cannon, M. (eds.) (2003). *The epidemiology of schizophrenia*. Cambridge University Press.

Hirsch, S.R. and Weinberger, D.R. (eds.) (2003). *Schizophrenia* (2nd edn). Blackwell Science, Oxford.

www.schizophreniaforum.org Schizophrenia Research Forum (sponsored by NARSAD, the Mental Health Research Association, through a contract with the National Institute of Mental Health).

References

1. Kraepelin, E. (1896). *Psychiatrie. Ein Lehrbuch für Studirende und Aerzte. Barth, Leipzig* (Reprint: edition 1976 by Arno Press Inc., New York).
2. Koller, J. (1895). Beitrag zur Erblichkeitsstatistik der Geisteskrankheiten in Canton Zürich. *Archiv für Psychiatrie*, **27**, 269–94.
3. Luxenburger, H. (1928). Vorläufiger Bericht über psychiatrische Serienuntersuchungen an Zwillingen. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, **116**, 297–326.
4. Brugger, C. (1931). Versuch einer Geisteskrankenzählung in Thüringen. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, **133**, 252–390.
5. Klemperer, J. (1933). Zur Belastungsstatistik der Durchschnittsbevölkerung. Psychosehäufigkeit unter 1000 stichprobenmässig ausgelesenen Probanden. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, **146**, 277–316.
6. Ödegaard, Ö. (1946). A statistical investigation into the incidence of mental disorders in Norway. *The Psychiatric Quarterly*, **20**, 381–401.
7. Essen-Möller, E., Larsson, H., Uddenberg, C.E., *et al.* (1956). Individual traits and morbidity in a Swedish rural population. *Acta Psychiatrica et Neurologica Scandinavica*, (Suppl. 100), 1–136.
8. Hagnell, O. (1966). *A prospective study of the incidence of mental disorder*. Svenska Bokforlaget, Lund.
9. Jablensky, A., Sartorius, N., Ernberg, G., *et al.* (1992). Schizophrenia: manifestations and course in different cultures. A World Health Organization ten-country study. *Psychological Medicine, Monograph Supplement*, **20**, 1–97.
10. World Health Organization. (1973). *The international pilot study of schizophrenia*, Vol. 1. WHO, Geneva.
11. Robins, L.N. and Regier, D.A. (eds.) (1991). *Psychiatric disorders in America. The Epidemiologic Catchment Area Study*. Free Press, New York.
12. Kessler, R.C., McGonagle, K.A., Zhao, S., *et al.* (1994). Lifetime and 12-month prevalence of DSM-IIIR psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Archives of General Psychiatry*, **51**, 8–19.
13. Wing, J.K., Cooper, J.E., and Sartorius, N. (1974). The measurement and classification of psychiatric symptoms. Cambridge University Press, Cambridge.
14. Robins, L.N., Wing, J., Wittchen, H.U., *et al.* (1988). The Composite International Diagnostic Interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry*, **45**, 1069–77.
15. Wing, J.K., Babor, T., Brugha, T., *et al.* (1990). SCAN. Schedules for clinical assessment in neuropsychiatry. *Archives of General Psychiatry*, **47**, 589–93.
16. Saha, S., Chant, D., Welham, J., *et al.* (2005). A systematic review of the prevalence of schizophrenia. *PLoS Medicine*, **2**, 413–33. (www.plosmedicine.org)
17. Strömgren, E. (1938). Beiträge zur psychiatrischen Erblehre, auf Grund von Untersuchungen an einer Inselbevölkerung. *Acta Psychiatrica et Neurologica Scandinavica*, (Suppl. 19), 1–86.
18. Bøjholm, S. and Strömgren, E. (1989). Prevalence of schizophrenia on the island of Bornholm in 1935 and in 1983. *Acta Psychiatrica Scandinavica*, **79**(Suppl. 348), 157–66.
19. Böök, J.A. (1953). A genetic and neuropsychiatric investigation of a North Swedish population (with special regard to schizophrenia and mental deficiency). *Acta Genetica*, **4**, 1–100.

20. Böök, J.A., Wettenberg, L., and Modrzewska, K. (1978). Schizophrenia in a North Swedish geographical isolate, 1900–1977: epidemiology, genetics and biochemistry. *Clinical Genetics*, **14**, 373–94.
21. Lin, T.Y., Chu, H.M., Rin, H., *et al.* (1989). Effects of social change on mental disorders in Taiwan: observations based on a 15-year follow-up survey of general population in three communities. *Acta Psychiatrica Scandinavica*, **79**(Suppl. 348), 11–34.
22. Crocetti, G.J., Lemkau, P.V., Kulcar, Z., *et al.* (1971). Selected aspects of the epidemiology of psychoses in Croatia, Yugoslavia, II. The cluster sample and the results of the pilot survey. *American Journal of Epidemiology*, **94**, 126–34.
23. Dube, K.V. and Kumar, N. (1972). An epidemiological study of schizophrenia. *Journal of Biosocial Science*, **4**, 187–95.
24. Rotstein, V.G. (1977). Material from a psychiatric survey of sample groups from the adult population in several areas of the USSR. *Zhurnal Nevropatologii I Psikhiatrii*, **77**, 569–74 (in Russian).
25. Keith, S.J., Regier, D.A., and Rae, D.S. (1991). Schizophrenic disorders. In *Psychiatric disorders in America. The Epidemiologic Catchment Area Study* (eds. L.N. Robins and D.A. Regier), pp. 33–52. Free Press, New York.
26. Jeffreys, S.E., Harvey, C.A., McNaught, A.S., *et al.* (1997). The Hampstead schizophrenia survey 1991. I. Prevalence and service use comparisons in an inner London health authority, 1986–1991. *The British Journal of Psychiatry*, **170**, 301–6.
27. Kebede, D. and Alem, A. (1999). Major mental disorders in Addis Ababa, Ethiopia. I. Schizophrenia, schizoaffective and cognitive disorders. *Acta Psychiatrica Scandinavica*, **100**, 11–7.
28. Jablensky, A., McGrath, J., Herrman, H., *et al.* (2000). Psychotic disorders in urban areas: an overview of the Study on Low Prevalence Disorders. *The Australian and New Zealand Journal of Psychiatry*, **34**, 221–36.
29. Waldo, M.C. (1999). Schizophrenia in Kosrae, Micronesia: prevalence, gender ratios, and clinical symptomatology. *Schizophrenia Research*, **35**, 175–81.
30. Arajärvi, R., Suvisaari, J, Suokas, J., *et al.* (2005). Prevalence and diagnosis of schizophrenia based on register, case record and interview data in an isolated Finnish birth cohort born 1940–1969. *Social Psychiatry and Psychiatric Epidemiology*, **40**, 808–16.
31. Wu, E.Q., Shi, L., Birnbaum, H., *et al.* (2006). Annual prevalence of diagnosed schizophrenia in the USA: a claims data analysis approach. *Psychological Medicine*, **36**, 1535–40.
32. Perälä, J., Suvisaari, J., Saarni, S.I., *et al.* (2007). Lifetime prevalence of psychotic and bipolar I disorders in a general population. *Archives of General Psychiatry*, **64**, 19–28.
33. Eaton, J.W. and Weil, R.J. (1955). *Culture and mental disorder. A comparative study of the Hutterites and other populations*. Free Press, Glencoe, IL.
34. Kendler, K.S., Gallagher, T.J., Abelson, J.M., *et al.* (1996). Lifetime prevalence, demographic risk factors, and diagnostic validity of nonaffective psychosis as assessed in a US community sample. *Archives of General Psychiatry*, **53**, 1022–31.
35. McGrath, J., Saha, S., Welham, J., *et al.* (2003). A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC Medicine*, **2**, 1–22. (www.biomedcentral.com/1741-7015/2/13)
36. Walsh, D. (1969). Mental illness in Dublin–first admissions. *The British Journal of Psychiatry*, **115**, 449–56.
37. Murphy, H.B.M. and Raman, A.C. (1971). The chronicity of schizophrenia in indigenous tropical peoples. Results of a twelve-year follow-up survey in Mauritius. *The British Journal of Psychiatry*, **118**, 489–97.
38. Lieberman, Y.I. (1974). The problem of incidence of schizophrenia: material from a clinical and epidemiological study. *Zhurnal Nevropatologii I Psikhiatrii*, **74**, 1224–32 (in Russian).
39. Helgason, L. (1977). Psychiatric services and mental illness in Iceland. *Acta Psychiatrica Scandinavica*, **53**(Suppl. 268), 1–140.
40. Castle, D., Wessely, S., Der, G., *et al.* (1991). The incidence of operationally defined schizophrenia in Camberwell, 1965–1984. *The British Journal of Psychiatry*, **159**, 790–4.
41. Rajkumar, S., Padmavati, R., Thara, R., *et al.* (1993). Incidence of schizophrenia in an urban community in Madras. *Indian Journal of Psychiatry*, **35**, 18–21.
42. Wig, N.N., Varma, V.K., Mattoo, S.K., *et al.* (1993). An incidence study of schizophrenia in India. *Indian Journal of Psychiatry*, **35**, 11–7.
43. Brewin, J., Cantwell, R., Dalkin, T., *et al.* (1997). Incidence of schizophrenia in Nottingham. *The British Journal of Psychiatry*,**171**, 140–4.
44. Mahy, G.E., Mallett, R., Leff, J., *et al.* (1999). First-contact incidence rate of schizophrenia on Barbados. *The British Journal of Psychiatry*, **175**, 28–33.
45. Bresnahan, M.A., Brown, A.S., Schaefer, C.A., *et al.* (2000). Incidence and cumulative risk of treated schizophrenia in the prenatal determinants of schizophrenia study. *Schizophrenia Bulletin*, **26**, 297–308.
46. Murthy, G.V.S., Janakiramaiah, N., Gangadhar, B.N., *et al.* (1998). Sex difference in age at onset of schizophrenia: discrepant findings from India. *Acta Psychiatrica Scandinavica*, **97**, 321–5.
47. Phillips, M.R., Yang, G., Li, S., *et al.* (2004). Suicide and the unique prevalence pattern of schizophrenia in mainland China: a retrospective observational study. *Lancet*, **364**, 1062–8.
48. Helgason, T. and Magnusson, H. (1989). The first 80 years of life. A psychiatric epidemiological study. *Acta Psychiatrica Scandinavica*, **79**(Suppl. 348), 85–94.
49. Ödegaard, Ö. (1980). Fertility of psychiatric first admissions in Norway, 1936–1975. *Acta Psychiatrica Scandinavica*, **62**, 212–20.
50. Brown, S. (1997). Excess mortality of schizophrenia. A meta-analysis. *The British Journal of Psychiatry*, **171**, 502–8.
51. Dupont, A., Jensen, O.M., Strömgren, E., *et al.* (1986). Incidence of cancer in patients diagnosed as schizophrenic in Denmark. In *Psychiatric case registers in public health* (eds. S.H. Ten Horn, R. Giel, and W. Gulbinat), pp. 229–39. Elsevier, Amsterdam.
52. Dalton, S.O., Mellemkjær, L., Thomassen, L., *et al.* (2005). Risk for cancer in a cohort of patients hospitalized for schizophrenia in Denmark, 1969–1993. *Schizophrenia Research*, **75**, 315–24.
53. Grinshpoon, A., Barchana, M., Ponizovsky, A., *et al.* (2005). Cancer in schizophrenia: is the risk higher or lower? *Schizophrenia Research*, **73**, 333–41.
54. Heilä, H. and Lönnqvist, J. (2003). The clinical epidemiology of suicide in schizophrenia. In *The epidemiology of schizophrenia* (eds. R.M. Murray, P. Jones, E. Susser, J. van Os, and M. Cannon), pp. 288–314. Cambridge University Press, Cambridge.
55. Mortensen, P.B. and Juel, K. (1993). Mortality and causes of death in first admitted schizophrenic patients. *The British Journal of Psychiatry*, **163**, 183–9.
56. Geddes, J.R. and Juszczak, E. (1995). Period trends in rate of suicide in first 28 days after discharge from psychiatric hospital in Scotland, 1968-92. *British Medical Journal*, **311**, 357–60.
57. Lawrence, D., Holman, C.D.J, Jablensky, A., *et al.* (2001). Increasing rates of suicide in Western Australian psychiatric patients: a record linkage study. *Acta Psychiatrica Scandinavica*, **104**, 443–51.
58. Jeste, D.V., Gladsjo, J.A., Lindamer, L.A., *et al.* (1996). Medical comorbidity in schizophrenia. *Schizophrenia Bulletin*, **22**, 413–30.
59. Susser, E., Valencia, E., and Conover, S. (1993). Prevalence of HIV infection among psychiatric patients in a New York City men's shelter. *American Journal of Public Health*, **83**, 568–70.
60. Martens, W.H. (2001). A review of physical and mental health in homeless persons. *Public Health Reviews*, **29**, 13–33.
61. Mortensen, P.B. (2003). Mortality and physical illness in schizophrenia. In *The epidemiology of schizophrenia* (eds. R.M. Murray, P. Jones, E. Susser, J. van Os, and M. Cannon), pp. 275–87. Cambridge University Press, Cambridge.

62. Lawrence, D.M., Holman, C.D.J., Jablensky, A.V., *et al.* (2003). Death rate from ischaemic heart disease in Western Australian psychiatric patients 1980–1998. *The British Journal of Psychiatry*, **182**, 32–6.
63. Goff, D.C., Sullivan, L.M., McEvoy, J.P., *et al.* (2005). A comparison of ten-year cardiac risk estimates in schizophrenia patients from the CATIE study and matched controls. *Schizophrenia Research*, **80**, 45–53.
64. Saari, K.M., Lindeman, S.M., Viilo, K.M., *et al.* (2005). A 4-fold risk of metabolic syndrome in patients with schizophrenia: the Northern Finland 1966 birth cohort study. *The Journal of Clinical Psychiatry*, **66**, 559–63.
65. Hyde, T.M., Ziegler, J.C., and Weinberger, D. (1992). Psychiatric disturbances in metachromatic leucodystrophy. *Archives of Neurology*, **49**, 401–6.
66. Murphy, K.C. and Owen, M.J. (1996). Minor physical anomalies and their relationship to the aetiology of schizophrenia. *The British Journal of Psychiatry*, **168**, 139–42.
67. Eaton, W.M., Hayward, C., and Ram, R. (1992). Schizophrenia and rheumatoid arthritis: a review. *Schizophrenia Research*, **6**, 181–92.
68. Murray, R.M., Grech, A., Phillips, P., *et al.* (2003). What is the relationship between substance abuse and schizophrenia? In *The epidemiology of schizophrenia* (eds. R.M. Murray, P. Jones, E. Susser, J. van Os, and M. Cannon), pp. 317–42. Cambridge University Press, Cambridge.
69. Koethe, D., Llenos, I.C., Dulay, J.R., *et al.* (2007). Expression of CB1 cannabinoid receptor in the anterior cingulate cortex in schizophrenia, bipolar disorder, and major depression. *Journal of Neural Transmission*, **114**, 1055–63.
70. Myers, C.S., Robles, O., Kakoyannis, N., *et al.* (2004). Nicotine improves delayed recognition in schizophrenic patients. *Psychopharmacology*, **174**, 334–40.
71. Henquet, C., Murray, R., Linszen, D., *et al.* (2005). The environment and schizophrenia: the role of cannabis use. *Schizophrenia Bulletin*, **31**, 608–12.
72. Semple, D.M., McIntosh, A.M., and Lawrie, S.M. (2005). Cannabis as a risk factor for psychosis: systematic review. *Journal of Psychopharmacology*, **19**, 187–94.
73. Curran, C., Byrappa, N., and McBride, A. (2004). Stimulant psychosis: systematic review. *The British Journal of Psychiatry*, **185**, 196–204.
74. De Leon, J. and Diza, F.J. (2005). A meta-analysis of worldwide studies demonstrates an association between schizophrenia and tobacco smoking behaviors. *Schizophrenia Research*, **76**, 135–57.
75. Zammitt, S., Allebeck, P., Dalman, C., *et al.* (2003). Investigating the association between cigarette smoking and schizophrenia in a cohort study. *The American Journal of Psychiatry*, **160**, 2216–21.
76. Boydell, J. and Murray, R. (2003). Urbanization, migration and risk of schizophrenia. In *The epidemiology of schizophrenia* (eds. R.M. Murray, P. Jones, E. Susser, J. van Os, and M. Cannon), pp. 49–67. Cambridge University Press, Cambridge.
77. Murray, C.J.L. and Lopez, A.D. (eds.) (1996). *The global burden of disease. A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020.* Harvard School of Public Health, Cambridge, MA.
78. Jablensky, A. (2000). Epidemiology of schizophrenia: the global burden of disease and disability. *European Archives of Psychiatry and Clinical Neuroscience*, **250**, 274–85.
79. Gottesman, I.I. (1991). *Schizophrenia genesis: the origins of madness.* W.H. Freeman, New York.
80. Lichtenstein, P., Björk, C., Hultman, C.M., *et al.* (2006). Recurrence risks for schizophrenia in a Swedish National Cohort. *Psychological Medicine*, **36**, 1417–25.
81. Kendler, K.S. and Eaves, L.J. (1986). Models for the joint effect of genotype and environment on liability to psychiatric illness. *The American Journal of Psychiatry*, **143**, 279–89.
82. Essen-Möller, E. (1935). Untersuchungen über die Fruchtbarkeit gewisser Gruppen von Geisteskranken. *Acta Psychiatrica et Neurologica*, **8**, 1–314.
83. Huxley, J., Mayr, E., Osmond, H., *et al.* (1964). Schizophrenia as a genetic morphism. *Nature*, **204**, 220–1.
84. Sipos, A., Rasmussen F., Harrison, G., *et al.* (2004). Paternal age and schizophrenia: a population based cohort study. *British Medical Journal*, **330**, 147–8.
85. Tramer, M. (1929). Über die biologische Bedeutung des Geburtsmonates, insbesondere für die Psychoseerkrankung. *Schweizerischer Archiv für Neurologie und Psychiatrie*, **24**, 17–24.
86. Mortensen, P.B., Pedersen, C.B., Westergaard, T., *et al.* (1999). Effects of family history and place and season of birth on the risk of schizophrenia. *The New England Journal of Medicine*, **340**, 603–8.
87. Torrey, E.F., Miller, J., Rawlings, R., *et al.* (1997). Seasonality of births in schizophrenia and bipolar disorder: a review of the literature. *Schizophrenia Research*, **28**, 1–38.
88. Mednick, S.A., Machon, R.A., Huttunen, M.O., *et al.* (1988). Adult schizophrenia following prenatal exposure to an influenza epidemic. *Archives of General Psychiatry*, **45**, 189–92.
89. Morgan, V., Castle, D., Page, A., *et al.* (1997). Influenza epidemics and incidence of schizophrenia, affective disorders and mental retardation in Western Australia: no evidence of a major effect. *Schizophrenia Research*, **26**, 25–39.
90. Grech, A., Takei, N., and Murray, R.M. (1997). Maternal exposure to influenza and paranoid schizophrenia. *Schizophrenia Research*, **26**, 121–5.
91. Cannon, M., Cotter, D., Coffey, V.P., *et al.* (1996). Prenatal exposure to the 1957 influenza epidemic and adult schizophrenia: a follow-up study. *The British Journal of Psychiatry*, **168**, 368–71.
92. Brown, A.S., Cohen, P., Harkavy-Friedman, J., *et al.* (2001). Prenatal rubella, premorbid abnormalities, and adult schizophrenia. *Biological Psychiatry*, **49**, 473–86.
93. Brown, A.S., Schaefer, C.A., Quesenberry, C.P., *et al.* (2005). Maternal exposure to toxoplasmosis and risk of schizophrenia in the offspring. *The American Journal of Psychiatry*, **162**, 767–73.
94. Cannon, T.D. (1997). On the nature and mechanisms of obstetric influence in schizophrenia: a review and synthesis of epidemiological studies. *International Review of Psychiatry*, **9**, 387–97.
95. Done, D.J., Johnstone, E.C., Frith, C.D., *et al.* (1991). Complications of pregnancy and delivery in relation to psychosis in adult life: data from the British perinatal mortality survey sample. *British Medical Journal*, **202**, 1576–80.
96. Buka, S.L., Tsuang, M.T., and Lipsitt, L.P. (1993). Pregnancy/delivery complications and psychiatric diagnosis: a prospective study. *Archives of General Psychiatry*, **50**, 151–6.
97. Cannon, M., Jones, P.B., and Murray, R.M. (2002). Obstetric complications and schizophrenia: historical and meta-analytic review. *The American Journal of Psychiatry*, **159**, 1080–92.
98. Zornberg, G.L., Buka, S.L., and Tsuang, M.T. (2000). Hypoxic-ischemia-related fetal/neonatal complications and risk of schizophrenia and other nonaffective psychoses: a 19-year longitudinal study. *The American Journal of Psychiatry*, **157**, 196–202.
99. Cannon, T.D., Rosso, I.M., Hollister, J.M., *et al.* (2000). A prospective cohort study of genetic and perinatal influences in the etiology of schizophrenia. *Schizophrenia Bulletin*, **26**, 351–66.
100. Bennedsen, B.E., Mortensen, P.B., Olesen, A.V., *et al.* (2001). Obstetric complications in women with schizophrenia. *Schizophrenia Research*, **47**, 167–75.
101. Jablensky, A.V., Morgan, V., Zubrick, S.R., *et al.* (2005). Pregnancy, delivery, and neonatal complications in a population cohort of women with schizophrenia and major affective disorders. *The American Journal of Psychiatry*, **162**, 79–91.
102. Cannon, T.D., Mednick, S.A., Parnas, J., *et al.* (1993). Developmental brain abnormalities in the offspring of schizophrenic mothers. *Archives of General Psychiatry*, **50**, 551–64.

103. Keshavan, M.S., Diwadkar, V.A., Spencer, S.M., *et al.* (2002). A preliminary functional magnetic resonance imaging study in offspring of schizophrenic patients. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, **26**, 1143–49.
104. Leask, S.J., Done, D.J., and Crow, T.J. (2002). Adult psychosis, common childhood infections and neurological soft signs in a national birth cohort. *The British Journal of Psychiatry*, **181**, 387–92.
105. Thaden, E., Rhinewine, J.P., Lencz, T., *et al.* (2006). Early-onset schizophrenia is associated with impaired adolescent development of attentional capacity using the identical pairs continuous performance test. *Schizophrenia Research*, **81**, 157–66.
106. Winterer, G., Egan, M.F., Raedler, T., *et al.* (2003). P300 and genetic risk for schizophrenia. *Archives of General Psychiatry*, **60**, 1158–67.
107. Calkins, M.E., Curtis, C.E., Iacono, W.G., *et al.* (2004). Antisaccade performance is impaired in medically and psychiatrically healthy biological relatives of schizophrenia patients. *Schizophrenia Research*, **71**, 167–78.
108. Zammitt, S., Allebeck, P., David, A.S., *et al.* (2004). A longitudinal study of premorbid IQ score and risk of developing schizophrenia, bipolar disorder, severe depression, and other nonaffective psychoses. *Archives of General Psychiatry*, **61**, 354–60.
109. Reichenberg, A., Weiser, M., Rabinowitz, J., *et al.* (2002). A population-based cohort study of premorbid intellectual, language, and behavioral functioning in patients with schizophrenia, schizoaffective disorder, and nonpsychotic bipolar disorder. *The American Journal of Psychiatry*, **159**, 2027–35.
110. Cannon, T.D., Mednick, S.A., and Parnas, J. (1990). Antecedents of predominantly negative- and predominantly positive-symptom schizophrenia in a high-risk population. *Archives of General Psychiatry*, **47**, 622–32.
111. Wahlberg, K.E., Wynne, L.C., Oja, H., *et al.* (1997). Gene-environment interaction in vulnerability to schizophrenia: findings from the Finnish adoptive family study of schizophrenia. *The American Journal of Psychiatry*, **154**, 355–62.
112. van Os, J., Hanssen, M., Bak, M., *et al.* (2003). Do urbanicity and familial liability coparticipate in causing psychosis? *The American Journal of Psychiatry*, **160**, 477–82.
113. Mischler, E.G. and Scotch, N.A. (1983). Sociocultural factors in the epidemiology of schizophrenia; a review. *Psychiatry*, **26**, 315–51.
114. Goldberg, E.M. and Morrison, S.L. (1963). Schizophrenia and social class. *The British Journal of Psychiatry*, **109**, 785–802.
115. Bhugra, D., Leff, J., Mallett, R., *et al.* (1997). Incidence and outcomes of schizophrenia in Whites, African-Caribbeans, and Asians in London. *Psychological Medicine*, **27**, 791–8.
116. Harrison, G., Glazebrook, C., Brewin, J., *et al.* (1997). Increased incidence of psychotic disorders in migrants from the Caribbean to the United Kingdom. *Psychological Medicine*, **27**, 799–806.
117. Selten, J.P., Slaets, J.P.I., and Kahn, R.S. (1997). Schizophrenia in the Surinamese and Dutch Antilean immigrants to the Netherlands: evidence of an increased incidence. *Psychological Medicine*, **27**, 807–11.
118. Selten, J.P., Slaets, J., and Kahn, R. (1998). Prenatal exposure to influenza and schizophrenia in Surinamese and Dutch Antillean immigrants to the Netherlands. *Schizophrenia Research*, **30**, 101–3.
119. Hutchinson, G., Takei, N., Fahy, T.A., *et al.* (1996). Morbid risk of schizophrenia in first-degree relatives of White and African-Caribbean patients with psychosis. *The British Journal of Psychiatry*, **169**, 776–80.
120. Jablensky, A. (2006). Subtyping schizophrenia: implications for genetic research. *Molecular Psychiatry*, **11**, 815–36.
121. Spitzer, R.L., Endicott, J., and Robins, E. (1978). Research diagnostic criteria. Rationale and reliability. *Archives of General Psychiatry*, **35**, 773–82.
122. Hutchinson, G., Takei, N., Bhugra, D., *et al.* (1997). Increased rate of psychosis among African-Caribbeans in Britain is not due to an excess of pregnancy and birth complications. *The British Journal of Psychiatry*, **171**, 145–7.

4.3.6 Aetiology

Contents

4.3.6.1 Genetic and environmental risk factors for schizophrenia

R. M. Murray and D. J. Castle

One thing that is certain about the aetiology of schizophrenia is that there is no single cause. This might reflect the fact that the schizophrenia construct itself is heterogeneous, such that specific subtypes might in the future be found to have specific causes. But it is more useful at this stage of our knowledge to conclude that, like other disorders such as ischaemic heart disease and diabetes mellitus, schizophrenia results from the cumulative effects of a number of risk factors. These may be crudely divided into the familial-genetic and the environmental, though there are clearly interactions between the two.

Familial–genetic risk

The most powerful risk factor for schizophrenia is having a relative afflicted with the disorder. Numerous studies have shown that the lifetime risk for broadly defined schizophrenia increases from about one per cent in the general population to about 10 per cent cent in first-degree relatives of patients with schizophrenia and to close to 50 per cent in those with two parents with the disorder.[1] However, familial aggregation does not prove that a condition is genetically transmitted; to look at this issue we need to turn to adoptee and twin studies.

Adoption studies

Adoptee studies offer the opportunity of separating the effects of familiality from genetics. In the first such study of schizophrenia, Heston and Denney[2] demonstrated that five out of 47 children of mothers with schizophrenia who were adopted away within a few days of their birth, later developed schizophrenia compared with none out of 50 adoptees with no family history of schizophrenia. Similar findings were reported from the Danish-American Study of Rosenthal *et al*[3] who found that a significantly higher proportion of the adopted-away offspring of parents with schizophrenia were classified as having schizophrenia or 'borderline schizophrenia', than were control adoptees. This study originated the concept of the schizophrenia spectrum disorder, which has come to include not only frank schizophrenia but also schizophreniform disorder, as well as schizotypal and possibly paranoid personality disorder.

In an extension of the Danish-American collaboration, Kety *et al.*[4] took all adoptees in Denmark who had schizophrenia and examined their biological and adoptive relatives; unlike the earlier adoption studies this one also used operational definitions of the

schizophrenia spectrum conditions. Fully 23.5 per cent of the biological first-degree relatives received a schizophrenia spectrum diagnosis compared with only 4.7 per cent of the biological relatives of normal control adoptees; the adoptive relatives of both groups of adoptees had very low rates of spectrum disorders.

Finally, Wender *et al.*[5] studied the grown-up children of normal individuals who, by mischance, had been placed with an adoptive parent who later developed schizophrenia, and found that they were at no increased risk of the disorder. Thus, adoption studies consistently indicate that the familial aggregation of schizophrenia is determined by individuals inheriting genes from someone with the disorder (or a related spectrum condition) rather than any effect of the intrafamilial culture (e.g. being brought up by a parent with schizophrenia).

Twin studies

Twin studies have come to the same conclusion. Gottesman,[1] who reviewed the literature, calculated the average probandwise concordance rate for broadly defined schizophrenia in monozygotic twins to be 46 per cent, compared with 14 per cent in dizygotic twins. This difference reflects that while monozygotic twins share all their genes, dizygotic twins share, on average, only half. Further evidence of the effect of heredity comes from the evidence that the concordance rate in 12 pairs of monozygotic twins who were reared apart was 58 per cent.[1]

The above twin studies preceded the introduction of operational definitions of schizophrenia. When studies with such definitions were carried out, the rates for both monozygotic and dizygotic twins were both lower, but the disparity between the two remained. Cardno *et al.*[6] examined 108 consecutive pairs of twins seen at the Maudsley Hospital in London, and reported probandwise concordance rates for DSM-IIIR schizophrenia of 42.6 per cent in monozygotic twins and 0 per cent in dizygotic twins.

What is the range of the clinical phenotype transmitted?

The fact that an individual can have the same genes as their co-twin with schizophrenia but have a better than evens chance of remaining non-psychotic indicates that it is not schizophrenia *per se* which is inherited but rather a susceptibility to it. Further evidence in support of this comes from a study which showed that the offspring of the identical but well co-twins of individuals with schizophrenia carry a risk of the disorder similar to that of the offspring of the affected twin.[7] Thus, the predisposition is transmitted without being expressed as schizophrenia.

As noted earlier, sometimes the predisposition may be expressed as non-psychotic spectrum disorders. In addition, family studies show that relatives of people with schizophrenia also show an increased risk of other psychotic conditions such as schizoaffective disorder, atypical and schizophreniform psychoses, and affective psychosis with mood-incongruent delusions. Thus, the clinical phenotype transmitted encompasses a range of psychotic conditions, as well as schizotypal personality disorder and paranoid personality. Within schizophrenia, researchers have asked whether different subtypes are differentially inherited. The results have in general been negative which is not surprising since clinicians know that an individual patient can appear predominantly hebephrenic on one admission and schizoaffective on another. However, there has been a consensus that paranoid schizophrenia is less familial than other types and is associated with a lower monozygotic twin concordance. Also, very late onset schizophrenia (late paraphrenia) appears to carry less genetic loading than early-onset types.

It has been repeatedly shown that schizophrenic symptoms can be summarized as three main factors: delusions and hallucinations (reality distortion), negative symptoms (psychomotor poverty), and disorganization or positive thought disorder.[8,9] Is schizotypal personality particularly closely related to one of these three core syndromes? Mata *et al.*[10] showed that schizotypal personality scores in non-psychotic relatives were significantly correlated with the presence of delusions and hallucinations in the probands; indeed, they were also correlated with premorbid schizoptypal traits in the childhood of the probands. Thus, it seems that certain families transmit schizotypal traits which manifest themselves in childhood; some family members remain schizotypal throughout life but in others this predisposition is compounded by other (genetic or environmental) factors so that the individual passes a threshold for the expression of delusions and hallucinations.

Genetic models

From the data reviewed above, we can conclude that schizophrenia cannot be explained by the inheritance of a single major gene. In any case, such simple Mendelian inheritance would be hard to square with the persistence of schizophrenia in the population. Since people with schizophrenia tend to reproduce less frequently than the rest of the population, one would have expected that a single major gene with such damaging consequences would have been selected out of the gene pool.

The evidence is compatible with oligogenic inheritance (a small number of genes involved) but most parsimonious is a polygenic model which postulates that a number of genes of small effect are involved. Support for this model comes from the fact that the risk to an individual increases with the number of affected relatives[1] and also that the monozygotic concordance rate is higher for those twins who had an early rather than late onset of psychosis.[6]

Family studies also show that the relatives of probands with an early onset have a higher morbid risk of psychosis than the relatives of late-onset patients.[11] These findings are compatible with the idea that schizophrenia is in part a developmental disorder and that some of the susceptibility genes may be involved in the control of neurodevelopment.[12]

Molecular genetic studies

Researchers have been using molecular techniques to seek the gene or genes that predispose to schizophrenia. The first technique to be used was that of linkage in which large families with several members affected with schizophrenia are studied to try and find a genetic marker that co-segregates with the disease. Two decades of linkage studies suggest that no gene can exist which increases the overall risk of schizophrenia by more than a factor of around three, and that, therefore, there are likely to be a number of susceptibility genes of small effect. This is the mode of transmission for other chronic disorders such as diabetes and hypertension, and, as with these disorders, the genetic basis of schizophrenia is beginning to be unravelled. In the past few years, findings from linkage studies have led on to detailed mapping studies of certain chromosomal

regions which have in turn implicated specific genes. Those for which there is most evidence currently are neuregulin and dysbindin.

Neuregulin[13]: An association between schizophrenia and a multi marker haplotype (a pattern of DNA within a gene) of Neuregulin 1 (**NRG1**) on chromosome 8p21–22 was found in an Icelandic sample in 2002 and soon replicated in a Scottish population. Subsequently, neuregulin has been implicated in other studies although the exact haplotype has varied in the different studies.

Dysbindin: Also in 2002, Straub *et al.*[14] reported, in Irish families, association between schizophrenia and several SNPs (single nucleotide polymorphisms) and multimarker haplotypes spanning the gene encoding dystrobrevin-binding protein 1 (**DTNBP1**), or dysbindin, located at chromosome 6p22.3. Some but not all other studies have replicated this association.

Another way of identifying susceptibility genes is through the study of chromosomal rearrangements. Thus, Blackwood, *et al.*[15] reported that a large Scottish pedigree showed strong evidence for linkage between a balanced chromosomal translocation (1, 11) (q42;q14.3) (two portions of the different chromosomes swapping positions with each other) and a broad phenotype consisting of schizophrenia, bipolar disorder, and recurrent depression. This translocation caused the disruption of a gene, termed disrupted in schizophrenia 1 (**DISC1**). Subsequent studies have examined DISC1 in Finnish and US samples, and have suggested that it may be a susceptibility gene for both schizophrenia and bipolar disorder.

A third approach, that of association studies, takes a gene that is suspected of involvement in the pathogenesis of the disorder and compares the frequency of its various alleles in a series of individuals with schizophrenia as opposed to a control group without schizophrenia. One such gene is the catecholamine O-methyl transferase (**COMT**) gene which has been extensively investigated because of its role in dopamine metabolism, especially in the prefrontal cortex.[16, 17] A mis-sense mutation (incorrect unit in the genetic code) generates a valine to methionine substitution at codon 158 (Val158Met), producing an unstable enzyme with reduced degradation of dopamine. The evidence that this polymorphism is in itself a susceptibility gene is uncertain but as we shall see later, it may compound other risk factors for schizophrenia.

Neuregulin, dysbindin, and DISC 1 are the most replicated putative susceptibility genes for schizophrenia but other plausible candidate genes identified by linkage and follow-up studies such as **G72** (D–amino acid oxidase activator, DAOA), have been suggested. G72 and several of the other putative risk genes appear to carry not only an increased risk of schizophrenia but also of bipolar disorder, and are thus congruent with the results of a twin study which suggested substantial genetic overlap between the two major psychoses.

Nevertheless, none of the above genes can yet be said to be 100 per cent proven as a cause of schizophrenia since there remain inconsistencies between the specific risk alleles and haplotypes among studies. It is unlikely that there is a simple relationship between carrying one risk allele and developing schizophrenia. Rather, an individual may need to carry a number of risk genes and be exposed to several environmental risk factors. In such a dynamic multifactorial model, several genes of small effect interact with each other and with time-specific exposure to environmental risk factors contribute to both the onset and outcome of schizophrenia.

Biological abnormalities in the relatives of people with schizophrenia

Relatives have been examined for some of the biological abnormalities which are found in their kin with schizophrenia. Thus, in the Maudsley Study of families multiply affected with schizophrenia, both the members with schizophrenia and those unaffected relatives who appeared to be transmitting the liability to the disorder (so-called obligate carriers) showed larger lateral ventricles than controls.[18, 19] McDonald *et al.*[20] went on to show that such families transmit a grey matter pattern that shows deficits in frontal and temporal areas and that the greater the genetic liability, the greater the deficit.

In the same Maudsley Family Study, the non-psychotic relatives exhibited other neurophysiological abnormalities such as an excess of delayed P300 event-related potentials; their prevalence was not as high as in the patients themselves but higher than in unaffected controls.[21] Those patients who showed an excess of saccadic distractability errors tended to have relatives with the same eye-tracking abnormalities.[22] The patients with schizophrenia and their well relatives from these multiply affected families also showed more integrative neurological abnormalities than controls.[23]

These findings suggest that what is being transmitted is not genes for schizophrenia *per se* but rather genes for a variety of characteristics (e.g. schizotypal personality, enlarged lateral ventricles, grey matter deficit, delayed P300, integrative neurological abnormalities) which may increase the risk of schizophrenia or at least be markers thereof. Individuals can inherit these characteristics without being psychotic; perhaps schizophrenia only ensues when an individual inherits a number of such endophenotypic abnormalities and passes a critical threshold of risk.[24]

Advancing paternal age in non-familial schizophrenia

An interesting finding first noted over 30 years ago is that schizophrenia is commoner in those whose fathers were old at the time they were born. One of the largest studies to demonstrate this comes from Sipos *et al.*[25] who studied the risk of schizophrenia in 7 54 330 people born in Sweden. The overall hazard ratio for developing schizophrenia increased with each 10 year increase in paternal age. This association between paternal age and schizophrenia has been repeatedly shown to be present in those with no family history of the disorder, but not in those with a positive family history. This stronger association between paternal age and schizophrenia in people without a family history raises the possibility that accumulation of de novo mutations in paternal sperm with ageing contributes to the risk of schizophrenia.

Environmental factors

It is evident from above that genes exert a probabilistic rather than a deterministic effect on the development of schizophrenia; environmental risk factors appear to be necessary for the disease to become manifest in many, if not all, cases.[26] But what are these environmental risk factors?

Pre- and perinatal complications

More than 20 studies have shown that patients suffering from schizophrenia are more likely to have a history of pre- or perinatal complications (collectively termed obstetric complications) than

are healthy subjects from the general population, patients with other psychiatric disorders, and their own healthy siblings.[27] Some of the studies which reported these findings were based upon interviews with patients' mothers asking them to recall their pregnancies; such interviews are obviously open to distortion by recall bias. However, similar findings have been reported by studies examining data collected in obstetric records at the time of birth of patients and controls.[28] Indeed a meta-analysis of large epidemiologically sophisticated studies which used contemporary records confirmed that there is modest but consistent effect of obstetric complications. [29]

Of course, it is possible that the excess obstetric complications in schizophrenia may be the consequence of some pre-existing abnormality. Since the foetus plays an active role in the normal progress of pregnancy and labour, foetal impairment induced by earlier abnormality may itself result in some perinatal complications. Also, some studies have shown that women with schizophrenia who become pregnant tend to have more obstetric complications, possibly owing to their behaviour during pregnancy, for example smoking and not attending antenatal visits.

The term 'obstetric complications' covers a broad range of obstetric events. An international study on 700 schizophrenic patients and a similar number of controls found that low birth weight, prematurity, and resuscitation at birth were particularly increased in the schizophrenic patients;[27] other complications that have been implicated include retarded foetal growth and rhesus incompatability. Thus, a common characteristic of most of the obstetric complications implicated is that they increase the risk of hypoxia.

Could hypoxic–ischaemic damage be the mechanism that increases the risk of later schizophrenia? Children who were subject to cerebral hypoxia at or before birth show an excess of abnormalities on MRI scan, of minor neurological signs, and of cognitive and behavioural problems, characteristics also found in many preschizophrenic children.[30] As one might predict, studies of monozygotic twins discordant for schizophrenia have shown that the affected twins have larger lateral ventricles and smaller hippocampi than their well co-twins;[31,32] furthermore, those twins who have been subjected to the most severe perinatal difficulties have the largest ventricles and smallest hippocampi.[33]

Similarly, Stefanis *et al.*[34] compared hippocampal volume in three groups, viz, schizophrenia patients with affected relatives but with no personal history of obstetric complications; schizophrenia patients with no affected relatives but who had a history of significant obstetric complications; and normal controls. Hippocampal volume was normal in the first schizophrenia group but reduced in the second group, implying that it is hypoxic-ischaemic damage rather than genetic predisposition alone that determines decreased hippocampal volume in schizophrenia.

Season of birth and maternal exposure to infection

Many studies have shown (in the Northern Hemisphere at least) that people born in late winter and spring are slightly more likely than expected to later develop schizophrenia. Since respiratory viral infections such as influenza tend to occur in autumn and winter, maternal infection might provide the explanation. A number of epidemiological studies have, therefore, addressed the question of whether maternal exposure to influenza during the second trimester of pregnancy is a risk factor for schizophrenia; some but not all studies have suggested that it is.[35] One study[36] reported an association between the presence of antibodies to the influenza virus in first trimester blood, but not during the other trimesters. The possibility that prenatal exposure to rubella may have a similar risk-increasing effect for schizophrenia has been raised, and a significant association has been reported with serologically-documented rubella exposure in gestation[37] Some studies have implicated other infectious agents such as herpes simplex, cytomegalovirus and toxoplasmosis, but there is as yet no consensus as to whether these findings are replicable or not.

Severe prenatal malnutrion appears to have an effect. Thus, children born following the Dutch Hunger Winter when the Nazi occupiers systematically starved the population were shown to have a higher risk of schizophrenia and this finding has recently been replicated in a Chinese population.[38,39]

Childhood risk factors

There is now a wealth of evidence attesting to the fact that a proportion of individuals who later manifest schizophrenia show abnormalities in their early development. The evidence for early developmental abnormalities in schizophrenia come from three main sorts of study:

- high-risk studies in which the offspring of parent(s) with schizophrenia are examined;
- follow-back studies where cases of schizophrenia are ascertained, and their early developmental trajectory plotted with the help of history from the individual and family, sometimes also including such evidence as school reports; and
- cohort studies, where birth cohorts are followed up prospectively, and individuals who later manifest schizophrenia are compared with the rest of the cohort in terms of their early development.

(a) High-risk studies

Studies of the offspring of mothers with schizophrenia, the so-called 'high-risk studies', show that between a quarter to a half show some deviation from normal in terms of their early development (reviewed by Davies *et al.*[40]) In the neonatal period, there is a tendency to hypotonia and decreased cuddliness; in infancy, milestones are delayed; in early childhood, there is poor motor co-ordination; and in later childhood, there are deficits in attention and information processing. Fish *et al.*[41] followed their cohort of 12 high-risk infants into adulthood. One developed schizophrenia and six showed schizotypal or paranoid personality traits; these authors coined the term 'pandysmaturation' to describe the abnormalities which included delayed motor milestones in the first two years of life.

(b) Follow-back studies

High-risk studies have been criticized on the basis that they are unrepresentative because only a minority of people who develop schizophrenia have a mother with the same illness. Therefore, a separate set of studies of representative groups of patients with schizophrenia have used maternal recall to document the early development of adults with schizophrenia. These have shown impairment of cognitive and neuromotor development and interpersonal problems. These findings are more commonly reported in males than females, and tend to be associated with an early onset of illness.[42] The findings are not specific to schizophrenia, being

reported also in the early development of some children who later manifest an affective psychosis.[43]

Of course, one of the major criticisms of follow-back studies is the likelihood of recall bias. Studies that have avoided this problem include those which have accessed IQ scores assessed prior to illness onset; these have shown that premorbid IQ is, on average, lower in those, particularly males, who later manifest schizophrenia.[44, 45]

Another source of material mapping early development has been childhood home videos, which have been reviewed by researchers 'blind' to whether the individual later manifested schizophrenia.[46] In comparison with their healthy siblings, the preschizophrenic children showed higher rates of neuromotor abnormalities (predominantly left-sided) and overall poorer motor skills; the group differences were significant only at two years of age.

(c) Cohort studies

Cohort studies have overcome many of the criticisms of follow-back studies. In an investigation of the 1958 British Perinatal Mortality cohort, comprising 98 per cent of all children ($n = 15\,398$) born in the United Kingdom in a certain week in March 1958, Done *et al.*[47] compared those who later manifested schizophrenia ($n = 40$), affective psychosis ($n = 35$), and neurotic illness ($n = 79$) with each other as well as with 1 914 randomly selected individuals with no history of mental illness. At age seven years, teacher ratings showed the preschizophrenic children to have exhibited more social maladjustment than controls; the effect was most marked in boys. The preaffective children differed little from normal controls, whilst the preneurotic children (expressly girls) showed some maladjustment (over- and under-reaction) at age 11 years.

In a similar study of the 1946 British Birth Cohort, Jones *et al.*[48] determined that 30 out of 4 746 individuals had, in adulthood, developed schizophrenia. This group was more likely than the rest of the cohort to show delayed milestones and speech problems, to have a lower premorbid IQ and lower education test scores at ages 8, 11, and 15 years, and to prefer solitary play at ages 4 and 6 years. Perhaps the most influential of all the cohort studies has been the Dunedin Birth Cohort Study, which followed the development of 1 037 children through the ages of 3 to 15 years, and assessed them again at the ages of 18, 21, and 26 years.[49] This study found that poorer motor development, poorer receptive language, and a lower IQ all increased the risk of subsequently developing schizophreniform disorder by age 26 years. The Dunedin cohort additionally provided evidence that a proportion of children who develop schizophrenia are already experiencing 'quasi-psychotic' phenomena by age 11 years.[50] These phenomena include beliefs that people are reading their minds or following or spying on them, or they are already hearing voices. Children with strong evidence of quasi-psychotic symptomatology were up to 16-times more likely to develop schizophreniform disorders by the age of 26 years; making these phenomena some of the most powerful early predictors of later psychosis.

Together, such studies provide compelling evidence for a tendency of individuals with schizophrenia to show abnormalities in development which antedate the onset of illness. The findings are compatible with the notion that subtle brain abnormalities (which may be genetically or environmentally mediated, or both) underpin schizophrenia. However, it is also possible that some of the childhood risk factors are independent and act in an additive manner to set individuals on an increasingly deviant trajectory towards schizophrenia.

Social and geographic risk factors

Recent dogma about schizophrenia has held that the incidence does not vary by time or place, even though such an occurrence would have made schizophrenia unique among diseases! Now this curious belief has been disproved by a raft of studies. In particular, a systematic review by McGrath *et al.*[50] concluded that the incidence of schizophrenia shows prominent worldwide variation (up to five-fold), and that it is about 40 per cent greater in men than women.

In 1939, Faris and Dunham[52] reported that an excess of individuals with schizophrenia was found in certain deprived inner-city areas. These authors suggested that social isolation in poor deprived parts of the city could precipitate schizophrenia. However, subsequently, their results were interpreted as a consequence of social drift, i.e. the idea that individuals with this illness 'drift' down the social scale.[53] This effect was postulated to result from not only the illness itself but also its prodroma and consequences such as loss of employment and estrangement from family. A related finding is that of lack of upward social mobility in individuals with schizophrenia. For example, Hollingshead and Redlich[54] reported that individuals with schizophrenia to be less likely than expected to attain the socio-economic status of their fathers.

More recently, research has focused on the apparent excess of individuals who later manifest schizophrenia, who actually start life in a setting which appears to increase the subsequent risk of schizophrenia. Kohn[55] stated that '... in all probability, lower class families produce a disproportionate number of schizophrenics' but the evidence concerning such 'social causation' is contradictory. Thus, Turner and Wagenfeld[56] reported fathers of schizophrenia patients to be themselves over-represented in lower socio-economic groups. However, Jones *et al.*[48] did not find this.

It may be that it is not so much poverty as being born or brought up in a city which increases the risk of the disorder. For example, Lewis *et al.*[57] found that Swedish conscripts who later manifested schizophrenia were 1.65 times more likely to have been born in urban than rural areas. Similarly, Marcelis *et al.*[58] reported that birth in an urban area of Holland carried twice the risk of later schizophrenia of birth in a rural area. Similar findings have come from Denmark where those individuals born in Copenhagen appear to have twice the risk of schizophrenia of those born in rural areas.[59] It is now generally accepted that the incidence is higher amongst those brought up in urban areas, and that the larger the town, and the longer the individual has lived in the city, the greater the risk. The exact mechanisms underlying this effect remain unclear.

Immigration

Since the classic study of Odegaard in 1932,[60] many studies have reported that migrants are at increased risk of schizophrenia. A recent meta-analysis of 18 studies of migrants from different backgrounds confirmed a weighted mean relative risk for first-generation migrants of 2.7 (95 per cent CI 2.3–3.2) and for second generation migrants, 4.5 (95 per cent CI 1.5–13.1). Risk was higher for migrants from lower socio-economic countries, and for black people moving into predominantly white societies.[61]

A notable example has come from a series of studies of African-Caribbeans resident in the United Kingdom, who show a markedly higher rate of schizophrenia than do their white

British-born counterparts.[62]This is in the absence of any increased risk to Caribbeans who remain in the West Indies.[63] The increase is striking. The large and sophisticated AESOP study of three English cities demonstrated a ninefold increase in the incidence of schizophrenia among African Caribbeans, and a six-fold increase among those of African origin.[64] Boydell *et al.*[65] further demonstrated that migrants were especially vulnerable if relatively isolated in localities where their own ethnic group were in a small minority. Of particular interest is that this increased risk also pertains to British-born offspring of Caribbean migrants, discounting an explanation in terms of migration stress alone. Furthermore, there is a marked increased risk in the siblings but not the parents of this second generation;[63] this suggests an environmental effect operating particularly upon this second generation.

Initial studies sought to ascertain any evidence of developmental disadvantage such as poor maternal nutrition, poor obstetric care, and possible maternal susceptibility to novel viruses. However, these studies have shown that, if anything, African-Caribbean schizophrenic patients in England show less evidence of neurodevelopmental insult than their white counterpart patients. Other research focuses on the possibility that a paranoid reaction to social disadvantage and discrimination may be one factor. The findings relating to skin colour potentially support the notion of perceived or real discrimination being an important variable. Other work suggests that people in certain migrant communities are particularly likely to be exposed to risk-increasing factors such as childhood adversity (e.g. parental separation) and adult social exclusion.

Life events

Brown and Birley[66] reported an excess of life events in the 3 weeks preceding schizophrenic relapse. Further studies were conflicting in their findings, possibly due to methodological problems. The study of Bebbington *et al.*[67] avoided many of the methodological pitfalls, assessing life events in 97 psychotic patients (52 with schizophrenia) and general population controls. There was a significant relationship between life events and onset or relapse of schizophrenia, although it was not as strong as for depressive psychosis. One possibility is that certain types of schizophrenic patients are particularly vulnerable to relapse following adverse life events. For example, Bebbington *et al.*[67] found females to be particularly prone, whilst van Os *et al.*[9] found life events to be associated with a less severe good-outcome illness.

There is also evidence that families who exhibit high 'expressed emotion' (comprising critical comments, hostility, and/or over-involvement) can provide an environment which enhances the risk of relapse in a family member with schizophrenia. Again, cause and effect are difficult to tease apart. Thus, it is possible that patients with more severe and intractable illnesses may induce more expressed emotion in their relatives. It is clear, though, that family interventions aimed at reducing levels of expressed emotion can be effective in reducing relapse rates in the individual.

Drug abuse

Numerous studies attest to the fact that illicit substance use is more prevalent in patients with schizophrenia than in the general population; estimates of the prevalence of such comorbidity in individuals with schizophrenia range from 20 to 60 per cent, and are consistently higher than in well controls.[68]

Whether illicit substances actually cause schizophrenia has been very contentious. The most robust methodology to consider this issue is a cohort design, and a number of such studies have now investigated whether premorbid exposure to cannabis is associated with an increased later risk of schizophrenia. Arseneault *et al.*[69] reviewed these studies and concluded that cannabis could be considered a cumulative casual factor in some cases of schizophrenia, operating in consort with other predisposing factors to 'tip the scales' in some individuals who might not otherwise have manifested the disorder. This literature needs to be seen in the light of the fact that the vast majority of people who use cannabis do not develop schizophrenia, and the majority of cases of schizophrenia are not caused by cannabis; it has been estimated that the population attributable fraction for cannabis and schizophrenia is of the order of 5 per cent to 8 per cent.

Similarly, although clinical wisdom suggests that illicit substance use has a negative impact on the longitudinal course of schizophrenia, there are few methodologically sound studies in this area.[70, 71] Indeed, even the finding of an excess of use, and the association of such use with a poor longitudinal course, is potentially explicable by confounding factors such as substance abuse by the patients who are more ill. On balance, though, it seems reasonable to conclude that illicit substances make the longitudinal course of illness worse, and that patients with schizophrenia should be strongly advised to seek help to cease such behaviours.

Risk factors, age of onset, and outcome

Individuals who have been exposed to certain risk factors for schizophrenia tend have an earlier onset of psychosis than those who have not. Thus, age of onset is earlier in those whose relatives show a high morbid risk of schizophrenia;[11] similarly, those twin pairs in which the schizophrenia has an early onset show the highest monozygotic concordance.[6]

Schizophrenia patients with an early age at onset of psychosis are also more likely than those with later onset to have had a history of exposure to obstetric complications,[72] while those who showed childhood deficits such as low IQ also tend to have an early onset.[45] Schizophrenia patients who abuse cannabis have also recently been shown to have an earlier onset than those who do not.

If a factor operates to increase the risk of schizophrenia and to bring on its onset, then it is logical to think that if it is still present then it will be associated with a poor outcome. Thus, a family history of schizophrenia, a history of obstetric complication, childhood low IQ, and continued drug abuse are all associated with a poor outcome. On the other hand, those patients who develop psychosis following stressful life-events tend to have a better outcome than those with no such precipitant.[9]

The risk factor model: Gene–environment interaction

Thus, one way of construing the aetiology of schizophrenia is to see individuals on a stress-vulnerability continuum in which genetic and environmental factors act in an additive manner until a threshold of liability for expression of psychosis is passed. An individual might, for instance, inherit a schizotypal personality but not develop frank psychosis unless exposed to some cerebral insult which

causes cognitive impairment; the sum of the two factors could produce the psychotic illness.

Assuming such a model in which a number of genes and environmental factors of small effect act additively, then the heritability of schizophrenia can be calculated to be between 66 and 85 per cent (i.e. a high proportion of liability to the disorder is under genetic influence). However, this assumes that the various factors operate additively, and much evidence is against this assumption. Rather, it seems that there is often an interaction between genetic susceptibility and environmental effects. As van Os and Marcelis[26] point out, it seems that certain individuals exposed to an environmental risk factor have a high risk of developing schizophrenia while others with a different genotype are at low risk.

Thus, the quality of upbringing can interact with genetic predisposition. For example, the Finnish Adoption Study has shown that when the offspring of women with schizophrenia are placed in a well-adjusted family, they have a lower risk of developing a schizophrenia spectrum disorder than if they are placed in a dysfunctional family, i.e the genotype renders the individual susceptible to the adverse effect of an adverse family environment.[73] Obstetric complications also appear to interact with, and compound, a genetic liability; the offspring of parents with schizophrenia are more likely to develop increased ventricular size following obstetric complications.

Similarly, there is evidence that individuals with a family loading for schizophrenia may be more susceptible to psychosis following abuse of cannabis.[74] The latter situation may be complicated by the possibility that individuals who inherit certain personality characteristics may be more likely to take drugs such as cannabis, i.e. their genotype renders them more liable to expose themselves to a factor to which they are genetically susceptible. Caspi *et al.*[75] investigating the Dunedin birth cohort study, presented evidence of a gene by environment (G x E) interaction between a functional polymorphism in the catechol-O-methyltransferase gene (COMT) and exposure to cannabis. The enzyme COMT has an essential role in the breakdown of dopamine in the prefrontal cortex. Caspi and colleagues showed that COMT moderates the influence of adolescent cannabis use, with at least a five-fold increased risk of developing schizophreniform disorder in cannabis users homozygous for the high activity Val allele (COMT); the Met/Met status offered relative protection (OR=1.1) while risk for heterozygotes was intermediate (OR=2.5). Furthermore, there was no correlation between the COMT genotype and cannabis use, indicating that COMT genotype does not influence cannabis consumption. This appears to be a the first clear example of a specific gene x environmental interaction predisposing for schizophrenia, but it remains to be replicated.

The implications

Having identified various risk factors for schizophrenia, we can proceed to consider the theoretical possibility of reducing the prevalence of certain risk factors and thus reducing the incidence of the disorder. From the point of view of public health, rare risk factors which have a big effect are much less important than common risk factors of even small effect. Thus, although familial risk has by far the biggest effect, it makes a smaller contribution to the total incidence of the disorder than environmental effects. Therefore, if all cases with an affected first-degree relative could be prevented, we would eliminate only about 10 per cent of the total cases. However, because being born in an urban area is so common, this small effect accounts for a much greater proportion of the population attributable fraction (33 per cent), i.e. if we could bring down the incidence of schizophrenia in cities to that in the countryside, we could theoretically eliminate one-third of all cases of the disorder. Of course, we could only do this if we knew what the critical urban factors were!

The importance of the evidence that early developmental factors are involved in the aetiology of schizophrenia lies in the fact that at least some of these are preventable. For example, advances in antenatal and perinatal care have reduced the frequency, and toxicity, of some obstetric complications. Similarly, vaccination programmes have reduced exposure to some viral infections in pregnancy and childhood. There could be a link between these developments and the decreased incidence of schizophrenia which has been observed in some western countries over recent decades.[76]

Nevertheless, most babies who are exposed to even severe obstetric complications will not later suffer from schizophrenia. Thus, there is at present no sense in attempting to improve antenatal care with the aim of reducing the occurrence of schizophrenia. There is one important exception. It is indisputable that the children of schizophrenic mothers have a higher risk of the disease if they are also exposed to obstetric complications. Therefore those women with schizophrenia who conceive must have the best possible antenatal care during their pregnancy, and steps should be taken to avoid any event (e.g. prolonged labour) which might lead to hypoxic damage to the baby.

The fact that preschizophrenic children show a number of impairments raises the possibility that predisposed individuals could be identified and 'rescued' by some intervention. Unfortunately, the childhood characteristics of such children are non-specific, and their predictive value for the later manifestation of the illness is too low to be of value for any preventative intervention, i.e. many children who show such deviation from normal in terms of early development do not later manifest schizophrenia, whilst other children who later develop schizophrenia have perfectly 'normal' early development. Furthermore, any such abnormalities must be seen in the total context of development, and it should be remembered that many such abnormalities are not static, but may be evident at some stages of development and not at others.

One might think that there is as yet little that can be done systematically to reduce the incidence of schizophrenia. However, one area where prevention is possible concerns drug abuse. The evidence concerning the abuse of cannabis is clear. If the population could be persuaded to avoid heavy use of cannabis, particularly the more potent varieties, then it is likely that a small but nevertheless significant proportion of cases of schizophrenia could be avoided.

Further information

Murray R.M., Jones P.B., Susser E., *et al.* (eds.) (2003) *The Epidemiology of Schizophrenia*. Cambridge: Cambridge University Press.

Van Os J., Marcelis, M. (1998) The ecogenetics of schizophrenia: a review. *Schizophrenia Research*, **32**, 127–35.

Harrison, P.J., Owen, M.J. (2003) Genes for schizophrenia? Recent findings and their pathophysiological implications. *Lancet*, **361**, 417–419.

Jones, P., Cannon, M. (1998) The new epidemiology of schizophrenia. *Psychiatric Clinics of North America*, **21**, 1–25.

References

1. Gottesman, I.I. (1991). *Schizophrenia Genesis: The Origins of Madness*. W.H. Freeman, New York.
2. Heston, L.L. and Denney, D. (1968). Interactions between early life experience and biological factors in schizophrenia. In *The transmission of Schizophrenia* (ed. D. Rosenthal and S. Kety), pp. 363–76. Pergamon Press, Oxford.
3. Rosenthal, D., Wender, P., Kety, S., *et al.* (1971). The adopted-away offspring of schizophrenics. *American Journal of Psychiatry*, **128**, 307–11.
4. Kety, S.S. *et al.* (1994). Mental illness in the biological relatives of schizophrenic adoptees. Replication of the Copenhagen study in the rest of Denmark. *Archives of General Psychiatry*, **51**, 442–55.
5. Wender, P.H., *et al.* (1974). Cross-fostering: a research strategy for clarifying the role of genetic and experimental factors in the etiology of schizophrenia. *Archives of General Psychiatry*, **30**, 121–8.
6. Cardno, A., Marshak, J., Coid, B., *et al.* (1999). Relationships between symptom dimensions and genetic liability to psychotic disorders in the Maudsley Twin Psychosis Series. *Archives of General Psychiatry*, **56**, 162–8.
7. Gottesman, I.I. and Bertelsen, A. (1989). Confirming unexpressed genotypes for schizophrenia: risks in the offspring of Fischer's Danish identical and fraternal discordant twins. *Archives of General Psychiatry*, **46**, 867–72.
8. Liddle, P.F. (1987). Symptoms of chronic schizophrenia. *British Journal of Psychiatry*, **151**, 145–51.
9. van Os, J. *et al.* (1994). The influence of life events on the subsequent course of psychotic illness: a follow-up of the Camberwell Collaborative Psychosis study. *Psychological Medicine*, **24**, 503–13.
10. Mata, I., Sham, P., and Murray, R.M. (2003) Schizotypal personality traits in nonpsychotic relatives are associated with positive symptoms in psychotic probands. *Schizophrenia Research*, **29**, 273–83
11. Sham, P. *et al.* (1994). Age at onset, sex and familial psychiatric morbidity in schizophrenia. Report from the Camberwell Collaborative Psychosis Study. *British Journal of Psychiatry*, **165**, 466–73.
12. Jones, P. and Murray, R.M. (1991). The genetics of schizophrenia is the genetics of neurodevelopment. *British Journal of Psychiatry*, **158**, 615–23.
13. Li D, Collier D.A. and He L. (2006) *Meta-analysis shows strong positive association of the neuregulin 1 (NRG1) gene with schizophrenia*. Molecular Genetics 1995–2002.
14. Straub, R.E., *et al.* (2002). Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. *American Journal of Human Genetics*. **71**, 337–48.
15. Blackwood, D.H.R., Fordyce, A., Walker, M.T., *et al.* (2001). Schizophrenia and affective disorders – cosegregation woth a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. *American Journal of Human Genetics*. **69**, 428–433.
16. Li T, Vallada H, Curtis D, *et al.* (1997). Catechol-O-methyltransferase Val158Met polymorphism: frequency analysis in Han Chinese subjects and allelic association of the low activity allele with bipolar affective disorder. *Pharmacogenetics*, **7**, 349–53.
17. Egan, M.F., *et al.* (2001). Effect of COMT Val108/158Met genotype on frontal lobe function and risk for schizophrenia. *Proceedings of the National Academy of Science, U.S.A.* **98**, 691–722.
18. Sharma, T. *et al.* (1998). Brain changes in schizophrenia: volumetric MRI study of families multiply affected with schizophrenia-the Maudsley Family Study 5. *British Journal of Psychiatry*, **173**, 132–8.
19. Sharma, T., Lancaster, E., Sigmondson, J., *et al.* (1999). Loss of cerebral asymmetry in familial schizophrenics and their relatives detected by magnetic resonance imaging. *Schizophrenia Research*, **40**, 111–20.
20. McDonald C, Bullmore E T. Sham P C, *et al.* (2004). Association of genetic risks for schizophrenia and bipolar disorder with specific and generic brain structural endophenotypes. *Archives of General Psychiatry*, **61**, 974–84
21. Frangou, S. *et al.* (1997). The Maudsley Family Study II: endogenous event-related potentials in familial schizophrenia. *Schizophrenia Research*, **23**, 45–53.
22. Crawford, T., Sharma, T., Puri, B.K., *et al.* (1998). Saccadic eye movements in families multiply affected with schizophrenia. *American Journal of Psychiatry*, **155**, 1703–10.
23. Griffiths, T.D., Sigmundsson, T., Takei, N., *et al.* (1998). Neurological abnormalities in familial and sporadic schizophrenia. *Brain*, **121**, 191–203.
24. Wickham, H. and Murray, R.M. (1997). Can biological markers identify endophenotypes predisposing to schizophrenia? *International Review of Psychiatry*, **9**, 355–64.
25. Sipos, A., Rasmussen, F., Harrison, G.T., *et al.* (2004), Paternal age and schizophrenia: a population based cohort study British Journal of Psychiatry, 329:1070.
26. van Os, J. and Marcelis, M. (1998). The ecogenetics of schizophrenia. *Schizophrenia Research*, **32**, 127–35.
27. Geddes, J.R., Verdoux, H., Takei, N., *et al.* (1999). Individual patient data meta-analysis of the association between schizophrenia and abnormalities of pregnancy and labour. *Schizophrenia Bulletin*, **25**, 413–23.
28. Hultman, C.M., Sparen, P., Takei, N., *et al.* (1999). Prenatal and perinatal risk factors for schizophrenia, affective psychosis and reactive psychosis. *British Medical Journal*, **318**, 421–5.
29. Cannon, M., Jones, P.B. and Murray, R.M. (2002). Obstetric complications and schizophrenia: historical and meta-analytic review. *American Journal of Psychiatry* **159**, 1080–92.
30. Stewart, A.L., Rifkin, L., Amess, P.N., *et al.* (1999). Brain structure, neurocognitive and behavioural function in adolescents who were born very preterm. *Lancet*, **353**, 1653–7.
31. Reveley, A.M., Reveley, M.A., Clifford, C.A., *et al.* (1982). Cerebral ventricular size in twins discordant for schizophrenia. *Lancet*, **i**, 540–1.
32. Suddath, R.L. *et al.* (1990). Anatomical abnormalities in the brains of monozygotic twins discordant for schizophrenia. *New England Journal of Medicine*, **322**, 789–94.
33. McNeill, T.F., Cantor-Graae, E., and Weinberger, D.R. (2000). Relationship of obstetric complications and differences in size of brain structures in monozygotic twin pairs discordant for schizophrenia. *American Journal of Psychiatry*, **157**, 203–12.
34. Stefanis, N., Frangou, S., Yakeley, J., *et al.* (2000). Hippocampal volume reduction in schizophrenia is secondary to pregnancy and birth complications. *Biological Psychiatry*, **46**, 697–702.
35. McGrath, J. and Murray, R.M. (1995). Risk factors for schizophrenia: from conception to birth. In *Schizophrenia* (ed. S. Hirsch and D. Weinberger), pp. 187–205. Blackwell Science, Oxford.
36. Brown, A. S., Begg, M. D., Gravenstein, S., *et al.* (2004a) Serologic evidence of prenatal influenza in the etiology of schizophrenia. *Archives of General Psychiatry*, **61**, 774–80.
37. Brown, A. S., Cohen, P., Greenwald, S. *et al.* (2000a) Nonaffective psychosis after prenatal exposure to rubella. *American Journal of Psychiatry*, **157**, 438–43.
38. Susser, E.S. and Lin, S.P. (1992). Schizophrenia after prensatal exposure to the Dutch hunger winter. *Archives of General Psychiatry*, **49**, 938–88.
39. St Clair, D., Xu, M., Wang, P., *et al.* (2005) Rates of adult schizophrenia following prenatal exposure to the Chinese famine of 1959–1961. *Journal of the American Medical Association*, **294**, 557–62.
40. Davies, N., Russell, A., Jones, P., *et al.* (1998). Which characteristics of schizophrenia predate psychosis? *Journal of Psychiatric Research*, **32**, 121–31.

41. Fish, B., Marcus, J., Hans, S.L., *et al.* (1992). Infants at risk for schizophrenia. *Archives of General Psychiatry*, **49**, 221–35.
42. Castle, D.J. and Murray, R.M. (1991). The neurodevelopmental basis of sex differences in schizophrenia. *Psychological Medicine*, **21**, 565–75.
43. Foerster, A., Lewis, S.W., Owen, M.J., *et al.*. (1991). Premorbid personality in psychosis: effects of sex and diagnosis. *British Journal of Psychiatry*, **158**, 171–6.
44. Aylward, E., Walker, E., and Bettes, B. (1984). Intelligence in schizophrenia: meta-analysis of the research. *Schizophrenia Bulletin*, **10**, 430–59.
45. Russell, A.J., Munro, J.C., Jones, P.B., *et al.*(1997). Schizophrenia and the myth of intellectual decline. *American Journal of Psychiatry*, **154**, 635–9. Letters and reply: *American Journal of Psychiatry*, **155**, 1633–7 (1998).
46. Walker, E.F., Savoie, T., and Davis, D. (1994). Neuromotor precursors of schizophrenia. *Schizophrenia Bulletin*, **20**, 441–51.
47. Done, J.D., Crow, T.J., Johnstone, E.C., *et al.* (1994). Childhood antecedents of schizophrenia and affective illness: social adjustment at ages 7 and 11. *British Medical Journal*, **309**, 699–703.
48. Jones, P.D., Rodgers, B., Murray, R.M., *et al.*(1994). Child developmental risk factors for adult schizophrenia. *Lancet*, **344**, 1398–402.
49. Cannon, M., Caspi, A., Moffitt, T.E., *et al.* (2002b). Evidence for early-childhood, pan-developmental impairment specific to schizophreniform disorder: results from a longitudinal birth cohort. *Archives of General Psychiatry*. **59**, 449–56.
50. Poulton, R., Caspi, A., Moffitt, T.E., *et al.* (2000). Children's self-reported psychotic symptoms and adult schizophreniform disorder: a 15-year longitudinal study. *Archives General of Phychiatry*, **57**, 1053–58.
51. McGrath J.J. (2006). Variations in the incidence of schizophrenia: data versus dogma. *Schizophrenia Bulletin*. **32**, 195–7.
52. Faris, R.B.L. and Dunham, H.W. (1939). *Mental disorders in urban areas. an ecological study of schizophrenia and other psychoses*. University of Chicago Press.
53. Goldberg, S.M. and Morrison, S.L. (1963). Schizophrenia and social class. *British Journal of Psychiatry*, **109**, 785–802.
54. Hollingshead, A.B. and Redlich, F.C. (1954). Schizophrenia and social structure. *American Journal of Psychiatry*, **110**, 695–701.
55. Kohn, M.L. (1975). Social class and schizophrenia: a critical review and reformulation. In *Annual review of the schizophrenic syndrome* (ed. R. Cancro). Brunner-Mazel, New York.
56. Turner, R.J. and Wagenfeld, M.O. (1967). Occupational mobility and schizophrenia: an assessment of the social causation and social selection hypothesis. *American Sociological Review*, **32**, 104–13.
57. Lewis, G., David, A., Andreasson, S., *et al.* (1992). Schizophrenia and city life. *Lancet*, **340**, 137–40.
58. Marcelis, M., Navarro-Mateu, F., Murray, R., *et al.* (1998). Urbanization and psychosis: a study of 1942–1978 birth cohorts in The Netherlands. *Psychological Medicine*, **28**, 871–9.
59. Mortensen, P.B., Pedersen, C.B., Westergaard, T., *et al.* (1999). Effects of family history and place and season of birth on risk of schizophrenia. *New England Journal of Medicine*, **340**, 603–8.
60. Odegaard, O. (1932). Emigration and insanity. *Acta Psychiatrica Scandinavica Supplementum*, **4**.
61. Cantor-Graae, E. And Selten, J-P. (2005). Schizophrenia and migration: a meta-analysis and review. *American Journal of Psychiatry*, **162**, 12–24.
62. van Os, J., Castle, J., Takei, N., *et al.* (1996). Psychotic illness in ethnic minorities: clarification from the 1991 census. *Psychological Medicine*, **26**, 203–8.
63. Hutchinson, G. *et al.* (1996). Morbid risk of psychotic illness in first degree relatives of white and African-Caribbean patients with psychosis. *British Journal of Psychiatry*, **169**, 776–80.
64. Fearon P, Kirkbride JB, Morgan C,*et al.* (2006). Incidence of schizophrenia and other psychoses in ethnic minority groups: results from the MRC AESOP Study. *Psychological Medicine*, **36**, 1541–50.
65. Boydell, J., van Os, J., McKenzie, K. (2001). Incidence of schizophrenia in ethnic minorities in London: ecological study into interactions with environment. *British Medical Journal*. **323**, 1336–8.
66. Brown, G.W. and Birley, J.L.T. (1968). Crises and life changes and the onset of schizophrenia. *Journal of Health and Social Behaviour*, **9**, 203–14.
67. Bebbington, P. *et al.* (1993). Life events and psychosis: initial results from the Camberwell Collaborative Psychosis Study. *British Journal of Psychiatry*, **162**, 72–9.
68. Meuser, K.T. *et al.* (1990). Prevalence of substance use in schizophrenia. *Schizophrenia Bulletin*, **16**, 31–56.
69. Arseneault, L., Cannon, M., Witton, J., *et al.* (2004). Cannabis as a potential causal factor in schizophrenia. In: *Marijuana and Madness* (ed. D.J. Castle and R.M. Murray), pp. 101–18. Cambridge: Cambridge University Press.
70. Turner, W.M. and Tsuang, M.T. (1990). Impact of substance use on the course and outcome of schizophrenia. *Schizophrenia Bulletin*, **16**, 87–96.
71. Grech, A. (1998). Drug abuse and psychosis in London and Malta. M.Sc. Thesis, University of London.
72. Verdoux, H. *et al.* (1997). Obstetric complications and age at onset in schizophrenia: an international collaborative meta-analysis of individual patient data. *American Journal of Psychiatry*, **154**, 1220–7.
73. Wahlberg, K.E., Wynne, L.C., Oja, H., *et al.* (1997). Gene-environment interaction in vulnerability to schizophrenia. *American Journal of Psychiatry*, **154**, 355–62.
74. McGuire, P., Jones, P., Harvey, I., *et al.* (1994). Cannabis and acute psychosis. *Schizophrenia Research*, **24**, 995–1011.
75. Caspi A, Moffitt T E, Cannon M, *et al.* (2005). Moderation of the effect of adolescent-onset cannabis use on adult psychosis by a functional polymorphism in the Catechol-O-Methayltransferase Gene: Longitudinal evidence of a gene X environment interaction. *Biological Psychiatry,* **57**, 1117–27.
76. Der, G., Gupta, S., and Murray, R.M. (1990). Is schizophrenia disappearing? Evidence from England and Wales 1952–1986. *Lancet*, **335**, 513–16.

4.3.6.2 The neurobiology of schizophrenia

Paul J. Harrison

The neurobiology of schizophrenia remains the subject of intense research activity. Here, the key issues and findings are described, divided into functional and structural aspects, and followed by a summary of the major neurobiological theories. Where possible, meta-analyses and systematic reviews are cited in preference to individual studies.

Functional neurobiology of schizophrenia

Dopamine

The dopamine hypothesis of schizophrenia has been neurochemically pre-eminent since the 1960s.(1) It proposes that the symptoms of schizophrenia result from dopaminergic overactivity, whether due to excess dopamine, or to an elevated sensitivity to it, for example because of an increased number of dopamine receptors. The hypothesis originated with two complementary observations: that effective antipsychotics were dopamine (D_2) receptor antagonists, and that dopamine-releasing agents such as amphetamine produce a paranoid psychosis.(2) It received support from various

findings of increased dopamine content and higher densities of D_2 receptors in post-mortem brain studies of schizophrenia, but proved difficult to refine or refute, for two main reasons. First, predictably, antipsychotics have marked effects on the dopamine system, confounding all studies of drug-treated subjects. Second, molecular biology revealed an unexpected complexity and diversity of dopaminergic genes, increasing the number of potential sites of dysfunction and mechanisms by which it might occur. For example, soon after the D_4 subtype of dopamine receptor was cloned, there were high profile reports that the receptor was up-regulated several-fold in schizophrenia, and might also be relevant for the actions of clozapine. However, neither suggestion was confirmed by further studies, and interest in this topic has subsided.

Despite these difficulties, substantial support for the dopamine hypothesis has now emerged, attributable largely to the availability of imaging-based methods to assess the dopamine system in the brain in vivo, free of medication, and post-mortem confounds. Notably, there is now strong evidence for a pre-synaptic dopamine abnormality, with several studies showing elevated dopamine synthesis, release, and higher dopamine receptor occupancy in the striatum.[3] The findings indicate a dysregulation and hyper-responsiveness of dopaminergic neurones in schizophrenia, sometimes referred to as 'hyperdopaminergia'. These abnormalities are present in patients with acute psychosis but not in patients in remission; recent data suggest that they may also occur in subjects in the prodrome of schizophrenia, and that they are localized to the associative parts of the striatum. It is not known whether the findings are specific to schizophrenia or common to other acute psychoses, whether they also affect dopamine pathways in the cerebral cortex, and whether they apply in all subjects with schizophrenia.[4] Neither is the cause of the hyperdopaminergia understood; one hypothesis is that it is downstream of a developmental deficit in the glutamatergic projections that regulate dopamine transmission,[5] another that it involves an imbalance between phasic and tonic modes of dopamine release.[6]

In addition to the excessive dopamine function associated with acute psychosis, there is also increasing evidence that deficiencies in dopamine transmission, especially in the dorsolateral prefrontal cortex, and genetically influenced, may underlie the working memory and allied cognitive deficits that occur in the disorder.[7] The relationship between these two facets of dopaminergic involvement in schizophrenia is not understood.

Glutamate

Phencyclidine and other non-competitive antagonists of the *N*-methyl-D-aspartate (NMDA) subtype of glutamate receptor produce a psychosis closely resembling schizophrenia.[8] This has driven the hypothesis of glutamatergic dysfunction in the disorder, particularly a disturbance of NMDA receptor-mediated glutamate transmission.[9] In support, drugs that enhance NMDA receptor function (via a variety of indirect mechanisms, since direct agonists are toxic) have some beneficial effects on positive, negative, and cognitive symptoms.[10] Also, impairment of NMDA receptor function in animal models, induced by either genetic or pharmacological manipulation, produces behavioural, structural, and neurochemical findings consistent with a 'schizophrenia-like phenotype'.[11,12] There is also a range of alterations in parameters of glutamate transmission in subjects with schizophrenia, including levels of glutamate receptors and of endogenous glutamate receptor modulators such as D-serine.[9,13] Interest in the glutamate system has been heightened with the realization that many of the putative susceptibility genes for schizophrenia have effects on NMDA receptors and their pathways[14] (see below).

Other neurotransmitters

A 5-hydroxytryptamine (5-HT, serotonin) involvement in schizophrenia was suggested because the hallucinogen lysergic acid diethylamide (LSD) is a 5-HT agonist. Recently, interest has focused on the 5-HT_{2A} receptor, for several reasons.[15] There is lowered 5-HT_{2A} receptor expression in the frontal cortex in schizophrenia, and a blunted neuroendocrine response to 5-HT_2 agonists; a high affinity for the receptor may contribute to the profile of atypical compared to typical antipsychotics, and variants in the gene are weakly associated with response to clozapine, and perhaps with schizophrenia. Elevated cortical 5-HT_{1A} receptors are also a replicated finding. Explanations for 5-HT involvement in schizophrenia include the trophic role of the 5-HT system in neurodevelopment, interactions between 5-HT and dopaminergic neurones, and impaired 5-HT_{2A} receptor-mediated activation of the prefrontal cortex.

GABA, the major inhibitory transmitter in the brain, has been implicated in schizophrenia, on the basis of findings of alterations in specific markers of GABAergic neurones and their connections as well as changes in GABA receptors.[16,17] The position of these alterations in the pathogenesis of schizophrenia is not known, as is also the case for the many other neurochemical differences that have been reported, e.g. in neuropeptides, endocannabinoids, muscarinic receptors, etc.[18]

Functional neuroimaging and cerebral activity

Cerebral activity in schizophrenia has been investigated by several methods, initially using positron emission tomography to measure regional cerebral blood flow and glucose utilization, and more recently using functional magnetic resonance imaging. The studies have addressed several questions: are there differences between cases and controls at rest, or correlations between patterns of activity and clinically defined groups of subjects, or does brain activation during the performance of cognitive tasks differ in those with and without the illness?

Hypofrontality—decreased activity in the frontal lobes—has been widely studied in schizophrenia since the first report in 1974. Results have broadly supported the notion, but with several important qualifications. The current view is that, whilst hypofrontality does occur in unmedicated subjects,[19] it is not an invariable finding, and may be related to clinical state.[20] Similarly, there are few other robust baseline differences in cerebral activity or perfusion between cases and controls. Instead, the focus has shifted towards the link between regional patterns of activation with specific symptoms, or with performance during cognitive tasks. Among the former category, a well-known example is that of Liddle *et al.*[21] who found that each of the three subsyndromes of chronic schizophrenia they had identified by factor analysis was associated with a different regional profile of cerebral blood flow. A relationship between superior temporal gyrus metabolic activity and auditory hallucinations has often been reported, sometimes lateralized to the left hemisphere.[22] Many other correlations between regional patterns of (de)activation and individual symptoms have also been reported.

A number of studies have investigated regional brain activation during the performance of various neuropsychological tests. For example, the hypofrontality of schizophrenia can be seen most clearly during working memory tasks, such as the Wisconsin Card Sort Test, which require activation of the frontal lobes, and at which patients are impaired. Conversely, when groups are matched for performance, subjects with schizophrenia show increased activation of these areas compared to controls, suggesting that they are less 'efficient' in how the information is processed, and requir greater 'effort' to achieve the same result. These issues illustrate that the situation is more complex than simply hypo- (or hyper-) frontality, but rather that there is a dynamic disturbance of frontal cortex function and regulation.[23] Beyond working memory and hypofrontality, a range of other specific correlations of this kind have been reported, but the key conclusion of this research is that cerebral dysfunction in schizophrenia is better conceptualized not as reflecting a static or single focal disorder, but as arising from abnormalities in distributed circuits linking specific cortical areas and subcortical nuclei. A prominent model is that of Andreasen,[24] who proposed the concept of 'cognitive dysmetria', in which deficits in activity in a circuit involving the cerebral cortex, thalamus, and cerebellum are key, and underlie the memory difficulties of schizophrenia. The view of the disorder as one of disturbed neural connectivity affecting multiple brain regions and their integration is supported by structural imaging and neuropathological data (see below). The model also highlights two other recent research themes: first, that brain areas beyond the 'traditional' ones (e.g. prefrontal cortex, hippocampus) and their interconnections are involved in the pathophysiology of the disorder; the most notable region of this kind is the cerebellum, formerly overlooked because of the erroneous view that it is entirely involved in motor control.[25] Second, the model places the cognitive deficits of schizophrenia centre stage in its pathophysiology, a view that was neglected for many years, but has regained prominence and is now widely cepted (Refs[26–28]; see Chapter 4.3.3).

Electrophysiology

A number of electrophysiological indices are altered in schizophrenia, and are relevant to the understanding of its neurobiology.[29] First, evoked potentials (electrical activity in the brain measured after a brief sensory stimulus); in particular, the P300 component is reduced and delayed in response to auditory and visual stimuli, indicative of impaired sensory processing.[30] Second, there is a high rate of eye movement abnormalities in schizophrenia, especially affecting smooth pursuit tracking, suggestive of impairment in the neural pathways subserving oculomotor control.[29] There are also differences in the cortical signal to noise ratio in the electroencephalogram, suggestive of an impairment of cortical information processing,[31] and consistent with the hypothesized abnormalities in cortical neural circuitry.

Structural neurobiology of schizophrenia

Finding the neuropathology of schizophrenia has been one of the major quests of biological psychiatry for over 100 years. Indeed, Alzheimer wrote a paper on the subject in 1897, 10 years before he described the disease that bears his name. However, whilst fundamental neuropathological discoveries were made in the dementias, there was no such progress for schizophrenia. In the past 20 years or so, the situation has changed. There is now compelling evidence that there is a neuropathology of schizophrenia, in the sense that there are statistically robust structural differences in the brains of patients with the disorder compared to normal subjects, both on structural imaging and at post-mortem (Table 4.3.6.2.1). On the other hand, the details and meaning of these changes are still elusive, and they are of limited clinical utility—they are not diagnostically specific, and they are only demonstrable when groups of cases and controls are compared.

Table 4.3.6.2.1 Morphological findings in schizophrenia

Replicated positive findings
Enlarged lateral and third ventricles
Decreased brain size and weight
Decreased cortical volume, especially temporal lobes
Fewer neurones in pulvinar thalamic nucleus
Decreased synaptic markers
Replicated negative findings
No increased incidence of Alzheimer's disease
No gliosis
Selected controversial findings
Increased density of cortical neurones
Smaller neurones
Reduced density of parvalbumin-positive interneurones
Aberrant distribution of white matter neurones
Fewer glia (oligodendrocytes)
Smaller mediodorsal thalamus with fewer neurones
Hemispheric asymmetry of pathology
Decreased dendritic markers
Effects of antipsychotic drugs on brain structure

Structural neuroimaging and macroscopic findings

The landmark study of Johnstone and colleagues showed, using computerized tomography, enlargement of the lateral ventricles in schizophrenia.[32] Although similar findings had been reported by pneumoencephalography, it was this paper which stimulated the field. It has been followed by many imaging studies, mostly in the last 20 years using magnetic resonance imaging,[33] and several meta-analyses. The latter show clearly that ventricular enlargement (with an average volume increase of ~40 per cent) is a feature of schizophrenia.[34–36] Accompanying this change there are decreases in cortical and whole brain volume of ~3 per cent,[36, 37] paralleled by a similar reduction of brain weight.[38] The regional localization of volume deficits is less clear, with different studies and meta-analyses implicating the hippocampus,[39] left superior temporal gyrus and medial temporal lobe (including hippocampus),[40] other regions of cerebral cortex, and thalamus.[41] For a narrative review of structural MRI studies, see Shenton *et al.*[42]

Structural brain changes are present in first episode patients.[43] Some differences are also present in subjects before they develop psychosis,[44,45] as well as in unaffected relatives,[46] indicating that part of the structural pathology is related to risk for schizophrenia (whether genetic or otherwise). Equally, other volumetric changes develop in high-risk subjects when they develop psychosis (e.g. hippocampal volume loss), suggesting that these changes are state-related rather than trait-related.[45, 47]

Two issues regarding structural imaging in schizophrenia remain controversial. First, the extent to which changes are progressive

after the onset of established illness.[48] The meta-analyses of the cross-sectional studies show no clear evidence of progression (in keeping with the stability of cognitive impairments, and the nature of the neuropathology to be described). On the other hand, several longitudinal studies do report greater shrinkage of various brain regions with time compared to control subjects, leading to pathophysiological theories related to aberrant plasticity and neurotoxicity.[49,50] The second controversy, which may be related to the first, concerns medication effects. Again, there is little consensus: there are positive and negative reports concerning effects of antipsychotics on whole or regional brain volumes, and some suggesting differential effects of typical versus atypical antipsychotics.[51]

Recent studies are using novel imaging methods to assess the status of white matter tracts in schizophrenia, to investigate hypotheses of aberrant anatomical connectivity.[52] A number of abnormalities have been demonstrated, broadly consistent with these notions,[53] although their interpretation (i.e. what is different functionally and/or anatomically) is not wholly clear.

The neuropathology is not degenerative

Despite the continuing uncertainties, the appreciation that there are structural brain changes in schizophrenia in terms of magnetic resonance imaging findings helped stimulate a new generation of morphometric and molecular post-mortem studies designed to determine the histological and cellular basis of the observations.

The most robust and important histological findings in schizophrenia are both negative.[54] Firstly, the neuropathology is not degenerative[55–57]: there are no lesions such as neurofibrillary tangles, amyloid plaques, or Lewy bodies, which would indicate the presence of any known neurodegenerative process.[56,57] Importantly, this conclusion even applies to the significant subgroup of elderly patients who develop dementia; the neuropathological basis for the dementia of schizophrenia is entirely unexplained.[56] Secondly, there is no excess of gliosis in the brains of patients with schizophrenia.[58] Gliosis, the proliferation and hypertrophy of astrocytes, is a sign of inflammation, injury, or other ongoing pathological processes. Hence the lack of gliosis is taken to denote that the disorder is likely to be neurodevelopmental in origin, affecting mechanisms involved in the normal maturation of the brain. Indeed, some recent studies suggest there is actually a decrease in the number or activity of some glial cells, an issue returned to later.

Morphometric and cytoarchitectural changes

Having ruled out these important possibilities, it has been difficult to pin down just what the histological changes are, and therefore the cellular basis for decreases in regional brain volumes. Nevertheless, the positive findings can be grouped together and viewed as broadly cytoarchitectural in nature—i.e. affecting the morphology and spatial organization of neurones and their processes (Ref.[54]; Table 4.3.6.2.1).

As a rule, the more dramatic (and well publicized) the initial finding, the less robust it has proved to be. For example, dysplasia (disorganized, misplaced, and misshapen neurones) in the entorhinal cortex was reported in 1986. Such a finding would be strongly suggestive of a prenatal developmental anomaly. However, subsequent studies have, at best, only partially replicated this observation. Similarly, a report that pyramidal neurones in the hippocampus are not aligned in their usual regular orientation, also indicative of a developmental disturbance, has not been consistently observed.

A decreased size of neurones, particularly in the hippocampus and prefrontal cortex, has been found in several studies. The size of a neurone is related to the volume of axon and dendrites, which it has to support, and also to its metabolic activity. Thus, the finding of smaller neurones in schizophrenia suggests the neurones may be receiving and making fewer, abnormal, or less-active connections. Support for this interpretation comes from studies of synaptic and dendritic markers, which have been reasonably consistent in showing decreases in the same brain areas.[59] It is unclear as to which specific populations of neurones and synapses are most affected in schizophrenia, and it may vary from one region to another; in both hippocampus and prefrontal cortex, there is evidence for involvement of excitatory (glutamatergic) pathways[60,61] as well subtypes of inhibitory (GABAergic) ones,[16,17] especially the class of parvalbumin-positive interneurones.[17] In addition to the neuronal pathology, several recent studies show reductions in markers of oligodendrocytes and their activity.[62] This type of glial cell is intimately involved in myelination, and contributes to synaptic homeostasis, and therefore their involvement in schizophrenia is in keeping with the occurrence of synaptic as well as white matter pathology.

A further area of interest is the thalamus, specifically the pulvinar and mediodorsal nuclei. Both have been found to be smaller and to contain fewer neurones in several studies of schizophrenia; in the case of the pulvinar nucleus, the evidence is amongst the most compelling of any brain region, coming from four methodologically rigorous studies (as well as complementary findings in the imaging literature, noted above), and with no corresponding negative studies.[63] These thalamic nuclei have extensive reciprocal connections with the prefrontal and temporal association cortices, and it is assumed that there is some causal link between the changes in each thalamic nucleus and its cortical partner.

Neuropathology and medication effects

Most patients studied neuropathologically were treated in life with antipsychotic drugs, and so the findings in schizophrenia are open to the criticism that they may have been caused by antipsychotic medication.[64] However, as noted above, many imaging studies show that the pathology, at least in terms of the gross alterations summarized in Table 4.3.6.2.1, is present in first episode and medication-naïve subjects. Also, in post-mortem studies, the reported neuropathological findings rarely if ever correlate with the extent of antipsychotic exposure. In some instances it may also be that medication ameliorates the disease effects. Nevertheless, the possibility that antipsychotic drugs have neuropathological effects should not be overlooked. Firstly, typical antipsychotics produce enlargement of, and synaptic structural alterations in, the basal ganglia (caudate, putamen, globus pallidus).[64] Secondly, a recent monkey study found that chronic administration of haloperidol or olanzapine at therapeutic levels led to decreased brain volume[65] along with increased neuronal density and decreased glial density,[66] thus reproducing several of the changes reported in schizophrenia.

Neurobiological theories of schizophrenia

Schizophrenia as a neurodevelopmental disorder

The neurodevelopmental model of schizophrenia is the prevailing pathogenic hypothesis.[5,67–69] Neurobiological data form an

important component of the evidence, along with epidemiological and other observations (Table 4.3.6.2.2). A specific version of the theory is that the pathology of schizophrenia originates in the second trimester *in utero*. An earlier timing is excluded since overt brain abnormalities would be seen if neurogenesis were affected, whilst the lack of gliosis has been taken to mean that the changes must have occurred prior to the third trimester when the gliotic response begins.[57] Other forms of the neurodevelopmental theory advocate abnormalities in processes such as myelination, synaptic pruning, and apoptosis, all of which continue long into post-natal life. Overall, a parsimonious view is that the neuropathological data are indicative merely of a basically developmental, as opposed to degenerative, disease process, but do not, in isolation, point to a particular mechanism or timing.

Schizophrenia as a disorder of connectivity

The functional imaging and histological data summarized above have together contributed to the emerging consensus that the pathophysiological basis of schizophrenia is one of connectivity.[24, 54, 70–72] The nature of the 'dysconnectivity' is not a simple lack, or gross mis-routing, of connections, but likely a subtle change in a more fine-grained aspect, such as the precise molecular composition, location, or activity of subpopulations of synapses.[71] Both intrinsic (local) and extrinsic (long-range) connections may be affected. The extent to which there is a structural basis to the aberrant connectivity remains uncertain,[73] but in so far as it exists, the histological basis of the syndrome is a difference in the circuitry, or 'wiring' of the brain, manifested by the cytoarchitectural differences in the morphology and organization of neurones and their synapses. If there is indeed a structural basis to the pathophysiology of the syndrome, it would help explain why many of the cardinal features are trait rather than just state abnormalities, and perhaps why individuals are vulnerable to relapse, in that a 'miswired' brain may be less able to respond rapidly, appropriately, or fully to environmental stressors.

Table 4.3.6.2.2 Evidence adduced for a neurodevelopmental basis to schizophrenia

- Usual age of onset in late adolescence and early adulthood
- Most environmental risk factors operate prenatally or in early childhood
- Neuromotor, intellectual, and behavioural differences in children many years prior to onset
- Structural brain changes and cognitive impairments present at, and prior to, onset of illness
- Nature of the neuropathological findings—lack of gliosis or neurodegeneration, plus cytoarchitectural disturbances
- Changes in expression of neurodevelopmental genes (e.g. reelin)
- Minor physical anomalies—for example, craniofacial dysmorphology and dermatoglyphics.
- Animal models—early lesions or pharmacological interventions lead to delayed 'schizophrenia-like' phenotypes

(Reprinted from PJ. Harrison, Schizophrenia susceptibility genes and neurodevelopment, *Biological Psychiatry*, **61**, 1119–20, copyright 2007, with permission from Elsevier)

For reviews and other references, see text.

Cerebral asymmetry and schizophrenia

Many neuropathological, neurochemical, neuropsychological, and electrophysiological studies of schizophrenia report lateralized abnormalities. Although there are also important negative findings, reductions in normal brain asymmetries, and a left hemisphere preference of the pathology, do seem more common than one would expect by chance.[74,75] Crow's influential theory sees a fundamental connection between schizophrenia, asymmetry, handedness, and language, causally linked to each other and to the same gene.[76] Alternatively, altered asymmetry in schizophrenia is viewed as an epiphenomenon of its developmental origins, a process which interferes with normal brain lateralization.

Susceptibility genes and neurobiology

Given its high heritability,[77] it can be assumed that genes are the major influence on the neurobiological features of schizophrenia, likely modified by the various environmental risk factors (see Chapter 4.3.6.1). The recent discovery of several probable susceptibility genes[14,78] now allows this question to be addressed more specifically, in terms of the normal functions of the genes, and how this is altered in those carrying the risk variants of the genes.

Neuregulin 1 (NRG1) is the best established susceptibility gene.[79] It encodes a family of proteins that have multiple roles in the nervous system, ranging from cell fate determination, to neuronal migration, neuronal-glial signalling, and NMDA receptor functioning.[80] As such, it is a good candidate gene for schizophrenia given the existing theories outlined above regarding aberrations in neurodevelopment, connectivity, and glutamate synaptic transmission. Equally, at present it is not clear which of these functions is actually affected in schizophrenia or explains the contribution that NRG1 plays in the disease process,[81] and determining this will not be simple. Initial data suggest an impairment of NRG1 signalling via its receptors,[82] and an alteration in the expression of one specific subtype of NRG1.[83,84]

An interaction with synaptic neurotransmission is also seen for most of the other leading susceptibility genes (Table 4.3.6.2.3), leading to the notion that this effect is a point of pathophysiological convergence of the genes.[14] Such a convergence is an attractive concept for several reasons, not least parsimony, but whilst there is some evidence to support it, it remains highly speculative. Indeed,

Table 4.3.6.2.3 Susceptibility genes and their neurobiological functions

Gene symbol	Functions include
NRG1	Multiple roles in brain development, synaptic plasticity, and glutamate signalling
DTNBP1	Glutamate release
DISC-1	Multiple roles in development, cell functioning, and synaptic signalling
PPP3CC	Critical molecule for integration of dopamine and glutamate signalling
DAOA	Affects metabolism of the NMDA receptor modulator D-serine
COMT	Regulation of dopamine function in frontal cortex

NRG1, neuregulin 1; DTNBP1, dysbindin; DISC-1, disrupted in schizophrenia-1; PPP3CC, calcineurin Aγ subunit; DAOA, D-amino acid oxidase activator COMT, catechol-*O*-methyltransferase.

Table 4.3.6.2.4 Key recent findings in the neurobiology of schizophrenia

- Elevated dopamine release and receptor occupancy during acute psychosis
- Confirmation (by meta-analysis) of structural brain changes, including in first-episode patients, and in unaffected relatives
- Exclusion of a neurodegenerative disease process
- Discovery of neuregulin and several other putative susceptibility genes, and of molecular mechanisms by which they may increase schizophrenia susceptibility

Table 4.3.6.2.5 Current key questions to be addressed in the neurobiology of schizophrenia

- Which, if any, of the neurobiological findings are clinically useful (i.e. will influence diagnosis, treatment, or prognosis)?
- What is the relationship between the dopaminergic changes, and those affecting other neurochemical systems and brain structure?
- Are there differing neurobiological substrates for the psychotic, cognitive, and negative components of schizophrenia?
- How do the genes interact with each other, and with the environmental risk factors, to produce their effects?

it will probably remain difficult to determine in detail the genetic contribution to the neurobiology of schizophrenia, given that there are many genes, each of small and complex effects.[85]

Summary

Significant progress has been made in understanding the neurobiology of schizophrenia over the past decade (Table 4.3.6.2.4). In particular, there is now good evidence for a dopaminergic dysfunction, and for structural brain changes that are present at, and in part before, the onset of illness. There is also emerging evidence for several susceptibility genes, accompanied by data suggesting mechanisms by which these genes contribute to the neurodevelopmental and other pathogenic processes that are thought to lead to schizophrenia. Whilst highlighting the progress, one must also acknowledge that much remains unknown (Table 4.3.6.2.5), and it is a moot point how and when the research advances will impact on the diagnosis, treatment, or prognosis of schizophrenia.

Further information

Lawrie, S.A., Weinberger, D.R., Johnstone, E.C. (2004). *Schizophrenia. From neuroimaging to neuroscience*. Oxford University Press, Oxford.

Schizophrenia Research Forum (www.schizophreniaforum.org)—links to all aspects of schizophrenia research, including up-to-date bibliographies, discussion forum, and a genetics database.

Weinberger, D.R. and Harrison, P.J. (2009). *Schizophrenia* (3rd edn). Wiley Blackwell, Oxford.

References

1. Matthyse, S. (1973). Antipsychotic drug actions: a clue to the neuropathology of schizophrenia? *Federation Proceedings*, **32**, 200–5.
2. Davis, K.L., Kahn, R.S., Ko, G., *et al.* (1991). Dopamine in schizophrenia: a review and reconceptualization. *The American Journal of Psychiatry*, **148**, 1474–86.
3. Laruelle, M. and Abi-Dargham, A. (1999). Dopamine as the wind of the psychotic fire: new evidence from brain imaging studies. *Journal of Psychopharmacology*, **13**, 358–71.
4. Abi-Dargham, A., Rodenhiser, J., Printz, D., *et al.* (2000). Increased baseline occupancy of D_2 receptors by dopamine in schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 8104–9.
5. Weinberger, D.R. (1987). Implications of normal brain development for the pathogenesis of schizophrenia. *Archives of General Psychiatry*, **44**, 660–9.
6. Grace, A.A. (1993). Cortical regulation of subcortical dopamine systems and its possible relevance to schizophrenia. *Journal of Neural Transmission*, **91**, 111–34.
7. Weinberger, D.R., Egan, M.F., Bertolino, A., *et al.* (2001). Prefrontal neurons and the genetics of schizophrenia. *Biological Psychiatry*, **50**, 825–44.
8. Javitt, D.C. and Zukin, S.R. (1991). Recent advances in the phencyclidine model of schizophrenia. *The American Journal of Psychiatry*, **148**, 1301–8.
9. Tsai, G. and Coyle, J.T. (2002). Glutamatergic mechanisms in schizophrenia. *Annual Review of Pharmacology and Toxicology* **42**, 165–79.
10. Tuominen, H.J., Tiihonen, J., and Wahlbeck, K. (2005). Glutamatergic drugs for schizophrenia: a systematic review and meta-analysis. *Schizophrenia Research*, **72**, 225–34.
11. Lipska, B.K. and Weinberger, D.R. (2000). To model a psychiatric disorder in animals: schizophrenia as a reality test. *Neuropsychopharmacology*, **23**, 223–39.
12. Tordjman, S., Drapier, D., Bonnot, O., *et al.* (2007). Animal models relevant to schizophrenia and autism: validity and limitations. *Behavioral Genetics*, **37**, 61–78.
13. Konradi, C. and Heckers, S. (2003). Molecular aspects of glutamate dysregulation: implications for schizophrenia and its treatment. *Pharmacology & Therapeutics*, **97**, 153–79.
14. Harrison, P.J. and Weinberger, D.R. (2005). Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Molecular Psychiatry*, **10**, 40–68.
15. Dean, B. (2003). The cortical serotonin$_{2A}$ receptor and the pathology of schizophrenia: a likely accomplice. *Journal of Neurochemistry*, **85**, 1–13.
16. Benes, F.M. and Berretta, S. (2001). GABAergic interneurons: implications for understanding schizophrenia and bipolar disorder. *Neuropsychopharmacology*, **25**, 1–27.
17. Lewis, D.A., Hashimoto, T., and Volk, D.W. (2005). Cortical inhibitory neurons and schizophrenia. *Nature Reviews Neuroscience*, **6**, 312–24.
18. Fallon, J.H., Opole, I.O., and Potkin, S.G. (2003). The neuroanatomy of schizophrenia: circuitry and neurotransmitter systems. *Clinical Neuroscience Research*, **3**, 77–107.
19. Andreasen, N.C., O'Leary, D.S., Flaum, M., *et al.* (1997). Hypofrontality in schizophrenia: distributed dysfunctional circuits in neuroleptic-naive patients. *Lancet*, **349**, 1730–4.
20. Spence, S.A., Hirsch, S.R., Brooks, D.J., *et al.* (1998). Prefrontal cortex activity in people with schizophrenia and control subjects. Evidence from positron emission tomography for remission of 'hypofrontality' with recovery from acute schizophrenia. *The British Journal of Psychiatry*, **172**, 316–23.
21. Liddle, P.F., Friston, K.J., Frith, C.D., *et al*, (1992). Patterns of cerebral blood flow in schizophrenia. *The British Journal of Psychiatry*, **160**, 179–86.
22. Weiss, A.P. and Heckers, S. (1999). Neuroimaging of hallucinations: a review of the literature. *Psychiatry Research: Neuroimaging*, **92**, 61–74.

23. Callicott, J.H., Mattay, V.S., Verchinski, B.A., *et al.* (2003). Complexity of prefrontal cortical dysfunction in schizophrenia: more than up or down. *The American Journal of Psychiatry*, **160**, 2209–15.
24. Andreasen, N.C. (1998). A unitary model of schizophrenia: Bleuler's 'fragmented phrene' as schizencephaly. *Archives of General Psychiatry*, **56**, 781–7.
25. Picard, H., Amado, I., Mouchet-Mages, S., *et al.* (2008). The role of the cerebellum in schizophrenia: an update of clinical, cognitive, and functional evidences. *Schizophrenia Bulletin*, **34**, 155–72.
26. Aleman, A., Hijman, R., De Haan, E.H.F., *et al.* (1999). Memory impairment in schizophrenia: a meta-analysis. *The American Journal of Psychiatry*, **156**, 1358–66.
27. Green, M.F. and Neuchterlein, K.H. (1999). Should schizophrenia be treated as a neurocognitive disorder? *Schizophrenia Bulletin*, **25**, 309–19.
28. Goldberg, T.E., David, A., and Gold, J.M. (2003). Neurocognitive deficits in schizophrenia. In *Schizophrenia* (2nd edn) (eds. S.C. Hirsch and D.R. Weinberger), pp. 168–86. Blackwell, Oxford.
29. Salisbury, D.F., Krljes, S., and McCarley, R.W. (2003). Electrophysiology. In *Schizophrenia* (2nd edn) (eds. S.C. Hirsch and D.R. Weinberger), pp. 298–309. Blackwell, Oxford.
30. Bramon, E., Rabe-Hesketh, S., Sham, P., *et al.* (2004). Meta-analysis of the P300 and P50 waveforms in schizophrenia. *Schizophrenia Research*, **70**, 315–29.
31. Winterer, G. and Weinberger, D.R. (2004). Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. *Trends in Neurosciences*, **27**, 683–90.
32. Johnstone, E.C., Crow, T.J., Frith, C.D., *et al.* (1976). Cerebral ventricular size and cognitive impairment in chronic schizophrenia. *Lancet*, **2**, 924–6.
33. Andreasen, N.C., Nasrallah, H.A., Dunn, V., *et al.* (1986). Structural abnormalities in the frontal system in schizophrenia. A magnetic resonance imaging study. *Archives of General Psychiatry*, **43**, 136–44.
34. Daniel, D.G., Goldberg, T.E., Gibbons, R.D., *et al.* (1991). Lack of a bimodal distribution of ventricular size in schizophrenia—a Gaussian mixture analysis of 1056 cases and controls. *Biological Psychiatry*, **30**, 887–903.
35. Lawrie, S.M. and Abukmeil, S.S. (1998). Brain abnormality in schizophrenia—a systematic and quantitative review of volumetric magnetic resonance imaging studies. *The British Journal of Psychiatry*, **172**, 110–20.
36. Wright, I.C., Rabe-Hesketh, S., Woodruff, P.W.R., *et al.* (2000). Meta-analysis of regional brain volumes in schizophrenia. *The American Journal of Psychiatry*, **157**, 16–25.
37. Ward, K.E., Friedman, L., Wise, A., *et al.* (1996). Meta-analysis of brain and cranial size in schizophrenia. *Schizophrenia Research*, **22**, 197–213.
38. Harrison, P.J., Freemantle, N., and Geddes, J.R. (2003). Meta-analysis of brain weight in schizophrenia. *Schizophrenia Research*, **64**, 25–34.
39. Nelson, M.D., Saykin, A.J., Flashman, L.A., *et al.* (1998). Hippocampal volume reduction in schizophrenia as assessed by magnetic resonance imaging—a meta-analytic study. *Archives of General Psychiatry*, **55**, 433–40.
40. Honea, R., Crow, T.J., Passingham, D., *et al.* (2005). Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. *The American Journal of Psychiatry*, **162**, 2233–45.
41. Konick, L.C. and Friedman, L. (2001). Meta-analysis of thalamic size in schizophrenia. *Biological Psychiatry*, **49**, 28–38.
42. Shenton, M.E., Dickey, C.C., Frumin, M., *et al.* (2001). A review of MRI findings in schizophrenia. *Schizophrenia Research*, **49**, 1–52.
43. Steen, R.G., Mull, C., McClure, R., *et al.* (2006). Brain volume in first-episode schizophrenia—systematic review and meta-analysis of magnetic resonance imaging studies. *The British Journal of Psychiatry*, **188**, 510–18.
44. Lawrie, S.M., Whalley, H., Kestelman, J.N., *et al.* (1999). Magnetic resonance imaging of brain in people at high risk of developing schizophrenia. *Lancet*, **353**, 30–3.
45. Pantelis, C., Velakoulis, D., McGorry, P.D., *et al.* (2003). Neuroanatomical abnormalities before and after onset of psychosis: a cross-sectional and longitudinal MRI comparison. *Lancet*, **361**, 281–8.
46. Boos, H.B.M., Aleman, A., Cahn, W., *et al.* (2007). Brain volumes in relatives of patients with schizophrenia—a meta-analysis. *Archives of General Psychiatry*, **64**, 297–304.
47. Velakoulis, D., Wood, S.J., Wong, M.T.H., *et al.* (2006). Hippocampal and amygdala volumes according to psychosis stage and diagnosis—a magnetic resonance imaging study of chronic schizophrenia, first-episode psychosis, and ultra-high-risk individuals. *Archives of General Psychiatry*, **63**, 139–49.
48. Weinberger, D.R. and McClure, R.K. (2002). Neurotoxicity, neuroplasticity, and magnetic resonance imaging morphometry—what is happening in the schizophrenic brain? *Archives of General Psychiatry*, **59**, 553–8.
49. Lieberman, J.A. (1999). Is schizophrenia a neurodegenerative disorder? A clinical and neurobiological perspective. *Biological Psychiatry*, **46**, 729–39.
50. Woods, B.T., Ward, K.E., and Johnson, E.H. (2005). Meta-analysis of the time-course of brain volume reduction in schizophrenia: implications for pathogenesis and early treatment. *Schizophrenia Research*, **73**, 221–8.
51. Lieberman, J.A., Tollefson, G.D., Charles, C., *et al.* (2005). Antipsychotic drug effects on brain morphology in first-episode psychosis. *Archives of General Psychiatry*, **62**, 361–70.
52. Davis, K.L., Stewart, D.G., Friedman, J.I., *et al.* (2003). White matter changes in schizophrenia—evidence for myelin-related dysfunction. *Archives of General Psychiatry*, **60**, 443–56.
53. Kubicki, M., McCarley, R., Westin, C.F., *et al*, (2007). A review of diffusion tensor imaging studies in schizophrenia. *Journal of Psychiatric Research*, **41**, 15–30.
54. Harrison, P.J. (1999). The neuropathology of schizophrenia. A critical review of the data and their interpretation. *Brain*, **122**, 593–624.
55. Harrison, P.J. (2003). Schizophrenia and its dementia. In *The neuropathology of dementia* (eds. M. Esiri, V. Lee, and J. Trojanowski), pp. 497–508. Cambridge University Press, Cambridge.
56. Arnold, S.E. and Trojanowski, J.Q. (1996). Cognitive impairment in elderly schizophrenia: a dementia (still) lacking distinctive histopathology. *Schizophrenia Bulletin*, **22**, 5–9.
57. Arnold, S.E., Trojanowski, J.Q., Gur, R.E., *et al.* (1998). Absence of neurodegeneration and neural injury in the cerebral cortex in a sample of elderly patients with schizophrenia. *Archives of General Psychiatry*, **55**, 225–32.
58. Roberts, G.W. and Harrison. P.J. (2000). Gliosis and its implications for the disease process. In *The neuropathology of schizophrenia* (eds. P.J. Harrison and G.W. Roberts), pp. 137–50. Oxford University Press, Oxford.
59. Eastwood, S.L. (2004). The synaptic pathology of schizophrenia: is aberrant neurodevelopment and plasticity to blame? *International Review of Neurobiology*, **59**, 47–72.
60. Lewis, D.A. and Gonzalez-Burgos, G. (2000). Intrinsic excitatory connections in the prefrontal cortex and the pathophysiology of schizophrenia. *Brain Research Bulletin*, **52**, 309–17.
61. Hof, P.R., Haroutunian, V., Copland, C., *et al.* (2002). Molecular and cellular evidence for an oligodendrocyte abnormality in schizophrenia. *Neurochemical Research*, **27**, 1193–200.
62. Harrison, P.J. and Eastwood, S.L. (2001). Neuropathological studies of synaptic connectivity in the hippocampal formation in schizophrenia. *Hippocampus*, **11**, 508–19.
63. Byne, W., Fernandes, J., Haroutunian, V., *et al.* (2007). Reduction of right medial pulvinar volume and neuron number in schizophrenia. *Schizophrenia Research*, **90**, 71–5.

64. Harrison, P.J. (1999). The neuropathological effects of antipsychotic drugs. *Schizophrenia Research*, **40**, 87–99.
65. Dorph-Petersen, K.A., Pierri, J.N., Perel, J.M., *et al.* (2005). The influence of chronic exposure to antipsychotic medications on brain size before and after tissue fixation: a comparison of haloperidol and olanzapine in macaque monkeys. *Neuropsychopharmacology*, **30**, 1649–61.
66. Konopaske, G.T., Dorph-Petersen, K.A., Pierri, J.N., *et al.* (2007). Effect of chronic exposure to antipsychotic medication on cell numbers in the parietal cortex of macaque monkeys. *Neuropsychopharmacology*, **32**, 1216–23.
67. Murray, R.M. and Lewis, S.W. (1987). Is schizophrenia a neurodevelopmental disorder. *British Medical Journal*, **295**, 681–2.
68. Weinberger, D.R. (1995). From neuropathology to neurodevelopment. *Lancet*, **346**, 552–7.
69. Lewis, D.A. and Levitt, P. (2002). Schizophrenia as a disorder of neurodevelopment. *Annual Review of Neuroscience*, **25**, 409–32.
70. McGuire, P.K. and Frith, C.D. (1996). Disordered functional connectivity in schizophrenia. *Psychological Medicine*, **26**, 663–7.
71. Frankle, W., Lerma, J., and Laruelle, M. (2003). The synaptic hypothesis of schizophrenia. *Neuron*, **39**, 205–16.
72. Stephan, K.E., Baldeweg, T., and Friston, K.J. (2006). Synaptic plasticity and dysconnection in schizophrenia. *Biological Psychiatry*,**59**, 929–93.
73. Konrad, A. and Winterer, G. (2008). Disturbed structural connectivity in schizophrenia—primary factor in pathology or epiphenomenon? *Schizophrenia Bulletin*, **34**, 72–92.
74. Holinger, D.P., Galaburda, A., and Harrison, P.J. (2000). Cerebral asymmetry. In *The neuropathology of schizophrenia* (eds. P.J. Harrison and G.W. Roberts), pp. 151–72. Oxford University Press, Oxford.
75. Sommer, I., Aleman, A., Ramsey, N., *et al.* (2001). Handedness, language lateralisation and anatomical asymmetry in schizophrenia—meta-analysis. *The British Journal of Psychiatry*, **178**, 344–51.
76. Crow, T.J. (2004). Cerebral asymmetry and the lateralization of language: core deficits in schizophrenia as pointers to the gene. *Current Opinion in Psychiatry*, **17**, 97–106.
77. Sullivan, P.F., Kendler, K.S., and Neale, M.C. (2003). Schizophrenia as a complex trait—evidence from a meta-analysis of twin studies. *Archives of General Psychiatry*, **60**, 1187–92.
78. Owen, M.J., Craddock, N., and O'Donovan, M.C. (2005). Schizophrenia: genes at last? *Trends in Genetics*, **21**, 518–25.
79. Li, D.W., Collier, D.A., and He, L. (2006). Meta-analysis shows strong positive association of the neuregulin 1 (*NRG1*) gene with schizophrenia. *Human Molecular Genetics*, **15**, 1995–2002.
80. Harrison, P.J. and Law, A.J. (2006). Neuregulin 1 and schizophrenia: genetics, gene expression, and neurobiology. *Biological Psychiatry*, **60**, 132–40.
81. Corfas, G., Roy, K., and Buxbaum, J.D. (2004). Neuregulin 1-erbB signaling and the molecular/cellular basis of schizophrenia. *Nature Neuroscience*, **7**, 575–80.
82. Hahn, C.G., Wang, H.Y., Cho, D.S., *et al.* (2006). Altered neuregulin 1-erbB4 signaling contributes to NMDA receptor hypofunction in schizophrenia. *Nature Medicine*, **12**, 824–8.
83. Law, A.J., Lipska, B.K., Weickert, C.S., *et al.* (2006). Neuregulin 1 transcripts are differentially expressed in schizophrenia and regulated by 5′ SNPs associated with the disease. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 6747–52.
84. Tan, W., Wang, Y., Gold, B., *et al.* (2007). Molecular cloning of a brain-specific, developmentally regulated neuregulin 1 (NRG1) isoform and identification of a functional promoter variant associated with schizophrenia. *The Journal of Biological Chemistry*, **282**, 24343–51.
85. Harrison, P.J. (2008). Schizophrenia genes: searching for common features, functions, and mechanisms. In *Cortical deficits in schizophrenia: from genes to function* (ed. P. O'Donnell), pp. 1–16. Springer, Norwell, MA.

4.3.7 Course and outcome of schizophrenia and their prediction

Assen Jablensky

Introduction

The course of schizophrenia is as variable as its symptoms. Systematic investigations of course of the psychoses were initiated by Kraepelin who believed that, in the absence of demonstrable brain pathology and aetiology, a common outcome into 'psychic weakness' of the clinical syndromes he grouped together as **dementia praecox** would provide a validity test for the disease entity. Later, Kraepelin revised his claim that the prognosis of dementia praecox was invariably poor and noted that 'permanent cures' had occurred in about 15 per cent of his cases.[1] Subsequent longitudinal studies have confirmed the striking variability of course as one of the salient characteristic of the 'natural history' of schizophrenia.

Methodological issues

The large number of studies on the course and outcome of schizophrenia published since the beginning of the twentieth century might suggest that the longitudinal aspects of the disorder are well established and exhaustively documented. Unfortunately, this is not the case since the methodological difficulties that accompany this type of research are complex and few studies have adequately dealt with all the major sources of error and confounding, including sample selection, definition of outcome, and diagnostic criteria used.[2,3]

The studies of the course and outcome of schizophrenia comprise statistical reports on admissions and discharges, long-term **follow-back studies** (in which the initial features of the cases and the course of the disorder are reconstructed retrospectively from admission records), and prospective investigations (in which patients are enlisted at an early stage of the disorder and followed up for a varying length of time). Each design is vulnerable to bias: admission and discharge statistics usually comprise patients at different stages of disease progression; follow-back studies rely on prevalence samples in which chronic cases tend to be over-represented; and **prospective studies**, though superior to other designs, tend to exclude patients who initially have diagnoses other than schizophrenia but are subsequently re-diagnosed as schizophrenic. The methodological issues that need to be considered in interpreting the results from longitudinal research into schizophrenia include the following.

Diagnosis

The use of either 'broad' or 'restrictive' definitions of schizophrenia may result in vastly different samples on which follow-up data are reported. Systems with an inbuilt illness duration criterion, such as **DSM-III**, **DSM-IIIR**, and **DSM-IV** which require at least 6 months of unremitting symptoms and a decline in functioning, are likely to overselect patients already developing a chronic course. The result would be a greater homogeneity of outcome at the cost of a compromised representativeness of the sample as regards the range of

possible outcomes of schizophrenia. Diagnostic systems which emphasize the cross-sectional features of the disorder, such as **ICD-10** (which requires 1 month's duration of clinically characteristic symptoms) avoid this limitation, possibly at the expense of including some cases of good prognosis that may be aetiologically or pathogenetically different from poor prognosis schizophrenia. However, until aetiology is elucidated, or validating biomarkers are established, the decision as to what constitutes 'true' schizophrenia will remain arbitrary. With regard to prognostic studies, less restrictive diagnostic systems have the advantage that a broad spectrum of outcomes would be available at the end point of prospective observation, allowing for subgroups to be identified and their characteristics related to the initial manifestations of the disorder and various risk factors.

Definitions and assessment of course and outcome variables

There is no single measure of course and outcome of a complex disorder such as schizophrenia. Blanket terms such as 'recovery', 'improvement', or 'deterioration' tend to conflate substantially different aspects of the evolution of the disorder over time. Most investigators today agree that course (comprising the pathways or trajectories of the disorder) and outcome (the net balance of the clinical and functional descriptors at the end point of observation) are multivariate composites. As a minimum, three domains that need not co-vary over time should be independently assessed: symptom severity; functional impairments including cognitive deficits, and disablement in social and occupational role performance. Each one of these can be further articulated into a number of areas or dimensions. In addition, one must consider extrinsic variables such as measures of environmental and treatment-related influences on course and outcome, as well as subjective experiences commonly described as 'quality of life'. Standardized, reliable instruments (interviews, inventories, rating scales) are required for the assessment of most variables. **Operational definitions** and criteria of relapse and remission have been proposed.[4,5] It should not be forgotten, however, that some of the richest sources of information are the perceptive, in-depth clinical case studies based on personal patient contact over many years. Collectively, such single case observations can generate hypotheses for testing in epidemiologically designed studies.

Length of follow-up

The evidence from previous research suggests that very different impressions of the course and outcome of schizophrenia would be gained depending on the duration of prospective observation and the degree of control over the inclusion of patients that are comparable in terms age and length of previous illness.

Cohort attrition

In any follow-up study, a proportion of cases will be lost to observation because of death, migration, refusal of contact, or other reasons for untraceability. Since such loss of subjects is likely to correlate with particular patterns of course and outcome, it is essential to estimate its possible effect (e.g. by statistical modelling) on the interpretation of the final results, especially if cohort attrition is greater than 15–20 per cent of the original sample.

Other aspects of study design

Variation in the sources of recruitment of cases (e.g. any admission to a treatment facility or catchment area sampling), and of information regarding course and outcome variables (e.g. face-to-face interviews or collateral data from case notes or informants), can obviously influence the results of any follow-up study. In addition, subtle variations in study design, such as whether investigators assessing patients' symptoms and functioning at any point in time are 'blind' to data from previous assessments, can bias the final results. Inclusion of a comparison group (e.g. patients with other psychotic disorders) would help evaluate the extent to which any observed patterns of course and outcome are specific to schizophrenia, whereas appropriate controls drawn from the general population can provide reference points for assessing social variables, such as occupational functioning, stressful life events, or habit-related behaviour such as substance use.

Statistical analysis

Longitudinal research poses a number of specific requirements with regard to data analysis. Thus, the problem of **multiple comparisons** is likely to arise when examining the data for significant associations; time series, survival, or path analysis may be required when observations are made and recorded at successive time points in the evolution of the disorder; and methods of unconfounding are called for at each step of the analysis of longitudinal data. While no single study up to date has met all the rigorous methodological requirements, a number of studies have succeeded in controlling at least some of the sources of bias and confounding. The results from previous research are, therefore, not strictly comparable in specific detail, but are informative as regards general trends and patterns.

The 'natural history' of schizophrenia before the neuroleptic era

Since the great majority of schizophrenic patients are today receiving pharmacological treatment, current and recent studies may not reflect the 'natural' course and outcome of the disorder. Studies in urban communities in Scotland[6] and India,[7] and a study in a rural community in China[8] estimated the proportions of never hospitalized schizophrenic patients at 6.7 per cent, 28.7 per cent, and 30.6 per cent, respectively. About half of the Scottish patients had been prescribed neuroleptics by their general practitioners while the Indian and Chinese patients had been virtually untreated. In all three settings the outcomes of these interesting samples (which presumably approximate the 'natural' history of the disorder) were heterogeneous but, except for a larger proportion of Chinese patients having marked psychotic symptoms, they did not differ much from the outcomes in the treated groups. In a historical study of 70 Swedish patients with first admissions in 1925, lifetime records were retrieved and re-diagnosed in accordance with DSM-III.[9] None of these patients had received neuroleptics. The final outcome was rated as good in 33 per cent (but no patient was considered as completely recovered), as 'profoundly deteriorated' in 43 per cent, and as intermediate in 24 per cent.

A long-term perspective on the course of schizophrenia ove successive generations is provided by a meta-analysis of 320 outcome studies on schizophrenia or dementia praecox published between 1895 and 1992 and including a total of 51 800 subjects.[10]

Overall, about 40 per cent of the patients were reported as improved after an average length of follow-up 5.6 years. There was a significant increase in the rate of improvement during the period 1956–1985 compared to 1895–1955, clearly related to the introduction of neuroleptic treatment, but a secular trend towards better outcomes with every successive decade had been present for much longer. Coupled with the virtual disappearance of the most malignant or 'catastrophic' forms of schizophrenia resulting in a profound defect state after a first psychotic episode, or death ('lethal catatonia'), these observations suggest that a transition to a less deteriorating course of the disorder had occurred prior to modern pharmacological treatment. Among the factors explaining this shift one should consider improvements in general care, progressive changes in attitudes and hospital regime which occurred in a number of institutions on both sides of the Atlantic in the 1930s and 1940s, as well as heightened expectations that psychosocial measures such as psychotherapy or rehabilitation could result in a cure in some cases.

Long-term prognosis

Results of course and outcome studies published over the last six decades are shown in Table 4.3.7.1. The studies have been selected on the basis of the length of follow-up (>5 years), effective sample size (>50), and intensity of follow-up and assessment to provide a global overview of the long-term course of schizophrenia.

Although the studies differ in their design (prospective, follow-back, or retrospective), their results have much in common.

Manfred Bleuler's monograph[11] is the account of an intensive study of 208 patients first admitted in 1942–1943 and personally followed up by the author for 22 years or until death. A recent re-interpretation of Bleuler's diagnoses in terms of DSM-IIIR, DSM-IV, and ICD-10 diagnostic criteria concluded that although some 30 per cent of the original cases would today meet criteria for schizoaffective disorder, the distribution of the types of long-term course did not change significantly.[12] Another 23-year follow-up of 504 patients admitted in 1945–1959 has been completed by Ciompi,[14] and Huber *et al.*[15] interviewed 289 surviving patients in Switzerland first admitted between 1900 and 1962 (median follow-up length 36.9 years).

Notwithstanding methodological constraints which apply to these studies, their findings are a unique record of what probably represents the closest approximation to the 'natural history' of schizophrenia. In summary, they indicate the following.

- Lasting recovery ('complete cure') occurred in 15 per cent to 26 per cent of the cases; 43 per cent had either remitted or exhibited mild residual abnormalities which did not interfere with their living in the community.
- Forty-four per cent were still in hospital and severe chronic states had developed in 14 to 24 per cent.

Table 4.3.7.1 Results of selected course and outcome studies in schizophrenia, 1972–2005

Author	Country	Sample size	Length of follow-up (years)	Proportion good outcome*
Bleuler (1972)[11]	Switzerland	208	23	20% Complete remission; 33% mild defect
Tsuang *et al.* (1979)[13]	USA	186	35	46% Recovered or improved significantly
Ciompi (1980)[14]	Switzerland	289	37	20% Recovered; 43% definitely improved
Huber *et al.* (1980)[15]	Germany	502	22	26% Recovered; 31% remission with mild defect
Harding *et al.* (1987)[16]	USA	118	32	62% Recovered or improved significantly
Ogawa *et al.* (1987)[17]	Japan	140	21–27	31% Recovered; 46% improved
Shepherd *et al.* (1989)[18]	UK	107	5	22% Recovered, no relapse
Johnstone *et al.* (1990)[19]	UK	530	3–13	14% Excellent; 18.5% very good social adjustment
Carone *et al.* (1991)[20]	USA	79	5	17% Complete remission
Marneros *et al.* (1992)[21]	Germany	249	25	Full remission in 24% ('broad') or 7% ('pure') schizophrenia
Thara *et al.* (1994)[22]	India	90 (first-onset cases)	10	12% Complete recovery; 62% remission
Mason *et al.* (1995)[23]	UK	67	13	17% Complete recovery; 52% remission
Wieselgren and Lindström (1996)[24]	Sweden	120	5	30% Good outcome
Wiersma *et al.* (1998)[25]	Holland	82	15	27% Complete; 50% partial remission
Ganev *et al.* (1998)[26]	Bulgaria	60	16	32% Complete; 5% partial remission
Gureje and Bamidele (1999)[27]	Nigeria	120	13	22% Unimpaired (social outcome); 19% some impairment
Finnerty *et al.* (2002)[28]	Ireland	67 (first-onset cases)	15	35% Complete remission; 46% partial remission
Thara (2004)[29]	India	90 (first-onset cases)	20	6% Complete recovery; 15% clinically stable
Lauronen *et al.* (2005)[30]	Finland	91 (birth cohort members)	To age 31 years	4% Complete recovery; 3% partial remission

*Descriptive categories used by the authors.

- In 50 per cent to 75 per cent of the patients, a clinically stable state set in after the fifth year since onset, with no significant further deterioration.
- Remitting course with multiple episodes and full remissions characterized 22 per cent of the patients; catastrophic course (rapid onset of chronic deterioration) was observed in 1 per cent to 4 per cent.
- The 20-year suicide rate was 14 per cent to 22 per cent.

Two American studies largely concur with these findings. In the Vermont study,[16] no less than 62 per cent of the cohort had achieved significant improvement or recovery after an average length of follow-up 32 years; the corresponding proportion in the Iowa 500 study[13] was 46 per cent.

The most striking finding from the long-term follow-up studies is the high proportion of patients who recover, either completely or with mild residual abnormalities, after decades of severe illness[31] This contrasts with the ingrained image of schizophrenia as an intractable, deteriorating illness that many clinicians tend to adopt on the basis of a limited follow-up horizon and patient samples selected for unfavourable course and treatment response. It is unlikely that the high percentage of recoveries in the long-term studies could be explained by cases of affective illness or brief transient psychoses misdiagnosed as schizophrenia (the retrospective re-diagnosis of cases according to DSM-III criteria in the American studies did not alter significantly the results). Similarly unlikely would be the attribution of all the good outcomes to the antipsychotic treatment many of these patients received in the later stages of their illnesses, since comparable proportions of improvement of recovery had been reported for patients who never received neuroleptics.[32] A tentative conclusion from such follow-up research would be that schizophrenia is not an invariably chronic deteriorating disorder, and that the progression of the disease can be arrested or even reversed at any stage. The causes of such reversibility remain poorly understood, but research focusing specifically on the recovering cases will undoubtedly provide essential clues for understanding the nature of schizophrenia.

The results of longitudinal studies published in the last decade generally tend to corroborate the pattern of outcomes outlined by the earlier studies. However, several recent studies suggest a trend of worsening clinical and social outcomes in patients with schizophrenia in both developed countries[28,30] and developing countries.[27,29] A 13-year follow-up of 120 Nigerian patients[27] reported much higher rates of severe impairment in social and occupational functioning than those found in the same region of the country by follow-up studies in the 1970–1980s.

Patterns and stages of the course of schizophrenia

The marked heterogeneity of the course of schizophrenia can be reduced to a limited number of patterns into which cases tend to cluster over time. In earlier long-term follow-up studies, eight different categories of course have been described by Bleuler[11] and by Ciompi,[14] and 12 by Huber *et al.*[15] These classifications were derived from empirical observation, rather than statistical modelling, and conflated into single categories the mode of onset, longitudinal aspects such as frequency and duration of psychotic episodes, remissions, and end states. Treating these various aspects of the longitudinal profile of the illness as independent dimensions has been recommended.[19] However, the complexity of statistical modelling of the course of schizophrenia is such that the development of a classification of course that would be both useful in clinical practice and rigorous in a statistical sense may not be easy to achieve. Therefore, a heuristic compromise between these two requirements should, as a minimum, define operationally and assess separately: (i) the number and duration of discrete episodes of illness; (ii) the predominant clinical features of each episode (e.g. psychotic or affective); (iii) the number and length of remissions and their quality (presence/absence of residual negative or deficit symptoms and signs). By combining these variables, several patterns of course have been derived that have found good empirical support in international follow-up studies:

1 single psychotic episode followed by complete remission;

2 single psychotic episode followed by incomplete remission;

3 two or more psychotic episodes, with complete remissions between episodes;

4 two or more psychotic episodes, with incomplete remissions between episodes;

5 continuous (unremitting) psychotic illness.

With some modifications, these longitudinal patterns have been incorporated into ICD-10 and DSM-IV as additional descriptors.

Although the components of the course patterns, such as episode, remission, residual symptomatology, etc. may not represent 'pure' dimensions, it is desirable to restrict the definition of **pattern of course** to clinical variables only, in order to be able to examine its correlations with risk factors and predictors, such as premorbid impairments, mode of onset, or social outcomes. Assessing social functioning independently of the clinical pattern of course is critical to the study of illness-environment interactions and the causes of disablement in schizophrenia.

At present it does not seem possible to define with any precision **discrete stages** in the progression of schizophrenic illnesses by using combined clinical and pathological criteria, as in cancer or cardiovascular disease. Nevertheless, a 'softer' form of staging is feasible since there is on the whole a good agreement between the results of different studies on the general pattern of course in schizophrenia. On the basis of long-term follow-up studies, the lifetime course of schizophrenia can be articulated into a premorbid phase (from birth to the onset of psychosis), a phase of acute or positive schizophrenic symptomatology, and a residual phase.[21] Various sub-stages have been proposed to describe in finer detail the pre-onset and early psychosis period.[33,34] For most practical and research purposes, a three-stage classification of **post-onset course** has been proposed:[24]

1 an early deteriorating phase (the first 5–10 years);

2 a middle (stabilization) phase;

3 a gradual improvement phase.

This model agrees well with the empirical evidence and could be useful in the collection and summarizing of data on individual risks and prognosis.

Geographical and cultural variation

Three prospective investigations initiated by the World Health Organization (WHO): the **International Pilot Study on**

Table 4.3.7.2 Two-year course and outcome features of 1070 patients with schizophrenia in the WHO 10-country study[37]

Course and outcome descriptor	% Patients in developing countries[1] (*n* = 467)	% Patients in developed countries[2] (*n* = 603)
Remitting, complete remissions	62.7	36.8
Continuous or episodic, no complete remission	35.7	60.9
Psychotic <5% of the follow-up	18.4	18.7
Psychotic >75% of the follow-up	15.1	20.2
No complete remission during follow-up	24.1	57.2
Complete remission for >75% of the follow-up	38.3	22.3
On antipsychotic medication >75% of the follow-up	15.9	60.8
No antipsychotic medication during follow-up	5.9	2.5
Hospitalized for >75% of follow-up	0.3	2.3
Never hospitalized during follow-up	55.5	8.1
Impaired social functioning throughout follow-up	15.7	41.6
Unimpaired social functioning >75% of follow-up	42.9	31.6

[1] Colombia, India, Nigeria.

[2] Czech Republic, Denmark, Ireland, Japan, Russia, United Kingdom, United States.

Table 4.3.7.3 Long-term (15- and 25-year) outcome in patient cohorts assessed in the International Study of Schizophrenia (ISoS)

Outcome variable	Incidence cohorts (*N* = 1171, including 15-year follow-up of the WHO 10-country cohort)	Prevalence cohorts (*N* = 462, including 25-year follow-up of 373 cases from the WHO IPSS)
Recovered at follow-up (Bleuler's criteria)	48.1	53.5
Not psychotic in the past 2 years	42.8	40.8
GAF-S[1] > 60	54.0	56.7
Working most of past 2 years	56.8	73.9
GAF-D[2] > 60	50.7	60.3
SMR[3] (range)	0.00–5.71[4]	1.04–8.88[5]

[1] Global assessment of functioning—symptoms scale.

[2] Global assessment of functioning—disability scale.

[3] Standard mortality ratio.

[4] Rochester (0.00), Moscow (1.41), Chandigarh urban (1.88), Prague (2.53), Chandigarh rural (3.02), Honolulu (3.13), Nottingham (3.31), Dublin (4.10), Nagasaki (5.71).

[5] Sofia (1.04), Cali (1.31), Madras (1.90), Agra (1.86), Beijing (2.97), Prague IPSS (3.84), Mannheim (5.55), Hong Kong (5.76), Groningen (8.88).

Schizophrenia (IPSS);[35,36] the 10-country study on **Determinants of Outcome of Severe Mental Disorders**;[37] and the study on **Assessment and Reduction of Psychiatric Disability**[38,39] laid the ground for a broad-based, cross-cultural evaluation of the course and outcome of schizophrenia. These studies comprise extensive initial and follow-up information on a total of 2736 patients in 16 countries, diagnosed with schizophrenia according to strict and comparable criteria. Identical or closely similar, standardized assessment procedures and instruments were employed, ensuring a high level of comparability across the multiple sites. Results of the WHO 10-country study (pooled data on 1070 patients in all the research sites) are presented in Table 4.3.7.2.

A more recent, transcultural investigation coordinated by WHO, the **International Study of Schizophrenia** (ISoS) involving 18 research centres in 14 countries,[40,41] achieved tracing 75 per cent of cases assessed in the earlier WHO studies referred to above, as well as additional cohorts from mainland China, Hong Kong, and India. Follow-up data were collected on a total of 1633 cases (surviving or dead), and 890 patients were re-interviewed at either 15- or 25-year follow-up since their first assessment. Key findings from this landmark study are presented in Table 4.3.7.3.

The following general conclusions can be drawn from the WHO studies.

1 There is a striking heterogeneity and variability of the course and outcome of schizophrenia, both across and within populations. Patients with similar clinical and diagnostic characteristics at baseline assessment develop a spectrum of outcomes ranging from stable clinical and social recovery after a single psychotic episode to chronic unremitting psychosis and severe impairment. Long-term follow-up studies lend credibility to the conclusion that a high proportion (over 30 per cent) of patients meeting the diagnostic criteria for schizophrenia have relatively favourable outcomes.

2 The frequencies of both relapses and remissions tend to increase over time: while at 2-year follow-up of the International Pilot Study of Schizophrenia (IPSS)[35] 11 per cent of the patients had experienced two or more psychotic episodes followed by complete remission, and another 18 per cent had two or more episodes followed by residual symptoms and impairments, the corresponding proportions at 5-year follow-up were 15 per cent and 33 per cent.[36] These trends are now bolstered by the findings of the International Study of Schizophrenia (ISoS)[40] which found a 48 per cent recovery rate at the 15-year follow-up and 54 per cent at the 25-year follow-up.

3 Regardless of the increasing relapse rate, the cumulative proportion of follow-up time during which patients have psychotic symptoms (as a percentage of the total follow-up time), tends to remain stable or decrease. At the end of the 5-year follow-up, 57 per cent of the patients had experienced a total of less than 9 months of active psychosis; only 22 per cent had been psychotic for 45–60 months. At 15-year and 25-year follow-up, 43 per cent and 41 per cent, respectively, have been free of active psychotic symptoms for the past 2 years.

4 The levels of social impairment assessed at 2 years changed very little during the subsequent years of follow-up. Overall, most of the observed change in the clinical state and social functioning of patients between the 2-year follow-up and the 5-year follow-up

was towards improvement rather than deterioration. This also is congruent with the findings at 15-year and 25-year follow-up.

5 While the percentage of patients with continuous, deteriorating illness was similar across the sites of the WHO studies, there were significant differences in the proportions of patients who achieved symptomatic and social recovery. In this respect, outcome was generally better in the developing countries. This unexpected finding of the first follow-up of the International Pilot Study of Schizophrenia patients,[35] who had been recruited from consecutive hospital admissions with the attendant possibilities of a selective bias, was subsequently replicated by the 10-country study which had an epidemiological design and recruited only first-contact patients from delimited populations.[37] The better course and outcome in the developing country areas could not be attributed to any particular subtype of the disorder, e.g. cases of acute onset, since it applied equally to the cases of slow, insidious onset. The main outcome difference across the study areas was in the occurrence and average length of symptom-free remissions. Remissions tended to be more frequent and to last longer in the developing countries. No single factor accounting for this difference could be identified and it is likely that complex interactions between illness and environment are involved that may include both population differences in **predisposing genes**,[42] and environmental or cultural factors, such as relative absence of an institutionalized role of 'the schizophrenic',[43] less intrafamilial **expressed emotion** towards the affected family member,[44] or better integration of the mentally ill person in the domestic economy in traditional rural communities. It should be noted, however, that the long-term WHO follow-up studies include patients whose onset of psychotic illness occurred decades ago, and that increasing social and economic stresses experienced by both rural and urban communities in many developing countries may have eroded the traditional support systems, resulting in worse outcomes, as suggested by several recent studies.

Whether the better outcome of schizophrenia in the developing countries is 'transportable' following migration to other settings, remains unclear. Data on immigrants treated for first episodes of schizophrenia in the United Kingdom suggest that while Asian patients have a lower relapse and readmission rate than British-born Whites, Afro-Caribbean's show a higher rate.[45] The marked social and family structure differences between the Asian and the Afro-Caribbean immigrant communities suggest that the likelihood of a more benign course in the new setting may depend on the degree to which the immigrant group has retained its traditional values and intra-group cohesion.

First episode psychosis

The recent focus on early detection and treatment of first episodes of psychosis, driven by theoretical considerations and clinical concerns, is supported by evidence suggesting that the course and outcome of the earliest stages of a schizophrenic illness may have a **pathoplastic effect** on its subsequent course. Specifically, the period between the first onset of psychotic symptoms and the initiation of treatment (**duration of untreated psychosis, DUP**) has been shown to correlate with increased time to remission and poor response to treatment.[46,47] Plausible clinical considerations have been proposed in support of the view that the first episode of psychosis represents a critical developmental transition that may impact the subsequent course of schizophrenia, possibly by inducing neurotoxic alterations in neural networks, thus preparing the ground for chronic illness.[48] An extension of this mode of thinking is the suggestion that a behavioural or pharmacological intervention prior to the onset of psychotic symptoms could delay or prevent the onset of schizophrenia.[49]

None of these hypotheses has been conclusively tested. However, a number of studies focusing on the **prodrome** and the earliest manifestations of psychosis have highlighted features such as a presymptomatic drop in cognitive performance and social functioning;[50] early co-occurrence of 'positive' and 'negative' symptoms;[37] as well as a general malleability of such dysfunction in response to appropriate behavioural interventions and low-dose, time-limited pharmacological treatment.[51–53] This suggests that clinical research bridging the gap between statistical investigations of risk factors or antecedents of disease and individual pathways to psychotic illness may have an important role to play in understanding and, ultimately, influencing the development and course of schizophrenia.

Prognosis of specific clinical symptoms and syndromes

Longitudinal studies suggest that the characteristic symptoms of schizophrenia tend to 'breed true', i.e. only a minority of patients are eventually reclassified into other disease categories because of a significant and lasting change in the predominant symptoms. However, the proportion of cases warranting a re-diagnosis seems to increase with the length of follow-up.

Depression in schizophrenia

In the WHO International Pilot Study of Schizophrenia,[35, 36] the proportion of patients with initial schizophrenic symptomatology who developed non-schizophrenic (mostly affective) episodes in the course of time increased from 3 per cent in the first 2 years to 17 per cent at the end of the 5-year follow-up. In contrast, subsequent episodes with schizophrenic features occurred in less than 10 per cent of the patients with an initial diagnosis of major depression. Depression is the most common non-schizophrenic syndrome co-ocurring with schizophrenia also in those patients who retain the essentially schizophrenic character of their illnesses. The proportion of patients who develop clear-cut episodes of major depression ranges from 15 per cent during a 5-year follow-up[54] to 24 per cent during a 12-year follow-up.[55] This is a much higher period prevalence than depression in the general population, which suggests that mood disorder may be an intrinsic part of the clinical spectrum of schizophrenia. Based on such data, a diagnostic rubric of **post-schizophrenic depression** has been added to the classification of schizophrenia in ICD-10.

First-rank (schneiderian) symptoms

Subjective thought disorder phenomena, such as thought broadcast or insertion, passivity ('replacement of will') experiences, and particular type of auditory hallucinations (third-person or commenting 'voices') were attributed 'first-rank' significance in the differential diagnosis between schizophrenic and affective psychoses by Kurt Schneider.[56] These symptoms are accorded special diagnostic weight in the current diagnostic criteria of both ICD-10

and DSM-IV. Although Schneider explicitly disclaimed any particular prognostic value for the first-rank symptoms, they have a strong tendency to recur in the course of schizophrenia. In the WHO 10-country study, patients with one or more first-rank symptoms on the initial examination had a three-fold increased risk of recurrence of such symptoms in subsequent episodes compared to patients with no first-rank symptoms at initial examination.[37]

Prognosis of schizophrenia subtypes

The evidence that each of the 'classic' subtypes of schizophrenia is associated with a characteristic pattern of course is generally weak but surprisingly good for some of the subtypes. Consistent differences have been reported between **paranoid**, **hebephrenic**, and **undifferentiated** schizophrenia (diagnosed according to DSM-III) on a long-term follow-up of 19 years.[57] Paranoid schizophrenia tended to have a remittent course, and to be associated with less disability, in contrast to hebephrenia which had an insidious onset and poor long-term prognosis. Undifferentiated schizophrenia occupied an intermediate position. In the WHO International Pilot Study of Schizophrenia,[35] four alternative groupings of the ICD-9 subtypes were tested by a discriminant function for differences with regard to six course and outcome measures. Clear discrimination was achieved between simple and hebephrenic schizophrenia grouped together, on the one hand, and the schizoaffective subtype on the other. However, the comparison of simple and hebephrenic schizophrenia with paranoid schizophrenia resulted in a considerable degree of overlap.

Better discrimination has been claimed for groups of patients diagnosed according to the criteria of Leonhard.[58] A 5–13 years follow-up study of 178 patients admitted with a diagnosis of schizophrenia and re-diagnosed according to the Leonhard's criteria as **systematic schizophrenia**, atypical (unsystematic) schizophrenia, **cycloid psychosis**, or **reactive psychosis**[59] resulted in marked outcome differences on blind assessment. While only 10 per cent of the cases in the two schizophrenia groups were judged to have 'recovered', the corresponding proportion in the cycloid and reactive psychoses group was 38 per cent. Conversely, the proportions of 'unimproved' cases were 49 per cent and 3 per cent.

The question whether good prognosis, remitting schizophrenia with an acute onset is a separate subtype that could be distinguished symptomatologically was addressed in the WHO 10-country study[37] by comparing 274 patients with an initial ICD-9 diagnosis of acute schizophrenic episode and 752 patients with other schizophrenia subtypes. The group of acute cases tended to be younger and had a lower male/female ratio, but was no different from the rest of the schizophrenic patients with regard to initial symptomatology. This argues against acute **schizophreniform** illness being a discrete syndrome, outside the clinical spectrum of schizophrenia.

The course and outcome data on **schizoaffective** disorders seem to support their placement within the broad category of schizophrenia. A retrospective and prospective study of 150 schizoaffective patients and 95 bipolar affective patients[60] established general similarities between the two groups but the schizoaffective cases were less likely to achieve a full remission and more likely to develop a residual state (in 57 per cent compared to 24 per cent in the bipolar group). An intermediate outcome between that of schizophrenia and bipolar affective disorder is a common finding in schizoaffective disorders.

Predictors of course and outcome

A wide range of variables have been explored as possible predictors of course and outcome in schizophrenia: (i) socio-demographic characteristics; (ii) features of the premorbid personality and premorbid functioning; (iii) family history of psychiatric disorder; (iv) history of past psychotic episodes and treatments; (v) substance use; (vi) characteristics of the onset; (vii) features of the initial clinical state and treatment response; and (viii) variables related to brain morphology and neurocognitive functioning. Many predictors have been independently replicated by different investigators and there is reasonable agreement on the general direction of their effects. However, definitions of both the independent (predictor) and the dependent (outcome) variable tend to vary across studies, and the statistical methods employed range from basic descriptive statistics (e.g. *x* per cent of the patients with characteristic *y* developed outcome *z*) to complex statistical models with capacity to quantify the independent contribution of individual variables to a specified outcome.

Table 4.3.7.4 lists the best predictors of the 2-year outcome in the WHO 10-country study[37] and Table 4.3.7.5 summarizes the findings about predictors of 15-year outcome for the subset of participants in the WHO 10-country study who were re-examined as part of the International Study of Schizophrenia.[40] Apart from the variable 'setting' (i.e. research centre), which is a proxy for an unspecified number of local area features ranging from population genetic background to the multiple facets of 'culture', the mode of onset of symptoms (acute versus insidious), drug abuse, and premorbid psychosocial functioning were the best predictors of the duration of psychotic episodes, achievement of remission, and social outcome at 2-year follow-up. Importantly, the total time with psychotic symptoms during the first 2 years post-onset emerged as the best predictor of 15-year outcome, highlighting the potential importance of interventions aiming to contain and minimize active psychosis during this critical stage.

Limitations of clinical prediction

The explanatory power of any predictor in schizophrenia (in terms of accounting for a proportion of the outcome variance) varies depending on sample size, setting, homogeneity of patient groups, and measurement error, but generally tends to be limited (rarely exceeding 30 per cent of the outcome variance). This suggests that no single background or premorbid characteristic of the person, and no clinical symptom or sign among the initial manifestations of the disorder, is strongly associated with its prognosis in the longer-term. Similarly to the genetic epidemiology of schizophrenia, where non-shared environmental influences account for a greater amount of variance than the shared environment, person-specific, emergent life events or changes in the mental state may have a similar or even greater impact on the outcome as the initial or premorbid predictors. Indeed, variables such as negative symptoms have been shown to gain in predictive power if they are assessed two or more years after the onset, or after the patients have received adequate treatment. The predictive capacity of other variables, for example, mode of onset, or a high index of expressed emotion, tends to become attenuated in the course of time. Thus, there is

Table 4.3.7.4 Best predictors of 2-year course and outcome in the WHO 10-country study (log-linear analysis of 1078 cases)

	Course and outcome variables			
Predictor	**Pattern of course**	**Time in psychosis**	**Remission/no remission**	**Social functioning**
Age	n.s.	n.s.	n.s.	*
Sex	*	*	**	**
Marital status	**	***	***	***
Acute versus gradual onset	***	***	***	***
Time since onset (duration of untreated psychosis)	n.s.	n.s.	**	***
First-rank symptoms at baseline	n.s.	n.s.	n.s.	n.s.
Adjustment in childhood	n.s.	n.s.	**	***
Adjustment in adolescence		***	***	***
Close friends	*	*	***	***
Street drug use	*	*	***	***
Setting (developing/developed country)	***	n.s.	***	***

n.s., Not significant.

*Significant at $p \leq 0.05$.

**Significant at $p \leq 0.01$.

***Significant at $p \leq 0.001$.

Table 4.3.7.5 Best predictors of 15-year outcome in the WHO 10-country study (stepwise linear regression analysis of 766 cases included in ISoS)

	Course and outcome variables			
Predictor	**GAF-S[1] (centre in the analysis)**	**GAF-S (area variables in the analysis)**	**GAF-D[2] (centre in the analysis)**	**GAF-D (area variables in the analysis)**
Percentage of time psychotic (first 2 years)	*	*	*	*
Setting (centre)	*	–	*	–
Blunted affect at initial examination	–	*	–	*
National health insurance available	–	–	–	*
Street drug use	–	–	–	*
Family involvement in care	–	–	–	*

[1] Global assessment of functioning—symptoms scale.

[2] Global assessment of functioning—disability scale.

* Significant at $p \leq 0.05$.

– Not significant or not applicable.

no fixed set of predictors of the course and outcome of schizophrenia, but rather a number of prognostic indicators which allow a judgement to be made about the probability of one or another type of course over a limited time period (usually not exceeding 5 years).

Short-term predictors

There is good agreement between different studies on the factors that help predict a relapse of psychotic symptoms after a period of stabilization or remission. By and large, the best predictor of relapse in the short-term remains the withdrawal of antipsychotic medication, usually due to non-compliance.[61] Heavy cannabis use has been shown to be associated with an increased risk of relapse in a dose-response relationship.[62] Other factors, such as stressful **life events**[63] and **expressed emotion** (EE) within the family,[44] have attracted considerable interest, both as independent predictors and as modifiers of the effects of pharmacological treatment. A high expressed emotion index, assessing a key family member's emotional over-involvement with, and concomitant criticism of the patient, has been found to be a reliable short-term predictor of psychotic relapse. However a limitation of the method is that it is only applicable to situations of intensive daily interaction between a patient and a carer (typically a family member) which may not be the case for many people with schizophrenia living in hostels or marginal accommodation. Moreover, the cross-cultural validity of the expressed emotion index still remains to be established.[64]

Medium- and long-range predictors

In first-episode cases, male sex, single marital status, premorbid social withdrawal and insidious onset have been shown to be relatively robust predictors of a poor outcome in the short- to medium-term (2–5 years), while female sex, being married, having social contacts outside the home, and acute onset predict a relatively good outcome. No consistent findings have been reported for age at onset as a predictor, and the long-term follow-up studies do not lend support to the view that an early onset is always associated with a poor prognosis. Similarly, a history of psychotic illness (including schizophrenia) in a first-degree relative does not necessarily predict a worse prognosis. On the contrary, in some studies[65] patients with a high familial load were found on follow-up to have a better outcome than 'sporadic' cases with no psychotic illness among their first-degree relatives.

A consistent finding of many studies is that the clinical symptoms in either the early, or the advanced stages of schizophrenia, have limited capacity to predict future course and outcome. An exception is the modest predictive power of clear-cut negative symptoms appearing early in the course of the disorder, or when assessed under the conditions referred to above.

The **socio-cultural setting**, i.e. a developing country or a developed country, was found to be among the best predictors of 2-year and 5-year outcome in the WHO studies.[35,36] It remained a

significant predictor of 15-year outcome in the International Study of Schizophrenia.[40] Exactly what factors may be underlying these marked cultural differences in the prognosis of schizophrenia remains an unresolved issue.

There is a growing interest in the predictive power of **neurocognitive functions**, such as verbal memory, working memory, processing speed, and sustained attention. Though positive results have been reported from a number of studies, the proportion of variance in outcome measures that could be explained by such factors varies from low 14 per cent[66] to as high as 60 per cent,[67] depending on sample selection, patients' age, and length of follow-up. Overall, there is increasing evidence that neurocognitive functioning at the early stages of a schizophrenic illness predicts significantly global psychosocial and occupational functioning in the medium-term. Neurocognition is therefore likely to be an increasingly important target for novel pharmacological and cognitive behavioural interventions in schizophrenia.

Recovery from schizophrenia

There is consistent evidence from longitudinal studies reviewed in this chapter that, notwithstanding the high risk of chronic disability, loss of developmental potential, and diminished quality of life associated with schizophrenia, there is a non-negligible proportion of people who meet the current diagnostic criteria for the disease but ultimately attain nearly complete recovery and a stable level of psychosocial functioning. The existence of a good outcome subgroup within schizophrenia has been known for a long time and the prevailing view is that it is not an artefact of misdiagnosis. Yet little focused research has been conducted to bring to light the characteristics and predictors of this clinical subpopulation. A retrospective study of 436 people diagnosed with schizophrenia in the United Kingdom[68] found that over a follow-up period of 6 years, 15.6 per cent had a single psychotic episode with complete remission. A 15-year prospective study of 145 patients in the United States[69] revealed that up to 25 per cent of DSM-III diagnosed schizophrenia patients had ceased on their own accord antipsychotic medication since the first 5 years of follow-up and the majority of them had remained symptom-free for the rest of follow-up. Common findings in these two studies were that the non-relapsing, high-functioning patients were characterized by a higher level of premorbid occupational achievement and social competence, were less likely to use street drugs,[68] had better insight and an internal 'locus of control'.[69] Further study of the implications of such indicators of better prognostic potential and internal resources should advance efforts to design management and treatment strategies reducing the disabling impact of the disorder.

Summary and conclusions

Studies conducted over many decades consistently demonstrate that schizophrenia presents a broad spectrum of possible outcomes and course patterns, ranging from complete or nearly complete recovery after acute episodes of psychosis to continuous, unremitting illness leading to progressive deterioration of cognitive performance and social functioning. Between these extremes, a substantial proportion of patients show an episodic course with relapses of psychotic symptoms and partial remissions during which affective and cognitive impairments become increasingly conspicuous and may progress to gross deficits. Although no less than one-third of all patients with schizophrenia have relatively benign outcomes, in the majority the illness still has a profound, lifelong impact on personal growth and development. The initial symptoms of the disorder are not strongly predictive of the pattern of course but the mode of onset (acute or insidious), the duration of illness prior to diagnosis and treatment, the presence or absence of comorbid substance use, as well as background variables such as premorbid adjustment (especially during adolescence), educational and occupational achievement, marital status, and availability of a supportive social network allow a reasonable accuracy of prediction in the short- to medium-term (2–5 years).

One of the most striking aspects of the longitudinal course of schizophrenia is the so-called 'terminal improvement'. A relatively high proportion of patients tend to improve substantially with ageing. What determines this long-term outcome is far from clear but the stereotype view of schizophrenia as an invariably progressive, deteriorating disorder does not accord well with the evidence. Similarly, a model of schizophrenia as a static **neurodevelopmental encephalopathy** decompensating post-adolescence under the influence of a variety of stressors fits only part of the spectrum of course patterns. In a significant proportion of cases, the disorder exhibits the unmistakable features of a shift-like process with acute exacerbations and remissions which may progress to severe deterioration or come to a standstill at any stage. Whether a single underlying pathophysiology can explain the variety of clinical outcomes, or several different pathological processes are at work, remains obscure. It has been suggested that the longitudinal course of schizophrenia should be seen as an open-ended, dynamic life process with multiple, interacting biological and psychosocial determinants. Obviously, such issues cannot be resolved by clinical follow-up studies alone, and require a strong involvement of neurobiological research in prospective investigations of representative samples of cases spanning the entire spectrum of course and outcomes. No such studies have been possible until recently, both because of the technical complexity of such an undertaking and because of the tendency to selectively recruit for biological investigations patients from the severe, deteriorating part of the spectrum. Overcoming such limitations will be essential to the uncovering of the mechanisms driving the 'natural history' of schizophrenia.

Further information

Hopper, K., Harrison, G., Janca, A., *et al.* (eds.) (2007). *Recovery from schizophrenia. An international perspective. A report from the WHO collaborative project the international study of schizophrenia*. Oxford University Press, New York.

References

1. Kraepelin, E. (1919). *Dementia praecox and paraphrenia*. Livingstone, Edinburgh.
2. van Os, J., Wright, P., and Murray, R.M. (1997). Follow-up studies of schizophrenia I: natural history and non-psychopathological predictors of outcome. *European Psychiatry*, **12**(Suppl. 5),327s–41s.
3. Olesen, A.V. and Mortensen, P.B. (2002). Readmission risk in schizophrenia: selection explains previous findings of a progressive course of disorder. *Psychological Medicine*, **32**, 1301–7.

4. Bebbington, P.E., Craig, T., Garety, P., *et al.* (2006). Remission and relapse in psychosis: operational definitions based on case-note data. *Psychological Medicine*, **36**, 1551–62.
5. Nuechterlein, K.H., Miklowitz, D.J., Ventura, J., *et al.* (2006). Classifying episodes in schizophrenia and bipolar disorder: criteria for relapse and remission applied to recent-onset samples. *Psychiatry Research*, **15**, 153–66.
6. Geddes, J.R. and Kendell, R.E. (1995). Schizophrenic subjects with no history of admission to hospital. *Psychological Medicine*, **25**, 859–68.
7. Padmavathi, R., Rajkumar, S., and Srinivasan, T.N. (1998). Schizophrenic patients who were never treated—a study in an Indian urban community. *Psychological Medicine*, **28**, 1113–17.
8. Ran, M., Xiang, M., Huang, M., *et al.* (2001). Natural course of schizophrenia: 2-year follow-up study in a rural Chinese community. *The British Journal of Psychiatry*, **178**, 154–8.
9. Jonsson, S.A.T. and Jonsson, H. (1992). Outcome in untreated schizophrenia: a search for symptoms and traits with prognostic meaning in patients admitted to a mental hospital in the preneuroleptic era. *Acta Psychiatrica Scandinavica*, **85**, 313–20.
10. Hegarty, J.D., Baldessarini, R.J., Tohen, M., *et al.* (1994). One hundred years of schizophrenia: a meta-analysis of the outcome literature. *The American Journal of Psychiatry*, **151**, 1409–16.
11. Bleuler, M. (1972). Die schizophrenen Geistesstörungen im Lichte langjähriger Kranken- und Familiengeschichten. G. Thieme, Stuttgart. English translation by S.M. Clemens (1978) The schizophrenic disorders. Long-term patient and family studies. Yale University Press, New Haven.
12. Modestin, J., Huber, A., Satirli, E., *et al.* (2003). Long-term course of schizophrenic illness: Bleuler's study reconsidered. *The American Journal of Psychiatry*, **160**, 2202–8.
13. Tsuang, M., Woolson, R., and Fleming, J. (1979). Long-term outcome of major psychoses. I. Schizophrenia and affective disorders compared with psychiatrically symptom-free surgical conditions. *Archives of General Psychiatry*, **36**, 1295–301.
14. Ciompi, L. (1980). Catamnestic long-term study on the course of life and aging of schizophrenics. *Schizophrenia Bulletin*, **6**, 606–18.
15. Huber, G., Gross, G., Schüttler, R., *et al.* (1980). Longitudinal studies of schizophrenic patients. *Schizophrenia Bulletin*, **6**, 592–605.
16. Harding, C., Brooks, G.W., Ashikara, T., *et al.* (1987). The Vermont longitudinal study of persons with severe mental illness. I. Methodology, study sample, and overall status 32 years later. *The American Journal of Psychiatry*, **144**, 718–26.
17. Ogawa, K., Miya, M., Watarai, A., *et al.* (1987). A long-term follow-up study of schizophrenia in Japan—with special reference to the course of social adjustment. *The British Journal of Psychiatry*, **151**, 758–65.
18. Shepherd, M., Watt, D., Falloon, I., *et al.* (1989). The natural history of schizophrenia: a five-year follow-up study of outcome and prediction in a representative sample of schizophrenics. *Psychological Medicine*, Monograph Supplement 15, 1–46, Cambridge University Press.
19. Johnstone, E.C., Macmillan, J.F., Frith, C.D., *et al.* (1990). Further investigation of the predictors of outcome following first schizophrenic episodes. *The British Journal of Psychiatry*, **157**, 182–9.
20. Carone, B.J., Harrow, M., and Westermeyer, J.F. (1991). Posthospital course and outcome in schizophrenia. *Archives of General Psychiatry*, **48**, 247–53.
21. Marneros, A., Deister, A., and Rohde, A. (1992). Comparison of long-term outcome of schizophrenic, affective and schizoaffective disorders. *The British Journal of Psychiatry*, **161**(Suppl. 18), 44–51.
22. Thara, R., Henrietta, M., Joseph, A., *et al.* (1994). Ten-year course of schizophrenia—the Madras longitudinal study. *Acta Psychiatrica Scandinavica*, **90**, 329–36.
23. Mason, P., Harrison, G., Glazebrook, C., *et al.* (1995). Characteristics of outcome in schizophrenia at 13 years. *The British Journal of Psychiatry*, **167**, 596–603.
24. Wieselgren, I.M. and Lindström, L.H. (1996). A prospective 1-5 year outcome study in first-admitted and readmitted schizophrenic patients: relationship to heredity, premorbid adjustment, duration of disease and education level at index admission and neuroleptic treatment. *Acta Psychiatrica Scandinavica*, **93**, 9–19.
25. Wiersma, D., Nienhuis, F.J., Sloof, C.J., *et al.* (1998). Natural course of schizophrenic disorders: a 15-year follow-up of a Dutch incidence cohort. *Schizophrenia Bulletin*, **24**, 75–85.
26. Ganev, K., Onchev, G., and Ivanov, P. (1998). A 16-year follow-up study of schizophrenia and related disorders in Sofia, Bulgaria. *Acta Psychiatrica Scandinavica*, **98**, 200–7.
27. Gureje, O. and Bamidele, R. (1999). Thirteen-year social outcome among Nigerian outpatients with schizophrenia. *Social Psychiatry and Psychiatric Epidemiology*, **34**, 147–51.
28. Finnerty, A., Keogh, F., O'Grady-Walsh, A., *et al.* (2002). A 15-year follow-up of schizophrenia in Ireland. *Irish Journal of Psychological Medicine*, **19**, 108–14.
29. Thara, R. (2004). Twenty-year course of schizophrenia: the Madras longitudinal study. *Canadian Journal of Psychiatry*, **49**, 564–69.
30. Lauronen, E., Koskinen, J., Veijola, J., *et al.* (2005). Recovery from schizophrenic psychoses within the northern Finland 1966 birth cohort. *The Journal of Clinical Psychiatry*, **66**, 375–83.
31. Auslander, L.A. and Jeste, D.V. (2004). Sustained remission of schizophrenia among community-dwelling older outpatients. *The American Journal of Psychiatry*, **161**, 1490–3.
32. Harrow, M. and Jobe, T.H. (2007). Factors involved in outcome and recovery in schizophrenia patients not on antipsychotic medications: a 15-year multifollow-up study. *The Journal of Nervous and Mental Disease*, **195**, 406–14.
33. Häfner, H., Maurer, K., Löffler, W., *et al.* (2003). Modeling the early course of schizophrenia. *Schizophrenia Bulletin*, **29**, 325–40.
34. McGorry, P.D., Hickie, I.B., Yung, A.R., *et al.* (2006). Clinical staging of psychiatric disorders: a heuristic framework for choosing earlier, safer and more effective interventions. *The Australian and New Zealand Journal of Psychiatry*, **40**, 616–22.
35. World Health Organization. (1979). *Schizophrenia. An international follow-up study*. Wiley, Chichester.
36. Leff, J., Sartorius, N., Jablensky, A., *et al.* (1992). The international pilot study of schizophrenia: five-year follow-up findings. *Psychological Medicine*, **22**, 131–45.
37. Jablensky, A., Sartorius, N., Ernberg, G., *et al.* (1992). Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization ten-country study. *Psychological Medicine* Monograph Supplement 20, Cambridge University Press, Cambridge.
38. Jablensky, A., Schwarz, R., and Tomov,T. (1980). WHO collaborative study of impairments and disabilities associated with schizophrenic disorders. *Acta Psychiatrica Scandinavica*, (Suppl. 285), **62**, 152–63.
39. Wiersma, D., Nienhuis, F.J., Giel, R., *et al.* (1996). Assessment of the need for care 15 years alter onset of a Dutch cohort of schizophrenic patients and an international comparison. *Social Psychiatry and Psychiatric Epidemiology*, **31**, 114–21.
40. Harrison, G., Hopper, K., Craig, T., *et al.* (2001). Recovery from psychotic illness: a 15- and 25-year international follow-up study. *The British Journal of Psychiatry*, **178**, 506–17.
41. Hopper, K., Harrison, G., Janca, A., *et al.* (eds.) (2007). *Recovery from schizophrenia. An international perspective. A report from the WHO collaborative project the international study of schizophrenia*. Oxford University Press, New York.
42. Freedman, M.L., Reich, D., Penney, K.L., *et al.* (2004). Assessing the impact of population stratification on genetic association studies. *Nature Genetics*, **36**, 388–93.
43. Waxler, N.E. (1979). Is the outcome for schizophrenia better in non-industrial societies? The case of Sri Lanka. *The Journal of Nervous and Mental Disease*, **167**, 144–58.
44. Leff, J.P., Wig, N.N., Bedi, H., *et al.* (1990) Relatives' expressed emotion and the course of schizophrenia in Chandigarh: a two-year follow-up of a first-contact sample. *The British Journal of Psychiatry*, **156**, 351–6.

45. Bhugra, D., Leff, J., Mallett, R., *et al.* (1997). Incidence and outcome of schizophrenia in Whites, African-Caribbeans and Asians in London. *Psychological Medicine*, **27**, 791–8.
46. Perkins, D., Gu, H., Boteva, K., *et al.* (2005). Relationship between duration of untreated psychosis and outcome in first-episode schizophrenia: a critical review and meta-analysis. *The American Journal of Psychiatry*, **162**, 1785–804.
47. Marshall, M., Lewis, S., Lockwood, A., *et al.* (2005). Association between duration of untreated psychosis and outcome in cohorts of first-episode patients: a systematic review. *Archives of General Psychiatry*, **62**, 975–83.
48. Lieberman, J.A. (1999). Is schizophrenia a neurodegenerative disorder? A clinical and neurobiological perspective. *Biological Psychiatry*, **46**, 729–39.
49. McGlashan, T.H. and Johannessen, J.O. (1996). Early detection and interbention with schizophrenia: rationale. *Schizophrenia Bulletin*, **22**, 201–22.
50. Johnstone, E.C., Ebmeier, K.P., Miller, P., *et al.* (2005). Predicting schizophrenia: findings from the Edinburgh High-Risk Study. *The British Journal of Psychiatry*, **186**, 18–25.
51. Szymanski, S.R., Cannon, T.D., Gallacher, F., *et al.* (1996). Course and treatment response in first-episode and chronic schizophreni a. *The American Journal of Psychiatry*, **153**, 519–25.
52. McGorry, P.D., Edwards, J., Mihalopoulos, C., *et al.* (1996). EPPIC: an evolving system of early detection and optimal management. *Schizophrenia Bulletin*, **22**, 305–26.
53. Menezes, N.M., Arenovich, T., and Zipursky, R.B. (2006). A systematic review of longitudinal outcome studies of first-episode psychosis. *Psychological Medicine*, **36**, 1349–62.
54. Sheldrick, C., Jablensky, A., Sartorius, N., *et al.* (1977). Schizophrenia succeeded by affective illness: catamnestic study and statistical enquiry. *Psychological Medicine*, **7**, 619–24.
55. Breier, A., Schreiber, J.L., Dyer, J., *et al.* (1991). National Institute of Mental Health longitudinal study of chronic schizophrenia. *Archives of General Psychiatry*, **48**, 239–46.
56. Schneider, K. (1959). *Clinical psychopathology* (trans. M.W. Hamilton). Grune & Stratton, New York.
57. Fenton, W.S. and McGlashan, T.H. (1991). Natural history of schizophrenia subtypes. I. Longitudinal study of paranoid, hebephrenic, and undifferentiated schizophrenia. *Archives of General Psychiatry*, **48**, 969–77.
58. Leonhard, K. (1999). *Classification of endogenous psychoses and their differentiated etiology* (2nd revised and enlarged edn). Springer, Wien, New York.
59. Stephens, J.H. and Astrup, C. (1963). Prognosis in 'process' and 'non-process' schizophrenia. *The American Journal of Psychiatry*, **119**, 945–53.
60. Angst, J., Felder, W., and Lohmeyer, B. (1980). Course of schizoaffective psychoses: results of a follow-up study. *Schizophrenia Bulletin*, **6**, 579–85.
61. Dencker, S.J., Malm, U., and Lepp, M. (1986). Schizophrenic relapse after drug withdrawal is predictable. *Acta Psychiatrica Scandinavica*, **73**, 181–5.
62. Linszen, D.H., Dingemans, P.M., and Lenior, M.E. (1994). Cannabis abuse and the course of recent-onset schizophrenic disorders. *Archives of General Psychiatry*, **51**, 273–9.
63. Brown, G.W. and Birley, J.L.T. (1968). Crises and life changes and the onset of schizophrenia. *Journal of Health and Social Behaviour*, **9**, 203–14.
64. López, S.R., Hipke, K.N., Polo, A.J., *et al.* (2004). Ethnicity, expressed emotion, attributions, and course of schizophrenia: family warmth matters. *Journal of Abnormal Psychology*, **113**, 428–39.
65. Johnstone, E.C., Macmillan, J.F., Frith, C.D., *et al.* (1990). Further investigation of the predictors of outcome following first schizophrenic episodes. *The British Journal of Psychiatry*, **157**, 182–9.
66. Milev, P., Ho, B.C., Arndt, S., *et al.* (2005). Predictive values of neurocognition and negative symptoms on functional outcome in schizophrenia: a longitudinal first-episode study with 7-year follow-up. *The American Journal of Psychiatry*, **162**, 495–506.
67. Evans, J.D., Heaton, R.K., Paulsen, J.S., *et al.* (2003). The relationship of neuropsychological abilities to specific domains of functional capacity in older schizophrenia patients. *Biological Psychiatry*, **53**, 422–30.
68. Rosen, K. and Garety, P. (2005). Predicting recovery from schizophrenia: a retrospective comparison of characteristics at onset of people with single and multiple episodes. *Schizophrenia Bulletin*, **31**, 1–16.
69. Harrow, M. and Jobe, T.H. (2007). Factors involved in outcome and recovery in schizophrenia patients not on antipsychotic medications: a 15-year multifollow-up study. *The Journal of Nervous and Mental Disease*, **195**, 406–14.

4.3.8 Treatment and management of schizophrenia

D. G. Cunningham Owens and E. C. Johnstone

Introduction

Historically, there was no shortage of interventions to 'treat' insanity, and later, schizophrenia. Most were palliative, barely effective and often barbaric. It was only with the development of antipsychotic drugs and evolution of trial methodology in the 1950s that a new era of care arrived.

With chronic or recurrent psychiatric disorders, 'treatment' and 'management' are not strictly synonymous. The former has a narrow, patient-specific and largely *symptom* focus, comprising traditional medical tools, especially medications, while the latter can be defined as encompassing a broader range of targets with techniques less specifically part of traditional 'medical' repertoires, including psychological, social, and behavioural interventions. In both its 'treatment' and 'management' aspects, schizophrenia has undergone a concerted therapeutic assault in the past few years but rather than introducing clarity into care, the recent literature might indicate that certainties which seemed so recently within our grasp, remain elusive.

Evidence-based medicine (EBM) has had a major impact on care recommendations for those suffering from schizophrenia. A number of national and international 'guidelines' have been published[1–3] and the Cochrane database provides reviews on a range of relevant issues. This move has been important in smoothing out variability in patient care but although influential and increasingly endorsed, the trend is not beyond criticism. Guidelines are largely derived from *efficacy* data, which may translate awkwardly to real-life *effectiveness*, while considering solely the 'bottom lines' of trials masks the qualitative problems inherent to design, execution, and analysis. Furthermore, over-adherence to guidelines diminishes the impact that *individual variability* within the patient pool makes to optimizing treatment choices while deskilling those charged with care of a complex condition where 'quality' requires greater skill than is usually credited to those who do it well.

Thus, while evidence-based guidelines can offer a *framework* for good care practices, they are *not* a substitute for a broad range of expertise in *individualized* care planning.

Evidence for efficacy

Drug treatment

Schizophrenia comprises a number of domains of disability that form useful targets for drug treatment. These include:

- 'positive' features
- 'negative' features
- cognitive ability
- affective symptomatology
- general behaviour

In addition, comprehensive treatment planning requires awareness of efficacy in:

- maintenance and
- treatment resistance

(a) 'Positive' features

In providing a scientific basis for the use of antipsychotic drugs in 'acute' schizophrenia, the NIMH/VA Collaborative Study comparing the efficacy of three phenothiazines of differing chemical type (chlorpromazine, thioridazine, trifluoperazine) cannot be bettered.[4] Although its conclusions referred specifically to *phenothiazines*, subsequent research justifies the generalization of its findings to 'antipsychotics', making this elegant study relevant still.

Five main points emerged:

1 *Antipsychotics produce a significantly greater improvement in patients with acute schizophrenic symptomatology than placebo.* This action on the positive (or 'acute') symptomatology is the *primary class action* and is beyond doubt. Reviewing the first two decades of antipsychotic use, Davis and Garver[5] found that in 86 per cent of controlled studies chlorpromazine was superior to placebo and that *all* 26 trials utilizing more than 500 mg/day reported definitely greater benefit than placebo. In this dose range *no* trials found only marginal benefits. A similar view can be drawn regarding the efficacy of licensed second-generation antipsychotics.

LEARNING POINT: The antipsychotic effect of antipsychotic drugs (or the magnitude of that effect) is *not* shared by other types of psychotropic agent such as sedatives, is not in need of further replication, and forms a valid basis for classification.

2 *Antipsychotics of (three) different chemical types do not differ in efficacy*. The conclusion that all antipsychotics have comparable 'acute' efficacy is *not* to say that clinically there is nothing to choose between them. Although sharing a primary class action, antipsychotics encompass diverse pharmacologies that in practice can overshadow the action they share in common. There are therefore many ways of viewing class members apart from the traditional subgrouping by chemical structure—e.g. clinical adverse effect profile, subdivided into *general* or *extrapyramidal* tolerability; pharmacodynamically in terms of potency, receptor-binding profile or binding characteristics ('loose' or 'tight'), etc. This is a large 'extended' family!

Table 4.3.8.1 Antipsychotic and placebo response rates in schizophrenia

	Antipsychotic	Placebo	%
Very much improved	16	1	15
Much improved	29	11	18
Improved	16	10	6
Slightly improved	31	31	0
Not improved	6	15	
Worse	2	33	

Data from NIMH Collaborative Study.[4]

LEARNING POINT: In choosing an antipsychotic drug for 'acute' treatment, tolerability not efficacy, is the key consideration.

3 *A substantial proportion of patients show limited or no response to antipsychotic drugs*. Overall, 61 per cent of those on active drug were considered 'improved' to 'very much improved'.[4] It has usually been agreed that standard drugs produce a satisfactory response in approximately 60–70 per cent of patients with 'acute' schizophrenia, while approximately 6–8 per cent do not respond at all.[6] However, if one includes placebo response rates a less favourable picture emerges, suggesting that significant benefits could be more realistically expected in approximately 40 per cent of patients (Table 4.3.8.1).

Not all those non-responders could be classified as 'treatment resistant' (this concept, now operationally defined, long post-dated the Collaborative Study), but it does mean that in the mixture of first episode and relapsed patients presenting in routine practice, expectations for 'acute' (i.e. 'positive') symptom resolution must be *realistic* and cannot necessarily be improved by increasingly aggressive drug treatments alone.

LEARNING POINT: While providing a necessary foundation for clinical improvement to *acute* psychotic symptomatology, expectations of antipsychotic drugs on this single domain must be realistic. Treatment objectives should be correspondingly broad-based.

4 *A rapid phase (over the first week) of improvement is followed by a slower, more protracted pattern of improvement (over several weeks).* This finding has recently been replicated but with a somewhat different interpretation to that traditionally adopted. In a review of placebo-controlled studies, Agid *et al.*[7] also found that standardized ratings of positive symptomatology began to decline in the first week and improvement was significantly greater in the first 2 weeks of treatment than in the subsequent two. This was interpreted as rejecting the conventional view that antipsychotic onset is delayed. In an early clinical test of the dopamine hypothesis, Johnstone *et al.*[8] compared the two isomers of flupenthixol (flupentixol), one a dopamine blocker, the other not. They found that antipsychotic efficacy between the two groups, although evident in the first week, only became *significant* in the third week, often taken as supportive of the delayed onset hypothesis. A compromise between these two viewpoints comes from other data from this study, which showed that the first week of treatment

was also a time of significant improvement in non-specific symptomatology, especially anxiety, in those exposed to the antidopaminergic isomer.[9]

LEARNING POINT: While early improvement on antipsychotics may comprise a component of primary efficacy, such early benefits should be interpreted circumspectly in treatment planning, especially dose modifications, as they may reflect *mainly* non-specific changes.

5 *Patients in placebo arms of antipsychotic drugs trials can—and do—show improvements in positive symptomatology*. This is a standard finding in clinical trials often overlooked in routine practice. Non-specific benefits that can accrue from the environment where treatment is undertaken can contribute to maximizing outcomes.

LEARNING POINT: Drug effects have a context!

Despite their limitations, antipsychotics are unquestionably *the* key element in acute treatment. In a unique study, May *et al.* compared response rates and outcomes in a large group of schizophrenic patients randomly assigned to one of five regimes: individual psychotherapy, antipsychotics, individual psychotherapy plus antipsychotics, ECT, and 'milieu' therapy (ward environment). Patients who received physical treatments did better with increased rates of discharge, reduced lengths of stay, and decreased need for additional treatments.[10] Two years after discharge, twice as many of those treated with antipsychotics as with psychotherapy were in employment, while in the 3 years post-discharge, drug-treated patients spent less time back in hospital.[11]

(b) 'Negative' features

The issue of whether antipsychotics exert *efficacy* on negative schizophrenic features rose to prominence in the 1980s as a result of the implications of Crow's Type 1/Type 2 hypothesis.[12] The inference of this is that efficacy is 'unlikely' owing to different pathophysiological substrates proposed for 'positive' and 'negative' states. Controversy has persisted, with claims that whatever the case may be for first-generation drugs, new generation antipsychotics do possess this action. Recent reviews, however, have been increasingly circumspect.[13]

One consequences of testing Crow's proposal was that 'negative' schizophrenic symptomatology came to be viewed as something ratable *cross-sectionally*. This was a radical shift for traditionally the 'defect state' was conceptualized as a set of phenomena evident mainly from *longitudinal* appraisal, relating to complex behavioural *signs* such as psychosocial and occupational functioning. This switch assumed that the varied states that can underlie 'negative' presentations can be distinguished using cross-sectional clinical means alone. This assumption has never been validated and seems to push at the limits of clinical examination.

A major confound in assessing trial data in this field is the bradykinesia of drug-related parkinsonism, a common, pervasive feature in those exposed to antipsychotics yet one whose boundaries remain poorly defined, especially for its *subjective* components.[14] Changing to drugs or doses with lower extrapyramidal liability, or improving psychotic phenomena that may underlie social withdrawal is a therapeutic action of sorts, but hardly addresses the question. Likewise, utilizing elaborate statistics (e.g. path analysis) to support efficacy for second-generation antipsychotics ignores the fact that the problem is clinical attribution, not analysis.

Table 4.3.8.2 The classification of 'negative' states in schizophrenia

Primary	Authentic schizophrenic state
Secondary	Positive symptomatology Depressed mood Extrapyramidal disorder Early dysphoria Bradykinesia Psychosocial isolation

After Carpenter et al.[29]

Carpenter and colleagues[15] raised awareness of this issue by emphasizing the varied clinical states that can present 'negatively', thereby introducing a differential diagnosis (Table 4.3.8.2). Furthermore, by emphasizing 'durability' in their definition of 'deficit' syndrome,[16] they reintroduced a longitudinal component, aligning this more with the historical concept of schizophrenic negativity. Interestingly, this group found clozapine's benefits on negative features to be confined to patients who did *not* conform to criteria for 'deficit' state.[17]

One cannot help wondering whether the quest for antipsychotic 'therapy' of negative states has been a wild goose chase, predicated on assessment that was reliable—but lacking validity. This view will be controversial, as will our conclusion that Crow's hypothesis has *not* yet been disproven—that it remains to be shown that antipsychotics as a class exert any therapeutic benefits on primary negative schizophrenic symptomatology.

(c) Cognitive ability

The importance of cognitive deficits underlying the overt symptomatology of schizophrenia is being increasingly highlighted. They comprise a valid endophenotype in predisposed individuals, general cognition suffers a decline in the shift to florid illness and specific cognitive impairments, especially in executive function and memory, may relate to structural changes in specific brain areas.[18] The pharmacological question is whether drug treatments can enhance cognitive performance and thereby promote benefits in other domains.

Evaluation of cognitive actions of antipsychotics faces major methodological problems and findings remain contradictory. Standard antipsychotics initially impair aspects of attention and motor behaviour which improve following continued exposure, though working memory and long-term recall do not appear to be fundamentally affected.[19] Paced performance tests tend to be affected while those that are untimed tend to be insensitive, perhaps reflecting subtle motor effects.[20]

While new generation drugs may be associated with marginally less cognitive impairment than standard drugs, data are inadequate for firm conclusions.[21] In the absence of consistent evidence of differential effects on a range of neuropsychological tests, it has been suggested that measurements of 'social competence' may represent a more appropriate target for study.[22]

Furthermore, in studies evaluating cognitive-enhancing agents as 'add on' therapy no consistent evidence of utility has emerged.[21,23]

(d) Non-specific symptomatology

Affective symptomatology is prominent in schizophrenia but the efficacy of antipsychotics on such features has received little attention.

Their utility in the treatment of *anxiety* and other manifestations of 'arousal' in acute episodes of illness has not been systematically addressed and rests largely on clinical wisdom.

There is however, a tradition in Continental Europe of attributing *antidepressant* actions to low-dose antipsychotics[24,25] (e.g. flupentixol, sulpiride L-enantiomer, amisulpride) when they may exert preferential actions at presynaptic (autoreceptor) dopaminergic sites. Data remain inconclusive but it is unlikely that such an action would be clinically useful, as presynaptic selectivity is lost at doses usually required for antipsychotic efficacy.

Depression is a common feature of untreated schizophrenia and resolves as positive psychotic phenomena diminish.[26] This probably does not reflect an antidepressant action but symptom covariation. Of greater concern is what was formerly referred to as the 'depressogenic' action of antipsychotics.[27] This again raises the difficulty of distinguishing between similar presentations of pathophysiologically different states. Van Putten and May described a dysphoric mood state in antipsychotic-treated patients ('akinetic depression'),[28] which resolved following administration of anticholinergic. This was most likely a subjective manifestation of bradykinesia.

(e) Behaviour

Behaviour can be variously disturbed in schizophrenia, though is seldom considered other than as part of a global assessment. While certain confrontational behaviours such as hostility, belligerence, and resistiveness do improve with antipsychotics,[29] this is usually attributed to improvement in positive symptoms. There is, however, evidence that certain types of behavioural disorder correlate with negative, not positive, features and may represent a distinct domain of disorder.[30] It seems likely that certain manifestations of behavioural disorganization represent independent dimensions of pathology with their own, predominantly negative, prognostic implications.[31]

(f) Maintenance

With mood disorders, 'relapse' (exacerbation of an ongoing episode) and 'recurrence' (emergence of a new episode) are well-defined. In schizophrenia, where full remission of acute symptomatology may not be a realistic treatment goal, the distinction is less clear. As the long-term aim is usually minimizing the likelihood of florid exacerbation in a disorder characterized by persisting symptomatology, the term 'maintenance' is preferable to 'prophylaxis'.

The efficacy of antipsychotic drugs in long-term maintenance of schizophrenic illness is beyond doubt.[32] This applies to first-episode patients and to those who have suffered multiple episodes. Nonetheless, it remains difficult to quantify the effect as published figures vary widely. Reviewing relevant trials (covering variable follow-up intervals), Janicak *et al.* concluded that on average 55 per cent of those on placebo relapsed compared to 21 per cent on active medication, providing overwhelming statistical support for the maintenance effect.[32] In qualification, maintenance studies tend to be biased towards patients who have already shown a degree of response and it is likely that the magnitude of this effect, at least over 12–24 months, is less than trial-based analyses suggest.

A crucial question for clinicians is how long maintenance medication should be continued. The evidence is clear. Relapse rates have been shown to be similar following cessation in groups maintained well for differing lengths of time, from months to years.[33] Furthermore, no difference in relapse has been found in those who responded well compared to those whose response was less good.[34] The implication is that, no matter the *duration* or the *quality* of well-being on antipsychotic maintenance, relapse is *inevitable* following discontinuation in those with an established relapsing-remitting illness pattern (i.e. two or more episodes), which comprise the majority.

Not only is relapse a characteristic inherent to these illnesses, it seems so too is *time* to relapse. Davis *et al.* showed that placebo relapses plotted over a 2-year period occurred along an exponential line, indicating a constant *rate* of relapse, calculated from pooled data at a steady 11.5 per cent per month.[35]

Thus, the clinician's position is clear—for maintenance of well-being following second or subsequent acute episodes, 'long-term' antipsychotic maintenance means 'lifelong'. Should patients decide to discontinue, past experience can offer an invaluable tool in predicting when relapse is likely.

Some patients see long-term maintenance as 'well-being' only of sorts, in which quality-of-life is unacceptably impaired. In such individuals, *targeted intervention*, where treatment is focused on prodromal relapse symptomatology, has intuitive appeal. Alas, the trial evidence does not support this as a general strategy. No controlled studies so far have found advantage in targeted intervention and in a meta-analysis, Davis *et al.* calculated 25 per cent relapse rates in those continuously treated, rising to 50 per cent in the targeted group.[36] While carefully selected individuals, who can work with family and psychiatrists, may prefer this approach, a further potential concern is that intermittent exposure to antipsychotics may increase liability to tardive dyskinesia.[14]

Intermittent treatment is, of course, a feature of poor compliance (also known as 'adherence' or 'concordance'), itself perhaps *the* major contributor to relapse. In terms of major medical events, antipsychotics have a highly favourable risk:benefit ratio but in terms of medically trivial but unpleasant, intrusive adverse effects, the risk:benefit ratio is *unfavourable*, something often overlooked. Ensuring a maximally effective *and* tolerable maintenance regime is a joint exercise, from which the doctor cannot be excused.

The evidence that depot formulations enhance compliance is strong, if largely indirect. Support comes from 'mirror image' studies, where time in hospital is compared prior to and after starting depot. Six such studies were unanimous in showing substantial reductions in time spent in the hospital after switching to depot (average reduction: 77.8 per cent).[36] There is nothing to suggest that this reflects additional *therapeutic* advantage inherent to depots, whose benefit lies simply in facilitating regular administration and objective monitoring of compliance. Long-acting injectable risperidone (not a 'depot' in the traditional sense) has so far received favourable assessment[37] but it is too early to provide head-to-head comparative data with conventionally formulated depots. Dosages are also an important consideration in maintenance, as those compatible with maximizing long-term tolerability and well-being are likely to be considerably *lower* than those necessary for acute symptom control. However, the data are insufficiently clear to provide specific guidance. Kane *et al.* showed that relapse rates were significantly higher in patients receiving fluphenazine decanoate in a dose of 1.25 to 5 mg two weekly compared to those receiving 25 mg,[38] yet Baldessarini *et al.* have calculated that in long-term treatment, half-maximal effective doses (ED50) may be as low as one-fifth to one-tenth those normally employed.[39] In the absence of

specifics, general principles must suffice. Kane and colleagues also showed that while relapse was more likely on low-dose regimes, neurological tolerability and psychosocial/quality-of-life parameters were superior.[38] Gradual pursuit of the minimal effective dose is all that can be recommended. Bearing in mind the exponential pattern of relapses, with a modal point at 3–5 months,[40] 'gradual' should equate to decrements at intervals of *months*, not weeks.

(g) Treatment resistance

The limitations of antipsychotic treatment in schizophrenia have been known since the 1960s but the concept of 'treatment resistance' only sprang to prominence with publication of the US multicentre clozapine study.[41] Within an operationally defined framework of 'resistance' (failure to respond to at least two antipsychotics of different chemical type administered in adequate dose for a minimum of 8 weeks), Kane *et al*. showed that 30 per cent of those on clozapine improved, while only 4 per cent on chlorpromazine/benztropine did likewise ($P < 0.001$).[41]

This has been interpreted as proving that clozapine possesses superior *efficacy* over standard agents. While this is one interpretation, it is not the only one. In this study, a chlorpromazine: clozapine dose equivalence of 2:1 was assumed, which might have disadvantaged chlorpromazine, allowing the interpretation that clozapine's advantage lies in its unique neurological *tolerability*. However, many other studies have confirmed clozapine's edge in patients who fail to respond satisfactorily to other antipsychotics and whatever the explanation, this is *real* added benefit. It does not mean however, that clozapine is a 'miracle' drug. While 30 per cent of such patients improving is welcome, the criteria for 'improvement' here were modest and subsequent review has failed to show substantial long-term benefits in higher level functioning, such as occupational ability.[42] Overall expectations of clozapine, a potentially difficult drug to administer, and for patients to tolerate, must be realistic.

Similar benefit has been claimed for other second-generation antipsychotics. However, there is *no* conclusive evidence that such advantages can be attributed to *any* other antipsychotic agent.

Management: psychological and psychosocial interventions

While nowadays, there is no suggestion that *the* core intervention in schizophrenia should be anything but medication, the limitations of medication alone in symptomatic, relapse prevention, and satisfaction/quality-of-life terms have long prompted interest in wider forms of management. Randomized-controlled studies of psychological and psychosocial interventions are complex and expensive to undertake and hold many potential problems—sample representativeness, high drop-outs, appropriateness of controls, fidelity to the intervention, blindness, etc. As a result, there have been many fewer such studies than of drugs, though in recent years new work, evaluating especially psychological interventions, has been published.

The major types of intervention include:

- Cognitive behaviour therapy
- Psychodynamic psychotherapy
- Social skills training
- Psychoeducation
- Family interventions

It is also convenient to consider aspects of service organization here.

(a) Cognitive behaviour therapy

As striking as the decline in dynamic psychotherapies over the past two decades has been the rise of *cognitive behaviour therapy* (CBT) as a clinical and research focus across psychiatry. Especially in the UK, this has extended to advocacy in schizophrenia. With over 20 randomized-controlled trials and five meta-analysis seeming to support its use, a place in management should be beyond doubt.[43]

The fact is, however, that doubt *does* remain. A recent Cochrane review[44] found that CBT did *not* reduce relapse and readmission compared to standard care (though it did decrease the risk of staying in hospital) and while it improved mental state over the medium term, after a year these slight benefits had disappeared. Continuous measures on mental state did not demonstrate consistent effects. Compared to supportive psychotherapy, CBT had no effect on relapse and when combined with a psychoeducational approach, no significant reduction in readmission rates relative to standard care alone could be demonstrated.

While some individual studies have shown impressive results, the powerful advocacy CBT has attracted as adjunctive management in schizophrenia seems, at this stage, disproportionate to the evidence base. Studies have been built around fundamental design flaws, most notably in relation to control conditions, allowing some to conclude that CBT can be shown to work only in poorly controlled trials and not in well controlled ones.[45] Even between studies showing benefit, it is difficult to discern what the most appropriate target(s) should be and what components ought to comprise an/the ideal CBT package. A further problem, specific to assessing CBT in group contexts, is inappropriate data analysis where independence of observations is universally assumed, something group interactions violate, with a resultant increase in Type 1 errors.[46] While CBT *may* hold promise, widespread endorsement of a resource intensive management would seem premature until more and better designed work reports.

A further application of CBT techniques in psychosis has been in enhancing compliance. While *compliance therapy*,[47] combining cognitive behaviour and motivational interviewing techniques, has shown promise, it has been insufficiently evaluated to support robust recommendations. A recent Cochrane review[48] identified only one study comparing this with non-specific counselling. No significant differences were found in overall 'non-compliance' rates or in mental state measures, attitudes to treatment, global functioning, or quality of life. Although at 1 and 2 year follow-ups, average number of days in hospital was reduced, this was not statistically significant. This study did not show any effect on insight, but other work has claimed that improvements can be achieved by short, insight-focused CBT interventions, but at the expense of increasing depression,[49] an observation also reported with non-CBT approaches targeting insight.[50]

Cognitive remediation (or rehabilitation), in which the desired end-point is not symptom reduction per se but improved *global* functioning via amelioration of cognitive deficits such as impaired vigilance, attention, and planning/decision-making, has also been applied in adjunctive management. Once again, while some individual findings are encouraging, data remain inconclusive.[51]

LEARNING POINT: While present evidence does support the use of CBT led interventions in adjunctive management of schizophrenia, the research is flawed and further, *well controlled* studies are necessary to determine a precise role.

(b) Psychodynamic psychotherapy

Unlike Freud, many analysts of the early-mid twentieth century were undaunted in pursuit of psychodynamic understanding of, and management for, schizophrenia but theories were universally unsupported by evidence. When assessed against supportive psychotherapy, no advantages could be demonstrated.[3] While some modest revival of interest may be detected, especially amongst advocates of 'early intervention', a recent review provided no support for such revisionism.[52] Furthermore, May's study[10,11] offers a cautionary warning of the potential for harm.

LEARNING POINT: Insight-orientated dynamic psychotherapy has no current place in the management of schizophrenia.

(c) Social skills training

As a result of the early age of onset, relapsing nature, and persistence of many clinical features, schizophrenia can potently disrupt smooth acquisition and evolution of skills essential for developing mature interpersonal relationships, occupational competence, and independent living. *Social* (or life) *skills training* evolved in the context of resettlement programmes aimed at discharging long-stay, institutionalized patients but in various forms remain widely practised. It is based on a structured learning-orientated approach to the acquisition of skills relevant to the individual and the demands of his/her environment.

Unfortunately, social skills training is difficult to evaluate, as this has become a 'blanket term' covering a wide range of applications and targets. While some studies have focused on rehearsal of activities of daily living, others concentrate on communication and conversational skills, and although some view improvement in symptoms as the underlying goal, for others the benefit lies with cognitive ability. Blindness of assessments is a major problem, though can be easily achieved using blinded video techniques. Thus, while individual studies have found improvements in assertiveness, general social competence, and even speed of discharge, with benefits extending to a widened social network and that generalize, a recent Cochrane review failed to find conclusive evidence of benefit.[53]

Illness self-management is part of several social skills training programmes but has been singled out as the focus of specialized techniques, comprising video modelling, role-play, and specific problem-solving combined with homework. While promising results have been reported,[54,55] this intensive approach requires further evaluation.

Vocational training (or rehabilitation) is not directly dependent on social skills training but is related to it. Although the majority of schizophrenic patients end up unemployed (up to 85 per cent in the US; 73 per cent in the UK),[56] specialist vocational training remains a scarce component of long-term illness management in most services. A review[56] concluded that *supported employment* models were significantly more effective than *pre-vocational training* in facilitating competitive employment, the latter being no better than standard community care. Supported employment was also associated with higher earnings and more hours working per month. It is sobering that even in trial contexts an average of only 34 per cent of individuals in supported employment were actually employed at 18 months, the comparable figure for those who received pre-vocational training being 12 per cent.[56]

LEARNING POINT: Although intuitively sound and generally appreciated by patients and families, 'social skills training' can only be given a firm evidence base with further well controlled studies in which individual components of therapy are 'teased out' for separate evaluation and specific end-points are genuinely blindly assessed.

(d) Psychoeducation

Much of the above comprises an element of 'education' in the widest sense but programmes have been advocated in which information exchange is *the* key intention. *Psychoeducation* can be targeted on the patient to improve outcomes, enhance compliance, and increase knowledge, including on early relapse recognition, thereby contributing to a better sense of well-being. Imparting factual material is also a fundamental component of many family interventions (see below).

In general, patients appreciate sessions in which their illness is explained, reinforcing the idea that some understanding is possible in situations which may seem incomprehensible. Furthermore, explaining bizarre experiences and beliefs in *illness* terms can help de-stigmatize preconceptions they themselves may hold.

Like most psychological and social interventions advocated as adjunctive strategies for schizophrenia, 'psychoeducation' is not a single procedure with standardized delivery, something that inherently limits systematic reviewing. Nonetheless, it has a favourable review in the Cochrane Library,[57] which found that relapse and readmission rates were significantly reduced (NNT = 9) at 9 to 18 months. Beyond this however, no effect was found on insight, medication-related attitudes or overall satisfaction with services by patients or families. Thus, while data to date are consistent with some *overall* benefits from psychoeducation, those that are proven remain few.

LEARNING POINT: Psychoeducation, as a blanket concept, is attractive to patients and carers and, because it is brief and inexpensive, to service providers, but its *therapeutic* impact may be limited. Future studies must define component elements and address complexities, such as the distinction between knowledge and understanding, and the vulnerability of learned material to degeneration.

(e) Family interventions

The therapeutics of schizophrenia broadened from sufferers to families and carers with the observation that criticism and hostility from a close relative ('expressed emotion': EE) was an important determinant of relapse. As families remain the key element of support for most patients, development of positive, constructive ways of helping them provide this has rightly formed a considerable research focus over the past 30 years. With regard to EE, the consensus is that reduction reduces relapse risk when combined with maintenance medication.[58] However, this appraisal is *not* unanimous and the importance of EE reduction is in need of modern systematic review.

Family interventions have broadened to include not only educational input about the illness, its consequences and service availability, but also psychosocial interventions aimed at developing an alliance (especially important during first presentations), reducing emotional distress, boundary setting, and instillation of realistic expectations of both patient and services. This component

heterogeneity again makes straight literature comparisons difficult. A recently updated Cochrane review[59] reported somewhat more equivocal findings than previously, suggesting that while family intervention *may* decrease relapse frequency, the number needed to be treated to prevent an episode of relapse (NNT) had risen from six in a previous review to eight. The authors emphasized that some negative studies may have been missed by their search, rendering even this figure tentative. There was clearer evidence of an effect in reducing hospital admission (NNT = 8) and a *likely* beneficial effect on compliance (NNT = 7). There was no effect on the tendency of individuals to drop out of care but a *likely* improvement in general social impairment and in levels of EE. The effect on suicide seemed neutral. While this may be taken as an endorsement of family interventions, the reviewers offer a salutary caution, suggesting that interested parties—clinicians, patients, policy makers—'cannot be confident on the effects of family intervention from the findings of this review'.

LEARNING POINT: Families generally welcome professional input dedicated to *them* and *their* needs at distressing times such as first diagnosis and subsequently, with realization of the implications. As a quality-of-care issue, such involvement is beyond the benefits trial evaluation can provide. However, in proposing what *works* therapeutically—and works *best*—further research is necessary to define specific elements of interventions that are not only appreciated but that contribute unequivocally to well-being of sufferers and family members.

So—using the *evidence*—what can be concluded from the above? In reviewing psychosocial management strategies advocated for those with schizophrenia and their relatives, two observations emerge. First, is a certain 'regression to the mean'—the more interventions are studied, the harder it becomes to replicate benefits enthusiastically identified early on. The second might be viewed as cynical but is worth stating nonetheless—namely, that in our quest to maximize the care of those with this fell disorder, it has been conclusively established that *drugs plus* 'something' produces better outcomes than *drugs alone*. What remains unclear is whether 'something' amounts to more than projecting high levels of professionalism and/or a common humanity or whether there is indeed a *specific* therapeutic component (or components) to any or all of these 'somethings'.

It is disappointing that after decades of research such a conclusion is still possible and future studies not only need to address the standardization of component element(s) of psychosocial management but must establish a supremacy for it/them before psychiatrists in routine practice and policy makers alike are able to invest not simply in services that others 'like' but that doctors can confidently endorse as having firm *evidence* of therapeutic benefit.

(f) Service organization

The rationale of 'community care' in promoting independence and choice is noble but its implementation has frequently been found wanting. A major thrust in community care management has been the introduction of *community mental health teams* (CMHTs), comprising a comprehensive range of disciplines that bring both collectivism to decision-making and varied expertise to service delivery. Compared to non-team care, CMHT management does promote a greater acceptance of treatment options but further advantages are hard to identify, though reductions in admission and suicides are *possible* benefits.[60]

It has long been clinical experience that not all patients with acute psychoses, including schizophrenia, require hospitalization and, although the research base (specifically in relation to schizophrenia) is slender, *home treatment* programmes have vocal advocates, partly because of the promise they hold for reduction in costly bed numbers. While patient satisfaction is usually high, early quality-of-life benefits may not be sustained and, owing to ongoing need for inpatient facilities, cost benefits may be limited long-term.[61,62] These are difficult services to organize, require intensive staffing and sophisticated multiagency working, and even then, staff morale can be hard to sustain.[63] They are highly selective in who they accept[64] and may work most effectively for those whose family units remain well-integrated and supportive,[65] such as adolescents, but doubts spanning all the potential benefits are sufficient to suggest caution in making empirically based leaps to such radical service reorganization in the absence of further data.

The policy of institutional closures that began in most developed countries in the 1960s/1970s did not result in reduced patterns of acute bed usage. In fact, *rising* admission rates caused concern that day and outpatient care alone were inadequate because of failure to maintain engagement or to meet complex needs. Case management and a variant, assertive community treatment (ACT), arose as ways of optimizing ongoing service involvement and ensuring coordination and delivery of care appropriate to individual clinical and social needs. Both aim to: maintain patients in contact with services; reduce frequency and durations of admissions; improve outcomes, especially social functioning and quality of life. Case management is essentially based on 'brokerage', where an *individual* member of the multi-disciplinary team is responsible for assessing needs, developing a care plan, arranging implementation and monitoring quality and engagement, and is also involved in an element of delivery. This is now seen as the least robust model and more complex variants have arisen. With ACT, the emphasis lies in *team* working.

Systematically reviewed, *case management* ensures that more people remain in contact with services but with the consequence of *increased* admission rates.[66] Furthermore, rather than shortening admissions, it may actually increase their duration. There is no evidence that it improves outcomes on any clinical or social variables. Compared to standard community care, patients receiving ACT are more likely to remain in contact with services, less likely to be admitted, and spend less time in hospital with other benefits in terms of accommodation, employment, and satisfaction.[67] However, mental state and social functioning are *not* improved and with overall costs accounted for, ACT is *not* less expensive. In comparison to case management, ACT has advantages in terms of reduced time in hospital but overall there are no cost benefits.[67] Few other comparisons are possible due to inadequate data.

The impression that case management should be dropped and ACT flourish[67] is somewhat oversimplified. The negative appraisal of case management as of 'questionable value' has been criticized[68] while the 'effectiveness' of ACT seems to rest on the recurrent problems of content and fidelity of delivery noted in relation to all the psychosocial interventions discussed here.

LEARNING POINT: Like drugs, organizational structures operate in a social and service development context and what applies to one patient group in one national or local environment might not reap similar benefits in another. Community-based

or otherwise, *the* weakness in the care of those with chronic, relapsing-remitting disorders like schizophrenia, characterized by autistic withdrawal and social dislocation, is failure to maintain engagement with specialist services. Any structure that fosters a proactive approach to engagement while maintaining staff morale is likely to be better appreciated and to provide better outcomes.

Treatment and management principles

The following is based on UK experience and may not translate completely across international borders where local traditions and organizational constraints may modify practice. It is only presented as an *outline* of issues to be considered and cannot provide a universal blueprint for care. As will become apparent, the authors would argue that such 'blueprints' do not best serve the interests of patients or the expertise of those implementing care.

Doctors confronted with a patient believed to be acutely psychotic must address three preliminary questions:

1 *Does the patient require admission or can they be managed as an outpatient?*

Some of the issues were aired above but from the *medical* point of view, there are a number of scenarios in which admission remains a priority (Table 4.3.8.3).

2 *Can the clinical situation be dealt with informally or are compulsory legal powers required?*

Table 4.3.8.3 Considerations in relation to admission policies in patients with acute schizophrenia

Supporting admission	Supporting non-admission
Unstable mental state Rapidly extending content Variable affect	Mental state disorder stable/slowly evolving
Imperative auditory hallucinations To harm self To harm others	Absent/no will to act
Marked affective change Suspiciousness, anger Depression	Affective change mild/amenable to reassurance
Behavioural disturbance Disorganization Dangerousness Commission Omission	Minimal behavioural disturbance/risk of harm
Cognitive disturbance Lack of insight Impaired attention/distractability Inability to comprehend advice Hopelessness/suicidal ideation	No barrier to engagement
Inadequate social support Living alone Vagrancy/neglect	Good social supports
Any (other) reason for non-compliance	Likelihood of compliance
Medical state Intercurrent physical illness Substance misuse/dependency	Physically fit
'Asylum'	Aversion to inpatient care

This will depend on the thorough *risk assessment* that must be a part of every examination. This must, of course, include the *health* risks, not solely those involving threat. The details of implementing compulsory detention for assessment and/or treatment will differ in different jurisdictions.

3 *What is the best first step in treatment?*

There is *no* single first–line drug for the treatment of acute schizophrenic episodes—nor, the authors would suggest, any first-line *type* of antipsychotic. Guidelines are fairly unanimous in recommending 'atypical' antipsychotics, especially in first episodes, and any recommendation qualifying this requires justification.

Despite its persistence, the term 'atypical' has never attained *pharmacological* credibility. It rests on a single *clinical* parameter—a perceived reduction in liability to promote extrapyramidal side-effects (EPS)—*quite specifically*, drug-induced parkinsonism. The problems surrounding a subclassification based on such a vague parameter are multiple (e.g. the boundaries of parkinsonism; inadequacies of rating schedules; discrepancies in dose equivalences between trial and comparator agents) making 'atypicality' of dubious scientific validity.[14] Carefully conducted trials, not sponsored by industry, have recently raised questions about putative advantages in EPS tolerability, even with high potency comparators such as haloperidol, when appropriate equivalence is used.[69] With quality-of-life parameters no different after up to 1 year, and no detectable patient preference,[70,71] objectively it is hard to see 'atypicality' as having any merits beyond product marketing. There is certainly little to support any *inherent* value for clinical decision-making beyond the fact that new generation drugs extend the options.

However, there is now strong *evidence* to challenge the blinkered 'algorithmic' prescribing guidelines can foster, especially in relation to 'atypical' antipsychotics. Results of the Clinical Antipsychotic Trials of Clinical Effectiveness (CATIE) study raise important issues about the tolerability profiles of different new generation compounds in relation to first-generation drugs[72] and challenge cost-effectiveness benefits.[73] These data form part of an emerging trend that, we contend, opens up once again the *full range* of antipsychotics—new and old—to consideration as treatment options.

An alternative to 'algorithmic' practice is to view guidelines as providing a *framework* only, within which *all* clinical information can be brought to bear in prescribing decisions. This approach takes its cue from drug regulation, where the appropriateness of granting a license is based on an *individual risk*: *benefit appraisal*. This is a *clinical* judgement, the outcome of which is dependent on *context* (e.g. not simply adverse effect burden but availability of alternatives). Patients, psychotic illnesses and drugs to treat them are each diverse, harbouring far greater differences than the few similarities they share. These diversities should be entered into the individual risk:benefit appraisal in making prescribing choices. So too should the *phase* of illness one is planning for—acute through to maintenance—as the risk:benefit appraisal may shift between these in particular individuals.

Some examples of issues for consideration in individual risk: benefit appraisals are shown in Table 4.3.8.4, but the professionalism psychiatry claims can only come from expertise in recognizing and accounting for the many varied possibilities.

Table 4.3.8.4 Individual risk:benefit appraisal in the use of antipsychotics : examples of some considerations

Presentation	Possible Strategies	Schedule
Behavioural disturbance 'positive' or 'negative' = arousal	Low potency standard (solo treatment)	'diminuendo' or 'crescendo'
Prominent non-specific affective symptomatology	Low potency standard Or New generation + benzodiazepine (potentiation)	Rapid 'crescendo' to tolerance
Marked insomnia	New generation + benzodiazepine Or low potency standard	'crescendo' to 6–8 hours nocturnal sleep + maximum of two hours in the day
High risk of psychiatric emergency	Low potency standard (+/- additional medication)	Slow 'diminuendo'
Tenuous engagement Poor past medication experiences (tolerability) Overfamiliarity with common regimes	New generation (solo treatment)	Slow 'crescendo'
Poor physical health Overweight Family history of CV disease	High potency standard	'crescendo'
Middle aged High CVS risk factors	High potency standard	Low dose regime assessed without change over protracted period
Prior or present EPS symptomatology	'loose binding' new generation (quetiapine)	Slow 'crescendo'
Established history of poor long-term engagement/ compliance	Long-acting injectable	Ultra-slow 'crescendo'

'diminuendo' = starting with higher doses and tailing down to tolerability

'crescendo' = starting with lower doses and building up to tolerability

Identifying goals and defining structure

The key to avoiding confusion and 'decision paralysis' in dealing with the complex clinical situations schizophrenia presents is to delineate the *structure* within which it is hoped to achieve a series of treatment/management *goals*. Although arbitrary, three 'phases' can be identified:

- Acute
- Post-acute
- Maintenance

The acute phase

This encompasses treatment during the maximally florid symptomatic period, corresponding to first presentation or subsequent acute exacerbations (Fig. 4.3.8.1). It is characterized by a significant shift to *illness*, though surprisingly, it remains unclear exactly what changes define this shift.

The goals include:

i) control of intrusive, non-specific symptomatology (e.g. anxiety, agitation, and especially insomnia);

ii) maximizing safety and well-being of the patient and others by containing chaotic, socially damaging behaviours;

iii) engaging the patient in therapeutic recommendations and (mental state permitting) gaining consent for treatment plans;

iv) implementing an appropriate foundation drug regime;

v) stabilizing positive symptomatology;

vi) preventing, or if unavoidable, treating psychiatric emergencies.

The risk:benefit appraisal at this early stage is driven by the first four of these, for the ultimate goal of acute phase treatment—stabilization of positive symptomatology—is, as noted, likely to be delayed. It is important not to overlook other goals that can be achieved quickly, especially those that can be held up as evidence of progress, such as improved sleep.

These goals may be achieved using a single antipsychotic, which is the *ideal*. In this regard, low potency first-generation drugs, such as chlorpromazine, have appeal because of low cost, extensive usage, wide dose flexibility and potent, if rapidly habituating, sedative properties. Flexible dose studies suggest the majority of responses will be achieved in the range of 500–600 mg/day[5] with some suggestion that first-episode patients may respond at lower doses. While low potency drugs have a justifiably admirable reputation in the treatment of presenting, acute phase symptomatology, they are not generally ideal as sole long-term treatment in view of the often intrusive effects of lingering sedation. As patients who have experienced benefit on a particular regime tend not to like changes, and such switches can be clinically problematic because of poor dose equivalence data, chlorpromazine alone is usually best reserved as an initial, short-term strategy.

High potency first-generation drugs, especially haloperidol, have also been widely used, especially in the United States, and while effective, contain a potential problem. High potency drugs are *safe* and tend to be used in higher doses than low potency ones (in one study, 4–6 times the low potency equivalent[74]). This undoubtedly follows from the fact that low potency compounds are inherently dose-limited by anti-autonomic and sedative actions. Unsurprisingly, liberal early use of high potency drugs is associated with higher rates of EPS (dystonias, akathisia, and parkinsonism).[14] Two points should be borne in mind in using high potency first-generation antipsychotics. Firstly, it has been shown in both clinical and functional imaging (PET) studies using D2 occupancy levels that usual minimum effective daily doses of haloperidol lie somewhere between 2 and 5 mg.[75,76] Secondly, even utilizing slightly higher doses, EPS need present no greater problems with haloperidol than they do with placebo or olanzapine.[69,75]

Nowadays, most clinicians tend to pursue a new generation drug as their first choice and this does have some practical advantages, not least single dosing and orodispersable formulations. However, while most new generation drugs have some sedative properties, these are clinically less prominent than with low potency first-generation compounds, so achieving early acute phase goals can be protracted. Furthermore, these drugs usually now come with defined protocols relating to starting doses and rate of increments which do not apply to older drugs. A number of guidelines address

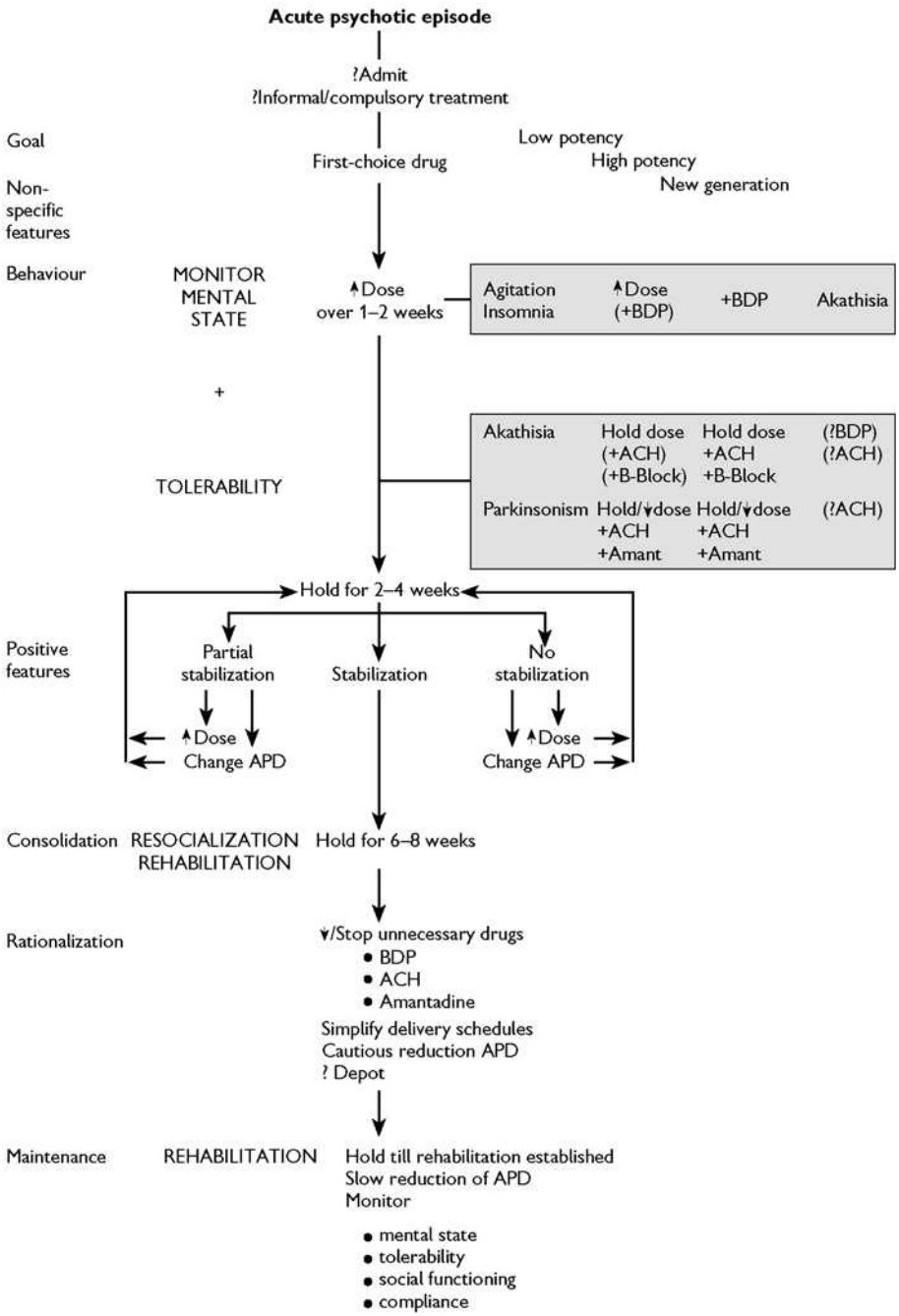


Fig. 4.3.8.1 Outline plan for treatment/ management of schizophrenia: APD, antipsychotic drug; BDP, benzodiazepine; ACH, anticholinergic; B-Block, ß-blocker; Amant, amantadine.

the potential impediment this introduces into achieving acute phase goals by recommending the adjunctive use of a sedative drug, especially benzodiazepines, to gain early control of intrusive non-specific symptomatology.[77] While this is sound, there is another, more traditional, approach—namely, the addition of a second, sedative antipsychotic.

This introduces the issue of polypharmacy, against which the literature has long (and rightly) railed. However, notwithstanding these many cautions, it might seem more logical to treat non-specific symptomatology that is *psychotically mediated* with an antipsychotic, even if one's goals are largely short-term and symptomatic, rather than to introduce a further class of medication whose benefits will not contribute to the fundamental treatment issue (see below) and may cause their own, unrelated problems, including behavioural 'dyscontrol'. Antipsychotic polypharmacy *must* be reviewed regularly, justified constantly and rationalized when possible. The real 'sin' of polypharmacy is its persistence in the treatment plan 'by default'. However, it must be acknowledged that behavioural disturbance and non-specific symptomatology may be so overwhelming that benzodiazepines may be useful in order to avoid over-rapid, excessive escalations of antipsychotic doses.

Even if not introduced as a fixed part of early treatment recommendations, prescription of additional antipsychotic on an 'as required' (or 'PRN') basis is virtually standard practice, allowing nursing staff to intervene at their discretion in the face of escalating symptomatology. This has no trial basis[78] and although a pragmatic solution to inherently unstable situations, must be utilized with discretion. Indications, maximum doses, frequencies, and modes of administration must all be written up separately and unambiguously. Such regimes should *not* be viewed merely as a means of 'keeping the peace' but should be used to inform judgements about how far short of practical requirements initial treatment recommendation fall, invaluable information to incorporate into each treatment review.

Antiparkinsonian medication should be automatically used with higher doses of high potency standard antipsychotics (=/> 10 mgms haloperiod daily or equivalent)[69]but not with lower doses or other acute phase choices.[14] While anticholinergics reduce the risk of acute EPS, only a minority will develop these to a clinically significant degree and antiparkinsonian drugs will interact with antipsychotics to interfere with their actions. These drugs have their own profile of antimuscarinic effects which can be minimized by choosing the most M1-selective compounds (e.g. biperiden, procyclidine) should be used.[14]

A further point relates to goal (iii) above. Patients are often more experienced psychopharmacologists than their doctors! In deciding treatment regimes in those with past histories, be advised by the patient. Medications in which they have confidence should always be one's first-line in treating acute relapses.

Prior to the emergence of goal (v) above, monitoring of acutely psychotic patients should be *frequent*, and preferably daily, each assessment resulting in a review and, if necessary, modification of the treatment plan. At this stage it is rarely useful to commence formal psychosocial, including educational, programmes with patients, whose cognitive difficulties will make it hard to engage meaningfully, though as part of one's ongoing dialogue, monitoring progress and addressing tolerability and compliance serves an elementary educational purpose. Likewise, one's interactions with families are best restrained to exchange of such basic information as the patient will permit (e.g. diagnosis and treatment) and to general support.

The boundary of acute phase of treatment is not rigidly defined, nor is it necessarily set by major symptomatic reversals. The ultimate aim is 'stabilization', as evidenced by the psychotic process becoming less 'active', delusions are recalled as yesterday's events, hallucinations become less preoccupying, behaviour more amenable. It may also be that sedation, previously an ally, becomes intrusive and a source of complaint.

The post-acute phase

This is characterized by the re-emergence of stability in both mental state and behaviour. The following goals should be considered:

i) consolidation of clinical improvements;

ii) rationalization of treatment regimes;

iii) resocialization.

The different domains of schizophrenic symptomatology do not all improve at the same rate and the first signs of amelioration should not be taken as evidence that the tide has turned. Global improvements need to mature into *specific* improvements in particular domains of symptomatology before this can be assumed. Practically, caution is important as over-rapid reductions in antipsychotic dosages before fundamental therapeutic changes have bedded down increases the risk of symptom exacerbation and although there is little evidence on the matter, clinical experience suggests that such very early setbacks are not only demoralizing but more difficult to stabilize, requiring higher doses for longer periods to recapture previous gains.

Nonetheless, the transition from acute to post-acute phase usually marks a pivotal point when the balance of appraisal shifts towards the increasing contribution of 'risk' to the risk:benefit assessment. Taking account of this shift is important in maximizing compliance, both short and long term. So an important step of post-acute treatment is about *gradual* change towards simplifying regimes and maximizing tolerability. The emphasis on 'gradual' is still important, as over-rapid reductions may be implemented ahead of relapse set in train by previous reductions.

In monitoring progress, two issues should be borne in mind. First, is a clear impression of the *criteria* on which 'improvement' are to be based. The traditional medical emphasis is on symptomatology, especially positive symptoms, but elimination of *all* psychotic phenomena, or development of *full* insight, may be unrealistic goals, particularly in those with established illness. Such attempts usually come at the expense of dose regimes significantly higher—and with correspondingly greater long-term risks—than are necessary for a good quality of life and maximum compliance.

Secondly, patients and families—and those who pay for care—often have unrealistic expectations of the timescale over which outcomes can be assessed in acute schizophrenic episodes. Considering both psychopathology and 'degree' of remission, Lieberman *et al.* found that the median time to 'remission' was 11 weeks, with a *mean* time of almost 37 weeks—and that, in a first-episode sample![79] Other studies have produced similar findings. Even if full 'remission' is not the goal of acute and post-acute treatment, the relatively protracted timescales over which 'benefit' is to be considered must be taken into account. There is *no* evidence that this process can be speeded up by escalating doses, which on the contrary, may prolong the situation by introducing unnecessary and intrusive adverse effects.

In this regard, it is worth introducing another concept from past studies. In the 1960s a number of authors described a phenomenon of so-called 'neuroleptic toxicity', an apparently paradoxical worsening of mental state with escalating antipsychotic doses.[80] In reviewing the dose–response literature, Baldessarini and colleagues suggested that the curvilinear relationships usually found may be explained by increasing extrapyramidal symptomatology.[39] Thus, with antipsychotic drugs, while 'less' may not necessarily be 'more', 'more is usually less'!

The post-acute phase is the time to introduce appropriate elements of wider 'management'. With the patient, this might include 'education' in a more formal sense than hitherto, tackling the nature of the condition and, most importantly, the key role of medication in its treatment, including addressing potential long-term recommendations. Regardless of one's interpretation of the literature on its specific benefits, CBT *principles* can be helpful in structuring goals and realistic pathways to attaining them and in re-instilling a sense of control and optimism. Families, too, can now more productively be brought into formal educational programmes, either singly or in groups.

The maintenance phase

The boundary between post-acute and maintenance phases is the least defined but is reached when remission—maximal improvements in all major domains of disorder—can reasonably be considered to have occurred. The major goals now are:

i) maximum well-being with minimum adverse effects;

ii) monitoring efficacy/effectiveness and tolerability;

iii) continuing or extending rehabilitation and social integration.

Attempts to reduce medication should still be cautious but, over the longer term more determined, as one seeks the *minimal effective dose*. As noted, evidence suggests that maintenance regimes can be considerably lower than those utilized for acute treatment but this aspect of long-term care is often omitted, perhaps on the basis that no one wishes to 'rock the boat'. The result is that maintenance may be facilitated by unnecessarily high doses that, in turn, may impede compliance and increase the liability to long-term adverse effects. These need not be simply neurological but may extend to wider domains of functioning. In the Northwick Park First Episodes study significantly more patients in the placebo group were found to have some clear achievement at 2-year follow-up compared to those on active antipsychotic.[81]

This finding referred to small numbers but raises the important question of what are the most satisfactory criteria by which to gauge long-term treatment response-domain-specific criteria or global outcomes. Whichever is selected, a second important shift occurs in the risk:benefit appraisal in this phase, the goal being the active elimination of as many components of 'risk' (i.e. side-effects) as possible.

Monitoring neurological tolerability should involve both enquiry into *subjective* adverse effects as well as examination for signs,[14] the most efficient being simply an assessment of the patient walking.

The most ambitious aim of maintenance comprises engagement in management geared to attaining the highest possible level of psychosocial and, where possible, occupational functioning for the patient, with carers able to exert the greatest degree of understanding and coping skills.

Depression, affecting up to 70 per cent of patients in the acute phase, tends to remit with the psychosis[26,82] but as many as one-third will develop depression in the maintenance phase. This *post-psychotic* (or post-schizophrenic) *depression* is likely to be as aetiologically heterogeneous as depression in other contexts but has been poorly studied including from a therapeutic perspective. Despite initial pessimism, there is evidence that such mood states respond to tricyclic antidepressants[83] though response is impeded in those experiencing residual or recurring psychotic symptoms on low maintenance antipsychotic regimes.[84]

Negative symptoms are present throughout the course of schizophrenia but are most likely to raise specific therapeutic issues during maintenance. An outline of care options is shown in Table 4.3.8.5.

Table 4.3.8.5 Outline management of 'negative' states in schizophrenia

Question	Intervention
1 Is the patient actively psychotic?	Start/increase antipsychotics Reduce levels of stimulation
If not: 2 Is there evidence of extrapyramidal side-effects?	Anticholinergics Amantadine Reduce antipsychotic doses Switch antipsychotic New generation Low ptency
If not: 3 Is there evidence of dysphoric mood?	Antidepressants Anxiolytics Reduce antipsychotics Supportive management Switch to new-generation antipsychotic
If not: 4 Has psychosis recently resolved?	Supportive management
If not: 5 Is the environment impoverished?	Resocialization Rehabilitation
If not: 6 Is the patient receiving long-term medication?	Reduce to reasonable maintenance dose Switch to new-generation antipsychotic Clozapine
Then: 7 Is the problem a 'deficit' state?	Adapt expectations to the patient's capabilities

After Carpenter *et al.* (29)

Prodromes

Some would say the above is far from comprehensive, missing *the* key element in the care of schizophrenia nowadays—prevention. In fact, they would argue, by concentrating on recommendations that only apply once the possibility of prevention has passed, we are submitting to traditional therapeutic pessimism, which is out-of-step with the optimism of the times.

Prodrome refers to features that, for any illness, characterize the difference between well-being and the state of illness evolution. The patient is not unwell as such but the journey to illness has commenced.

In schizophrenia research, 'prodrome' applies to two scenarios. The first, and perhaps most therapeutically relevant, relates to second and subsequent episodes where the clinical team may have a basis for prevention—not of the illness but of the episode (i.e. *relapses*). The universally negative trial evidence in relation to targeted intervention on early relapse symptomatology has been mentioned but this may be one of those situations in which trial *efficacy* does not translate well to *real-life* situations. Maybe the patients selected for such approaches must be 'targeted' as much as the symptomatology!

Early symptom recognition should certainly be a key part of the education of both patients and their families, a task helped by the fact that as a rule, psychotic episodes run 'true to form'. Non-specific and positive features recognized from an earlier episode can be recruited to help identify emergence of subsequent episodes. Especially sensitive, is a change in sleep pattern. As was noted,

if patients have stopped maintenance medication, knowledge of the *time* to relapse following previous cessations can be useful in high lighting the 'critical period' for subsequent relapse. Such knowledge can be empowering, especially for families and carers and even though trial support is lacking, clinical experience suggests it can sometimes be useful in facilitating swift reintroduction of medication or increasing dosages from maintenance to treatment levels.

Relapse prevention through early symptom recognition is *not*, however, from where optimism currently springs. This comes from the second application of 'prodrome'—to the early phase of *illness*, not episode, development.

The '*early intervention*' movement has swelled to an influential grouping within both the research and policy arms of psychiatry. Its origins lie in a very real concern—the delay that many schizophrenic patients experience between the first signs of illness and entry into specialist care. This so-called 'duration of untreated psychosis' (DUP) is on average 1–2 years but can be longer[85] and has been linked to adversity of outcome, initially on the theoretical basis of some factor mediating neurotoxicity,[86] though this remains unsupported.

While a degree of consensus is possible on what the key elements *might* comprise, it is as yet impossible to construct a valid model of what 'early intervention' *should* comprise and thereafter to measure fidelity.[87] Nonetheless, its principles have been enthusiastically adopted and development of services is government policy in many countries, including the UK. This does however, remain controversial—not at the *quality-of-care* level, where the aims of improving awareness and service access are inherently sound, but at the *scientific* level, where evidence supporting improved outcomes remains weak.

Recent systematic reviews do point to modest benefits in terms of a lower symptom burden and delayed readmission in the short-intermediate term.[88,89] However, assessment of DUP is invariably retrospective and although it may be done with reliability, validity remains suspect. Also, the link may be confounded by, for example, some illness characteristic that mediates both delay in entering services and poor outcome, making interpretation difficult. The authors' own work with those at high risk of developing schizophrenia has pointed to the problems of attributing a 'prodromal' psychotic state solely to the emergence of what are traditionally considered 'psychotic' symptoms and to the non-specific nature of those symptoms that do seem to point to a later formal diagnosis.[90]

The precise delineation of pathway(s) to illness remains to be refined to a degree that provides meaningful positive predictive values, thereby avoiding the awkward issue of unnecessary and potentially risky interventions in those whose 'operationally defined normality' is merely different, not necessarily prepsychotic. Continued investment in 'early intervention' services must for the present be driven more by quality-of-care considerations than an evidential base.

Psychiatric emergencies

In dealing with acutely ill psychotic patients, one must always bear in mind the 'unpredictability factor' and the *potential* for aggressive or violent outbursts during acute symptomatic 'shifts'. Risk assessment of dangerousness is an imprecise science but should be incorporated into all routine clinical assessments during acute phase treatment—and conclusions *documented*. Caution should be exercised with patients who are profoundly suspicious, verbally aggressive, resistant to engagement, whose presentation does not allow for comprehensive mental state examination, whose clinical condition is complicated by substance misuse and especially, those who have a past history of assaultive or threatening behaviour.

Principles of wider management are crucial in *avoiding* emergency situations including the quality of the (ward) environment, staff:patient ratios, etc., but even with high levels of vigilance, pre-emptive plans and good quality management, emergency situations may still occur and must be dealt with decisively.

An outline plan relating to emergency situations is shown in Fig. 4.3.8.2.

Poor response and treatment resistance

Where a patient has not responded satisfactorily to an adequate dose (600–800 mg/day chlorpromazine or its equivalent) of antipsychotic for an adequate period of time (at least 6–8 weeks) they might be considered a 'poor responder' but not yet 'treatment resistant'. Several strategies have been suggested in this situation.

Antipsychotics

Conventionally, the first approach is to modify the antipsychotic regime by:

1 increasing first-choice drug to a high-dose schedule (up to 1000 mg chlorpromazine or its equivalent) for 6–8 weeks;

2 changing to a drug of different chemical type in standard dose ranges for 6–8 weeks. (Current research would suggest this should include changes between not just old and new drugs, but in the other direction too);

3 increasing the dose of the second-choice drug to a high-dose schedule for 6–8 weeks.

As far as the literature is concerned, it is only when at least one, and preferably two, of these steps have failed that the illness should be considered 'treatment resistant'.[41]

Despite adoption in routine practice, there is little evidence that such manoeuvrings are of themselves effective. More *time* may still be the crucial factor. There is some evidence that in those switching from a new generation drug, slightly better results may be achieved with risperidone and olanzapine than quetiapine or ziprazidone,[91] though this might be saying simply that, when tolerated, relatively high potency compounds do better than relatively low potency ones.

In clinical practice, especially when external pressures and a sense of therapeutic confusion have clouded therapeutic goals, antipsychotic doses can escalate 'by default'. The issue of 'neuroleptic toxicity' has been mentioned and where high-dose regimes cannot be specifically *justified*, it is worth *reducing* to average or low doses and assessing response. Furthermore, patients showing poor response may benefit from addition of a *depot*, even when compliance is not in doubt, possible advantages perhaps relating to adverse pharmacokinetic parameters, such as poor absorption or enhanced metabolism with complex or high-dose oral regimes.

IMPENDING EMERGENCY

Non-drug intervention	Drug intervention
• Talking down • Distraction • Seclusion	Antipsychotic (with sedative properties) orally e.g. chlorpromazine 50 – 100mgs (liquid/tabs) haloperidol 5 – 10mgms Low-distribution benzodiazepine e.g. lorazepam 1 – 2mgms

Review 30 –60 minutes
If no response – repeat
Or
If no response or no initial co-operation

• Talking down • Distraction • Seclusion • Monitor physically	(Sedative) antipsychotic IM e.g. haloperidol 5-10mgs olanzapine 10mgms chlorpromazine 50-100mgms (with care in frail elderly or drug naïve) Low-distribution benzodiazepine e.g. lorazepam 1-2mgms

Revise treatment plan – start/increase baseline antipsychotic
Review 30 –60 minutes
If no response – repeat (with higher dose ranges, vital signs permitting)
Revise treatment plan

ESTABLISHED EMERGENCY

Non-drug intervention	Drug intervention
• Seclusion • Talking down • Monitor vital signs	sedative antipsychotic IM (as above) in adequate dose ± Low distribution benzodiazepine as above, administered separately (depending on severity of incident) Or zuclopentixol acetate (Acuphase) IM * High potency antipsychotic IV e.g. haloperidol 5 – 10mgms

Review 30 – 60 minutes
Revise treatment plan
Repeat if necessary (vital signs permitting)

* should be used with care in the frail or drug naïve

Fig. 4.3.8.2 Outline plan for treatment/management of psychiatric emergencies.

Adjunctive medications

Simplifying any complexities that may have entered into treatment is a useful strategy when response is poor, such as reducing or stopping anticholinergics or other drugs, such as antidepressants, where possible. For kinetic (and possibly dynamic) reasons, such drugs may be acting against the primary therapeutic aim.

While a number of other drugs have been recommended for adjunctive treatment of suboptimally responding schizophrenia, there is inadequate evidence to support any of them.

Lithium, independent of its actions on mood, is not 'antipsychotic'[92] but has been recommended for patients with schizoaffective disorders. The evidence is inconclusive and its use is best considered empirical in those with prominent affective symptoms.[93] There is some evidence of a more rapid improvement in symptoms with *valproate* augmentation but any benefits seem to be transient.[94] This is more than can be said for *carbamazepine*, whose widespread adjunctive use comes with no supporting evidence, though such studies that have been done have been small.[95] A recent Cochrane review of the role of *lamotrigine* did provide tentative support from small, poor quality studies, suggesting that PANSS total and positive and negative subscale scores significantly decline on lamotrigine compared to placebo.[96] Further work is required before clear recommendations can be made.

At one time, *benzodiazepines* were advocated in both the sole and adjunctive treatment of schizophrenia, usually in high doses

(~100 mg/day diazepam). Benefits beyond simple sedation have been hard to find and systematic reviews highlight a small number of small, supporting studies.[97] While, as noted, benzodiazepines are widely used for treating psychotically mediated acute behavioural disturbance, such evidence as there is suggests little difference between them and antipsychotics.[97]

Finally, it is worth remembering that the theory behind the introduction of ECT, however flawed, related to *schizophrenia* and there is some evidence to support its use in this condition, especially when rapid global improvement is required and when antipsychotic response has been limited.[98] However, as in depression, its benefits are usually short-lived and do not substitute for an effective long-term medication strategy.

When the patient satisfies criteria for 'treatment resistance', the evidence is, as noted, overwhelming that *clozapine* is better than any other drug regime.[41] There is *no* evidence that any other antipsychotic, old or new, shares its edge and in this regard, clozapine remains a *unique* compound.

Of the psychosocial interventions, only CBT has been suggested as possibly helpful in modifying symptoms and improving outcomes in those with 'treatment resistance'[99] though further work is required to confirm these tentative findings.

Concluding remarks

The above might be interpreted as inferring that nothing much has changed in the treatment and management of schizophrenia, which remains a somewhat pessimistic, even unrewarding, area of therapeutic endeavour: one that is regressing far less moving on. This is far from our experience and the opposite of the impression we wish to create.

Certainly, as far as drug treatments are concerned, no single agent or type of agent now seems more satisfactory across the board than any other, but the challenge to 'atypicality' as a valid subgrouping of antipsychotics does not limit options—rather it *broadens* them, restoring to the treatment repertoire the wide range of choices that is the key to individualized care planning. With psychosocial interventions, there does remain more work to be done in *proving* absolute efficacy and/or effectiveness and the relative place of each, but in service development and care planning, risk:benefit appraisal is sophisticated enough to encompass what has qualitative *value* as well as what is quantitatively *proven*.

As in all branches of medicine where chronic and relapsing disease is encountered, restoring order on chaos, fostering engagement and lighting a way forward when none may be obvious are for the highest levels of skill, in which evidence-based practice can provide the direction but not yet the specific path. We are fortunate in now having available to us the greatest *ever* range of interventions to bring to the care of those who suffer from this most complex and fell disorder. None is comprehensive, all have limitations, but if we wish to provide quality care, care that accounts for the multifarious manifestations patients present, it is our duty to apply not only the experience of others but of ourselves too.

There is no 'quick fix' in gaining competence in the treatment and management of schizophrenia—and, as yet, no curative 'holy grail' either. But there is, more than ever, the opportunity for clinicians to demonstrate *real* expertise in moulding the range of therapeutics now at our disposal. If that is not reason for medical optimism, what is!

Further information

There is at present a sense of 'flux' in care recommendations for those who suffer from schizophrenia and the authors would recommend that some of the articles below should be *read* as opposed to just 'referred to'. Both the American[2] and the Australian[3] guidelines are good examples of present trends in this approach, though illustrate the slight differences of emphasis that even 'evidence' permits. The CATIE study[72] is seminal and mandatory reading for all those involved in treatment planning (the background to this is presented in *Schizophrenia Bulletin* (2003, vol. 29(1))—the CUtLASS study likewise.[71] Both mark a change from simply efficacy-based studies to pragmatic, 'everyday' designs that is likely to intensify. For a background to the importance of absolute doses of standard comparator drugs in efficacy trials and dose equivalence issues, the studies of Rosenheck *et al.*[69] and Strakowski *et al.*[70] are informative, as is that of Baldessarini *et al.*[74] from an earlier period. Crow's original paper on the Type 1/Type 2 dichotomy[12] is historically important (in the 1980s/1990s, the most cited source of research-testable hypotheses in psychiatry) and helpful in understanding what was being proposed and the limitations of subsequent efforts to disprove it.

The authors would contend that behind much of the confusion in the clinical psychopharmacology of the antipsychotic drugs in recent years lurks the long shadow of parkinsonism. This, rather than tardive dyskinesia, has always been the extrapyramidal side-effects issue, partly because of its pervasive presence but also because of its wide and ill-defined boundaries. Those wishing to familiarize themselves with more than just the basics are referred to the book by Owens,[14] one of the few texts to specifically present drug-induced parkinsonism in both its subjective and objective components for primarily a psychiatric audience.

Finally, all practicing psychiatrists nowadays should be familiar with the address of the Cochrane Library (www.thecochranelibrary.com) and its Database of Systematic Reviews and should feel comfortable referring to it themselves, not just second-hand. It presents quality syntheses of complex material of relevance to many key areas of psychiatric practice.

References

1. National Institute for Clinical Excellence. (2002). *Schizophrenia: core interventions in the treatment and management of schizophrenia in primary and secondary care*. NICE, London.
2. Lehman, A.F., Lieberman, J.A., Dixon, L.B., *et al.* (2004). Practice guideline for the treatment of patients with schizophrenia (2nd edn). *The American Journal of Psychiatry*, **161**(Suppl. 2), 1–56.
3. McGorry, P. (2005). Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for the treatment of schizophrenia and related disorders. *The Australian and New Zealand Journal of Psychiatry*, **39**, 1–30.
4. Cole, J.O. and the NIMH Psychopharmacology Service Center Collaborative Study Group. (1964). Phenothiazine treatment in acute schizophrenia. *Archives of General Psychiatry*, **10**, 246–61.
5. Davis, J.M. and Garver, D.L. (1978). Neuroleptics—clinical use in psychiatry. In *Handbook of psychopharmacology* (eds. L.L. Iversen, S. Iversen, and S. Snyder), pp. 129–64. Plenum Press, New York.
6. Tuma, A.H. and May, P.R.A. (1979). And if it doesn't work, what next? A study of treatment failures in schizophrenia. *The Journal of Nervous and Mental Disease*, **167**, 566–71.

7. Agid, O., Kapur, S., Arenovich, T., *et al.* (2003). Delayed-onset hypothesis of antipsychotic action: a hypothesis tested and rejected. *Archives of General Psychiatry*, **60**, 1228–35.
8. Johnstone, E.C., Crow, T.J., Frith, C.D., *et al.* (1978). Mechanism of the antipsychotic effect in the treatment of acute schizophrenia. *Lancet*, **i**, 848–51.
9. Johnstone, E.C. (1979). The clinical implications of dopamine receptor blockade in acute schizophrenia. In *Proceedings of an international symposium on neuroleptics and schizophrenia* (ed. J.M. Simister), pp. 21–8. Lundbeck, Luton.
10. May, P.R.A., Tuma, A.H., Yale, C., *et al.* (1976). Schizophrenia: a follow-up study of the results of five forms of treatment. II. Hospital stay over two to five years. *Archives of General Psychiatry*, **33**, 481–6.
11. May, P.R.A., Tuma, A.H., Yale, C., *et al.* (1981). Schizophrenia: a follow-up study of the results of five forms of treatment. *Archives of General Psychiatry*, **38**, 776–84.
12. Crow, T.J. (1980). Molecular pathology of schizophrenia: more than one disease process. *British Medical Journal*, **280**, 66–8.
13. Erhart, S.M., Marder, S.R., and Carpenter, W.T. (2006). Treatment of schizophrenia negative symptoms: future prospects. *Schizophrenia Bulletin*, **32**, 234–371.
14. Cunningham Owens, D.G. (1999). *A guide to the extrapyramidal side-effects of antipsychotic drugs*. Cambridge University Press, Cambridge.
15. Carpenter, W.T., Heinrichs, D.W., and Alphs, L.D. (1985). Treatment of negative symptoms. *Schizophrenia Bulletin*, **11**, 440–52.
16. Carpenter, W.T., Heinrichs, D.W., and Wagman, A.M.I. (1988). Deficit and non-deficit forms of schizophrenia: the concept. *The American Journal of Psychiatry*, **145**, 578–83.
17. Breier, A., Buchanan, R.W., Kirkpatrick, B., *et al.* (1994). Effects of clozapine on positive and negative symptoms in outpatients with schizophrenia. *The American Journal of Psychiatry*, **151**, 20–6.
18. Owens, D.G.C. and Johnstone, E.C. (2006). Precursors and prodromata of schizophrenia: findings from the Edinburgh high risk study and their literature context. *Psychological Medicine*, **36**, 1501–14.
19. King, D.J. (1990). The effects of neuroleptics on cognitive and psychomotor function. *The British Journal of Psychiatry*, **157**, 799–811.
20. King, D.J. (1994). Psychomotor impairment and cognitive disturbances induced by neuroleptics. *Acta Psychiatrica Scandinavica*, **89**(Suppl. 380), 53–8.
21. Harvey, P.D. and McClure, M.M. (2006). Pharmacological approaches to the management of cognitive dysfunction in schizophrenia. *Drugs*, **66**, 1465–73.
22. Harvey, P.D., Patterson, T.L., Potter, L.S., *et al.* (2006). Improvement in social competence with short-term atypical antipsychotic treatment: a randomised, double-blind comparison of quetiapine versus risperidone for social competence, social cognition, and neuropsychological functioning. *The American Journal of Psychiatry*, **163**, 1918–25.
23. Ferreri, F., Agbokou, C., and Gauthier, S. (2006). Cognitive dysfunctions in schizophrenia: potential benefits of cholinesterase inhibitor adjunctive therapy. *Journal of Psychiatry & Neuroscience*, **31**, 369–76.
24. Salminen, J.K., Lehtonen, V., Allonen, H., *et al.* (1980). Sulpiride in depression: plasma levels and effects. *Current Therapeutic Research*, **27**, 109–15.
25. Margakis, B.P. (1990). A double-blind comparison of oral amitriptyline and low-dose intramuscular flupenthixol decanoate in depressive illness. *Current medical Research and Opinion*, **12**, 51–7.
26. Knights, A. and Hirsch, S.R. (1981). 'Revealed' depression and drug treatment for schizophrenia. *Archives of General Psychiatry*, **38**, 806–11.
27. de Alarcon, R. and Carney, M.W.P. (1969). Severe depressive mood changes following slow-release intramuscular fluphenazine injection. *British Medical Journal*, **3**, 564–7.
28. Van Putten, T. and May, P.R.A. (1978). 'Akinetic depression' in schizophrenia. *Archives of General Psychiatry*, **35**, 1101–7.
29. Klein, D. and Davis, J.M. (1969). *Diagnosis and drug treatment of psychiatric disorders*. Williams and Wilkins, Baltimore, MD.
30. Owens, D.G.C. and Johnstone, E.C. (1980). The disabilities of chronic schizophrenia: their nature and factors contributing to their development. *The British Journal of Psychiatry*, **136**, 384–95.
31. Miller, P.M., Johnstone, E.C., Lang, F.H., *et al.* (2000). Differences between patients with schizophrenia within and without a high security psychiatric hospital. *Acta Psychiatrica Scandinavica*, **102**, 12–18.
32. Janicak, P.G., Davis, J.M., Preskorn, S., *et al.* (1993). *Principles and practice of psychopharmacology*. Williams and Wilkins, Baltimore, MD.
33. Johnson, D.A.W. (1979). Further observations on the duration of depot neuroleptic maintenance therapy in schizophrenia. *The British Journal of Psychiatry*, **135**, 524–30.
34. Morgan, R. and Cheadle, J. (1974). Maintenance treatment of chronic schizophrenia with neuroleptic drugs. *Acta Psychiatrica Scandinavica*, **50**, 78–85.
35. Davis, J.M., Shaffer, C.B., Killan, G.A., *et al.* (1980). Important issues in the drug treatment of schizophrenia. *Schizophrenia Bulletin*, **6**, 70–87.
36. Davis, J.M., Metalon, L., Watanabe, M., *et al.* (1993). Depot antipsychotic drugs: place in therapy. *Drugs*, **47**, 741–73.
37. Kane, J.M., Eerdekens, M., Lindenmayer, J.-P., *et al.* (2003). Long-acting injectable risperidone: efficacy and safety of the first long-acting atypical antipsychotic. *The American Journal of Psychiatry*, **160**, 1125–32.
38. Kane, J.M., Rifkin, A., Woerner, M., *et al.* (1983). Low dose neuroleptic treatment of outpatient schizophrenics. *Archives of General Psychiatry*, **40**, 893–6.
39. Baldessarini, R.J., Cohen, B.M., and Teicher, M.H. (1990). Pharmacological treatment. In *Schizophrenia—treatment of acute episodes* (eds. S.T. Levy and P.T. Ninan), pp. 61–118. American Psychiatric Press, Washington, DC.
40. Hirsch, S.R. (1986). Clinical treatment of schizophrenia. In *The psychopharmacology and treatment of schizophrenia* (eds. P.B. Bradley and S.R. Hirsch), pp. 286–339. Oxford University Press, Oxford.
41. Kane, J.M., Honigfeld, G., Singer, J., Meltzer, H.Y., and the Clozaril Collaborative Study Group. (1988). Clozapine for the treatment-resistant schizophrenic: a double-blind comparison with chlorpromazine. *Archives of General Psychiatry*, **45**, 789–96.
42. Wahlbeck, K., Cheine, M.V., and Essali, A. (1999). Clozapine versus typical neuroleptic medication for schizophrenia. *Cochrane Database of Systematic Reviews*, Issue 4. Art. No.: CD000059. DOI: 10.1002/14651858.CD000059.
43. Kingdon, D. (2006). Psychological and social interventions for schizophrenia: robust evidence supports a wide range, including cognitive therapy. *British Medical Journal*, **333**, 212–13.
44. Jones, C., Cormac, I., Silveira de Mota Neto, J.I., *et al.* (2004). Cognitive behaviour therapy for schizophrenia. *Cochrane Database of Systematic Reviews*, Issue 4. Art. No.: CD000524. DOI: 10.1002/14651858.CD000524.pub2.
45. McKenna, P.J. (2006). Cognitive behaviour therapy is not effective. *British Medical Journal*, **333**, 353.
46. Baldwin, S.A., Murray, D.M., and Shadish, W.R. (2005). Empirically supported treatments or type 1 errors? Problems with the analysis of data from group-administered treatments. *Journal of Consulting and Clinical Psychology*, **73**, 924–35.
47. Kemp, R., Kirov, G., Everitt, B., *et al.* (1998). Randomised controlled trial of compliance therapy: 18 month follow-up. *The British Journal of Psychiatry*, **172**, 413–19.
48. McIntosh, A.M., Conlon, L., Lawrie, S.M., *et al.* (2006). Compliance therapy for schizophrenia. *Cochrane Database of Systematic Reviews*, Issue 3. Art. No.: CD003442. DOI: 10.1002/14651858.CD003442.pub2.
49. Rathod, S., Kingdon, D., Smith, P., and Turkington, D. (2005). Insight into schizophrenia: the effects of cognitive behavioural therapy on the components of insight and association with sociodemographics—data on a previously published randomised controlled trial. *Schizophrenia Research*, **74**, 211–19.

50. Carrol, A., Fattah, S., Clyde, Z., *et al.* (1999). Correlates of insight and change in schizophrenia. *Schizophrenia Research*, **35**, 247–53.
51. Hayes, R.L. and McGrath, J.J. (2000). Cognitive rehabilitation for people with schizophrenia and related conditions. *Cochrane Database for Systematic Reviews*, (3), CD000968. DOI: 10.1002/14651858. CD000968.
52. Malmberg, L. and Fenton, M. (2001). Individual psychodynamic psychotherapy and psychoanalysis for schizophrenia and severe mental illness. *Cochrane Database of Systematic Reviews*, Issue 3. Art. No.: CD001360. DOI: 10.1002/14651858.CD001360.
53. Robertson, L., Connaughton, J., and Nicol, M. (1998). Life skills programmes for chronic mental illnesses. *Cochrane Database of Systematic Reviews*, Issue 3. Art. No.: CD000381. DOI: 10.1002/14651858.CD000381.
54. Eckman, T.A., Wirshing, W.C., Marder, S.R., *et al.* (1992). Technique for training schizophrenic patients in illness self-management: a controlled trial. *The American Journal of Psychiatry*, **149**, 1549–55.
55. Linszen, D., Dingemans, P., Van der Does, J.W., *et al.* (1996). Treatment, expressed emotion and relapse in recent onset schizophrenic disorders. *Psychological Medicine*, **26**, 333–42.
56. Crowther, R., Marshall, M., Bond, G., *et al.* (2001). Vocational rehabilitation for people with severe mental illness. *Cochrane Database of Systematic Reviews*, Issue 2. Art. No.: CD003080; DOI: 10.1002/14651858.CD003080.
57. Pekkala, E. and Merinder, L. (2002). Psychoeducation for schizophrenia. *Cochrane Database of Systematic Reviews*, Issue 2. Art. No.: CD002831.DOI: 10.1002/14651858.CD002831.
58. Leff, J., Kuipers, L., Berkowitz, R., *et al.* (1982). A controlled trial of social intervention in the families of schizophrenic patients. *The British Journal of Psychiatry*, **141**, 121–34.
59. Pharoah, F., Mari, J., Rathbone, J., *et al.* (2006). Family intervention for schizophrenia. *Cochrane Database of Systematic Reviews*, Issue 4. Art. No. CD000088. DOI: 10.1002/14651858.CD000088.pub2.
60. Tyrer, P., Coid, J., Simmonds, S., *et al.* (1998). Community mental health teams (CHMTs) for people with severe mental illnesses and disordered personality. *Cochrane Database of Systematic Reviews*, Issue 4. Art. No. CD000270. DOI: 10.1002/14651858. CD000270.
61. Marks, I.M., Connolly, J., Muijen, M., *et al.* (1994). Home-based versus hospital-based care for people with serious mental illness. *The British Journal of Psychiatry*, **164**, 179–94.
62. Grawe, R.W., Falloon, I.R.H., Widen, J.H., *et al.* (2006). Two years of continued early treatment for recent-onset schizophrenia: a randomised controlled study. *Acta Psychiatrica Scandinavica*, **114**, 328–36.
63. Audini, B., Marks, I.M., Lawrence, R.E., Connolly, J., and Watts, V. (1994). Home-based versus out-patient/in-patient care for people with serious mental illness. Phase II of a controlled study. *The British Journal of Psychiatry*, **164**, 204–10.
64. Harrison, J., Alam, N., and Marshall, J. (2001). Home or away: which patients are suitable for a psychiatric home treatment service. *Psychiatric Bulletin*, **25**, 310–13.
65. Fitzgerald, P. and Kulkarni, J. (1998). Home-oriented management programme for people with early psychosis. *The British Journal of Psychiatry*, **172**(Suppl. 33), 39–44.
66. Marshall, M., Gray, A., Lockwood, A., *et al.* (1998). Case management for people with severe mental disorders. *Cochrane Database of Systemic Reviews*, Issue 2. Art. No.: CD000050. DOI: 10.1002/14651858.CD000050.
67. Marshall, M. and Lockwood, A. (1998). Assertive community treatment for people with severe mental disorders. *Cochrane Database of Systematic Reviews*, Issue 2. Art. No.: CD001089. DOI: 10.1002/14651858.CD001089.
68. Rosen, A. and Teesson, M. (2001). Does case management work? The evidence and the abuse of evidence-based medicine. *The Australian and New Zealand Journal of Psychiatry*, **35**, 731–46.
69. Rosenheck, R.A., Perlick, D., Bingham, S., *et al.* (2003). Effectiveness and cost of olanzapine and haloperidol in the treatment of schizophrenia: a randomised controlled trial. *The Journal of the American Medical Association*, **290**, 2693–702.
70. Strakowski, S.M., Johnson, J.L., DelBello, M.P., *et al.* (2005). Quality of life during treatment with haloperidol or olanzapine in the year following a first psychotic episode. *Schizophrenia Research*, **78**, 161–9.
71. Jones, P.B., Barnes, T.R.E., Davis, L., et al. (2006). Randomised controlled trial of the effect on quality of life of second-vs-first generation antipsychotic drugs in schizophrenia: Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1). *Archives of General Psychiatry*, **63**, 1079–87.
72. Lieberman, J.A., Stroup, T.S., McEvoy, J.P., *et al.* (2005). Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *The New England Journal of Medicine*, **353**, 1209–23.
73. Rosenheck, R.A., Leslie, D.L., Sindelar, J., *et al.* (2006). Cost-effectiveness of second-generation antipsychotics and perphenazine in a randomised trial of treatment for chronic schizophrenia. *The American Journal of Psychiatry*, **163**, 2080–9.
74. Baldessarini, R.J., Katz, B., and Cotton, P. (1984). Dissimilar dosing with high-potency and low-potency neuroleptics. *The American Journal of Psychiatry*, **141**, 748–52.
75. Oosthuizen, P., Emsley, R.A., Turner, J., *et al.* (2001). Determining the optimal dose of haloperidol in first-episode psychosis. *Journal of Psychopharmacology*, **15**, 251–5.
76. Kapur, S., Zipursky, R., Roy, P., *et al.* (1997). The relationship between D2 receptor occupancy and plasma levels on low dose oral haloperidol: a PET study. *Psychopharmacology*, **131**, 148–52.
77. Taylor, D., Kerwin, R., and Paton, C. (eds.) (2005). *The Maudsley 2005–2006 prescribing guidelines*. Taylor and Francis, London.
78. Whicher, E., Morrison, M., and Douglas-Hall, P. (2002). 'As required' medication regimes for seriously mentally ill people in hospital. *Cochrane Database of Systemic Reviews*, Issue 1. Art. No.: CD003441. DOI: 10.1002/14651858. CD003441.
79. Lieberman, J., Jody, D., Geisler, S., *et al.* (1993). Time course and biologic correlates of treatment response in first-episode schizophrenia. *Archives of General Psychiatry*, **50**, 369–76.
80. Bishop, M.P., Gallant, D.M., and Sykes, T.F. (1965). Extrapyramidal side effects and therapeutic response. *Archives of General Psychiatry*, **13**, 155–61.
81. Johnstone, E.C., Macmillan, F.J., Frith, C.D., *et al.* (1990). Further investigations of the predictors of outcome following first schizophrenic episodes. *The British Journal of Psychiatry*, **157**, 182–9.
82. Birchwood, M., Iqbal, Z., Chadwick, P., *et al.* (2000). Cognitive approach to depression and suicidal thinking in psychosis. I. Ontogeny of post-psychotic depression. *The British Journal of Psychiatry*, **177**, 516–21.
83. Siris, S.H., Morgan, V., Fagerstrom, R., *et al.* (1987). Adjunctive imipramine in the treatment of post-psychotic depression: a con trolled study. *Archives of General Psychiatry*, **44**, 533–9.
84. Siris, S.H., Pollack, S., Bermanzohn, P., *et al.* (2000). Adjunctive imipramine for a broader group of post-psychotic depressions in schizophrenia. *Schizophrenia Research*, **44**, 187–92.
85. McGlashan, T.H. (1999). Duration of untreated psychosis in first episode schizophrenia: marker or determinant of course. *Biological Psychiatry*, **46**, 899–907.
86. Lieberman, J.A. (1999). Is schizophrenia a neurodegenerative disorder? A clinical and neurobiological perspective. *Biological Psychiatry*, **46**, 729–39.
87. Marshall, M., Lockwood, A., Lewis, S., *et al.* (2004). Essential elements of an early intervention service for psychosis: the opinions of expert clinicians. *BMC Psychiatry*, **4**.
88. Marshall, M., Lewis, S., Lockwood, A., *et al.* (2005). Association between duration of untreated psychosis and outcome in cohorts of first-episode patients: a systematic review. *Archives of General Psychiatry*, **62**, 975–83.

89. Perkins, D.O., Gu, H., Boteva, K., *et al.* (2005). Relationship between duration of untreated psychosis and outcome in first-episode schizophrenia: a critical review and meta-analysis. *The American Journal of Psychiatry*, **162**, 1785–804.
90. Owens, D.G.C., Miller, P., Lawrie, S.M., *et al.* (2005). Pathogenesis of schizophrenia: a psychopathological perspective. *The British Journal of Psychiatry*, **186**, 386–93.
91. Stroup, T.S., Lieberman, J.A., McEvoy, J.P., *et al.* (2006). Effectiveness of olanzapine, quetiapine, risperidone and ziprasidone in patients with chronic schizophrenia following discontinuation of a previous atypical antipsychotic. *The American Journal of Psychiatry*, **163**, 611–22.
92. Johnstone, E.C., Crow, T.J., Frith, C.D., *et al.* (1988). The Northwick Park 'functional' psychosis study: diagnosis and treatment response. *Lancet*, **ii**, 119–26.
93. Leucht, S., McGrath, J., and Kissling, W. (2003). Lithium for schizophrenia. *Cochrane Database of Systematic Reviews*, Issue 3. Art. No.: CD003834. DOI: 10.1002/14651858.CD003834.
94. Basan, A. and Leucht, S. (2003). Valproate for schizophrenia. *Cochrane Database of Systematic Reviews*, Issue 3. Art. No.: CD004028. DOI: 10.1002/14651858.CD004028.pub2.
95. Leucht, S., McGrath, J., White, P., *et al.* (2002). Carbamazepine for schizophrenia and schizoaffective psychoses. *Cochrane Database of Systematic Reviews*, Issue 3. Art. No.: CD001258. DOI: 10.1002/14651858.CD001258.
96. Premkumar, T.S. and Pick, J. (2006). Lamotrigine for schizophrenia. *Cochrane Database of Systematic Reviews*, Issue 4. Art. No.: CD005962. DOI: 10.1002/14651858.CD005962.pub2.
97. Volz, A., Khorsand, V., Gillies, D., *et al.* (2007). Benzodiazepines for schizophrenia. *Cochrane Database of Systematic Reviews*, Issue 1. Art. No.: CD006391. DOI: 10.1002/14651858.CD006391.
98. Tharyan, P. and Adams, C.E. (2005). Electroconvulsive therapy for schizophrenia. *Cochrane Database of Systematic Reviews*, Issue 2. Art. No.: CD000076. DOI: 10.1002/14651858.CD000076.pub2.
99. Pilling, S., Bebbington, P., Kuipers, E., *et al.* (2002). Psychological treatments in schizophrenia. I. Meta-analysis of family intervention and cognitive behaviour therapy. *Psychological Medicine*, **32**, 763–82.

4.3.9 Schizoaffective and schizotypal disorders

Ming T. Tsuang, William S. Stone, and Stephen V. Faraone

Introduction

This chapter focuses on two disorders in the schizophrenia 'spectrum': schizoaffective disorder and schizotypal personality disorder. The emphasis includes the clinical features, classification, diagnosis, epidemiology, aetiology, course, prognosis, and possibilities for prevention for each disorder. Some aspects will be underscored to reflect controversial issues, such as the heterogeneity apparent in each condition. Such issues relate to the accurate classification of the disorders, which is important for at least two reasons. First, it is essential to develop reliable and valid diagnostic criteria in order to study the aetiology of the disorders and then utilize that knowledge to develop rational and testable intervention strategies. Heterogeneity adds variance to the process that reduces both the reliability of diagnosis and also the statistical power of experimental designs to detect intervention/treatment effects. Second, the development of newer generations of psychopharmacological treatments holds the promise of matching more appropriate and efficacious medications with specific syndromes or types of symptoms. This trend underscores the importance of differential diagnosis in determining what treatment a patient will receive. Heterogeneity within a diagnostic category complicates achievement of this goal. Another area to be emphasized involves the goal of early interventions, in addition to palliative treatments for these disorders. In contrast, other areas such as the genetic aetiology of schizoaffective disorder and schizotypal personality disorder, and treatments for schizoaffective disorder, will receive less emphasis here, to avoid redundancies with other chapters in this volume. Each disorder will be considered separately, starting with a review of schizoaffective disorder, the more severe of the two spectrum conditions.

Schizoaffective disorder

Clinical features

Schizoaffective disorder afflicts patients having schizophrenic and affective symptoms. Either they have affective symptoms of sufficient severity and chronicity to exclude an uncomplicated diagnosis of schizophrenia, or they show features of schizophrenia that are sufficient to exclude an uncomplicated diagnosis of an affective disorder.[1] These types of symptoms may or may not occur simultaneously, which underscores the importance of viewing the course of the illness longitudinally in addition to its cross-sectional presentation. Symptom clusters that are primarily affective or primarily schizophrenic predominate at different times.

Compared to patients with schizophrenia, patients with schizoaffective disorder often (though not always) demonstrate relatively high levels of premorbid function,[2,3] but nevertheless show significant premorbid weaknesses in multiple cognitive and clinical functions.[4] Patients with schizoaffective disorder also tend to show more identifiable precipitating events. The nature of the precipitating stressor may vary widely; for example it may be physical (e.g. recently giving birth or experiencing a head injury) or interpersonal (e.g. change in an important relationship). The clinical course of the disorder is often characterized by a periodic, rapid onset of symptoms that shows a relatively high degree of remission after several weeks or months. As Vaillant pointed out in the 1960s, many of these patients 'recover' completely after an episode, and resume their lives at premorbid levels of function.[5] As will be noted further below, the clinical features of some cases of schizoaffective disorder mainly resemble those of schizophrenia, while the features of other cases are more similar to those of bipolar disorder. Regardless of the subtype or variant of the disorder, however, the mortality rate is of special concern. Rates of death due mainly to suicide or accident show elevations in this disorder that are similar to those observed in schizophrenia and in major affective disorders.[6]

In general, schizoaffective disorder is more common in females than in males.[3] The age of onset varies, but tends to be younger than that of unipolar or bipolar disorder. Tsuang *et al.* found the median age of onset for schizoaffective disorder was 29 years, which was significantly lower than groups with bipolar or unipolar affective disorder, but similar to a group with schizophrenia.

Marneros *et al.*[2] also reported that a median age of onset of 29 years for schizoaffective disorder was lower than the median age for groups with affective disorders (35 years), but reported that it was higher than a group with schizophrenia (24 years). In contrast, Reichenberg *et al.*[4] reported no differences in the age of first hospitalization between patients with schizophrenia, schizoaffective disorder, or non-psychotic bipolar disorder. These differences between studies reflect differences in both the diagnostic criteria employed, and the heterogeneity of the disorder.

Classification

The classification of schizoaffective disorder has always been controversial. Kraepelin reported in 1919 that patients with both affective and schizophrenic symptoms complicated the differential diagnosis due to the 'mingling of morbid symptoms of both psychoses'. Kasanin first employed the term 'acute schizophrenic psychoses' in 1933 to describe a group of patients who experienced a rapid onset of emotional turmoil and psychotic symptoms, but who recovered after several weeks or months.[3] In other words, the symptoms appeared similar to schizophrenia during periods of exacerbation, but unlike schizophrenia, they showed a greater tendency to remit between episodes. These features sparked an ongoing debate by the 1960s about the proper classification of schizoaffective disorder. Much of this discussion involved the following proposals:

1 It was a type of schizophrenia (e.g. 'remitting schizophrenia');
2 It was a type of affective disorder;
3 It was a unique disorder that was separate from both schizophrenia and bipolar disorder;
4 It reflected an arbitrary categorization of clinical symptoms that masked a continuum of pathology between schizophrenia and affective illness;
5 It contained a heterogeneous collection of 'interforms' between schizophrenia and affective disorder (i.e. symptoms of both disorders).

The last possibility is not mutually exclusive of the first four; for example, one or more variants of schizoaffective disorder may be related closely to schizophrenia, while another may be related more closely to an affective disorder.

The puzzle has yet to be solved. Family and outcome studies provide useful ways of assessing the relative merits of each of the possibilities outlined above. These approaches are informative and will be reviewed below, although interpretations of such studies are complicated at times by the use of different diagnostic criteria across investigations.

(a) Family studies

Family studies provide an important tool for assessing the relationship between disorders. They are a type of genetic study that assumes that related disorders will co-aggregate more frequently among biologically related individuals than they would in the general population. Thus, a disorder is more likely to be in the schizophrenia spectrum if it occurs more frequently among the relatives of schizophrenic patients, compared with suitable controls. Similarly, a disorder is more likely to be in the affective spectrum if it occurs more frequently among the relatives of patients with affective disorders. Evidence for the inclusion of schizoaffective disorder in the schizophrenia spectrum is discussed in greater detail elsewhere (see Chapter 4.3.6.1). Only representative findings pertinent to the present discussion about the classification of schizoaffective disorder will be summarized here.

Bertelsen and Gottesman[7] summarized a series of seven family studies published between 1979 and 1993, using structured diagnostic criteria. Analyses of risk to the development of schizophrenia, schizoaffective disorder, and affective disorder in the first-degree relatives of patients with schizoaffective disorder, were included. In all seven studies, the relatives showed a higher risk of developing an affective disorder than of developing schizoaffective disorder. In five of the seven studies the risks of developing schizophrenia was equal to or greater than the risk of developing schizoaffective disorder. Thus, the relatives of schizoaffective patients showed generally higher risks of developing disorders other than the one with which they were diagnosed. These findings were consistent with a heterogeneous view of schizoaffective disorder, in which individual cases represented subtypes of either schizophrenia or of affective disorder. The findings were also consistent with the possibility that schizoaffective disorder represents a chance collection of 'interforms' between schizophrenia and affective disorder.

These findings were not consistent with the view that schizoaffective disorder represented a continuum between the other two disorders, because in that case, the rate of schizoaffective disorder in first-degree relatives would have been higher, compared with the rates at which these relatives developed schizophrenia or affective disorder. The findings were also inconsistent with the possibility that schizoaffective disorder represented a unique disorder that was independent of either schizophrenia or an affective disorder. In that case, the first-degree relatives of patients with schizoaffective disorder should show relatively high rates of schizoaffective disorder itself, but relatively low rates of the other disorders. In the series of studies reviewed by Bertelsen and Gottesman,[7] the morbid risk for schizoaffective disorder itself ranged from 1.8 to 6.1 per cent in first-degree relatives of patients with schizoaffective disorder, which was still higher than the rate observed in the general population (see the section on epidemiology below). These results, taken together with the higher risks for both schizophrenia and affective disorder, suggest that schizoaffective disorder is a heterogeneous condition. Recent reviews of family studies, including those that considered depressed (i.e. unipolar) and bipolar subtypes, have also underscored both the heterogeneity of schizoaffective disorder, and the controversial nature of its classification.[8,9]

(b) Outcome studies

A majority of outcome studies show that schizoaffective disorder has a better course than schizophrenia, but a poorer course than affective disorder.[10–12] For example, Tsuang and colleagues reviewed 10 outcome studies reported between 1963 and 1987 that assessed patients with either schizoaffective disorder or schizophrenia.[10] Global, marital, social, occupational, hospital course, and symptom dimensions of outcome were measured. In each category, patients with schizophrenia showed poorer outcomes. In contrast, their review of 11 outcome studies comparing schizoaffective disorder with affective disorder showed that affective disorder was associated with equal or better outcomes on almost all dimensions. Thus, despite differences in methodology and diagnostic criteria,

schizoaffective disorder was frequently associated with clinical outcomes that were intermediate between those associated with schizophrenia and those related to affective disorder.

Other researchers reported similar findings. Kendler *et al.*, for example, showed intermediate levels of clinical impairment for schizoaffective disorder in an epidemiological family study.[13] Marneros *et al.* reported on outcomes as part of the Cologne Longitudinal study, using modified DSM-III-R diagnoses.[14] The outcomes were measured by symptoms in five dimensions (psychotic symptoms, reduction of energetic potential, qualitative and quantitative disturbances of affect, and other disturbances of behaviour) that persisted for at least 3 years. Consistent with the pattern described thus far, poor outcomes in the schizoaffective group occurred at a rate (49.5 per cent of the sample) that was intermediate between those observed in the schizophrenic (93.2 per cent) and affective groups (35.8 per cent), and differed significantly from both of them. In a more recent study, Jäger *et al.* studied 241 patients at the time of their first hospitalization, and then again 15 years later.[15] Similar to these other examples, schizoaffective subjects presented a clinical picture that was less impaired than the one shown by schizophrenic subjects, but more impaired than the one shown by affective subjects.

While these studies show schizoaffective disorder to have intermediate outcomes generally, there are categories in which it resembles schizophrenia or affective disorder more closely. For example, Samson *et al.*[10] and Reinares *et al.*[12] noted that outcomes for schizoaffective disorder were equivalent to those for affective disorder in several dimensions. Marneros *et al.*, showed that 70 per cent of a schizoaffective group was rated as good or excellent on a measure of social adjustment, which did not differ significantly from 84 per cent of an affective group who received the same rating.[12] Both groups differed significantly from a schizophrenic group, however, in which only 44 per cent of the group demonstrated good or excellent outcomes. Moreover, the schizoaffective and affective disorder groups did not differ on a rating scale of psychological impairments (e.g. body language, affect display, conversation skills, and cooperation), although both were rated as significantly less impaired than the schizophrenic group.

Other studies, however, such as Kendler *et al.*[13] reported similarities between some types of psychotic symptoms in schizoaffective disorder and schizophrenia, including the severity of delusions and positive thought disorder, and the frequency of hallucinations. Each of these groups showed higher levels of these symptoms than an affective disorders group. Hizdon *et al.* showed recently that individuals with schizoaffective disorder did not differ from individuals with schizophrenia on basic cognitive measures of executive function, memory, and processing speed, although the schizoaffective group did perform better on measures of social cognition.[16] Reichenberg *et al.* showed that individuals with schizophrenia and schizoaffective disorder who were assessed premorbidly performed similar to each other but lower than individuals who later developed non-psychotic bipolar disorder, on tests of non-verbal and verbal intellectual function, and on tests of basic reading and reading comprehension.[4]

These overall differences in outcome serve to validate the classification of schizoaffective disorder as a separate syndrome further. Its heterogeneity, however, raises the issue of whether such intermediate outcomes might reflect the mean of a combination of mainly good and mainly poor outcomes. This in turn leads to the question of whether schizoaffective disorder can be subtyped in a useful and valid manner. If so, are better and worse outcomes associated with different variants of the syndrome?

Vaillant suggested in the 1960s that prognostic indicators, including a good premorbid level of adjustment, the presence of precipitating factors, an acute onset, confusion, the presence of affective symptoms, and a familial history of affective disorder (or the absence of a schizophrenic history), could predict remission in approximately 80 per cent of cases of 'remitting schizophrenia'.[17] The inclusion of affective symptoms and a positive family history for affective illness on the list contributed (later) to hypotheses that variants of schizoaffective disorder were related to affective illness and to better outcomes. In contrast, variants associated more with schizophrenic symptoms or family history were associated more with schizophrenia and with relatively poor outcomes.[18]

There have been a variety of attempts to subtype schizoaffective disorders, based on whether affective or schizophrenic symptoms predominate. The validity of many of these attempts, however, is inconclusive. Bertelsen and Gottesman noted, for example, that at best, relatives of individuals with affective type schizoaffective disorder, or schizophrenic type schizoaffective disorder, showed only trends towards higher rates of affective disorder or schizophrenia, respectively.[7] Similarly, Kendler *et al.* did not detect different rates of schizophrenia or affective illness in first-degree relatives of patients with schizoaffective disorder when the patients were subtyped into bipolar and depressive subgroups.[13] Moreover, the subtypes did not predict differences in outcomes.

Conversely, a latent class analysis of psychotic patients from the Roscommon study showed that most cases of DSM-III-R schizoaffective disorder were categorized in either a bipolar schizomania class ($n = 19$), or in a schizodepression class ($n = 13$), rather than in schizophrenia ($n = 1$), major depression ($n = 0$), schizophreniform ($n = 3$), or hebephrenia ($n = 3$) classes.[19] Moreover, Reinares *et al.* reviewed evidence showing that bipolar and depressive subtypes differed from each other in ways consistent with differences between bipolar and unipolar affective disorders.[12] For example, the bipolar schizoaffective subtype was associated with more total episodes, more episodes with shorter periods and cycles, and higher frequency of cycles. Higher numbers of cycles were associated with poorer long-term outcomes. Taken together, these studies show at least some recent support for the subtyping of schizoaffective disorder into mainly affective and mainly schizophrenic variants.

Other factors associated with poor outcomes include poor inter-episode recoveries,[13] persistent psychotic symptoms in the absence of affective features, poor premorbid social adjustment, chronicity, a higher number of schizophrenia-like symptoms,[20] and the presence of schizoaffective mixed states.[12]

Diagnosis and differential diagnosis

The DSM-IV diagnosis of schizoaffective disorder[1] is listed in the category of 'schizophrenia and other psychotic disorders'. The major feature of the disorder is that, in addition to meeting the clinical criteria for schizophrenia (criterion A), an individual must also experience a major depressive, manic, or mixed episode concurrently. In addition, in the same period of illness, a patient must experience symptoms of psychosis (hallucinations and/or

delusions) for a period of at least 2 weeks, in the absence of mood-related symptoms (criterion B). Nevertheless, affective symptoms must comprise a substantial portion of total duration of the illness (criterion C), and symptoms may not be attributable to either substance use or to a major medical condition (criterion D). Two subtypes of the disorder, including bipolar type and depressive type, may be diagnosed.

The criteria for schizoaffective disorder in ICD-10 are similar to those in DSM-IV. The essential requirement is that prominent symptoms of affective disorder and prominent symptoms of schizophrenia are present together for at least 2 weeks. Depressive, manic, and mixed subtypes are recognized.

The differential diagnosis includes, most prominently, either schizophrenia or affective disorder, which may be differentiated in part by consideration of the longitudinal criteria (criteria B and C), in addition to the cross-sectional criteria (criterion A). The presence of conditions relating to general medication and substance use should also be considered in the differential diagnosis.

Epidemiology

The epidemiological status of schizoaffective disorder is somewhat uncertain compared with schizophrenia, largely because of dilemmas related to the diagnosis and classification of the disorder. To help in the standardization of data from different studies, representative incidence and prevalence estimates will be emphasized from recent investigations that utilized research diagnostic, DSM-III-R or DSM-IV criteria.

(a) Incidence

Earlier studies showed that new cases of 'schizomanic' patients (i.e. manic patients who also demonstrated schizophrenic or paranoid symptoms) numbered approximately 1.7 per 100 000 per year.[20] This was less than the 4 per 100 000 per year shown by 'schizodepressive' patients. The number of schizoaffective cases in this study exceeded the number of manic patients, and made up half of the number of schizophrenic cases. Since then, Tien and Eaton analysed data from the Epidemiologic Catchment Area study for three non-overlapping groups with psychotic symptoms.[21] One of these groups comprised individuals with 'psychotic affective syndrome', which was similar to schizoaffective disorder except that most members of the group (59 per cent) demonstrated psychotic symptoms only in conjunction with a mood disturbance (essentially DSM-III-R mood disturbance with psychotic symptoms). The incidence of this disorder was 1.7 per 1000 per year, which was approximately equal to the rate for schizophrenia (2.0 per 1000 per year). Even if the 59 per cent of the group who met the criteria for a mood disorder with psychotic features was excluded, the remaining 41 per cent would still comprise a higher incidence rate than that detected by earlier studies. Differences in sampling procedures (treated versus non-treated samples) may have contributed to the differences observed in the rates. More importantly, however, these studies showed that schizoaffective disorder occurred at 50 to 85 per cent of the rate of schizophrenia, thus confirming that patients with this disorder comprise a clinically significant population. One current but long-standing issue involves questions about the temporal stability of incidence rates in schizophrenia-related disorders, as reflected by reports of both increases and decreases.

(b) Prevalence

Until recently, prevalence estimates for schizoaffective disorder relied mainly on samples that were treated in clinics or other psychiatric settings. Because a variety of factors influence the decision to enter and remain in treatment, the estimates varied substantially. For example, Okasha reviewed studies that reported rates varying between 2 and 29 per cent.[8] A recent epidemiological study in Finland using 8028 people who were at least 30 years old showed a lifetime prevalence rate of 0.32 per cent for schizoaffective disorder (compared to 0.87 per cent for schizophrenia), which accounted for 10.5 per cent of all psychotic disorders.[22] This is a lower estimate than many earlier studies reported, and likely results from a combination of factors (including a narrowing of diagnostic criteria, and increased utilization of multiple sources of information such as case notes and registers, in addition to interview data) that together have improved diagnostic accuracy. Prevalence estimates of putative schizoaffective subtypes remain subject to the same inconsistencies of diagnosis and selection factors that affect schizoaffective disorder itself. Not surprisingly, there is little consensus about whether manic or schizophrenic subtypes predominate (see also Tsuang *et al.*[20]).

(c) Review of evidence

Treatments for schizoaffective disorder are the same as those for schizophrenia and affective disorders alone. As the nature and efficacy of those treatments are discussed elsewhere, they will not be considered here. Rather, this section will focus on management issues related to the need to treat symptoms of both disorders simultaneously, or sequentially.

Management

The authors have found it useful to consider psychopharmacological treatment in terms of putative subtypes, including affective type schizoaffective disorder and schizophrenic type schizoaffective disorder.

Treatment of schizoaffective disorder, affective subtype, will include antipsychotic medication (e.g. clozapine, risperidone quetiapine, ziprasidone, or olanzepine), particularly if psychotic symptoms are present. In addition, antidepressants, mood stabilizers (e.g. lithium), or anticonvulsants (e.g. valproate or carbamazepine) may be useful with this group. It will be necessary in such cases to weigh the potential risks of such medications, such as elevated toxicity, against the potential benefits.

In schizoaffective disorder, schizophrenic subtype, combination treatments may also be more effective than a single treatment. We find, however, that antipsychotic treatments alone may be more efficient in many cases. This is particularly true if affective symptoms (i.e. depression) are largely secondary to the experience of having a psychotic condition, and its attendant interpersonal, social, and financial difficulties. In these cases, remediation of the psychotic symptoms may also have the effect of easing the affective problems. For other cases, which include more of a treatment-refractory depression, antipsychotic medication may be augmented with lithium (or another mood stabilizer) or antidepressant medication. Moreover, electroconvulsive therapy may reduce mortality rates in schizoaffective patients.

The authors note that it may be difficult at times to distinguish the affective subtype from the schizophrenic subtype, especially in the presence of florid psychotic symptoms. In these cases,

treatment decisions may rest on the presenting symptoms of the patient. Treatment during intermorbid periods is in part dependent on the presence or absence of psychotic symptoms. As noted above, psychotic episodes in this period are associated with relatively poorer outcomes, and are likely to require chronic antipsychotic therapy.

Schizotypal personality disorder

Clinical features

Like schizoaffective disorder, schizotypal personality disorder is a complex and chronic condition that includes some, but not all, of the features of schizoaffective disorder and schizophrenia. Most notably, persistent psychosis is not part of the syndrome, although mild forms of thought disorder may occur, such as magical thinking or ideas of reference (as opposed to delusions of reference, which indicate psychosis). Moreover, brief episodes of psychosis may occur in times of stress, but will not persist.

Schizotypal patients show pervasive deficits in social and interpersonal traits. They often demonstrate aloofness, poor eye contact, affective constriction, and suspiciousness. Consequently, close interpersonal relationships are either avoided, or cause discomfort and anxiety. These individuals usually have few friends. Not surprisingly, schizotypal patients are often deficient in accurately sensing social cues or affective signals from others. Although they can interact with people when necessary, they often prefer not to, and do not become more comfortable in social situations with time.

Schizotypal patients may also show magical thinking, ideas of reference, unusual perceptions (e.g. sensing the presence of another person, or that people are talking about them), and/or perceptual illusions (e.g. often perceiving a dimly lit lamp-post as a person). Both their social deficits and these cognitive–perceptual problems contribute to an overall impression of oddness. However, this feature may occur independently of other clinical symptoms,[23] and manifest itself in odd speech or unusual appearance. The oddness or eccentricities evident in these patients are often ego syntonic (i.e. they are not experienced as problems). Moreover, schizotypal patients show deficits in attention, long-term verbal memory, and executive functions. These deficits are qualitatively similar to those seen in schizophrenia (and schizoaffective disorder), but like many other clinical manifestations of this disorder, they are quantitatively milder.

Like schizophrenia, schizotypal personality disorder is often evident by early adulthood, but schizotypal traits may be evident in late childhood or adolescence. Once it appears, the disorder tends to show a chronic course, but one that includes periodic exacerbations and attenuations of symptoms. A recent study that followed individuals with schizotypal personality disorder for 2 years showed that paranoid thoughts and unusual perceptual experiences were among the most stable and least malleable DSM-IV symptoms, while the most changeable were odd behaviours and restricted affect.[24] The former symptoms were thus more trait-like, and the latter were more intermittent. Consistent with these findings, the same group also showed that in the course of 2 years (with treatment), 61 per cent of schizotypal patients no longer met DSM-IV diagnostic criteria for the disorder.[25] With a more stringent definition of improvement (12 months with two or less symptoms meeting criteria), the rate of remission dropped to 23 per cent. These studies show that both the severity and the expression of the disorder vary over time and probably, as a function of treatment.

Classification

In contrast to the controversy surrounding the classification of schizoaffective disorder, family, twin, and adoption studies clearly support the view that schizotypal personality disorder is best classified in the schizophrenia spectrum.[26] Nevertheless, it is a complex and chronic disorder that in all likelihood, is also heterogeneous. Kendler pointed out that this heterogeneity was at least partly related to the two primary methods used to study the disorder.[27] One of these involves the 'clinical method', which identifies patients with mild forms of schizophrenic or psychotic-like symptoms. This type of patient, for example, is often characterized by relatively high levels of positive psychiatric symptoms (e.g. magical thinking and perceptual distortions).

In contrast, the 'family research method' identifies relatives of patients with schizophrenia who have subtle, schizophrenia-like symptoms. Features associated more with familial than with clinical schizotypal personality disorder include a predominance of negative symptoms (e.g. social withdrawal and impairment, and higher levels of anxiety and poor rapport), cognitive impairments (e.g. impaired language comprehension, eye-tracking, and attentional dysfunctions), and elevated rates of schizophrenia and related disorders in family members.[26] Thaker *et al*. reported that familial and clinical schizotypal personality disorders were similar on measures of physical or social anhedonia,[28] and that some neuropsychological deficits were also associated with both groups.[26]

The concept of familial schizotypal disorder is particularly important because it may share a common genetic basis with schizophrenia. Paul Meehl first proposed the term 'schizotaxia' to describe the genetic vulnerability to schizophrenia, and suggested that individuals with schizotaxia would eventually develop either schizotypal personality disorder or schizophrenia, depending on the protection or liability afforded by environmental circumstances. As the concept evolved, Meehl reformulated it to allow for the possibility that some people with schizotaxia would develop neither schizophrenia nor schizotypal personality disorder. In fact, evidence now shows that the clinical symptoms observed in many non-psychotic, first-degree relatives of people with schizophrenia are similar to those observed in familial schizotypal personality disorder.[26] Psychiatric features in such relatives frequently include an aggregation of negative symptoms that are qualitatively similar to, but milder than, those often cited in schizophrenia.[29] Positive symptoms, however, are usually less evident in these relatives than they are in schizophrenia or in schizotypal personality disorder. Neuropsychological impairments in biological relatives of people with schizophrenia are also qualitatively similar to, but milder than, those seen in people with schizophrenia.[26] In particular, these neuropsychological deficits frequently include problems in working memory/attention, long-term verbal memory, and concept formation/abstraction.

Faraone *et al*. recently suggested a reformulation of Meehl's concept of schizotaxia that focuses on these features of negative symptoms and neuropsychological deficits.[26] Unlike schizotypal personality disorder, which occurs in less than 10 per cent of the adult relatives of patients diagnosed with schizophrenia, the basic

symptoms of schizotaxia occur in 20 to 50 per cent of adult relatives, suggesting further that the genetic liability to schizophrenia does not lead inevitably to schizophrenia, schizotypal personality disorder, or schizoid personality disorder.

Diagnosis and differential diagnosis

The DSM-IV criteria for schizotypal personality disorder include a 'pervasive pattern of social deficits' and 'cognitive or perceptual distortions' and behavioural 'eccentricities' (criterion A).[1] At least five of nine specific symptoms (e.g. ideas of reference, constricted affect, odd behaviour, or appearance) must be present to satisfy this criterion. These symptoms must occur by early adulthood. They must not occur exclusively during the course of four other conditions, including schizophrenia, a mood disorder with psychotic features, any other psychotic disorder, or a pervasive developmental disorder (criterion B).

The differential diagnosis includes a variety of other disorders. A key difference between schizotypal personality disorder and schizophrenia, a psychotic mood disorder, or another psychotic condition involves the transient nature of psychotic symptoms in schizotypal personality disorder. It may be distinguished from developmental communication disorders by a lack of compensatory means (e.g. gestures) of communicating, and it may be distinguished from autistic or Asperger's disorders by the relatively greater deficits in social awareness and frequent presence of stereotyped behaviours in those syndromes. Schizotypal personality disorder may be confused with several other personality disorders, but can be distinguished from them. In particular, it differs from schizoid personality disorder by its pattern of cognitive–perceptual distortions, and by the odd appearance or behaviour shown frequently by schizotypal patients. The pattern of schizotypal symptoms also differs from that manifested in borderline personality disorder, although there are similarities between these conditions. Schizotypal personality disorder differs from borderline personality disorder, however, in that psychotic-like symptoms and social isolation are more likely to persist in the absence of affective turmoil, and schizotypal individuals are less likely to display the impulsive and manipulative traits that are often associated with borderline personality disorder.

Epidemiology

(a) Incidence

To the authors' knowledge, there continue to be no published incidence studies for schizotypal personality disorder.

(b) Prevalence

A review by Lyons showed that prevalence rates for schizotypal personality disorder in non-clinical samples ranged from 0.7 to 5.1 per cent, with a median near 3.0 per cent.[30] Higher rates occurred in clinical samples—2.0 to 64.0 per cent, with a median of 17.5 per cent. More recently, Torgersen *et al.* reported a rate of 0.6 per cent for DSM-III-R in a community sample, which is lower than the rates in studies reviewed by Lyons.[31] Similar to recent prevalence rates reported for schizoaffective disorder (described above), more recent studies have tended to show lower rates than earlier studies. In contrast to non-clinical samples, the prevalence of schizotypal personality disorder among the relatives of schizophrenic individuals is as high as 10 per cent.[32]

Treatment

(a) Review of evidence

There is, unfortunately, a dearth of outcome studies involving psychotherapy, psychosocial, or psychopharmacological treatments for schizotypal personality disorder. Older published studies often show methodological limitations (e.g. small samples, subjects with mixed diagnoses, inadequate controls, and problems with internal validity), or provide outcome data on only limited aspects of the disorder. Despite these caveats, it is clear that few treatment gains are evident in earlier studies. This is particularly true of studies that utilized psychodynamically oriented psychotherapy, either alone or in combination with other treatments (e.g. group therapy or art therapy) as the primary treatment modality.[33] Recent evidence for the efficacy of psychotherapy for personality disorders is more promising, but is limited mainly to other personality disorders.[34]

Several earlier studies investigated the usefulness of medications in treating schizotypal personality disorder, although they typically employed small numbers of subjects, combined samples of schizotypal and borderline personality disorders, and showed little clinical improvement.[33] Typical antipsychotic drugs, in particular, were proposed to reduce positive symptoms or depressed mood in times of acute stress, but the high incidence of adverse side effects discouraged their widespread use at other times, including the more chronic stable (i.e. non-crisis) phases of the disorder. Other types of medication, including fluoxetine, have shown generally non-specific effects of treatment.

Hymowitz *et al.* administered a low dose of haloperidol, a first-generation antipsychotic medication, to 17 outpatients with DSM-III diagnoses of schizotypal personality disorder, for 6 weeks.[35] The initial dose of 2.0 mg was intended to rise to 12.0 mg, but side effects prevented increases beyond a mean dose of 3.6 mg. Even with lower doses, 50 per cent of the sample withdrew from the study because of side effects. The 17 subjects who completed 2 weeks of the protocol improved somewhat in ratings of ideas of reference, odd communications, social isolation, and overall functioning.

More recently, Koenigsberg *et al.* employed a double-blind protocol to administer low doses of risperidone (0.25 mg/day—2.0 mg/day), a second-generation antipsychotic medication, to 25 patients with DSM-IV schizotypal personality disorder, for 9 weeks.[36] Compared to a placebo control group, patients who received risperidone demonstrated significant reductions in positive and negative symptoms, with no difference in dropout rates between groups. These findings are encouraging and consistent with evidence described above that schizotypal symptoms are amenable to change.[24,25] Hopefully, findings like this will stimulate additional research into pharmacological treatments for this disorder.

Management

Patients with schizotypal personality disorder often view their worlds as odd and threatening places and may require extended courses of treatment. Although trust and rapport with the therapist are often difficult to establish in schizotypal personality disorder, the therapeutic relationship may be used to mitigate the marked deficits in interpersonal relationships that characterize this syndrome.[37] The frequent occurrence of paranoia and suspiciousness, together with social aloofness and constricted affect, may

make exploratory psychotherapeutic approaches less effective than supportive cognitive behavioural therapies. In fact, these patients may only seek treatment to alleviate circumscribed problems, like anxiety or somatic complaints. Approaches that emphasize concrete, interim goals, and stipulate explicit means of attaining them, thus have the best chances of success. Because individuals with this disorder are vulnerable to decompensation during times of stress and may experience transient episodes of psychosis, they may also benefit from techniques to facilitate stress reduction (e.g. relaxation techniques, exercise, yoga, and meditation). Fortunately, some people with schizotypal features are likely to seek treatment in times of stress.[38] In the short-term, brief courses of antipsychotic treatment may be useful if symptoms of psychosis appear.

Cognitive problems are also frequently amenable to concrete goal-oriented approaches to treatment. Patients benefit from understanding their cognitive strengths and weaknesses because it helps them confront and cope with long-standing difficulties in their lives. For example, problems in attention, verbal memory, or organizational skills contribute to failures in educational, occupational, and social endeavours, while reinforcing negative self-images and increasing performance anxiety. Knowledge of circumscribed cognitive problems allows patients to reframe their difficulties in a more positive manner, and facilitate selection of realistic personal, educational, and occupational goals. Moreover, specific cognitive deficits are often subject to partial remediation. For example, standard procedures will attenuate deficits in the acquisition, organization, and retrieval of new information (e.g. writing information down in a 'memory notebook', using appointment books or planners, and rehearsing new information, among others). In some instances, the documentation of specific cognitive deficits (e.g. attention) can lead to academic accommodations in school (e.g. more time to take exams), which will help individuals function closer to their intellectual potentials.

Possibilities for prevention

At present, most early intervention programmes involve secondary prevention, which includes the early identification and treatment of clinical (usually psychotic or psychotic-like) symptoms. While intervention is necessary to alleviate clinical symptoms at any point during the disorder, it is particularly important early on because it might alter the course of the illness. Patients treated with antipsychotic medication during their first or second hospital admission, for example, show better outcomes than those who are not treated until later in the course of their disorders.

Primary prevention, which involves treatment before the disorder manifests itself clinically, is not yet available for schizoaffective disorder, schizotypal personality disorder, or other disorders in the schizophrenia spectrum. To develop such treatments, it will be necessary to predict who is most likely to develop a disorder. There are a few encouraging approaches, including ongoing 'high-risk' studies that follow the offspring of schizophrenic parents longitudinally.[39] Such studies help to identify traits early in life that predict which individuals are most likely to experience emergent clinical symptoms in adulthood. This type of study is particularly important because it can facilitate the formation of homogeneous high-risk groups, which in turn can facilitate the development of focused prevention strategies.

With our current knowledge, it is difficult to justify preventive treatments—especially medication—for people without symptoms. The authors have argued elsewhere, however, that if people in high-risk groups (like first-degree biological relatives of patients with schizophrenia) show clinically meaningful symptoms that can be organized into valid liability syndromes, then intervention attempts may become appropriate.[26] The authors proposed this course of action for people with 'schizotaxia', and suggested preliminary research guidelines.[40]

Eventually, prevention will be a primary therapeutic approach for the treatment of disorders in the schizophrenia spectrum. While primary prevention has yet to occur, the authors are optimistic that high-risk studies, progress in secondary prevention, and progress in discovering the genetic aetiology of the schizophrenia spectrum, will facilitate primary prevention strategies eventually.

Further information

Raine, A. (2006). Schizotypal personality: neurodevelopmental and psychosocial trajectories. *Annual Review of Clinical Psychology*, **2**, 291–326.

Evans, J.D., Heaton, R.K., Paulsen, J.S., McAdams, L.A., Heaton, S.C., and Jeste, D.V. (1999). Schizoaffective disorder: a form of schizophrenia or affective disorder? *The Journal of Clinical Psychiatry*, **60**, 874–82.

Tsuang, M.T., Stone, W.S., and Lyons, M.J. (2007). Toward prevention of schizophrenia: early detection and intervention. In *Recognition and prevention of major mental and substance abuse disorders* (eds. M.T. Tsuang, W.S. Stone, and M.J. Lyons), pp. 213–37. American Psychiatric Publishing, Inc, Washington DC.

References

1. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (*DSM-IV*). American Psychiatric Association, Washington, DC.
2. Marneros, A., Deister, A., and Rohde, A. (1990). Sociodemographic and premorbid features of schizophrenic, schizoaffective and affective psychoses. In *Affective and schizoaffective disorders* (eds. A. Marneros and M.T. Tsuang), pp. 130–45. Springer-Verlag, Berlin.
3. Tsuang, M.T., Simpson, S.J.C., and Fleming, J.A. (1986). Diagnostic criteria for subtyping schizoaffective disorder. In *Schizoaffective psychoses* (eds. A. Marneros and M.T. Tsuang), pp. 50–62. Springer-Verlag, Berlin.
4. Reichenberg, A., Weiser, M., Rabinowitz, J., *et al.* (2002). A population-based cohort study of premorbid intellectual, language and behavioral functioning in patients with schizophrenia, schizoaffective disorder, and nonpsychotic bipolar disorder. *The American Journal of Psychiatry*, **159**, 2027–35.
5. Vaillant, G. (1962). The prediction of recovery in schizophrenia. *The Journal of Nervous and Mental Disease*, **135**, 534–43.
6. Simpson, J. (1988). Mortality studies in schizophrenia. In *Nosology, epidemiology and genetics of schizophrenia*, Vol. 3 (eds. M.T. Tsuang and J.C. Simpson), pp. 245–73. Oxford, Elsevier Science Publishers B.V., Amsterdam, New York.
7. Bertelsen, A. and Gottesman, I.I. (1995). Schizoaffective psychoses: genetical clues to classification. *American Journal of Medical Genetics* (*Neuropsychiatric Genetics*), **60**, 7–11.
8. Okasha, A. (2007). The concept of schizoaffective disorder: utility versus validity and reliability—a transcultural perspective. In *The overlap of affective and schizophrenic spectra* (eds. A. Marneros and A.H. Akiskal), pp. 104–32. Cambridge University Press, Cambridge.
9. Coryell, W. (2007). Phenomenological approaches to the schizoaffective spectrum. In *The overlap of affective and schizophrenic spectra* (eds. A. Marneros and A.H. Akiskal), pp. 133–44. Cambridge University Press, Cambridge.

10. Samson, J.A., Simpson, J.C., and Tsuang, M.T. (1988). Outcome of schizoaffective disorders. *Schizophrenia Bulletin*, **14**, 543–54.
11. Angst, J. (1986). The course of schizoaffective disorders. In *Schizoaffective psychoses* (eds. A. Marneros and M.T. Tsuang), pp. 63–93. Springer-Verlag, Berlin.
12. Reinares, M., Vieta, E., Benabarre, A., *et al.* (2007). Clinical course of schizoaffective disorders. In *The overlap of affective and schizophrenic spectra* (eds. A. Marneros and A.H. Akiskal), pp. 145–55. Cambridge University Press, Cambridge.
13. Kendler, K.S., McGuire, M., Gruenberg, A.M., *et al.* (1995). Examining the validity of DSM-III-R schizoaffective disorder and its putative subtypes in the Roscommon Family Study. *The American Journal of Psychiatry*, **152**, 755–64.
14. Marneros, A., Rohde, A., and Deister, A. (1998). Frequency and phenomenology of persisting alterations in affective, schizoaffective and schizophrenia disorders: a comparison. *Psychopathology*, **31**, 23–8.
15. Jager, M., Bottlender, R., Strauss, A., *et al.* (2004). Fifteen-year follow-up of ICD-schizoaffective disorders compared to schizophrenia and affective disorders. *Acta Psychiatrica Scandinavica*, **109**, 30–7.
16. Fiszdon, J.M., Richardson, R., Greig, T., *et al.* (2007). A comparison of basic and social cognition between schizophrenia and schizoaffective disorder. *Schizophrenia Research*, **91**, 117–21.
17. Vaillant, G. (1964). Prospective prediciton of schizophrenic remission. *Archives of General Psychiatry*, **11**, 509–18.
18. Levitt, J.J. and Tsuang, M.T. (1988). The heterogeneity of schizoaffective disorder: implications for treatment. *The American Journal of Psychiatry*, **145**, 926–36.
19. Kendler, K.S., Karkowski, L.M., and Walsh, D. (1998). The structure of psychosis: latent class analysis of probands from the Roscommon family study. *Archives of General Psychiatry*, **55**, 492–9.
20. Tsuang, M.T., Levitt, J.J., and Simpson, J.C. (1995). Schizoaffective disorder. In *Schizophrenia* (eds. S.R. Hirsch and D.R. Weinberger). Blackwell Scientific, Oxford.
21. Tien, A.Y. and Eaton, W.W. (1992). Psychopathologic precursors and sociodemographic risk factors for the schizophrenia syndrome. *Archives of General Psychiatry*, **49**, 37–46.
22. Perala, J., Suvisaari, J., Saarni, S.I., *et al.* (2007). Lifetime prevalence of psychotic and bipolar I disorders in a general population. *Archives of General Psychiatry*, **64**, 19–28.
23. Battaglia, M., Cavallini, M.C., Macciardi, F., *et al.* (1997). The structure of DSM-III-R schizotypal personality disorder diagnosed by direct interviews. *Schizophrenia Bulletin*, **23**, 83–92.
24. McGlashan, T.H., Grilo, C.M., Sanislow, C.A., *et al.* (2005). Two-year prevalence and stability of individual DSM-IV criteria for schizotypal, borderline, avoidant, and obsessive-compulsive personality disorders: toward a hybrid model of axis II disorders. *The American Journal of Psychiatry*, **162**, 883–9.
25. Grilo, C.M., Sanislow, C.A., Gunderson, J.G., *et al.* (2004). Two-year stability and change of schizotypal, borderline, avoidant and obsessive-compulsive personality disorders. *Journal of Consulting and Clinical Psychology*, **72**, 767–75.
26. Faraone, S.V., Green, A.I., Seidman, L.J., *et al.* (2001). Schizotaxia: clinical implications and new directions for research. *Schizophrenia Bulletin*, **27**, 1–18.
27. Kendler, K.S. (1985). Diagnostic approaches to schizotypal personality disorder: a historical perspective. *Schizophrenia Bulletin*, **11**, 538–53.
28. Thaker, G., Moran, M., Adami, H., *et al.* (1993). Psychosis proneness scales in schizophrenia spectrum personality disorders: familial vs. nonfamilial samples. *Psychiatry Research*, **46**, 47–57.
29. Tsuang, M.T., Gilbertson, M.W., and Faraone, S.V. (1991). Genetic transmission of negative and positive symptoms in the biological relatives of schizophrenics. In *Positive vs. negative schizophrenia* (eds. A. Marneros, M.T. Tsuang, and N. Andreasen), pp. 265–91. Springer-Verlag, New York.
30. Torgersen, S., Kringlen, E., and Cramer, V. (2001). The prevalence of personality disorders in a community sample. *Archives of General Psychiatry*, **58**, 590–6.
31. Lyons, M.J. (1995). Epidemiology of personality disorders. In *Textbook in psychiatric epidemiology* (eds. M.T. Tsuang, M. Tohen, and G.E. Zahner), pp. 407–36. Wiley-Liss, New York.
32. Battaglia, M. and Torgersen, S. (1996). Schizotypal disorder: at the crossroads of genetics and nosology. *Acta Psychiatrica Scandinavica*, **94**, 303–10.
33. Tsuang, M.T., Stone, W.S., and Faraone, S.V. (1998). Overview for treatment for schizotypal and schizoid personality disorders: present and future. *NOOS Aggiornamenti in Psichiatria*, **4**, 201–15.
34. Verheul, R. and Herbrink, M. (2007). The efficacy of various modalities of psychotherapy for personality disorders: a systematic review of the evidence and clinical recommendations. *International Review of Psychiatry*, **19**, 25–38.
35. Hymowitz, P., Frances, A., Jacobsberg, L.B., *et al.* (1986). Neuroleptic treatment of schizotypal personality disorders. *Comprehensive Psychiatry*, **27**, 267–71.
36. Koenigsberg, H.W., Reynolds, D., Goodman, M., *et al.* (2003). Risperidone in the treatment of schizotypal personality disorder. *The Journal of Clinical Psychiatry*, **64**, 628–34.
37. Bender, S. (2005). The therapeutic alliance in the treatment of personality disorders. *Journal of Psychiatric Practice*, **11**, 73–87.
38. Poreh, A. and Whitman, D. (1993). MMPI-2 schizophrenia spectrum profiles among schizotypal college students and college students who seek psychological treatment. *Psychological Reports*, **73**, 987–94.
39. Stone, W.S., Faraone, S.V., Seidman, L.J., *et al.* (2005). Searching for the liability to schizophrenia: concepts and methods underlying genetic high-risk studies of adolescents. *Journal of Child and Adolescent Psychopharmacology*, **15**, 403–17.
40. Tsuang, M.T., Stone, W.S., Seidman, L.J., *et al.* (1999). Treatment of nonpsychotic relatives of patients with schizophrenia: four case studies. *Biological Psychiatry*, **41**, 1412–18.

4.3.10 Acute and transient psychotic disorders

J. Garrabé and F.-R. Cousin

Historic introduction

Acute and transient psychotic episodes have been described since the end of the nineteenth century. Descriptions have varied from one country to another, so that the exact nosology has not yet been established. The links between acute psychoses (generally defined as having brief obvious psychotic symptomatology) and chronic psychoses (schizophrenic psychoses and psychoses with persistent delusions) are still under discussion.

For instance, Sections F20 and F21 in ICD-10[(1)] are devoted to 'Schizophrenia, schizotypal and delusional disorders'. A specific diagnostic category named 'Acute and transient psychotic disorders' is included, distinct from Schizophrenia (F20), Schizotypal disorder (F21), Persistent delusional disorder (F22), Induced delusional disorder (also called *folie à deux*) (F24), and Schizoaffective disorder (F25).

In this textbook, the acute and transient psychotic disorders (Chapter 4.3.10) appear in the section dedicated to schizophrenia,

which also includes schizotypal disorders and schizoaffective disorders (Chapter 4.3.9). However, this section is clearly distinguished from the chapter in which the persistent delusional disorders are discussed (Chapter 4.4). These taxonomic divergences are justified more by the history of acute psychoses than by scientific findings.

In the nineteenth century, German psychiatrists had already described *akute primäre Verruckheit*,[2] termed *paranoia acuta* by Karl Westphal. In 1876 (published in 1878), Westphal used this term to describe an acute form of paranoia with an outburst of perceptual hallucinations, consisting mostly of hallucinatory voices and delusions, with clouding of consciousness. In 1890, Meynert repeated the clinical description but named the condition amentia.[3] Sigmund Freud chose this type of acute delusion with hallucinations for his psychoanalytic conception of psychosis.[4]

In the sixth edition of his textbook, published in 1899, Kraepelin[5] included all the paranoias under dementia praecox, and in the eighth edition (1908–1915) he combined manic and melancholic periodic disorders in a single group, leaving acute psychosis with no place between these two diagnostic categories.

In 1911, Bleuler[6] replaced the single disease dementia praecox by the concept of a group of schizophrenias of various clinical forms. He noticed that schizophrenia often began with an acute excitatory episode lasting from a few hours to a few years. He described a wide variation of outcome of acute forms of psychosis, but he separated acute schizophrenias from simple schizophrenia as he believed that acute forms do not necessarily end in deterioration.

In 1916, based on Karl Jaspers' psychopathology, the Danish psychiatrist Wimmer[7] described psychogenic psychosis as a reactive psychosis arising after psychosocial trauma. Mayer-Gross,[8] who proposed an organic aetiology for schizophrenia, described 'oneiroid states' consisting of acute psychotic symptomatology with no other specific organic features.

In 1961, Leonhard[9] used Kleist's concept of marginal psychosis (*Randpsychosen*) to develop his description of 'cycloid psychoses' as endogenous psychoses separate from schizophrenic psychoses and from manic and melancholic psychoses. These cycloid psychoses tend to have a benign and periodic course.

Earlier (1933), Kasanin[10] had described 'acute schizoaffective psychoses', raising questions about the links between schizophrenic and affective diseases.

Langfeldt[11] suggested that observation for 5 years was required to be able to distinguish schizophrenia and what he called schizophreniform psychosis. This long-term observation is a reminder of the Bleulerian concept of acute schizophrenias which could last for several years. Epidemiological studies have led to the presence in modern classifications of a group of acute schizophreniform psychoses under the rubric 'Schizophreniform disorder' (DSM-IV Section 295.40) or 'Acute psychotic disorder schizophrenic-like' (ICD-10 Section F23.2).

In France, the concept of *bouffée délirante* led naturally to a specific class of acute psychoses. In 1895, Magnan[12] and his disciple Legrain[13] described *bouffée délirante* or *délire d'emblée* (immediate delusion) within the polymorphic delusions of the chronic insane. This concept is based on Morel's theory of degeneration, commonly accepted in the nineteenth century. The question of whether there is a susceptibility or a predisposition to the occurrence of an acute psychosis remains unanswered.[14]

In 1954, Ey[15,16] described the development of the concept of *bouffées délirantes* and of acute psychoses with hallucinations from the time of Magnan to a symposium devoted to the clinical subdivision of schizophrenic psychoses held at the First World Congress of Psychiatry in 1950, where the various ideas were discussed by Langfeldt, Karl Leonhard, and Aubrey Lewis (Table 4.3.10.1).

Table 4.3.10.1 Historical development of the terminology of acute and transient psychotic disorders

Historic term	Current terminology
1876 Westphal *Akute primäre Verruckheit* *Paranoia acuta*	F23.3 Other acute predominantly delusional psychotic disorder
1890 Meynert Amentia	
1895 Magnan and Legrain *Bouffées délirantes*	F23.0 Acute polymorphic psychotic disorder without symptoms of schizophrenia
1899 Kraepelin Dementia praecox	F20.0 Schizophrenia
1909–1913 Kraepelin Paranoia	F22.0 Persistent delusional disorder
1911 Bleuler Acute-onset forms of schizophrenia	F23.1 Acute polymorphic psychotic disorder with symptoms of schizophrenia F23.2 Acute schizophrenia-like psychotic disorder
1916 Wimmer Psychogenic psychosis	F23.3 Other acute predominantly psychotic disorder
1924 Mayer-Gross *Oneroide Erlebnisform*	F23.3 Acute schizophrenia-like psychotic disorder
1933 Kasanin Acute schizoaffective psychoses	F25 Schizoaffective disorders
1939 Langfeldt Schizophreniform states	F23.2 Acute schizophrenia-like psychotic disorder
1954 Ey *Bouffées délirantes et psychoses hallucinatoires aigues*	F23.0 Acute polymorphic psychotic disorder without symptoms of schizophrenia
1961 Leonhard Cycloid psychoses	F23.0 Acute polymorphic psychotic disorder without symptoms of schizophrenia

Clinical description: psychopathology

The heterogeneous group of acute and transient psychotic disorders are characterized by three typical features, listed below in descending order of priority:

- suddenness of onset (within 2 weeks or less);
- presence of typical syndromes with polymorphic (changing and variable) or schizophrenic symptoms;
- presence of associated acute stress (stressful events such as bereavement, job loss, psychological trauma, etc.).

The onset of the disorder is manifested by an obvious change to an abnormal psychotic state. This is considered to be abrupt when it occurs within 48 h or less. Abrupt onset often indicates a better outcome. Full recovery occurs within 3 months and often in a shorter time (a few days or weeks). However, a small number of patients develop persistent and disabling states.

The general (G) criteria for these acute disorders in DCR-10 (Diagnostic Criteria Research of ICD) are as follows.

G1 There is acute onset of delusions, hallucinations, incomprehensible or incoherent speech, or any combination of these. The time interval between the first appearance of any psychotic symptoms and the presentation of the fully developed disorder should not exceed 2 weeks.

G2 If transient states of perplexity, misidentification, or impairment of attention and concentration are present, they do not fulfil the criteria for organically caused clouding of consciousness as specified for F05, criterion A.

G3 The disorder does not satisfy the symptomatic criteria for manic episode (F30), depressive episode (F32), or recurrent depressive disorder (F33).

G4 There is insufficient evidence of recent psychoactive substance use to satisfy the criteria for intoxication (F1x.0), harmful use (F1x.1), dependence (F1x.2), or withdrawal states (F1x.3 and F1x.4). The continued moderate and largely unchanged use of alcohol or drugs in the amounts or with the frequency to which the individual is accustomed does not necessarily exclude the use of F23; this must be decided by clinical judgement and the requirements of the research project in question.

G5 There must be no organic mental disorder (F00–F09) or serious metabolic disturbances affecting the central nervous system (this does not include childbirth). (This is the most commonly used exclusion clause.)

A fifth character should be used to specify whether the acute onset of the disorder is associated with acute stress (occurring 2 weeks or less before evidence of first psychotic symptoms):

- F23.x0 without associated acute stress and
- F23.x1 with associated acute stress.

For research purposes it is recommended that change of the disorder from a non-psychotic to a clearly psychotic state is further specified as either abrupt (onset within 48 h) or acute (onset in more than 48 h but less than 2 weeks).

Six categories of acute psychoses are presented in ICD-10, and we shall discuss them in order.

F23.0 acute polymorphic psychotic disorder without symptoms of schizophrenia

The diagnostic criteria are based on the classical symptoms of the true *bouffée délirante* described by Magnan and Legrain.

(a) Suddenness of onset

Bouffée délirante occurs over a period of a few hours or days, usually to young adults and often women in their 30s. The onset of the delirious episode is 'like a thunderbolt in a serene sky'. This aphorism from Legrain has the same meaning as the French classical expression *délire d'emblée* (immediate delusion).

Although premonitory symptoms, such as increasing perplexity and anxiety, may occur, the delusions start suddenly and are always accompanied by a break-up in the individual psychic life. If the onset is preceded by a stressful or traumatic event, such as resettlement or acculturation, this may take place some months previously and the outburst of the delirious episode is delayed. The fifth code character of category F23 is used to specify whether acute stress is associated with the onset of the disorder (e.g. F23.00 has no associated acute stress).

(b) Polymorphic psychotic symptoms

The delusional themes are varied and include grandeur, persecution, influence, possession, body transformation (depersonalization), derealization, or world alteration; these themes change with time and may combine. Other symptoms are also varied, including hallucinations, illusions, interpretations, and intuitions.

(c) The emotional state

As a consequence of the delusions the patient experiences mood change and emotional turmoil (happiness, ecstasy, anxiety, irritability). However, the criteria for manic episode, depressive episode, schizoaffective disorder, and schizophrenia are not satisfied.

Consciousness fluctuates with the delirious convictions and changes of emotion. There is a specific disorientation with respect to time and place—the passage of time (*temps vécu* according to Eugène Minkowski) and 'temporality' (*Sein und Zeit* according to Ludwig Binswanger) are disturbed. This disorientation, described by Karl Jaspers as a 'first delirious experience' (*Erlebnisse*), has to be understood as a dream-like state.

Ey[15,16] differentiated the acute psychoses in terms of the specific alteration in the perception of time rather than their transient course. According to Jacksonian ideas, the acute psychoses are the expression of a destructuring of consciousness to levels related to each acute psychosis.

(d) The duration of the delirious experience

In ICD-10, the criterion of a duration of less than a month distinguishes other categories from schizophrenia (F20) and manic or depressive episodes (F30 and F32).

(e) Short recovery time

In most cases recovery from the acute psychotic disorder occurs within a few weeks or months. However, long-term prognosis is difficult because of the risk of relapse into either a repeated episode or a more chronic disease. If resolution of the symptoms has not occurred after 3 months, the diagnosis should be changed to persistent delusional disorder (F22) or non-organic psychotic disorder (F28).

(f) Suggested criteria

Pull *et al.*[17] have suggested the following empirical criteria for *bouffée délirante*.

- Abrupt or acute onset, with no previous psychiatric disorder except other similar episodes.
- Absence of chronicity: the active stages disappear completely in a few weeks or months. Relapses can occur, but there is no psychiatric disorder between consecutive episodes.
- Specific symptoms: delusions and/or any type of hallucination, depersonalization and/or derealization with or without confusion, and affective disturbance manifesting as depression or euphoria. The symptoms change from day to day and even from hour to hour.
- There is insufficient evidence for organic mental disorder, alcoholism, or drug addiction. The exclusion clauses are less restrictive in ICD-10, since a moderate, continued, and unchanged use of alcohol or drugs in habituated individuals does not exclude the diagnosis.

- The true acute psychotic disorder occurs without any associated psychosocial stress factor. When psychosocial stress factors are found, there is only a temporal link with the so-called 'reactive' acute psychosis.

(g) Long-term evolution

Bleuler[6] described one-third of cases of acute schizophrenic psychoses as single episode, one-third as recurrent episodes with repetition of the same acute and transient psychoses (either manic or depressive), and one-third following a course which developed as schizophrenia.

Between 1976 and 1989, Metzger and Weiber[18] studied 885 cases of acute psychoses. Using the criteria of Pull *et al.*[17] they found group 303 cases of genuine *bouffée délirante* (two-thirds female, one-third male, average age of 32 years). They followed the course of 191 cases (over an average period of 6.2 years): 34 per cent did not relapse, 24 per cent had recurrent and transient episodes, 34 per cent developed schizophrenia, and over 7 per cent developed a periodic affective disorder (manic and depressive states). The relapse or chronic course rate was higher in the group without triggering factors ($n = 92$) than in the group with triggering factors ($n = 99$). The difference between true *bouffée délirante* (no triggering factors) and other acute and transient psychotic disorders raises questions about their pathogenesis.

In the first 2 years, it is essential to distinguish *bouffée délirante* from schizophrenia[19] and other acute psychoses.[20] Follow-up during this period must be very careful.

F23.1 acute polymorphic disorder with symptoms of schizophrenia (*bouffée délirante* or cycloid psychosis with symptoms of schizophrenia)

This diagnostic category combines the symptoms of acute polymorphic psychotic disorder with some typical symptoms of schizophrenia (F20) present for most of the time. However, the schizophrenic symptoms are not precisely listed. F23.1 can be a provisional diagnosis, which is changed to schizophrenia if the criteria of F20 persist more than a month.

Acute polymorphic disorder with symptoms of schizophrenia satisfies the general criteria for acute and transient psychotic disorders:

- acute onset of less than 2 weeks
- polymorphic delusions and hallucinations or perceptual disturbances leading to incomprehensible or incoherent speech
- clouding of consciousness with impairment of attention or concentration, disorientation, perplexity, etc.
- emotional turmoil and affective symptoms (depressed mood, euphoria, anxiety, irritability) without the symptomatic criteria for manic–depressive or recurrent depressive disorders
- rapid changes of the type and intensity of symptoms
- no evidence of causation by organic or psychoactive substances.

It is also associated with some schizophrenic symptoms which are present most of the time:

- mental automatism (thought echo, insertion, withdrawal, or broadcasting)
- control, influence, passivity referred to body movements, thoughts, actions, or sensations
- hallucinations with commentary
- catatonic behaviour
- negative symptoms.

The ICD-10 clinical criteria give no information about psychotic or schizophrenic symptoms or about the action of antipsychotic drugs on these symptoms.

Leonhard[9] described cycloid psychosis as an episode with clouding of consciousness and a marked alteration of thinking. Many authors have reported follow-up studies of cycloid psychoses,[21–23] which confirm the better prognosis of cycloid psychoses than of schizophrenias and schizoaffective disorders.

F23.2 acute schizophrenia-like psychotic disorder (schizophreniform psychosis)

This acute psychotic disorder lasts for less than a month and is mostly schizophrenic. The polymorphic psychotic symptoms are stable (no rapid changes, no emotional turmoil or confusion), sometimes with emotional instability.

The duration criterion is the most important. This category is a provisional diagnosis and appears to include such disparate descriptions as oneirophrenia (oneiroid states or *Erlebnisform*[8]), schizophrenic reaction (DSM.IV 298.8, Brief reactive psychoses), and schizophreniform psychosis.[11] In ICD-10, if the first episode lasts for more than a month, it has to be considered as an acute onset of schizophrenia.

The Scandinavian psychiatric school[24] justify the retention of this category because of the very good and rapid recovery, and have tried to determine factors in the personal and family history predicting the onset of schizophrenia.

Schizophreniform disorder remains in DSM-IV (295.40) because the evidence linking it to typical schizophrenia remains unclear, but the duration criterion is less restrictive (up to 6 months). Features suggesting a good prognosis are onset within 4 weeks, confusion at the height of the psychotic episode, previously good social and occupational functioning, and absence of blunted or flat affect.

F23.3 other acute predominantly delusional psychotic disorders

The main clinical features of this category are delusions and hallucinations. The foreground delusions are mostly persecutory (delusions that the person or close relatives are being malevolently treated or are under external influence, with thought disturbances); auditory hallucinations are present in the background. Despite their stability, these psychotic features do not meet the criteria for schizophrenia (F20).

As for F23.0, the duration of this acute predominantly delusional psychotic episode must be less than 3 months. If the persecutory delusions persist for more than 3 months, the diagnosis changes to persistent delusional disorders (F22). This development from F23.3 to F22 is reminiscent of the classical *paranoia acuta*. Thus, both paranoid reaction and psychogenic paranoid psychosis are included in F23.3. Paranoid reaction must be distinguished from induced delusional disorder or *folie à deux* (F24), in which the delusions of the 'dominated' patient disappear when the two people are separated (see Chapter 4.4).

If the background auditory hallucinations persist for more than 3 months, the diagnosis is changed to other non-organic psychotic

disorders (F28). This diagnostic category is defined by exclusion: the persistent hallucinatory disorder does not meet the criteria for schizophrenia (F20), persistent delusional disorders (F22), acute and transient psychotic disorders (F23), psychotic types of manic episode (F30.2), or severe depressive episode (F32.3). F28 also corresponds to chronic hallucinatory psychosis, as explicitly noted in ICD-10.

F23.8 other acute and transient psychotic disorders

This category includes any other acute and transient psychotic disorders with no evidence of organic cause that are not classifiable under previous F23 categories, such as ephemeral delusions or hallucinations and undifferentiated excitement.

F23.9 acute and transient psychotic disorder unspecified

Brief psychotic disorder (298.8) is defined in DSM-IV as an episode of acute and transient psychotic disorders (delusions and hallucinations with disorganized speech and behaviour) which lasts at least a day but less than a month with eventual full return to previous level of functioning. If the symptoms occur after stressful events in the person's life, brief psychotic disorder with marked stressor(s) has to be specified. If the symptoms occur within 4 weeks post-partum, brief psychotic disorder with post-partum onset has to be specified.

Cultural variants

Other forms of acute psychoses have been observed in both traditional and developing countries, with high prevalence in Asia, Africa, and Latin America. These brief psychotic episodes are culture-bound syndromes, often with immediate precipitating stress or life events.[25] There is disorganized behaviour, delusions, thought disorders, confusion, and mood disorders, usually with full recovery and no relapse in a 1-year follow-up.

The culture-specific disorders[26] and their potentially related syndromes are often acute and transient. The status of these culture-reactive disorders is controversial and needs more clinical and epidemiological research. The mode of assignment to categories in ICD-10 does not suggest category F23.

Appendix I of DSM-IV (Outline for cultural formulation and glossary of culture-bound syndromes) lists 25 syndromes, with a glossary mostly using the local terms (seven in Hispanic languages, five in English, and one in French). *Bouffée délirante*, described only in West Africa and Haiti, is defined as episodes resembling brief psychotic disorder and classified in F23.0. Messich and Lin[25] have suggested that the whole group of culture-bound syndromes should be classified as acute and transient psychotic disorders, although this is justified only for a very few such as *amok* (dissociative episode with persecutory ideas and aggressive behaviour from Malaysia), *shin-byung* (Korean dissociation and possession), and spell (trance state in southern United States).

ICD-10 includes the two Malaysian syndromes *koro* and *latah* as well as *dhat* (India) in F48.8, Other specified neurotic disorders.

International follow-up studies have shown that cultural factors can influence the course and prognosis of acute psychotic disorders. In 1979, the World Health Organization compared the course of schizophrenia (295), psychotic depression, mania, and other psychoses in different cultures, using the ICD-9 criteria for the diagnoses. The outcome for the schizophrenic group was better in emerging countries than in the industrialized world. These results probably explain the individualization of category F23 in ICD-10.

Some authors[27] have suggested that short-lived psychotic episodes are expressions of overcharged mechanisms of defence, or of individual psychological fragility. The brief psychosis is an understandable development of the psychic life of the subject and has a cathartic effect.

Culturally related syndromes are discussed further in Chapter 4.16.

Treatment

Short-term treatment

Acute psychotic syndromes require early hospitalization in either an inpatient psychiatric unit or a crisis centre. These syndromes are to be considered as psychiatric emergencies. The decision to admit to hospital is taken in order to make a careful physical and mental examination clinical evaluation, to separate the patient from his or her environment, to provide a reassuring setting, and to prevent any suicidal or aggressive tendencies.[28]

The goals are to prevent auto or heteroagressivity (suicidal potential, affective symptoms, agitation, aggressive behaviour, command hallucinations, etc.), to reduce the acute psychotic symptoms, to suppress the causal factors and to establish an early therapeutic alliance with the patient and his family. Antipsychotic drugs medications are prescribed.[29] Some clinicians wait for a day or two before starting neuroleptic therapy in order to eliminate an organic cause (a general medical condition or substance abuse disorder can be present with acute and transient psychoticsymptoms) and prescribe benzodiazepines rather than neuroleptics. More often, however, antipsychotic treatment starts immediately after physical, electrophysical, radiological assessments, and laboratory tests (CBC, blood electrolytes, cholesterol, triglycerides, toxicology screen, etc.) to evaluate health status.

The choice of antipsychotic medication depends on the clinician's experience and the clinical features. Second-generation antipsychotic medications (amisulpride, aripiprazole, olanzapine, quetiapine, risperidone, ziprasidone) are commonly prescribed as first-line treatment; clozapine is reserved to schizophrenia with high suicidal potential or to resistant schizophrenia. First-generation antipsychotic medications (chlorpromazine, haloperidol, etc.) are second choice or adjunctive medications.

In cases of major anxiety or agitated behaviour, short-acting sedative drugs neuroleptics such as first-generation antipsychotics chlorpromazine (100–500 mg/day), (loxapine: 50–300 mg/day), with or without benzodiazepines (lorazepam) can be prescribed or levomepromazine (25–300 mg/day) are chosen, or zuclopenthixol acetate (100–300 mg every 3 days by intramuscular injection) is used as a short-acting depot antipsychotic. Parenteral administration (standard intramuscular administration) may be required if the patient refuses oral medication, or if a rapid effect is required because the patient is seriously uncooperative or is too dangerously disturbed.

Predominance of delusions and hallucinations indicates a high-potency antipsychotic agent as haloperidol (5–15 mg/day) or flupenthixol (80–200 mg/day).

Benzodiazepines may be given to potentiate the action of the neuroleptics. Alprazolam (0.5–4 mg/day), clorazepate (50–200 mg/day), and lorazepam (2–5 mg/day) produce rapid sedation in acutely psychotic patients if they are used with a neuroleptic. Some clinicians prefer the combination of two neuroleptics (haloperidol–levomepromazine, haloperidol–cyamemazine).

New compounds with fewer adverse effects can be used (amisulpride, 800–1200 mg/day; olanzapine, 5–20 mg/day; quetiapine, 75–500 mg/day; risperidone, 2–10 mg/day).

In culture-bound syndromes the prescription of antidepressants is often recommended as primary treatment.

The dosage may be adjusted from low doses and gradually increased, or adjusted to the standard dose after a first loading dose. Frequent monitoring to assess drug response and adverse effects (extrapyramidal side-effects, orthostatic hypotension, anticholinergic effects, and temperature dysregulation) is essential, and corrections and prophylactic prescriptions (e.g. antiparkinsonian medications) may be necessary.

Sociotherapy (occupational or intensive) and psychotherapy (reality–adaptive–supportive) are indicated depending on the state of the patient and his environment, with individual, family, or rehabilitation care.

Continuation treatment

The effectiveness of psychopharmacotherapy is usually manifested in the first 6 weeks, with improved sleep, regression of agitation, recovery from anxiety and delusion, and finally disappearance of the psychotic features. When there is no recovery or improvement either another antipsychotic drug should be used or the dosage of the first increased. Worsening of the symptoms, serious side-effects, or a poor response to pharmacotherapy may lead to the main indications for electroconvulsive therapy.

If mood disorders or cyclic episodes occur, treatment with antidepressants, mood stabilizers (lithium or valproate), or an anticonvulsant drug (carbamazepine) may be indicated. Care must be taken to distinguish between a post-neuroleptic depression and the development of a (schizo)affective disorder.

Prevention of recurrence

The possibility that psychotic symptoms may re-emerge has to be borne in mind during the first 2 years of follow-up. Low-dosage pharmacotherapy must be maintained for 1 or 2 years after recovery. During this long-term follow-up, periodic assessment and effective clinical care with social and psychological therapy are essential.

Advice about management

Patients are often hospitalized under constraint because they do not acknowledge the disorder. The initial non-compliance leads to the frequent use of first-generation antipsychotic medications classic intramuscular neuroleptics. After remission recovery, a switch to a second-generation antipsychotic medication newer antipsychotic drug, which is better tolerated, helps to ensure the acceptance of long-term treatment when the psychotic symptoms have disappeared.

In general, psychotherapy and psychosocial care are more effective in an outpatient setting after symptomatic remission recovery has started. A good relationship between patient and psychiatrist together with collaboration with the family practitioner and social workers improve the long-term prognosis. If resources allow, psychotherapy by a trained practitioner, behavioural therapy, or family therapy may be combined with a low-dose pharmacotherapy.[30]

Further information

There are no substantial sources of information in the English language. The following will be of use to those who read French or Spanish. Others should seek further information using an information retrieval system such as Medline.

Weibel, H. and Metzger, J.Y. (2007). Psychoses délirantes aiguës. *Encyclopédie Médico—chirurgicale. Psychiatrie*, 37–230 A 1O 15 p. 142 ref.

Segarra, E.R., Eguiliz, U.I., and Gutierrez, F. (2007). Trastorno esquizofreniforme y trastorno esquizoafectivo. 158 ref. in *Trastornos psicoticos* (*457–592*) Coordinador M. Roca Benassar. Fundacion española de Psiquiatria y Salud mental. Barcelona: Ars medica.

References

1. World Health Organization. (1992). *International statistical classification of diseases and related health problems,* 10th revision. WHO, Geneva.
2. Bercherie, P. (1980). L'école d'Illenau. In *Les fondements de la clinique* (ed. Le Seuil), pp. 119–28. Ornicar, Paris.
3. Meynert, T. (1890). *Klinische vorlesungen über psychiatrie auf wissenschaftlichen grundlagen*. Braumüller, Vienna.
4. Freud, S. (1917). Metapsychologie Ergänzung Zurtraumlehre. In *Gesammelte werke*. Internationales Psychoanalytischer Verlag, Leipzig, 1947.
5. Kraepelin, E. (1899). *Psychiatrie* (6th edn). Barth, Leipzig.
6. Bleuler, E. (1911). Dementia praecox oder Gruppe der Schizophrenien. In *Aschaffenburg Handbuch des Psychiatrie*. Deuticke, Leipzig.
7. Wimmer, A. (1916). Psykogene sindssygdomsformer. In St. Hans Hospital, Jubilee Publication. Gad, Copenhagen.
8. Mayer-Gross, W. (1924). Selbschilderungen der Verwirrtheit (Die oneroïde Erlebnisform). Stringler, Berlin.
9. Leonhard, K. (1961). Cycloid psychoses: endogenous psychoses which are either schizophrenic or manic-depressive. *Journal of Nervous and Mental Diseases*, **107**, 633–48.
10. Kasanin, J. (1933). The acute schizo-affective psychoses. *American Journal of Psychiatry*, **13**, 97–126.
11. Langfeldt, G. (1939). *The schizophreniform states*. Oxford University Press, London.
12. Magnan, V. (1893). *Leçons cliniques sur les maladies mentales*. Bataille, Paris.
13. Legrain, M. (1895). *Du délire des dégénérés*. Progrés Medical, Paris.
14. Samuel-Lajeunesse, B. and Heim, A. (1994). Psychoses délirantes aigues. In *Encyclopedie de medecin et de chirurgie*, 37 230, A10. Editions Techniques, Paris.
15. Ey, H. (1954). Bouffées délirantes et psychoses hallucinatoires aigues. In *Etudes psychiatriques*, Vol. III. Desclée de Brower, Paris.
16. Ey, H. (1955). Psychoses délirantes aigues (bouffées délirantes aigues, psychoses hallucinatoires aigues, états oniroïdes). In *Encyclopedie de medecin et de chirurgie*, 37 230, A10, pp. 1–6. Editions Techniques, Paris.
17. Pull, C.B., Pull, M.C., and Pichot, P. (1987). Des critères empiriques français pour les psychoses. II-Consensus des psychiatres français et définitions provisoires. *Encéphale*, **13**, 53–7.
18. Metzger, J.-Y. and Weibel, H. (1991). Les bouffées délirantes. *Congrès de psychiatrie et de neurologie de langue française*. Masson, Paris.
19. World Health Organization. (1979). *Schizophrenia: an international follow-up study*. Wiley, Chichester.
20. Guilloux, J. (1987). Psychoses délirantes aigues. Statut nosologique et evolution. *Revue Française de Psychiatrie*, **5**, 9–13.

21. Perris, L. (1974). A study of cycloid psychoses. *Acta Psychiatrica Scandinavica*, **253**(Suppl.), 1–75.
22. Brockington, I.F., Perris, L., Kendell, R.E., *et al.* (1982). The course and outcome of cycloid psychoses. *Psychiatry in Medicine*, **12**, 97–105.
23. Barcia, D. (1998). *Psicosis cicloides (psicosis marginales, bouffées délirantes)*. Triacastela, Madrid.
24. Strömgren, E. (1963). Schizophreniform psychosis. *Acta Psychiatrica Scandinavica*, **41**, 483–9.
25. Messich, J. and Lin, K.M. (1995). Acute and transient psychotic disorders and culture-bound syndromes. In *Comprehensive textbook of psychiatry* (eds. H. Kaplan and B. Sadock) (6th edn), Vol. 1, pp. 1049–57. Williams and Wilkins, Baltimore, MD.
26. World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders-diagnostic criteria for research, Annex 2*. World Health Organization, Geneva.
27. Cousin, F.R. and Vanelle, J.M. (1987). Défense et illustration du concept de bouffée délirante aiguë. *Information Psychiatrique*, **63**, 315–21.
28. American Psychiatric Association (2004). Practice guideline for the treatment of patients with schizophrenia. *Supplement to the American Journal of Psychiatry*, **161**.
29. Janicak, P., Davis, J., Preskorn, S., *et al.* (1997). Management of acute psychosis. In *Principles and practice of psychopharmacotherapy* (eds. P.G. Janicak, J.M. Davis, S.H. Preskorn, and F.J. Ayd) (2nd edn), pp. 110–39. Williams and Wilkins, Baltimore, MD.
30. Rochet, T., Daléry, J., and De Villard, R. (1995). Troubles psychotiques aigus et transitoires. In *Thérapeutique psychiatrique* (eds. J.L. Senon, D. Sechter, and D. Richard), pp. 797–803. Hermann, Paris.

4.4

Persistent delusional symptoms and disorders

Alistair Munro

Introduction

Delusional disorder (DSM-IV 297.1 and ICD-10 F22)[1,2] is a psychotic illness with some superficial resemblances to schizophrenia from which, however, it is quite distinct. It presents with a stable and well-defined delusional system, which is typically 'encapsulated' within a personality, which retains many normal aspects, unlike the situation in schizophrenia in which there is widespread personality disorganization in addition to the psychotic features. Nevertheless, although many normal aspects of the personality are preserved, the individual's way of life becomes progressively distorted by the intensity and intrusiveness of the delusional beliefs. Hallucinations may be present but are not usually prominent. This is a chronic disorder, probably lifelong in most instances, which retains an unjustified reputation for being untreatable. Because of the nature of their delusions, many patients are unwilling to accept that they have a mental disorder or that they require psychiatric treatment but, if they can be persuaded to cooperate and accept appropriate medication, the condition can be shown to respond to treatment in a remarkably high proportion of cases.

Delusional disorder used to be known as 'paranoia', and the terms are virtually synonymous. Paranoia and its related disorders were regarded as an important group of psychiatric illnesses until the early part of the twentieth century. Then, because of changing diagnostic and classificatory approaches, especially a tendency to overdiagnose schizophrenia, the diagnosis of paranoia all but disappeared from standard classificatory systems. In 1987, paranoia was again officially recognized by DSM-IIIR but was renamed delusional (paranoid) disorder—since simplified to delusional disorder. It is the only officially acknowledged member of the old group of paranoid illnesses appearing in DSM-IV and ICD-10.

Although the diagnosis of paranoia all but ceased for many years, the illness and its sufferers did not disappear. When the phenomena of the disorder came to attention the patient was either labelled as schizophrenic or else a specific feature of the delusional phenomenology was seized upon and spurious diagnoses were described. Thus we have a multiplicity of apparently disparate diagnoses such as de Clérambault's syndrome (delusional erotomania), the Othello syndrome (delusional jealousy), querulant paranoia (a form of persecutory delusional disorder), monosymptomatic hypochondriacal psychosis (delusional disorder with somatic preoccupations), and many others. The result has been an extraordinarily scattered literature with cases recorded in a variety of medical and non-medical sources, but very few in psychiatric journals until recently. Since DSM-IIIR there has been a serious attempt to resolve the confusion and to diagnose paranoia/delusional disorder by its own intrinsic features, but many problems still bedevil the nomenclature.

Jaspers, in discussing paranoia, said: 'Why are the paranoics as defined by Kraepelin so rare, yet when they do occur they are so typical?' This remains true because there are striking similarities from case to case and the illness' features clearly distinguish it from other psychoses, yet many psychiatrists continue to label it erroneously.

DSM-IV and ICD-10 provide criteria to differentiate delusional disorder as an illness *sui generis* and these are now widely accepted. This section adopts that official approach but with two caveats. The first is that the descriptions are bald and not very helpful to the clinician who has not actually seen cases of the disorder. The second is that the category of delusional disorder (persistent delusional disorders in ICD-10) may well be overrestrictive as it stands. However, some well-respected authorities take a somewhat different approach, regarding 'delusional disorders' as all psychiatric illnesses with delusions and then subcategorizing according to the underlying syndrome, which might be severe mood disorder, schizophrenia, actual delusional disorder, etc. Therefore the reader of any text must be aware of a particular author's criteria for diagnosis in this area.

Emil Kraepelin (1856–1926) clearly described paranoia and included it in a continuum of illnesses with delusional features, especially paraphrenia and paranoid schizophrenia. This so-called 'paranoid spectrum' will be briefly alluded to later. Paranoid schizophrenia continues to be a widely used diagnosis but usually in the context of schizophrenia. Paraphrenia is not officially acknowledged in DSM-IV or ICD-10 but cases fitting its traditional description are quite commonly seen in practice. The present author regards it as a significant entity and the reader is encouraged to become familiar with descriptions to be found elsewhere.

At present, 'delusional disorder' is both an illness category and essentially the only syndrome contained within that category. In recent years, another diagnosis—delusional misidentification syndrome (**DMIS**)—has come into increasing prominence. Originally described in 1923 by Capgras and Reboul-Lachaux[3] as

an illness in which the individual is delusionally convinced that someone familiar in the environment has been replaced by an almost exact double, this 'Capgras syndrome' led a rather marginal existence in the literature for many years. Lately, however, there have been considerably more case-reports of better quality and clinical subtypes have been established. Most importantly, sound psychological and neuropathological work has increasingly shown significant cerebral pathologies in a high proportion of sufferers.

DMIS is not currently recognized by DSM or ICD but in many respects it resembles delusional disorder and should certainly be included in an expanded category of that disorder.

Finally, there is an important phenomenon which is found in association with all illnesses with delusions, especially delusional disorder. This is named 'shared psychotic disorder' in DSM-IV and 'induced delusional disorder' in ICD-10, but is often still referred to by its long-established name *folie à deux*. Here, the primary patient has a bona fide delusional illness and a secondary patient has come to accept the delusional beliefs as true. The secondary patient is usually a highly impressionable individual living in prolonged close contact with the other; he or she is not truly deluded, but retains the beliefs tenaciously as long as the intimate relationship is maintained. A less common variety is when two people each have genuine delusional disorders and, through close proximity, come to share identical abnormal beliefs. *Folie à deux* is not uncommon and, as will be explained later, there are very practical reasons why the clinician should be aware of its possible presence and the ways in which it may influence management of the case.

The paranoid spectrum[4]

Since Kraepelin's time there has been a tacit acceptance by many psychiatrists of a spectrum simplified as:

delusional disorder—paraphrenia—paranoid schizophrenia.

Somewhat anecdotally, the literature suggests that approximately 10 per cent of cases of delusional disorder or paraphrenia will deteriorate to schizophrenia though, in general, most cases of delusional disorder remain diagnostically stable in the long term. Several reports have indicated that, as one moves to the delusional disorder end, a family history of schizophrenia becomes progressively less common. The risk for schizophrenia in the close family of a case of delusional disorder appears to be much the same as in the general public. In paranoid schizophrenia the family history of schizophrenia is approximately half as common as in other schizophrenias and profound disintegration of personality is less frequent.

When dealing with cases in this general area the clinician should bear in mind the concept of a paranoid spectrum. This, plus knowledge of constituent illnesses, will make it easier to distinguish delusional disorder from superficially similar conditions, a matter of considerable importance when considering treatment and prognosis.

Problems of nomenclature

Although English-speaking psychiatrists (and most members of the public) use the word 'paranoid' to mean 'persecutory', strictly speaking it just means 'delusional'. In many writings on 'paranoia' and 'paranoid' disorders, authors do not make it clear whether delusions are present or not in their cases.

Unfortunately, with the passage of time, the term 'paranoid' has come to be used so loosely that it has lost any meaningful clinical connotation. Paranoia should now be regarded as an historical usage, pretty well synonymous with delusional disorder.

The word 'paranoid' is still used in the official diagnoses of paranoid schizophrenia and paranoid personality disorder. The former is acceptable because the illness has delusions as a prominent feature, but it is quite illogical in describing a personality disorder, which cannot have delusions. Since it is unlikely that the personality disorder will be renamed soon, the reader should be aware of such pitfalls in our psychiatric terminology and consequently the need for ultra-careful case-descriptions.

Although the **form** of delusional disorder is remarkably characteristic, the delusional **contents** and the ways in which cases come to attention are extremely varied, and this has led to an extraordinarily complex history. The core description, that of paranoia, gradually crystallized in the latter half of the nineteenth century and was definitively delineated by Kraepelin, who recognized subtypes with delusional contents of grandiosity, persecution, erotomania, and jealousy, and also allowed for the possibility of a hypochondriacal content. He clearly differentiated paranoia from *dementia praecox* (i.e. schizophrenia). Kraepelin later doubted whether hallucinations could be present: in fact, non-prominent hallucinations are now acceptable and in every other respect Kraepelin's century-old definition of paranoia still largely serves to describe present-day delusional disorder.[5]

Subsequently, Kraepelin[5] introduced the concept of paraphrenia, an illness similar to paranoid schizophrenia but with significantly better preservation of affect and of personality. As already mentioned, he regarded paranoia, paraphrenia, and paranoid schizophrenia as a relatively discrete group of illnesses, later referred to as the paranoid spectrum.

It was later found that a minority of cases of paranoia and paraphrenia eventually deteriorated to schizophrenia and this somewhat illogically led to these diagnoses being progressively ignored. Despite this, speculation on the nature of delusions continued, most notably by Jaspers (1883–1969),[6] Kretschmer (1888–1964),[7] and Freud (1856–1939) and his followers. These speculations contributed a good deal to the descriptive phenomenology of delusions but whereas we know a good deal about delusional contents, we understand little about the origin of delusions or of delusional illnesses, or the reasons for their unique features.[8] Unfortunately, as much of the writing on delusions appeared when most psychoses, and certainly paranoia, had no effective treatments, writers usually dwelt on the untreatability of paranoia, a pessimistic view that persists but is no longer warranted.

From the 1970s onwards, interest in paranoia reappeared and a more optimistic view of treatment emerged. Since its renaissance as delusional disorder in DSM-IIIR in 1987, paranoia has again become a respectable diagnosis. Not only that, it has subsumed several quasi-disorders which were undoubtedly delusional but which had been described superficially on the strength of their delusional content alone. Several of these have already been noted (see p. 281).

Nowadays the clinical description of delusional disorder is well established, but adequate case series are rare and scientific investigations are in their infancy, except in the case of the diagnosis which still remains officially unrecognized, delusional misidentification syndrome, in which underlying brain abnormalities are

commonly demonstrable. The separateness of delusional disorder from schizophrenia is beyond doubt, but its relationship to the other constituents of the paranoid spectrum still has to be determined. Delusional disorder is no longer regarded as rare, but many years of neglect have left many psychiatrists sadly unaware of its characteristic features.

Delusions: clinical aspects

A delusion may be defined very loosely as a mistaken idea which is held unshakably by the patient and which cannot be corrected. As will be seen, this is not a satisfactory definition, although it may be a useful starting point for clinical recognition of a delusional process. This brief exposition is concerned to facilitate clinical recognition and not to dwell on psychopathological theories, which are dealt with in detail elsewhere in this book.

It is a widely held opinion that delusions are qualitatively different from normal ideas or beliefs and have an all-or-nothing aspect. The DSM-IV definition initially seems to accept this viewpoint, stating that a delusion is 'A false belief based on an incorrect inference about external reality that is firmly sustained despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary. The belief is not one ordinarily accepted by other members of the person's culture or subculture'. But the definition goes on to say that it is often difficult to distinguish between a delusion and an overvalued idea (in which there is an unreasonable belief or idea but not held with such pathological certitude as in a delusion), and that 'Delusional conviction occurs on a continuum' from normal to abnormal. These two statements markedly lessen the initial description of the absolute nature of the delusional wrongness.

The definition of delusion by Mullen[9] based on the earlier description by Jaspers is widely quoted and its implications are largely accepted by DSM-IV and ICD-10. He characterizes delusions as follows:

1 They are held with absolute conviction.

2 The individual experiences the delusional belief as self-evident and regards it as of great personal significance.

3 The delusion cannot be changed by an appeal to reason or by contrary experience.

4 The content of delusions is unlikely and often fantastic.

5 The false belief is not shared by others from a similar socio-economic group.

Clinicians widely employ the terminology on delusions introduced by Jaspers, for example when they use terms such as 'primary' and 'secondary' delusions, 'delusional mood' (*Wahnstimmung*), and 'delusional memory'. These concepts are of some descriptive and possibly heuristic value, but they do not prove particularly helpful in distinguishing delusions from overvalued ideas in individual cases, nor in deciding whether a particular delusional phenomenon is specific to a given mental disorder.

In a sense, all delusions are secondary in that they are the product of a pathological process in the brain which, in most cases, we can only guess at. It is sometimes useful to differentiate clinically between the 'primary' or 'autochthonous' delusion, which appears fully fledged and relatively suddenly, and the 'secondary' delusion, which is a further development within the delusional system and may sometimes seem to be the individual's way of rationalizing his delusional beliefs although, of course, the rationalization must necessarily be filtered through a mind already thought-disordered and affected by delusions. For example, the initial belief may be that the police are watching him night and day; the secondary delusion 'explains' that this is because he has secret information about aliens which the authorities do not wish divulged. The better organized the delusions, the more convincing are the 'explanations', even to outsiders.

Not all primary delusions arise suddenly and, in fact, it must be presumed that in most cases the suddenness is more apparent than real. Almost certainly, unless the delusion is the result of an acute brain dysfunction such as may follow a head injury or delirium, there is a lead-up process, which may be accompanied by the aforementioned *Wahnstimmung*, a mood state compounded of anxiety, perplexity, and a sense of impending crisis. When the delusion crystallizes, the delusional mood often disperses and is replaced by a sense of revelation and of certainty. It seems likely that this phenomenon occurs in a proportion of delusional disorder patients and it often happens that, at the moment of revelation, some coincidental but irrelevant circumstance is picked upon to explain the appearance of the new belief. For example, a media event, a thunderstorm, a chance telephone call, etc., may thereafter be, in the patient's mind, the 'cause'.

While we regard delusions as one of the most characteristic elements of all the psychotic illnesses and a sine qua non in the diagnosis of delusional disorder, clear-cut description, and delineation have proved elusive despite many years of study and experiment.[10] In fact, it would seem that none of the characteristics of delusion which we traditionally accept stand up completely to scientific scrutiny. In particular, nowadays the so-called bizarreness of a delusion has been shown to have little or no distinguishing value.[11]

Much of the classical work on delusions was done in pretreatment times when the chronic condition was readily available for study in institutions. In the present era our aim is to diagnose psychotic disorders as early as possible, sometimes even before frank delusions are evident, and to begin treatment at once. Neuroleptics rapidly interfere with many psychopathological processes; they certainly suppress delusions, although not necessarily permanently. Of course this makes ongoing experimental observations of delusions, especially of the acute variety, all but impossible in clinical circumstances. Psychiatrists find themselves in the paradoxical situation of diagnosing illness because of the presence of delusions whose scientific validity is largely unsubstantiated, and then causing these to disappear before they can be verified properly. Nevertheless, until we have more objective means of making diagnoses it remains essential that, as far as we can, we recognize delusions when they occur and separate them from other abnormal psychopathological appearances.

How can a clinician deal with this? Firstly it seems inescapable that he or she be both experienced and insightful. Given these qualities, it often does seem possible to have an informed sense of whether a belief is true or false and, if the latter, whether it is being held with delusional intensity. A key element in the decision is a comparison between the patient's current beliefs and those he

habitually held, and here a corroborative account from an informed outside source is usually necessary.

The observer's educated suspicion that a delusion is present is the starting point, but it is evident that that suspicion has to be aroused by the context of the apparently delusional idea because, no matter how isolated it appears to be, it nearly always occurs in the setting of a mental disorder whose other features may indicate a specific psychiatric diagnosis. Illogically, instead of recognizing the delusion and using it to make a definite diagnosis, we develop the conviction that we are dealing with a probable psychosis and thereafter judge all the patient's utterances in light of that. While he may indeed be experiencing delusions, it is essential that we do not automatically assume that anything the psychotic individual says has of necessity to be of a delusional nature.

We must accept that we cannot be absolute in our recognition of a delusion. In addition to the illness context we base our estimate on a series of nuances, no one of which is pathognomonic but an accumulation of which becomes increasingly convincing. The abnormalities to be sought are as follows:

1 An idea or belief is expressed with unusual persistence or force.

2 As far as we can tell, the idea is not typical of the individual's previously prevailing thinking and is not shared by his or her social community.

3 The idea appears to exert an undue influence on the person's life and consequently the way of life is altered to an extraordinary degree.

4 Despite the significance to the patient of the belief, he or she often displays secretiveness or resentment when questioned about it.

5 The individual tends to be humourless and oversensitive about the belief.

6 There is a quality of 'centrality'; no matter how strange the belief or its consequences, the patient rarely questions that incredible things are happening to him or her. For example, why should a perfectly ordinary harmless person be singled out for constant surveillance by the security agencies? But this is simply accepted.

7 Attempts to contradict the belief are likely to arouse an inappropriately strong emotional reaction, often with irritability and hostility and with a superciliousness that may be a form of grandiosity.

8 On reflection the belief appears unlikely to the observer, but at the time of history-taking the vehemence of its expression may temporarily disguise its improbability.

9 The patient is so emotionally overinvested in the idea that it swamps other elements in the psyche, and many everyday activities are neglected.

10 If the delusion is acted out, uncharacteristic behaviours, sometimes involving violence, will occur which may be partly understood in terms of the abnormal belief.

11 Others who know the patient well will usually observe that his or her thinking and behaviour are alien, unless *folie à deux* is present when, paradoxically, the other person's denials of abnormality are themselves possible confirmation of the presence of delusion.

12 An odd feature of delusions is that, no matter how strongly they are held, when the patient is given the opportunity to obtain real proof he or she persistently evades accepting the opportunity.

13 One must always look for the features which frequently accompany delusions, especially suspiciousness, hauteur, grandiosity, evasiveness, and eccentric or threatening behaviour, as well as evidence of thought disorder, mood change, and hallucinations.

Particular features of delusions in delusional disorder

In addition to any of the above, in delusional disorder we find several other elements, which are of importance in leading to the diagnosis:

1 The delusional system is stable and is expressed or defended with intense affect and with highly rehearsed arguments. The form of logic used by the patient is very consistent but the propositions are based on false premises. Since the individual is so focused on his beliefs and is so self-assured, he often succeeds in making the enquirer feel inept.

2 The delusional system is markedly 'encapsulated', so that the beliefs therein and their accompanying symptoms are to a considerable extent separated from the rest of the personality which retains a good deal of normal function. However, the compelling force of the delusions often overshadows these normal aspects and this is increasingly so with advancing chronicity of the illness, when the tendency to express and act out the delusions may well increase.

3 When the individual is preoccupied with the delusional system there is strong emotional and physiological arousal, but when he or she is engaged on neutral topics, the arousal abates and an ordinary conversation can take place. Switching between normal and abnormal 'modes', sometimes very rapidly, is virtually pathognomonic of delusional disorder.

4 Because of the encapsulation of the delusions and the normal–abnormal switch just described, the patient may have phases of relative normality interspersed with psychotic periods. The switch can occur spontaneously or as a result of external provocation; the two are difficult to disentangle because the hypervigilant individual may perceive provocation in almost anything. Since it is a chronic illness the symptoms never remit, but if they are temporarily in the background the patient may converse and function almost normally and may have sufficient quasi-insight to keep the delusions concealed for the moment. Total denial of mental abnormality and resistance to psychiatric referral are almost universal in cases of delusional disorder and lead to severe underestimation of the illness's frequency.

5 As a result of the features just described, many delusional disorder patients can continue to exist in society, sometimes with very abnormal but harmless beliefs but in other instances with highly malignant delusions, which they may or may not act out.

6 As will be repeatedly emphasized, delusional disorder must be diagnosed on the **form** of the illness and the content of the delusion is not used to make the primary diagnosis. On the other hand, the particular content is employed to categorize into subgroups, as will shortly be described.

Delusional disorders: clinical features

Official diagnostic criteria

The DSM-IV and ICD-10 criteria are shown in Tables 4.4.1 and 4.4.2, respectively.

As will be seen, the DSM-IV and ICD-10 descriptions are very similar in overall outline but with a number of rather striking minor differences. The following specific items should be noted:

1 DSM-IV uses the term 'non-bizarre' delusions; this criterion has been shown to have little or no validity.[11]

2 DSM-IV allows the presence of tactile and olfactory hallucinations, while ICD-10 mentions only auditory hallucinations; in practice most modalities may be represented but the important point is that they are relatively non-prominent and usually parallel to the content of the delusion(s).

3 DSM-IV says that delusions should have been present for 1 month and ICD-10 insists on 3 months. Both are guesses, but ICD-10 is probably right to err on the side of caution and it provides category F22.8 as a temporary niche until the definitive diagnosis emerges.

4 Both classifications exclude delusional illnesses due to organic brain disorder, medical illnesses, medication effects, or psychoactive substance abuse. In essence this is correct, especially in an illness of acute onset. However, as will be noted later, an apparently typical delusional disorder may arise as a long-term complication of any of these factors.

5 DSM-IV and ICD-10 agree emphatically that delusional disorder is not schizophrenia and DSM-IV notes that general functioning is not impaired. Both say that mood disturbance may accompany the delusional illness but is not a cause of it.

6 The list of subtypes according to delusional content is similar in both classifications, although ICD-10 adds self-referential and litigious themes.

Table 4.4.1 DSM-IV delusional disorder (297.1)

Principal features
(a) Non-bizarre delusions of at least 1 month's duration
(b) Criterion A for schizophrenia has never been met, although tactile and olfactory hallucinations may be acceptable if they are related to the delusional theme
(c) Apart from the impact of the delusion(s) or its consequences, functioning is not markedly impaired and behaviour is not obviously odd or bizarre
(d) Concurrent mood episodes, if present, are brief relative to the duration of the delusional disorder
(e) The disturbance is not the direct outcome of a drug or medication or of a medical disorder
Subtypes
Erotomanic
Grandiose
Jealous
Persecutory
Somatic
Mixed (allowing for the presence of more than one of the foregoing)
Unspecified or other

Table 4.4.2 ICD-10 persistent delusional disorders

Delusional disorder (F22.0)
Principal features
(a) A delusion or set of related delusions, other than those described as typically schizophrenic, must be present; the most common are persecutory, grandiose, hypochondriacal, jealous, or erotic
(b) The delusion(s) must be present for at least 3 months
(c) The general criteria for schizophrenia are not fulfilled
(d) There are no persistent hallucinations, but there may be transitory or occasional auditory hallucinations that are not speaking in the third person or making a running commentary
(e) Depressive symptoms or episodes may be intermittently present, but the delusional symptoms must persist at times when there is no disturbance of mood
(f) There must be no evidence of primary or secondary organic mental disorder or of a psychotic disorder due to psychoactive substance use
Subtypes
Persecutory
Litigious
Self-referential
Grandiose
Hypochondriacal
Jealous
Erotomanic
Other persistent delusional disorders (F22.8)
This is a residual category for persistent disorders with delusions that do not fully meet the criteria for delusional disorder or schizophrenia. Illnesses with prominent delusions accompanied by persistent hallucinatory voices or by psychotic symptoms insufficient to satisfy the criteria for schizophrenia are included here. A delusional disorder of less than 3 months' duration is coded under Acute and Transient Psychotic Disorders (F23) until proven otherwise.

7 Neither classification specifies that the essence of delusional disorder is a highly organized delusional system, largely encapsulated from normal aspects of the personality, although DSM-IV hints at this when it comments that functioning is not markedly impaired and behaviour is not obviously odd or bizarre. Neither comments that the patient can demonstrate alternating 'normal' and 'delusional' modes.

8 The ICD-10 category of 'other persistent delusional disorders' is vaguely described and is largely a catch-all heading or, as mentioned above, a temporary holding station. However, it could conceivably be used for the time being to subsume the unofficial delusional disorder diagnoses of paraphrenia and delusional misidentification syndrome.

9 Overall, DSM-IV and ICD-10 give rather laconic descriptions of delusional disorder and it will be necessary to flesh them out with relevant clinical details. This will be done after the next section on aetiological considerations.

General aetiological considerations in delusional disorders

It must be stressed that knowledge of aetiology in delusional disorder is scanty and highly speculative, largely because so little modern research has been conducted. What follows is an outline,

and certain other factors will be noted when we come to consider some of the illness.

(a) Genetic factors

Changes in definitions of paranoia/delusional disorder over the years and the frequent confusion with schizophrenia make most studies all but impossible to interpret. Conclusions are inferential rather than evidence based. However, it seems well established[12] that delusional disorder and paranoid schizophrenia are less directly inherited than other forms of schizophrenia, and that there is little or no evidence of a genetic link between delusional disorder and schizophrenia.

There may be genetic links with certain severe personality disorders, especially of the paranoid and schizoid varieties, but these are difficult to substantiate. There does seem to be an excess of such disorders in relatives and premorbidly in delusional disorder patients themselves. It is suggested that paranoid and schizoid traits are particularly liable to lead to social isolation and aggravation of delusional tendencies.[13,14]

(b) Organic brain factors

Recent evidence from the study of delusional misidentification syndrome (see later) indicates that delusions of a very specific type may arise in association with certain well-defined brain insults. There are strong hints, but much less supportive evidence, to suggest that organic brain factors may also be important in cases of delusional disorder. For example, head injury may lead to the development of marked paranoid symptoms, and there is a long-established association between chronic alcoholism and pathological jealousy.[15] Old age itself may be linked to the onset of symptoms typical of delusional disorder, and early evidence of brain changes, especially in subcortical areas, is starting to appear in studies of various kinds of senile 'paranoid' illness.[16–18] Amphetamine and cocaine abuse[19] can induce delusional illness, as can therapeutic drugs, including L-dopa and methyldopa,[20] at times. Delusional illness induced by the brain effects of AIDS infection has been documented.[21]

Gorman and Cummings[22] have proposed that delusional illnesses of organic origin have underlying features in common, particularly temporal lobe or limbic involvement and an excess of dopamine activity in certain areas of the brain.

If organic factors predominate in a particular case, delusions must be seen as a secondary feature of an organic brain disorder. However, if the organic factors are subtle and of long duration, the clinical appearances may be those of a quite typical delusional disorder which, interestingly, may well respond to neuroleptic treatment as effectively as idiopathic cases. (In fact, 'idiopathic' may simply denote organicity at a more subtle level.) It is very possible that organic brain factors are much more common than we suspect in delusional disorder, especially in young males who have previously abused alcohol or drugs or have suffered a head injury in the past, and in older patients (more commonly female) who suffer from effects of an ageing brain.[23,24]

(c) Interplay with mood factors

We have already seen that DSM-IV and ICD-10 agree that mood symptoms may accompany delusional disorder but not cause it. Delusional and mood disorders are separate illnesses with their own natural histories and responses to treatment, yet there is a complex relationship between them, as is also the case with mood disorder and schizophrenia. For example, it is well documented that some cases of apparently typical mood illness, unipolar or bipolar, can progress to delusional disorder or schizophrenia over time. Conversely, cases which appear to be delusional disorder but with an episodic course may prove to be bipolar illness. There are a number of anecdotal reports of delusional disorder responding to antidepressant treatment, and it is more than likely that these represent a failure to recognize the true nature of a mood disorder associated with delusions.

Both depressive disorder and mania may be complicated by delusions. On the other hand, mood symptoms, especially dysphoria with anxiety, are a common complication of delusional disorder, while individuals with the grandiose subtype may show elation, which mimics mania but is far more sustained. In recovering delusional disorder, one may see postpsychotic depression of varying degrees of severity and this is described later. Suicide is not unknown in delusional disorder but its frequency is undetermined.

In many delusional disorder patients the illness is profoundly isolating and sets them at odds with the rest of the society, which often generates suspiciousness, dejection, anxiety, and agitation in the individual. It seems that a vicious circle results whereby the delusion induces distress and physiological overarousal which, in turn, reinforces the strength of the delusion and progressively diminishes reality input.

(d) Psychodynamic theories of causation

The psychodynamic literature continues to discuss aspects of 'paranoia' but often fails to differentiate clearly between trait, symptom, personality disorder, and psychotic illness. Most of the emphasis is on the persecutory aspect of paranoia, with only occasional references to other types of delusional content. Since psychotherapists rarely treat psychotic patients, their experience of delusional phenomena must actually be rare and their knowledge of the features correspondingly scanty. Their theoretical bias is to interpret the origins of paranoia in terms of psychological maldevelopment, ignoring the increasing weight of evidence that faulty brain mechanisms are involved. One must read the psychoanalytic literature on this particular topic with an ultracritical attitude, since it usually fails to provide adequate illness definitions or clear case reports and generates explanatory theories which are unjustifiably presented as proven facts.

(e) Conclusions regarding aetiology

No systematic research on paranoia took place for more than half a century and modern investigations into delusional disorder are only beginning to appear. Therefore it is premature to propose specific aetiological theories. However, a gathering weight of evidence does suggest a localized and relatively circumscribed brain disorder associated with the possible influence of abnormal neurotransmitter activity, probably involving dopamine overactivity. Whatever the original basis of delusional disorder, it certainly seems that provocative influences such as head injury, alcohol abuse, and ill effects of drugs may play a part, whereas speculation about psychological causations suggest that this is at most a secondary influence. There is an urgent need for the study of extended case series utilizing modern neurophysiological and neuropsychological investigative methods.

Delusional disorder: general features and introduction to the subtypes

We have already outlined the diagnostic criteria for delusional disorder in DSM-IV and ICD-10 and have amplified these with descriptions of many of the clinical phenomena associated with the illness. It has been emphasized that this is a stable and readily recognizable disorder, provided that the clinician is informed of the essential criteria and has dealt with at least several cases to familiarize him- or herself with its very characteristic 'feel'. With this experience it becomes much more possible to delve under the prominent symptoms related to delusional content and to discern the underlying form of the illness. However, it is the predominant delusional content in an individual case, and the symptoms and behaviours related to this, which decide how a patient will present for assessment. Therefore we shall consider the main subtypes in some detail. It cannot be stressed enough that these are not separate types of illness, but variants on a single psychopathological theme.

All cases of delusional disorder occur in clear consciousness and have a stable and persistent delusional system which is relatively encapsulated. Since much of the personality remains remarkably intact, a considerable degree of social functioning is retained in many cases. The patient experiences a heightened sense of self-reference within the delusional context and ordinary events take on unusual significance. He or she clings to the delusion with fervid intensity and spurns any suggestion that a mental illness is present. Outside the delusional system the patient shows quite normal thinking, affect, and behaviour, but there is a marked tendency for gradual pushing to one side of these normal aspects. The retention of such a degree of normality makes the illness totally different from schizophrenia.

Earlier it has been indicated that the DSM-IV criterion of non-bizarreness is unhelpful, although in all cases of delusional disorder the delusions are relatively well structured, coherent, and consistent, and the logic would often be acceptable if it weren't that its basic premises are irrational. Many affected individuals can maintain overtly normal activities, at least in public, but increasing pressure of the delusion tends to cause corresponding responses in behaviour; these may be channelled socially, as in hypochondriacally deluded patients who utilize medical resources, albeit excessively, or antisocially, as in the aggression of the jealously deluded individual. Mood abnormalities are common as a response to the effects of the illness.

Hallucinations do occur in some cases and may affect any modality, but they are often difficult to assess and to differentiate from delusional misinterpretations and illusions. Widespread persistent hallucinations in more than one sensory sphere should make one cautious about the diagnosis of delusional disorder.

The illness appears to affect men and women approximately equally, but it is not clear if this is true of all subtypes. Despite older assertions that the illness is restricted to the middle-aged and elderly, the age of onset can actually be from late adolescence to extreme old age, with male patients appearing on average to experience earlier initiation. Some patients behave in an eccentric or fanatical fashion and, as a group, delusional disorder sufferers are excessively likely to be unmarried, divorced, or widowed, probably reflecting restriction of affective responses and some isolative tendencies. Despite this, the condition can be compatible with marriage and continued employment. The premorbid personality is usually described as asocial and there may indeed be an excess of long-standing schizoid and paranoid personality disorders. However, when a patient makes a good recovery there may be little evidence of this, and it is possible that in some cases a 'personality disorder' is actually the prolonged and insidious prodrome of the illness.

Onset may be gradual or acute. In the latter the patient often identifies a precipitating stressor, which is difficult to confirm (e.g. the person who has a delusion of skin infestation may attribute it to a single insect bite many years previously). While most individuals are secretive about their abnormal beliefs or express them by such means as physical complaints or legal processes, a certain number actually utilize them, perhaps within the context of an extreme religious sect or by becoming an excessively insistent agitator on some social issue. Disinhibited and overtly aggressive behaviour seems more likely to occur in males, at times leading to clashes with the authorities.

In all cases of delusional disorder, no matter what the nature of the delusional theme, the investigator should look for the relatively unique feature of the illness—the patient's ability to move between normal and delusional modes of thinking. In the former there is relatively calm mood, reasonable rapport, and appropriate emotional responses, whereas in the latter there is overalerting, suspiciousness, and the sense that the person is being remorselessly driven by the delusional beliefs. This situation is difficult for the inexperienced observer to comprehend, since it is inconceivable to most people that someone who can appear perfectly rational at one moment can almost instantaneously change to a possessed irrational being—and then back again just as quickly. In a sense the same patient is both sane and insane, and when in the latter mode may be ultrapersuasive about the acceptability of his or her beliefs. One may imagine the plight of a lawyer faced with a client who has committed some uncharacteristically outrageous act as a result of a delusion, who can then discuss his case with apparent insight and logic, and even genuine remorse, but who nevertheless remains totally self-justifying. As a corollary, the client will usually deny the possibility of mental illness and often refuses to cooperate with psychiatric assessment. He may also refuse to cooperate with the legal process, to his knowing detriment.

Delusional disorder, when it was known as paranoia, often had a bad reputation because patients were regarded as angry, suspicious, accusatory, and potentially violent. Some undoubtedly are, but as we consider the various subtypes nowadays we realize that many sufferers, perhaps the majority, lead lives of internalized despair in progressively isolated circumstances. Anger and suspiciousness are often secondary, at least in part, to the perceived neglect of their overwhelming concerns. The illness is chronic and self-reinforcing, and it is likely that only a minority of cases are recognized or helped. Psychiatry does not have an impressive record of helpfulness towards this group of patients.

The subtypes of delusional disorder

As previously noted, DSM-IV recognizes five main subtypes based on the predominant delusional themes: the erotomanic, grandiose, jealous, persecutory and somatic, plus mixed and unspecified types. ICD-10 also recognizes these subtypes, and adds litigious and self-referential categories. Here, the litigious variety is included within the persecutory group and self-referential cases are not given separate status since self-reference is, in reality, a feature of the illness as a whole and prominent in all cases.

When delusional disorder was resurrected in DSM-IIIR, single delusional themes were emphasized, but the mixed category in DSM-IV accepts the reality that, for example, a hypochondriacal individual can also feel persecuted and an erotomanic patient can be extremely grandiose. Also, we shall find that there are considerable individual variations within the overall themes, so that in the somatic subtype there are cases involving different body systems. Yet the range of major themes does not appear to be all that wide and we have no explanation for this relative restriction in their number. The 'unspecified' category in DSM-IV allows us to accommodate any case whose delusional theme is unusual and leaves a door open to the discovery of other themes in the future.

In presenting the subtypes, relatively more attention will be given to the somatic form. This should not be taken as an indication that this is the most common variant; rather, it happens to be the one which has been best documented in the recent psychiatric literature. Other types of delusional presentation are much more often described in non-psychiatric and non-medical sources, where the fundamental nature of the illness may be overlooked, and so we are only beginning to correlate such descriptions with modern findings on delusional disorder.

Delusional disorder: persecutory and litigious subtypes

In most people's minds the persecutory type of delusional disorder is the archetype of 'paranoia' and it is usually assumed that it is the most common variety. Therefore it is surprising to find that the literature, while full of speculation, is very lacking in good descriptions of the phenomenology of the illness and, apart from unreliable psychoanalytic theory, says relatively little about persecutory delusions themselves.

Clinical features

By definition the illness is a chronic psychotic disorder with a well-systematized delusional system and with relative sparing of the surrounding personality. The persecutory threats may be perceived simply as coming from 'them', but can elaborate to descriptions of the most labyrinthine plots involving a variety of known and unknown adversaries. The beliefs are extremely stable and usually increase in complexity with the passage of time. There is heightened awareness and misinterpretation of neutral environmental cues and, not unnaturally suspiciousness, extreme anxiety, and irritability are present. Elements of grandiosity are not uncommon, with the individual accepting that he or she is the centre of focused and malignant attention that would be inexplicable to the normal person. As the illness progresses there is a tendency to involve an increasing number of people in the persecutory system, not uncommonly relatives, physicians, law-enforcement agencies, aspects of government, and others.

As with other subtypes of delusional disorder, many individuals are able to conceal their increasingly insistent delusions for some time, but because of fear of harm they are likely to isolate themselves more and more. If they live alone they may come to be regarded as eccentrics, but if they remain in contact with society the suspicion and anger must eventually become evident, so that interactions with family, social agencies, or the authorities become increasingly confrontational.[24] Despite the reputation of 'paranoia' for violence, only a small proportion of these individuals resort to threat or assault, but with those who do the danger may be profound as the individual is without reservation in his beliefs and will act as though genuinely under severe provocation. Disinhibition may at times be engendered by alcohol or drug use, which makes such situations even more volatile.

Even in a long-standing illness, islands of normal functioning remain; despite this there is little or no insight and the patient resists any psychological explanation for his beliefs. He usually refuses to see a psychiatrist voluntarily; many patients of this kind are encountered in a forensic setting only after an outburst of unacceptable behaviour and are minimally cooperative.[25]

Litigious variety of the persecutory subtype (querulous paranoia)[26,27]

In some individuals with delusional disorder there is a profound and persistent sense of having been wronged in some way, and these people endlessly and repetitively seek redress, sometimes personally but often through the legal system. In a proportion of cases there may initially have been a genuine grievance and there may also have been unsatisfactory recompense, but the subsequent pursuit of 'justice' becomes never-ending and also becomes self-reinforcing because no satisfactory resolution is possible.

This group may not be large but it generates considerable media publicity. Reports of cases naturally tend to be in the literature of the legal profession, the law-enforcement agencies, and, to some extent, forensic psychiatry, but rarely from general psychiatry. Because the individual appears relatively high functioning apart from his delusional beliefs, the complaintive behaviour may be regarded as mere eccentricity for a long time. As in many cases of delusional disorder, the immediate complaint and behaviour may seem coherent and not unreasonable but over time their ongoing, never-ending, and extraordinarily demanding quality begin to raise the suspicion of severe underlying psychopathology. Even then, unless the person begins to be perceived as a threat, little may be done and prolonged harassment of officialdom and the legal system may be tolerated for surprisingly long periods. In some national communities (e.g. Germany and the Scandinavian countries) there are legal provisions to stop 'barratry' or unreasonable use of the law by declaring an individual a querulous litigant.

Goldstein[28] has described three typical ways in which 'litigious paranoia' presents. The first is the 'hypercompetent defendant' who knows and uses the letter of the law up to and beyond its limits but pays no heed to its spirit. The second is the 'paranoid party in a divorce proceeding' who is often consumed with jealousy and pursues vendettas against the ex-spouse, the lawyers on both sides, and even the judge. The third is the 'paranoid complaining witness' who endlessly initiates litigation despite repeated adverse judgements. All such individuals pursue their grievances in a driven manner, see conspiracy in every corner, and are often quite unscrupulous in their single-mindedness, blatantly bending facts to fit with their beliefs. Since they hold the delusional belief with total conviction, they can accept no counterargument or contrary facts. In the past, persistent litigation was virtually a preserve of the rich, but many modern societies provide a variety of avenues for complaintiveness and will even support complaint procedures, and so abnormally litigious behaviour appears to be on the increase.[29]

Diagnosis of the persecutory subtype

All the features of a delusional disorder as previously described are present. In this subtype, wariness, irritability, suspiciousness, and threatening behaviour are especially prominent, and both impulsive and planned violence may occur. Gaining confidence is extremely difficult, but if this succeeds, the more normal aspects of the individual's personality may become apparent and one may also perceive how chronically anxious and overalerted he or she is.

Differential diagnosis

The illness must be distinguished especially from the following:

- paranoid schizophrenia
- paranoid and antisocial personality disorders
- substance-related disorders
- organic brain disorders, including early dementia and some epileptic disorders
- obsessive-compulsive disorder.

Epidemiology

Virtually nothing is known of the frequency and distribution of persecutory delusional disorder. As with other subtypes it occurs in both sexes, but male cases are probably overreported because of a readier tendency to violence and antisocial acts. The literature is biased by the reporting of the most overt cases, often in the news media or through the courts. It is open to speculation how many cases avoid diagnosis; as noted, relatively few come to the clinical attention of psychiatrists other than in forensic work.

Course and prognosis

Delusional disorder is very chronic, and it is presumed that cases of the persecutory subtype are as likely as others to be lifelong and to show increasing psychopathology with the passage of time. In a proportion of cases there is always a risk of violence and illegal behaviour. Since cooperation with assessment and treatment is usually minimal, the overall figures for prognosis must be bad, but we have no reliable data to confirm this.

Forensic complications[30,31]

If someone with a generalized psychotic disorder like schizophrenia becomes sufficiently disorganized, functioning in the community becomes impossible. In contrast, many delusional disorder patients retain a sufficient grasp of reality to continue existing in society, sometimes indefinitely. However, this does not imply that their illness is quiescent. Intellectual ability, capacity for reasoning, and the form of thought remain relatively intact, but the delusional process worsens. They retain the ability to brood on their beliefs so that normal thought processes and delusions interweave, as do normal and abnormal behaviours. Anger may express itself explosively, but some individuals carry out violent actions in a very calculated way, believing that a just vengeance is being exacted. Afterwards there may be real regret and a clear awareness of a wrong having been committed against society, but the actions are seen as justifiable and necessary. The patient is usually aware that by societal standards his deed is legally and morally wrong, but that awareness resides within the normal non-delusional aspect of his mental functioning. Within the confines of the delusional system, the person unswervingly believes that it was obligatory for him to behave as he did.

In such cases, the judge and jury are placed in a quandary, made worse by the individual's frequent arrogance (related to grandiosity), self-justification, and ambivalent expression of regret. The ability to acknowledge the wrongness of one's action in general terms and even to show remorse for it, while also asserting that it was necessary to carry it out, may well be regarded as indicating wilfulness or hypocrisy. Then, paradoxically, culpability may be determined by the content of the delusion, although this has minor relevance to the disinhibition of behaviour. Thus, as Goldstein has pointed out, if the person felt threatened because of a delusional belief and reacted, as he genuinely perceived, in self-defence, his degree of blame may be adjudged to be low. But if he were equally deluded and carefully plotted revenge, this might be seen as highly culpable. Such a distinction cannot be defended logically either in the clinical situation or at law.

Delusional disorder defies any definition of insanity in black or white terms; it is both black and white. Because few psychiatrists, even in the forensic field, are familiar with its detailed characteristics, psychiatry has had limited success in educating the legal profession about the subtleties of the illness or the conundrum that delusions can induce such abnormal behaviour in an individual who superficially appears rational and for significant periods of time is effectively sane even though the illness is always lurking there.

Treatment of the persecutory subtype

Treatment is discussed later in this chapter in the section on overall treatment aspects.

Delusional disorder: somatic subtype (monosymptomatic hypochondriacal psychosis)

Modern society, especially in developed countries, is preoccupied with health concerns. While much of this is positive, there is no doubt that many people worry excessively about health matters and a proportion of these show pathological self-concern. This can shade into hypochondriasis, in which there is a persistent conviction of illness in the absence of objective evidence, with misinterpretation of bodily sensations as disease and with inability to accept reassurances. In many cases the individual shows some degree of body image disturbance, sometimes of extreme degree.[32] Usually we think of hypochondriasis as referring to physical complaints, but nowadays it seems that an increasing number of people are also liable to complain of psychological disorder.

Hypochondriasis is common and may be a personality trait, but it can also be an accompaniment to many psychiatric illnesses, both delusional and non-delusional. It is the presenting feature of the somatic subtype of delusional disorder and in different patients we see many varieties of alteration of body image expressed in delusional terms. Certain themes of delusional content tend to predominate and this has meant an unfortunate proliferation of descriptive names scattered across a fragmented literature, leading to many difficulties in conceptualizing the subtype and in separating it from other psychiatric disorders with prominent hypochondriasis. As with all subtypes of delusional disorder, the clinician

must bear in mind the advice already given that, for the diagnosis of delusional disorder, it is the characteristic form of the illness, which is of prime importance, not the content of the delusional beliefs. The hypochondriasis in delusional disorder may superficially resemble that of somatoform disorder, psychotic depression, or obsessive-compulsive disorder, but careful investigation will reveal an illness very different from these.

Clinical features

We shall consider the manifestations of the somatic subtype under four major theme areas:

1 delusions involving the skin;

2 delusions of ugliness or misshapenness (dysmorphic delusions);

3 delusions of body odour or halitosis;

4 miscellaneous.

(a) Delusions involving the skin[33]

In the delusion of skin infestation, the patient insists that he has organisms, usually insects, crawling over the surface of the skin and sometimes burrowing into the skin or under the nails. In most instances he cannot see the creatures, but sometimes there may be graphic descriptions. This may represent a visual hallucination but more usually seems to be a vivid ideational projection.

The delusion of parasites burrowing deeply under the skin is often attributed to worm-like parasites, and internal body sensations or the rippling of small superficial muscles are misinterpreted as evidence of their activities. Sometimes the patient believes that the worms have spread throughout the body or intermittently migrate from place to place.

In the delusion of discrete foreign bodies under the skin or nails, these bodies are occasionally described as inanimate, but generally the patient says they are seed-like or believes that they are parasite eggs. In some individuals this is associated with an irresistible urge to pick, and multiple deep excoriations may result. Such people are sometimes labelled as having 'neurotic excoriations' or factitious disorder, but in fact the picking behaviour is delusionally motivated and is an irresistible urge to stem the invasion of the parasites.

Chronic cutaneous dysaesthesia[34] is an unremitting burning sensation of the skin or mucosae, sometimes generalized but at other times largely confined to complaints of glossodynia or vulvodynia. A minority of these patients appear to have a delusional disorder.

A subgroup of patients with trichotillomania and onychotillomania[35] have delusional illnesses, and the hair-pulling or nail-picking may be part of the attempt to rid themselves of parasites.

In all the above presentations, the delusion and its associated behaviours typically occur in the setting of many well-retained personality features and the patient can often make very clear-cut and apparently rational complaints, convincing the many physicians they attend, at least for a time, that actual physical disease is present. However, no somatic treatment works and the complaining becomes increasingly shrill and unreasonable. The sufferer cannot be persuaded that infestation is not present and often becomes very angry at the perceived incompetence of the dermatologists he has visited.

Usually the story of the infestation is presented in great detail, perhaps involving an original event such as an insect bite. 'Proof' is presented by displaying skin lesions, deformed nails, bald patches, etc. The 'matchbox' or 'pill-bottle' sign, in which the patient produces a small container in which 'insect corpses' or 'eggs' are kept, is typical; the contents nearly always turn out to be dried mucus, skin scrapings, or pieces of lint. Often, there is incessant cleaning of self and surroundings, and repeated demands may be made to local authorities or pest-control agencies for disinfestation of the home. At times bizarre and even dangerous self-treatment is resorted to, such as applying boiling water or corrosive substances to the skin. The more normal part of the psyche is dominated by shame or fear of passing on the infestation, so that progressive social isolation tends to occur, with attendance on doctors as virtually the only outside activity.

(b) Dysmorphic delusions

'Dysmorphophobia', an old term which implies a morbid fear of being deformed, is still sometimes employed to describe cases in this category but should be abandoned since it has been so loosely used to denote both delusional and non-delusional complaints as well as a variety of very different illnesses.[36] In the present context we are considering only cases typical of delusional disorder, which present with a false belief of ugliness or deformity. In some instances there may indeed be some minor deformity, but the complaining and demand for alleviation are out of all proportion and expressed with delusional intensity.

A specific feature is often singled out by the complainer, such as an overlong nose, prominent ears, over-large or undersized breasts, dissatisfaction with the appearance of the genitalia, a skin blemish, or some other. However, in other cases the total body is perceived as abnormal, and there is evidence that a small group of apparent cases of anorexia nervosa and bulimia nervosa may have an underlying delusional disorder.

Many of the patients with dysmorphic delusions go from surgeon to surgeon demanding cosmetic procedures and usually being refused, but if the surgeon does not perceive the illogicality of the complaint an operation may take place. While some successes have been reported, the general consensus is that most cases need psychiatric rather than surgical intervention and unnecessary operations may seriously worsen the mental disorder in the long term.

It is sometimes very difficult to distinguish cases of delusional disorder of somatic subtype from severe somatization disorder, and claims have been made that there is a continuum between these illnesses.[37] The evidence for this is minimal and a diagnostic distinction is essential since treatments of the two disorders are very different.

(c) Delusion of smell or of halitosis[38]

In this category it is often very difficult to distinguish between delusions and hallucinations of smell or taste. The term 'olfactory reference syndrome' is often used to describe olfactory delusions, but in fact it should properly only refer to hallucinatory experiences. Sometimes the deluded patient will say that he or she has not actually experienced the odour, which is usually unpleasant, but 'knows' that it is present because of remarks made by others or their avoidant behaviour. In other cases the stench is described graphically and consistently (like 'burning rubber' or 'faeces') and here a hallucination may be present. There may be no explanation, or else the smell may be attributed to escaping flatus, abnormal sweat secretion, or sinus or dental problems leading to halitosis

etc. As is typical, an unending and escalating search for a physical cure occurs.

(d) Miscellaneous delusional contents

Presumably there is an almost infinite possibility of different themes, but in practice their numbers are somewhat limited. The following have been described.

(i) Dental

Although his dentition is satisfactory, the patient insists that his dental bite is abnormal and obtains repeated corrective treatments from successive dentists, none of which works. This has been termed the 'phantom bite syndrome', and may sometimes be associated with complaints of facial pain for which no physical basis is apparent. There may also be delusional complaints of deformity of the jaw or abnormality of the temporomandibular joint.

(ii) Delusion of transmitting non-sexual diseases

Some patients may be convinced that they are causing illnesses in others (e.g. tuberculosis), and they will cite as evidence, for example, that everyone starts coughing when they enter a room.

(iii) Delusion of sexually transmitted disease[39]

Hypochondriasis is, of course, rampant around the topic of sexually transmitted disease. A subgroup of delusional disorder patients develop the conviction that they have venereal disease, often when there is no evidence of risk-taking behaviour having occurred. In the past syphilis was probably the greatest fear, but nowadays it is usually AIDS. Repeated tests showing negative serology have no reassuring effect. Interestingly, a few cases of actual AIDS have been described in which a delusional illness with hypochondriasis has emerged, usually due to direct effects of the virus on the brain.

Differential diagnosis of delusional disorder: somatic subtype

First, the presence of a significant physical disorder must be excluded (although it is possible for a physical illness and delusional disorder to coexist). The illness must be distinguished from the following:

- paranoid schizophrenia
- substance-related disorders (e.g. itching related to alcohol-related liver failure, cocaine abuse, etc.)
- organic brain disorders
- severe depressive disorder with hypochondriacal delusions
- somatoform disorders, especially body dysmorphic disorder
- obsessive-compulsive disorder
- factitious disorder

Epidemiology

Cases usually present in medical and surgical practices and much less often in a psychiatric context. We have no idea of the frequency because non-psychiatrists make a variety of diagnoses, often untranslatable in psychiatric terms. However, the somatic subtype of delusional disorder is certainly not uncommon, and this is increasingly being revealed as consultation–liaison psychiatry develops.

These cases make a strong impression on physicians and surgeons because of their insistence and unreasonable demands. To date, dermatologists have been most aware of the nature of the delusional complaining and, in some cases, have learned to treat the deluded patients satisfactorily with appropriate medication. Infectious and tropical disease specialists also have an awareness, as do gastroenterologists and some dentists, and they are gradually referring more cases for psychiatric help. Plastic and cosmetic surgeons see a considerable number of cases with dysmorphic delusions, but it is still rather uncommon for them to seek psychiatric consultations. Since the patient with delusional disorder generally refuses to visit a psychiatrist willingly, it is often necessary for us to consult on the other specialists' territory in order to offer practical help and to obtain a better idea of the illness's frequency.

From what we know, the somatic subtype affects both sexes approximately equally and the age of onset may be from late adolescence to extreme old age. The illness is more common in the unmarried, divorced, and widowed.

Course and prognosis

Typically the illness is long term with a tendency to worsen with time. Some patients eventually lapse into a rather apathetic state, and some attempt or commit suicide in chronic despair, but the majority continue to move from doctor to doctor demanding treatment on their own deluded terms.

Treatment of the somatic subtype

Treatment is discussed later in this chapter in the section on overall treatment approaches.

Delusional disorder: jealousy subtype[40]

This is sometimes known as the Othello syndrome, but the term is not recommended as it lacks specificity.

The phenomenon of jealousy

Jealousy can arise in various contexts, but here we shall deal with sexual jealousy. This is a virtually universal human emotion, especially when a rival is attempting to lure away someone's sexual partner. Males and females are equally prone to jealousy but may express it differently; Mullen and Martin[41]suggest that men are mainly concerned with losing the partner whereas women worry about the effect of infidelity on the ongoing relationship.

Broadly there are three levels of jealousy. Normal jealousy is understandable in terms of the situation and the individual's perception of it, and its expression can range from pique to severe rage. How it is expressed is largely related to temperament; some people habitually vent anger with slight provocation and others usually bottle up their feelings. On the whole, men tend to act out their jealous anger more physically.

Neurotic jealousy occurs where the mood and its mode of expression are relatively normal but owing to non-psychotic psychiatric illness, including personality disorder, the reaction is impulsive and excessive. Although the individual is reacting to an overvalued idea rather than a delusion, this type of jealousy can be irrational and quite persistent, and may be expressed dangerously.

Psychotic jealousy, as in delusional disorder, is characterized by a fixed delusional belief which cannot be swayed by reasoned argument or presentation of contrary evidence. This is the most

alarming type since there is no dissuasion and there is an inexorability about the way that the individual accuses, controls, and even stalks the victim. Since the accusations are usually untrue, the latter is bewildered by them, but occasions do arise when a partner actually has been unfaithful and it is then very difficult for the observer to know at first how much of the jealousy is justified and how much is delusional. Eventually the savageness and unreasonableness of the accusations reveal themselves as undoubtedly abnormal, but meanwhile a frightened partner will have suffered enormous abuse and possibly repeated assault.

When does jealousy become pathological?

Jealousy, which appears justifiable is regarded as normal, although perhaps not laudable, and it will usually be accepted by society if its manifestations are not antisocial. Nowadays we increasingly disapprove of jealous violence, whether provoked or not, but in some communities there is still acknowledgement of the *crime passionel*, the crime committed out of jealous love. However, this is an excuse extended only to males, and the jealous woman who commits assault or murder is usually treated more harshly.

Cobb[42] proposed the following as clinical features of pathological jealousy, whether it be neurotic or psychotic.

1 The jealous thinking and behaviour are unreasonable in expression and intensity.

2 The jealous individual is convinced of the partner's guilt but the evidence is dubious to others.

3 A recognizable psychiatric illness is present which could plausibly be associated with abnormal jealousy.

4 In a proportion of cases, the jealous person has habitual personality characteristics of jealousy, suspiciousness, and overpossessiveness.

5 The jealousy persists unduly and reinforces itself.

6 Pathological jealousy is usually focused on one specific person.

In neurotic jealousy, which in some ways resembles obsessive-compulsive disorder, there is high self-awareness of the emotion and sometimes of its irrationality. In delusional jealousy the person is totally at one with the belief, which has come to occupy much of his or her time. Counterargument or contrary evidence is rejected, yet in delusional disorder the individual may be so high functioning that he or she is totally convincing to outsiders and may even be able to brainwash the innocent victim into admission of guilt, a form of *folie à deux*.

The impact of pathological jealousy

Delusional jealousy is anguishing to the sufferer and even more so to the sexual partner who is accused of infidelity. The latter is subjected to escalating emotional abuse, and indignation, protest, and proof of innocence are unavailing. Physical violence, especially by males, is common[43] and in a proportion of cases finally ends in homicide, sometimes followed by the suicide of the perpetrator.[44] Subjected to prolonged threat, many victims are too terrified to speak up, and some become housebound in a vain attempt to prevent accusations of philandering. From time to time the situation reveals itself when the desperate partner attempts suicide and talks to a helping professional when recovering.

Clinical features

The person's belief in the other's infidelity is absolute and brooks no contradiction. There is much associated irritability, despondency, and, in some cases, aggressiveness. An ever-increasing proportion of time is spent searching for spurious 'proofs', and 'clues' are pounced upon and misinterpreted; for example, an innocent stain is declared to be semen. The victim is put through endless interrogations and is kept under constant surveillance.

Paradoxically, when the jealous individual is questioned closely about his or her specific charges, details prove vague, there is dismissiveness, and there are self-justifying repetitive assertions. Evidence is always about to be produced but rarely materializes. Strangely too, the jealous person often avoids taking the action, which might provide definite proof of guilt or innocence, and this passivity in the midst of intensiveness may be evidence of some volitional defect.

As noted, delusional jealousy is more commonly reported in men, but this is probably an artefact due to their greater likelihood of violence. Also, there is a link between chronic alcohol abuse, as well as amphetamine and cocaine abuse, and delusions of jealousy, and it is known that these substance abuses are more common in males.

Epidemiology

The overall prevalence of abnormal jealousy and the specific prevalence of the delusional disorder subtype are unknown and, because of fear on the part of the victim, both are invariably underreported. Both heterosexual and homosexual cases occur, and family patterns of jealous behaviour have been described, but there is little evidence of direct inheritance.

Course and prognosis

The condition may appear gradually or suddenly, but even when the onset seems rapid there may have been a previous period of rumination and perplexity of varying duration. When the delusion crystallizes the perplexity vanishes and the patient is then totally sure of his belief.

Delusional jealousy is typical in being chronic and often lifelong. Without treatment the prognosis is poor and the danger to the victim is ever present. Most patients refuse psychiatric help and unfortunately may only receive it after incarceration for a violent crime.

Differential diagnosis

The illness must be distinguished from the following:

- actual marital or sexual problems, including spousal infidelity
- mental handicap, where a simple-minded person may develop a 'crush' and be unable to understand that the other person does not reciprocate, or else enters into a sexual relationship and cannot cope with the partner's motives and behaviours
- schizophrenia, especially of the paranoid type
- major mood disorder with delusions, either depressive or manic
- personality disorder, especially of the paranoid, antisocial, borderline, histrionic, and narcissistic types
- obsessive-compulsive disorder

- substance abuse (which may complicate any of the other differential diagnoses)
- organic brain disorders, including dementias and some epileptic disorders
- sexual dysfunction may lead to fears that a normal partner is seeking satisfaction elsewhere

It should be noted that an important part of the diagnostic process, an accurate collateral history, may be impossible to obtain in cases of delusional jealousy because of the victim's fears.

Forensic complications

In cases of identified physical abuse in a relationship one possibility that must always be considered is delusional disorder of jealous type. Severe assault and even murder are not uncommon, and the physician has a duty to warn and protect the partner if these dangers seem real, perhaps divulging confidential information if necessary. Of course the patient denies that his beliefs are unjustified and may present his case more convincingly than the terrified victim can. If involuntary committal is necessary it may be very difficult to sustain, partly because of the individual's ability to maintain a pseudonormal facade and often because he or she threatens litigation.

Occasionally, cases of stalking, usually of females by deluded males, are jealousy related and the victim is nearly always well aware of the stalker's identity in these instances.

Treatment

This will be discussed when considering the overall treatment approach to delusional disorder. If successful treatment can be achieved, the couple may require considerable psychotherapy and counselling to re-establish a trusting and fear-free relationship.

Delusional disorder: erotomanic subtype[45–48]

In erotomania the individual has strong erotic feelings towards another person and has the persistent, unfounded belief that this other person is deeply in love with him or her. The belief is usually delusional, though a small number of non-delusional cases have been reported. Occasionally the imagined lover does not actually exist, but more often he or she is a real person who is unaware of the situation. The phenomenon is often referred to as de Clérambault's syndrome, but this usage is obsolete and can be misleading since it is used to describe erotomanic manifestations in a number of different mental illnesses.

In the older literature it was claimed that erotomanic delusions were largely confined to women, especially isolated and frustrated elderly spinsters, but more and more cases of male erotomania are being reported nowadays.[49] In both sexes the majority of cases described involve heterosexual emotions, but homosexual erotomania is now well documented in both males and females.

Clinical features

The patient yearns for another person and has the unshakeable belief that these feelings are reciprocated. The person is often socially unattainable, may be of higher social status, and can be a celebrity. There has rarely been close contact and the love object will usually be unaware of the situation, but despite this the patient believes that the other initiated the imagined relationship, often with covert signals or utterances. Many patients experience strong erotic, even orgasmic, feelings, but some insist that the 'relationship' is platonic and that the other person is maintaining a non-sexual attitude of watchful protectiveness.

In many instances the patient makes no attempt to get in touch with the love object, perhaps writing letters or buying gifts but not sending them. When given a chance to make actual contact he or she will frequently avoid doing so and will make spurious excuses such as not wanting to offend the other person's spouse. In those cases where the patient does attempt contact, false reasons are presented to explain the almost inevitable rejection that results.

Since this is erotomania in the setting of delusional disorder, the illness will have the typical form of a tightly knit delusional system with preservation of relatively normal personality features and with greater or lesser ability to continue functioning in society. There is often enough insight or inhibition present for the patient to keep the delusional beliefs concealed. However, at times he or she may be profoundly angered by being 'inexplicably' rejected and may act this out, occasionally dangerously. This is more likely to occur in males.

The onset of erotomania can be gradual or apparently sudden. Hallucinations are sometimes present but are not prominent, although the patient may be encouraged by 'hearing' the other person express passionate feelings. Occasionally, the presence of tactile hallucinations leads the patient to believe that a lover has paid a visit during the night (sometimes picturesquely referred to as the 'incubus syndrome').

Diagnosis and differential diagnosis

Many covert examples necessarily go unrecognized and so there is a bias towards diagnosis of cases with some sort of acting out behaviour. Otherwise the most common situation is one where the patient, after years of silent suffering, becomes unhappy enough to be treated for depression and then, during sympathetic history-taking, lets the delusional belief slip out. There is often much accompanying anguish and perhaps anger, and of course the beliefs are regarded as indisputable. Obviously, if the patient has been very secretive, a confirmatory history may be impossible to obtain. In married patients the spouse may be totally unaware of delusions which have lasted for years.

The following disorders may be associated with secondary erotomanic features:

- Schizophrenia, especially paranoid, in which the erotomania coexists with other delusions, florid hallucinations, and more widespread thought disorder.
- Major mood disorder, in depressive or manic phases.
- Organic brain disorders, including epilepsy, post-head-injury states, following long-term substance abuse, senile dementia, and possibly as a side effect of steroid treatment.
- Mental handicap, in which misunderstanding occurs regarding another's feelings or intentions. However, we must remember that the mentally handicapped are liable to sexual abuse and we must not unthinkingly dismiss sexually laden remarks that they may make about other individuals. Conversely, we must also remember that mental handicap can coexist with psychotic disorders and delusional expressions.

- Delusional misidentification syndrome has occasionally been described with erotomanic features.
- Non-delusional erotomanic beliefs may emerge in unstable individuals, sometimes complicating transference in the course of psychotherapy. If associated with histrionic traits there may be florid acting out, but the beliefs do not have the qualities of a delusion.

Epidemiology

Nothing is known of the frequency of erotomania in general, or of the erotomanic subtype of delusional disorder. As will be noted below, the more dangerous aspects of the illness are proving to be not uncommon.

Course and prognosis

Without treatment this is a chronic illness, which is likely to worsen gradually over time.

Forensic complications[50]

Males who irrationally act out their erotomanic delusions are usually diagnosed as schizophrenic but some prove to have delusional disorder. Often the overt behaviour is in the nature of harassment, but even without violence the individual's persistent intrusiveness and incorrigibility can be thoroughly alarming to the victim, who is bewildered by the situation and by the other's accusations of duplicity.[51]

Severely aggressive behaviour can lead to assault, kidnapping, and even murder, sometimes of the love object but at other times of an acquaintance who is viewed as a rival. A manifestation, which has gained much recent publicity is that of victim stalking, and in a considerable number of cases the victim has no idea who is carrying out the stalking.

While women are generally less prone to aggressive acting out of their delusions, they may sometimes demonstrate their false beliefs in devastating ways. For example, a deluded woman may claim publicly that a physician, counsellor, or teacher has demonstrated strong erotic feelings towards her. This belief may be the result of a delusional memory. If she has an undeteriorated personality, is coherent, totally believes her own story, and presents it with typical vehemence and persistence, it may be virtually impossible to persuade the public and the authorities that the accusations are totally false. Any professional person dealing with deluded patients must be aware of abnormal transference emotions that may arise in the patient during treatment, usually of a heterosexual nature but sometimes homosexual. Great circumspection is then required and the therapist must immediately seek collegial (and possibly legal) help.

Treatment

Treatment is discussed later in the section on overall treatment of delusional disorder.

Delusional disorder: grandiose subtype[23]

This is the least well-described variant of delusional disorder, not surprisingly in view of its nature. An individual who is habitually elated, even exalted, and who may believe himself or herself rich or powerful is unlikely to seek help, especially psychiatric help. If he or she remains sufficiently high functioning to remain in the community, the delusions may be undetected; indeed some people capitalize on their beliefs by belonging to fringe organizations, apocalyptic religious groups, or doomsday sects. Sometimes these groups develop malignant qualities, especially under a deluded but charismatic leader, and one cannot minimize the dangerous qualities of the forceful megalomaniac whose grandiosity is alloyed with persecutory anger. Like-minded and impressionable people are readily drawn in and a kind of mass shared psychotic disorder may result; comparisons with Nazi Germany are not inapt.

Clinical features

This disorder often only becomes apparent over time and with observation. The few cases that we identify tend to fall into two categories. The first are those whose state of bliss is so profound that they totally neglect self-care. The rest are usually seen in custody after they have committed an offence under delusional influence. The characteristic underlying features of a delusional disorder have already been well described.

Differential diagnosis

The illness must be distinguished from the following.

- Mania, in which grandiosity is associated with euphoria, overactivity, and, at times, irritability and suspiciousness. As the mood is often volatile and/or phasic, so the grandiose features are unstable.
- Schizophrenia in which there is marked incongruity between ecstatic affect and relative thought poverty.
- Organic brain disorders, especially affecting the prefrontal cerebral lobes, which cause labile mood, disinhibited behaviour, and some degree of cognitive deficit. Cerebral syphilis (general paralysis of the insane) used to be the best-known exemplar.
- Antisocial personality disorder in which the individual feels above the law and may express grandiose ideas and behaviours. In these cases one finds evidence of lifelong impulsivity, lack of remorsefulness, and usually a long history of delinquency.

Epidemiology

We have virtually no information. The illness can occur in either sex and apparently at any age from adolescence onwards. It may appear gradually or suddenly.

Course and prognosis

As far as we know, the grandiose subtype is as chronic and unremitting as the other subtypes of delusional disorder. For many years the presence of grandiosity in any psychiatric disorder has been regarded as a bad prognostic factor. In delusional disorder this may be so, because a grandiose delusional system is particularly likely to be associated with spurning of treatment. Even if treatment begins to be effective, the abandoning of highly pleasurable beliefs may not be welcomed by some patients.

There may be forensic complications if grandiose delusions are acted out.

Treatment

This is discussed later when considering the overall treatment of delusional disorder.

Delusional disorder: mixed and unspecified subtypes

There is little to be added regarding these categories. In DSM-IV it is accepted that more than one delusional theme can exist side by side so that, provided that the form of the illness is that of delusional disorder, it is acceptable to have combined themes of, for example, hypochondriasis and persecution, persecution and grandiosity, erotomania and jealousy, etc.

The unspecified type is a residual category and again the illness must have the form of a delusional disorder, but the delusional content is one other than those specifically listed. No reliable data exist on either the mixed or the unspecified subtypes.

Other disorders with persistent delusions

As mentioned previously, ICD-10 has a category for 'other persistent delusional disorders' (F22.8), but this is too loosely worded to delineate any coherent clinical entity. Paraphrenia, as has been noted before (p. 283) is regarded by some, including the present author, to be a candidate for inclusion in an expanded category of delusional disorders but currently is not receiving official notice. Therefore, the only condition we shall consider here is delusional misidentification syndrome.

Delusional misidentification syndromes (DMIS)[52]

The abilities to recognize individual faces and to discriminate between different faces are fundamental human processes and normally we are extraordinarily adept at them. Changes in a familiar facial appearance can be unsettling and even frightening, in both children and adults. A great deal of sophisticated neurophysiological and neuropsychological investigation has been carried out on normal and abnormal face-recognition processes.

A number of clearly defined neurological disorders are associated with very specific abnormalities of face recognition.[53] Here we shall emphasize those cases in which a delusion of misidentification is the principal symptom of the disorder and in which the form or structure of the illness is in many ways similar to that of delusional disorder. These are the delusional misidentification syndromes. However, it is important to note that superficially similar presentations may occur as secondary features in cases of schizophrenia, severe mood disorder, or dementia, and in these we refer to a misidentification phenomenon rather than syndrome.

(a) Clinical features

There are four main variants of DMIS:

1 the Capgras syndrome, in which the individual falsely perceives that someone in his environment, usually a close relative or friend, has been replaced by an almost exact double

2 the Frégoli syndrome, where the patient believes that one or more individuals have altered their appearances to resemble familiar people, usually to persecute or defraud him or her

3 intermetamorphosis, in which the patient believes that people around have exchanged identities so that A becomes B, B becomes C, and so on

4 the syndrome of subjective doubles, where the patient is convinced that exact doubles of him- or herself exist, a kind of *Doppelgänger* phenomenon

Additional alternative forms have been described and sometimes features of more than one variant occur in an individual case, especially in the subjective doubles phenomenon.

Although the patient is convinced of the deception, he is often aware that something is wrong and that the replacements are subtly incorrect. Many sufferers are extremely distressed and frightened and fear impending harm. In some cases, they may become enraged and attack the 'impostor' with considerable violence. Their belief is of delusional intensity and they usually cannot be dissuaded by argument or by demonstration of contrary proof.

It is of particular interest that the misidentification is not indiscriminate but involves a limited number of usually familiar people. In some cases, substitution also involves places and objects. An admixture of depersonalization and derealization is not unusual, especially in the earlier stages.

(b) Classification

Delusional misidentification syndromes have not been included in DSM-IV or ICD-10, though if misidentification occurs in the setting of a psychotic illness like schizophrenia it is regarded as a feature of that illness. However, if it is the principal aspect of a psychosis it should be regarded as a disorder *sui generis*. In those cases where there is a discrete delusional system occurring in clear consciousness and within a relatively intact personality, it would seem logical to assign it as a new subcategory within Delusional Disorder (DSM-IV) or Persistent Delusional Disorders (ICD-10). In cases where organic brain disease is prominent, the proper assignment would be to Mental Disorders due to a General Medical Condition (DSM-IV) or Organic, Including Symptomatic, Mental Disorders (ICD-10).

(c) Diagnosis and differential diagnosis

The diagnosis is based on recognition of the delusional nature of the belief, the accompanying agitation, and uncharacteristic behaviours, possibly including violent attacks on others. A full neurological investigation is mandatory.

Differential diagnosis includes:

- Schizophrenia
- delusional disorder, persecutory subtype
- major mood disorder with delusions
- organic brain disorder
- substance abuse disorders.

(d) Epidemiology

The frequency of DMIS is unknown but an increasing number of cases are being reported. The disorder occurs in both sexes and across a wide age range, but particularly in the middle-aged and elderly.

(e) Aetiology

When DMIS was first recognized, attempts were made to explain its symptoms on psychodynamic grounds. Such theories have been almost totally undermined of late by the increasing recognition of significant brain pathology in a high proportion of cases.

Nowadays we have many reports on brain dysfunction in DMIS[18,54,55] but large case series are lacking. There is some consensus that abnormalities of the right cerebral hemisphere, especially temporoparietal, are especially likely to be present but are not

inevitable, and lesions in other cerebral locations have been found. The lesions can be regarded as causal of specific abnormalities of face recognition, a function mediated especially in the right hemisphere. Also, there appears to be dissociation of sensory information from its normal affective accompaniment and failure of suppression of inappropriately repetitive behaviours (also a right-sided function). These last two features are very typical of delusional disorders in general.

We know little of the biological substrate of delusional symptoms, but one proposal is that there may be a dysfunction of the limbic–basal ganglia mechanisms, with particular emphasis on dopamine overactivity. In DMIS there seems to be a breakdown in integration of information between the right parietotemporal cortex, the limbic system, and certain basal ganglia, resulting in the specific misidentification, associated with inappropriate emotion and inability to suppress abnormal thoughts and behaviours. Very tentatively it is possible to postulate a complex brain mechanism which normally integrates sensory and affective impulses and downregulates repetitive behaviour, and whose malfunction results in delusional beliefs, altered judgement, overintense mood, and inability to change or develop insight. The particular delusional content might be determined by the specific site within the mechanism at which the brain dysfunction occurred.[56]

Although the above is simply an attempt at a paradigm, it is also a model with potential for the study of delusional and concomitant phenomena by modern neurobiological investigative methods. It also allows us to conjecture about the general similarities and specific differences between DMIS and delusional disorder.

(f) Course and prognosis

Although DMIS may appear insidiously, it not infrequently appears relatively suddenly in a previously normal individual, presumably related to the underlying cerebral pathology. Where brain damage is substantial, the prognosis is that of the brain disease. If the brain dysfunction is more subtle and does not remit, the delusional symptoms may become chronic. Forensic complications may occur if the patient becomes violent,[47] and a small number of murders by patients have been reported in DMIS.

(g) Treatment

Acute treatment may involve sedation and antipsychotic medication. Ongoing treatment is by maintenance doses of an antipsychotic with possible addition of an anticonvulsant. Psychological counselling may be beneficial as the patient improves.

Folie à deux: a phenomenon which may accompany illnesses with delusions[57,58]

This phenomenon is listed as a psychiatric disorder in DSM-IV (Shared Psychotic Disorder, 297.3) and in ICD-10 (Induced Delusional Disorder, F24) but there is a conceptual difficulty in regarding it as a psychotic illness in its own right, as will be discussed shortly.

Folie à deux is a venerable term used to describe a situation in which mental symptoms, usually but not invariably delusions, are communicated from a psychiatrically ill individual (the 'primary patient') to another individual (the 'secondary patient') who accepts them as truth. As noted, DSM-IV and ICD-10 refer to this by different names and there have been several confusing changes of official terminology over the years. The older name, which is used as an alternative by DSM-IV, is well known to most psychiatrists and is used here by preference. However, *à deux* may sometimes be a misnomer since several people can be involved, and then we read of *folie à trois*, *folie à plusieurs*, *folie à ménage*, etc.

Taking the dyad as the classical situation, the two people are usually closely associated, especially husband–wife, siblings, or parent–child, and often live in social isolation. The content of the shared belief depends on the predominant delusion(s) of the primary patient and can include convictions of persecution, delusional parasitosis, belief in having a child who does not exist, misidentification delusions, and many others. There have been descriptions of shared persecutory and apocalyptic beliefs in quasi-religions and cults led by a charismatic leader with gullible followers. In many shared delusional examples there is a sense of antagonism from 'them' who may be defined or who may be what Cameron[59] referred to as the 'paranoid pseudocommunity', the hovering 'they' who carry out oppression which is evident to the sufferers but not to normal others.

Once thought rare, *folie à deux* has been increasingly described in the literature. Milder cases may not be recognized and, also, many delusional people strive to avoid psychiatric referral; collusion between primary and secondary patient in this has been noted. The physician should be aware of *folie à deux* and never overlook it.

'Shared psychotic' and 'induced delusional' both imply that two members of the dyad are psychotic. In delusional disorder the primary patient is psychotic but the recipient of the beliefs usually is not. Most often, the latter is highly impressionable and dependent and adopts the false beliefs because of their intense and unceasing transmission by the primary patient. Social isolation, accentuated by induced mistrust of 'them', discourages adequate reality testing. Thus one can say that the content of the secondary patient's false belief derives from psychotic thinking, but he or she is not usually psychotic.

Nearly all cases of *folie à deux* are reported in association with schizophrenia, delusional disorder, severe depressive disorder with delusions, or early dementia, but it is probable that the condition also sometimes coexists with non-psychotic illnesses such as obsessive-compulsive disorder, somatoform disorder, and histrionic–dissociative personality disorder, in which the beliefs are intensely held and communicated but are not delusional. This makes the DSM-IV and ICD-10 terms even less appropriate.

Subtypes

In practice, the great majority of cases are of the type described, sometimes known as *folie imposée* because the belief is impressed on the secondary patient.

However, one occasionally sees an alternative presentation, the so-called *folie simultanée*, in which two predisposed people develop illnesses with delusions and through long and over-close association come to share identical false beliefs. This is said to be likelier when there is a genetic link (for example two unmarried siblings) or when older people have lived together in considerable isolation for many years.

Classification

Folie à deux is included with schizophrenia and other psychotic disorders in DSM-IV and with schizophrenia, schizotypal, and delusional disorders in ICD-10, and is regarded as a separate psychotic illness. It would be better to treat it as an important clinical phenomenon, which may be associated with other identifiable mental disorders. However, rather than encourage pedantic argument, it is best to alert the clinician to its existence, its frequency, and, as will be seen, its importance.

Diagnosis

The phenomenon is recognized by the identical nature of the two individuals' false beliefs, their gross affective investment in these, and their refusal to accept alternatives even when overwhelming proof is presented. Careful history-taking will usually readily distinguish the primary from the secondary patient, but this can sometimes prove difficult. In the much less common *folie simultanée*, the distinction is largely irrelevant.

Epidemiology

This is unknown except that, by definition, it will occur most in association with an illness characterized by delusional beliefs or severely overvalued ideas, especially under isolated conditions, and it is by no means rare. It is extremely important for the clinician to be on the lookout for it. He or she may be convinced that a patient is expressing delusional ideas but be thoroughly perplexed when an apparently rational relative unswervingly supports these. This can lead to serious mismanagement of the case. Conversely, recognition of *folie à deux* may solve a baffling diagnostic problem and result in appropriate care for both individuals.

Aetiology

Folie à deux appears to arise from a combination of the following:

1 innate impressionability and marked dependence on the primary patient
2 personality traits such as suggestibility, low initiative, poor reality testing, etc. in the secondary patient
3 in some cases, low intelligence in the secondary patient
4 the intensity of the abnormal beliefs expressed by the primary patient
5 the length of time over which the abnormal beliefs have been imposed
6 the degree of social isolation

Treatment

In *folie imposée* the logical approach is to identify the primary patient and treat his or her mental disorder adequately. It may also be helpful to separate the two individuals for a time, for example by hospitalizing the primary patient. With both people, every attempt must be made to reduce social isolation and to reintroduce them to reality. If the primary patient's delusions improve with treatment, the secondary patient's beliefs usually also improve. It is rarely appropriate to treat the secondary patient with antipsychotic medication, although this is sometimes mistakenly done.

In *folie simultanée*, both patients require neuroleptic treatment.

Theoretically, treatment is straightforward but in practice it can be problematic. For example, the primary patient with delusional disorder often resists psychiatric help, and subterfuge and resistance by both individuals is common. In group situations, for example a cult, this resistance is likely to be widespread and intense, and will be justified by the participants in terms of religious, social and legal rights, which they claim are being suppressed. The propagators should be separated from the recipients as much as possible, and the treatment team has to expend much time and diplomacy in gaining some confidence and a degree of cooperation. Direct challenge of the beliefs in any shared delusional situation is usually totally counterproductive.

Mass suicide is a reported outcome of shared delusions in some cult situations and any danger of this must be countered with great urgency.[60]

Treatment of delusional disorder

There are special aspects to the treatment of delusional disorder, which need emphasis, especially since many clinicians are unfamiliar with them. The first is that we must realize that delusional disorder is indeed treatable, despite frequent pessimistic statements to the contrary. In fact, the greatest difficulty is not treatment responsiveness, but persuading the patient that he needs psychiatric help because his delusions militate against this. With careful diagnosis and an approach that encourages the patient to cooperate, delusional disorder can respond to treatment, which in all cases will primarily be by neuroleptics.

General aspects of the treatment of delusions[61]

We usually aim to treat the illness of which the delusion is a part, but there is good evidence that delusions themselves, as well as hallucinations, can be considerably modified by a psychological approach. In severe psychotic illnesses the initiation of psychological treatment usually has to follow the initial controlling of serious symptoms with medications or, on occasion, electroconvulsive therapy. Thereafter, a cognitive behavioural approach or, to a much lesser extent, conventional psychotherapy can help the individual to reduce preoccupation with false beliefs, become less isolated, and reorientate towards reality.[62,63] But there is no good evidence that psychological methods by themselves can completely eliminate established delusions.

Since many illnesses are associated with delusions we have to tailor the psychopharmacological approach to suit each particular condition. In delusional disorder, the schizophrenias, and schizoaffective disorder, neuroleptics are the mainstay, with antidepressants, mood stabilizers, and electroconvulsive therapy sometimes playing subsidiary roles.

The rate of symptom response to treatment in a psychotic disorder is not uniform. For example, hallucinations often resolve quite quickly, but delusions can be much more persistent. Despite vigorous treatment they can last for many months and, in some patients, never fully remit. If the patient continues to be deluded, non-compliance with treatment is likely to be present, especially in delusional disorder where the individual is often expert at concealing his or her lack of cooperation. At present it appears that all the subtypes of delusional disorder are potentially responsive to treatment. If treatment fails, always consider non-compliance before abandoning the current approach.

Treatment approaches

Since so many delusional disorder patients actively resist seeing a psychiatrist, it is best to see them in a non-psychiatric setting if possible, for example in the office of the referring specialist or family physician. The psychiatrist who treats cases of delusional disorder needs much patience and tact, and it is common to spend one or more sessions first gaining the individual's confidence and finally persuading him to give a psychotropic medication a trial. Many of them argue vehemently and with well-organized pseudologic against the premise that they are mentally ill and use all kinds of sophistry to deny the need for a neuroleptic, but a calm and persistent approach will gain cooperation in a good proportion of cases.

Whatever neuroleptic is prescribed it is essential to begin with the lowest effective dose. This dose is only raised if required and then very gradually to avoid side effects which are guaranteed to prompt withdrawal from treatment. The patient should be seen at least once a week as an outpatient in the initial stages. Inpatient treatment is not often indicated, although forensic cases will nearly always be seen in an institutional setting.

With a positive treatment response it is not unusual to see minor improvements in a few days, such as reduced agitation, a slight increase in well-being, improved sleep, and a little less preoccupation with the delusion. On average it is about 2 weeks before the delusional system is significantly ameliorated, but in some patients this may take 6 weeks or more.

Quite often if this degree of improvement occurs the patient decides that there is no further need of treatment and stops it. Within days or weeks there is an inevitable return of the delusion with its accompanying agitation and preoccupation. It is then that the psychiatrist must be available to encourage resumption of medication. Although at this stage the patient still believes in his delusions, the experience of improvement followed by relapse makes a deep impression and, given trust in the therapist, often leads to long-term cooperation.

It is striking that good recovery is often relatively rapid and can be surprisingly complete, even when the illness has been present for many years. Some patients return to a considerable degree of intrapsychic, interpersonal, and occupational functioning, with little evidence of the personality disorder that is supposed to be so prevalent in delusional disorder. Also, many patients require surprisingly little counselling or psychotherapy in resuming a reasonable life, although these should always be available if required. Such results suggest that this profound illness may be due to a relatively circumscribed brain abnormality and also that, in some cases at least, a very insidious onset may cause initial symptoms, which mimic personality disorder.

In most cases, treatment has to be continued for an indefinite period since delusional disorder is potentially a lifelong illness. Naturally the drug dosage should be the lowest, which keeps symptoms under control and this maintenance dose is often very low indeed. Perhaps up to one-third of patients can eventually be weaned successfully from medication, but we have no means of predicting who these will be, so that any reduction in effective treatment must be carried out with extreme caution. Sadly, a proportion of relapses are due to injudicious withdrawal of treatment by a physician and we must assume the need for treatment to be permanent unless proved otherwise. It is interesting that successfully treated patients, whether on maintenance drugs or not, keep a lookout for subsequent recurrences themselves and may report that tension-inducing circumstances provoke some reappearance of symptoms. Such patients may then request to have their medication resumed or increased.

There is no necessary correlation between acquired insight into the desirability of taking one's medication and true insight into the illness itself. Many patients never fully accept the psychotic nature of their experience, but as long as they are benefiting from treatment and are functioning reasonably there is nothing to be gained from challenging them on this. If, despite treatment, the delusions remain intrusive, cognitive behavioural therapy and counselling should be available,[62,63] but intrusive psychotherapy must be avoided.

Overall, the best-attested treatment results refer to the somatic subtype, with smaller literatures on the erotomanic, jealousy, and persecutory subtypes, and virtually nothing on the grandiose form. A wide variety of neuroleptics have been reported, but for a considerable time pimozide, a diphenylbutylpiperidine, was the drug of first choice.[64] Recently a study by Manschrek and colleagues[65] has apparently shown that some of the newer atypical antipsychotics can produce comparable results. Antidepressants and benzodiazepines are ineffective as first-line treatments and monoamine oxidase inhibitor antidepressants are absolutely contraindicated.

Currently, the best estimate of treatment outcomes is that, if the diagnosis is correct, the patient compliant, and the treatment adequate, recovery (defined as 'return to full function with total or near-remission of symptoms') occurs in approximately 50 per cent of all patients, and a further 30 per cent will show significant but less complete improvement. An unknown, but considerable, proportion of those who show no improvement can be assumed to have been non-compliant. Although mostly based on non-blind trials, these figures are culled from a worldwide literature, which does show considerable consistency.

Recognition and treatment of postpsychotic depression

Ten per cent or more of delusional disorder patients whose illness responds to neuroleptics experience significant degrees of mood disorder during recovery, sometimes very severe depression with suicidal risks.[66] This has also been noted in recovering schizophrenics. Various explanations have been proposed such as a medication side effect or perhaps the achievement of painful insight. The most likely reason is neurochemical, due to rapid changes in neurotransmitter balance.

If the neuroleptic is withdrawn, the depression tends to improve but the delusions return. Therefore the proper approach is to continue with the minimum effective dose of neuroleptic and to add an antidepressant drug in an effective dose. Occasionally, in extremely severe cases, electroconvulsive therapy is required. Subsequently the neuroleptic is continued but, in most instances, the antidepressant can be gradually reduced.

All cases of delusional disorder should be monitored for the possible emergence of mood symptoms during recovery and treatment immediately instituted if necessary. If suicidal symptoms appear, inpatient admission is highly recommended.

Conclusions

Paranoia/delusional disorder is unique in psychiatry in that it is virtually a newly discovered illness, yet much of the fundamental descriptive work was done a century or more ago. This long hiatus means that most practitioners have little knowledge or experience of the disorder, and the few who are aware of it usually only see a small part of the fabric. The dermatologist treats a case of delusional parasitosis, the cosmetic surgeon has an impossible patient wth a dysmorphic delusion, the lawyer does not know what to do with a totally unreasonable litigant, the police officer has to deal with a jealous murderer or an erotomanic stalker, and the personnel officer has an employee who is convinced his fellow workers are persecuting him, etc. How can we draw all this scattered material together and add it to the psychiatric literature to make a whole cloth? The answer is largely by consciousness raising and education.

Kendler, an authority in this field, has said, 'The paranoid disorders may be the third great group of functional psychoses, along with affective disorder and schizophrenia'.[67] If he is correct, it is imperative that we hone our diagnostic and treatment skills in order to improve the help we might offer to delusional disorder sufferers and to facilitate research which is so badly needed.

Further information

Bhugra, D. and Munro, A. (eds.) (1997). *Troublesome disguises: underdiagnosed psychiatric syndromes*. Blackwell Science, Oxford.

Cash, T.F. and Pruzinsky, T. (eds.) (1990). *Body images: development, deviance and change*. Guilford Press, New York.

Freeman, D., Bentall, R., and Garety, P. (2008) *Persecutory delusions: assessment, theory and treatment*. Oxford University Press, Oxford.

Garety, P.A. and Hemsley, D.R. (1994). *Delusions: investigations into the psychology of delusional reasoning*. Maudsley Monograph No. 36. Oxford University Press, Oxford.

Manschrek, T.C. (ed.) (1992). Delusional disorders. *Psychiatric Annals*, **22**, 225–85.

Munro, A. (1999). *Delusional disorder*. Cambridge University Press, Cambridge.

Sedler, M.J. (ed.) (1995). Delusional disorders. *Psychiatric Clinics of North America*, **18**, 199–425.

Sharma, V.P. (1991). *Insame jealousy*. Mind Publications, Cleveland, TN.

References

1. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
2. World Health Organization. (1992). *International statistical classification of diseases and related health problems*, 10th revision. WHO, Geneva.
3. Capgras, J. and Reboul-Lachaux, J. (1923). L'illusion des 'sosies' dans un délire systématisé chronique. *Bulletin de la Societé Clinique de Médecine Mentale*, **2**, 6–16.
4. Munro, A. (1997). Paraphrenia. In *Troublesome disguises: underdiagnosed psychiatric syndromes* (eds. D. Bhugra and A. Munro), pp. 91–111. Blackwell Science, Oxford.
5. Kraepelin, E. (1896). *Lehrbuch der Psychiatrie* (5th edn). Barth, Leipzig.
6. Jaspers, K. (1963). *General psychopathology* (7th edn) (trans. J. Hoenig and M. Hamilton). Manchester University Press, Manchester.
7. Kretschmer, E. (1927). *Der sensitiver Beziehungswahn* (2nd edn). Springer, Berlin.
8. Garety, P.A. and Hemsley, D.R. (1994). *Delusions: investigations into the psychology of delusional reasoning*. Maudsley Monograph No. 36. Oxford University Press, Oxford.
9. Mullen, P. (1979). Phenomenology of disordered mental function. In *Essentials of postgraduate psychiatry* (eds. P. Hill, R. Murray, and G. Thorley), pp. 25–54. Academic Press, London.
10. Butler, R.W. and Braff, D.L. (1991). Delusions: a review and integration. *Schizophrenia Bulletin*, **17**, 633–47.
11. Flaum, M., Arndt, S., and Andreasen, N.C. (1991). The reliability of 'bizarre' delusions. *Comprehensive Psychiatry*, **32**, 59–65.
12. Farmer, A.E., McGuffin, P., and Gottesman, I.I. (1987). Searching for the split in schizophrenia: a twin study perspective. *Psychiatric Research*, **13**, 109–18.
13. Kendler, K.S. and Gruenberg, A.M. (1982). Genetic relationship between paranoid personality disorder and the 'schizophrenic spectrum' disorders. *The American Journal of Psychiatry*, **139**, 1185–7.
14. Kendler, K.S. (1987). Paranoid disorders in DSM III, a critical review. In *Diagnosis and classification in psychiatry* (ed. G.L. Tischler), pp. 57–83. Cambridge University Press, Cambridge.
15. Michael, A., Mirza, S., Mirza, K.A.H., *et al.* (1995). Morbid jealousy in alcoholism. *The British Journal of Psychiatry*, **167**, 668–72.
16. Feinstein, A. and Ron, M.A. (1990). Psychosis associated with demonstrable brain disease. *Psychological Medicine*, **20**, 793–803.
17. Almeida, O.P., Howard, R., Förstl, H., *et al.* (1992). Late paraphrenia: a review. *International Journal of Geriatric Psychiatry*, **7**, 543–8.
18. Flint, A.J., Rifat, S.L., and Eastwood, M.R. (1991). Late-onset paranoia: distinct from paraphrenia? *International Journal of Geriatric Psychiatry*, **6**, 103–9.
19. Connell, P.H. (1958). *Amphetamine psychosis*. Maudsley Monograph No. 5. Chapman & Hall, London.
20. Morimotom, K., Miyatake, R., and Nakamura, M., (2002). Delusional disorder: molecular genetic evidence for dopamine psychosis. *Neuropsychopharmacology*, **26**, 794–801.
21. Reilly, T.M. and Batchelor, D.H. (1991). Monosymptomatic hypochondriacal psychosis and AIDS (letter). *The American Journal of Psychiatry*, **148**, 815.
22. Gorman, D.G. and Cummings, J.L. (1990). Organic delusional syndrome. *Seminars in Neurology*, **10**, 229–38.
23. Munro, A. (1988). Delusional (paranoid) disorder: etiologic and taxonomic considerations. The possible significance of organic brain factor in etiology of delusional disorders. *Canadian Journal of Psychiatry*, **33**, 171–4.
24. Munro, A. (1998). *Delusional disorder*. Cambridge University Press, Cambridge.
25. Kennedy, H.G., Kemp, L.I., and Dyer, D.E. (1992). Fear and anger in delusional (paranoid) disorder: the association with violence. *The British Journal of Psychiatry*, **160**, 488–92.
26. Ungvari, G.S. and Hollokoi, R.I.M. (1993). Successful treatment of litigious paranoia with pimozide. *Canadian Journal of Psychiatry*, **38**, 4–8.
27. Rowlands, M.W.D. (1988). Psychiatric and legal aspects of persistent litigation. *The British Journal of Psychiatry*, **153**, 317–23.
28. Goldstein, R.L. (1987). Litigious paranoids and the legal system: the role of the forensic psychiatrist. *Journal of Forensic Sciences California*, **32**, 1009–15.
29. Freckelton, I. (1988). Querulent paranoia and the vexatious complainant. *International Journal of Law and Psychiatry*, **11**, 127–43.
30. Reavis, D.J. (1995). *Ashes of Waco: an investigation*. Simon & Schuster, New York.
31. Weightman, J.M. (1984). *Making sense of the Jonestown suicides: a sociological history of the people's temple*. Edwin Mellen Press, Lewiston, NY.
32. Pruzinsky, T. (1990). Psychopathology of body experience: expanded perspectives. In *Body images: development, deviance and change* (eds. T.F. Cash and T. Pruzinsky), pp. 170–89. Guilford Press, New York.

33. Van Moffaert, M. (1992). Psychodermatology: an overview. *Psychotherapy and Psychosomatics*, **58**, 125–36.
34. Koblenzer, C.S. and Bostrom, P. (1994). Chronic cutaneous dysesthesia syndrome: a psychotic phenomenon or a depressive symptom? *Journal of the American Academy of Dermatology*, **30**, 370–4.
35. Stein, D.J. and Hollander, E. (1992). Low-dose pimozide augmentation of serotonin reuptake blockers in the treatment of trichotillomania. *The Journal of Clinical Psychiatry*, **53**, 123–6.
36. Munro, A. and Stewart, M. (1991). Body dysmorphic disorder and the DSM-IV: the demise of dysmorphophobia. *Canadian Journal of Psychiatry*, **36**, 91–6.
37. McElroy, S.L., Phillips, K.A., Keck, P.E., *et al.* (1993). Body dysmorphic disorder: does it have a psychotic subtype? *The Journal of Clinical Psychiatry*, **54**, 389–95.
38. Malasi, T.H., El-Hilu, S.R., Mirza, I.A., *et al.* (1990). Olfactory delusional syndrome with various aetiologies. *The British Journal of Psychiatry*, **156**, 256–60.
39. Bhanji, S. and Mahony, J.D.H. (1978). The value of a psychiatric service within the venereal disease clinic. *British Journal of Venereal Disease*, **54**, 266–8.
40. Crowe, R.R., Clarkson, C., Tsai, M., *et al.* (1988). Delusional disorder: jealous and nonjealous types. *European Archives of Psychiatric and Neurological Science*, **237**, 179–83.
41. Mullen, P.E. and Martin, J. (1994). Jealousy: a community study. *The British Journal of Psychiatry*, **164**, 35–43.
42. Cobb, J. (1984). Morbid jealousy. In *Contemporary psychiatry* (ed. S. Crown), pp. 68–79. Butterworths, London.
43. Mowat, R.R. (1966). *Morbid jealousy and murder*. Tavistock Publications, London.
44. Fishbain, D.B. (1986). Suicide pacts and homicide. *The American Journal of Psychiatry*, **143**, 1319–20.
45. Gillett, T., Eminson, S.R., and Hassanyeh, F. (1990). Primary and secondary erotomania: clinical characteristics and follow-up. *Acta Psychiatrica Scandinavica*, **82**, 65–9.
46. Segal, J.H. (1989). Erotomania revisited: from Kraepelin to DSM-IIIR. *The American Journal of Psychiatry*, **146**, 1261–6.
47. Silva, J.A., Leong, G.B., Weinstock, R., *et al.* (1994). Delusional misidentification syndromes and dangerousness. *Psychopathology*, **27**, 215–19.
48. Raschka, L.B. (1979). The incubus syndrome, a variant of erotomania. *Canadian Journal of Psychiatry*, **24**, 549–53.
49. Goldstein, R.L. (1987). More forensic romances: de Clérambault's syndrome in men. *Bulletin of the American Academy of Psychiatry and the Law*, **15**, 267–74.
50. Purcell, R., Pathé, M., and Mullen, P.E. (2005). Association between stalking victimisation and psychiatric morbidity in a random community sample. *The British Journal of Psychiatry*, **187**, 416–20.
51. Pathé, M. and Mullen, P.E. (1997). The impact of stalkers on their victims. *The British Journal of Psychiatry*, **170**, 12–17.
52. Ellis, H.D. and Young, A.W. (1990). Accounting for delusional misidentifications. *The British Journal of Psychiatry*, **157**, 239–48.
53. Weinstein, E.A. (1994). The classification of delusional misidentification syndromes. *Psychopathology*, **27**, 130–5.
54. Cummings, J.L. (1985). Organic delusions: phenomenology, anatomical correlations, and review. *The British Journal of Psychiatry*, **142**, 184–97.
55. Förstl, H., Burns, A., Jacoby, R., *et al.* (1991). Neuroanatomical correlates of clinical misidentification and misperception in the senile dementia of the Alzheimer type. *The Journal of Clinical Psychiatry*, **52**, 268–71.
56. McAllister, T.W. (1992). Neuropsychiatric aspects of delusions. *Psychiatric Annals*, **22**, 269–77.
57. Arnone, D., Patel, A., and Giles, M.Y.T. (2006). The nosological significance of folie à deux: a review of the literature. *Annals of General Psychiatry*, **5**, 11–22.
58. Silveira, J.M. and Seeman, M.V. (1995). Shared psychotic disorder: a critical review of the literature. *Canadian Journal of Psychiatry*, **40**, 389–95.
59. Cameron, N. (1959). The paranoid pseudocommunity revisited. *American Journal of Sociology*, **65**, 57–61.
60. Myers, P.L. (1988). Paranoid pseudocommunity beliefs in a sect milieu. *Social Psychiatry and Psychiatric Epidemiology*, **23**, 252–5.
61. Fennig, S., Fochtmann, L.J., and Bromet, E.J. (2005). Delusional and shared psychotic disorder. In *Comprehensive textbook of psychiatry* (8th edn) (eds. H.J. Kaplan and B.J. Saddock), pp. 1525–33.
62. Kingdon, D., Turkington, D., and John, C. (1994). Cognitive behaviour therapy of schizophrenia. *The British Journal of Psychiatry*, **164**, 581–7.
63. Silva, S.P., Kim, C.K., Hofmannn, S.G., *et al.* (2003). To believe or not to believe: cognitive and psychodynamic approaches to delusional disorder. *Harvard Review of Psychiatry*, **11**, 20–9.
64. Munro, A. and Mok, H. (1995) An overview of treatment in paranoia/delusional disorder. *Canadian Journal of Psychiatry*, **40**, 616-22
65. Manschrek, T.C. and Khan, N.L. (2006). Recent advances in the treatment of delusional disorder. *Canadian Journal of Psychiatry*, **51**, 114–19.
66. Mandel, M.R., Severe, J.B., Schooler, N.R., *et al.* (1982). Development and prediction of postpsychotic depression in neuroleptic-treated schizophrenics. *Archives of General Psychiatry*, **39**, 197–203.
67. Kendler, K.S. (1982). Demography of paranoid psychosis (delusional disorder): a review and comparison with schizophrenia and affective illness. *Archives of General Psychiatry*, **39**, 890–902.

4.5

Mood disorders

Contents

4.5.1 Introduction to mood disorders

John R. Geddes

Mood and disorders of mood

The concept of mood is difficult to define. In psychiatry, it has come to mean a pervasive emotional tone varying along an axis from happiness to sadness—and perhaps anxiety. The boundaries between normal and abnormal mood are equally difficult to define. Nonetheless, there is usually no doubt about the most extreme manifestations of low mood, **depression**, or elevated mood, **mania**.

Early history

Descriptions of variations in mood which go beyond normal limits and are associated with functional impairment are present in the oldest writings of mankind.[1] The ancient Greeks identified that mood disorders were diseases of the body, rather than the effects of supernatural spirits and identified the link between elevation of mood and states of despondency or depression. The Hippocratics also identified that mental disorders were located in the brain. This insight was lost for 2000 years under the influence of Galen's humoral theory which held that melancholia was due to an excess of black bile and mania due to an excess of yellow bile. During this period, attempts to put forward empirical theories in both Western and Eastern civilizations often fell foul of increasingly dominant religious dogma.

Development of modern psychiatric nosology

In Europe, during the Enlightenment of the seventeenth and eighteenth centuries, reason and empiricism once again emerged. In *Anatomy of Melancholy*, published in 1632, Richard Burton provides a comprehensive review of previous writings on mood disorder.[2] Burton, who writes with the penetrating insight of someone with extensive personal experience of mood disorder, clearly makes the link between mood elevation and enhanced creativity, cycling with periods of low mood, when all pleasure is lost:

> When I go musing all alone
> Thinking of divers things fore-known.
> When I build castles in the air,
> Void of sorrow and void of fear,
> Pleasing myself with phantasms sweet,
> Methinks the time runs very fleet.
> All my joys to this are folly,
> Naught so sweet as melancholy.
> When I lie waking all alone,

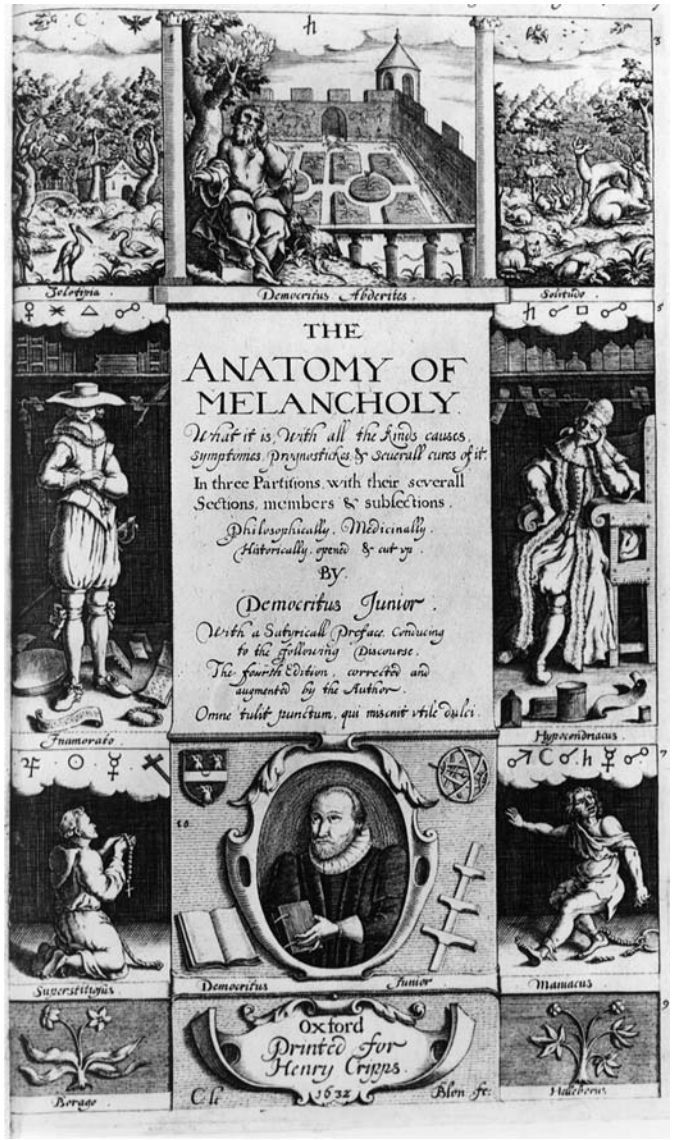


Fig. 4.5.1.1 Frontispiece to Burton's *Anatomy of Melancholy.*

Recounting what I have ill done,
My thoughts on me then tyrannise,
Fear and sorrow me surprise,
Whether I tarry still or go,
Methinks the time moves very slow.
All my griefs to this are jolly,
Naught so mad as melancholy.

(extract from The Author's Abstract of Melancholy, *Anatomy of Melancholy*)

Burton's work is an erudite and comprehensive review of the work on mood disorders until the early seventeenth century, although of course not systematic in the modern sense. Following Burton, Thomas Willis (1621–1675), Sedleian Professor of Natural Philosophy at Oxford University, perhaps better known for his description of the eponymous circle of Willis, is now recognized as one of the first to (re)localize psychiatric disorders within specific body organs, primarily the brain, rather than due to the circulation of bodily humours. In *Cerebri anatomi* (1664) Willis writes[3]:

> Melancholy is a complicated distemper of the brain and heart. For as melancholick people talk idly, it proceeds from the vice or fault of the brain and the inordination of the animal spirits dwelling in it, but as they become very sad and fearful, this is deservedly attributed to the passion of the heart. But we cannot here yield to what some physicians affirm, that melancholy doth arise from a melancholick humour. Melancholy being a long time protracted, passes oftentimes into stupidity, or foolishness, and sometimes into madness.

This identification of the physical brain as the location of the involved pathological processes initiated the modernist project of the scientific study of serious mood disorder, using contemporarily available scientific methods to generate insights into diagnosis, aetiology, course, and treatment. The first steps in this project were diagnosis and classification. Modern psychiatric diagnostic systems can be traced to post-Enlightenment Europe—in particular, France. The development of the mental hospital system throughout Europe provided the populations of patients for the early psychiatrists to study, as they created and refined diagnostic systems. In 1854, Esquirol described *la folie circulaire* which remains a recognizably modern description of bipolar disorder.[4] Despite this, the prevalent view throughout the nineteenth century was that manic and depressive states were separate entities. At the turn of the twentieth century, Kraepelin distinguished *dementia praecox* from *manic-depressive insanity*[5] Kraepelin emphasized the episodic course of the latter, the relatively benign prognosis and the family history of mood disorder.

Expanding the scope of mood disorder and recognition of diagnostic heterogeneity

In the twentieth century, diagnostic systems that had derived from the populations of large mental hospitals were challenged by the recognition that mental disorders were widespread in milder forms in the general population. Under the influence of Freud, Bleuler expanded Kraepelin's original concept of dementia praecox to include some of the less severe forms that he identified in the general population and renamed the disorder *schizophrenia* to reflect his views of the fundamental psychology of the psychosis.[6] A similar process to this occurred with the mood disorders, eventually leading to the recognition that milder, albeit still severe enough to cause impairment of function, forms of depression, and anxiety were the commonest forms of mental disorder in the general population. Indeed, after the World Health Organization and World Bank's Global Burden of Disease study in the 1990s, it became clear that unipolar depression was the among the most important causes of disability worldwide.[7] With the expansion in the scope of mood disorders, it was recognized that there is also heterogeneity in the way in which mood disorders manifest themselves. There was debate about existence of subtypes of depression, often based on the severity of symptoms and hypothesized links to an underlying brain process. For example, Gillespie's *reactive depression* in which the depressive disorder develops in response to external circumstances and remains subject to influence by external influences during its course, can be contrasted with *endogenous* and *melancholic* subtypes which were held to arise from a primary neurobiological abnormality.

The distinction between unipolar and bipolar disorders

Despite having some descriptive clinical utility, most postulated depressive subgroups have remained of uncertain status. The distinction between unipolar and bipolar disorders, however, has become accepted and is included in modern diagnostic systems. Initially, during the twentieth century, diagnostic systems continued to maintain the Kraepelinian approach to mood disorder with *manic-depressive insanity* including both severe mania and depression.—i.e. illnesses that were characterized only by depressive episodes and illnesses which included both poles.

The school of Kleist and Leonhard identified the heterogeneous nature of patients with manic-depressive illness—some had both manic and depressive episodes while others had only depressions. They coined the terms bipolar and unipolar to describe these two forms. Carl Perris and Jules Angst later provided some empirical validation for the separation on the basis of family history.[8,9]

Modern diagnostic systems—the birth of diagnostic criteria

Under the influence of psychoanalytic thought, psychiatric diagnoses had become very vague and unreliable by the 1960s—particularly in the United States where psychoanalysis was particularly influential. Following the observation that schizophrenia was more prevalent in the United States than in United Kingdom, the US–UK Diagnostic project showed conclusively that the apparent differences appeared to be due to the differences in diagnostic practice rather than true differences in prevalence.[10] The US–UK Diagnostic project highlighted the unacceptable reliability of psychiatric diagnoses which fuelled the arguments of critics of psychiatry. It was recognized that more reliable diagnostic systems were required which were explicitly based on evidence of validity. The defining feature of these systems was the use of explicit diagnostic criteria and, historically, the most important of these was the third edition of the Diagnostic and Statistical Manual of the American Psychiatric Association.[11]

International classification systems were based on compromise between national views rather than evidence and were therefore slower to change. The ninth edition of the World Health Organization's International Classification of diseases retained the concept 'Manic-Depressive Illness' which included both unipolar and bipolar disorders. However, the distinction was finally made in the 10th edition in 1993.[12]

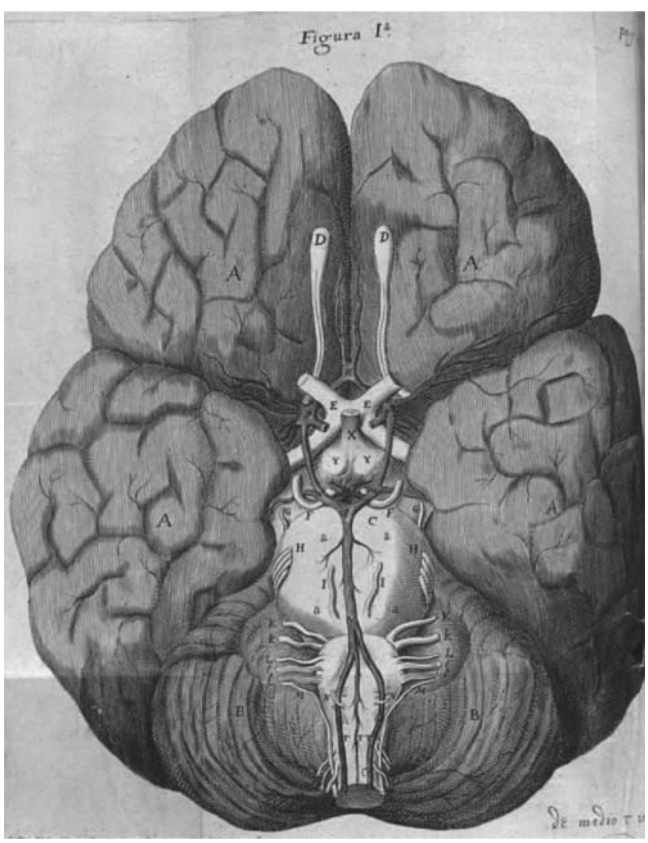


Fig. 4.5.1.2 Illustration by Christopher Wren from *Cerebi anatome* by Thomas Willis.

Refining the concept—subgroups of unipolar and bipolar disorder

As well as the fundamental Kraepelinian distinction between mood disorders and schizophrenia and the subsequent distinction between unipolar and bipolar mood disorders, there have been several attempts to subclassify mood disorders.

Unipolar disorder

A crucial distinction has been made between *endogenous* and *reactive* forms of unipolar depressive disorder (see above). Current classifications do not emphasize the distinction because it is now recognized that both typical endogenous and reactive clinical pictures can be related to external stressors. Nonetheless, the concept of a melancholic subtype remains.

Bipolar disorder

As more is known about the epidemiology of bipolar disorder, it has become apparent that some people do not experience manic episodes, but nonetheless do experience episodes of mood elevation that are clearly noticeable to themselves and others (hypomania)- as well as depressive episodes. To accommodate these heterogeneous clinical pictures, bipolar disorder has been divided into bipolar disorder, type 1 (mania ± depression) and bipolar disorder, type 2 (hypomania + depression). A focus of intense current research is the concept of bipolar spectrum disorder—in other words, the recognition that there is probably a continuum of mood phenomena from normal mood through to extreme mania.

Future developments in the classification of mood disorders

We can expect more changes in the diagnostic classification as knowledge continues to accrue. Our current classifications remain entirely descriptive, based on cross-sectional clinical symptoms and longitudinal course. As such, the classifications have been therapeutically useful—we now have increasing evidence for the diagnosis-specific effects of treatments. In the future, as out knowledge of the basic neurobiology of mood disorder increases, incremental or fundamental changes may be required.

Further information

Goodwin, F.K. and Jamison, K.R. (2007). *Manic-depressive illness: bipolar disorders and recurrent depression* (2nd revised edn). Oxford University Press, USA.

References

1. Goodwin, F.K. and Jamison, K.R. (1990). *Manic depressive illness*. Oxford University Press, New York.
2. Burton, R (1632). *Anatomy of Melancholy.* http://onlinebooks.library.upenn.edu/webbin/gutbook/lookup?num=10800
3. Willis, T. (1664). *Cerebri anatome: cui accessit nervorum descriptio et usus*. J. Martyn & J. Allestry, London.
4. Esquirol, J.E.D. (1838). *Des maladies mentales*. Balliere, Paris.
5. Kraepelin, E. (1896). *Lehrbuch der psychiatrie* (5th edn). Barth, Leipzig.
6. Bleuler, E. (1950). *Dementia praecox or the group of schizophrenias*. International Universities Press, New York.
7. Murray, C.J. and Lopez, A.D. (1997). Global mortality, disability, and the contribution of risk factors: global burden of disease study. *Lancet*, **349**, 1436–42.
8. Perris, C.A. (1966b). Study of bipolar (manic-depressive) and unipolar recurrent depressive psychoses I. Genetic investigation. *Acta Psychiatrica Scandinavica*, **42**(Suppl. 194), 15–44.
9. Angst, J. (1966). *Zur Ätiologie und Nosologie endogener depressiver Psychosen*. Springer, Berlin.
10. Cooper, J.E., Kendell, R.E., Gurland, B.J., *et al.* (1972). *Psychiatric diagnoses in New York and London*. Maudsley Monograph No. 20. Oxford University Press, London.
11. American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd edn DSM-III). APA, Washington, DC.
12. World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders*. World Health Organization, Geneva.

4.5.2 Clinical features of mood disorders and mania

Per Bech

Introduction

The clinical features of mood disorders are dimensional, i.e. distributed according to their severity.[1] The categorical approach as manifested in the DSM-IV or ICD-10 does not, however, preclude dimensional descriptions, because in DSM-IV as well as in ICD-10 the categories or diagnoses are essentially defined by minimum and maximum cut-off scores on the symptomatic states.

Fig. 4.5.2.1 shows a coordinate system in which the ordinate represents the dimension of manic states and the abscissa the dimension of depressive states. The cut-off scores refer to the standardization of the isometric rating scales for mania and depression.[1–3] Jamison[4] has argued that the term 'bipolar' perpetuates the notion that: '. . . depression exists rather tidily segregated in its own pole, while mania clusters off neatly and discretely on another. This polarisation of two clinical states flies in the face of everything that we know about the fluctuating nature of manic-depressive illness . . . and it minimises the importance of mixed manic-depressive states'. William James referred to the stream of consciousness to emphasize its continuity in contrast to its conception as a series of discrete states. However, William James actually confessed that during his own depressive episodes, his mood states,[5] to a large extent, blocked his own stream of consciousness.[6] In the perspective outlined by Jamison,[4] polarity in the clinical world is not two opposites that contradict each other by a logical relationship of juxtaposition. Polarity should rather, as shown in Figs. 4.5.2.1 and 4.5.2.2, be considered at a level of psychological intercorrelations in which clinical mania and depression exist by virtue of each other involving both negative and positive correlations.

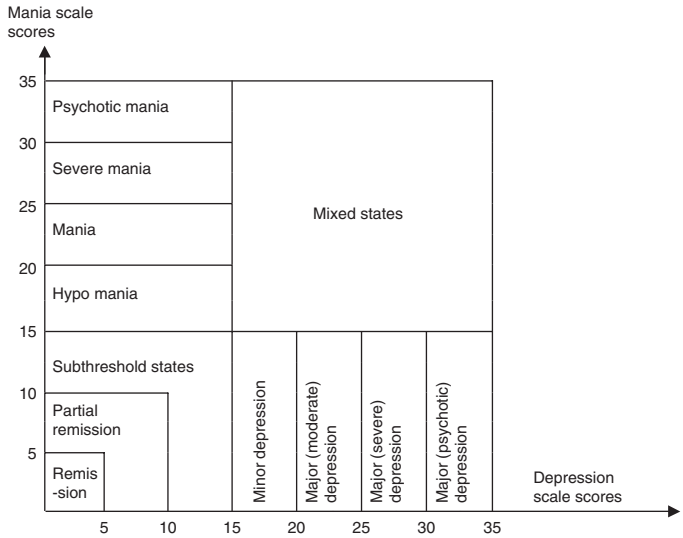


Fig. 4.5.2.1 Patient state fluctuations within the spectrum of mood states from the subthreshold states over minor mood states (hypomania or minor depression) to major states (mania/psychotic mania or major depression without or with psychotic symptoms) and to the mixed states.

The symptom rating scales shown in Fig. 4.5.2.1 measure the severity of mania, depression, or mixed states and have a time frame of 3 days. Clinically, this is the minimum for the measurement of the spectrum of mood states ranging from subthreshold states to states of psychotic severity.

Ultrashort states of mood swings are often seen without any reference to mood disorders. In one of Henry James' masterpieces, The Ambassadors, his autobiographical hero has a tendency towards being introverted in the morning, while in the afternoon and evening he is more extraverted, like a man ' . . . who, elately finding in his pocket more money than usual . . . ', though without spending a lot. This 24-h 'cyclothymia' between introversion and extraversion displays too mild a symptomatology to be part of a 'bipolar' disorder and might be referred to as a temperamental neuroticism. Thus, Eysenck's original questionnaire for measuring neuroticism included items of being moody without any apparent reason or being inclined to having frequent ups and downs in mood.[7] Neuroticism now seems to include subclinical, temperamental, low, negative affectivity (worrying, gloomy, dysphoric, and hostile).

Fig. 4.5.2.2 shows another coordinate system in which the ordinate is representing the dimension of mania severity and the abscissa the dimension of depression severity, but covering the lifelong correlation of the courses of manic and depressive episodes,

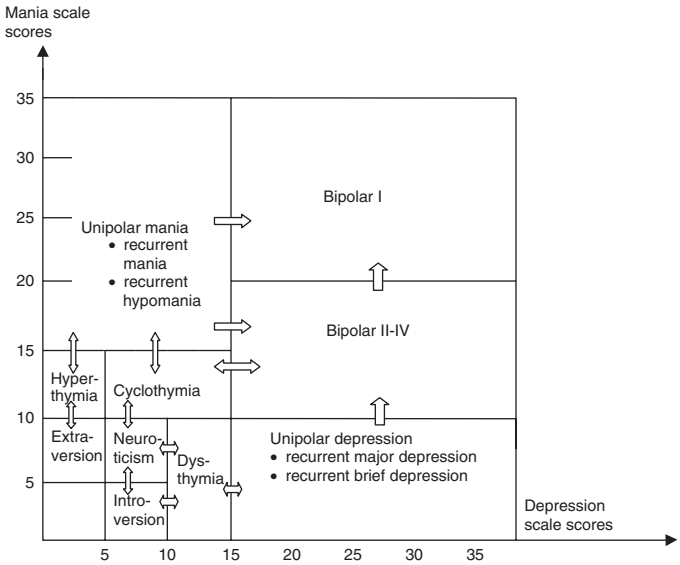


Fig. 4.5.2.2 Patient state lifetime fluctuations within the spectrum of mood polarity syndromes or disorders from the temperamental neuroticism over minor disorders (dysthymia or cyclothymia) to major unipolar disorders (hypomania/mania or brief/major depression) and to the various bipolar disorders.

i.e. the longitudinal diagnosis of mood polarity disorders. While a unipolar course of manic episodes without any depressive episodes is very rare,[8] the course of unipolar depressive episodes without manic episodes is much more frequent. Angst[9] has criticized the DSM-IV definitions of depressive episode disorders (a duration of at least 2 weeks) and of manic episode disorder (a duration of at least 1 week). Thus, Angst has shown that these episode duration criteria are not only too narrow (not sensitive enough), but they also lack empirical evidence (not validated). The spectrum of mood polarity as shown in Fig. 4.5.2.2 is an attempt to refer to the DSM-IV definitions, though modified with reference to Angst.[9] The mood polarity disorder as shown in Fig. 4.5.2.2 is defined by the highest score a given patient has obtained in the coordinate system at any time. Unipolar depression, however, with a score of 10 or less on the mania scale, as shown in Fig. 4.5.2.2, remains a lifelong uncertain diagnosis. Thus, Angst[9] has demonstrated a persistent risk of 1–2 per cent per year of a diagnostic change from unipolar to bipolar disorder.

The dimensional approach has been valid also in regard to personality disorder. As discussed by Angst,[9] only around 15 per cent of the general population seem to report no lifelong personality disorders, and he therefore calls these persons 'supernaturals'. Thus, nearly everyone has some kind of personality disorder, and within the spectrum of mood personality disorders, extraversion, introversion, neuroticism, hyperthymia, dysthymia, and cyclothymia are to be considered subthreshold disorders as shown in Fig. 4.5.2.2.

It has recently been argued that when taking into account both the subthreshold levels of symptoms and the short states of 2–3 days' duration rather than the whole DSM-IV episodes, there appears to be a linear correlation between mania and depression in the course of illness in many patients with mood disorders.[10]

The depressive episode: duration and severity

Table 4.5.2.1 shows the DSM-IV and ICD-10 depressive symptoms for the diagnosis of major depression, which to a large extent covers the rating scale dimension in Fig. 4.5.2.1. Thus, the individual symptoms should be present most of the day and nearly every day during an episode. Kendler and Gardner[11] have shown that the risk of developing a new major depressive episode (i.e. of a duration of 14 days or more) is as high for patients with major depressive symptoms lasting from 5 to 13 days as for patients with symptoms lasting from 14 to 59 days.

The study by Kendler and Gardner[11] has also demonstrated that patients with a subthreshold quantity of depressive symptoms (i.e. just below five out of the nine DSM-IV symptoms listed in Table 4.5.2.1 or minor depression in Fig. 4.5.2.1) had the same risk of developing a new major depressive episode as patients fulfilling the symptomatic criteria of major depression. It has been shown that approximately 50 per cent of the patients fulfilling the ICD-10 category of mild depression also fulfil the criteria for DSM-IV major depression.

To illustrate how a major depressive episode often develops on a continuum of depressive symptoms from the first prodromes of decreased positive well-being (introversion) to the major depressive episode, a layman's description is shown in Box 4.5.2.1. These autobiographical notes of the late William Styron (1925–2006) describe his first episode of depression at the age of 60.[12] As indicated, the symptom of anxiety (not included in DSM-IV or ICD-10) is a very important symptom of major depression from its very onset. When suicidal impulses developed, he was admitted to Yale New Hospital.

Table 4.5.2.1 Depression symptoms as included in DSM-IV and ICD-10

	Symptoms of depression	DSM-IV	ICD-10
1	Depressed mood most of the day, nearly every day	+	+
2	Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day	+	+
3	Loss of energy or fatigue nearly every day	+	+
4	Loss of confidence or self-esteem	–	+
5	Unreasonable feelings of self-reproach or excessive or inappropriate guilt, nearly every day	+	+
6	Recurrent thoughts of death or suicide, or any suicidal behaviour	+	+
7	Diminished ability to think or concentrate, or indecisiveness, nearly every day	+	+
8	Psychomotor agitation or retardation nearly every day	+	+
9	Insomnia or hypersomnia nearly every day	+	+
10	Change in appetite (decrease or increase with corresponding weight change)	+	+

+ indicates that the symptom is included; – indicates that the symptom is not included.

Box 4.5.2.1 The stages from decreased positive well-being through mild depression to major depression without psychotic features

- The shadows of nightfall seemed more sober, my mornings were less buoyant, walks in the woods became less zestful, and there was a moment during my working hours when a kind of panic and anxiety overtook me, just for a few minutes, accompanied by a visceral queasiness . . .
- . . . As the disorder gradually took full passion of my system, I began to conceive that my mind itself was like one of those outmoded small-town telephone exchanges, being gradually inundated by flood-waters.
- . . . I particularly remember the lamentable near disappearance of my voice . . . The libido also made an early exit . . . food, like everything else within the scope of sensation, was utterly without savor.
- . . . My few hours of sleep were usually terminated at three or four in the morning, when I started up into yawning darkness . . . I'm fairly certain that it was during one of these insomnia trances that there came over me the knowledge that this condition would cost me my life, if it continued on such a course . . . I had not conceived precisely how my end would come. In short, I was still keeping the idea of suicide at bay . . . What I had begun to discover is that the grey drizzle of horror, induced by depression, takes on the quality of physical pain . . .

(from W. Styron (1990), *Darkness visible. A memoir of madness*, Random House, New York.)

Major depression with or without melancholia

Negative beliefs such as 'loss of self-esteem' or 'inappropriate guilt' are among the most specific core symptoms of major depression. Inappropriate guilt can be experienced as a punishment for past misdeeds (prior to the current episode of depression). The prevailing element of negative beliefs is a sense of loss which is associated with lower self-esteem experienced retrospectively. The symptom which discriminates best between anxiety states and major depressive disorder is guilt.

However, also states of anxiety with worrying and panic are important symptoms of depression. Anxiety is among the core items of the Hamilton Depression Scale.[13] Patients suffering from subthreshold depression experience less anxiety than patients with major depression. Another important symptom of depression included in the depression dimension of Fig. 4.5.2.1, but not in Table 4.5.2.1, is emotional and social withdrawal. Within the flux between mania and depression, this is probably another specific symptom (the emotionally intrusive behaviour in mania is its opposite pole).

Both in DSM-IV and ICD-10, major depressive states can be further specified as melancholic or somatic syndromes. In earlier descriptions (including Freud's 'Mourning and melancholia'), endogenous or somatic depression is distinguished from psychogenic or reactive depression by 'early morning awakening' and 'depression regularly worse in the morning'. These two signs are the only features of somatic or melancholic depression not included in the list of symptoms in Table 4.5.2.1. Strictly speaking, diurnal variation of symptoms is not a symptom itself, but rather a description of the fluctuation. The most 'somatic' symptom in Table 4.5.2.1 is change in body weight (Styron had lost 20 to 25 pounds over a period of 6 weeks, when the illness developed into a major depression).[14]

Styron's depression (Box 4.5.2.1) included the somatic feature of early morning awakening and suicidal thoughts. The latter are not just a consequence of the other symptoms. Styron described how during depression he could still keep '…the idea of suicide at bay…'. At a later stage (not shown in Box 4.5.2.1), just before he was admitted to hospital, Styron tried to write a suicide letter. Suicidal thoughts were often present late at night, when anxiety symptoms had lifted.

Measurements of social behaviour and subjective distress have shown that acute major depression is one of the most disabling and distressing of medical disorders.[6] The constant mental pain and the suicidal symptoms seriously affect quality of life. The suicidal risk in major depression is especially high when psychomotor retardation is improving in the course of treatment. The treating physician or the relatives typically observe improvement in the depressive symptoms before the patient does, because psychomotor retardation improves before mood state or hopelessness do. The risk is especially high in socially isolated people. Major depression has the highest risk of suicide of all mental disorders, and all patients with major depression should be assessed for the risk of suicide. It has been shown that patients with unipolar depression have higher suicide rates than patients with bipolar I and bipolar II disorders. However, concerning the core items of severe depressive states, no differences in the intensity of symptoms have been seen between unipolar and bipolar patients.

Major depression with psychotic features

According to DSM-IV or ICD-10, the term 'psychotic depression' is not synonymous with endogenous or melancholic depression. This agrees with Hamilton[13] who used the term 'psychosis' to refer to the severity of symptoms. As stated by Hamilton:[13] ' . . . a schizophrenic patient, who has delusions is not necessarily worse than one who has not, but a depressive patient who has is much worse . . . '.

Recurrent depression

Recurrent major depression

After a single episode of major depression, around 85 per cent of patients experience recurrent episodes. While the first episode of major depression is often provoked by a negative life event such as loss of job, retirement, marital separation or divorce, subsequent episodes are often unprecipitated (but positive life events can also provoke depression). Depressive episodes typically increase in frequency and duration as they return.[3]

Recurrent brief depression

The symptoms of recurrent brief depression, first described by Angst, are similar to those of major depression (Table 4.5.2.1) with regard to both number and severity. Recurrent brief depression is a

state of major depression lasting 2–3 days. Its diagnosis has not been adopted fully in DSM-IV, but it is included in ICD-10. It should be distinguished from recurrent suicidal behaviour, for example in patients with borderline personality disorder.

Seasonal depression

Seasonal depression is seen most frequently in winter, and less frequently in summer. In DSM-IV, seasonal depression has been adopted as a specifier (rather than a diagnostic category) which can be applied not only to recurrent depression but also to bipolar disorder. The seasonal episodes (e.g. winter depression) have to outnumber any non-seasonal depressive episodes in the same patient. In ICD-10 only seasonal depression is briefly mentioned, but only in an annex for disorders under consideration.

According to DSM-IV, the symptoms of seasonal depression are similar to those of major depression. However, it has been shown that the symptoms differ from those of major depression, with hypersomnia, overeating, carbohydrate craving, and weight gain (often referred to as atypical depression).

Atypical depression

DSM-IV atypical depression has a specifier which can be applied both to major depression and to bipolar I and bipolar II, but also to dysthymia. The core item is mood reactivity (i.e. the mood brightens in response to actual or potential positive events), while the associated symptoms are hypersomnia, overeating, and weight gain, leaden feelings in arms or legs, and interpersonal rejection sensitivity. It has been shown that mood reactivity is more often seen in bipolar II disorders than in unipolar depression disorders.

The manic episode: duration and severity

Table 4.5.2.2 shows the symptoms of mania according to DSM-IV and ICD-10, which to a large extent covers the rating scale dimension in Fig. 4.5.2.1. The episode criteria for hypomania and mania are shorter than those for major depression. For hypomania, the symptoms should have lasted at least 4 days, and it has been suggested to accept as little as 2 days (e.g. for states of hypomania).[15] In DSM-IV, but not in ICD-10, the category bipolar II disorder has been adopted (Fig. 4.5.2.2), referring to patients with a major depressive episode, who previously have experienced episodes of hypomania. It has been shown that patients with cyclothymia have the same risk of developing bipolar II depression as patients with hypomania. It has been proposed to extend the bipolar spectrum to include a category of bipolar III to refer to depressed patients who have developed hypomania episodes during treatment with antidepressant medication,[16] while bipolar IV refers to depressed patients having a substance-induced hypomania.

Table 4.5.2.2 Manic symptoms as included in DSM-IV and ICD-10

	Symptoms of mania	DSM-IV	ICD-10
1A	Elevated mood	+	+
1B	Irritable mood	+	+
2	Increased self-esteem or grandiosity	+	+
3	Decreased need for sleep	+	+
4	Increased talkativeness	+	+
5	Flight of ideas	+	+
6	Distractibility	+	+
7A	Increased social activities or contacts	+	+
7B	Psychomotor agitation	+	+
8	Risk taking behaviour	+	+
9	Increased sexual activities	–	+

+ indicates that the symptom is included; – indicates that the symptom is not included.

Patient-reported questionnaires have recently been published to identify previous episodes of hypomania in depressed patients. It has been shown that the Hypomania Checklist[17] is superior to the Mood Disorder Questionnaire in identifying patients with bipolar II disorder.

The problem with self-reported questionnaires to measure manic symptoms is that many patients do not perceive their hypomanic symptoms as pathological. They may describe themselves as 'normal' or their response pattern may show that they play 'the manic game'.

Studies with clinician-rated mania scales have shown that the various symptoms in Table 4.5.2.2, when quantified, can be rank-ordered in one single dimension of severity, analogously to the depressive symptoms in Table 4.5.2.1. This is illustrated by Box 4.5.2.2, which lists the three stages of mania from the longitudinal study by Carlson and Goodwin[18] on hospitalized patients in the untreated stage of their illness.

Hypomania

Hypomania can be the first stage of a spiralling upswing of mood (Box 4.5.2.2). According to DSM-IV, hypomania is more in concordance with the following description by Jamison[4] than with Box 4.5.2.2: ' . . . When you're high it's tremendous. The ideas and feelings are fast and frequent . . . Shyness goes, the right words and gestures are suddenly there, the power to captivate others a felt certainty . . . Sensuality is pervasive and the desire to seduce and be seduced irresistible'. The shyness or introversion seen in mild depression or dysthymia contrast with the lack of shyness and extraversion seen in the hypomanic patient.

As described in Box 4.5.2.2, the dysphoric or irritable mood is often a core symptom of hypomania, which in the next stage of mania develops further, to cooperation difficulties and impulsive hostility. Although hypomania might cause hyperactiveness in social functioning,[15] the more or less hidden hostility often causes marked impairment in the hypomanic individual's ability to pursue some necessary task or to maintain an acceptable contact with family members. In this respect, therefore, dysphoric hypomania may cause as much clinically significant impairment in social functioning as cyclothymia.

Mania without psychotic features

In mania, the elevated spirit seen in hypomania is often mixed with irritability and hostility. Jamison[4] has described the change: 'Humor and absorption on friends' faces are replaced by fear and concern. Everything previously (in the hypomanic state) moving with the grain is now against—you are irritable, angry, frightened, uncontrollable . . . '.

Box 4.5.2.2 The three stages of the acute manic episode as observed in untreated inpatients

- **Hypomania**

  Increased well-being and/or irritable mood, more busy, pressured speech, makes more telephone calls, seductive

- **Mania**

  Nearly always pleasant and cheerful. Occasionally losing insight and cooperation, impulsive, angry, very hyperactive, less sleep, more pressure of speech, makes repeated telephone calls, racing thoughts, more expansive, some grandiosity

- **Mania with psychotic features**

  Emotionally labile, can be very angry, very intrusive. Uncooperative, severely agitated, no sleep, very talkative, loud, flight of thoughts, grandiosity, religious delusions 'hearing God', sexually very preoccupied

(modified from G.A. Carlson and F.K. Goodwin (1973), The stages of mania. A longitudinal analysis of the manic episode, *Archives of General Psychiatry*, **28**, 221–8.)

The psychomotor symptoms of mania are restlessness and less need for sleep. There is pressure of speech; the patient talks more and in a louder voice. There is intrusive behaviour, arguments, and attempts to dominate others. Expansiveness is manifested as increased self-esteem; for example, the patient clearly overestimates his or her own capacities or hints at unusual abilities. Jamison[4] has described how in periods of mania she did not worry about money: 'The money will come from somewhere; I am entitled; God will provide. Credit cards are disastrous, personal cheques even worse . . . mania is a natural extension of the economy . . . So I bought precious stones, elegant and unnecessary furniture, three watches within an hour (in the Rolex rather than Timex class) . . . '.

To be diagnosed as a manic episode, the disorder should last at least a week. The criteria for mania are elevated or clearly irritable mood, and at least three of the symptoms listed in Table 4.5.2.2. These symptoms should be severe enough to cause marked impairment of occupational functioning. Hospital admission is often needed to prevent the patient from harming him/herself or others.

The psychomotor restlessness and the pressured speech in the dysphoric hypomanic state present a clinical picture of agitated depression, i.e. a mixed state (Fig. 4.5.2.1). The taste for talking is observed more frequently in females than in males, but otherwise no significant differences are seen between males and females in the manic dimension (Fig. 4.5.2.1).

Mania with psychotic features

Psychotic states of mania are characterized by greater pressure of speech, more open hostility, severe agitation, no need for sleep, flight of thoughts, severe distractibility, and grandiose delusions. In younger people, psychotic mania is often misdiagnosed as schizophrenia.

In hospital, the increased social contact of manic patients is clearly different from the emotional bluntness of schizophrenics. The intrusive behaviour seen in severe mania is of an extremely dominating and manipulative nature, out of context with the setting. Hamilton found this to be one of the most important mania symptoms. Secondary persecutory delusions often develop. The expansive religious delusion 'hearing God' should be differentiated from the schizophrenic patient's religious hallucinations.

Both DSM-IV and ICD-10 differentiate between mood-congruent psychotic symptoms (such as grandiose delusions of religion and voices supporting the patient's superhuman powers) and mood-incongruent psychotic symptoms (which are often the secondary delusions of persecution mentioned above).

Mixed states

The mixture of manic and depressive symptoms is the essential feature of mixed states (Fig. 4.5.2.1). It is not necessarily a bipolar course of symptoms, but the mixture of depressive and manic symptoms nearly every day. DSM-IV requires that the duration of a mixed episode should be at least 1 week; ICD-10 requires at least 2 weeks.

Kraepelin described transitory moods of depression in acute manic states. Transitory moods of depression have been recorded in manic patients by use of a rating scale administered by the nursing staff. Such short-lived states of 'depression' in patients with acute mania should be referred to as 'microdepressions' and not mixed episodes. Winokur described 'microdepressions' very clearly:

> . . . If one allows a manic patient to talk, one will note that he shows fleeting episodes of depression embedded within mania (microdepressions). He may be talking in grandiose and extravagant fashion and then suddenly for 30 seconds breaks down to give an account of something he feels guilty about . . . His eyes will fill with tears but in 15 to 30 seconds he will be back to talking in his expansive fashion.

Rapid cycling

DSM-IV has a specifier for rapid cycling, which refers to patients with bipolar I or bipolar II disorders, who have experienced at least four episodes during the previous 12 months. The episode has to be demarcated by partial or full remission for at least 2 months or a switch to an episode of opposite polarity.

Conclusion

The clinical spectrum of the states of depression and mania has been described in Fig. 4.5.2.1 by the symptomatic dimensions of severity as validated by clinician-rated scales. Thus, symptom severity is a key issue of the spectrum of mood states.

The spectrum of mood polarity disorders covering the longitudinal diagnosis of manic and depressive episodes is shown in Fig. 4.5.2.2. Recent research has demonstrated how important it is to recognize subthreshold states of mania and depression, as they can have a major impact on both social functioning and quality of life, since many patients with mood disorders spend much time in subthreshold disorders, i.e. cyclothymia, dysthymia, or neuroticism.

While the severity spectrum of the states of mania and depression has been accepted as evidence-based, we still lack a validation of the mood polarity spectrum in long-term follow-up studies.

Acknowledgement

I wish to thank Jules Angst and Rasmus W. Licht for their valuable comments and suggestions during the revision of this chapter.

Further information

The greatest resource book for research is Goodwin, F.K. and Jamison, K.R. (2007). *Manic depressive illness* (2nd edn). Oxford University Press, Oxford.

Other sources for further information are:

Marneros, A. and Goodwin, F.K. (eds.) (2005). *Bipolar disorders: mixed states, rapid cycling and atypical forms*. Cambridge University Press, Cambridge.

Stein, D.J., Kupfer, D.J., and Schatzberg, A.F. (eds.) (2006). *Textbook of mood disorders*. American Psychiatric publishing, Washington, DC.

Ghaemi, S.N. (2003). *The concepts of psychiatry. A pluralistic approach to the mind and mental illness*. Johns Hopkins University Press, Baltimore.

References

1. Bech, P. (1993). *Rating scales for psychopathology, health status and quality of life*. Springer Verlag, Berlin.
2. Bech, P. (2005). The Bech-Rafaelsen Mania and Melancholia Scales in clinical trials: a 25-year review of their use as outcome measure in bipolar and unipolar patients. In *Focus on bipolar disorder research* (ed. M.C. Brown), pp. 131–51. Nova Science Publishers, Hauppauge, New York.
3. Goodwin, F.K. and Jamison, K.R. (2007). *Manic-depressive illness* (2nd edn). Oxford University Press, New York.
4. Jamison, K.R. (1995). *An unquiet mind. A memoir of moods and madness*. Knopf, New York.
5. James, W. (1890). *Principles of psychology*. Holt, New York.
6. Bech, P. (1998). *Quality of life in the psychiatric patient*. Mosby-Wolfe, London.
7. Eysenck, H.J. (1958). A short questionnaire for the measurement of two dimensions of personality. *Journal of Applied Psychology*, **42**, 14–17.
8. Klerman, G.L. (1981). The spectrum of mania. *Comprehensive Psychiatry*, **22**, 11–20.
9. Angst, J. (2007). The bipolar spectrum. *The British Journal of Psychiatry*, **190**, 81–6.
10. Cassano, G.B., Rucci, P., Frank, E., *et al.* (2004). The mood spectrum in unipolar and bipolar disorder: arguments for a unitary approach. *The American Journal of Psychiatry*, **161**, 1264–9.
11. Kendler, K.S. and Gardner, C.O. (1998). Boundaries of major depression: an evaluation of DSM-IV criteria. *The American Journal of Psychiatry*, **155**, 172–7.
12. Styron, W. (1990). *Darkness visible. A memoir of madness*. Random House, New York.
13. Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, **23**, 56–62.
14. West, J.L.W (1998). *William Styron. A life*. Random House, New York.
15. Benazzi, F. (2007). Bipolar disorder—focus on bipolar II disorder and mixed depression. *Lancet*, **369**, 935–45.
16. Klein, D.N., Shankman, S.A., and McFarland, B.R. (2006). Classification of mood disorders. In *Textbook of mood disorders* (eds. D.J. Stein, D.J. Kupfer, and A.F. Schatzberg), pp. 17–32. American Psychiatric Publishing Inc., Washington, DC.
17. Angst, J., Adolfsson, R., Benazzi, F., *et al.* (2005). The HCL-32: towards a self-assessment tool. *Journal of Affective Disorders*, **88**, 217–33.
18. Carlson, G.A. and Goodwin, F.K. (1973). The stages of mania. A longitudinal analysis of the manic episode. *Archives of General Psychiatry*, **28**, 221–8.

4.5.3 Diagnosis, classification, and differential diagnosis of the mood disorders

Gordon Parker

Introduction

Varying expressions of mood disorders make for difficulties in definition, diagnosis, and classification. DSMIV and ICD10 formal classifications with decision rules (see Chapter 1.9) provide a structure but their underlying models may or may not be valid. This chapter therefore considers how mood disorders can be variably conceptualized and structured—an issue of intrinsic importance but also influencing identification of causes and management. Some definitional and boundary issues are first detailed prior to considering sub-typing and differential diagnostic issues.

Definitions

(a) Depression

The term *depression* is extremely broad, variably defining an affect, mood states, disorders, or syndromes—as well as disease states. A depressed 'affect' usually occurs in response to a specific situation and is defined as a transient and non-substantive state of feeling 'depressed', 'sad', or 'blue'.

A *depressed mood* is more pervasive, more likely to be experienced as unusual or atypical, associated with negative ideation (e.g. hopelessness, helplessness, pessimism about the future), and may influence behaviour. Its quintessential construct is lowering of the individual's intrinsic level of self-esteem, with the extent of self-esteem lowering roughly equating to the severity of the mood state. Experienced by most people, it generally lasts only minutes to days in non-clinical situations.

A *depressive condition* (be it a disorder, syndrome, or disease) is generally distinguished by a longer duration, more clinical (and more pathological) features, and distinct social impairment. A duration criterion ensures that a transient depressed mood state does not alone establish psychiatric 'case' status, with a minimum duration of 2 weeks capturing most conditions other than the so-called 'adjustment disorders'. Additional clinical features (detailed shortly) inform us about severity (e.g. 'major' and 'minor' depressive disorders) and sub-typing, while the social impairment criterion further cleaves 'normal' mood states from clinical depressive conditions.

At times, depressive conditions are described as *primary or secondary*, a distinction necessarily imprecise. We comfortably concede 'secondary depression' when depression emerges during the course of a substantive psychiatric condition (e.g. schizophrenia) or medical condition, or following certain aetiologically defined triggers (e.g. substance abuse). However, as depression is commonly contributed to by other psychiatric disorders (e.g. severe anxiety states) and primary psychosocial factors, it might be logical to also call these 'secondary' depressive disorders, and yet this rarely occurs. The term 'secondary depression' therefore generally imputes a substantive primary condition with depression as a clear-cut consequence.

(b) Bipolar/unipolar depression specifics

Turning from cross-sectional to longitudinal definition, the course specifier 'bipolar' is applied to those having had at least one manic or hypomanic episode, whether preceded or not by a depressive episode. Originally, Leonhard[1] introduced the concept of 'monopolar' (or 'unipolar') depression to distinguish those who had episodes of the melancholic sub-type of depression, but no manic episode. Regrettably, the term 'unipolar' depression is now used to define a residual (i.e. non-bipolar) category, so heterogeneous as to be of limited meaning and utility.

(c) Mania/hypomania

As described in Chapter 4.5.2, such conditions are the converse of depression and fundamentally represent hedonistic, high energy states. Here self-esteem is almost invariably increased, the mood generally infectious, the individual energized or 'wired', disinhibited, with creativity and religiosity often enhanced, while psychotic features may be present.

Distinguishing 'hypomania' and 'mania' is imprecise in the formal classificatory systems, as noted shortly. To some theorists, the presence of psychotic features determines manic (as against hypomanic) status. Others subscribe to a dimensional model. For example, Goodwin and Jamison[2] suggest that hypomania and mania differ little in mood components, but that cognition, perception, and behaviour differ in severity and manifestation.

(d) Bipolar categories

In recent years, bipolar disorder has been principally sub-categorized into bipolar I and bipolar II expressions, with 'manic' and 'hypomanic' episodes, respectively, defining the 'highs'. The term 'bipolar III' refers to a manic or hypomanic 'switch' on exposure to—or cessation of—an antidepressant drug and may reflect a pure drug effect and/or a vulnerability to switching in those with a latent bipolar condition. Numerous other bipolar categories (e.g. IV, V, and VI) have been proposed in the last few decades.[3] Many describe a 'hyperthymic' bipolar type (where the individual tends to be frequently cheerful, overly talkative, extroverted, self-assured, and full of ideas). Whether this is merely an exuberant personality style or a mild or sub-clinical expression of bipolar disorder remains to be clarified. The growth in bipolar sub-types has led to the dimensional concept of a 'bipolar spectrum'.

(e) 'Mixed states'

Here the individual with a bipolar disorder shows depressive features during a manic episode or manic features during a depressive episode. While sometimes used to describe the transition from one polar mood disturbance to another, it more commonly refers to the coterminous presence of manic and depressive features. Clinically, such patients more tend to report perturbing agitation rather than elevated mood in conjunction with depressive symptoms.

Depressive disorders: contrasting models

Unitary or binary?

The extended debate as to whether the depressive disorders are best conceptualized as comprising one or more distinct disorders warrants overview. The 'unitarian' model presupposes one depressive disorder, varying essentially by severity. The strict 'binarian' view postulated two separate types (i.e. 'endogenous'/'psychotic' versus 'neurotic'/'reactive'). There were a number of ascriptions to the 'endogenous' (now termed 'melancholic') type. Firstly, as indicated by its naming, its determinants weighted genetic and other biological factors rather than exogenous psychosocial factors. Secondly, that it had a distinctive pattern of ('endogeneity') symptoms and signs—noted shortly. Thirdly, that it showed a preferential response to physical treatments (e.g. antidepressant drugs and ECT) and less responsivity to psychotherapy. By contrast, the second 'neurotic' or 'reactive' depressive type was viewed as more reflecting depression emerging as an interaction of a predisposing personality style and precipitating life-event stressors.

The debate was strongly influenced by Lewis's clinical study[4] finding no clear demarcation between depressive types, examined both cross-sectionally and longitudinally, thus delivering support to the unitarian view. The introduction of multivariate statistical approaches led to the debate being reactivated in the 1960s, with the so-called Newcastle School arguing strongly that their analyses supported a binary view. In a representative paper, Kiloh and Garside[5] used a factor-analytic strategy to argue for separate 'endogenous' and 'neurotic' depressive conditions. However, factor analysis is not ideal for developing a typology, in that it produces dimensions (here of symptoms) rather than groupings of patients. Subsequently, more appropriate strategies have been used, such as cluster analysis[6] and latent class analysis,[7] and with those studies providing some support for separate classes. Critics suggest, however, that such classes or subgroups could still be determined by severity or, even if sub-types can be identified, question whether sub-classification has any management importance.[8–10]

This latter challenge is fundamental, taking us to the heart of any consideration of the diagnosis and classification of the depressive disorders. To the unitarians, as depression essentially varies only by degree, treatment choices (e.g. electroconvulsive therapy (ECT), antidepressant drugs, psychotherapy, or cognitive behavioural therapy) are commonly decided on the basis of severity. The opposing argument—for conceding sub-types—was well put by Kendell,[11] who drew on historical analogies. For example, he noted that distinguishing between cardiac and renal forms of 'dropsy' allowed prediction of those who would respond to digitalis.

Thus, if there are valid depressive sub-types, the contribution of putative psychosocial and biological risk factors may vary across each, and exert differential responses to differing treatment modalities. If the sub-typing model is valid, forcing homogeneity by creating dimensionally based categories such as 'major depression' will ensure muddied results. As noted by Hickie,[12] numerous studies of patients with DSM-defined 'major depression' have failed to demonstrate any coherent pattern of neurobiological changes, replicate key biological correlates, and demonstrate any specific pattern of treatment response outside inpatient treatment settings.

Approaches in the classificatory systems

How then have the official classificatory systems addressed such a substantive issue? In developing the DSM-III system,[13] the working group was required to make a decision on the competing unitarian or binarian models. While the binarians were at the door, they had, until then, failed to prove their case and the DSM-III committee chose a compromise. Thus, DSM-III depression classification was predicated on an initial dimensional component (i.e. 'major' versus 'minor' disorders). If criteria for a major disorder were met, second-order and more categorical decisions about the

presence of melancholia or psychotic depression were specified. This model proved unsatisfactory for melancholia. For example, Zimmerman and colleagues[14] noted that the DSM-III melancholia criteria set, unlike the definition provided in the predecessor (DSM-II), 'did not predict treatment response'. Thus, the DSM-III-R[15] criteria set for melancholia was revised to include complete recovery after previous episodes, previous good response to somatic treatments, and no significant personality disturbance, to overcome the lack of predictive validity by building into the definition some of the 'givens' held by many clinicians about melancholia. However, the criteria set for melancholia developed for DSM-IV returned essentially to the DSM-III set, with limitations considered below. The contrasting system, ICD-10, is essentially based on a stricter dimensional or unitarian view of the depressive disorders—comprising 'severe', 'moderate', and 'mild' disorders.

During the extended debate as to whether a categorical and more 'biological' type of depression exists—it was variably termed 'endogenous', 'endogenomorphic', 'autonomous', and 'melancholic' depression. The last is probably preferable as numerous studies have quantified few or no differences in the likelihood of those with 'endogenous' and 'non-endogenous' depression reporting antecedent life events, so arguing against any term weighting 'internal' or 'external' causes.

Whether psychotic (or delusional) depression is a 'severe' form of melancholia or a separate entity also remains problematic. DSM-III had a category 'major depression with psychotic features' for use when delusions or hallucinations are present or when there is 'depressive stupor (the individual is mute and unresponsive)', thus viewing 'psychotic depression' as a sub-type of the generic 'major depression' category rather than a sub-type of melancholia. While 'depressive stupor' may be a useful marker or proxy for the condition, this criterion was not retained in DSM-III-R or DSM-IV, but is included in ICD-10. Two points argue for psychotic depression as a distinct entity: the presence of psychotic features, and its poor response to antidepressant medication alone and to neuroleptic medication alone in comparison to high responsiveness to their combination.[16]

A strict interpretation of the 'binary' view would place the non-psychotic, non-melancholic depressive conditions in a pure second class. However, rather than view this as a pure 'type', this class is best regarded as a heterogeneous residue category (i.e. non-melancholic depression), with its heterogeneity expressed widely—across aetiological factors, clinical expression, and natural and treated history.

Classification of affective mood disorders

Formal classification—depressive disorders

Both ICD-10 and DSM-IV have multiple conditions and specifiers. The ICD-10 system allows mild and moderate depressive episodes (with or without a 'somatic syndrome' conceptualized as reflecting 'melancholic' features), and severe depressive episode (with or without psychotic symptoms). There are separate codes for a similar set of 'recurrent' disorders, while several 'persistent' mood disorders (including cyclothymia and dysthymia) and residual conditions are listed. DSM-IV has two principal 'stem' disorders (major depressive episode and dysthymia), with the first having a number of optional specifiers including 'with' melancholic, catatonic, psychotic, or atypical features, as well as including disorders showing longitudinal patterns of rapid cycling or a seasonal pattern. Both systems have categories for affective disorders secondary to organic disease, while DSM-IV includes mood disorders due to a general medical condition or substance use, or occurring in the post-partum period. Both classificatory systems include adjustment disorders with depression.

Formal classifications are therefore built principally on severity, features of current episode, patterns of disorder expression over time, as well as persistence and recurrence. Few diagnoses are consistent across the ICD-10 and DSM-IV systems and, while each provides definitions that allow a 'shared language' to be used by clinicians and researchers, the extent to which their severity-weighted groupings capture 'meaningful' depressive sub-types remains problematic.

For example, and as detailed elsewhere,[17] 'major depression' has come to be viewed as an entity, sufficient in and of itself for testing antidepressant therapies and to generate treatment recommendations. Limitations to such a model become apparent if we consider the analogy of 'major breathlessness', which could be a transient consequence of acute exercise, or reflect quite differing pathological processes (e.g. asthma, pneumonia, or a pulmonary embolus) benefiting from quite differing treatment approaches. Thus, a diagnosis of 'major depression' or 'clinical depression' is, in reality, a first-level domain diagnosis, and benefiting from secondary specification. The latter tends to proceed on the basis of severity, but alternative and more categorical models have long been proposed as considered elsewhere in this chapter.

Formal classification—bipolar disorders

The DSM-IV definition effectively requires an initial or previous manic episode for bipolar I disorder, while bipolar II disorder requires hypomanic episodes and one or more previous episode of major depression.

To meet DSM-IV diagnostic status, manic episodes must have lasted 7 days and hypomanic episodes 4 days. Both ICD-10 and DSM-IV have course specifiers for bipolar disorder containing 10 and 4 subgroups, respectively. In addition to the number of subgroups, differences include a greater emphasis on distinguishing bipolar I and II in DSM-IV, and cyclothymia being listed as a 'bipolar disorder' in DSM-IV as against being a 'persistent' mood disorder overlapping with a personality style in ICD-10. Distinguishing 'hypomania' and 'mania' is regrettably imprecise in the formal classificatory systems. Both DSM-IV[18] and ICD-10[19] disallow a diagnosis of hypomania if psychotic features are present but, conversely, do not require psychotic features for a diagnosis of mania. DSM-IV lists essentially similar clinical criteria (and criteria number cut-off) for hypomanic and manic episodes, but distinguishes mania by the presence of marked impairment in social functioning (risking subjective judgement), requirement for hospitalization (which is logically more a consequence than a defining criterion although it may have some proxy value), and the presence of psychotic features in a manic episode. As noted earlier, ICD-10 views hypomania as 'an intermediate state without delusions, hallucinations, or complete disruption of normal activities'.

Thus, formal 'cleavage' between bipolar I and II (and constituent manic and hypomanic states) is largely dimensional in relation to clinical features and with some logical fallacies. Further, duration

criteria (i.e. at least 7 days for mania and at least 4 days for hypomania in DSM-IV) do not appear sustainable. In recent years there have been many studies[20] indicating that clinical definition of bipolar disorder is not dependent on the duration of the highs, and that imposing DSM-IV duration criteria for both mania and hypomania may exclude a significant percentage of those with true bipolar disorder.

A sub-typing model for classifying depression

As detailed earlier, there are intrinsic difficulties in classifying depression according to any single model when it is a term encompassing normal mood states through to possibly categorical diseases, and when any imposition of a severity-based model raises problems about how to differentiate meaningful groups (e.g. 'cases' from 'non-cases'). A personal mixed model is now detailed for consideration—one shaped by clinical experience and supported by research findings. It is described in line with the 'reasoning steps' that a clinician might employ in assessing a potential depressive disorder.

(a) Step 1: Is a depressive disorder present?

For all the depressive disorders, the first building block generally requires evidence of a depressed mood (although some with a melancholic or psychotic depression may deny 'depression'). Useful questions include the following: 'Do you feel depressed'?, 'Has there been any change in your self-esteem or from the way you generally value yourself'?, and 'Are you being more self-critical or harder on yourself than usual'?

The next clinical priority is to determine if the depression is sufficiently severe as to warrant 'case' status, and here the DSM-IV criteria for a major depressive episode have common acceptance in terms of representative symptoms and duration criteria. That criteria set lists the following:

- four mood items (depressed mood, loss of interest or pleasure, feelings of worthlessness or inappropriate guilt, recurrent thoughts of death, and suicidal ideation)
- weight changes
- sleep disturbance
- fatigue
- impaired concentration
- psychomotor disturbance

A positive diagnosis requires five or more of the nine, evidence of functional impairment, and a minimum duration of 2 weeks.

(b) Step 2: If a depressive disorder is present, what sub-type?

If 'caseness' criteria are met, the next decision should be to determine the diagnostic sub-type. We favour a three-class hierarchical model (i.e. respectively psychotic, melancholic, and a heterogeneous non-melancholic class)—hierarchical in the sense that those in the two highest classes have class-specific features as well as possessing features of the subordinate classes. Clinical assessment then allows a sequencing process to diagnosis.

In essence: does this individual have a melancholic or non-melancholic depression; are there psychotic features indicative of a psychotic depression? That sequence will now be detailed. The left half of Fig. 4.5.3.1 details the so-called 'structural model'. A shared

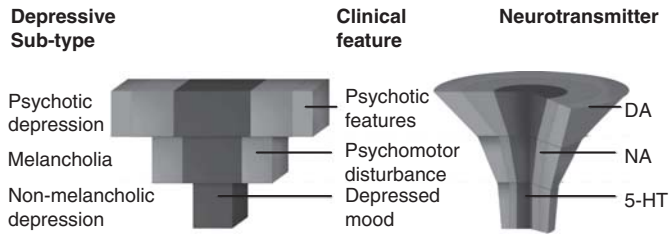


Fig. 4.5.3.1 The structural and functional model of three depressive classes. DA = dopamine; NA = noradrenalin; 5-HT = serotonin.

mood disorder component is present and varying in severity across the three principal sub-types. However, sub-type distinction proceeds on the basis of two class-specific components—psychomotor disturbance and psychotic features.

(c) Step 3: Differentiating melancholic and non-melancholic depression

In DSM-IV, the 'melancholic features specifier' requires, in addition to a base diagnosis of major depression, either one of two A criteria and three (or more) of six B criteria, with most of the latter comprising so-called 'endogeneity symptoms'. However, some such DSM criteria have limitations. Firstly, criterion A (loss of pleasure and/or lack of mood reactivity) is generally met by both 'melancholic' and 'non-melancholic' patients. Secondly, some criterion B features are vague. For example, 'distinct quality' is defined as a mood different to that experienced after 'the death of a loved one', and is a negative definition (akin to defining 'psychiatry' as 'not cardiology'). A second, 'excessive or inappropriate guilt', is a concatenated descriptor, subsuming the excessive expression of normal guilt, as well as guilt held at an overvalued or delusional level. Others are non-specific (e.g. early morning wakening, significant anorexia, or weight loss), being common in other psychiatric conditions (e.g. anxiety disorders) as well as in other expressions of depression. Further, as DSM-IV-defined major depression and melancholia share several similar features (e.g. anhedonia, psychomotor disturbance, weight loss), even those two conditions are poorly cleaved.

Features of melancholia: Any improvement on DSM-IV distinction of melancholia would benefit from identification of features specific to—or distinctly over-represented in—melancholia. Nelson and Charney[21] undertook a review of 33 studies using multivariate statistical approaches to identify melancholic or 'endogeneity symptoms'. They found no support for appetite/weight loss and insomnia, little support for early morning wakening and 'distinct quality', but some support for a severely depressed and non-reactive mood, loss of interest in pleasurable activities (or anhedonia), and psychotic features. The most strongly associated feature was psychomotor change (with retardation more consistently associated than agitation). When Rush and Weissenburger[22] examined nine diagnostic systems for diagnosing melancholia or endogenous depression, the only common criterion in all nine was psychomotor retardation (with psychomotor agitation included in six). Our research[23] has established that the specificity of psychomotor disturbance to 'melancholia' is dependent on measuring it as a sign, with that diagnostic weighting more recently detailed and endorsed by others.[24] Thus, and returning to a hierarchical model,

differentiation between the non-melancholic and melancholic disorders (on the basis of clinical features) appears assisted principally by the specific feature of behaviourally evident psychomotor disturbance. As measured by the sign-based CORE system,[23] psychomotor disturbance is reflected along three dimensions—impaired cognitive processing and motor retardation and agitation, although components are not mutually exclusive. For example, those with significant agitation may have it present for much of the time or, and more commonly, have a base of retardation with intermittent epochs of agitation.

In younger individuals with true melancholia, overt psychomotor disturbance is less distinctive, although they still tend to report distinct concentration problems. In addition to such 'signs', symptoms seemingly over-represented in melancholia include: distinct anergia often preventing the individual from getting out of bed to bathe; anergia and mood distinctly worse in the morning (i.e. diurnal variation); and an anhedonic and non-reactive mood.

The non-melancholic disorders have no specific features—apart from some (e.g. mood reactivity) that are the converse of their expression in melancholic disorder (i.e. non-reactive mood). Thus, these conditions are effectively diagnosed by excluding the two higher order conditions of melancholic and psychotic depression.

(d) Step 4: Implications of the distinction

Classification should never be sterile. It should at least provide us with a lexicon and, ideally, inform us about management nuances. We suggest that there are important treatment implications associated with distinguishing melancholic and non-melancholic disorder. The right half of Fig. 4.5.3.1 offers a 'functional model' that operates in parallel with the 'structural model'. As detailed,[25] it assumes that if there is any neurotransmitter perturbation underpinning the non-melancholic disorders, then it is principally serotonergic in origin, shaping the hypothesis that there is no advantage in proceeding beyond a narrow-action Selective Serotonin Reuptake Inhibitor (SSRI) antidepressant. It further assumes a greater noradrenergic contribution (and possibly dopaminergic contribution) to melancholia, and shapes the hypothesis that broad-action antidepressants will overall be more effective than SSRI antidepressants—and any psychotherapy—for melancholic depression. Both hypotheses have been supported in a number of trials and naturalistic studies[26] arguing the importance of distinguishing the melancholic sub-type.

(e) Step 5: Distinguishing psychotic depression

In psychotic (or delusional) depression, the central mood component is generally even more severe than in melancholic depression but many will deny or minimize a depressed mood. A number of 'endogeneity symptoms' are also frequently more severe (particularly non-reactive mood, anhedonia, and constipation). One frequent feature in melancholic depression (i.e. diurnal variation of mood) is, however, rarely present at episode nadir, as the patient is more likely to be persistently depressed across the days. Observable psychomotor disturbance is present and generally distinctly more severe than in melancholic depression. In some, the combination of the cognitive processing problems and motor change (retardation in particular) can give the impression of a dementing process, and provides an example of 'pseudodementia'.

The key class-specific feature, however, is the presence of psychotic features. Delusions are almost invariably present while hallucinations (auditory most commonly) are present in 10 to 20 per cent. DSM-IV subdivides delusions and hallucinations as 'mood congruent' (where themes of guilt, disease, death, nihilism, and personal inadequacy dominate) and 'mood incongruent' (where psychotic features appear independent of the depressive theme and might include persecutory themes, delusions of control, as well as thought insertion and thought broadcasting) states. It is important to emphasize that mood-incongruent states are common and do not, by themselves, necessarily challenge a diagnosis of delusional depression or argue for a schizophrenic illness of necessity.

A significant percentage (approximately one-third) of patients with psychotic depression develop constipation. In many it is a primary feature (not merely a side-effect of psychotropic medication), and may serve as a nidus of delusional interpretation (e.g. the depressed patient who believes that their bowels have turned to concrete, or that they have a bowel cancer).

When psychomotor disturbance is extremely severe, it may not be possible for psychotic features to be elicited, particularly if the patient is mute. Many patients are diffident about revealing psychotic material, and here indirect questions can often be useful. In particular, pursuing the presence of 'guilt' can assist, with guilt here defined as a sense of self-blame and not merely self-criticism, together with a sense of remorse for wrong acts or omissions that are independent of any concern about potential evaluation by others. In psychotic depression, the guilt is more likely to be held at the level of an overvalued idea or at a formally delusional level. If direct inquiry does not elicit delusional material, then asking 'Do you feel any sense that you deserve to be punished'? can help elicit previously unexpressed psychotic material.

Our 'functional model' (Fig. 4.5.3.1) argues that there is likely to be a greater dopaminergic (than noradrenergic and serotonergic) contribution to psychotic depression, a hypothesis supported by meta-analyses[16,27] quantifying this condition as showing a 25 per cent response to an antidepressant alone, 33 per cent to an antipsychotic alone, and 80 per cent to their combination and to ECT.

(f) Step 6: If a non-melancholic depressive disorder, can this class be sub-typed?

There is no generally accepted sub-typing system for this residual group, and where symptoms reflecting the lower order mood construct dominate the clinical picture. In the absence of any class-specific features, any categorical sub-typing model is unlikely to be valid, and dimensional models more appropriate. But what are the salient constructs for dimensionalizing? DSM-IV and ICD-10 proceed largely on a severity dimension for the mood disorder (e.g. 'major' and 'minor') but also on patterns of recurrence and persistence, all appropriate candidate dimensions.

Historically, terms such as 'neurotic depression' and 'reactive depression' were used, with the former emphasizing a pre-morbid style of neuroticism and high anxiety, and the latter defining depression developing largely in response to life-event stressors. This suggests another set of candidate constructs, dimensionalizing both personality and severity of stress components—and modelling their interaction.

Several earlier studies argued for some clinical utility emerging from such a model. Thus, an early factor-analytic study[28] suggested both a 'hostile' type (evidenced by irritability as well as anxiety) and an 'anxious-tense' type. Blashfield and Morey[29]

reviewed 11 cluster analytic studies suggesting separate 'hostile' and 'anxious' depressive subgroups. In an extensive review of the then published studies, Roth and Barnes[30] suggested three principal subgroups, with depression associated with a personality disorder, in addition to 'hostile' and 'anxious' depression.

While such 'hostile' and 'anxious' subgroups have been identified for a lengthy period, clear and consistent descriptions are lacking. Grinker *et al.*[31] described those with 'hostile' depression as unappreciative, actively angry, provocative, and making excessive demands of—and complaints about—their therapists, suggesting a personality disorder contribution. The second ('anxiety') subgroup is variably interpreted as defining either those with an anxious personality or temperament, or the presence of significant coterminous anxiety symptoms when primarily depressed.

The model we favour is one that respects historical description (i.e. 'reactive' versus 'neurotic') but develops the 'personality' contribution beyond the simple diagnostic allocation of 'neurotic depression'.

Terms such as 'reactive depression' or 'situational depression' (or as used in DSM-IV and ICD-10, 'adjustment disorder') concede that some individuals develop a non-melancholic depressive disorder largely or purely as a consequence of a stressful life event, which may produce, trigger, and/or maintain depressive episodes. In an empirical study,[32] the authors were unable to establish clinical, family history, and even life-event stress differences between those with 'situational' and 'non-situational' major depression. Our research[33] indicated that for the acute reactive 'disorders', the impact emerged less from the severity of the stressor and more from its perceived 'meaning' or 'salience'. We suggested a cognitive 'lock and key model', whereby individuals may be predisposed by perturbing developmental events such as a highly judgemental parent (creating 'locks'). In adult life, exposure to a mirroring situation (e.g. a judgemental boss) might act as a 'key' for precipitating a 'reactive' depression. For more chronic 'situational depression' scenarios, here we presume that the stressor is a chronic one and/or that the stressor induces a 'learned helplessness' mind set in the individual, where they believe that it does not matter what they do or attempt to do—there will be no impact on outcome—and they develop a sense of 'powerlessness' along with depressive symptoms.

A spectrum model

In modelling non-melancholic disorders reflecting a personality contribution, we suggest the utility of a 'spectrum' model, a term variably used but which argues for a continuum between temperament/personality style and symptom states,[34] or some level of inter-dependency. This spectrum model views certain biological factors as shaping temperament and personality style—which then shape surface marker symptoms during a non-melancholic depressive episode. An earlier research report[35] identified those presenting with an 'irritable/hostile' depression as being more likely to have a cluster B personality and to report 'acting out' behaviours when stressed (i.e. demonstrating a 'short fuse' response to stress). By contrast, those with an 'anxious depressive' spectrum disorder appeared more likely to internalize anxiety. They tended to have shown shyness and behavioural inhibition in childhood, to have high rates of lifetime anxiety disorders, score high on trait anxiety, view themselves as 'worriers', 'nervy', or 'tense', and to rate as having a cluster C personality style. When stressed, they were somewhat more likely to 'act in' by becoming quiet, retiring to their room, crying, and 'stewing'. Thus, anxiety was evident both in the temperament pattern and in the prominent symptom profile when depressed—demonstrating the 'spectrum'. The suggested profile of these spectrum disorders is not only important for clinical consideration, but in facilitating research into possible neurobiological determinants and to consider any treatment specificity. For example, Blashfield and Morey[36] reviewed studies indicating that 'anxious depressives respond well to major and minor tranquillizers but not to tricyclics, while hostile depressives show little improvement with conventional drug therapies'. Further, Fava *et al.*[37] reported that anxious depressives were more likely to be responders to a selective serotonin reuptake inhibitor (SSRI) than other depressive expressions (including 'hostile depression').

We have more recently pursued the 'spectrum model' for non-melancholic disorders beyond the two candidate groupings (i.e. hostile/irritable and anxious depressive) considered above. Both clinical observation and literature review indicated eight predisposing personality styles[38] but an application study[39] suggested that the spectrum model could be supported up to a six-factor personality model. The latter spectrum model is now detailed.

Those who scored high on the personality measure of 'anxious worrying' were, when depressed, indecisive, and self-blaming, as well as feeling anxious and tense. Those high on the 'irritability/snappiness' personality dimension were likely, when depressed, to be irritable and impulsive. Rather similar personality dimensions of 'personal reserve' and 'social avoidance' identified some differential coping responses apart from a shared social withdrawal response, with the former avoiding others and losing interest in people, while those high on social avoidance reported avoiding pleasurable activities. Those high in 'perfectionism' were more likely to focus on trying to solve the problem and to seek distraction. The sixth personality style benefits from more detailed consideration as it informs us about the nature of 'atypical depression', a condition which—while listed in DSM-IV and held to show a specific response to monoamine oxidase inhibitor antidepressant drugs—has been quite variably interpreted over time.[40] In that review, we argued for the primacy of the personality style of 'rejection sensitivity'. In our spectrum study,[39] those high on 'rejection sensitivity' were distinctly more likely to report some of its characteristic features (e.g. food cravings, hypersomnia), but also a set of cognitive extensions of their personality style (e.g. a tendency to feel abandoned, rejected, and put down) as well as distinct self-consolatory strategies (e.g. spending money, seeking support, and even crying).

While such a spectrum model is not a categorical one (in circumscribing 'pure' depressive conditions) nor sufficiently consistent in expression to deserve endophenotypic status, such 'fuzzy set' patterning may have clinical implications in terms of assisting a richer diagnostic formulation and in providing therapies that address any predisposing personality style, rather than merely treat the surface marker of non-melancholic 'depression'. Much research will be required to determine if candidate expressions show any differential response to contrasting antidepressant strategies, necessary to ensure that classification has clinical utility.

A sub-typing model for classifying bipolar disorders

As considered earlier, the formal official classificatory systems fail to provide pristine cleavage between bipolar I and bipolar II disorders. We briefly describe a model that supports an historical approach to this issue (by preserving the definition of mania—and bipolar I status—to those who have had psychotic features during a 'high') and which links the bipolar disorders with psychotic and melancholic depression.

The 'isomer model' illustrated in Fig. 4.5.3.2 builds on studies suggesting that 'bipolar depressive episodes' are highly likely to be psychotic or melancholic in type, albeit with some differences in that individuals not infrequently report 'atypical features' of hypersomnia and food cravings rather than more characteristic melancholic features of insomnia and appetite loss. The model was further shaped by findings derived in one clinical sample.[41] Firstly, bipolar patients who had experienced a psychotic 'high' had a 50 per cent chance of having a depressive episode with psychotic features. Secondly, of the bipolar patients who had never been psychotic during a high, none had experienced psychotic features.

The 'isomer model' therefore posits two contrasting 'mirror image states'. Those with bipolar states oscillate across parameters of energy and mood (elevated in highs, depressed in lows). More specifically, individuals with bipolar II disorder oscillate between non-psychotic depression and hypomania—never experiencing psychotic depression or psychotic mania. By contrast, those with bipolar I disorder have had psychotic manic episodes (by definition) and, when depressed, are at increased risk of having psychotic episodes.

The advantage to the model is that it respects and formalizes the historical view of regarding the presence or absence of psychotic features during a high as distinguishing mania and hypomania (and respective bipolar I and II disorders). Further, it suggests that aetiological studies might well focus on pursuing the nature of two oscillators—one of mood and energy, and one of psychotic features. Thirdly, it offers a template for pursuing management options for the two differing bipolar disorders. Currently, there are many treatment guidelines for managing bipolar I—but only informal templates for managing bipolar II disorder[42]—and it may be unwise to extrapolate such bipolar I guidelines for managing bipolar II disorder.

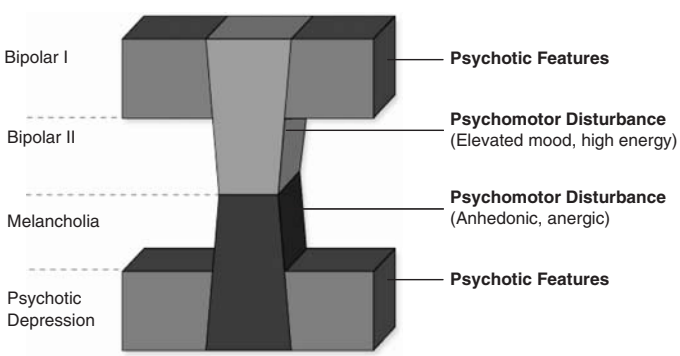


Fig. 4.5.3.2 The isomer model for conceptualizing bipolar I and bipolar II conditions. (Reproduced from Parker *et al.*[41])

Differential diagnosis and ascertainment difficulties

Depression

The three key features of lowered self-esteem, increased self-criticism, and a depressed mood distinguish depression phenomenologically from states such as grief or bereavement—where there is a distinct sense of 'loss' of something valued, but no primary 'loss' of self-esteem. They also assist phenomenological differentiation from anxiety, where the individual is more likely to report a sense of insecurity, fear, apprehension, worry, panic, or of 'going mad'. Subsequent questioning of clinical symptoms and observation should clarify if a clinical syndrome exists, its diagnostic sub-type and—from a longitudinal perspective—whether the individual has a unipolar or bipolar course.

At the practical level, such a definitional approach may fail in certain groups. A percentage of those with psychotic or melancholic depression appear more emotionally blunted and 'flat' rather than depressed. They may deny depression, self-blame and worthlessness, and instead note a lack of feeling (or vitality), sluggishness, enervation, or emphasize physical states, such as anergia. Less commonly, they may evidence 'corporization' by Schneider,[43] with reports of pain or physical sensations in the head, chest, or stomach. As noted earlier, others may have such profound psychomotor disturbance that they do not respond to questioning and appear as if they have a dementia (here a so-called 'pseudodementia'). In such instances, pursuit of proxy items (severe psychomotor disturbance, pathological guilt, overvalued ideas) may assist. If unsuccessful, the diagnosis may require corroborative reports and clinical observation, as well as certain investigations (e.g. CT or MRI scans, single-photon emission CT scans, EEG) to exclude a dementia. In a percentage of the elderly, however, a depressive episode and dementia may coexist (particularly a vascular dementia), and reflect a shared aetiological process.

Secondly, there are some patients and certain cultures where psychological issues are either denied or expressed somatically, although careful and directed questioning about central depressive descriptors will usually clarify the possibility. If unsuccessful, diagnostic clarification may require corroborative reports and clinical observation.

Thirdly, it is commonly difficult to define 'clinical depression' in those with a medical illness. Here, general depression criteria sets risk false-positive diagnoses by including certain features which may be secondary to the medical illness (e.g. fatigue, insomnia, anorexia) rather than reflecting depression *per se*. Common corrective strategies include the 'aetiological approach', where only symptoms viewed as independent of the medical condition are counted, the 'exclusive approach', eliminating potentially confounded items such as anorexia, the 'inclusive approach', where all symptoms are counted irrespective of their origin, and the 'substitutive approach', excluding features that could be due to the medical illness and substituting features such as social withdrawal and crying.

In some cases where depression has been established, the salient clinical difficulty is in determining whether depression is or is not the primary disorder in those with concomitant major medical problems, excessive alcohol intake, organic central nervous system disease, and certain psychiatric conditions (e.g. anxiety states, depressive personality disorder). Clinical judgement is generally

required with two alternate logical approaches: either weighting the disorders hierarchically or sequentially. The hierarchical approach assumes that the more severe disorder is the primary one, while the sequential approach weights the antecedent condition (e.g. organic disorder, schizophrenia, anxiety). Acceptance of one approach does not logically bind the clinician to any therapeutic consequence (such as necessarily treating only the more severe or primary condition).

Mania/hypomania

The differential diagnoses for manic states essentially include other psychotic conditions (e.g. schizophrenia, drug-induced psychosis) and, rarely, a primary organic state. While cross-sectional dissection of the phenomenology can be helpful, there is wisdom in also weighting the longitudinal course. Thus, those with manic states are more likely to describe complete restoration of function between episodes (of mania and/or depression), while this is less likely for those with schizophrenia. Definitive distinction is not always possible, and a diagnosis of 'schizoaffective' disorder may then be appropriate. In severe mania, an 'organic' picture may be suggested, and require exclusion of a dementia or delirium.

The differential diagnosis of hypomanic states is often quite difficult. Questioned about having 'highs' often elicits defensive response from those with true states, while some depressed people will present a remission to a euthymic state as a 'high'. Highly creative people may affirm many hypomanic descriptors when possessed by the muse (e.g. less need for sleep, feeling creative, and overconfident, being enthused and energized), as may those with a distinctly extroverted or cyclothymic personality when stimulated. Some patients with a cluster B personality style (especially of the borderline type) may also describe mood states that approximate to hypomania. Clarification is probably best assisted by a sequence of strategies, including asking the individual about times when, neither depressed nor feeling normal, they have states of feeling overly 'energetic and wired', that they have an appropriate number of concomitant criterion features during such highs, that there was a 'trend break' where 'highs' became a new phenomenon (most commonly mid-adolescence or later), and that—during their highs—any usual level of anxiety melts away, and by interviewing a corroborative witness.

Conclusions

Current formal classificatory systems list a large number of mood disorders, with criteria designed to assist diagnostic reliability. Most reflect attempts to create classes on the basis of severity-weighted dimensional models. Terms such as 'major depression' and 'unipolar depression'—as well as 'bipolar spectrum disorder'—have achieved acceptance in recent years, for such diagnoses are easily made (and easily reified), but the limitations inherent to their heterogeneity should not be ignored. Until the dissonance between the formal classifications and clinician-derived models has been resolved, practitioners should proceed by recognizing the advantages and limitations to competing approaches. A functional classificatory system should have clinical utility, going beyond mere description and informing the clinician about treatment differentiation.

Further information

Parker, G. and Manicavasagar, V. (2005). *Modelling and managing the depressive disorders. A clinical guide*. Cambridge University Press, Cambridge.

Parker, G. (2007). *Modelling and managing bipolar II disorder: a clinical guide*. Cambridge University Press, Cambridge (in press).

http://www.blackdoginstitute.org.au

References

1. Leonhard, K. (1957). *The classification of endogenous psychoses* (trans. R. Berman). Irvington, New York.
2. Goodwin, F.K. and Jamison, K.R. (1990). *Manic-depressive illness*. Oxford University Press, New York.
3. Phelps, J. (2008). The bipolar spectrum. In *Modelling and managing bipolar II disorder. A clinical guide* (ed. G. Parker), pp. 15–45, Cambridge University Press, Cambridge.
4. Lewis, A. (1934). Melancholia: a clinical survey of depressive states. *Journal of Mental Science*, **80**, 277–378.
5. Kiloh, L.G. and Garside, R.F. (1963). The independence of neurotic depression and endogenous depression. *The British Journal of Psychiatry*, **109**, 451–63.
6. Paykel, E.S. (1971). Classification of depressed patients: a cluster analysis derived grouping. *The British Journal of Psychiatry*, **118**, 275–88.
7. Parker, G., Hadzi-Pavlovic, P., Mitchell, P., *et al.* (1994). Defining melancholia: properties of a refined sign-based measure. *The British Journal of Psychiatry*, **164**, 316–26.
8. Parker, G. (2006). The DSM classification of depressive disorders: debating its utility. *Canadian Journal of Psychiatry*, **51**, 871–3.
9. Parker, G. (2006). Through a glass darkly: the disutility of the DSM nosology of depressive disorders. *Canadian Journal of Psychiatry*, **51**, 879–86.
10. Goldney, R.D. (2006). The utility of the DSM nosology of mood disorders. *Canadian Journal of Psychiatry*, **51**, 874–7.
11. Kendell, R.E. (1989). Clinical validity. In *The validity of psychiatric diagnosis* (eds. L.N. Robins and J.E. Barrett), pp. 305–23. Raven Press, New York.
12. Hickie, I. (1996). Issues in classification: III. Utilising behavioural constructs in melancholia research. In *Melancholia: a disorder of movement and mood. A phenomenological and neurobiological review* (eds. G. Parker and D. Hadzi-Pavlovic), pp. 38–56. Cambridge University Press, New York.
13. American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd edn). American Psychiatric Association, Washington, DC.
14. Zimmerman, M., Black, D.W., and Coryell, W. (1989). Diagnostic criteria for melancholia. The comparative validity of DSM-III and DSM-III-R. *Archives of General Psychiatry*, **46**, 361–8.
15. American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd edn). American Psychiatric Association, Washington, DC.
16. Spiker, D.G., Weiss J.C., Dealy, R.S., *et al.* (1985). The pharmacological treatment of delusional depression. *The American Journal of Psychiatry*, **142**, 430–5.
17. Parker, G. (2005). Beyond major depression. *Psychological Medicine*, **35**, 467–74.
18. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
19. World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders—clinical descriptions and diagnostic guidelines*. World Health Organization, Geneva.

20. Parker, G. (2008). Defining and measuring bipolar II disorder. In *Modelling and measuring bipolar II disorder. A clinical guide*. Cambridge University Press, Cambridge, 46–60.
21. Nelson, J.C. and Charney, D.S. (1981). The symptoms of major depressive illness. *The American Journal of Psychiatry*, **138**, 1–13.
22. Rush, A.J. and Weissenburger, J.E. (1994). Melancholic symptom features and DSM-IV. *The American Journal of Psychiatry*, **151**, 489–98.
23. Parker, G. and Hadzi-Pavlovic, D. (1996). *Melancholia: a disorder of movement and mood. A phenomenological and neurobiological review*. Cambridge University Press, New York.
24. Taylor, M.A. and Fink, M. (2003). Catatonia in psychiatric classification: a home of its own. *The American Journal of Psychiatry*, **160**, 1233–41.
25. Malhi, G., Parker, G., and Greenwood, J. (2005). Structural and functional models of depression: from sub-types to substrates. *Acta Psychiatrica Scandinavica*, **111**, 94–105.
26. Parker, G. (2007). Defining melancholia: the primacy of psychomotor disturbance. *Acta Psychiatrica Scandinavica*, **115**(Suppl. 433), 21–30.
27. Parker, G., Roy, K., Hadzi-Pavlovic, D., *et al.* (1992). Psychotic (delusional) depression: a meta-analysis of physical treatments. *Journal of Affective Disorders*, **24**, 17–24.
28. Overall, J.E., Hollister, C.E., Johnson, M., *et al.* (1966). Nosology of depression and differential response to drugs. *The Journal of the American Medical Association*, **195**, 946–8.
29. Blashfield, R.K., and Morey, L.C. (1979). The classification of depression through cluster analysis. *Comprehensive Psychiatry*, **20**, 516–27.
30. Roth, M. and Barnes, T.R.E. (1981). The classification of affective disorders: a synthesis of old and new concepts. *Comprehensive Psychiatry*, **22**, 54–77.
31. Grinker, R.R., Miller, J., Sabshin, M., *et al.* (1961). *The phenomena of depressions*. Harper and Row, New York.
32. Hirschfeld, R.M.A., Klerman, G.L., Andreasen, N.C., *et al.* (1985). Situational major depressive disorder. *Archives of General Psychiatry*, **42**, 1109–14.
33. Parker, G., Gladstone, G., Roussos, J., *et al.* (1998). Qualitative and quantitative analyses of a 'lock' and 'key' hypothesis of depression. *Psychological Medicine*, **28**, 1263–73.
34. Cassano, G.B., Michelini, S., Shear, M.K., *et al.* (1997). The panic–agoraphobic spectrum: a descriptive approach to the assessment and treatment of subtle symptoms. *The American Journal of Psychiatry*, **154**(Suppl.), 27–38.
35. Parker, G., Hadzi-Pavlovic, D., Roussos, J., *et al.* (1998). Non-melancholic depression: the contribution of personality, anxiety and life events to subclassification. *Psychological Medicine*, **28**, 1209–19.
36. Blashfield, R.K., and Morey, L.C. (1979). The classification of depression through cluster analysis. *Comprehensive Psychiatry*, **20**, 516–27.
37. Fava, M., Uebelacker, L.A., Alpert, J.E., *et al.* (1997). Major depressive subtypes and treatment response. *Biological Psychiatry*, **42**, 568–76.
38. Parker, G., Manicavasagar, V., Crawford, J., *et al.* (2006). Assessing personality traits associated with depression: the utility of a tiered model. *Psychological Medicine*, **36**, 1131–9.
39. Parker, G. and Crawford, J. (2007). A spectrum model for depressive conditions: extrapolation of the atypical depression prototype. *Journal of Affective Disorders*, **103**, 155–163.
40. Parker, G., Roy, K., Mitchell, P., *et al.* (2002). Atypical depression: a reappraisal. *The American Journal of Psychiatry*, **159**, 1470–9.
41. Parker, G., Hadzi-Pavlovic, D., and Tully, L. (2006). Distinguishing bipolar and unipolar disorders: an isomer model. *Journal of affective Disorders*, **96**, 67–73.
42. Parker, G. (2008). *Modelling and managing bipolar II disorder. A clinical guide*. Cambridge University Press, Cambridge.
43. Schneider, K. (1920). The stratification of emotional life and the structure of the depressive states. *Zentrallblat für Neurologie, Psychiatrie*, **59**, 281.

4.5.4 Epidemiology of mood disorders

Peter R. Joyce

The Global Burden of Disease, which is a comprehensive assessment of mortality and disability from diseases and injuries in 1990 and projected to 2020, highlights the importance of mood disorders for the world. Using the measure of disability-adjusted life years, it was determined that unipolar major depression was the fourth leading cause of disease burden in the world. It was also projected that, in the year 2020, unipolar major depression would be the second leading cause of disease burden in the world. Disability-adjusted life years is based on both mortality and disability. If one looks at disability alone, then unipolar major depression was the leading cause of disability in the world in 1990, and bipolar disorder was the sixth leading cause. Across the world, 10.7 per cent of disability can be attributed to unipolar major depression and, in developed countries, unipolar major depression contributes to nearly 20 per cent of disease burden in women aged from 15 to 44 years.[1]

The mood disorders have received considerable attention in psychiatric epidemiology over the last 25 years. These received particular attention in the five-site United States National Institutes of Mental Health Epidemiologic Catchment Area Study (**ECA**), as well as in the epidemiological studies in other countries around the world that used the ECA methodology. Mood disorders also received particular attention in the National Comorbidity Survey (**NCS**) in the United States, in the National Psychiatric Morbidity Survey of Great Britain, and most recently in the World Mental Health Survey (**WMH**) across many countries. Thus, there is substantial data from around the world on the epidemiology of these disorders. In addition, many of the population-based twin registries, such as in Virginia (USA), have also paid particular interest to mood disorders and have the additional advantage of being able to consider genetic as well as environmental risk factors.

Bipolar disorders

Diagnostic issues

While classical bipolar disorder with episodes of euphoric mania interspersed with episodes of depression is one of the clearest clinical syndromes in psychiatry, the boundaries of bipolar disorder remain contested. As case definition is central to epidemiology, all the contested boundaries of bipolar disorder could influence prevalence rates and our understanding of risk factors. Some of the major boundary issues for bipolar disorder include the overlap of bipolar disorder with psychotic features, with schizoaffective disorder and schizophrenia, and the overlap of bipolar disorder with unipolar major depression when patients who present primarily with depression have brief or mild episodes of hypomania. There is also an overlap of bipolar disorder with apparent personality disorder, especially Cluster B personality disorders such as borderline and narcissism, and the issue of when hyperthymic personality merges into bipolar disorder.[2,3] When bipolar disorder is comorbid with substance abuse there are also important diagnostic issues.

Another important issue in determining caseness of bipolar disorder for epidemiological surveys is symptom pattern and duration. A number of the diagnostic instruments for assessing bipolarity in population surveys limit the questions on mania to a type of symptom profile characterized by euphoria, grandiosity, increased energy, and decreased sleep. Whether the commonly used epidemiology interviews adequately detect those individuals who have manic episodes characterized by irritability, anger, and activation is very debatable. The other key diagnostic issue is what criteria are used to categorize the minimum duration for hypomania; is four days too long, is even two days too long? Furthermore, as insight is sometimes impaired in hypomania and mania, and as these are low prevalence disorders, the accuracy of case detection of bipolar disorders in populations remains an issue for further research.[4]

Prevalence

Population studies such as the ECA, and its related cross-national studies, and the NCS reported that the lifetime prevalence of bipolar disorder varies from 0.3 to 1.5 per cent. The NCS data include only bipolar I data, while the ECA includes bipolar I and bipolar II disorder.[4,5] In all studies, the six-month prevalence is not much lower than the lifetime prevalence of bipolar disorder. These findings reflect the high degree of chronicity and/or recurrence associated with bipolar disorder. Broader definitions of mania/hypomania have resulted in lifetime prevalence rates increasing to about 4 per cent.[6]

In these population studies, the mean age of onset of bipolar disorder has varied from 17 to 27 years. However, as age of onset is not normally distributed, the mean is a slightly misleading variable; in clinical samples, while the mean age of onset may be in the twenties, the most common age of onset are the teenage years.

In bipolar disorder, the prevalence in males and females is similar. This is in contrast to the reasonably consistent female excess found in major depression.

Comorbidity

In the NCS, all identified bipolar I individuals suffered from at least one, and often up to three or more, comorbid disorders. The most common comorbid disorders included the full range of anxiety disorders, alcohol and drug dependence, and conduct disorder or other antisocial behaviours.

Alcohol and drug abuse and/or dependence are commonly comorbid with bipolar disorder. Old studies found that binge drinking was especially common in bipolar individuals and that this binge pattern of drinking was more associated with manic episodes than with depressive episodes. Clinical studies find that bipolar patients with comorbid substance dependence are less compliant with prescribed mood stabilizers and have more frequent hospital readmissions. Stimulant abuse/dependence rates are especially increased in bipolar disorder.

Individuals with bipolar disorder have the full range of anxiety disorders, including phobias, panic disorder, and obsessive–compulsive disorder. Perhaps surprisingly, comorbid rates of these anxiety disorders tend to be higher in bipolar disorder than in major depression.

Another area of high comorbidity with bipolar disorder is that of childhood conduct disorder and attention deficit disorder. One of the issues in understanding this high rate of comorbidity is whether childhood conduct disorder and/or childhood attention-deficit disorder are sometimes the first manifestations or precursors of bipolar disorder. Certainly, if the pattern of conduct-disorder symptoms or attention-deficit symptoms is episodic rather than consistent over time, the issue becomes whether these are not early manifestations of bipolar disorder rather than truly independent comorbid conditions. The other key diagnostic controversy in this area is the status of juvenile or childhood bipolar disorder.

Use of health services

In the ECA study, 39 per cent of those with bipolar I or bipolar II disorders received outpatient psychiatric treatment within 1 year and about 10 per cent would receive inpatient treatment within a 6-month period. In the NCS study, 45 per cent of those with bipolar disorder had received psychiatric treatment in the previous 12 months; although 93 per cent reported lifetime treatment for their bipolar disorder. However, both of these studies suggest that more than half the individuals with bipolar disorder are not currently in psychiatric treatment and, given the high morbidity and mortality associated with bipolar disorder, this is of major concern.[4]

Risk factors for bipolar disorders

In considering the risk factors for bipolar disorder, it is useful to separate risk factors into those that are risk factors for lifetime vulnerability (for example genetic factors) and those that are risk factors for the onset of an episode of depression or mania (for example, life events). Thus, in determining risk factors for lifetime vulnerability, genetic factors constitute the largest single risk factor. However, if one is considering who is vulnerable to an episode of mania over the next six months, genetic factors will play a relatively smaller part and predictions may be best based on other factors such as past history, childbirth, being treated for depression with antidepressant medication, and the approach of spring or summer. Genetic risk factors are discussed further in Chapter 4.5.5.

Although organic factors, such as some type of central nervous system damage, are unusual risk factors in young adults, in late-onset bipolar disorder (age of onset more than 50 years) organic disease of the central nervous system is an increasing factor for the development of mania. In younger adults, AIDS and head injury are two important aetiological factors in a limited number of cases of bipolar disorder.

Risk factors for manic episodes in people with bipolar disorder

A range of other biological factors are particularly relevant risk factors to the onset of episodes of illness, but they may contribute a relatively small part to lifetime vulnerability. Many women have their first episode of depression or mania in the postpartum period. While a limited number of women may have manic episodes limited to the postpartum period, postpartum episodes of mania are more commonly part of a long-term bipolar disorder and these women will have episodes both precipitated by childbirth and at other times in their life. Indeed, in the postpartum period, having a history of bipolar disorder is one of the strongest risk factors for the development of a postpartum psychosis.

There is substantial evidence that seasonal patterns influence the onset of manic and depressive episodes. There are consistent

findings of an excess of manic episodes in late spring and early summer. To date, however, the nature of the environmental factors that influence this late spring, early summer peak of manic episodes is less clear.

There is also substantial evidence that disruptions of normal biological rhythms may precipitate the onset of manic or depressive episodes. This has been documented in relation to international travel involving east–west or west–east travel with disruption of circadian rhythms. Disruption of circadian rhythms through shift-work or other factors, which disrupt the normal sleep cycles, may also be important triggers to the onset of episodes of mania. These findings have led to the development of a social rhythm metric, as an adjunct to interpersonal psychotherapy (interpersonal social rhythms therapy) as a treatment for individuals with bipolar disorder.

Adverse life events have been well documented to be precipitants of manic episodes, as well as depression. It appears that life events are more likely prior to the first or second episode of mania and are less likely later in the course of illness. The critical factor in life events triggering mania may be whether there is associated sleep disruption, rather than the 'psychological' meaning of the event.

Depressive disorders

Diagnostic issues

A key issue for the epidemiology of depressive disorders is defining the boundaries of major depression and dysthymia. Depressive symptoms in the community are common, and defining both the symptom count and the duration at which depressive symptoms count as part of a clinical disorder is arbitrary. When Kendler and Gardner[7] examined the boundaries of major depression as defined by DSM-IV in a population-based twin sample of women, they found that, if a twin had four or fewer depressive symptoms, syndromes composed of symptoms involving no or minimal impairment, and episodes lasting less than 14 days, then the individual's co-twin was still at an increased risk of major depression. Kendler and Gardner concluded that they could find no empirical support for the DSM-IV requirement of duration for two weeks, five symptoms, or clinically significant impairment. These authors suggested that major depression, as articulated by DSM-IV, may be a diagnostic convention imposed on a continuum of depressive symptoms of varying severity and duration. Wainwright *et al.*,[8] using data from the National Psychiatric Morbidity Survey of Great Britain, have also suggested that research should move beyond a binary decision of case versus non-case, and utilize a probablistic measure of psychiatric case status, replacing the arbitrary threshold with a smooth transition. This type of approach allows the benefits of syndrome diagnosis to be retained, while not falling into the dilemma of an arbitrary threshold that lacks validity.

Provided that one accepts the arbitrary definition of major depression, then determining the rates of current depressive disorders is not especially problematic. However, there are major methodological issues involved in determining whether an individual has ever had a lifetime episode of major depression. Lifetime prevalence rates vary from 4.4 per cent in the United States ECA study, to 17.1 per cent in the NCS, and to over 30 per cent in Kendler's Virginia twin sample of women. In part, subjects in the community may forget or fail to report past episodes of major depression (recall bias), and the manner in which the questions are asked may importantly influence lifetime rates of depression. In the Diagnostic Interview Schedule, which was used in the ECA, respondents were asked about lifetime symptoms, a lifetime diagnosis was made, and then recency of the lifetime diagnosis was determined. More recent diagnostic interview schedules, such as the Composite International Diagnostic Interview, first ask about current depressive symptoms and then, having 'primed' individuals about depressive symptoms, go on to enquire about past depressive episodes. Interviews that follow the schedule of 'priming' before asking about past episodes appear to obtain considerably higher rates of lifetime major depression. Determining lifetime rates of depression with greater precision is an important task, as the vulnerability to depression conferred by risk factors such as genetic factors and childhood experiences may be wrongly estimated if lifetime rates of major depression are imprecise. For instance, when Kendler *et al.*[9] examined the heritability of major depression and corrected for the moderate reliability of a lifetime diagnosis of major depression, the heritability estimate increased from 40 per cent to over 70 per cent. As concluded by Kendler, major depression is not a disorder of high reliability and moderate heritability, but is a diagnosis of moderate reliability and high heritability.

DSM-IV allows major depression to be further subclassified into subtypes, such as melancholia, atypical, psychotic, and by severity and recurrence. Most of the traditional epidemiology studies have tended to ignore the issue of subtyping major depression. Recently, however, the issue of the atypical depression subtype has received particular attention in the study of the Virginia twins and in the NCS. In both these studies, latent class analysis suggests that atypical depression is a distinct subtype with several distinctive features, such as higher rates of parental alcohol- and drug-use disorders, higher interpersonal dependency, and higher rates of conduct disorder. If risk factors for atypical depression are, in part, distinct from risk factors for other subtypes of major depression, then for epidemiology to contribute to an understanding of aetiology it will be important to undertake further work on depressive subtypes.[10]

Prevalence

In the ECA, the six-month prevalence of major depression across five sites was 2.2 per 100. In ECA equivalent studies the six-month prevalence rate ranged up to 5.3 per 100. In the NCS, the 1-month prevalence of a major depressive episode was 6.1 per 100.[11] In the National Psychiatric Morbidity Surveys of Great Britain, the one-week rate of a depressive episode was 2.1 per 100.[12] Together, these studies would suggest that the current rate of major depression is in the realm of two to five per cent.

The estimates of the lifetime rate of major depression are much more variable. The lowest rate reported is 4.4 per 100 from the ECA study, while, in the study of Virginia twins, the lifetime rate of major depression is over 30 per cent. It is reasonable to believe that the true lifetime rate of major depression is probably in the realm of 10 to 20 per 100, but caution should be exercised in expressing lifetime rates of depression with undue precision.

These rates of major depression may also be lower if the rate of bipolar disorder is higher. Isolated clinical studies have found that one in two, not one in ten, individuals presenting with depressive disorders have features of bipolar spectrum disorders. If these

figures are correct, then this would presumably lower the rates of major depression, but would correspondingly increase the rates of bipolar disorders.

Over the past decade, one of the controversial findings in the epidemiology of major depression has been whether the rates of depression are increasing, and whether it is occurring at a younger age. Despite methodological concerns about the reliability of lifetime major depression, studies across countries have reasonably consistently documented an increasing rate of major depression with an earlier age of onset.[13] As mood disorders are the single largest risk factor for suicide, it is also of note that, in most Western countries, the rate of suicide, especially in young adults, increased considerably from the 1970s to 1990s, although the suicide rate is now declining in many countries. This could, however, reflect better recognition and treatment of depression.

Risk factors

(a) Genetics

There is now substantial evidence that genetic factors are of major importance as risk factors for vulnerability to major depression. While traditional estimates have put the heritability at about 40 per cent, when Kendler *et al.*[9] allowed for the moderate reliability of the diagnosis of major depression, the heritability estimate increased to 70 per cent. Of greater interest is that the genes for major depression do not appear to be unique for depression, but overlap with the genes for anxiety and the genes for neuroticism.[14, 15] The greater prevalence of depression in women may be due to the strong association of anxiety and neuroticism with depression, and that the higher rates of anxiety and neuroticism in women lead to higher rates of depression.

(b) Gender

One of the most consistent findings in the epidemiology of major depression is that the ratio of women to men is approximately 2:1. This increased rate of major depression in women arises during puberty, as in childhood there is a slightly higher prevalence of depression in boys than girls. The timing of this transition in rates by gender is related to biological puberty rather than just to age.

(c) Childhood experiences

Early theorizing suggested that the loss of a parent in childhood increased the later risk for major depression; although many studies have examined this issue, they have inconsistently found it to be a risk factor for adult depression. However, studies that examine the nature of child–parent attachment using a measure such as the Parental Bonding Instrument have consistently found that a lack of parental care is associated with increased rates of depression.[16] More recently, childhood sexual abuse has been established as a risk factor for adult major depression.

However, cumulative childhood disadvantage almost certainly poses a greater risk to later depression than any single childhood variable in isolation. Thus, if studies only look at single childhood risk factors, they may miss the full impact of global childhood adversity. The converse of childhood risk factors is childhood resilience, and it is probable that one good relationship with an adult and high intelligence in the child may, in part, protect from other adversities.

(d) Personality

There has been a long history of interest in the likelihood that people with certain personality traits are more vulnerable to depression than others. It is likely that those individuals who are unduly anxious, impulsive, and obsessional may have increased rates of later major depression. The best data exists for neuroticism, which emerges as a clear risk factor for the later development of depressive and anxiety disorders. However, as already mentioned, the same genes seem to contribute to the development of neuroticism and to later anxiety and depressive disorders.

(e) Social environment

There has been considerable interest in the role of marital status as a risk factor for major depression. For men, it appears clear that married men have the lowest rate of depression, while separated or divorced men have the highest rates of major depression. In women, the association is slightly less clear, but in the ECA study the same findings applied for women as for men. Understanding the nature of the association between marital status and rates of depression is more problematic. If personality is a risk factor for depression, then the same traits could interfere with the ability to marry or to stay married. There is little doubt that depression sometimes contributes to marital maladjustment and separation or divorce. Finally, the stresses associated with divorce and separation could increase the likelihood of an episode of depression occurring.

In the classic and influential work of George Brown on working-class women, having three or more children, a lack of paid employment, and the lack of a confident were risk factors for the development of an episode of depression. Subsequent studies have inconsistently replicated the risk factors of having children or lack of paid employment, but have supported the finding that the lack of a confident increases the risk of depression.

It is well established that adverse life events, particularly those characterized by loss, increases the risk of an episode of major depression. Interestingly, however, the life events which may constitute the greatest risk may be 'dependent' rather than 'independent' life events. The increased vulnerability to an episode appears to last for a period of two to three months following such an event.

Early thinking about depression suggested that there would be those depressions that occurred for largely biological reasons and those precipitated by adverse life events; however, it is now clear that such a dichotomous view is incorrect. Kendler *et al.*[17] showed that there is a significant genotype by environment interaction in the prediction of onset of major depression. They proposed that genetic factors influence the risk of onset of depression, in part, by increasing the sensitivity of individuals to the depression-inducing effects of stressful life events.

(f) Physical illness

Having a chronic or severe physical illness is associated with an increased risk for depression. The mechanisms behind this increased risk may vary depending upon the physical disorder. In disorders such as Parkinson's disease, it is possible that there are shared neurotransmitter abnormalities between Parkinson's disease and depression. In post-stroke depression, there is good evidence that the location of the lesion, at least in part, contributes to the rate of depression, which suggests a neuroanatomical/neurotransmitter connection between the physical illness and the likelihood of depression. For non-central nervous system

disorders, such as acute myocardial infarction, diabetes, and cancers, the mechanism for this association is less clear. However, at least in the case of patients with cancer, most do not suffer from major depression and, if they do, the key risk factors are family history and a past history of depression. This suggests that the stress associated with a serious or chronic physical illness may act by bringing out an individual's lifetime vulnerability to depression.

An integrated aetiological model

The ultimate purpose of studying risk factors for depressive disorders is to contribute to the development of an integrated aetiological model. The most promising research in this area has been performed by Kendler and colleagues on twins from the Virginia Twin Register.[18] In this study, both female–female and male–male twin pairs of known zygosity have been assessed on a series of occasions at longer than one-year intervals. A range of predictor variables; including genetic factors, parental warmth, childhood parental loss, childhood disorders, lifetime traumas, neuroticism, self esteem, social support, past depressive episodes, recent difficulties, and recent stressful life events have been examined to see how they contributed prospectively to the development of an episode of major depression over the next 12 months. In considering the results from this study, it is important to bear in mind the limitations of this landmark study, especially the fact that they were predicting the onset of an episode over 12 months and not predicting lifetime episodes. However, despite these caveats, Kendler and colleagues developed a model that predicted over 50 per cent of the variance in the liability to develop major depression in the next twelve months. The strongest predictors to depression were as follows:

- stressful life events
- genetic factors
- previous history of major depression
- neuroticism

It is of note that some of the risk factors exerted these effects directly, while other effects were largely indirect. Thus, 60 per cent of the effect of genetic factors on liability to depression was direct, but the remaining 40 per cent was indirect and largely mediated by past episodes of depression, stressful life events, and neuroticism. Variables such as perceived parental warmth had no direct effect on liability to develop an episode of major depression, but did impact upon neuroticism, a history of a past depressive episode, recent difficulties, and lifetime traumas.The most comparable prospective studies looking at risk factors for the development of major depression have been undertaken during the postpartum period, which is a time of increased risk of depression. In this special case, the most consistent risk factors are family history and a past history of depression, and there is lesser support for a lack of social support, neuroticism, and complications during childbirth.

As one of the key tasks of epidemiology is to contribute to an understanding of aetiology, models that integrate risk factors are important strategies for further research. They provide clinicians with predictive power, and can also guide intervention studies to prevent the onset of episodes of depression.

Comorbidity

One of the important contributions of epidemiology to the study of mood disorders over the past twenty five years has been the recognition of the extent to which depression and other psychiatric disorders are often comorbid. In both the ECA study and NCS, over two-thirds of all individuals identified as having an episode of major depression also met the criteria for one or more other psychiatric disorders. Not surprisingly, the most common comorbid disorders are anxiety disorders and substance-abuse disorders. In the NCS, the anxiety disorders with the highest odds ratios indicating comorbidity were generalized anxiety disorder, panic disorder, and post-traumatic stress disorder. It is also important to note that for most anxiety disorders, with the exception of panic disorder, the anxiety disorder usually predates the onset of the depressive disorder.[19] This is of considerable importance, as the risk factors for pure major depression differed from the risk factors for comorbid major depression. Furthermore, the cohort effects of increasing rates of major depression were largely attributable to increasing rates of comorbid major depression, rather than to increasing rates of pure major depression. These results raise important issues for prevention, as it may well be that targeting young people with anxiety disorders and could be a major step to the prevention of the development of later major depressive disorders.

The second key area of comorbidity with major depression is with alcohol dependence. Data from the Virginia Twin Register suggest that part of this comorbidity is due to shared genetic factors, although there is also a smaller common environmental risk factor to both disorders.

Another area of considerable comorbidity with major depression is the personality disorders. The comorbidity between major depression and these disorders is receiving considerable attention in clinical samples but, to date, there are only limited data in epidemiological samples on the importance of these patterns of comorbidity.

Use of health services

One of the major challenges for psychiatry presented by epidemiological studies of depression has been the consistent finding that the majority of cases of depression in the community are not recognized, diagnosed, nor treated. In the ECA study, it was found that 65 to 70 per cent of people with depression had visited a health professional in the last 6 months, but only 15 to 20 per cent had had a visit for a mental health reason and only about 10 per cent had seen a mental health specialist. Ormel *et al.*[20] found that patients with depression who present with largely somatic rather than psychological symptoms are extremely unlikely to be recognized by general practitioners. Even if major depression is recognized in the primary care setting, it is often not adequately treated.[21]

References

1. Murray, C.J.L. and Lopez, A.D. (1996). *The global burden of disease and global health statistics*. Harvard University Press, Boston, MA.
2. Akiskal, H.S. (2003). Validating 'hard' and 'soft' phenotypes within the bipolar spectrum: Continuity or discontinuity? *Journal of Affective Disorders*, **73**, 1–5.

3. Blacker, D. and Tsuang, M.T. (1992). Contested boundaries of bipolar disorder and the limits of categorical diagnosis in psychiatry. *American Journal of Psychiatry,* **149**, 1473–83.
4. Kessler, R.C., Rubinow, D.R., Holmes, C., *et al.* (1997). The epidemiology of DSM-III-R bipolar I disorder in a general population survey. *Psychological Medicine,* **27**, 1079–89.
5. Weissman, M.M., Bland, R.C., Canino, G.J., *et al.* (1996). Cross national epidemiology of major depression and bipolar disorder. *Journal of the American Medical Association,* **276**, 293–9.
6. Angst, J. (1998). The emerging epidemiology of hypomania and bipolar II disorder. *Journal of Affective Disorders,* **50**, 143–51.
7. Kendler, K.S. and Gardner, C.O. (1998). Boundaries of major depression: an evaluation of DSM-IV criteria. *American Journal of Psychiatry,* **155**, 172–7.
8. Wainwright, N.W.J., Surtees, P.G., and Gilks, W.R. (1997). Diagnostic boundaries, reasoning and depressive disorder, I. Development of a probabilistic morbidity model for public health psychiatry. *Psychological Medicine*, **27**, 835–45.
9. Kendler, K.S., Neale, M.C., Kessler, R.C., *et al.* (1993). The lifetime history of major depression in women. Reliability of diagnosis and heritability. *Archives of General Psychiatry,* **50**, 863–70.
10. Sullivan, P.F., Prescott, C.A. and Kendler, K.S. (2002). The subtypes of major depression in a twin registry. *Journal of Affective Disorders*, **68**, 273–84.
11. Blazer, D.G., Kessler, R.C., McGonagle, K.A., *et al.* (1994). The prevalence and distribution of major depression in a national community sample: the National Comorbidity Survey. *American Journal of Psychiatry,* **151**, 979–86.
12. Jenkins, R., Lewis, G., Bebbington, P., *et al.* (1997). The National Psychiatric Morbidity Surveys of Great Britain—initial findings from the household survey. *Psychological Medicine*, **27**, 775–89.
13. Cross National Collaborative Group (1992). The changing rate of major depression. Cross national comparisons. *Journal of the American Medical Association,* **268**, 3098–105.
14. Kendler, K.S., Neale, M.C., Kessler, R.C., *et al.* (1992). Major depression and generalized anxiety disorder. Same genes (partly) different environment. *Archives of General Psychiatry*, **49**, 716–22.
15. Andrews, G., Stewart, G., Allen, R., *et al.* (1990). The genetics of six neurotic disorders: a twin study. *Journal of Affective Disorders*, **19**, 23–9.
16. Parker, G. (1983). Parental 'affectionless control' as an antecedent to adult depression. A risk factor delineated. *Archives of General Psychiatry,* **40**, 956–60.
17. Kendler, K.S. and Karkowski-Shuman, L. (1997). Stressful life events and genetic liability to major depression: genetic control of exposure to the environment? *Psychological Medicine*, **27**, 539–47.
18. Kendler, K.S., Gardner, C.O. and Prescott, C.A. (2002). Toward a comprehensive developmental model for major depression in women. *American Journal of Psychiatry*, **159**, 1133–45.
19. Kessler, R.C., Nelson, C.B., McGonagle, K.A., *et al.* (1996). Comorbidity of DSM-III-R major depressive disorder in the general population: results from the US National Comorbidity Survey. *British Journal of Psychiatry,* **168**, 17–30.
20. Ormel, J., Koeter, M.W.J., Van den Brink, W., *et al.* (1991). Recognition, management and course of anxiety and depression in general practice. *Archives of General Psychiatry,* **48**, 700–6.
21. Schulberg, H., Block, H.R., Madonia, M.J., *et al.* (1996). Treating major depression in primary care practice, eight-month clinical outcomes. *Archives of General Psychiatry,* **53**, 913–19.

4.5.5 Genetic aetiology of mood disorders

Pierre Oswald, Daniel Souery, and Julien Mendlewicz

Introduction

Advances towards the understanding of the etiological mechanisms involved in mood disorders provide interesting yet diverse hypotheses and promising models. In this context, molecular genetics has now been widely incorporated into genetic epidemiological research in psychiatry. Affective disorders and, in particular, bipolar affective disorder (BPAD) have been examined in many molecular genetic studies which have covered a large part of the genome, specific hypotheses such as mutations have also been studied. Most recent studies indicate that several chromosomal regions may be involved in the aetiology of BPAD. Other studies have reported the presence of anticipation in BPAD and in unipolar affective disorder (UPAD).[1–3] In parallel to these new developments in molecular genetics, the classical genetic epidemiology, represented by twin, adoption and family studies, provided additional evidence in favour of the genetic hypothesis in mood disorders. Moreover, these methods have been improved through models to test the gene-environment interactions.

In addition to genetic approaches, psychiatric research has focused on the role of psychosocial factors in the emergence of mood disorders. In this approach, psychosocial factors refer to the patient's social life context as well as to personality dimensions. Abnormalities in the social behavior such as impairment in social relationships have been observed during episode of affective disorders, and implicated in the etiology of affective disorders. Further, gender and socio-economic status also emerged as having a possible impact on the development of affective disorders. Finally, the onset and outcome of affective disorders could also be explained by interactions between the social life context and the individual's temperament and personality. The importance of temperament and personality characteristics in the etiology of depression has been emphasized in various theories, although disagreement exists with regard to terminology and the etiology.

While significant advances have been done in these two major fields of research, it appears that integrative models, taking into account the interactions between biological (genetic) factors and social (psychosocial environment) variables offer the most reliable way to approach the complex mechanisms involved in the etiology and outcome of mood disorders. This chapter will review some of the most promising genetic and psychosocial hypotheses in mood disorders that can be integrated in interactive models.

Genetic epidemiology of mood disorders

The various strategies available to investigate genetic risk factors in psychiatric disorders belong to the wider discipline of genetic epidemiology. This combines both epidemiological and genetic investigations and has the primary objective of identifying the genetic and non-genetic (environmental) causes of a disease. Genetic epidemiological data in affective disorders has come mostly

from family, twin, adoption and segregation (within families) studies. Family, twin and adoption studies are the mainstay in establishing the genetic basis of affective disorders. These methods firstly demonstrated that genetic factors are involved in the aetiology of these disorders.[4] Twin and adoption data may also be used to investigate the relative contributions of genetic and environmental factors to the aetiology of a disease.[5] The exact contribution of these factors is not yet firmly understood for affective disorders but some studies provide contributing findings. The study of adoptees who are separated from their biological parents has consistently favoured the gene-environment hypothesis in the aetiology of diverse psychiatric disorders.[6] In adoption studies, both UPAD and BPAD have been described to be more frequent in biological relatives of the adopted subject, suffering from affective disorders than in adopted relatives.[7]

The diagnostic validation and the structure of the genetic and environmental risk factors in mood disorders are also approached in twin studies.[8] From the landmark study of Rosanoff *et al.* in 1934 to more recent works from Mc Guffin *et al.*[9] and Kieseppa *et al.*,[10] concordance rates for BPAD are higher in monozygotic twins (20–100 per cent) than in dizygotic twins (0–38 per cent).

Molecular genetics in affective disorders

The rapid advance in molecular genetic techniques over the last decade has generated a large database of DNA markers across the whole human genome and has enabled chromosomal regions throughout almost the entire genome to be studied in affective disorders. These studies have been performed mainly using linkage and association methodologies. Current linkage and association methods investigate heritable factors at a molecular genetic level, and enable genes to be mapped.[11] These approaches are mostly applied to BPAD, which is considered to be the 'core' phenotype in affective disorders. Linkage analysis tests the hypothesis that a linkage relationship exists between a known genetic marker and a trait which is known to be genetically determined but has not yet been mapped on a chromosome.[12] Two genetic loci are linked if they are located closely together on a chromosome. In linkage analysis, the distance between a marker locus and the gene under investigation is used for gene mapping. This method was originally designed to explore a major single genetic transmission and to evaluate the extent of co-segregation between genetic markers and the phenotype investigated in pedigrees. The major problems which linkage methodology face when applied to affective disorders are the complex aetiology and inheritance patterns. More than one locus are probably involved in susceptibility to these disorders, and the exact mode of transmission is not known. Mis-specification of the genetic parameters of the phenotype may lead to errors in linkage studies.[12] Furthermore, the linkage approach fails to detect minor gene effects which contribute to genetic susceptibility to the disorder.[13] More recently, genome-wide linkage studies have been performed on samples of families with multiply affected members.[14]

The association method offers an alternative strategy of studying genetic factors involved in complex diseases in which the mode of transmission is not known.[15] The association strategy does not require genetic parameters to be known (non-parametric method). The purpose of association studies is to compare frequencies of genetic marker alleles in patient and control populations in order to detect linkage disequilibrium. Linkage disequilibrium between the disease locus and the marker tested is defined as a level of concordance between the two loci which is higher than would be expected by chance. The major reason for this is their proximity on a chromosome. The major advantage of association studies is that they can detect genes with minor effects other than a single major locus (SML). The major limitation of this approach is that spurious associations between a genetic marker and a disorder may result from variations in allele frequency between cases and controls observed if the two populations are ethnically different (population stratification). It is important in this case to compare populations which are homogenous in their ethnic background. A further major difficulty in association studies is the interpretation of the precise meaning of the association observed.[16] The result may be interpreted as linkage disequilibrium between the disease locus and the associated marker allele(s). Alternatively, the associated marker may be interpreted as a susceptibility factor which is directly involved in the disease. The candidate gene approach in association studies is a useful method to investigate linkage between markers and diseases. A candidate gene refers to a region of the chromosome which is potentially implicated in the aetiology of the disorder concerned. The possibility of false positive results must be taken into account, as a very large number of candidate genes now exist. The probability that each of these genes is involved in the aetiology of the disorder is relatively low.

Linkage studies in affective disorders (See[17] for review)

From more than two decades of linkage studies, it seems that several chromosomal locations have been associated with affective disorders, sometimes with conflicting results. Mendlewicz *et al*[18] first reported possible genetic linkage between manic depression and coagulation Factor IX (F9) at Xq27 in 11 pedigrees. Another region of interest seems to be the chromosome 18 where the pericentromeric region was suggested to carry susceptibility genes. The chromosome 11 has been thoroughly investigated in AD but showed contradictory results. Chromosomes 4, 6, and 10 were also investigated with conflicting and/or unreplicated results. Darier's disease (keratosis follicularis), a rare autosomal dominant skin disorder associated with increased prevalence of epilepsy and mental retardation, whose gene was mapped on chromosome 12 (12q23–24.1), was found to cosegregate with BPAD in one pedigree. This result was replicated in several family studies. Genome-wide linkage analyses provide an accurate tool to study regions of interest. In BPAD, early positive and promising results were contradicted by further analyses. This fact is not surprising, since these studies were performed on small samples sizes, insufficient to replicate modest linkage signals.[19] Meta-analyses were thus performed on BPAD to increase the power to detect modest linkage signals.[14] Bipolar loci with evidence of linkage were found on the following arms: 4p, 6p, 6q, 9p, 10q, 12q, 13 q, 14q, 17q, 18p–q, 21q, 22q.[14,20] McQueen *et al.*[21] found susceptibility loci on chromosomes 6q and 8q by using a combined analysis of eleven linkage studies. A recent study from Schumacher *et al.*,[22] in four European samples, confirmed previously reported loci, 4q31 and 6q24, and provided evidence for a new linkage locus, 1p35-36.

Candidate genes in affective disorders

Serotonin markers

Dysfunction of the serotoninergic system has long been suspected in major depression and related disorders. Depression can successfully be treated with selective drugs which target serotonin receptors. The serotonin transporter may also be involved in susceptibility to affective disorders and in the response to treatment with these drugs. Most recent replication studies did not support these initial positive findings. This has been the case for 5HTT.[17] The tryptophan hydroxylase (TPH1) gene, which codes for the rate limiting enzyme of serotonin metabolism, is also an important candidate gene for affective disorders and suicidal behavior. Bellivier *et al*[23] reported a significant association between genotypes at this marker and BPAD, no association was found with suicidal behaviour. In a previous study in depressed patients suicidal behaviour has been associated with one variant of this gene.[24] The tryptophan hydroxylase isoform (TPH2) showed an association between BPAD and suicidality.[25]

Other candidate genes

Among other pathways, DRD2, DRD3, DRD4 and DAT1 were largely studied and replicated. Unfortunately, results remain conflicting. Recent studies have implicated neurotrophic factors in the underlying disease processes of affective disorders. Brain-derived neurotrophic factor (BDNF), the most abundant of the neurotrophins in the brain, enhances the growth and maintenance of several neuronal systems, serves as a neurotransmitter modulator, and participates in plasticity mechanisms such as long-term potentiation and learning.[26] Although promising, BDNF did not confirm its role in the pathophysiology of affective disorders.[27,28] Several new candidate genes from the well-known molecular cascades have been tested: PIK3C3 in the intracellular signalling pathway; PCDH11Y, a proto-cadherin and GSK3β, a target molecule of lithium. Finally, studies of circadian rhythm-related genes showed promising results in BPAD, such as ARNTL.[29,30]

Anticipation and expanded trinucleotide repeat sequences

Anticipation implies that a disease occurs at a progressively earlier age of onset and with increased severity in successive generations. This may explain the non-Mendelian pattern of inheritance observed in some inherited diseases. Anticipation has been found to correlate with specific mutations in these syndromes: expanded trinucleotide repeat sequences. An expanded repeat sequence is unstable and may increase in size between family members, leading to increased disease severity of the disorder.

Anticipation has been described in BPAD and in UPAD.[1,3] One study highlighted an association between Cysteine-Alanine-Glycine (CAG) trinucleotide repeats and BPAD illness in Swedish and Belgian patients with affective disorder.[31] CAG repeats have been detected by the Repeat Expansion Detection method (RED-method). This hypothesis has also been tested in a family sample of two-generation pairs with BPAD.[32] A significant increase in CAG repeats between parents and offspring generations was observed however, when the phenotype increased in severity, i.e. changed from major depression, single episode or unipolar recurrent depression to BPAD. This is the first evidence of genetic anticipation in BPAD families and should be followed by the identification of loci within the genome containing triplet repeats. CTG 18.1 on chromosome 18q21.1 and ERDA 1 on chromosome 17q21.3 are two repeat loci recently identified but were not found to be associated with BPAD.[33] A newly identified CTG/CAG repeat was found to be associated with BPAD.[34]

Phenotype definition

Facing the heterogeneity of results, it has been hypothesized that genetic factors could explain some symptoms or clinical features of the syndromes, such as severity of the disease, age at onset or gender predominance. Early-onset, and more specifically pediatric-onset, BD has been suggested to have its own pattern of genetic susceptibility factors. Family studies have consistently found a higher rate of BD among the relatives of early-onset BD patients than in relatives of later-onset cases.[35] Among the most recent studies, Faraone *et al.*[36] found, in a genome-wide scan, 3 regions of interest that may influence age at onset of mania in BPAD. Geller *et al.*[37] found an association between BDNF and BPAD with early-onset. Massat *et al.*[38] provided evidence for the influence of HTR2C in early-onset BPAD

More specific neurophysiologic, neuroimaging, neurocognitive, or neurochemical trait measures might identify homogeneous groups of patients. These 'traits' are called 'endophenotypes' and are believed to represent the genetic liability of the disorder among non-affected subjects.[39] Endophenotypes in BD are difficult to define. Circadian rythms, stress reactivity and appetite regulation have been proposed. Bipolar patients are also suggested to show inappropriate emotional responsiveness. Using emotional facial stimuli, depressed BD patients show impaired recognition of happy and sad facial expressions.[40] These findings confirm a particular pattern of characteristics in BD, and suggest that genetic factors may explain these characteristics, rather than the whole clinical picture.

Shared genetic predisposition between BPAD and schizophrenia

If BPAD and schizophrenia (SCZ) are distinguishable, they may share some characteristics. Indeed, family studies show partial overlap in familial susceptibility for these two conditions.[41] Evidence for linkage of both BPAD and SCZ were found on 18p11, 13q32, 10p14, 22q11–13 and 6p22.2.[42–44] More interestingly, two genes showed promising results in molecular genetic studies in these two conditions. From the first report from Hattori *et al.*,[45] G72, found on 13q34 and encoding d-amino acide oxidase activator (DAOA), was found to be associated with delusion or psychosis, rather than with the entire bipolar or schizophrenic clinical pictures.[46] Although robust, these results were confirmed in a recent meta-analysis from Detera-Wadleigh *et al.*[47] Largely distributed in neurones, DISC1 (Disrupted in Schizophrenia 1) interacts with many proteins and is related to several neuronal functions. Thomson *et al.*[48] found a robust association between DISC1 and BPAD, but this result needs confirmation. Theses cases of shared genetic predisposition emphasizes the need, in future classification systems such as DSM-V, to focus on classification that may more closely represent expression of underlying biologic systems.[49]

Pharmacogenetics

One of the main difficulties in clinical practice is the inability to know *a priori* which psychotropic drug will be best suited for each case. Therefore, several groups worldwide try to overcome that obstacle by searching genetic markers that might be predictive of treatment response.

BPAD

The first studies on the relationship between response to lithium and family history have been published in the 1970s, supporting an association between a family history of BPAD and satisfactory response to treatment. Mendlewicz *et al.*[50] first reported a study of 36 patients through a double blind study of lithium prophylaxis. They found that 66 per cent of the responders to lithium had at least one first-degree relative with BPAD, and that only 2–1 per cent of the lithium nonresponders had a first-degree relative with BPAD. Lipp *et al.*[51] first reported an association between DRD2 and non-response to lithium. Several studies have followed (See[52] for review). Positive results were found in TPH and 5HTT.[53,54]

UPAD

Previous studies of an association between poor antidepressant response in depressive patients and 5HTT were largely replicated and could be considered as a robust finding (see[55] for exhaustive review and discussion). Among more recent findings, GNB3 (beta3 subunit of G protein) and DAT1 were found to be associated with treatment response. Binder *et al.*[56] found an association between FKBP5, which plays an important role in the glucocorticoid receptor function, and rapid response to antidepressant. Interestingly, this group of patients was characterized by a high rate of relapses. Finally, the large STAR*D (Sequenced Treatment Alternatives to Relieve Depression) project provided recent promising data for HTR2A 5HTT and GRIK4, which codes for the kainic acid-type glutamate receptor KA1 and treatment response.[57–59]

Psychosocial factors in affective disorders

Impairment in social relationships, dysfunctional cognition, gender, economic status, and temperament has been suggested as involved in the emergence of mood disorders. However, empirical studies on psychosocial factors of patients with affective disorders examine psychosocial features assessed after recovery from or/and at the time of episodes of affective disorders. These retrospective studies might not be able to distinguish between premorbid psychosocial patterns and those which result from previous episodes of illness. Further, longitudinal studies focusing on the role of psychosociological factors have involved predictions of recurrence or exacerbation of symptomatology in previously affected people, but not regarding the onset of the diseases. Thus, the demonstration of temporal antecedence to the initial onset of affective disorder is extremely difficult.[60] Thus, the conclusions in terms of etiological psychosocial factor are limited.

Impairment in social and familial relationships

Difficulties in social functioning are concomitant to depressive disorders.[61] The concept of social support has been widely used to predict general health and more specifically psychiatric symptoms.[62] Previous research revealed that the degree of integration in a social network, or structural support, have a direct positive effect on well-being, reducing negative outcomes in both high and low stress life events. Among depressed individuals, dysfunction in social activities has been found to persist long time after remission from the depressive episode.[63] The social dysfunctioning concerns more specifically marital relationships, parental, and familial relationships.

The relationship between marital disturbance and affective disorders has received increased attention over the past decades. First, descriptive studies have suggested that marital conflict correlates highly with concomitant depression,[64] and marital therapy has been found to be effective in reducing the symptoms of depression, alone as well as in combination with pharmacotherapy. Further, previous research found dysfunctional patterns of communication in couples with a depressed spouse.[65, 66] The lack of a confiding and intimate relationship leaves individuals vulnerable to depression.[67, 68] Finally, marital distress may also exacerbate difficulties experienced in extramarital relationships,[69] thereby increasing introverted behavior and social isolation. In similar manner, the absence of a marital partner may hasten the onset of depression among vulnerable individuals.[70]

The parental relationships seem also to have a great impact in the course of affective disorders. A variety of authors have emphasized the importance of the quality of early experiences with parents in the development of adult depression. Beck first, explicitly attributes the development of negative cognition and negative schemata of self to critical, disapproving parents.[71]

Dysfunctional cognition

According to the helplessness model of depression[73] vulnerability to depression derives from a habitual style of explaining the causes of life events, known as attributional style. A large body of research found that individuals suffering of depression think more negatively than healthy individuals. Specifically, depressed patients have a tendency to make internal, stable, and global causal attributions for negative events, and to a lesser extent, the attribution of positive outcomes to external, specific, and unstable causes. In other words, depressed patients have a low self-esteem.[74,75] Thus, when thinking about the self, past, current and future circumstances, depressed patients emphasize the negative, and this process is likely to contribute to the perpetuation of their depressed mood.

Gender

Evidence for sex differences in responses to depression comes from a large number of studies. Women are consistently reported to have greater prevalence of affective disorders than men.[76,77] First, women may experience two important periods, known to be associated with higher rates of depression: pregnancy and post-partum. The prevalence of major or minor depression among pregnant women ranges from 7 per cent to 26 per cent.[78] Depression during pregnancy is a strong predictor of postpartum depression. The prevalence of postpartum depression ranges from 10 per cent to 15 per cent in the first year after childbirth.[79] Besides these two specific conditions, the reasons for this sex difference are unclear, and are as likely to be social as biological.

Divergences in the number and quality of social and occupational roles have been proposed to explain the greater prevalence of affective disorders among women. In the context of marital relationship, previous research has indicated that for men, marriage confers a protection against illness, while it appears to be associated with higher rates of depression for women.[80] There has been some evidence that within the marriage the traditional female role is limiting, restricting, which may lead to depression.[81,82] For example, the role of child caretaker has consistently been shown to be associated with both high levels of stress and a higher incidence of depression for women.[83]Women are found to have more depressive symptoms when there are young children in the home, and this tends to increase in an almost linear fashion according to the number of children in the household.[84] Further, since women who are employed outside the home also tend to be responsible for household chores,[85] the notion that differentiation in occupational roles could partially explain the prevalence of depression for women is supported.

Socio-economic status

Many studies have reported that low socio-economic status is associated with high prevalence of mood disorders.[86] Since a long time, in social psychiatry, the 'social causation' and 'social selection' hypotheses have been formulated to explain the role of the low socio-economic status in the disease. The causation hypothesis suggests that the stress associated with low social position, that is exposure to adversity and lack of resource to cope with difficulty, may contribute to the development of the affective disorder[87] while the social selection hypothesis argued that genetically predisposed persons drift down to or fail to rise out of such positions.[88,89] Thus, the social selection hypothesis emphasizes the genetic interpretation of cause, while social causation hypothesis focuses on the etiologic role of the environment. Few longitudinal data sets are available to test the causal hypothesis. Nevertheless, there is evidence that disadvantaged socio-economic status, poverty, or education and occupation can be considered as risk factors for mood disorders.[90,91] Nevertheless, Bruce and Hoff found that the effect of poverty is substantially reduced when controlling for degree of isolation from friends and family, suggesting that social isolation mediates some of the relationships between economic status and mood disorders.[92]

In summary, a positive relationship has been found between socio-economic status and vulnerability to affective disorders, with higher rates of vulnerability found among individuals with lower educational and social achievement levels.

Temperament and behaviour

Temperament has been defined in terms of differences in the adaptative systems, that is differences in reactivity and self-regulation to the social context.[93–95]

The model of temperament developed by Eysenck approaches temperament in terms of cortical arousal.[96] Eysenck suggested that individuals differ in their basic arousability and therefore, in their optimal level of stimulation. These physiological differences give rise to the primary personality dimension of introversion-extraversion. Introverts are said to possess relatively reactive reticular systems, and thus to attain their optimal level of cortical arousal at relatively low level of stimulation. As a result of their low optimal arousal level, introverts are expected to prefer and seek out mild forms of stimulation and to avoid more intense and novel forms of stimulation. In contrast, extraverts are said to possess relatively unreactive reticular systems, to have correspondingly high optimal levels of cortical arousal, and to therefore, approach more intense and novel forms of stimulation.

The differences between the Cloninger's model and other models are that Cloninger assumes relationships between biogenic amine neurotransmitters (norepinephrine, serotonin, and dopamine) and personality dimensions. Specifically, Cloninger defined temperament dimensions in terms of individuals' differences in associative learning in response to novelty, danger or punishment, and reward. Further, he hypothesized a positive correlation between serotoninergic activity and harm avoidance, dopaminergic activity and novelty seeking, and finally between noradrenergic activity and reward dependence. According to this author, these aspects of personality denote traits that are usually considered temperament factors because they are heritable, manifest early in life, and apparently involved in learning. The possible tridimensional combinations of extreme (high or low) variants on these basic stimulus response characteristics correspond closely to the traditional descriptions of personality disorders. The specific relationship between temperament and mood disorder is not yet understood satisfactorily. Studies have been done regarding the Tridimensional Personality Questionaire (TPQ) scores in relation to mood disorder, the data available suggest that depressed patients have elevated harm avoidance scores.[97–100]

The possible role of candidate genes has been investigated in personality. Association between a personality trait (Novelty Seeking) and the 7 repeat allele in the locus for Dopamine receptor D4 gene (DRD4) has been observed in a group of 124 unrelated Israeli normal subjects.[101] Novelty Seeking was assessed from the Tridimensional Personality Questionnaire (TPQ).[102] An association was also observed between similar personality traits and long alleles of DRD4 gene in 315 subjects, mostly male siblings from United States.[103] This last study utilized the NEO Personality Inventory (NEO-PI-R)[104] from which TPQ Novelty-Seeking scores can be estimated. More recent studies also suggest a pattern of influence on temperamental dimension exerted by serotonergic and dopaminergic genes.[105] They studies, even not definitive, suggest that the contribution of these polymorphisms to the clinical presentation of mood disorders could be mediated by an influence on personality differences.

The gene-environment hypothesis

The availability of molecular genetic findings in affective disorders offers new directions in this research field. It is now possible to consider the gene-environment hypotheses using the DNA as the genetic liability variable. In primate models, early experiences of maternal separation were found to confer increased risk of depression during adult age.[106] Barr *et al.* found that infant rhesus monkeys showing a specific polymorphism in 5HTT were more likely to engage in rough play than were individuals without this polymorphism.[107] In humans, a landmark study demonstrated that a functional polymorphism in the promoter region of 5HTT gene moderated the influence of stressful life events on depression and suicidal behaviour.[108] This study was replicated

with conflicting results using different methodologies.[109–112] However, these results support the notion that a combination of genetic predisposition and specific life events may interact to facilitate the development of affective disorders.

Conclusion

The complexity and heterogeneity of affective disorders is a major limitation for gene-environment studies.This could be attributed to their non-Mendelian mode of inheritance. BPAD and UPAD are, in fact, phenotypes which do not appear to exhibit classic Mendelian recessive or dominant inheritance involving a single major locus. The presence of both environmental as well as genetic factors and phenotypic diversity also represent important problems when dealing with these diseases. After the era of enthusiasm due to first results from linkage and association studies, the lack of replication and the identification of potential methodological biases led to a period of pessimism. However, recent technological advances allow for the analysis of hundreds of components in a biological system simultaneously. Gene expression micro-arrays may analyse the expression of hundreds of genes in a specific tissue. More specifically, micro-array technologies measure levels of messenger ribonucleic acid (mRNA), an intermediate product between gene and protein.[113] The levels of these 'transcripts' are compared between a population suffering from a specific disease and control subjects. The analysis of mRNA transcripts, instead of DNA regions, may provide additional information on genetic regulation processes of illness. Ultimately, the understanding of neurobiological processes underlying affective disorders may help developing therapeutic and prevention strategies.

Further information

Serretti, A. and Mandelli, L. (2008). The genetics of bipolar disorder: genome 'hot regions', genes, new potential candidates and future directions. *Mol Psychiatry*, **13**, 742–71

The Wellcome Trust Case Control Consortium http://www.wtccc.org.uk

References

1. McInnis, M.G., McMahon, F.J., Chase, G.A. *et al.* (1993). Anticipation in bipolar affective disorder. *American Journal of Human Genetics*, **53**, 385–90.
2. Nylander, P.-O, Engstrom, C., Chotai, J., *et al.* (1994). Anticipation in Swedish Families with Bipolar Affective Disorder. *Journal of Medical Genetics*, **9**, 686–89.
3. Engström, C., Johansson, E.L., Langström, M., *et al.* (1995). Anticipation in unipolar affective disorder. *Journal of Affective Disorder*, **35**, 31–40.
4. Mendlewicz, J. (1994) The search for a manic depressive gene: from classical to molecular genetics. *Progress in Brain Research,* **100**, 225–59.
5. Vieland, V.J., Susser, E. and Weissman, M.M. (1995) Genetic epidemiology in psychiatric research. In *Genetics of Mental Disorders Part I: Theoretical aspects. Baillière's Clinical Psychiatry, International Practice and Research* (eds. G.N. Papadimitriou and J. Mendlezicz), pp. 19–46. Vol. 1, No 1, Bailliere Tindall, London.
6. Cadoret, R.J., Winokur, G., Lengbehn, D., *et al.* (1996) Depression spectrum disease, I: the role of gene-environment interaction. *American Journal of Psychiatry*, **153,** 892–9.
7. Mendlewicz, J., Rainer, J.D. (1977). Adoption study supporting genetic transmission in manic--depressive illness. *Nature,* **268**(5618), 327–9.
8. Kendler, K.S., Eaves, L.J., Walters, E.E. *et al.* (1996) The identification and validation of distinct depressive syndromes in a population-based sample of female twins. *Archives General of Psychiatry,* **53,** 391–9.
9. McGuffin, P., Rijsdijk, F., Andrew, M., (2003). The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Archives General of Psychiatry*, **60**(5), 497–502.
10. Kieseppä, T., Partonen, T., Haukka, J., *et al.* (2004). High concordance of bipolar I disorder in a nationwide sample of twins. *American Journal of Psychiatry*, **161**(10), 1814–21.
11. Weiss, K.M. (1993) Genetic variation and human disease: principles and evolutionary approaches. Cambridge University Press, Cambridge, pp.117-148.
12. Ott, J. (1991) Analysis of human genetic linkage (2nd edn.). Johns Hopkins University Press, Baltimore.
13. Propping, P., Nothen, M.M., Fimmers, R. *et al.* (1993). Linkage versus association studies in complex diseases. *Pychiatric Genetics,* **3,** 136–9.
14. Maier, W., Zobel, A., Rietschel, M,. (2003). Genetics of schizophrenia and affective disorders. *Pharmacopsychiatry*, **36** (Suppl 3), S195–202.
15. Hodge, S.E. (1993) Linkage analysis versus association analysis: distinguishing between two models that explain disease-marker associations. *American Journal of Medical Genetics,* **53,** 367–84.
16. Hodge, S.E. (1994) What association analysis can and cannot tell us about the genetics of complex disease. *American Journal of Medical Genetics,* **54,** 318–23.
17. Oswald, P., Souery, D., Mendlewicz, J. (2003). Molecular genetics of affective disorders. *International Journal of Neuropsychopharmacol,* **6**(2), 155–69.
18. Mendlewicz, J., Simon, P., Sevy, S., *et al.* (1987). Polymorphic DNA marker on chromosome and manicdepression. *Lancet*, **1**, 1230–2.
19. Suarez, B.K. and Hampe, C.L. (1994). Linkage and association. *American Journal of Human Genetics*, **54**(3), 554–9
20. Green, E.K., Raybould, R., Macgregor, S., *et al.* (2005). Operation of the schizophrenia susceptibility gene, neuregulin 1, across traditional diagnostic boundaries to increase risk for bipolar disorder. *Archives General of Psychiatry*, **62**(6), 642–8.
21. McQueen, M.B., Devlin, B., Faraone, S.V., *et al.* (2005). Combined analysis from eleven linkage studies of bipolar disorder provides strong evidence of susceptibility loci on chromosomes 6q and 8q. *American Journal of Human Genetics*, **77**(4), 582–95
22. Schumacher, J., Kaneva, R., Jamra, R.A., *et al.* (2005). Genomewide scan and fi ne-mapping linkage studies in four European samples with bipolar affective disorder suggest a new susceptibility locus on chromosome 1p35–p36 and provides further evidence of loci on chromosome 4q31 and 6q24. *American Journal of Human Genetics*, **77**(6), 1102–11.
23. Bellivier, F., Leboyer, M., Courtet, P., *et al.* (1998) Association between the tryptophan hydroxylase gene and manic-depressive illness. *Archives of General Psychiatry*, **55**, 33–7.
24. Mann, J.J., Malone, K.M., Nielsen, D.A., *et al.* (1997) Possible association of a polymorphism of the tryptophan hydroxylase gene with suicidal behavior in depressed patients. *American Journal of Psychiatry*, **154,** 1451–3.
25. De Luca, V., Mueller, D.J., Tharmalingam, S., *et al.* (2004). Analysis of the novel TPH2 gene in bipolar disorder and suicidality. *Molecular Psychiatry*, **9**(10), 896–7.
26. Murer, M.G., Yan, Q., Raisman-Vozari R Brain-derived neurotrophic factor in the control human brain, and in Alzheimer's disease and Parkinson's disease. *Progress in Neurobiology*, **63**(1), 71–124.
27. Oswald, P., Del-Favero, J., Massat, I., *et al.* (2004). Non-replication of the brain-derived neurotrophic factor (BDNF) association in bipolar affective disorder: a Belgian patient-control study. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics,* **129**(1), 34–5.
28. Schumacher, J., Jamra, R.A., Becker, T., *et al.* (2005). Evidence for a relationship between genetic variants at the brain-derived neurotrophic

factor (BDNF) locus and major depression. *Biological Psychiatry*, **58**(4), 307–14.
29. Nievergelt, C.M., Kripke, D.F., Barrett, T.B., (2006). Suggestive evidence for association of the circadian genes PERIOD3 and ARNTL with bipolar disorder. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics,* **141**(3), 234–41.
30. Mansour, H.A., Wood, J., Logue, T., *et al.* (2006). Association study of eight circadian genes with bipolar I disorder, schizoaffective disorder and schizophrenia. *Genes, Brain, and Behavior*, **5**(2), 150–7.
31. Lindblad, K., Nylander, PO., De bruyn, A., *et al.* (1995) Expansion of trinucleotide CAG repeats detected in Bipolar Affective Disorder by the RED-(rapid expansion detection) method. *Neurobiology of disease*, **2,** 55–62.
32. Mendlewicz, J., Lipp, O., Souery, D. *et al.* (1997) Possible maternal genomic imprinting on expended trinucleotide CAG repeats in bipolar affective disorder. *Biological Psychiatry*, **42**, 1115–22.
33. Mendlewicz, J., Souery, D., Del-Favero, J., *et al.* (2004). Expanded RED products and loci containing CAG/ CTG repeats on chromosome 17 (ERDA1) and chromosome 18 (CTG18.1) in trans-generational pairs with bipolar affective disorder. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics,* **128**(1), 71–5.
34. Tsutsumi, T., Holmes, S.E., McInnis, M.G., *et al.* (2004). Novel CAG/ CTG repeat expansion mutations do not contribute to the genetic risk for most cases of bipolar disorder or schizophrenia. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics,* **124**(1), 15–9.
35. Faraone, S.V., Glatt, S.J., Tsuang, M.T. (2003). The genetics of pediatric-onset bipolar disorder. *Biological Psychiatry*, **53**, 970–7.
36. Faraone, S.V., Glatt, S.J., Su, J., *et al.* (2004). Three potential susceptibility loci shown by a genome-wide scan for regions infl uencing the age at onset of mania. *American Journal of Psychiatry*, **161**(4), 625–30.
37. Geller, B., Badner, J.A., Tillman, R., *et al.* (2004). Linkage disequilibrium of the brain-derived neurotrophic factor Val66Met polymorphism in children with a prepubertal and early adolescent bipolar disorder phenotype. *American Journal of Psychiatry*, **161**(9), 1698–700.
38. Massat, I., Lerer, B., Souery, D., *et al.* (2007). HTR2C (cys23ser) polymorphism infl uences early onset in bipolar patients in a large European multicenter association study. *Molecular Psychiatry*, **12**(9), 797–8.
39. Leboyer, M., Bellivier, F., Nosten-Bertrand, M., *et al.* (1998). Psychiatric genetics: search for phenotypes. *Trends in Neurosciences*, **21**, 102–5.
40. Rubinow, D.R. and Post, R.M. (1992). Impaired recognition of affect in facial expression in depressed patients. Biological Psychiatry, **31**, 947–53.
41. Berrettini, W. (2003). Evidence for shared susceptibility in bipolar disorder and schizophrenia. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics*, **123**(1), 59–64.
42. Berrettini, W. (2000). Susceptibility loci for bipolar disorder: overlap with inherited vulnerability to schizophrenia. *Biological Psychiatry*, **47**(3), 245–51.
43. Walss-Bass, C., Escamilla, M.A., Raventos, H., *et al.* (2005). Evidence of genetic overlap of schizophrenia and bipolar disorder: linkage disequilibrium analysis of chromosome 18 in the Costa Rican population. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, **139**(1), 54–60.
44. Schulze, T.G., Buervenich, S., Badner, J.A., *et al.* (2004). Loci on chromosomes 6q and 6p interact to increase susceptibility to bipolar affective disorder in the national institute of mental health genetics initiative pedigrees. *Biological Psychiatry*, **56**(1), 18–23.
45. Hattori, E., Liu, C., Badner, J.A., *et al.* (2003). Polymorphisms at the G72/G30 gene locus, on 13q33, are associated with bipolar disorder in two independent pedigree series. *American Journal of Human Genetics*, **72**(5), 1131–40.
46. Chen, Y.S., Akula, N., Detera-Wadleigh, S.D., *et al.* (2004). Findings in an independent sample support an association between bipolar affective disorder and the G72/G30 locus on chromosome 13q33. *Molecular Psychiatry,* **9**(1), 87–92.
47. Detera-Wadleigh, S.D. and McMahon, F.J. (2006). G72/G30 in schizophrenia and bipolar disorder: review and meta-analysis. *Biological Psychiatry*, **60**(2), 106-14.
48. Thomson, P.A., Wray, N.R., Millar, J.K., *et al.* (2005). Association between the TRAX/ DISC locus and both bipolar disorder and schizophrenia in the Scottish population. *Molecular Psychiatry*, **10**(7), 657–8, 616.
49. Merikangas, K.R. and Risch, N. (2003). Will the genomics revolution revolutionize psychiatry? *American Journal of Psychiatry*, **160**(4), 625–35.
50. Mendlewicz, J., Fieve, R.R. and Stallone, F. (1973). Relationship between the effectiveness of lithium therapy and family history. *American Journal of Psychiatry,* **130**, 1011–13.
51. Lipp, O., Mahieu, B., Souery, D., *et al.* (1997). Molecular genetics of bipolar disorders: Implication for psychotropic drugs. *European Neuropsychopharmacology*, **7**(suppl 2), 112–3.
52. Oswald, P., Souery, D. and Mendlewicz, J. (2006). Pharmacogenetics of bipolar disorders. In *Psychopharmacogenetics* (eds. P. Gorwood and M. Hamon), pp. 75–100, Springer, USA.
53. Serretti, A., Lilli, R., Lorenzi, C., *et al.* (1999). tryptophan hydroxylase gene and response to lithium in mood disorders. *Psychiatry Research*, **33**, 371–7.
54. Serretti, A., Lilli, R., Mandelli, L., *et al.* (2001). Serotonin transporter gene associated with lithium prophylaxis in mood disorders. *Pharmacogenomics Journal*, **1**(1), 71–7.
55. Serretti, A. and Artioli, P. (2004). From molecular biology to pharmacogenetics: a review of the literature on antidepressant treatment and suggestions of possible candidate genes. *Psychopharmacology (Berl)*, **174**(4), 490–503.
56. Binder, E.B., Salyakina, D., Lichtner, P., *et al.* (2004). Polymorphisms in FKBP5 are associated with increased recurrence of depressive episodes and rapid response to antidepressant treatment. Nature Genetics, **36**(12), 1319–25.
57. McMahon, F.J., Buervenich, S., Charney, D., *et al.* (2006). Variation in the gene encoding the serotonin 2A receptor is associated with outcome of antidepressant treatment. *American Journal of Human Genetics*, **78**(5), 804–14.
58. Hu, X.Z., Rush, A.J., Charney, D., *et al.* (2007). Association between a functional serotonin transporter promoter polymorphism and citalopram treatment in adult outpatients with major depression. *Archive General of Psychiatry*, **64**(7), 783–92.
59. Paddock, S., Laje, G., Charney, D., *et al.* (2007). Association of GRIK4 with outcome of antidepressant treatment in the STAR*D cohort. *American Journal of Psychiatry*, **164**(8), 1181–8.
60. Depue, R.A. and Monroe, S.M. (1986). Conceptualization and measurement of human disorder in life stress research: the problem of chronic disturbance. *Psychological Bulletin,* **99**, 36–51.
61. Hirschfeld, R.M.A, Klerman, G.L., Clayton, P.J., *et al.* (1983). Assessing personality: Effects of the depressive state on trait measurement. *American Journal of Psychiatry,* **140**, 695–9.
62. Kendler, K.S. (1997) Social support: a genetic epidemiologic analysis. *American Journal of Psychiatry,* **154**, 1398–404.
63. Shapira, B. *et al.* (1999). Social adjustment and self-esteem in remitted patients with unipolar and bipolar affective disorder: a case-control study. *Comprehensive Psychiatry,* **40**, 24–30.
64. Crowther, J.H. (1985). The relationship between depression and marital maladjustment: A descriptive study. *Journal of nervous and mental disease,* **173**, 227–31.
65. Kahn, J., Coyne, J.C., and Margolin, C. (1985). Depression and marital disagreement: The social construction of despair. *Journal of social and Personal Relationships,* **2**, 447–61.

66. Biglan, A., Hops, H., Sherman, L., *et al*. (1985). Problem-solving interactions of depressed women and their husbands. *Behavior Therapy,* **16**, 431–51.
67. Brown, G.W. and Prudo, R. (1981). Psychiatric disorder in a rural and urban population: 1. Etiology of depression. *Psychological medicine,* **11**, 581–99.
68. Costello, C.G. (1982). Social factors associated with depression: A retrospective community study. *Psychological Medicine,* **12**, 329–39.
69. Coyne, J.C., and DeLongis, A. (1986). Going beyond social support: The role of social relationships in adaptation. *Journal of Consulting and Clinical Psychology,* **54**, 454–60.
70. Brown, G.W., and Harris T. (1978). *Social origin of depression*. London: Free Press.
71. Beck, A.T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. Harper & Row, New York.
72. Peterson, C. and Seligman, M.E.P. (1984). Causal explanations as a risk factor for depression: Theory and evidence. *Psychological Review,* **91**, 347–74.
73. Peterson, C. and Seligman, M.E.P. (1984). Causal explanations as a risk factor for depression: Theory and evidence. *Psychological Review,* **91**, 347–74.
74. Tracy, A., Bauwens, F., Martin, F., *et al.* (1992). Attributional style and depression: a controlled comparison of remitted unipolar and bipolar patients. *British Journal of Clinical Psychology,* **31**, 83–4.
75. Pardoen, D., Bauwens, F., Tracy, A., (1993). Self-esteem in recovered bipolar and unipolar out-patients. *British Journal of Psychiatry,* **163**, 755–62.
76. Bebbington, P.E., *et al*. (1998) The infl uence of age and sex on the prevalence of depressive conditions: report from the National Survey of Psychiatric Morbidity. *Psychological Medicine,* **28**, 9–19.
77. Kroenke, K., and Spitzer, R.L. (1998). Gender differences in the reporting of physical and somatoform symptoms. *Psychosomatic Medicine,* **60**, 150–5.
78. Hobfoll, S.E., Ritter, C., Lavin, J., *et al.* (1995). Depression prevalence and incidence among inner-city pregnant and postpartum women. *Journal of Consulting and Clinical Psychology*, **63**(3), 445–53.
79. Afifi, M. (2007).Gender differences in mental health. *Singapore Medical Journal,* **48**(5), 385–91.
80. Weissman, M.M. (1987). Advances in psychiatric epidemiology: Rates and risks for major depression. *American Journal of Public Health,* **77**, 445–51.
81. Gove, W.R., and Tudor, J.F. (1973). Adult sex roles and mental illness. *American Journal of Sociology,* **78**, 1308–14.
82. Ramsey, E.R. (1974). Boredom: the most prevalent american disease. *Harpers,* **249**, 12–22.
83. Thoits, P.A. (1986). Multiple identities: Examining gender and marital status differences in distress. *American Sociological Review,* **51**, 259–72.
84. Radloff, L.S. (1975). Sex differences in depression: The effects of occupation and marital status. *Sex Roles,* **1**, 249–65.
85. Rosenfield S. (1992). The costs of sharing: wives' employment and husbands' mental health. *Journal of Health and Social Behavior,* **33**, 213–25.
86. Dohrenwend, B.P., *et al*. (1992). Sosioeconomic status and psychiatric disorders: the causation-selection issue. *Science,* **255**, 946–52.
87. Faris, R.E.L., and Dunham, W. (1939). *Mental disorders in urban areas*. University of Chicago Press, Chicago.
88. Jarvis, E. (1971). Insanity and idiosy in Massachusetts: report of the commission of Lunacy, 1855. Harvard University Press, Cambridge.
89. Odegaard, O. (1956). The incidence of psychoses in various occupations. *International Journal of Social Psychiatry*, **2**, 85–104.
90. Bruce, M.L., Takeuchi, D.T., and Leaf, P.J. (1991). Poverty and psychiatric status: Longitudinal evidence from the New Haven Epidemiologic Catchment Area Study. *Archives of General Psychiatry,* **48**, 470–4.
91. Gallo, J.J., Royall, D.R., and Anthony, J.C. (1993). Risk factors for the onset of depression in middle age and later life. *Social Psychiatry Psychiatr Epidemiology,* **28**, 101–8.
92. Bruce, M.L., and Hoff, R.A. (1994). Social and physical health risk factors for fi rst-onset major depressive disorder in a community sample. *Social Psychiatry and Psychiatric Epidemiology,* **29**, 165–71.
93. Cloninger, C.R., Svrakic, D.M., and Przybeck, T.R. (1993). A psychobiological model of temperament and character. *Archives of General Psychiatry,* **50**, 975–90.
94. Derryberry, D., and Rothbart M.K. (1984). Emotion, attention, and temperament. In *Emotion, Cognition, & Behavior* (eds. C.E. Izard, J. Kagan, and R.B. Zajonc), pp 133–66. Cambridge University Press, New York.
95. Derryberry, D, and Rothbart, M.K. (1997). Reactive and effortful processes in the organization of temperament. *Developmental Psychopathology,* **9**, 633–52.
96. Eysenck, H.J. (ed.) (1981). *A model of personality*. Springer- Verlag, New York.
97. Joffe, R.T., Bagby, R.M., Levitt, A.J., *et al*. (1993). The Tridimensional Personality Questionnaire in major depression. *American Journal Psychiatry,* **150**, 959–60.
98. Kleifield, E.I., Sunday, S., Hurt, S., *et al*. (1994). The effects of depression and treatment on the Tridimensional Personality Questionnaire. *Biological Psychiatry,* **36**, 68–70.
99. Nelsen, M.R., and Dunner, D.L. (1995). Clinical and differential diagnostic aspects of treatment-resistant depression. *Journal of Psychiatry Research,* **29**, 43–50.
100. Chien, A.J., and Dunner, D.L. (1996). The Tridimensional Personality Questionnaire in depression: state versus trait issues. *Journal of Psychiatry Research,* **30**, 21–7.
101. Ebstein, R.P., Novick, O., Umansky, R., *et al*. (1996) Dopamine D4 receptor (D4DR) exon III polymorphism associated with the human personality trait of Novelty-Seeking. *Nature Genetics,* **12**, 78–80.
102. Cloninger, C.R. (1987) A systematic method for clinical description and classifi cation of personality variants. *Archives of General Psychiatry,* **44**, 573–88.
103. Benjamin, J., Li, L., Paterson, C., *et al*. (1996). Population and familial association between the D4 dopamine receptor gene and measures of Novelty-Seeking. *Nature Genetics*, **12**, 81–4.
104. Costa, P.T.J and McCrae, R.R. (1992). In Revised NEO Personality Inventory (NEOPI-R) and NEO Five Inventory (NEO-FFI) professional manual. *Psychological Assessment Ressources*, Odessa, FL.
105. Serretti, A., Mandelli, L., Lorenzi, C., *et al*. (2006). Temperament and character in mood disorders: influence of DRD4, SERTPR, TPH and MAO-A polymorphisms. *Neuropsychobiology*, **53**(1), 9–16.
106. Lesch, K.P., Bengel, D., Heils, A. *et al*. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science,* **274**, 1527–31.
107. Barr, C.S., Newman, T.K., Schwandt, M., *et al*. (2004). Sexual dichotomy of an interaction between early adversity and the serotonin transporter gene promoter variant in rhesus macaques. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(33), 12358–63.
108. Caspi, A., Sugden, K., Moffitt, T.E., *et al*. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*, **301**(5631), 386–9.
109. Eley, T.C., Sugden, K., Corsico, A., *et al*. (2004). Gene-environment interaction analysis of serotonin system markers with adolescent depression. *Molecular Psychiatry*, **9**(10), 908–15.
110. Kaufman, J., Yang, B.Z., Douglas-Palumberi, H., *et al.* (2004). Social supports and serotonin transporter gene moderate depression in maltreated children. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(49), 17316–21.

111. Grabe, H.J., Lange, M., Wolff, B., *et al.* (2005). Mental and physical distress is modulated by a polymorphism in the 5-HT transporter gene interacting with social stressors and chronic disease burden. *Molecular Psychiatry*, **10**(2), 220–4.
112. Wilhelm, K., Mitchell, P.B., Niven, H., *et al.* (2006). Life events, first depression onset and the serotonin transporter gene. *British Journal of Psychiatry*, **188,** 210–5.
113. Konradi, C. (2005). Gene expression microarray studies in polygenic psychiatric disorders: applications and data analysis. *Brain Research. Brain Research Reviews*, **50**(1), 142–55.

4.5.6 Neurobiological aetiology of mood disorders

Guy Goodwin

Introduction

Neurobiology provides an explanation of behaviour or experience at the level, either of systems of neurones or individual cells. The current era of progress is driven by contemporary cognitive neuroscience and a rapid evolution in the platform technologies of imaging and genetics. These will allow us to improve our accounts of the functional anatomy of the component elements of mood and its disorder, their functional neurochemistry and, in all probability, give meaning to what a cellular account of depressive illness may eventually describe. This chapter will offer a partial and personal view of these developments to date.

There are now authoritative models of causation in mood disorder, established from well designed, large-scale twin studies (see Chapter 4.5.5). These inform the classical formulation of mood disorder as requiring a vulnerability, a precipitating factor or factors, and maintaining factors which prevent spontaneous recovery. Neurobiology will be addressed under these headings.

Vulnerability to mood disorder

The key vulnerability factors appear to be genes, temperament (also in substantial part genetic), and early adversity. There has been limited work on the neurobiology of these risk factors, as opposed to the vast effort to understand the depressed phenotype. However, for potential prevention either of onset or relapse, such factors appear more logical targets for current research effort and will be covered first. Success in depression would parallel that seen in moving the management of heart disease from the acute episode of infarction to the treatment of metabolic risk factors.

(a) Genetics

Neurobiology has informed the genetic search for candidate genes, starting with the human serotonin transporter (SERT) gene (see Chapter 4.5.5). There has been a terrific proliferation of possible genetic effects deriving from neurobiological theories designed either to explain elements of the actions of psychotropic drugs, the depressed phenotype or from animal experiments. The latter are limited by the validity of animal models of depression per se. Some of the former will be noticed below.

Genes making small contributions to the risk of psychiatric disorder are emerging from direct analysis of the genome (see Chapter 2.4.2). Consistent findings must inform biological investigations in future. At this point it is uncertain whether insights will come from studying variation in individual genes, as has often been assumed, or from a much more complex understanding of cellular function regulated only in part by genetic variation. On the latter assumption the role of genetic hits is to direct attention to processes which may go wrong in the relevant disease. For mood disorder, these seem likely to be developmental or related to stress regulation.

(b) Temperament

The way in which genes may regulate the expression of vulnerability traits is suggested by animal studies. For example, when animals are selected for differences in emotional behaviour they also show different hypothalamic–pituitary–adrenal (HPA) axis function. Specifically, Roman high- and low-avoidance rats differentially acquire a two-way active avoidance response in a shuttle box. High-avoidance animals show greater prolactin and HPA axis responsivity to stress compared with low-avoidance animals. However, young Roman strain rats show identical HPA axis reactivity, although prolactin responses and behaviour are different.[1] In other words, reactivity to the environment may share a measure of common genetic control across physiological and behavioural domains, but HPA abnormality per se develops secondary to emotional experience, or at least is magnified by it.

In human studies, neuroticism is an old psychological construct often criticized as reflecting an average or habitual mood state rather than a truly independent risk. We have studied extremes of the dimension (high and low N) in young subjects before the onset of depression and in older groups who may or may not have experienced depressive episodes. Interestingly, high neuroticism with or without a history of depression is associated with increased awakening cortisol[2] in mature subjects, but not in subjects under 20 years of age, echoing the rodent finding. Thus, N has a purely biological consequence that develops with emotional experience, but is independent of depression per se.

What the neuroticism construct has also lacked hitherto has been a plausible psychological dimension. Cognitive bias relevant to the onset of depression can be detected in young high N subjects. In emotional categorization and memory tasks, high N volunteers were faster to classify dislikeable self-referent personality characteristics and produced fewer positive memory intrusions. They also had a higher threshold for identifying happy faces. This suggests the hypothesis that risk for depression is largely manifest as reduced positive processing of emotional information[3]; increased negative processing appears to develop only after the actual experience of depression. Neural biases underlying this behaviour are even more readily detected.[4] Our hypothesis is that high neuroticism is not just an habitual low mood but is **biologically** founded in negative biases in attention, processing, and memory for emotional material. Indeed, there is now genetic evidence favouring a common genetic locus in human beings and rodent.[5] How emotional bias translates into either low-level symptoms or a full mood episode will be of great interest. Furthermore, depressive episodes per se appear to have an impact on brain function, and increase the risk of further relapse (see below).

(c) Early adverse experience

Adverse childhood experience was identified in genetically uncontrolled studies as a risk factor predisposing women to subsequent

depression (Chapter 4.5.5) and has been confirmed in genetically informative designs.[6] In a clinical context, such developmental or social effects are usually viewed as separable from biology. Indeed, their very existence is usually taken to validate a 'social' approach to psychiatry. From a more unified point of view, however, one would predict measurable neurobiological consequences. In fact, such effects have proved to be more profound than most biologists anticipated.

Variations in maternal care produce individual differences in neuroendocrine responses to stress in rats. The offspring of mothers that exhibited more licking and grooming of pups during the first 10 days after birth showed, in adult life, reduced plasma ACTH and corticosterone responses to acute stress.[7] In addition, there was increased hippocampal glucocorticoid-receptor messenger RNA (**mRNA**) expression, enhanced glucocorticoid feedback sensitivity, and decreased levels of hypothalamic corticotrophin-releasing hormone (**CRH**) mRNA. Greater early maternal attention also substantially reduced subsequent behavioural fearfulness in response to novelty, increased benzodiazepine receptor density in the amygdala and locus coeruleus, increased α_2-adrenoreceptor density in the locus coeruleus, and decreased CRH receptor density in the locus coeruleus. Thus, maternal care serves to programme behavioural responses to stress in the offspring by altering the development of the neural systems that mediate fearfulness.

When BALB/cByJ mice were raised by an attentive C57BL/6ByJ dam, their excessive stress-elicited HPA activity was reduced, as were their behavioural impairments. However, cross-fostering the more resilient C57BL/6ByJ mice to an inattentive BALB/cByJ dam failed to elicit behavioural disturbances. In other words, vulnerable offspring may have their problems exacerbated by maternal behaviour, while early-life manipulations may have less obvious effects in relatively hardy animals.[8] Whether separation or stress paradigms in rodents can be taken as precise models of the mechanisms underlying the risk of mood disorder or other psychiatric problems cannot yet be decided, but their general relevance to the human case seems obvious. At present, data in human subjects is limited but findings that relate to the better characterized animal models are emerging.[9]

In fact, epidemiological data have linked increased risks of cardiovascular, metabolic, neuroendocrine, and psychiatric disorders in adulthood with an adverse *foetal* environment as well. Glucocorticoid excess may be the mechanism.[10] Low-birthweight babies have higher plasma cortisol levels throughout adult life, which suggests a permanent change in HPA function. Whether such effects and later effects of environmental stress in childhood can in part mediate co-morbidity between a range of psychiatric and physical disorders is of growing contemporary interest. It is unclear how, over- or underactivity in stress regulation contributes to psychiatric disorder: both appear to be implicated since awakening cortisol responses may be blunted in subjects with early adversity[9] or enhanced in at risk neurotic individuals.

Gene–environment interaction is the likely basis of the neurobiology of mood disorder. In general terms this must be correct. Either the genetic/biological or the environmental factors could be targets for prevention. Whether the genetic mechanisms can be brought into sufficient focus to allow specific new pathways to be identified remains the major current challenge. It is often assumed that mediating characteristics or the endophenotype may have a simpler genetic architecture than the disease itself: unfortunately, the evidence so far gives reason for caution. This debate is currently very polarized between optimists (see Chapter 2.5.3 by Meyer-Lindenberg & Goldberg) and pessimists (see Chapter 2.4.2 by Flint). The genetic and developmental routes into distal common pathways regulating stress responses may be very numerous. Disorders that are both common and very variable in expression, such as depression, may turn out to have little specificity that is worth talking about. Every illness may be an ensemble of many specific factors, none of which is individually going to lead to a more focused treatment or a better prediction of treatment response.

Precipitating factors: the neurobiology of life events

Like early adversity, the role of life events in depression has been affirmed in genetically controlled studies. Life events are relevant to almost all first episodes of depression, but are less significant in its recurrence. The biology of life events is subsumed in the biology of stress, at best a clumsy term. In human studies it will be always difficult to isolate the critical ingredients of a particular psychological stress from the individual differences that stressed individuals bring to their experience. There have been few recent contributions to the field of direct relevance to depression.[11] However, a key clinical feature of the illness course in depression is the association of life events most strongly with first episodes of depression. Subsequent episodes appear to need a less substantial environmental trigger, as if the patient becomes sensitized.[12] Patients with a strong family history may effectively be presensitized. Accordingly the effect of life events and the brain changes that occur with repeated or chronic illness is of great relevance to prevention and reduction of the risk of future episodes.

Maintaining factors: biological studies of the depressed state

In the majority of biological studies of affective disorder, patients have been studied when ill and compared with normal controls. Over the years, this kind of design has produced a range of positive findings, usually of modest effect. It remains true to say that no biological changes have ever been found that distinguish between depressed patients and controls better than does the clinical assessment of the patients. What is also curious, and not a little tantalizing, is the impression that some symptoms may, in part, represent biological adaptations directed to put things right. Thus, on the one hand, there may be consistent effects upon hormone secretion or sleep that represent phenomena of illness. On the other, deliberate changes in hormone status or sleep deprivation may modify the state of depression. Depression is also so common in its less severe forms, that it is tempting to see it as a biologically adaptive mechanism in response to loss or social defeat. Informative animal analogues might be expected to exist, but theoretical comparisons with other biological models such as early separation in primates or hibernation in bears are limited by the species gap.[13]

However, what makes depression the clinical burden it is, remains its tendency to persist and sometimes become chronic. The biological factors contributing to this are still poorly understood, but they would provide an obvious target for novel drug development. In general it is not yet obvious which symptoms of acute depression are related to this key biology and which are either irrelevant or even adaptive. If there is now a consistent interest, it has been

stimulated by the gradual acceptance that some cells divide to produce neurones in the mature brain, especially in the hippocampus. It is very tempting to suppose that the plastic effects maintaining the unwanted brain state in depression may be related to neurogenesis or its failure, which is a beautiful hypothesis requiring confirmation by direct evidence.

(a) The depressed state: functional anatomy

Perfusion or metabolic imaging can indirectly detect changes in neuronal activity (see Chapters 2.3.6 and 2.3.8). Signals can be well localized, but their meaning is ambiguous. They may reflect either reversible changes in function or a semi-permanent loss of neuronal connectivity. Reductions in function in anterior brain structures have been typical in major depression. Hypoperfusion tends to be greatest in frontal, temporal, and parietal areas and most extensive in older patients; high Hamilton scores tend to be associated with reduced perfusion.[14] Reductions in frontal areas may be more likely in patients with impoverished mental states. Thus, neuropsychological testing in major depression shows evidence of slowing in motor and cognitive domains, with additional prominent effects on mnemonic function that are most marked in the elderly. These effects are correlated with reduced frontal perfusion in the elderly. In younger patients, there may actually be increased perfusion in the frontal and cingulate cortex. Metabolic increases in the cingulate gyrus have been associated with a good treatment response.[15] Highly localizing findings have been unusual, however. The only exceptions have been within-subject changes on recovery in the mesial frontal cortex and perhaps the basal ganglia.[14]

There has been a dramatic expansion of imaging studies of emotional processing in normal volunteers, now usually with fMRI (see Chapters 2.3.8 and 2.5.4). It is well summarized by meta-analysis of over 300 such emotion induction and cognitive task. Emotion induction resulted in inferior medial activation and cognitive tasks resulted in dorsolateral activation.[16] However, the broad spread of precise loci of activation means that localization within the frontal lobes has proceeded little further. It may explain the diffuse reports typical of the depression literature. Nevertheless, a focus on limbic activity has led to quite specific, quasi-neurological hypotheses about connectivity in frontal areas and to treatment innovation: deep brain stimulation adjacent to subgenual cingulated cortex (Brodmann area 25).[17] How effective, and how localized this treatment effect really is, will be an important challenge to the field. However, it underlines that 'functional imaging' of brain perfusion primarily informs anatomy.

Isotope-based imaging of receptor occupation could more plausibly offer mechanistic understanding of psychiatric disorder. In depression, it has progressed with the availability of suitably informative ligands. However, the field generally tends to employ small sample sizes, and fundamental advances are difficult to identify. Single-photon emission tomography (**SPET or SPECT**) with the dopamine $D_{2/3}$ ligand [123I]IBZM showed increased binding in the striatum.[18] There were significant correlations between IBZM binding in the left and right striatum and measures of reaction time and verbal fluency, but not of mood as such. This finding has been confirmed with a PET ligand.[19] Increased $D_{2/3}$ binding in the striatum probably reflects a reduced dopamine function, whether due to a reduced release or secondary upregulation of receptors. Binding to the 5-HT1a receptor appears to be reduced in unipolar depression, an effect also present in recovered atients.[20]

In recent years, new SPET and PET ligands for the serotonin and dopamine transporter have become available (see Chapter 2.3.6 by Grasby). For the serotonin transporter in acute depression, the story is not consistent.[21,22] Binding to the dopamine transporter appears to correlate with depressive symptoms in healthy volunteers.[23] Hence trait effects may confound state effects and vice versa. Isotope-based imaging has been slow to develop a wide choice or availability of ligands, hence its role has been largely to follow rather than stimulate new ideas. Its specificity does mean that it can critically test hypotheses about specific receptors.

Such ligands have not yet made an impact on treatment strategies, as dopamine receptor ligands have for the antipsychotics. However, there are interesting preliminary conclusions: for example, drugs that bind to the serotonin transporter appear to saturate the site at therapeutic doses and increase the availability of dopamine reuptake sites.[24]

In summary, functional imaging has served to implicate frontal and limbic rather than posterior brain areas, in broad confirmation of anatomical conclusions derived from observing the effects of lesions or brain stimulation. Relevant neuropsychological challenges are now being incorporated into imaging protocols and we have the first example of an imaging-led treatment innovation—deep brain stimulation. Finally, 'functional' abnormalities may importantly predict structural abnormality in depression.

(b) Neuroendocrine challenge tests

Secretion of hormones in the anterior pituitary is under control, both direct and indirect, of central neuronal cell bodies that may project relatively widely within the brain. The secretion of a given hormone in response to specific precursors or agonists for individual neurotransmitter receptors has been proposed as a way of testing the security of such connections. Hormone secretion provides a bioassay of the system of interest. There is a measure of consensus about the findings in major depression, which, indeed, forms the most consistent basis for our understanding of disturbed neurotransmission in depression. However, the approach no longer leads the neurobiology, and merits consideration instead, in the more specific context of neuroendocrine function (see Chapter 2.3.3). The main findings are described below.

Neuroendocrine drug challenge suggests attenuated serotonergic function and increased cholinergic function in depression. Reduced responses to adrenergic and dopaminergic challenge also suggest impaired neurotransmission. Interpretation of tests with agonists is always difficult, because blunting may occur in an overactive system that has been downregulated. In addition, if the secretion of the assay hormone itself is actually directly affected by the state of depression, interpretation in terms of specific neurotransmitter abnormalities may be misleading. This is a particular problem for ACTH/cortisol responses (see below). In fact, enthusiasm for neuroendocrine surrogate markers of monoamine transmission within the brain has probably diminished in recent years, but the paradigm of drug challenge nevertheless remains interesting. We must assay brain responses of the monoamine projections more centrally involved in mood regulation.

(c) Hypercortisolaemia

About half of all patients with major depression have a raised cortisol output, which tends to return to normal on recovery. It is most consistently associated with an 'endogenous' pattern of

illness (see Chapter 4.5.3). While cortisol is always regarded as a 'stress' hormone, and is secreted in response to various types of acute stress, the stresses that commonly result in long-term hypercortisolaemia are poorly understood. The idea that there is a relatively specific link between chronic high cortisol levels and mood disorder is notably persistent. In major depression there is peripheral hypertrophy of the adrenal glands, measurable in MRI body scans, and an enhanced response to corticotrophin. The MRI change, like the hypercortisolaemia itself, reverses on recovery.[25]

Suppression of cortisol secretion occurs normally via glucocorticoid receptor-mediated inhibitory feedback to the hypothalamus; it is readily produced by dexamethasone, which is a potent exogenous glucocorticoid (the dexamethasone suppression test (**DST**)). For example, Non-suppression of endogenous cortisol after dexamethasone occurs in Cushing's disease. It implies either reduced feedback and/or enhanced central drive to release cortisol. It was initially observed that the 1-mg DST showed high specificity (96 per cent) and sensitivity (67 per cent) as a putative diagnostic test for melancholia.[26] At this point of time the result attracted intense interest, but has since proved difficult to generalize. The high specificity established against normal controls was less against other patient groups. Thus, DST non-suppression has not been accepted as a diagnostic test. This failed effort to give medical respectability to psychiatric diagnosis came to devalue what remains an important observation. Non-suppression usually reflects hypercortisolaemia, which is itself a robust phenomenon of mood disorder that requires explanation like any other core biological symptom. Other symptoms that we identify as part of the depressive syndrome lend themselves less easily to investigation. The DST also has potential clinical uses beyond diagnosis. DST non-suppression predicts a low placebo response rate to drug treatment,[27] and hypercortisolaemia predicts a low rate of clinical response to psychological intervention.[28]

It remains unclear whether cortisol contributes to the clinical state of depression by a direct action on the brain. Exogenous cortisol administration is associated with affective symptoms, and chronic excessive cortisol secretion commonly appears to produce depressive symptoms in Cushing's disease. An HPA axis programmed to hypersecrete cortisol under stress could be a pathogenic mechanism explaining why depression or mania develops. This view has provoked efforts to treat mood disorder by inhibition of cortisol synthesis with metyrapone or blocking the post-synaptic receptors. The effects of such manipulations appear primarily, and unexpectedly, to influence cognitive function more than mood per se. Thus the anti-glucocorticoid, mifepristone improved spatial working memory in bipolar depression[29] and the anti-mineralocorticoid, spironolactone significantly impaired selective attention and delayed recall of visuospatial memory in healthy volunteers without effects on CCK-induced panic anxiety.[30]

There is a final twist: when depressed patients are given large doses of cortisol they tend to show acute mood enhancement[31] and oral dexamethasone has been reported to elevate mood in major depression, especially in hypersecretors.[32] This leads to the converse hypothesis that an HPA axis appropriately adapted to chronic stress early in development might be unable to mount a normal effective response to acute stress later in life. Cortisol may then be seen as a euphoriant (or antidepressant), and hypercortisolaemia as an antidepressant response of the stress-regulating mechanisms of the brain. Based on this view, all cortisol levels seen in depression may be set inappropriately low for the ongoing stress, however high or low they are compared with the normal range.

Whether one supposes cortisol levels to be set too high or too low in depression, it remains inconvenient that either a suppression or an augmentation of steroid effect seems, initially at least, to elevate mood. A way out of this complication may lie in cortisol's action on two receptors in the brain (the glucocorticoid and mineralocorticoid receptors) that may have opposite actions. However, we still need better-controlled replicated data on the effects of steroid manipulations, both in at-risk subjects and in major depression. It is also possible that peripheral cortisol levels are largely irrelevant to the brain and that receptor regulation may critically modulate their central action: one challenging hypothesis is that antidepressants work through changing receptor disposition.[33]

An increased cortisol production is associated with an increased release of hypothalamic β-endorphin and probably a pulsatile increase in ACTH. The paraventricular nucleus of the hypothalamus represents the highest level of dedicated neurones in the HPA axis. The neurosecretory cells of the paraventricular nucleus release the peptides CRH and AVP into the portal hypophyseal blood. These hormones in turn stimulate the release of ACTH from the anterior pituitary. Major depression is characterized by a blunted ACTH response to CRH, an elevated level of CRH in the cerebrospinal fluid, and increased numbers of neurones expressing CRH mRNA in the paraventricular nucleus of the hypothalamus postmortem.[34] CRH is not confined to the paraventricular nucleus, but is expressed in a variety of other central nuclei whence it can produce anxiogenic behavioural effects. CRH receptors, which exist in two forms, are widely distributed in the hypothalamus and cortex. A related peptide, urocortin, has a similar pharmacology. Knocking out the CRH-1 receptor gene in mice impaired the HPA stress response and reduced anxiety-like behaviour. Non-peptide antagonists of CRH action, and of other peptide hormones implicated in stress responses have been taken very seriously as putative anxiolytics or antidepressants.[35] If effective, they will be among the first of a new generation of truly novel treatments based on peptide neurotransmission. The failure to see new compounds of this general kind by now is disappointing, and in the case of a neurokinin antagonist, aprepitant, there has been a high profile failure in major depression.[36]

(d) Thyroid abnormalities

In unselected major depression, thyroid hormone levels are usually normal, but there may be abnormalities of the thyrotropin (thyroid-stimulating hormone) response to thyrotropin-releasing hormone. The thyrotropin response is blunted in a significant number of patients, but this effect is poorly understood and has few accepted clinical associations. In contrast, a subgroup of patients may show an enhanced thyrotropin response with normal thyroid hormone levels (referred to as grade II hypothyroidism). These associations and the use of thyroid hormones in treatment suggest that there is more to be learned in this area (see Chapter 4.5.8).

(e) Sleep disturbance

Sleep is often disturbed in depression but in a variety of ways. Early-morning waking is the most typical in endogenous or melancholic depression, with the sleep patterns in such patients being similar to those seen in patients with mania. Trouble getting to sleep, frequent awakenings, and unsatisfactorily prolonged sleep

are also commonly seen in depression. Like other biological manifestations of the disorder, the extent to which sleep is simply a consequence of the state of depression or a contribution to its biology is uncertain. Patients with severe depression or mania may respond to sleep deprivation with a transient elevation in mood. It implies that the sleep–wake cycle is directly involved with mood regulation and its disorder.

In severe depression (melancholia) the typical effects are a reduction in the total length of slow-wave sleep and a shortened latency in the appearance of rapid eye movement (**REM**) or dreaming sleep.[37] The cholinergic projections from the hindbrain may be REM-ON cells, while serotonergic and noradrenergic cells may be REM-OFF cells. The disturbed sleep of depression could be due to an increased cholinergic and/or a decreased serotonergic/noradrenergic drive; simplistic though it sounds, the experimental evidence is supportive. Depressed patients challenged with a cholinergic agonist in the second non-REM period enter REM significantly faster than psychiatric and normal control subjects. The reduced sensitivity of the noradrenergic system is suggested because clonidine fails to suppress REM in depressed patients compared with controls.[38] Tryptophan depletion (to attenuate 5-HT function) partially mimics the changes seen in depression in recovered patients.[39]

Sleep tends to recover on recovery from depression, and the tricyclic antidepressants in particular suppress REM sleep. However, sleep disturbance may be an early predictor of relapse, and disturbed sleep parameters predict a poor response to cognitive behaviour therapy.[40] Indeed, depressed patients may have inherently weak slow-wave sleep processes because unaffected subjects with a family history of depression show reduced slow-wave sleep and increased REM density in the first sleep cycle[41].

Interest in sleep as a fundamental key to understanding mood disorder has waned in the last two decades. However, its neurobiology is increasingly well understood, and its time may come again.

(f) Monoamine metabolite turnover

The earliest studies to investigate the actions of tricyclic antidepressants highlighted their actions on the turnover of the monoamine metabolites in animal brain. The 'monoamine theory of depression' proposed the reduced functioning of monoamine transmission in depression. Therefore it was natural to seek relevant measures of monoamine chemistry in the cerebrospinal fluid of patients and controls. The study of what became irreverently known as 'neural urine' and indeed of urine itself, since peripheral measures of monoamine turnover are also potentially relevant, virtually defined a decade of biological psychiatry in the 1970s and 1980s. Drugs had similar effects on neurotransmitter turnover as seen in animal studies, demonstrating that the human techniques were sufficiently sensitive. Indeed the monoamine theory is, at its best, a theory about drug action because the monoamine and metabolite changes produced by illness in patients have proved remarkably unconvincing.[42] The findings for the noradrenaline metabolite MHPG and the 5-HT metabolite 5-hydroxyindoleacetic acid were negative. The dopamine metabolite homovanillic acid did show the predicted decrease, but only significantly in women. There were trends to modest increases in all the major metabolites in mania. Although disappointing, cerebrospinal fluid studies could never reflect the activities of smaller groups of neurones localized in areas critical for the modulation of mood. Such a focus is only possible in isotope imaging (PET or SPET) or better post-mortem studies of the brain.

(g) Tryptophan depletion

The most convincing evidence that 5-HT is intimately involved in mood disorder has come from depletion of tryptophan, the amino acid precursor of 5-HT. The level of tryptophan in both peripheral blood and the brain can be driven to very low levels by a short-term low-protein diet and subsequent loading with large neutral amino acids. These compete with tryptophan for access to the brain amino acid transporter and also increase its peripheral metabolism, which results in the reduced synthesis and release of 5-HT. Initial observations appeared to bear primarily on the mechanism of drug action. Thus, patients who had recovered from major depression while taking a serotonin-selective reuptake inhibitor experienced a clear-cut return of severe symptoms lasting for several hours after tryptophan depletion. This finding has now been critically extended to patients with a history of recurrent major depression who were euthymic but not taking any medication.[43] Prominent objective symptoms of retardation and cognitive distortion returned in a stereotyped and severe way, reflecting previous symptoms. The effects on mood in patients who have had a previous episode of depression are qualitatively different from the more minor changes seen in normal female controls or even subjects with a strong family history. This may imply the formation of a form of neurobiological template, which increases the vulnerability to subsequent relapse or recurrence. The immediacy of the link between neurotransmitter function and symptoms may be the reason why patients with recurrent major depression need long-term treatment with antidepressant drugs to remain well.

(h) Does mood disorder have a functional neuropathology?

Severe mood disorder is virtually defined by its frequent recurrence or its chronicity. The first episodes of severe depression occur more frequently with increasing age and tend to be more refractory to treatment. Severe mood disorder is associated with ventricular enlargement and sulcal prominence.[44] Late-onset depression is characterized by pronounced impairments in most areas of cognitive function, in particular executive function and processing speed and is increasingly regarded as having a quasi-neurological quality. Indeed, there is an increased rate of white matter lesions, perhaps related to vascular disease, in older patients.[44] The relationship between cognitive deficits and underlying neuropathological changes requires further examination. Elderly patients with early-onset depression demonstrate greater preservation of executive functioning and processing speed, which may reflect partially distinct disease processes possibly mediated by different neuropathological mechanisms.[45] The key hypothesis must be that it is the particular pattern of functional disruption resulting from any cellular pathology that increases the risk of depression. It may be reasonable to describe such a change as a functional neuropathology.

In younger patients, the issue is whether depression per se leads to a functional neuropathology. In patients with unusual refractoriness and chronicity, MRI scanning again suggested reduced grey matter parameters, most significantly in the left hippocampus but also more diffusely in the left parietal and frontal association cortices. Left hippocampal grey matter density was correlated with measures of verbal memory, supporting the functional significance

of the imaging changes. In contrast, patients with severe illnesses fully responsive to treatment showed no differences from controls. Any finding in the chronic group could predate the onset of depression, or be the result of the illness process or its treatment. It is fashionable to attribute structural changes in depression to hypercortisolaemia, but in this study that was not the explanation.[46] A failure of BDNF, related neurogenesis or loss of synaptic plasticity is also a possibility. Reduced hippocampal volume is a relatively consistent finding in many studies of modest size which have also implicated inter-linked structures in basal ganglia and thalamus.[47]

Rather surprisingly, a correlation between lifetime duration of illness and memory performance was also seen in a very large outpatient sample studied after recovery from a discrete episode.[48] It favours a toxic link between the burden of depression and cognition, which has implications for public health. It also means that the mechanisms associated with very severe depression are also relevant in less severe ambulant forms.

Post-mortem studies of the brain in mood disorder have been rare and are limited by tissue availability. Such studies in elderly depression have greater potential validity than the much more numerous investigations of schizophrenia. The Stanley Neuropathology consortium has made samples of tissue widely available from small but well-characterized patient series. In the hippocampus, the most consistent findings are of reduced GABA function and abnormal measures of synaptic density or neuronal plasticity.[49] Such studies have seldom focused on other 'candidate regions' such as the inferior frontal or cingulate cortex or amygdale.[50] Several studies suggest a particular involvement of glial cells.[51] Since glia support the energy requirements of neurones, and their deficient function could account for aspects of the imaging abnormalities found in these disorders: elevated levels of glucocorticoids acting on glia could change their function, or glial changes could represent responses to primary neuronal withdrawal (see also Chapter 2.3.5).

Post-mortem studies can also address the neurochemistry, perhaps more directly and completely than other methods. Normal ageing is accompanied by a decline in a variety of indices of monoamine function including presynaptic markers of 5-HT innervation. In a small series of depressed suicides, there were 54 per cent fewer neurones in the dorsal raphe nucleus expressing SERT mRNA compared with controls.[52] Whether a reduced serotonergic innervation is the critical change that increases the vulnerability to mood disorder of patients with advancing years is not yet established. If so, the potential for MDMA to have long-term effects in heavy users is real and worrying.[53]

In suicide, post-mortem findings have broadly paralleled those in depression, with an important emphasis on 5-HT metabolism and neurotransmission (see Chapter 4.15.3). Whether 5-HT neurotransmission, perhaps like that involving the other monoamines, represents a functional domain implicated independently in a variety of psychiatric syndromes and behaviours remains to be well established.

Conclusions

Mood disorder has an important neurobiological basis. This stretches from a vulnerability, which seems to be attributable to polymorphism in genes critical to stress regulation, through the impact that early experience has on the subsequent programming of the brain for stress responses, to the final responsiveness when encountering particular personal adversity in later life. Biological studies have highlighted the role of key brain areas within the limbic system such as the cingulate cortex and amygdala. We are still a long way from understanding, with any precision, the critical connections and cellular mechanisms, but the function of monoamine neurones generally, and of serotonergic projections in particular, is closely associated with mood regulation. Peptide neurotransmitters have long seemed likely to play a central role in stress regulation, but their potential as targets for antidepressant drug action are yet to be fulfilled. Finally, observations in the most chronic illnesses and in the elderly with depression have highlighted the possibility of a functional neuropathology underlying severe mood disorder. Depression seems to be critically related to the evolving story around neurogenesis in the brain. It is perhaps appropriate that its resolution will require fundamental advances in brain science: psychiatry has always posed, or anyway implied, the most demanding of scientific questions: how does the brain work?

Further information

American College of Neuropsychopharmacology: 5th Generation of progress. Available at: http://www.acnp.org/Default.aspx?Page=5thGenerationChapters

References

1. Castanon, N., Perez-Diaz, F., and Mormede, P. (1995). Genetic analysis of the relationships between behavioral and neuroendocrine traits in Roman high and low avoidance rat lines. *Behavior Genetics*, **25**(4), 371–84.
2. Portella, M.J., Harmer, C.J., Flint, J., *et al.* (2005). Enhanced early morning salivary cortisol in neuroticism. *The American Journal of Psychiatry*, **162**(4), 807–9.
3. Chan, S.W.Y., Goodwin, G.M., and Harmer, C.J. (2007). Highly neurotic never-depressed students have negative biases in information processing. *Psychological Medicine*, **37**, 1281–92.
4. Chan, S.W.Y., Norbury, R., Goodwin, G.M., *et al.* (2008). Risk for depression is associated with exaggerated neural responses to fearful facial expressions of emotion. *British Journal of Psychiatry* (in Press).
5. Fullerton, J., Cubin, M., Tiwari, H., *et al.* (2003). Linkage analysis of extremely discordant and concordant sibling pairs identifies quantitative-trait loci that influence variation in the human personality trait neuroticism. *American Journal of Human Genetics*, **72**(4), 879–90.
6. Kendler, K.S., Kuhn, J.W., and Prescott, C.A. (2004). Childhood sexual abuse, stressful life events and risk for major depression in women. *Psychological Medicine*, **34**(8), 1475–82.
7. Liu, D., Diorio, J., Tannenbaum, B., *et al.* (1997). Maternal care, hippocampal glucocorticoid receptors, and hypothalamic–pituitary–adrenal responses to stress. *Science*, **277**(5332), 1659–62.
8. Caldji, C., Tannenbaum, B., Sharma, S., *et al.* (1998). Maternal care during infancy regulates the development of neural systems mediating the expression of fearfulness in the rat. *Proceedings of the National Academy of Sciences of the United States of America*, **95**(9), 5335–40.
9. Heim, C., Plotsky, P.M., and Nemeroff, C.B. (2004). Importance of studying the contributions of early adverse experience to neurobiological findings in depression. *Neuropsychopharmacology*, **29**(4), 641–8.
10. Seckl, J.R. and Meaney, M.J. (2006). Glucocorticoid 'programming' and PTSD risk. *Annals of the New York Academy of Sciences*, **1071**, 351–78.
11. Biondi, M. and Picardi, A. (1999). Psychological stress and neuroendocrine function in humans: the last two decades of research. *Psychotherapy and Psychosomatics*, **68**(3), 114–50.

12. Kendler, K.S., Thornton, L.M., and Gardner, C.O. (2000). Stressful life events and previous episodes in the etiology of major depression in women: an evaluation of the 'kindling' hypothesis. *The American Journal of Psychiatry*, **157**(8), 1243–51.
13. Tsiouris, J.A. (2005). Metabolic depression in hibernation and major depression: an explanatory theory and an animal model of depression. *Medical Hypotheses*, **65**(5), 829–40.
14. Goodwin, G.M. (1997). Neuropsychological and neuroimaging evidence for the involvement of the frontal lobes in depression. *Journal of Psychopharmacology*, **11**(2), 115–22.
15. Mayberg, H.S., Brannan, S.K., Mahurin, R.K., *et al.* (1997). Cingulate function in depression: a potential predictor of treatment response. *Neuroreport*, **8**(4), 1057–61.
16. Steele, J.D. and Lawrie, S.M. (2004). Segregation of cognitive and emotional function in the prefrontal cortex: a stereotactic meta-analysis. *Neuroimage*, **21**(3), 868–75.
17. Mayberg, H.S., Lozano, A.M., Voon, V., *et al.* (2005). Deep brain stimulation for treatment-resistant depression. *Neuron*, **45**(5), 651–60.
18. Shah, P.J., Ogilvie, A.D., Goodwin, G.M., *et al.* (1997). Clinical and psychometric correlates of dopamine D-2 binding in depression. *Psychological Medicine*, **27**(6), 1247–56.
19. Meyer, J.H., McNeely, H.E., Sagrati, S., *et al.* (2006). Elevated putamen D-2 receptor binding potential in major depression with motor retardation: an [C-11] raclopride positron emission tomography study. *The American Journal of Psychiatry*, **163**(9), 1594–602.
20. Bhagwagar, Z., Rabiner, E.A., Sargent, P.A., *et al.* (2004). Persistent reduction in brain serotonin(1A) receptor binding in recovered depressed men measured by positron emission tomography with [C-11]WAY-100635. *Molecular Psychiatry*, **9**(4), 386–92.
21. Meyer, J.H. (2007). Imaging the serotonin transporter during major depressive disorder and antidepressant treatment. *Journal of Psychiatry & Neuroscience*, **32**(2), 86–102.
22. Staley, J.K., Sanacora, G., Tamagnan, G., *et al.* (2006). Sex differences in diencephalon serotonin transporter availability in major depression. *Biological Psychiatry*, **59**(1), 40–7.
23. Newberg, A., Amsterdam, J., and Shults, J. (2007). Dopamine transporter density may be associated with the depressed affect in healthy subjects. *Nuclear Medicine Communications*, **28**(1), 3–6.
24. Shang, Y.L., Gibbs, M.A., Marek, G.J., *et al.* (2007). Displacement of serotonin and dopamine transporters by venlafaxine extended release capsule at steady state—a [I-123]2 beta-carbomethoxy-3 beta-(4-iodophenyl)-tropone single photon emission computed tomography imaging study. *Journal of Clinical Psychopharmacology*, **27**(1), 71–5.
25. Rubin, R.T., Phillips, J.J., Sadow, T.F., *et al.* (1995). Adrenal gland volume in major depression: increase during the depressive episode and decrease with successful treatment. *Archives of General Psychiatry*, **52**(3), 213–18.
26. Carroll, B.J., Feinberg, M., Greden, J.F., *et al.* (1981). A specific laboratory test for the diagnosis of melancholia. Standardization, validation, and clinical utility. *Archives of General Psychiatry*, **38**(1), 15–22.
27. Ribeiro, S.C.M., Tandon, R., Grunhaus, L., *et al.* (1993). The DST as a predictor of outcome in depression: a meta-analysis. *The American Journal of Psychiatry*, **150**(11), 1618–29.
28. Thase, M.E., Dube, S., Bowler, K., *et al.* (1996). Hypothalamic–pituitary–adrenocortical activity and response to cognitive behavior therapy in unmedicated, hospitalized depressed patients. *The American Journal of Psychiatry*, **153**, 886–91.
29. Young, A.H., Gallagher, P., Watson, S., *et al.* (2004). Improvements in neurocognitive function and mood following adjunctive treatment with mifepristone (RU-486) in bipolar disorder. *Neuropsychopharmacology*, **29**(8), 1538–45.
30. Otte, C., Moritz, S., Yassouridis, A., *et al.* (2007). Blockade of the mineralocorticoid receptor in healthy men: effects on experimentally induced panic symptoms, stress hormones, and cognition. *Neuropsychopharmacology*, **32**(1), 232–8.
31. Goodwin, G.M., Muir, W.J., Seckl, J.R., *et al.* (1992). The effects of cortisol infusion upon hormone secretion from the anterior pituitary and subjective mood in depressive illness and in controls. *Journal of Affective Disorders*, **26**(2), 73–83.
32. Dinan, T.G., Lavelle, E., Cooney, J., *et al.* (1997). Dexamethasone augmentation in treatment-resistant depression. *Acta Psychiatrica Scandinavica*, **95**(1), 58–61.
33. Pariante, C.M., Thomas, S.A., Lovestone, S., *et al.* (2004). Do antidepressants regulate how cortisol affects the brain? *Psychoneuroendocrinology*, **29**(4), 423–47.
34. Mitchell, A.J. (1998). The role of corticotropin releasing factor in depressive illness: a critical review. *Neuroscience and Biobehavioral Reviews*, **22**(5), 635–51.
35. Nemeroff, C.B. and Vale, W.W. (2005). The neurobiology of depression: inroads to treatment and new drug discovery. *Journal of Clinical Psychiatry*, **66**, 5–13.
36. Keller, M., Montgomery, S., Ball, W., *et al.* (2006). Lack of efficacy of the substance P (neurokinin(1) receptor) antagonist aprepitant in the treatment of major depressive disorder. *Biological Psychiatry*, **59**(3), 216–23.
37. Berger, M. and Riemann, D. (1993). REM sleep in depression—an overview. *Journal of Sleep Research*, **2**(4), 211–23.
38. Schittecatte, M., Garcia Valentin, J., Charles, G., *et al.* (1995). Efficacy of the 'clonidine REM suppression test (CREST)' to separate patients with major depression from controls; a comparison with three currently proposed biological markers of depression. *Journal of Affective Disorders*, **33**(3), 151–7.
39. Moore, P., Gillin, J.C., Bhatti, T., *et al.* (1998). Rapid tryptophan depletion, sleep electroencephalogram, and mood in men with remitted depression on serotonin reuptake inhibitors. *Archives of General Psychiatry*, **55**(6), 534–9.
40. Thase, M.E., Simons, A.D., and Reynolds, C.F. III. (1993). Psychobiological correlates of poor response to cognitive behavior therapy: potential indications for antidepressant pharmacotherapy. *Psychopharmacology Bulletin*, **29**(2), 293–301.
41. Lauer, C.J., Schreiber, W., Holsboer, F., *et al.* (1995). In quest of identifying vulnerability markers for psychiatric disorders by all-night polysomnography. *Archives of General Psychiatry*, **52**(2), 145–53.
42. Schatzberg, A.F., Samson, J.A., Bloomingdale, K.L., *et al.* (1989). Toward a biochemical classification of depressive disorders. X.Urinary catecholamines, their metabolites, and D-type scores in subgroups of depressive disorders. *Archives of General Psychiatry*, **46**(3), 260–8.
43. Smith, K.A., Fairburn, C.G., and Cowen, P.J. (1997). Relapse of depression after rapid depletion of tryptophan. *Lancet*, **349**(9056), 915–19.
44. Videbech, P. (1997). MRI findings in patients with affective disorder: a meta-analysis. *Acta Psychiatrica Scandinavica*, **96**(3), 157–68.
45. Herrmann, L.L., Goodwin, G.M., and Ebmeier, K.P. (2007). The cognitive neuropsychology of depression in the elderly. *Psychological Medicine*, **37**, 1693–702.
46. Shah, P.J., Ebmeier, K.P., Glabus, M.F., *et al.* (1998). Cortical grey matter reductions associated with treatment-resistant chronic unipolar depression: controlled magnetic resonance imaging study. *The British Journal of Psychiatry*, **72**(172), 527–32.
47. Sheline, Y.I. (2003). Neuroimaging studies of mood disorder effects on the brain. *Biological Psychiatry*, **54**(3), 338–52.
48. Gorwood, P., Corruble, E., Falissard, B., *et al.* (2008). Toxic effects of depression on brain function: impairment of delayed recall reflects the cumulative length of the depressive disorder in a large sample of depressed out-patients. *American Journal of Psychiatry,* **165**, 731–9.
49. Knable, M.B., Barci, B.M., Webster, M.J., *et al.* (2004). Molecular abnormalities of the hippocampus in severe psychiatric illness: postmortem findings from the Stanley neuropathology consortium. *Molecular Psychiatry*, **9**(6), 609–20.

50. Harrison, P.J. (2002). The neuropathology of primary mood disorder. *Brain*, **125**, 1428–49.
51. Cotter, D.R., Pariante, C.M., and Everall, I.P. (2001). Glial cell abnormalities in major psychiatric disorders: the evidence and implications. *Brain Research Bulletin*, **55**(5), 585–95.
52. Arango, V., Underwood, M.D., Boldrini, M., *et al.* (2001). Serotonin 1A receptors, serotonin transporter binding and serotonin transporter mRNA expression in the brainstem of depressed suicide victims. *Neuropsychopharmacology*, **25**(6), 892–903.
53. Green, A.R. and Goodwin, G.M. (1996). Ecstasy and neurodegeneration. *British Medical Journal*, **312**(7045), 1493–4.

4.5.7 Course and prognosis of mood disorders

Jules Angst

The importance of course

Ever since Kahlbaum's monograph 1863[1] the course and outcome of mental disorders have played important roles as criteria and validators of psychiatric classification. The prognosis is fundamental for doctor and patient when deciding whether to start long-term prophylactic medication and, at a later stage, whether to stop a successful long-term treatment. Course is a crucial factor in estimating the social consequences, costs, suicide risk, and mortality associated with mood disorders.

The description of course includes the age of onset, episode length, recurrence of episodes, residual symptoms between episodes and outcome (remission, chronicity, death). These aspects are covered in this chapter.

Stability of the diagnoses of mood disorders

Mood disorders can be roughly sub-classified into unipolar mania, bipolar disorder, and unipolar depression.[2] The three groups differ significantly as regards family history, personality, course and outcome, including mortality. Unipolar mania has not yet been studied extensively, and for this reason will not be dealt with here.

Distinguishing between bipolar disorder and unipolar depressive disorder is hampered by the fact that the diagnosis of unipolar depression is always uncertain. Many depressives are hidden bipolar patients: a long-term follow-up study over 27 years showed a constant rate of diagnostic change from depression to hypomania of 1.25 per cent per year of follow-up. As a consequence of this diagnostic instability, the exact ratio of bipolar to unipolar depressive subjects in the population is unknown; modern estimates range from 1:5 to 1:1. The discussion of bipolar disorder has therefore to take account of its unipolar counterpart.

Bipolar disorder

Onset

In patients admitted to hospital between 1913 and 1940 and not treated by electroconvulsive therapy or modern psychotropic drugs, bipolar disorder clearly manifested earlier than unipolar depression; this finding is confirmed by modern community studies. In most patients bipolar disorder begins during adolescence but in some cases may already manifest in childhood. Unfortunately pediatric psychiatry cannot yet provide prospective data from large representative community studies on the onset and course of bipolar disorder starting in childhood or adolescence.

In the offspring of bipolar parents social functioning up to the age of 18 develops normally before the onset of their illness.

Bipolar disorders usually begin as depression, and it takes a further 5 years on average until the first manic syndrome manifests.[3] There may be unspecific prodromal symptoms in the form of mood lability, vegetative lability, somatization, or being hyper-alert or easily excited; there may also be discrete cognitive impairment present before the onset of the affective disorder.[4] After the onset of the disorder social functioning often begins to be impaired.

The first depressive and manic manifestations are commonly mild, brief, or uncharacteristic, and are often only diagnosed in retrospect after years.

Prospective epidemiological studies in adolescents and young adults found the onset of bipolar disorder to occur in the teens (with means and medians around 15 years or later), whereas studies of hospitalized patients date its onset in retrospect in the early 20s or in the 30s.

Bipolar-I illness manifests earlier than bipolar-II and psychotic bipolar disorder. Late-onset bipolar disorder is extremely rare but does occur and may be associated with specific neuropathology. An early age of onset of the disorder, usually manifesting as depression, is correlated with suicidality, comorbid substance-use disorder and a rapid cycling course.

A two-peak distribution of the age of onset has sometimes been described for both bipolar disorder and depression in men and women, with no specific association in women between the second peak and the menopause.

Duration of episodes

Most episodes are short, but 10 to 20 per cent become chronic (lasting more than 2 years); the distribution of episode length is log normal, and therefore percentiles and not averages should be used as parameters. Using data collected a century ago on the natural length of episodes of mania and bipolar disorder, mainly among hospitalized patients, it is possible to compute a median length of 4 to 6 months for mania and 5 to 6 months for bipolar disorder. These figures do not differ from those obtained today despite a wide range of antimanic and antidepressant treatments. Among hospitalized bipolar patients episode length (median) was 4.2 months; 25 per cent of bipolar episodes lasted more than 7.3 months.

About 20 to 30 per cent of episodes are *biphasic* (mania with subsequent switch into depression, or depression with subsequent, switch into hypomania/mania); such high switch rates were already observed before the introduction of electroconvulsive therapy and antidepressants. An effective treatment does not induce a switch but increases (compared to placebos) the rates of responders; and the response is a precondition for the natural switch.

Recurrence of episodes

Recurrence is typical of mood disorders. It can be described by the number of episodes, the length of intervals (measured from remission to the onset of a new episode), and the length of cycles

(measured from the beginning of one episode to the beginning of the next). In prospective studies, time to the onset of a new episode is frequently used as a parameter for survival analyses and frailty analyses of recurrence.

In both bipolar disorder and unipolar depression the time from the first to the second episode is on average much longer than from the second to the third episode and so on. This progressive shortening of cycles and free intervals then levels off and fluctuates around a certain (but still variable) individual limit. Most published data on interval length or cycle length are methodologically flawed because they have not been corrected for the number of episodes/cycles observed. Nonetheless, multiple episodes obviously follow each other in more rapid succession than a few episodes distributed over a lifetime. Statistically, a normal distribution of cycle length can be obtained by log *n* transformation. Even after taking episode numbers into account, there is a clear intra-individual trend to a progressive shortening of cycle length, as demonstrated by frailty analyses[5] dimming the prognosis for both bipolar disorder and unipolar depression. Initial cycle length tends to be shorter in late-onset than in early-onset mood disorders, increasing the risk of recurrence in the elderly.

Precipitating events play an important role in the onset of the first few affective episodes; thereafter recurrence seems to become gradually autonomous with stressful events contributing little or nothing to the process. Stressors may not only precipitate episodes but also increase a pre-existent vulnerability, sensitizing the individual and thereby making him or her more vulnerable to further episodes (kindling effect). In bipolar illness there is no difference in the quality or quantity of stressors precipitating depressive and manic episodes; a legacy or the loss of a relative can induce depression or mania. The sensitivity to stressors has also a genetic component.

Over a patient's lifetime his condition continuously fluctuates on a dimension of severity, which ranges from psychotic, via major and minor syndromes (cyclothymic and minor bipolar disorders), cyclothymic temperament within the norm, symptomatic to symptom-free.

The NIMH Collaborative Depression Study with annual assessments of outpatients over an average of 13 years, demonstrated that bipolar-II patients spent slightly more time with symptoms/syndromes (33 per cent) than bipolar-I patients (27 per cent). In both subtypes of bipolar disorders depressive periods were three times more common than manic periods[6] but bipolar-I patients suffered more from psychotic features. In a 25-year follow-up study of hospitalized mood disorder patients, manic and depressive episodes were about equally present in bipolar-I patients, whereas in bipolar-II patients the course was dominated by depression.

Daily assessments of the course by the life-chart methodology over more than 3 years confirmed that bipolar outpatients spent a three-fold greater amount of time in depression than hypomania. But it was also shown that bipolar-I patients spent significantly more time in hypomania than their bipolar-II counterparts but an equal time in depression; in more than half the time the patients were euthymic.

Over lifetime bipolar patients experience twice as many episodes as unipolar depressives, a difference which is not explained by the manifestation of manic episodes in addition to depression. The total number of episodes observed depends on the length of observation. In a 22- to 26-year follow-up study, bipolar patients experienced a median of 10 episodes, but depressive patients only four. A family history of mood disorders increases recurrence. The proportion of mania to depression remains fairly stable across multiple episodes, but over their lifetime patients spend more time in depression than in mania.

Outcome

(a) Incomplete remission of episodes

Remission after bipolar episodes is frequently incomplete in terms of symptomatic and functional recovery. Residual symptoms are common in patients in both psychiatric and general practice settings and bipolar subjects identified in community studies. Residual depressive symptoms are more impairing than hypomanic symptoms, which may even enhance functioning.[7] The chronic residual symptoms are mainly depressed mood, anxiety, and somatic disturbances, such as insomnia, hypersomnia, headaches, neurasthenic complaints, reduced libido, and gastrointestinal symptoms. Functional recovery was found to develop later than symptomatic recovery. Short-term outcome is less favourable in patients with agitation, rapid cycling, poor premorbid functioning, comorbidity with anxiety disorders, social phobia, substance use, OCD, obesity, personality disorders, sexual trauma, abuse, and behaviour disorders in childhood.[8] Manic versus mixed episodes do not differ in outcome after 1 year.

(b) Long-term course and outcome

The long-term course and outcome of bipolar disorders is characterized by high recurrence rates, frequent residual symptoms between episodes; compared to depression they carry a higher risk of suicide attempts but lower risk of suicides.

Bipolar disorder has a poorer outcome than depression and there is no burn-out with age. After a follow-up of 22 to 26 years, definitive recovery (at least 5 years with good social adaptation) was found in 25 per cent of 186 depressive subjects, whereas the figure was only 16 per cent of the 220 bipolar patients; a chronic course lasting at least 2 years without remission was present in 12 to 14 per cent of depressive and bipolar patients. A chronic course is associated with early life adversity, including sexual and physical abuse.

Comorbidity with alcoholism, a factor known to correlate with poorer outcome and increased mortality, was found in 30 per cent of the bipolar patients. Modern treatment has improved the outcome of mood disorders by reducing chronicity and rehospitalization.

Recurrence (number of hospital admissions) may increase the risk of dementia.

Mortality

Mortality is expressed by the standardized mortality ratio (**SMR**) in comparison with the normal population (SMR = 1.0). Data on hospitalized psychiatric patients in Sweden[9] give an overall SMR for bipolar men of 2.5 and for women 2.7. The high mortality is a consequence not only of suicides (SMR = 15.0 for males and 22.4 for females) but of most other causes of death (excluding cancer): infections, endocrine, cardiovascular, cerebrovascular, respiratory and gastrointestinal disorders, homicides, accidents, traffic accidents and secondary substance-use disorders (mainly alcohol use disorder), etc. This is true for both men and women.

Consequences for treatment

Incomplete recovery and high recurrence are the main problems in the treatment of bipolar disorders. In the treatment of an acute episode the primary goals are full recovery and the prevention of relapses. The length of treatment of an acute episode depends on its estimated spontaneous duration, which may be derived from earlier episodes. In case of doubt, 6-months' maintenance treatment after recovery is the rule. Long-term prophylaxis should be maintained lifelong.

Most time is spent in depression, therefore a combination of a mood stabilizer with an antidepressant is often indicated. The same is true for the long-term medication of the disorder, where combined treatments are the rule.

Special attention should be given to suicidal bipolar patients, who should preferably be treated with drugs shown to be antisuicidal (lithium or clozapine combined often with antidepressants). A lifelong study demonstrated a three- to five-fold reduction of suicides. Comorbidity with alcohol use disorder increased the mortality but not the suicide rate.

Major depressive disorder (MDD)

Onset

Unlike bipolar disorder, depression may start at any time of life and has therefore a later mean age of onset. In the United Kingdom a mean age of onset of 33 years was found in hospitalized patients. In a large United States study of outpatients the mean was 29.4 years, but in 53 per cent of cases the onset was before the age of 21. The distribution of the age of onset is bimodal, with peaks in the 30s and 50s.

Prospective data suggest that MDD very often begins as subthreshold depression; this is especially true for late-onset cases, where the full MDD episodes also tend to last longer and to become chronic more often. Age of onset and earlier episodes assessed in retrospect are subject to dramatically false recall.

There is no true dichotomy between early-onset and late-onset depression but a continuous distribution accompanied by a systematic decrease in genetic vulnerability (morbid risk among first-degree relatives) and an increase in precipitation by environmental factors. The correlation of age of onset with the genetic component was found in both patient and community samples. Childhood traumata create a vulnerability which promotes earlier onset and higher comorbidity.

Duration of episodes

The length of depressive episodes is log normally distributed; therefore simple mean values are meaningless. Compared to bipolar disorder, episodes of depressive disorders last about 1 month longer (median duration of 5.4 months); and about 18 to 25 per cent of patients develop chronic depression with a minimum duration of 2 years. In the general population, among whom there are many untreated cases of depression, episodes were found to be shorter; the 25th, 50th, and 75th percentiles were 4, 8, and 16 weeks respectively for recurrent episodes.

Chronicity is clearly correlated with age and persistent cognitive deficits; it is common in the elderly but relatively rare in adolescents. As is in bipolar disorder chronicity is associated with early life adversity, including sexual and physical abuse.

Recurrence of episodes

Major depression is recurrent in about 85 per cent of cases; compared to bipolar disorders, however, depressive patients experience only half as many episodes over their lifetime. The cycle length (time from the start of an episode to the start of a subsequent episode) is consequently longer than in bipolar disorders. There is also a systematic shortening of cycle length with the increasing number of episodes, as shown by frailty analyses. The precipitation of episodes by life events—frequent initially—decreases as the number of episodes grows; the periodicity becomes increasingly autonomous. A twin study suggests that undesirable life events play a significant role in the recurrence of depression in women with a low genetic risk.

In a recent large representative record study in Denmark (N = 20 350 first admissions) unipolar depressives had strikingly lower recurrence rates (hospitalizations) than bipolars, the rates for both correlating with the number of previous episodes.[10] The authors concluded: 'The course of severe unipolar and bipolar disorder seems to be progressive in nature despite the effect of treatment and irrespective of gender, age and type of disorder'. Risk factors for recurrence are previous recurrence, long duration of episodes, late onset, age, severity, and incomplete remission.

Residual symptoms represent a strong risk factor for further recurrence; a survival analysis by Paykel *et al.*[11] found a three-fold higher risk of recurrence (76 per cent) in patients with residual symptoms than in those without (25 per cent). This sub-threshold depressive morbidity is clinically relevant and a clear risk factor for future recurrence and suicidality, especially in the elderly.

Outcome

In long-term follow-up studies, 43 to 52 per cent of depressed outpatients became symptom-free between their episodes, and the other half continued to suffer from dynamically fluctuating residual syndromes or symptoms. The corresponding cross-sectional status may be diagnosed as dysthymia, recurrent brief depression, minor depression, residual syndromes and symptoms, or as full recovery. Initial severity and comorbidity are positively correlated with poor outcome in terms of poorer functioning and incomplete remission. In a British study residual symptoms of major depression, defined by a score of eight or more on the 17-item Hamilton Depression scale, were found in 32 per cent of 60 patients 12 to 15 months after remission.[11] In a large cross-cultural study (N = 968) conducted over 9 months, between 25 and 48 per cent of cases experienced complete remission,[12] rates which also depended on the severity of the depression and comorbidity. Severe residual symptoms correlate with long-term morbidity, impaired social functioning at work and in relationships and suicidal behaviour, especially in the elderly, amongst whom remission is present in only about one third of cases.

Recent evidence from a 30-month prospective community-based cohort study of 75-year-old subjects suggests that a history of depression may increase the risk of senile dementia,[13] which is compatible with findings from Denmark that recurrence (number of hospital admissions) increases the risk of dementia.[14]

Mortality

Unipolar depressives have twice the mortality risk of the general population. The SMR (standardized mortality ratio) for suicide is

21 for males and 27 for females;[9] these figures are higher than those for bipolar disorder. Many other causes of death are also more common among unipolar depressives than in the general population but to a lesser extent than among bipolar patients.

The frequently quoted suicide rate of 12 to 19 per cent is only valid for selected hospitalized patient samples, which, by definition, include many suicidal patients. In community and outpatient samples, suicides account for a considerably smaller percentage of deaths; but long-term data is not available yet.

A lifelong follow-up study showed a suicide-preventive effect of administering low-dose antidepressant medication to severe unipolar depressives.

Course of other subtypes of mood disorders

Dysthymia

Subjects with a depressive personality disorder are especially prone to develop dysthymia. Dysthymia is by definition a chronic form of depression; nevertheless dysthymic patients have a similar outcome to major depressives: in outpatients the 5-year recovery rate was 53 per cent. Dysthymia is highly comorbid with other psychiatric disorders, especially with major depression (which it may precede or follow), so worsening the prognosis.

Minor depression

In the general population the recurrence rates of minor depression are comparable to those of major depression, a fact confirmed by survival analyses.[15] A diagnostic change from minor to major depression, or the reverse, is frequent during the course of mood disorders. Minor depression increases about five-fold the risk for the development of major depression. Primary minor depression, like depressive symptoms in general, is a significant risk factor for major depression. It can also represent a residual state of major depression and is a strong risk factor for further recurrence.

Among the elderly minor depression is common as a residual state. Both minor and major depression should be considered seriously as a target for preventive intervention and treatment; full recovery should be the goal.[15]

Seasonal affective disorder

Many patients experience depressive episodes mainly in autumn and winter; mania tends to occur more often in summer. Seasonal affective disorder (SAD) remained seasonal in 70 per cent of 43 cases followed up over 2 to 5 years and in 42 per cent over 8.8 years.[16] The diagnostic stability of SAD was fairly good (26 to 57 per cent). A large number of patients developed seasonal subthreshold (subsyndromal) depression, whereas full remission was present in only 15 to 20 per cent.

Rapid cycling mood disorder

There is no generally accepted definition of rapid cycling; it is usually defined by the occurrence of at least four episodes per year, counting arbitrarily a biphasic episode as two episodes. Rapid cycling occurs almost exclusively in bipolar disorder; it is more frequent in females and in the bipolar-II subtype. Rapid cycling often manifests at an early age and increases the risk of suicide attempts but does not appear to represent a final course pattern of bipolar disorder: it is often a transient, non-familial manifestation of the disorder. In a prospective follow-up conducted over approximately 3 years the diagnosis was stable in about half the cases and the other half became simply recurrent (non-rapid cycling); in a control group 10 per cent of the non-rapid-cycling patients converted to rapid cycling.[17] Studies on the long-term prognosis are inconclusive.

Consequences for treatment

Antidepressant treatment can shorten the time to recovery. Therapeutic decisions on the length of acute treatment will depend on the length of the individual's previous episodes and on the average episode length observed in follow-up studies. The length of affective episodes has probably not changed in 100 years. Antidepressants cannot shorten the episodes but can minimize the symptoms. Treatment should be maintained for the full duration of episodes, which are frequently masked, otherwise relapses must be expected. Full symptomatic and functional remission to the premorbid level should be the goal of treatment.

As with acute treatment, the choice of a long-term prophylactic medication should take into consideration the previous individual course of the disorder plus the general scientific knowledge of course and prognosis, and keep in mind the increased mortality, especially the high suicide risk associated with depression. Recurrence is also a feature of mild cases, but in contrast to severe cases the suicide mortality is probably low.

Over a patient's lifetime, each new recurrence is associated with a new risk of suicide and requires long-term prophylaxis with lithium, combined with low-dose antidepressants. This is especially necessary in the presence of suicidality in the previous or in the family history. If further studies confirm that the recurrence of affective episodes increases the risk of dementia, such a long-term prophylaxis may become even more important.

Conclusions

Bipolar disorder and depression are serious illnesses responsible for most suicides in the population and are recurrent lifelong in most cases. The remission between episodes is often incomplete, increasing the risk of recurrence. Compared to depression bipolar disorder is twice as recurrent and complicated by higher comorbidity with multiple somatic and psychiatric disorders, especially with substance-use disorder which shortens life expectancy even more. Most patients spend about half their lifetime in good health but the other half in largely depressive mood states fluctuating on the broad severity spectrum. Sub-diagnostic morbidity has been recognized as clinically very relevant and in need of permanent treatment. Great progress in the study of continuous fluctuations has been made by the introduction of the life-chart methodology and by computer-assisted daily assessments, methods which hold further promise.[18] Treatment should focus on all manifestations of the illness, including minor morbidity, in order to achieve full recovery to the premorbid level of functioning; in a long-term perspective the primary goal remains the prevention of recurrence, secondary substance-use disorders and suicides.

Future studies should also try to reduce the dementia associated with affective disorders.[13,19,20]

Acknowledgement

I wish to thank Per Bech for his valuable comments and suggestions during the revision of this chapter.

Further information

Goodwin, F.K. and Jamison, K.R. (2007). *Manic-depressive illness* (2nd edn). Oxford University Press, New York.

Maj, M., Akiskal, H.S., Lopez-Ibor, J.J., *et al.* (eds.) (2002). *Bipolar disorder*. John Wiley & Sons, Chichester.

Marneros, A. (ed.) (2004). *Das Neue Handbuch der Bipolaren und Depressiven Erkrankungen*. Georg Thieme, Stuttgart.

References

1. Kahlbaum, K. (1863). *Die Gruppirung der psychischen Krankheiten und die Eintheilung der Seelenstörungen*. A.W. Kafemann, Danzig.
2. Angst, J. (2007). The bipolar spectrum. *The British Journal of Psychiatry*, **190**, 189–91.
3. Hillegers, M.H.J., Reichart, C.G., Wals, M., *et al.* (2005). Five-year prospective outcome of psychopathology in the adolescent offspring of bipolar parents. *Bipolar Disorders*, **7**, 344–50.
4. Christensen, M.V., Kyvik, K.O., and Kessing, L.V. (2006). Cognitive function in unaffected twins discordant for affective disorder. *Psychological Medicine*, **36**, 1119–29.
5. Kessing, L.V., Hansen, M.G., Andersen, P.K., *et al.* (2004). The predictive effect of episodes on the risk of recurrence in depressive and bipolar disorders—a life-long perspective. *Acta Psychiatrica Scandinavica*, **109**, 339–44.
6. Judd, L.L., Schettler, P.J., Akiskal, H.S., *et al.* (2003). Long-term symptomatic status of bipolar I vs. bipolar II disorders. *International Journal of Neuropsychopharmacology*, **6**, 127–37.
7. Judd, L.L., Akiskal, H.S., Schettler, P.J., *et al.* (2005). Psychosocial disability in the course of bipolar I and II disorders. A prospective, comparative, longitudinal study. *Archives of General Psychiatry*, **62**, 1322–30.
8. Paykel, E.S. (1998). Remission and residual symptomatology in major depression. *Psychopathology*, **31**, 5–14.
9. Ösby, U., Brandt, L., Correia, N., *et al.* (2001). Excess mortality in bipolar and unipolar disorder in Sweden. *Archives of General Psychiatry*, **58**, 844–50.
10. Kessing, L.V., Andersen, P.K., Mortensen, P.B., *et al.* (1998). Recurrence in affective disorder. I. Case register study. *The British Journal of Psychiatry*, **172**, 23–8.
11. Paykel, E.S., Ramana, R., Cooper, Z., *et al.* (1995). Residual symptoms after partial remission: an important outcome in depression. *Psychological Medicine*, **25**, 1171–80.
12. De Almeida Fleck, M.P., Simon, G., Herrman, H., *et al.* (2005). Major depression and its correlates in primary care settings in six countries. 9-month follow-up study. *The British Journal of Psychiatry*, **186**, 41–7.
13. Fischer, P., Zehetmeyer, S., Jungwirth, S., *et al.* (in press). Risk factors for Alzheimer dementia in a community-based birth-cohort at age 75. *Dementia and Geriatic Cognitive Disorders.*
14. Andersen, K., Lolk, A., Kragh-Sorensen, P., *et al.* (2005). Depression and the risk of Alzheimer disease. *Epidemiology*, **16**, 233–8.
15. Kessler, R.C., Zhao, S., Blazer, D.G., *et al.* (1997). Prevalence, correlates, and course of minor depression and major depression in the National Comorbidity Survey. *Journal of Affective Disorders*, **45**, 19–30.
16. Schwartz, P.J., Brown, C., Wehr, T.A., *et al.* (1996). Winter-seasonal affective disorder: a follow-up study of the first 59 patients of the NIMH seasonal studies program. *The American Journal of Psychiatry*, **153**, 1028–36.
17. Bauer, M.S., Calabrese, J., Dunner, D.L., *et al.* (1994). Multisite data reanalysis of the validity of rapid cycling as a course modifier for bipolar disorder in DSM-IV. *The American Journal of Psychiatry*, **151**, 506–15.
18. Bauer, M., Grof, P., Rasgon, N., *et al.* (2006). Self-reported data from patients with bipolar disorder: impact on minimum episode length for hypomania. *Journal of Affective Disorders*, **96**, 101–5.
19. Angst, J., Gamma, A., Gerber-Werder, R., *et al.* (2007). Does long-term medication with lithium prevent or attenuate dementia in bipolar and depressed patients? *International Journal of Psychiatry in Clinical Practice*, **11**, 2–8.
20. Kessing, L.V. and Andersen, P.K. (2007). Does the risk of developing dementia increase with the number of episodes in patients with depressive disorder and in patients with bipolar disorder? *Journal of Neurology, Neurosurgery, and Psychiatry*, **75**, 1662–6.

4.5.8 Treatment of mood disorders

E. S. Paykel and J. Scott

Evidence

Medication and physical treatments

(a) Acute treatments for depression

(i) Antidepressants: general issues

The first modern antidepressants became available in the late 1950s, coinciding with introduction of randomized controlled trials in psychiatry. Tightening of licensing requirements has ensured good efficacy for new antidepressants. Overall efficacy and speed of response of most antidepressants are similar. In addition to meta-analyses of specific antidepressant classes, there have been meta-analyses confirming efficacy in specific patient groups or disorder subtypes, including dysthymia[1] and the elderly.[2]

There are moderate, but not very large, effect sizes. About 30 per cent more subjects respond well on antidepressants than on placebo, partly because good response is often seen in placebo groups, probably not due to the placebo but to spontaneous improvement and non-specific treatment effects such as those of seeing a helping figure, supportive psychotherapy, and hospitalization. Non-specific outcome tends to be worse among severely ill patients, but here also antidepressants may be less effective. Placebo-controlled trials are still mandated by regulatory authorities for new antidepressants, since comparisons between active drugs have high risk of type 2 error.

Table 4.5.8.1 lists the principal antidepressants and recommended doses, omitting minority drugs limited to a small number of countries. Some of the antidepressants listed are not available in all countries. Recommended doses depend partly on national authorities and readers should check the situation in their own country.

(ii) Reuptake inhibitors

Tricyclic antidepressants. In addition to the tricyclics listed in Table 4.5.8.1, maprotiline and amoxapine are available in some countries. There have been many older efficacy reviews, with

Table 4.5.8.1 Antidepressant medications*

Drug	Usual dose range (mg/day)**
Reuptake inhibitors	
Tricyclics	
Amitriptyline	75–300
Clomipramine	75–300
Desipramine	75–300
Dosulepin (dothiepin)	75–225
Doxepin	75–150
Imipramine	75–300
Lofepramine	70–210
Nortriptyline	75–150
Proptriptyline	15–60
Trimipramine	75–300
Serotonin and noradrenaline reuptake inhibitors (SNRIs)	
Duloxetine	60
Venlafaxine	75–375 (check)
Selective serotonin reuptake inhibitors (SSRIs)	
Citalopram	20–60
Escitalopram	10–20
Fluoxetine	20–80
Fluvoxamine	50–300
Paroxetine	20–60
Sertraline	50–200
Selective noradrenaline reuptake inhibitor (NARI)	
Reboxetine	4–12
Monoamine oxidase inhibitors	
Irreversible monoamine oxidase inhibitors	
Isocarboxazid	30–60
Phenelzine	45–90
Tranylcypromine	20–30
Reversible monoamine oxidase inhibitor (RIMA)	
Moclobemide	300–600
Others	
Bupropion	200–450
Mianserin	30–90
Mirtazepine	15–45
Trazodone	150–600

*This list is not fully comprehensive because of new developments and national differences.

**Official dose recommendations vary between countries and should always be checked.

further studies accumulating as tricyclics have been included in placebo-controlled trials of new antidepressants.

Earlier views that tricyclics were more effective in endogenous and psychotic depressives have not been confirmed and effects extend across a broad spectrum of depressives, extending more widely into anxiety disorders, panic disorder, and obsessive-compulsive disorder. Tricyclic use has lessened in recent years in favour of antidepressants with fewer side effects.

Serotonin and noradrenaline reuptake inhibitors (SNRIs). Selective SNRIs share the proposed therapeutic mechanisms of tricyclics but with fewer and different side effects, including the serotoninergic effect of nausea, and for venlafaxine at high dose, risk of blood pressure elevation and cardiac arrhythmia, with toxicity in overdose. There is good efficacy evidence for venlafaxine, with possible superiority over SSRIs,[3] and emerging evidence for duloxetine.[4] A further SNRI in some countries is milnacipran.

Selective serotonin reuptake inhibitors (SSRIs). Meta-analyses of SSRIs[5] show efficacy comparable with tricyclics, but lower rates of side effects and discontinuation. There is conflicting evidence as to whether SSRIs are less effective than tricyclics and SNRIs in severe depression. SSRIs and clomipramine, the most serotoninergic tricyclic, are more effective than noradrenergic tricyclics in obsessive-compulsive disorder.

There have been vigorous debates regarding effects of antidepressants, particularly SSRIs, on suicidal behaviour, with many studies.[6] Actual suicide does not appear to be increased and may be reduced. There is some evidence that suicidal feelings and attempts may be increased in the early weeks after starting, particularly for SSRIs, with development of tension or agitation, and in children and adolescents. On the other hand SSRIs and other newer antidepressants are considerably safer in overdose than tricyclics and SNRI. Warnings were issued by regulatory authorities in the United States and the United Kingdom when the evidence started to emerge. Most strongly the findings argue for careful clinical surveillance of patients prescribed antidepressants, particularly in the early phases of treatment when risk of suicidality has long been recognized as increased. This is in any case important in depressed patients.

Noradrenaline reuptake inhibitors. Only one newer selective NARI is available, reboxetine. It is clearly superior to placebo.[7]

(iii) Monoamine oxidase inhibitors (MAOIs)

Irreversible MAOIs. There are few MAOIs available, reflecting hypertensive and other interactions and limited use. The older MAOIs bind irreversibly to the enzyme, and new enzyme needs to be synthesized over 1–2 weeks to reverse the effect. Controlled trials[8] show superiority to placebo in depression, and in anxiety disorders. Often high doses are necessary and the hydrazine MAOIs, phenelzine, and isocarboxazid, show better clinical response in slow acetylators. Tranylcypromine has additional stimulant effects, and has been viewed as more effective but with more risk of interactions.

From the late 1950s, there were suggestions of particular efficacy in atypical depression, variously regarded as non-endogenous depression, anxiety disorder with depression, or as a pattern of reversed vegetative symptoms with increased appetite, increased sleep, evening worsening, reactivity, and other features, a meaning currently predominant in the United States. Evidence is not strong, but comparative trials of phenelzine and tricyclics point to better effects than tricyclics with anxiety disorders and reversed vegetative symptoms.[8] On the other hand MAOIs are mostly used as second-line drugs by psychiatrists, where other antidepressants have failed, irrespective of clinical picture.

Reversible competitive inhibitors. The reversible MAO-A selective drug moclobemide can dissociate from the enzyme and be displaced by substances with higher affinity, including tyramine. It shows superiority to placebo,[9] but not at doses below 450 mg daily, and evidence is best for 600 mg. Many clinicians view moclobemide as less effective than older MAOIs.

(iv) Other antidepressants

Bupropion. Bupropion is relatively stimulant and is epileptogenic. It is licenced for depression in the United States, and in some other countries as an aid to smoking cessation.

Mianserin. Mianserin, an older drug which blocks alpha-2 autoreceptors, is sedative in side effects and carries a definite although low risk of agranulocytosis.

Mirtazepine. Mirtazepine blocks alpha-2, 5HT2, and 5HT3 receptors. It has been shown superior to placebo and is also a sedative.[10]

Trazodone. Trazodone, an older drug is relatively sedative carries a risk of priapism in males.

New classes of antidepressants, not yet licenced, are continually being sought.[11] Agomelatine, an agonist at melatonin MT1 and MT2 receptors and a 5HT2C antagonist has shown evidence of efficacy. Efforts have been made to develop drugs, which inhibit cortisol secretion. Development pathways for new antidepressants are long and failures due to weak treatment effects or major adverse effects are common.

(b) Electroconvulsive therapy (ECT)

ECT, the earliest of modern treatments, is still the most effective in severe depression. In a meta-analysis.[12] It has been found more effective than simulated ECT in blind trials, and more effective than pharmacotherapy. Bilateral ECT appears more effective than unilateral but this may not be true at adequate stimulus intensity. Best effects occur in psychotic depression with delusions or psychomotor retardation. There is also some evidence, which is not conclusive, that ECT may benefit mania.

Other physical treatments

(a) Bright light

Bright light, reviewed in Chapter 6.2.10.2 is the established treatment for seasonal affective disorder. Several studies in non-seasonal depressions as adjunctive treatment have suggested some benefit, although the effect may not be sustained.[13]

(b) Repetitive transcranial magnetic stimulation (TMS)

TMS is reviewed in Chapter 6.2.10.3. There is clear evidence of superiority of left prefrontal TMS compared with sham therapy but the degree of benefit appears weak. There is still uncertainty regarding optimal stimulation parameters. At present the evidence is not sufficient for widespread use of TMS in clinical practice.

(c) Vagus nerve stimulation

Stimulation of afferent left cervical vagus nerve fibres by a stimulator implanted in the chest wall is approved by regulatory authorities in a number of countries for resistant epilepsy and in the United States for resistant depression. Controlled evidence is still limited.[14]

(d) Deep brain stimulation

Chronic stimulation of white matter tracts adjacent to the subgenual cingulate region by implanted electrodes has been reported to produce striking remission in a small number of patients with resistant depression, associated with marked reduction in local cerebral blood flow.[15] Further experience is still needed.

Longer-term treatment

(a) Continuation treatment

In recent years, it has become apparent from follow-up studies that the long-term outcome of depression is still often problematic. It is customary to distinguish between relapse, or early symptom return, and later recurrence of a new episode.[16] In parallel, drug treatment after the acute episode has been divided into earlier continuation treatment, to prevent relapse, and longer-term maintenance treatment to prevent recurrence.

There have been many controlled trials of continuation treatment on active drug for 6 to 8 months after the acute episode against withdrawal to placebo, showing substantial benefit from continuation.[17] A fluoxetine study with staged withdrawal showed benefit of continuation for 24 and 38 weeks, but not 62 weeks, suggesting that routine continuation should be for 9–12 months rather than 6 months.[18] A controlled trial of early lithium withdrawal after augmentation showed very high relapse rate indicating a need for lithium continuation[19] and there is evidence for continuation drugs after ECT.

(b) Maintenance treatment

Longer-term studies in unipolar depression have shown clear benefit from maintenance antidepressants,[17] although recurrence rates in drug-treated patients may be high.

Withdrawal reactions requiring temporary restarting can occur if antidepressants are stopped abruptly, particularly after high doses for long periods, with malaise, coryza-like symptoms, vomiting, and diarrhoea. This may occur with a variety of antidepressants including SSRIs, tricyclics, and SNRIs.

Trials of lithium in maintenance treatment of severe recurrent unipolar depressives also show benefit over placebo.[20] Comparative efficacy against antidepressants is not clear.

(c) Acute treatments for bipolar disorder

In the treatment of bipolar depression there is a good evidence for efficacy of antidepressants.[21] The major risks are precipitation of mania and rapid cycling, which appear to occur more with tricyclics and venlafaxine than with other antidepressants, suggesting the preferential use of SSRIs, MAOIs, or ECT and covering with lithium. Lamotrigine has also been found effective.[22] For lithium alone as an antidepressant the evidence is less clear-cut.

In the treatment of acute mania, lithium has been extensively reviewed and the evidence is good.[23] Anticonvulsants have been increasingly evaluated, with good controlled trial evidence in mania for valproate and carbamazepine, but not lamotrigine or other anticonvulsants.[22] There is also controlled trial evidence for antipsychotics, both older and newer.[24] Available trials do not clearly point to choice of treatment.

(d) Maintenance treatment of bipolar disorder

Controlled trials of maintenance lithium against placebo in bipolar disorder show clear reduction of recurrences, particularly of mania, with weaker effects on depression.[25] There is also reduction of suicide in mood disorder more generally.

There are high rates of early recurrence, particularly of mania, when lithium is discontinued after long-term use, particularly if discontinuation is rapid.[26] This mandates slow withdrawal in practice. One study.[27] has indicated greater benefit on residual

symptoms for doses producing blood levels 0.8 to 1.0 mmol/l than for 0.4 to 0.6 mmol/l. This is important since follow-up studies of bipolar patients show that subsyndromal symptoms are common.

There have been fewer long-term studies of anticonvulsants.[22] For valproate evidence is suggestive rather than conclusive. For lamotrigine two controlled trials have confirmed prophylaxis of bipolar depression. Although carbamazepine has been used for longer, evidence from placebo-controlled trials is mainly for acute treatment of mania.

Several manufacturer-sponsored trials have indicated prophylactic effects of olanzapine in maintenance. Older antipsychotics have long been used as adjunctive treatment in bipolar maintenance.

Psychological treatments

There is high public demand for psychotherapies. Guidelines are less well developed and less robustly based than for pharmacotherapy. The concept of empirically supported psychotherapies (EST) assists definition of the evidence base.[28]

(a) Acute treatment of depression

(i) Psychological treatments: general issues

The therapy models can broadly be divided into 'process-orientated' therapies such as psychodynamic or supportive–expressive therapies, and 'outcome-orientated' therapies, such as cognitive behaviour therapies (CBT) and interpersonal therapy (IPT), which primarily focus on symptom reduction. The latter EST are protocol or guideline driven, with a manual describing the therapy in detail. In large-scale randomized controlled treatment trials (RCTs) they have been found superior to pill or psychological placebo-controlled treatments or efficacious as antidepressants or other established treatments.[28] The EST are usually brief (12–20 sessions), delivered by a trained therapist, establishing a working alliance and using an individualized case formulation to focus on specific 'here and now' problems, encouraging between session 'homework' tasks to enable development of new coping skills.

(ii) Meta-analyses

The two best-evaluated therapies are CBT (including behaviour therapies) and IPT, with many meta-analyses of a large number of RCTs. There are fewer studies of dynamic therapies. Early outcome studies and resultant meta-analyses were hampered by use of 'completer' samples rather than 'intent-to-treat' analyses, but found specific psychological therapies, particularly CBT, more effective than other verbal therapies or waiting-list controls. A meta-analysis of 29 carefully selected controlled trials,[29] using intent to treat analyses, found efficacies of individual CBT (response rate, 50 per cent), behavioural therapy (55 per cent), and IPT (52 per cent) in the treatment of the acute episode were not significantly different to pharmacotherapy (58 per cent). Brief dynamic psychotherapies, mainly group therapies, were less effective (35 per cent) and only marginally superior to waiting-list controls (30 per cent). CBT was less effective in a group format, behavioural therapy was equally effective, and IPT was more beneficial when a significant other took part in therapy.

A recent meta-analysis[30] found CBT and IPT equally effective in reducing depressive symptoms in heterogeneous groups of patients with depressive symptoms and syndromes, and both were superior to counselling, other therapies or control treatments in primary care. When compared in meta-analyses to an active treatment, however, CBT and IPT show only equivalent efficacy or marginal superiority. Almost all RCTs of psychological therapies have been in outpatient depressives, not above moderate severity and comparisons with pharmacotherapy do not represent severe or inpatient disorder.

(iii) Cognitive therapy

Earlier CBT studies suggested it was effective over a range of severity and endogenicity. However, in the three-centre NIMH study,[31] which sought to overcome earlier methodological flaws, a secondary analysis found CBT less effective than pharmacotherapy in the more severe disorders within the outpatient range studied. Another study found that individuals with mild or moderately severe depression did equally well with either 8 or 16 sessions, but those with severe depression showed better response with 16 sessions. A large RCT in subjects with severe major depressive disorder[32] found no difference overall between antidepressants and CBT, but also significantly greater effects for antidepressants at the trial centre where the CBT therapists were less experienced.

(iv) IPT

IPT Has been found effective in moderate numbers of RCTs, in a variety of situations, including combination studies with antidepressants and some situations where medication as a first-line treatment for depression is not easily feasible (pregnancy, postpartum, adolescence, and in a developing country).[33] The NIMH study[31] is the only one that has compared IPT with a pill-placebo control. Interpersonal therapy was nearest to the effectiveness of imipramine but a little weaker, and IPT was more beneficial than CBT in patients with more severe outpatient depression.

(v) Other psychotherapies

Fewer acute controlled trials of other therapies have been published. Two large-scale RCTs of individual behavioural therapy found respectively greater efficacy than insight-orientated therapy, which disappeared by 3 months follow-up, and no differences from several other treatments.

The only published RCTs on marital therapy mainly draw on behavioural approaches with behavioural marital therapy or CBT both more effective than waiting-list controls in one study and an advantage for behavioural marital therapy over CBT one study of depressed individuals with marital discord. Couples therapy was more effective than pharmacotherapy in a study of depressed patients living with a critical partner. An RCT of family therapy in inpatients with affective disorder[34] found benefit on role function for female patients on discharge, and for bipolars but not unipolars on follow-up.

Counselling is often offered alone or in combination with medication, particularly in primary care or non-specialist mental health settings. Recent meta-analyses[30] suggest limited benefits, with superiority over treatment as usual or waiting-list control conditions lost beyond about 3 months. Individuals who are socially isolated and lack a confidante may particularly benefit.

(vi) Combined psychotherapy and pharmacotherapy

Data on whether combination of psychotherapy and pharmacotherapy bestows additional benefit over either treatment alone are not conclusive, but some studies suggest a small advantage at the

level of severity usually studied, both for CBT and for IPT. However, in a meta-analysis of eight outcome studies, neither CBT, behavioural therapy, nor IPT added to antidepressants were any more effective than pharmacotherapy alone.[35]

In a large RCT of antidepressant and cognitive behavioural analysis system of psychotherapy (CBASP) given alone, or in combination for major depressive episode persisting for longer than 1 year[36] the group receiving the combined treatment had significantly greater overall reduction in depressive symptoms and attainment of remission as compared with single therapies.

In other studies, treatment with an antidepressant alone or in combination with IPT has been found more effective than IPT alone in dysthymics, and greater benefit of combined treatment has been reported in an RCT of group CBT and medication. These studies indicated that combined treatments improved quality of life over and above individual treatments. There is also some evidence that social adjustment is particularly benefited by IPT.

(b) Longer-term psychological treatment

(i) Continuation and maintenance trials

Psychological treatments are more expensive than antidepressants, but the balance could change if they reduced relapse rates.[37] Naturalistic follow-ups suggest this, but are hard to interpret because the original RCTs were not designed or powered to evaluate long-term outcome.

The use of continuation and maintenance psychotherapy is a relatively new concept. The key maintenance trial for IPT[38] assigned subjects with recurrent depression to imipramine, placebo, or monthly IPT with or without medication. There was a highly significant effect for antidepressants on recurrence rates and a modest effect (at the end of year 1 and year 3) for IPT. The first study of IPT, by its originators 15 years earlier, did not find it reduced relapse, compared with antidepressant continuation.

Five longer-term RCTs have now shown relapse and recurrence reduction with CBT.[37] Three of these, in individuals in remission using mindfulness or other techniques have only found benefit in subjects with a previous history of repeated episodes.

At least 20 per cent of people with an initial episode of major depressive disorder do not recover within 2 years, and those with residual depressive symptoms have a high risk of relapse. In a RCT[39] in 158 subjects with persistent residual depressive symptoms following major depression assigned to antidepressant continuation or antidepressant plus CBT, at 18-month follow-up relapse rates in the CBT plus antidepressant group were reduced by 45–50 per cent compared with the control group. Relapse prevention persisted for 3.5 years. Similar findings have been obtained in three small RCTs.[40]

(c) Bipolar disorder

Until very recently, research in psychological treatments for bipolar disorder was limited. Prior to 1995 a number of small studies, few of them randomized trials, suggested that adherence to medication was improved.[41] More recently good RCTs have been completed. The earlier of these found some benefit from psychological treatments in preventing relapse, with a hint that the briefer interventions are more effective in preventing mania than depression. A meta-analysis of relapse rates[42] found an odds ratio (OR) for relapse in the intervention as compared to the control group of 0.31. Three recent large efficacy trials used CBT, family therapy, or group psycho-education, respectively. A separate meta-analysis of these[42] found a similar odds ratio for relapse, with more benefit for depressive than manic/hypomanic relapses.

The MRC multicentre RCT in the United Kingdom[43] was one of the few pragmatic effectiveness trials, with over 250 subjects randomized to CBT plus usual psychiatric treatment or usual treatment alone. Over 50 per cent of the sample had a recurrence by 18 months, with no overall significant differences between groups, but a significant interaction with number of previous episodes such that CBT was significantly more effective than treatment as usual in subjects with fewer than 12 previous episodes, but less effective in those with more episodes.

One variant of IPT has been evaluated in a large RCT, interpersonal and social rhythm therapy (IPSRT), which includes an approach to stabilize social rhythms.[44] It was found to reduce bipolar relapse rates more than did IPT. The full advantages and disadvantages of different forms of psychotherapy for bipolar disorders are yet entirely clear, but further evidence is expected to emerge soon.

Management

Treatment of unipolar disorder

(a) General aspects: where to manage

The goals of depression treatment are to alleviate acute symptoms, to restore psychosocial functioning, and to prevent relapse and recurrence. Crucial decisions are the selection of an intervention and treatment setting. These involve four key issues: severity of the disorder including risk of self-harm, availability of treatments (specific antidepressants, trained therapists), patient preference, and nature of any associated difficulties. Recent guidelines include those of the UK National Institute for Health and Clinical Excellence[45] and of CANMAT.[46]

(b) Treatment setting

Very severe depression with definite suicidal risk is best managed in inpatient facilities to allow careful monitoring. With less suicidal risk and good social support severe cases can be managed with intensive community support, partial hospitalization, day care, or combined outpatient and home-based care. Moderate or mild depression should be managed in outpatient settings, unless treatment is complicated by severe physical illness, or non-response requires more detailed assessment.

(c) Primary or secondary care

Worldwide, most depressed patients are treated in primary care or general medical settings. Cases are referred to specialist mental health services because the disorder is more severe, chronic, treatment resistant, or because other difficulties, such as alcohol misuse or marital problems, complicate the clinical picture.

(d) Medication or psychological therapy

Although severe depression may respond to psychotherapy alone, recovery is slower than with drugs or ECT. Psychotherapy may be used in addition, rather than alone. In milder depressions, choice depends partly on patient preference and availability of psychotherapies, although when major depression criteria are reached antidepressants should preferably not be withheld. In the complex situations requiring specialist referral it is often necessary to use

combinations of drugs and psychosocial approaches. Although physical and psychological treatments are described separately in this section, treatment rarely involves only prescribing, but education and support of patient and family are important aspects of any clinical management.

Medication

(a) When to treat

The most important indications for use of medication are severity and persistence. A severity threshold has been shown a little below major depression at which tricyclic antidepressants start to show superiority over placebo.[47] SSRIs may show benefit in minor depression.[48] Antidepressants are also superior to placebo in dysthymia. In mild depressive episodes highly reactive to major stress, and in acute grief, prognosis for spontaneous resolution is often good, and medication may be delayed, provided that improvement is occurring. Impairment of function and suicidal feelings are other indications to treat. Recent guidelines recommend use in depression of moderate severity, major depression, or ICD-10 depressive episode.

(b) What antidepressant to choose

Since SSRIs are comparatively free of side effects and are not costly if out of patent, they are generally recommended as first choice antidepressants. Where there is a previous history of response to a specific drug or class, the best first choice is that antidepressant.

With few exceptions, symptom pattern is not a good guide to treatment. Effects of antidepressants extend widely across the spectrum of depression and in to anxiety disorders. The place of SSRIs in very severe depression is still debated. MAOIs or SSRIs or are a reasonable first choice in atypical depression. With comorbid obsessive-compulsive disorder the SSRIs, SNRIs, or clomipramine are preferable to noradrenergic antidepressants. Light therapy is indicated for seasonal affective disorder, alone or with antidepressant.

(c) How to use an antidepressant

There is a delay in clinical antidepressant effects of 1 to 3 weeks or longer, although some improvement may be seen earlier. An antidepressant needs to be continued for a minimum duration of 6 weeks at adequate or high dose before being regarded as ineffective.

Since side effects often occur more in the early weeks, with later tolerance, build-up of dose over 2 to 3 weeks is useful. For fluoxetine, the exceptionally long half-life of the active metabolite means that blood levels build-up for some weeks, even on a standard dose. Standard dosage regimes fail to allow for the considerable interindividual variability in pharmacokinetic mechanisms that occurs. Low doses are usually ineffective. Where response is not occurring and there are only minor side effects doses should be increased towards maximum, except in the elderly, who are vulnerable to poor elimination and to side effects.

Dose division during the day can be based on pharmacology. Half-lives of most antidepressants, combined with delay in therapeutic effects, mean that one dose per day may be adequate, but for most, two doses per day are better and three doses may be useful for some patients. Moclobemide, which is easily displaced from MAO and metabolized, should be given in three doses daily. For sedative antidepressants, administration of two-thirds of the dose at night may avoid hypnotics. Doses at bedtime of the more stimulant antidepressants, including SSRIs and older MAOIs, should be avoided.

Common side effects of SSRIs and SNRIs are nausea, and other gastrointestinal disturbances, insomnia, and sometimes tension and restlessness. Blood pressure should be monitored with high dose venlafaxine and it should not be used with hypertension or risk of ventricular arrhythmias. Common side effects of tricyclics clinically are sedation and anticholinergic side effects of dry mouth, blurred near vision, urinary retention, orthostatic hypotension, and confusion in the elderly. Dose-limiting side effects of MAOIs are hypotension, insomnia, and ankle oedema and tyramine containing foods and interacting drugs must be avoided.

(d) What to use if the first choice does not work

The evidence base for second choice of antidepressant after poor response to the first is weak. The large non-blind partially randomized STAR*D study,[49] has shown few differences between different choices after weak first response to citalopram.

The major pragmatic options are switch to a different antidepressant or combination. Switch should be to an antidepressant of a different class, or to a broad action SNRI. If this is not adequate, lithium augmentation is the best-supported change. Where depression is severe ECT is an alternative second choice.

(e) What to use in special situations

(i) Suicidal risk

Major suicidal risk requires consideration of lethality of particular antidepressants in overdose. Tricyclics and SNRIs should be avoided. SSRIs and most other newer antidepressants are comparatively safe in overdose, with careful monitoring for increased risk early in treatment.

(ii) Psychotic depression

Evidence as to choice of antidepressant is weak, but suggests SNRIs or tricyclics, rather than SSRIs. Combination with an antipsychotic may be useful, but antipsychotics are not adequate alone.[50] ECT is an alternative.

(iii) Medical illness

Antidepressant treatment with medical disorders is often difficult, because of toxicity due to illness, high blood levels from impaired metabolism, side effects, and drug interactions. Cardiac problems indicate use of an SSRI or other newer non-tricyclic non-SNRI antidepressant, and treatment after myocardial infarction may enhance survival. MAOIs lower blood pressure as a dose-related effect and should be avoided.

Concomitant epilepsy may be controlled by adjustment of anticonvulsant dose. Tricyclics and bupropion are epileptogenic and should be avoided. Epileptic potential is usually less with newer antidepressants but is often not clear, and the only antidepressants established not to be epileptogenic are older MAOIs.

(iv) The elderly

The elderly are particularly liable to side effects. SSRIs and newer antidepressants are preferable to tricyclics.

(v) Children

Antidepressants are not first choice treatments for depression in children and adolescents, for whom psychological therapy is preferable. Where the depression is not improving and is severe, an

antidepressant may be needed. Tricyclics are not effective.[51] There is more evidence for SSRIs, best for fluoxetine,[52] but the risk-benefit ratio is argued. For adolescents combination of SSRIs with CBT may be useful. The risk of increased suicidality early in treatment mandates careful observation. The FDA issued an Advisory in 2004 but not a contraindication.

(vi) Pregnancy and lactation

Tricyclics have been used in pregnancy for many years and do not carry risk of foetal malformation. For SSRIs the current evidence is contradictory and for some newer drugs evidence is lacking. Lithium and anticonvulsants carry some risk of foetal malformation and are contraindicated in the early months of pregnancy. Psychotropic drugs used at the time of delivery may produce complications of anaesthesia and foetal sedation and if possible should be withdrawn temporarily. Most psychotropic drugs appear in breast milk in small quantities. Breast feeding should be discussed with the patient.

Electroconvulsive therapy

Use of ECT varies internationally. Many psychiatrists use it as a first choice treatment in severe depression with psychomotor retardation or mood-congruent depressive delusions, or where an antidepressant has failed, and for moderately severe depressions which have not responded to one or two courses of antidepressant.

The UK National Institute for Clinical Excellence has recommended use only to achieve rapid improvement where a trial of other treatment has failed or the condition is potentially life threatening, in severe depression, catatonia, and prolonged or severe mania.[53] Detailed recommendations on how ECT should be administered also can be consulted.[54]

Non-response and resistant depression

Controlled trials of treatments for depressives who do not respond to first or second treatments are limited. There is good evidence for lithium augmentation of antidepressants[55] and limited evidence for potentiation by triiodothyronine (T3).[56] A variety of other augmenters have been tried including tryptophan, pindolol, buspirone, and combinations of antidepressants such as SSRI-bupropion, SNRI-mirtazepine, tricyclic-MAOI, with weak evidence.

We therefore depend on clinical experience.[57] If there is still limited response after several antidepressant treatments, a systematic approach to resistant depression should be adopted:

1 Reassess the situation thoroughly, with full history, assessment, and laboratory investigation of thyroid function to ask the following questions: (a) Is the diagnosis correct? Wrong diagnosis is in practice unusual. (b) Are there perpetuating factors in personality, family environment, or the social setting? Commonly, where depression has been long-term, secondary role loss (including work), and family adaptations to a non-functioning member mean there are no roles or relationships for the patient to return to and remission cannot occur, or is transient, unless psychotherapy, family therapy, and rehabilitation are employed to change the situation. (c) Is hypothyroidism impairing response? This may develop as a result of earlier use of lithium.

2 Consider previous treatment. Failure to use high or maximal doses of antidepressants when response is not occurring can lead to apparently resistant depression.

3 Employ drug and other physical treatments.

4 As remission occurs, introduce psychotherapeutic, cognitive, and rehabilitative interventions.

The treatment decisions depend on what has been used before. Start with the most promising antidepressant suggested by the history and push to a very high dose. If there is no good clue venlafaxine may be useful, with monitoring of blood pressure for elevation as high doses are reached.

The next intervention is augmentation. Lithium is easiest and best. Blood levels required for augmentation are not established. If response occurs, the lithium should be continued for some months, although not as long as antidepressant. Care should be taken to monitor for a serotonin syndrome (excitability, restlessness, temperature elevation), with serotoninergic antidepressants.

T3 may produce fewer side effects than lithium but evidence is weaker and care must be exercised if there are cardiac problems. L-tryptophan may potentiate MAOIs and the more serotoninergic uptake inhibitors. Its effect is weak, but may be combined with lithium. Pindolol, which blocks 5-HT_{1A} autoreceptors, can accelerate speed of antidepressant response but increased amount of response is less clear.

At one time tricyclic-MAOI combinations were often used and occasionally still have a place, combined with lithium. SSRIs, SNRIs, and clomipramine risk of serotonin syndrome and are contraindicated. Doses should be increased gradually. More ordinarily 1 to 2 weeks should intervene between stopping an older MAOI and starting a reuptake inhibitor.

When vigorous medication regimens have still not produced a response, it is often helpful to add bilateral ECT to maximal drug therapy, even when ECT alone has not previously helped. The practical places for vagus nerve stimulation and deep brain stimulation are not yet clear. In intractable severe chronic depression, psychosurgery still has an occasional place, followed by active rehabilitation.

Longer-term treatment

(a) Continuation treatment: a routine

Continuation antidepressant for 9–12 months should be routine following response to acute treatment. Antidepressant or lithium are also advisable after ECT. The continuation dose should initially be the same as the acute treatment dose. After 2 to 3 months this may be reduced if side effects are a problem, but only by a small amount, to avoid symptom return.

Before the antidepressant is withdrawn the patient should also be completely free of residual symptoms for at least 4 months. These symptoms or history of previous relapse or recurrence suggest longer continuation. Withdrawal should then be carried out slowly, over 2 to 3 months, to minimize the risk of relapse, and of withdrawal symptoms.

In some patients, after withdrawal, or achievement of a low dose, depressive symptoms return. Full dose should be resumed, followed by continuation for a further period of 9 months to a year. Some of these patients relapse again on later drug withdrawal, and long-term maintenance should then be considered.

(b) Maintenance treatment

Longer-term maintenance is indicated where there have been several recurrences, such as two episodes in the last 2 years, or three

episodes in 5 years.[16,45] This also depends on the severity of episodes, their potential impact on personal life, family life, and career, and the patient's preferences. Discussion is required, to reach a joint decision.

For most unipolar depressives the maintenance treatment choice will be whichever antidepressant has been effective and well tolerated. Where antidepressants have not been fully effective, lithium or combination are required. Antidepressant doses should be the same or a little lower than those needed for acute treatment in the particular patient. The length of maintenance is harder to specify, and it depends on the history. Two to three years will be a minimum, longer where remission has been difficult to achieve or recent episodes more frequent, 5 years or more where risk is greater, and indefinite or lifelong where two or three attempts to withdraw have been followed by another episode within a year.[16] Withdrawal of antidepressants, lithium, or other drugs after long-term maintenance should always be gradual. Where withdrawal is followed soon after by a recurrence, longer-term maintenance is indicated, and where this sequence has been repeated, lifelong treatment.

Psychological treatments

The psychological management of acute depression ranges across basic clinical management, psychoeducation for the individual and partner or family, supportive psychotherapy, to formal psychotherapy.

(a) Clinical management

Clinical management is a key component of the care of all individuals with depression regardless of the specific treatment. It includes education about the nature of the disorder, its polarity, course and prognosis, treatment options, their advantages and disadvantages, delay in benefits, probable side effects and how to manage these, and planned length of treatment. This dialogue can help build a strong treatment alliance, overcome misconceptions such as fears of addiction to antidepressants, and improve medication adherence. Involving 'significant others' also in some sessions facilitates information and engagement in the treatment process, clarifies what aspects of care are the clinician's responsibility, and can help reduce tension in interpersonal relationships resulting from the patient's depression. For many depressed individuals the reduction in symptoms from this basic approach and medication restores the previous level of functioning. They are again able to use their coping skills to resolve personal problems and no other form of psychological input is needed.

(b) When to use psychological therapies

If additional strategies are needed, a number of factors should guide their choice, including treatment setting, severity, chronicity, complexity, patient preferences, and availability of therapists. Most management approaches are multifaceted and in a dynamic state so the divisions below, structured by severity of depression, although useful may be somewhat artificial.

(i) Mild depression

For individuals with milder depression, usually treated in primary care, psychological approaches can be used alone as alternatives to medication, and there is little advantage in combined treatment. Many patients express their own treatment preferences. Some primary care physicians offer extended treatment sessions for depression, but many utilize counsellors. A course of counselling, lasting 8–10 sessions, is particularly useful to patients who lack a confidante. Benefits appear in the first 3 months, but there is no evidence of longer-term benefits compared to other treatments or 'watchful waiting'[45] (regular appointments with a clinician offering monitoring and support without medication or other interventions).

The next option may be CBT or IPT. There are no robust predictors of differential response to these as opposed to medication so choice often depends on availability of a suitable therapist. Difficulties in timely access have led to development of computerized CBT and of self-help approaches. The individual is offered a computer version of therapy or a written manual to guide self-therapy (bibliotherapy), with possible 1–3 sessions from a therapist to explain the process and overcome hitches. In other circumstances, including management of depressed pregnant women, telephone versions of IPT have been developed. There is not yet consensus on the efficacy of these alternatives but a trial may be used in mild depression.

(ii) Moderate depression

Formal psychotherapy may be offered to individuals with moderate depression as the only treatment, or preferably combined with medication. Usually this comprises 15–20 sessions of individual therapy, but since more than 20 per cent of couples report marital discord in association with depression, marital or family approaches should also be considered. Psychological treatments may also be required where medical conditions or needed medications contraindicate antidepressants, or the patient refuses medication. Other important considerations include previous coping skills, premorbid stressors requiring problem-solving approaches, ability to articulate emotions and difficulties. If there are doubts regarding suitability, referral may be offered for assessment. Patient preference for therapy does not imply benefit, and dropout rates are about similar to those for medication. Also consider whether needs can be met through a time-limited input or if longer-term support is preferable through day services or a community support worker. Therapists adherence to the therapy model, their level of expertise and skill and the provision of regular and adequate supervision to the practitioner are other factors and may account for some 30 per cent of the variance in patient's improvement.[58]

(iii) Severe depression or complex presentations

Research evidence suggests that CBT or IPT may be used alone in severe depression if delivered by an expert therapist. However improvement is slower than with antidepressants and in practice these are usually combined with medication, which is commenced first. In severe depressions it may be difficult for an individual to concentrate until the mental state has been stabilized by medication. Patients who fail to respond adequately to antidepressants alone require reassessment. In many instances medication has improved the vegetative symptoms of depression, but the individual manifests psychological vulnerabilities, such as Axis II comorbidity or long-standing, low self-esteem, or social stressors, acting as maintenance factors and amenable to therapy. The evidence suggests that IPT or CBT are the most useful therapies to combine with antidepressants in unipolar disorders, but the more severe or complex the case, the more important therapist factors become.[58] More severe or chronic disorders also require more prolonged therapy.

Continuation and maintenance therapy

Some patients may benefit from regular but less frequent sessions of continuation or maintenance CBT or IPT to prevent relapse, with about eight additional sessions over the course of a year, to explore whether the techniques learnt in therapy are being used well and to identify potential triggers of relapse or recurrence. It is unrealistic to offer this to all those who receive an acute course of IPT or CBT, but it may be considered for those at high risk of relapse or recurrence due to residual symptoms, history of highly recurrent episodes, ongoing severe psychosocial stressors, refusal to take continuation medications. A full course of CBT may also be offered now, rather than during the acute illness, if remission is partial with residual symptoms.

Treatment of bipolar disorder

Recent guidelines on treatment of bipolar disorder, providing supporting evidence, include those from the UK National Institute for Health and Clinical Excellence,[59] the CANMAT consensus statement,[60] and the APA Practice Guideline.[61]

(a) Treating acute mania

Most mood stabilizers and antipsychotics have acute antimanic effects, but lithium or valproate are the drugs of first choice. Since response may take 7 to 14 days, an antipsychotic is also often required, with effects within days in controlling the acute manic symptoms. The older antipsychotic, haloperidol should not be used with lithium due to risk of neurotoxicity syndrome. Benzodiazepines such as lorazepam and clonazepam may also be used temporarily to treat hyperactivity, insomnia, and agitation. The loss of insight, impaired judgement, hyperactivity, and disinhibited behaviour of manic patients often mean that hospitalization is required to ensure patient safety and commence treatment. For severe mania, ECT may occasionally be used.

(b) Treating bipolar depression

Treatment of bipolar depression is in principle similar to unipolar, but with mood stabilizer cover. If severity indicates antidepressant use, an SSRI, bupropion, or MAOI should be used, but not a tricyclic or SNRI, because of higher risk of mania and cycling. A mood stabilizer should be used in addition. Lamotrigine may produce benefit alone. If cycling occurs in spite of this, an antipsychotic may be added, and the antidepressant cautiously reduced. ECT may be indicated, as in unipolar depression, and does not appear to lead to rapid cycling.

(c) Continuation and maintenance treatment for bipolar disorder

Continuation treatment is needed after acute bipolar episodes, using the single medication or combination that has been used in acute treatment. Bipolar disorder is more recurrent than unipolar disorder, and mania can have catastrophic effects on personal life. The threshold for maintenance treatment is therefore lower and it should be used for 1 to 2 years following a first episode of mania and longer after bipolar recurrences.

Choice of mood stabilizer includes an increasingly wide range. Probably lithium is still the first choice, in the absence of major side effects. Doses should achieve blood levels 12 h after the last dose of 0.4 to 0.8 mmol/l, but sometimes higher levels are needed to prevent recurrences, and to reduce residual symptoms. After some months blood levels are often very stable and only require occasional monitoring. Advice needs to be given on circumstances, which may disturb levels such as dehydration, gastrointestinal upset, travel, hot climates. Thyroid function should be monitored for hypothyroidism every 6 to 12 months.

Poor or partial responses are not uncommon, and valproate is a first choice alternative. Lamotrigine may be helpful where depression is a main feature. In treatment-resistant cases, two mood stabilizers may be required, particularly lithium and valproate. Antipsychotics, particularly olanzapine, may also be added or substituted if response is poor.

Antidepressants can often be withdrawn slowly after mood stabilizers are well established, but may be necessary to prevent chronic or recurrent depression. Antidepressants should not usually be used alone for maintenance in view of the risk of mania. A few patients require multiple combinations of mood stabilizers, an antipsychotic, and an antidepressant.

Length of maintenance treatment varies, from a few years, to lifelong where several recurrences have followed withdrawal. Withdrawal of lithium should always be slow, over 3 to 6 months, to avoid rebound mania. It is best to do the same for the anticonvulsants.

(d) Rapid cycling

Treatment of established rapid cycling may be difficult. It is important to avoid a cycling alternation of antidepressant and antimanic medication, following too late after mood changes. The patient should be established on mood stabilizers, usually in high dose and often more than one in combination. Antipsychotic may be added. Antidepressants are better avoided but may be needed in addition, used cautiously and with slow dose increase and an effort to establish a constant dose. Thyroid function should be checked initially and regularly.

(e) Pregnancy

Management in pregnancy is complicated by the risk of foetal malformations with lithium and anticonvulsants. These should be withdrawn for the first trimester. Antipsychotics and antidepressants may be used if necessary. Olanzapine may be appropriate but safety of newer antipsychotics and antidepressants is less well established than for older ones. In later pregnancy lithium and anticonvulsant mood stabilizers may need to be reinstituted. Medication should be withdrawn at the time of delivery to avoid effects on the foetus, but started again as soon as possible thereafter, as risks of relapse are increased postpartum.

Psychological treatments

(a) Acute treatment of mania

Psychotherapy has not been evaluated in manic patients and the evidence for benefit in bipolar disorders is for relapse reduction when delivered during euthymia. Access to therapy in most countries is limited and it is important to use clinical management, supportive therapy, and psychoeducational sessions as part of general treatment. Use should be made of systematic care packages to address psychosocial problems, to facilitate recognition of early warning symptoms of relapse, and to implement self-monitoring and self-management strategies.[59] Support should also include education about the disorder and treatment options, and help in

adjusting to the diagnosis and should, if possible, include the patient's 'significant others'. About six sessions of family therapy may have beneficial effects extending beyond discharge.

Therapy with individuals with bipolar disorders can be challenging, requiring skilled and experienced therapists who also understand aetiology and pharmacology. There is a trend towards more cost-efficient group approaches, which allow individuals to observe the strengths and weaknesses of coping strategies employed by peers. However, a key study of group psychoeducation found benefits only with a structured group with a coherent treatment plan and not with unstructured groups without expert professional leadership.

The therapies currently found effective to treat acute bipolar depressive episodes or to reduce the risk of future relapse in RCTs do not differ markedly in either their overall efficacy or their core content. They incorporate four key components: psychoeducation and adjustment to the disorder, lifestyle regularity and 'harm reduction' (reducing substance misuse), enhancing medication adherence, and developing self-management skills to reduce the risk of relapse by recognizing potential triggers or prodromal signs or symptoms.

There are no consistent predictors of therapy response. There is some consensus that patients with Axis I or II comorbidity, particularly substance misuse or borderline personality disorder will benefit from an extended course of therapy. Likewise, where there have been many previous episodes of bipolar disorder, patients are unlikely to be able to assimilate and implement all the problem-solving and coping strategies being offered, in a time-limited therapy. Even where therapy does not clearly reduce relapse, many recipients value the opportunity to discuss their problems in detail, reporting improved quality of life and social functioning. Clinically, if the goal is relapse prevention, it may be appropriate to particularly target for therapy those individuals at an earlier stage of their 'bipolar career', whom evidence suggests may benefit more.[43]

Longer-term treatment of bipolar disorders

Supportive or other psychological therapies may extend over considerable periods of time. People with affective disorders also form 15 to 20 per cent of the long-stay patient populations of United Kingdom and United States mental hospitals. Rehabilitation techniques, as applied to schizophrenia and other severe mental disorders, are important for instilling hope and developing day-to-day living skills in people with chronic or recurrent affective disorders. Lastly, individuals with chronic health problems are at high risk of non-adherence with medication. There is an important role for psychotherapy in enhancing the acceptance of long-term medication[23,62] and there is emerging evidence that psychological interventions targeted directly or indirectly at enhancing medication adherence can have a beneficial impact on outcome in bipolar disorders.

Further information

Beck, A.T. (2005). The current state of cognitive therapy: a 40-year retrospective. *Archives of General Psychiatry*, **62**, 953–9.

Canadian Psychiatric Association and the CANMAT Depression Work Group. (2001). Clinical practice guidelines for the treatment of depressive disorders. *Canadian Journal of Psychiatry*, **46**(Suppl. 1), 5S–90S.

NICE. (2004). *Depression. Management of depression in primary and secondary care*. NICE Clinical Guideline 23 National Institute for Clinical Excellence, London.

NICE. (2006). *Bipolar disorder. The management of bipolar disorder in adults, children and adolescents, in primary and secondary care*. NICE Clinical Guideline 38 National Institute for Health and Clinical Excellence, London.

Stein, D.J., Lerer, B., and Stahl, M. (2005). *Evidence based psychopharmacology*. Cambridge University Press, Cambridge.

Weissman, M.M., Markowitz, J.C., and Klerman, G.L. (2000). *Comprehensive guide to interpersonal psychotherapy.* Basic Books, New York.

References

1. De Lima, M.S., Hotopf, M., and Wessely, S. (1999). The efficacy of drug treatments for dysthymia: a systematic review and meta-analysis. *Psychological Medicine*, **29**, 1273–89.
2. Wilson, K., Mottram, P., Sivanranthan, A., *et al.* (2001). Antidepressants versus placebo for the depressed elderly. *Cochrane Database of Systematic Reviews*, (1). CD000561. DOI: 1O.1002/14651858.CD000561.
3. Smith, D., Dempster, C., Glanville, J., *et al.* (2002). Efficacy and tolerability of venlafaxine compared with selective serotonin reuptake inhibitors and other antidepressants: a meta-analysis. *The British Journal of Psychiatry*, **180**, 396–404.
4. Hudson, L.L., Wohlreich, M.M., Kajdasz, D.K., *et al.* (2005). Safety and tolerability of duloxetine in the treatment of major depressive disorder: analysis of pooled data from eight placebo-controlled clinical trials. *Human Psychopharmacology*, **20**, 327–41.
5. MacGillivray, S., Arroll, B., Hatcher, S., *et al.* (2003). Efficacy and tolerability of selective serotonin reuptake inhibitors compared with tricyclic antidepressants in depression treated in primary care: systematic review and meta-analysis. *British Medical Journal*, **326**, 1014.
6. Goldney, R.D. (2006). Suicide and antidepressants: what is the evidence? *The Australian and New Zealand Journal of Psychiatry*, **40**, 381–5.
7. Schatzberg, A.F. (2000). Clinical efficacy of reboxetine in major depression. *The Journal of Clinical Psychiatry*, **61**(Suppl. 10), 31–8.
8. Paykel, E.S. (1990). Monoamine oxidase inhibitors: when should they be used. In *Dilemmas and difficulties in the management of psychiatric patients* (eds. K. Hawton and P. Cowen), pp. 17–30. Oxford University Press, Oxford.
9. Paykel, E.S. (1995). Clinical efficacy of reversible and selective inhibitors of monoamine oxidase A in major depression. *Acta Psychiatrica Scandinavica*, **91**(Suppl. 386), 22–7.
10. Anttila, S.A. and Leinonen, E.V. (2001). A review of the pharmacological and clinical profile of mirtazapine. *CNS Drug Reviews*, **7**, 249–64.
11. Norman, T.R. (2006). Prospects for the treatment of depression. *The Australian and New Zealand Journal of Psychiatry*, **40**, 394–401.
12. The UK ECT Review Group. (2003). Efficacy and safety of electroconvulsive therapy in depressive disorders: a systematic review and meta-analysis. *Lancet*, **361**, 799–808.
13. Tuunainen, A., Kripke, D.F., and Endo, T. (2004). Light therapy for non-seasonal depression. *Cochrane Database of Systematic Reviews*, (2). CD004050. DOI: 1O.1002/14651858.CD004050.pub2.
14. Nemeroff, C.B., Mayberg, H.S., Krahl, S.E., *et al.* (2006). VNS therapy in treatment-resistant depression: clinical evidence and putative neurobiological mechanisms. *Neuropsychopharmacology*, **31**, 1345–55.
15. Mayberg, H.S., Lozano, A.M., Voon, V., *et al.* (2005). Deep brain stimulation for treatment-resistant depression. *Neuron*, **45**, 651–60.
16. Paykel, E.S. (2001). Continuation and maintenance therapy in depression. *British Medical Bulletin*, **57**, 145–59.
17. Geddes, J.R., Carney, S.M., Davies, C., *et al.* (2003). Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. *Lancet*, **361**, 653–61.
18. Reimherr, F.W., Amsterdam, I.D., Quitkin, F.M., *et al.* (1998). Optimal length of continuation therapy in depression: a prospective assessment during long-term fluoxetine treatment. *The American Journal of Psychiatry*, **155**, 1247–53.

19. Bauer, M., Bschor, T., Kunz, D., *et al.* (2000). Double-blind, placebo-controlled trial of the use of lithium to augment antidepressant medication in continuation treatment of unipolar major depression. *The American Journal of Psychiatry*, **157**, 1429–35.
20. Cipriani, A., Smith, K., Burgess, S., *et al.* (2006). Lithium versus antidepressants in the long-term treatment of urupolar affective disorder. *Cochrane Database of Systematic Reviews*, (4). CD003492. DOl: 10.1002/14651858. CD003492. pub2.
21. Gijsman, H.J., Geddes, J.R., Rendell, J.M., *et al.* (2004). Antidepressants for bipolar depression: a systematic review of randomized, controlled trials. *The American Journal of Psychiatry*, **161**, 1537–47.
22. Bowden, C.L. and Karren, N.U. (2006). Anticonvulsants in bipolar disorder. *The Australian and New Zealand Journal of Psychiatry*, **40**, 386–93.
23. Goodwin, F.K. and Jamison, K.R. (2007). *Manic depressive illness. Bipolar Disorders and Recurrent Depression* (2nd edition). Oxford University Press, New York.
24. Jones, R.M., Thompson, C., and Bitter, I. (2006). A systematic review of the efficacy and safety of second generation antipsychotics in the treatment of mania. *European Psychiatry*, **21**, 1–9.
25. Geddes, J.R., Burgess, S., Hawton, K., *et al.* (2004). Long-term lithium therapy for bipolar disorder: systematic review and meta-analysis of randomized controlled trials. *The American Journal of Psychiatry*, **161**, 217–22.
26. Baldessarini, R.J., Tondo, L., Floris, G., *et al.* (1997). Reduced morbidity after gradual discontinuation of lithium treatment for bipolar I and II disorders: a replication study. *The American Journal of Psychiatry*, **154**, 551–3.
27. Keller, M.B., Lavori, P.W., Kane, I.M., *et al.* (1992). Subsyndromal symptoms in bipolar disorder: a comparison of standard andlow serum levels of lithium. *Archives of General Psychiatry*, **49**, 371–6.
28. Chambless, D.L. and Hollon, S.D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, **66**, 7–18.
29. Agency for Health Care Policy and Research. (1993). Depression Guideline Panel. *Depression in primary care: treatment of major depression*, Vol. **2**. Clinical Practice Guideline Number 5. USDHHS, Public Health Service, Rockville, MD.
30. Churchill, R., Hunot, V., Comey, R., *et al.* (2001). A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological therapies in depression. *Health Technology Assessment*, **5**, No 35. Southampton: HTA.
31. Elkin, I., Shea, M., Watkins, J., *et al.* (1992). National Institute of Mental Health treatment of depression collaborative treatment programme. *Archives of General Psychiatry*, **46**, 971–82.
32. DeRubeis, R., Hollon, S., Amsterdam, J., *et al.* (2005). Cognitive therapy v medication in the treatment of moderate to severe depression. *Archives of General Psychiatry*, **62**, 409–16.
33. Weissman, M. (2007). Recent non-medication trials of interpersonal psychotherapy for depression. *International Journal of Neuropsychopharmacology*, **10**, 117–22.
34. Clarkin, J, Glick, I, Haas, G., *et al.* (1990). A randomized clinical trial of inpatient family intervention: results for affective disorder. *Journal of Affective Disorder*, **18**, 17–28.
35. Depression Guideline Panel. (1993). *Clinical practice guideline No 5: depression in primary care. Treatment of major depression*, Vol 2, pp. 71–123. US Department of Health and Human Services, AHCPR Publications, Rockville, MD.
36. Keller, M.B., McCullough, J.P., Klein, D.N., *et al.* (2000). A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. *The New England Journal of Medicine*, **342**, 1462–70.
37. Paykel, E.S. (2007). Cognitive therapy in relapse prevention in depression. *The International Journal of Neuropsychopharmacology*, **10**, 131–6.
38. Frank, E., Kupfer, D., Perel, J., *et al.* (1990). Three year outcomes for maintenance therapies in recurrent depressions. *Archives of General Psychiatry*, **47**, 1093–9.
39. Paykel, E.S., Scott, J., Cornwall, P.L., *et al.* (2005). Duration of relapse prevention after cognitive therapy in residual depression: a follow-up of controlled trial. *Psychological Medicine*, **35**, 59–68.
40. Fava, G.A., Ruini, C., Rafanelli, C., *et al.* (2004). Six-year outcome of cognitive behavior therapy for prevention of recurrent depression. *The American Journal of Psychiatry*, **161**, 1872–6.
41. Scott, J. and Gutierrez, M.J. (2004). The current status of psychological treatments for bipolar disorders: a systematic review of relapse prevention. *Bipolar Disorders*, **6**, 498–503.
42. Scott, J., Colom, F., and Vieta, E. (2006). A meta-analysis of relapse rates with adjunctive psychological therapies compared to usual psychiatric treatment for bipolar disorders. *The International Journal of Neuropsychopharmacology*, **9**, 1–7.
43. Scott, J., Paykel, E., Morriss, R., *et al.* (2006). Cognitive behavioural therapy for severe and recurrent bipolar disorders: randomised controlled trial. *The British Journal of Psychiatry*, **188**, 313–20.
44. Frank, E., Kupfer, D.J., Thase, M.E., *et al.* (2005). Two-year outcomes for interpersonal and social rhythm therapy in individuals with bipolar I disorder. *Archives of General Psychiatry*, **62**, 996–1004.
45. NICE. (2004). *Depression. Management of depression in primary and secondary care*. NICE Clinical Guideline 23 National Institute for Clinical Excellence, London.
46. Canadian Psychiatric Association and Canadian Network for Mood and Anxiety Treatments (CANMAT). (2001). Clinical guidelines for the treatment of depressive disorders. *Canadian Journal of Psychiatry*, **46**(Suppl. 1), 5S–90S.
47. Paykel, E.S., Hollyman, J.A., Freeling, P., *et al.* (1988). Predictors of therapeutic benefit from amitriptyline in mild depression: a general practice placebo-controlled trial. *Journal of Affective Disorders*, **14**, 83–95.
48. Judd, L.L., Rapaport, M.H., Yonkers, K.A., *et al.* (2004). Randomized, placebo-controlled trial of fluoxetine for acute treatment of minor depressive disorder. *The American Journal of Psychiatry*,**161**, 1864–71.
49. Valenstein, M. (2006). Keeping our eyes on STAR*D. *The American Journal of Psychiatry*, **163**, 1484–6.
50. Wijkstra, J., Lijmer, J., Balk, F.J., *et al.* (2006). Pharmacological treatment for unipolar psychotic depression: systematic review and meta-analysis. *The British Journal of Psychiatry*, **188**, 410–5.
51. Hazell, P., O'Connell, D., Heathcote, D., *et al.* (2002). Tricyclic drugs for depression in children and adolescents. *Cochrane Database of Systematic* Reviews, (2). CD002317. DOl: 1O.1002/14651858. CD002317.
52. Whittington, C.J., Kendall, T., Fonagy, P., *et al.* (2004). Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. *Lancet*, **363**, 1341–5.
53. NICE. (2003). *Guidance on the use of electroconvulsive therapy*. Technology Appraisal 59 National Institute for Clinical Excellence, London.
54. Royal College of Psychiatrists. (1995). *Council report CR39: the ECT handbook. The second report of the Royal College of Psychiatrists special committee on ECT*. Royal College of Psychiatrists, London.
55. Bauer, M. and Dopfmer, S. (1999). Lithium augmentation in treatment-resistant depression: meta-analysis of placebo-controlled studies. *Journal of Clinical Psychopharmacology*, **19**, 427–34**.**
56. Aronson, R., Offman, H.J., Joffe, R.T., *et al.* (1996). Triiodothyronine augmentation in the treatment of refractory depression. A meta-analysis. *Archives of General Psychiatry*, **53**, 842–8.
57. Kennedy, N. and Paykel, E.S. (2004). Treatment and response in refractory depression: results from a specialist affective disorders service. *Journal of Affective Disorders*, **81**, 49–53.
58. Roth, T. and Fonagy, P. (1996). *What works for whom?* pp. 57–103. Guilford Press, London.
59. NICE. (2006). *Bipolar disorder. The management of bipolar disorder in adult: children and adolescents, in primary and secondary care.* NICE Clinical Guideline 38 National Institute for Health and Clinical Excellence, London.

60. Yatham, L.N., Kennedy, S.H., O'Donovan, C., *et al.* (2005). Canadian network for mood and anxiety treatments. Canadian Network for Mood and Anxiety Treatments (CANMAT) guidelines for the management of patients with bipolar disorder: consensus and controversies. *Bipolar Disorders*, **7**(Suppl. 3), 5–69.
61. American Psychiatric Association. (2002). *Practice guideline for the treatment of patients with bipolar disorder* (2nd edn). American Psychiatric Press, Washington.
62. Tacchi, M.J. and Scott, J. (2005). *Medication adherence in schizophrenia and bipolar disorders*, p. 110. John Wiley & Sons, London.

4.5.9 Dysthymia, cyclothymia, and hyperthymia

Hagop S. Akiskal

Subthreshold affective conditions, personality, and temperament

Long before psychiatry moved to the outpatient arena in the latter part of the twentieth century, psychiatrists had observed milder mood disturbances among the kin of patients hospitalized for endogenous or psychotic depressions or mania. Some were described as sullen, morose, or otherwise moody, without discrete episodes; others reported self-limited episodes, but often went untreated. With the advent of modern treatments, practitioners are being increasingly consulted by patients presenting with attenuated affective disturbances. Although the relationship of these ambulatory mood states and more classical severe affective disorders has not been resolved, there is emerging sleep electroencephalography (**EEG**) and familial-genetic evidence[1–3] that a continuum exists between them. Along the same lines, studies conducted in the United States and Germany[4,5] into what were once described as 'neurotic' depressions have revealed a progression to more endogenous, psychotic, or bipolar switching. For these and related reasons, current official classification systems such as the ICD-10 and DSM-IV, have dropped the neurotic-endogenous dichotomy. Sceptics would perhaps argue that the new categorization of depressive disorders into dysthymic and major subtypes is not much of an improvement. Nonetheless, the new terminology has drawn attention to a large universe of human suffering that had been neglected in the past, and the conceptualization of dysthymia as a variant of mood disorder has had a far-reaching impact on diagnostic and therapeutic habits of clinicians worldwide.[6] The emerging concept of the bipolar spectrum, which does include manic, cyclic depressive (bipolar II), cyclothymic, hyperthymic and related conditions, is beginning to have a similar impact on practice.[7]

The subthreshold mood disorders are not only in continuum with more pathological mood states, but they also provide a bridge with normal affective conditions. In this context, temperament, as a construct encompassing affective personalities, is currently enjoying a renaissance as one of the possible substrates for the origin of mood disorders. Temperament classically refers to an adaptive mixture of traits which, in the extreme, can lead to illness or modify the expression of superimposed affective states. The subthreshold conditions covered in this chapter represent the extreme expressions of these temperaments. A new self-administered instrument, the TEMPS-A,[8] now validated in 10 language versions, is being used internationally to measure the classical constructs of depressive, cyclothymic, hyperthymic, and irritable, as well as anxious temperaments.

In the current literature, various terms such as 'minor affective states', 'intermittent depression', 'hysteroid dysphoria', and 'atypical depression' are often used for subthreshold disorders.[9] These terms are avoided here, because in contemporary practice these conditions are at least as 'typical' as major mood disorders: their impact on the sufferer is not time-limited, nor minor, and involves more than a state of demoralization and moral foible. The following passage from Sir Aubrey Lewis[10] is *à propos*:

> . . . Severe emotional upsets ordinarily tend to subside, but mild emotional states . . . tend to persist, as it were, autonomously. Hence the paradox that a gross blatant psychosis may do less damage in the long run than some meager neurotic incubus: a dramatic attack of mania or melancholia, with delusions, wasting, hallucinations, wild excitement may have far less effect on the course of man's life than some deceptively mild affective illness which goes on so long that it becomes inveterate. The former comes as a catastrophe and when it has passed the patient takes up his life again . . . while with the latter he may never get rid of his burden.

It is a curious fact that most subthreshold affective conditions, while symptomatologically attenuated, tend to pursue a chronic course. This raises the question, partially addressed in this chapter, whether these conditions in their trait expressions might serve some useful function, even as they burden the individual with cares and instability which could predispose to full-blown affective disease. By their very chronicity, these subaffective conditions pose difficult conceptual and clinical questions about their differentiation from personality disorders.[11] Sceptics might argue that subthreshold affective conditions are nothing more than personality disorders and/or expressions of 'neuroticism'. Actually, a close examination of the Eysenck personality inventory, which ranges over a large terrain of depressiveness, anxiousness, emotionality, and mood lability among others, reveals low-grade intermittent affective symptomatology.[12] And at least one genetic investigation has reported that neuroticism and major depression in women share substantial genetic underpinnings.[13] Nonetheless, clinicians have always preferred categorical constructs, because neuroticism and related personality constructs do not do justice to the rich clinical phenomenology of disorders within the subaffective realm. I finally wish to point out that terms like 'neurotic', 'psychopathic', or 'personality disorder' used as epithets to describe a person have pejorative connotations. They tend to describe what is negative about someone, whereas 'temperament' refers to the optimum mixture of both liabilities and assets regarding a human being, thereby rendering therapeutic work possible with relatively little countertransference. This is particularly relevant when we consider the distinct possibility that the many dysthymic and cyclothymic individuals might otherwise be labeled 'borderline.'[11]

The dysthymic spectrum

History

The term 'dysthymia' (meaning 'bad mood') originated in classical Greek and is still in current use in that country with the same connotation.

In the Hippocratic School, it was considered as part of the broader concept of melancholia (meaning 'black bile'). A temperament predisposed to melancholia was also delineated, and referred to individuals who were lethargic, brooding, and insecure. It was not until the early nineteenth century that dysthymia was reintroduced into medicine by the German physicians, Stark and Fleming to describe depressions in inpatients that pursued a chronic course.[14] Eventually, dysthymia came to subsume all mood disorders. The major residue of dysthymia in the latter sense in Europe today is the French rubric of *les dysthymies*, as a synonym for *troubles de l'humeur*; the DSM-IV or ICD-10 'dysthymic disorder' in that country is translated as *le trouble dysthymique*.

The more direct lineage of our current usage of the term dysthymia is to be found in the latter part of the nineteenth century in the work of Kraepelin, who delineated the depressive disposition as one of the constitutional foundations of affective episodes. The condition often began early in life, such that by adolescence many showed an increased sensitivity to life's sorrows and disappointments: they were tormented by guilt, had little confidence in their abilities, and suffered from low energy. As they grew into adulthood, they experienced 'life with its activity [as] a burden which they habitually [bore] with dutiful self-denial without being compensated by the pleasures of existence'. In some, these temperamental peculiarities were so marked that they could be considered 'morbid without the appearance of more severe, delimited attacks . . .' (clearly foreshadowing the modern concept of trait dysthymia). In other cases, recurrent melancholia arose from this substrate without definite boundaries (again anticipating the concept of 'double-depression').

Subsequently, Kurt Schneider in his opus *Psychopathic Personalities* devoted considerable space to a depressive type whose entire existence was entrenched in suffering. Building on this rich phenomenological tradition, our research in Memphis[15] helped in operationalizing the core characteristics of such patients encountered in contemporary practice: gloomy, sombre, and incapable of having fun; brooding, self-critical, and guilt-prone; lack of confidence, low self-esteem, preoccupation with failure; pessimistic, easily discouraged; easy to tire, sluggish, and bound to routine; non-assertive, self-denying, and devoted; shy and sensitive. These traits have excellent internal consistency and discriminatory ability.[16] Similar concepts have also appeared in the Japanese literature,[17] with particular emphasis on self-critical attitudes, persistence in work habits, and devotion to others. Finally, the French construct of *la depression constitutionelle*[18] has emphasized the lethargic aspects with a sense of inadequacy. A self-rated scale in all of these languages[8] now can assist in reliable and valid assessment of depressive temperament traits.

The classical tenet in psychiatry has been that affectively ill patients recover from their acute episodes with relatively little symptomatic residua and dysfunction. Community psychiatry, which has given renewed visibility to the temperamentally expressed low-grade fluctuating depressive disorders, has challenged this classic view. With the advent of DSM-III, such patients are now officially designated as 'dysthymic.' In the ICD-10 classification, the low-grade depressive baseline is considered the main pathology; only an occasional superimposed depressive episode is permitted, provided that it is mild. In DSM-IV, at least two patterns have been described: pure dysthymia uncomplicated by major depression and a more prevalent pattern of dysthymia complicated by major depressive episodes that could be even moderate or severe in intensity (and which has been dubbed 'double depression').

The mystery of this incapacitating depressive subtype—long recognized, but only recently sanctioned in official diagnostic manuals—is that, in their habitual condition, sufferers lack the classical 'objective' or 'major' signs of acute clinical depression, such as profound changes in psychomotor and vegetative functions. Instead, patients consult their doctors for more fluctuating complaints consisting of gloominess, lethargy, self-doubt, and lack of *joie de vivre*; they typically work hard, but do not enjoy their work; if married, they are deadlocked in bitter and unhappy marriages which lead neither to reconciliation nor separation; for them, their entire existence is a burden: they are satisfied with nothing, complain of everything, and brood about the uselessness of existence. As a result, in the past those who could afford it were condemned to the couch for what often proved to be interminable analysis. The legitimization of dysthymia as a clinically significant variant of affective disorder in both the United States and WHO classifications has helped the cause of more cost-effective treatments.

To sum up, for nearly 2500 years physicians have described individuals with a low-grade chronic depressive profile marked by gloominess, pessimism, low enjoyment of life, relatively low drive, yet endowed with self-critical attitudes and suffering for others.[19]This constellation is as much a virtue as it is a disposition to melancholy, and many dysthymic patients presenting clinically have various admixtures of major depression. This is compatible with a spectrum-concept of depressive illness, which defines various degrees of severity.

Clinical picture and diagnostic considerations[20]*

Diagnostic criteria for dysthymia in both DSM-IV and ICD-10 stipulate a two-year duration of low-grade depressive symptoms, exclusive of such indicators of severity as suicidality and psychomotor disturbances. Dysthymia is distinguished from chronic major depressive disorder by the fact that it is not a sequel to well-defined major depressive episodes. Instead, patients often complain that they have always been depressed. Most are of early onset (less than 20 years). A late-onset subtype first manifesting after the age of 50 is much less prevalent and has not been well characterized clinically, but it has been identified largely through studies in the community.

At their best, dysthymic individuals invest whatever energy they have in work, leaving none for leisure or social activities. According to Tellenbach, such dedication to work represents overcompensation against depressive disorganization. Kretschmer had earlier suggested that such persons were the 'backbone of society,' devoting their lives to jobs that require dependability and great attention to detail. These features represent the obsessoid facet of dysthymia. Such individuals may seek outpatient counseling and psychotherapy for what some clinicians might consider 'existential depression': individuals who complain that their life lacks lustre, joy, and meaning. Others present clinically because of an intensification of their gloom to the level of clinical depression; history of lifelong low-

* Unless otherwise specified, in the remainder of this chapter, references to concepts, historical developments, and research covering dysthymia and cyclothymia through the year 2000 can be found in this centenary review of Kraepelinian psychiatry.[20]

grade depressive symptoms would distinguish them from episodic major depressive patients.

The proverbial dysthymic patient will often complain of having been 'depressed since birth'. In the eloquent words of Kurt Schneider, "they view themselves as belonging to an 'aristocracy of suffering". These hyperbolic descriptions of suffering in the absence of more objective signs of depression earn such patients the label of 'characterological depression'. The description is further reinforced by the fluctuating depressive picture that merges imperceptibly with the patient's habitual self, leading to the customary clinical uncertainty as to whether dysthymic disorder belongs to the affective or personality disorder domains.

At their worst, patients with low-grade depression having an intermittent course can present such instability in their life, including suicidal crises, that some clinicians would entertain the diagnosis of borderline personality disorder. This is not consistent with the classic picture of dysthymia arising from a temperamental type with more mature ego structure described above. Depressives with unstable (that is to say, 'borderline') personality structure more often belong to the irritable cyclothymic–bipolar II spectrum.

The greatest overlap of dysthymia is with major depressive disorder, but differs from it in that symptoms tend to outnumber signs (more subjective than objective depression). Thus, marked disturbances in appetite and libido are uncharacteristic, and psychomotor agitation or retardation is not observed. Nonetheless, subtle 'endogenous' features are not uncommonly reported: inertia, lethargy, and anhedonia that are characteristically worse in the morning. Because many patients with dysthymia presenting clinically fluctuate in and out of a major depression, the core DSM-IV criteria for dysthymia tend to emphasize vegetative dysfunction, whereas the alternative criterion B for dysthymia in a DSM-IV appendix lists cognitive symptoms; although the latter appear more characteristic of trait dysthymia, the DSM-IV field trial could not demonstrate their specificity for dysthymia.

A Milan-San Diego collaboration of a large sample from community and primary-care medical settings revealed that negative mood (by definition), along with low energy, poor concentration, low self-esteem, sleep and appetite disturbance, and hopelessness (in descending order) were the most common symptoms of dysthymia. These data suggest that the cognitive and somatic symptoms are not easily separable in practice. None the less, this study did raise the possibility that factors could be discerned along two different axes: 'negative affectivity' and 'lassitude with poor concentration'. In our experience, patients loading on the latter factor often complain of hypersomnia and may exhibit subtle bipolar signs; alternatively, they might have some link to the poorly defined constructs of neurasthenia, chronic fatigue syndrome, and fibromyalgia. In terms of differential diagnosis, patients with chronic fatigue syndrome present with disabling fatigue and, typically, deny depressive symptoms; patients with fibromyalgia complain of pain; by contrast, the typical patient with dysthymia cannot stop relating to the physician his or her litany of depressive symptoms. Polysomnography, though not yet definitive, may shed some light on differentiating fibromyalgia from dysthymia proper.

Although dysthymic disorder represents a more restricted concept than does its parent, neurotic depression, it is still quite heterogeneous. Anxiety is not a necessary part of its clinical picture, yet dysthymia is sometimes diagnosed in patients with anxiety and neurotic disorders. That clinical situation is perhaps to be regarded as a secondary or 'anxious dysthymia' or, as some British authors seem to prefer, as part of a 'general neurotic syndrome' (an implicit partial return to the now defunct concept of neurotic depression).

Table 4.5.9.1 The core characteristics of dysthymia

Long-standing subthreshold depression of a fluctuating or persistent nature
Gloomy and joyless disposition
Brooding about the past and guilt prone
Low drive and lethargy
Low self-esteem and preoccupation with failure
Identifies suffering as part of the habitual self

Summarized from Akiskal.[19]

The clinical picture of dysthymic disorder that emerges from the foregoing descriptions is quite varied, with many who fluctuate in and out of major depression, whereas in others the pathology is woven into the habitual self. Prospective follow-up supports a continuum between temperament, dysthymia and major depression. These considerations suggest that a clinically satisfactory operationalization of dysthymia must include both symptoms and trait characteristics (Table 4.5.9.1). The following vignette illustrates this more prototypical form of dysthymic suffering.

Case Study: This 37-year-old never-married male teacher presented with the complaint that he was 'tired of living' and was considering 'ending it all'. He said that much of his life had been 'wasted', he had never known any joy, and that all human existence for him was a 'tragic mistake of God'. He was known to be a dedicated and talented teacher, but he felt all his efforts had been 'useless and in vain'. He said he probably was 'born depressed', because he had not known any happiness and that the only utility he could have for mankind was 'to serve as a specimen to be researched—to shed light on human misery'. Although he conceded that some women found him interesting, even intellectually stimulating, he said he could not enjoy physical intimacy, that even orgasm lacked passion; nonetheless, he masturbated frequently, fantasizing about married female teachers—only to feel guilty. We could not document any major affective episodes. He stated that he had always functioned at a 'mediocre level' (which was at variance with the good feedback students had given him year after year); but did admit he 'appreciated work, because there was nothing else to do'. He denied alcohol and drug habits. There had never been any periods of hypomania, but one of his maternal aunts had been treated for a 'cyclical depression' and was apparently doing well on lithium. The patient's mother was a sombre serious work-oriented woman who had raised three children and had done voluntary work for the church, but had no depressive complaints. His father had died from a coronary attack, but his side of the family was otherwise unremarkable.

Although both DSM-IV and ICD-10 omit suicidal preoccupations in their diagnostic criteria for dysthymia, as testified by the above case, this is what often brings patients to clinical attention.

Course[20]

An insidious onset of depression dating back to late childhood or the teens, preceding any superimposed major depressive episodes by years, even decades, is the most typical developmental

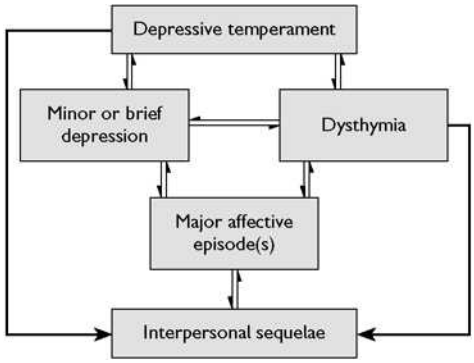


Fig. 4.5.9.1 Diagram to show putative relationships within a broad depressive spectrum.

background of dysthymic disorder. A return to the low-grade depressive pattern is the rule following recovery from superimposed major depressive episodes, if any, hence the designation 'double depression'.

Few studies have studied the phenomenology of dysthymia in childhood. DSM-IV does not seem to distinguish between childhood and adult dysthymia, yet current clinical experience indicates that the main symptoms in childhood dysthymia include irritability, low self esteem, fatigue, low mood, guilt, poor concentration, anhedonia, and hopelessness; as in adults, comorbid anxiety disorders were prevalent; suicidality was more common in adolescents. These findings should be useful in dysthymic children in future studies. A long-term prospective Pittsburgh study of prepubertal children has revealed an episodic course of dysthymia with remissions and exacerbations, and eventual complication by major depressive episodes, as well as hypomanic, manic, or mixed episodes postpubertally. Persons with dysthymic disorder presenting clinically as adults tend to pursue a chronic 'unipolar' course, which may or may not be complicated by major depression: they rarely develop spontaneous hypomania or mania. However, when treated with antidepressants, some adult patients with dysthymia may experience brief hypomanic switches that typically disappear when the antidepressant dose is decreased. Although ICD-10 and DSM-IV would not 'allow' the occurrence of such switches in dysthymia, systematic clinical observation have verified their occurrence in between 10 and 30 per cent of dysthymic patients. In that special dysthymic subgroup, the family histories are typically positive for bipolar disorder. Such patients, often conforming to the double depressive pattern, represent a clinical bridge between major depressive disorder and bipolar II.

A 12-year NIMH prospective study has shown that patients with major depressive disorder spent 44 per cent of their course in low-grade depression (versus 15 per cent of time in major depressive episodes). This suggests that major depression, dysthymia, or otherwise subsyndromal depression constitute somewhat artificial conventions on the threshold and duration of depressive illness, representing alternative manifestations of the same diathesis. In this context, residual intermorbid depressive symptoms have been confirmed as being strongly predictive of a rapid relapse into a new major depressive episode. Various 'major' and 'minor' depressive conditions described in DSM-IV and its appendix must not be viewed as distinct depressive subtypes, but part of a symptomatic continuum. Fig. 4.5.9.1 shows a diagram of these putative relationships within a broad depressive spectrum.

Epidemiology[20]

From 1966 to 1980, *Index Medicus* listed no more than 10 articles on chronic depressions. Since 1980, when dysthymia was first introduced in DSM-III, at least 500 articles have appeared on chronic depression, mostly on dysthymia. This phenomenal growth in research interest parallels the increasing public health significance of this disorder. It is estimated that 3–5 per cent of the world population is suffering from dysthymia. Like major depressive disorders, dysthymia is twice as common in women as in men. Because of its chronicity, dysthymia is among the most prevalent psychiatric conditions in clinical practice. Dysthymia is more disabling, as far as quality of life in social and personal areas, work, and leisure, than depression in the setting of a severe anxiety disorder like agoraphobia. Celibacy is also common in early-onset dysthymia, but not for long; modern successful treatments often lead to a change in marital status!

UCLA research in primary care has focused on depressive symptoms falling short of the major depressive threshold, as far as symptom intensity is concerned, as well as falling short of the two year duration criterion for dysthymia. Despite its chronicity, 50 per cent of people remain unrecognized by general practitioners. Despite the low-grade nature of their depressive complaints, these patients report high degrees of morbidity and impairment in a variety of health domains and quality of life, including 'bed days' (namely, the number of days per year they stayed ill in bed). Actually, these impairments are generally more pronounced than those of patients with a variety of medical conditions, such as hypertension, diabetes, arthritis, and chronic lung disease; only coronary artery disease exceeded the disability of low-grade depression in several domains. Stroke has recently been added on this list.

In light of the foregoing developments, both the World Psychiatric Association[21] and the World Health Organization[22] have developed programmes to address the challenges of educating general practitioners in the proper recognition and treatment of dysthymia.

Aetiological considerations[20]

Some sensitivity to suffering, a cardinal feature of the depressive temperament, represents an important attribute in a species like ours, where caring for young and sick individuals is necessary for survival. This temperament, historically the *Anlage* of dysthymia, in the extreme often leads to clinical depression. The constitutional viewpoint, while dominant in the early part of the twentieth century, gradually disappeared from psychiatric thinking. One reason was that Kurt Schneider preferred to conceptualize such conditions as 'psychopathy', by which he meant abnormal personality development. Independently, Freud's disciples took this one step further and, eventually in outpatients, all milder depressions with a tendency to chronicity came to be considered as the expressions of a character neurosis. In support for this position, these authors could point to the long-standing nature of the interpersonal difficulties in the lives of these individuals. When, and if, antidepressants were prescribed, they were given in homeopathic doses; worse, many patients received stimulants or benzodiazepines rather than genuine antidepressants. Failure to respond to these incorrectly chosen pharmaceutical agents seemed to further reinforce the notion of a 'character' defect.

Several lines of observation during the latter part of the 20th century have challenged the concept of 'character neurosis' as an explanation for low-grade depression, and thereby forced a return to the more classical European concept of temperament with its biological underpinnings. First, in a 1980 Memphis study of rapid eye movement (REM) latency (normally 90 min, measured from sleep onset to the first REM period) conducted in 'depressive characters' who were not in a state of major depression, we reported that REM latency was less than 60 min, and REM was redistributed to the early part of the night (which was the reverse of what we observed in chronic anxious patients). Moreover, a family history for major affective illness (including bipolar) was significantly high in short-REM latency patients. (The reverse was true for those with familial alcoholism and sociopathy.) The sleep findings were so reminiscent of those seen in major affective illness that we were compelled to give our patients systematic open trials with desipramine and nortriptyline (the best-tolerated secondary amine tricyclics in those days) or lithium carbonate if antidepressants failed (based on the observation of familial bipolar disorder in some). Nearly 40 per cent remitted, of whom one out of three developed brief hypomania. The sleep findings have been replicated in other laboratories. Furthermore, a Hungarian study has shown that patients with dysthymia experience transient lifting of their mood with sleep deprivation. Other studies have shown high rates of affective illness in a systematic familial investigation of dysthymic probands. There also exist dysthymic patients whose lifelong suffering and discontent appear, in retrospect, a legacy of an unsatisfactory childhood marked by deprivation or abuse at the hands of alcoholic and/or sociopathic parents or step-parents. Although it is clinically attractive to invoke the notion of 'learned helplessness' secondary to such inescapable childhood traumata, an alternative hypothesis is that the helplessness of these individuals might develop secondary to an inherited diathesis which biases these children's early experiences in a dysphoric direction.

As for neuroendocrine markers, thyroid-releasing hormone-thyroid-stimulating hormone challenge and electrodermal activity similar to those with major depressive disorders are the main findings; by contrast, dexamethasone suppression and catecholamine metabolism are essentially unaltered in dysthymia. These observations, along with the REM latency findings, suggest that dysthymia represents trait depression. Coupled with the family history data, this traitness can be postulated to be of constitutional origin. Certainly, the occurrence of major affective episodes in the long-term course of dysthymia, in both community and clinical samples, is in line with this position. It is, therefore, of great theoretical and practical significance that shortened REM latency has been reported in the offspring of the affectively ill. More recently a variety of other biological findings have been reported in dysthymia, further strengthening the link with major depression: low testosterone and adrenal-gonadal steroid levels,[24] neuro-immune abnormalities,[25] effects of prenatal maternal dysthymia on foetal growth,[26] as well as small genual corpus callosum volume[27] and enlarged amygdalar volume.[28] Coupled with s-allele polymorphism of the serotonin transmitter,[29,30] that subthreshold depression, however defined, is not a 'neurotic' phenomenon in the traditional psychodynamic sense of the term, but part of the spectrum of depressive illness! Table 4.5.9.2 summarizes the foregoing links between dysthymia and major depressive disorder, and which support its inclusion within the family of mood disorders.

Table 4.5.9.2 Evidence for considering dysthymia as a variant of major depressive disorder

Familial affective loading
Phase advance of rapid eye-movement sleep
Diurnality of inertia, gloominess, and anhedonia
Thyroid-releasing hormone-thyroid-stimulating hormone challenge test abnormalities
Low testosterone
Lowered interlukin-1-beta production
Amydalar enlargement
s-allele in serotonin transporter gene
Prospective course complicated by recurrent major depressive episodes
Positive response to selected thymoleptics
Treatment-emergent hypomania

Updated from Akiskal.[19]

There are also exist medical and neurological factors that may contribute to dysthymic symptom formation. Actually, joint medical-neurological and non-affective psychiatric disease is often contributory to extreme refractoriness among the chronic depressive states of these patients. Such patients are at risk for suicide, especially those with epilepsy or progressive degenerative neurological disease. Interestingly, living with a medically disabled spouse or family member, too, can be associated with some chronicity of depression.

The emergence of pathogenetic understanding, as outlined above, is all the more impressive, given the controversies on the very nature of dysthymia and its legitimacy as a nosological entity.[31]

Treatment[20]

The trait nature of dysthymia can be further observed in the fact that dysthymia often pursues an unrelenting course towards chronicity. Thus, spontaneous recovery has been shown to occur in no more than 13 per cent of subjects in the community over 1 year. In outpatient clinics, the outcome is somewhat better, but this is probably due to the treatment received and a longer follow-up.

Most classes of antidepressants have been shown to be effective in dysthymia in double-blind studies (Table 4.5.9.3). The rationale for using classic antidepressants such as tricyclic compounds in our mood clinic was our observation of shortened REM latency in

Table 4.5.9.3 Major controlled pharmacological trials in dysthymia worldwide*

Reference	Country	Medication
Vallejo *et al.* (1987)	Spain	Phenelzine versus imipramine
Kocsis *et al.* (1987)	USA	Imipramine versus placebo
Stewart *et al.* (1989)	USA	Phenelzine versus placebo
Thase *et al.* (1996)	USA	Sertraline versus imipramine
Vanelle *et al.* (1998)	France	Fluoxetine versus placebo
Lecrubier *et al.* (1996)	France	Amisulpride versus placebo
Versiani *et al.* (1997)	Brazil	Moclobemide versus imipramine

*Summarized from references[19,20]

dysthymic subjects in our sleep laboratory. Irreversible monoamine oxidase inhibitors such as phenalzine were used because of the belief that non-classical depressions respond preferentially to this class of drugs; the same can be said for the reversible inhibitors of type A monoamine oxidase. Amisulpride was tried, because it reverses 'negative' symptoms in schizophrenia and, by analogy, it was hypothesized that the low motivation and lethargy seen in some patients with dysthymia reflected a shared underlying dimensional transnosological pathology. The selective serotonin re-uptake inhibitors (SSRIs) were used empirically, because of good tolerance compared with the tricyclic antidepressants, and later it was suggested that improvement in dysthymia correlates with normalization of serotonergic indices. The foregoing trials, conducted in different countries during the past 12 years represent the most eloquent evidence for the increasing worldwide acceptance of the concept of dysthymia as a clinically significant variant of affective disorder.[21,22]

The treatment of dysthymia should continue in most cases for 2 years or more. Tricyclic antidepressants have too many side-effects in clinically effective doses (desipramine equivalent of 150 mg or more per day). Given dietary and medication prohibitions, monoamine oxidase inhibitors are also not practical as first-line drugs. Overall good tolerance in long-term use, despite sexual side-effects, has made the SSRIs first-line intervention treatment for dysthymia; given that many people with dysthymia are young individuals who should be eager to form families, their acceptance of long-term SSRI use is an indication that the alleviation of the depressive suffering of dysthymia is genuine and far outweighs the sexual dysfunction. However, 75 to 150 mg bupropion-SR can be taken in the morning on the desired day of sexual union, but preferably no more than once a week. Moclobemide also spares sexual function, but seems more effective in anxious and milder cases of dysthymia. Amisulpride, which rarely causes amenorrhoea and/or galactorrhoea, is otherwise well tolerated in dysthymia in the more lethargic forms of the illness seen in general medical practice.

The dosage of nearly all antidepressants in dysthymia is in the full range for that recommended for major depression (20–40 mg for fluoxetine). In the case of amisulpride, the dosage is low (25–50 mg), because at this dosage the drug is a dopamine agonist, believed to be the necessary ingredient for its mechanism of action in dysthymia. Both dysthymia and double depression respond equally well, and the duration of underlying dysthymia does not seem to matter. The main difference in treatment for these two course patterns is that dysthymia need not be treated for a lifetime, but double depression should probably be treated indefinitely. Women seem to have a preferentially better response to SSRIs, which have the added benefit of treating the premenstrual accentuation of dysthymic symptoms. Borderline thyroid function (e.g. a high baseline thyroid-stimulating hormone level) preferentially occurs in women with dysthymia, so that these women would benefit from thyroid augmentation (levothyroxine 175 mg/day) of the antidepressant. In those patients who oversleep in the morning, terminal sleep deprivation and/or morning phototherapy represent useful adjuncts to antidepressants. Although there are no controlled studies in children, our clinical experience indicates that SSRIs often prove effective in this population, with the appropriate dosage reduction for body-weight. In adults, concurrent personality disturbances (for instance, obsessoid, avoidant, dependent, and hostile features) do not compromise therapeutic responses. Indeed, more often than not, such personality disturbances recede with the successful alleviation of dysthymic suffering; social function improves in tandem. (However, extremely hostile patients, who may meet symptomatological criteria for 'dysthymia' but whose irritable dysphoria more appropriately belongs to the cyclothymic domain, are best managed with mood-stabilizing anticonvulsants, for example divalproex 600 to 1200 mg/day.)

In a Memphis study, we have shown that, with the judicious use of the foregoing modalities in private practice, three out of four patients with dysthymia engulfed in gloom for much of their lives had sustained remissions for five or more years. Depressive episodes *and* suicidal preoccupation and/or crises were prevented, in tandem with the alleviation of the dysthymia. Approximately 15 per cent experienced 'overcorrection' of their dysthymia in the direction of mild hypomania; this is particularly likely in the presence of inhibited-social phobic traits as part of dysthymia, and when the family history is bipolar. The hypomania is typically short-lived, and tends to disappear when the antidepressant dose is reduced; in some cases, it is necessary to provide lithium (600–900 mg/day) or valproate augmentation (500–750 mg/day). The question has been raised whether selective serotonin-reuptake inhibitors, in particular fluoxetine, change the personality in a hyperthymic direction. In our experience, most observed changes are compatible with adaptive behaviour that emerge as a result of alleviation of depressive suffering; the more distinctly protracted hypomanic changes nearly always require familial bipolar diathesis. It is, nonetheless, true that with SSRIs we have entered an era of 'dimensional psychopharmacology', whereby the clinician could dose the patient to a desired end from a functional standpoint. Many become care-less rather than careless. The present author has also encountered some patients treated with SSRIs who view a life without cares as negative; in such cases, one should opt for very low doses and a more gradual lifting of the dysthymia, and help the patient adjust to a new self-image of normalcy.

As for psychotherapy, there is little credible evidence for its efficacy as monotherapy in the treatment of dysthymia. Actually, some female mental health experts have argued that exploration of one's mental inadequacies, in the 'passive' psychoanalytical situation, is particularly negative for women; the more 'active' cognitive behavioural approaches, which encourage thinking, and behaviours reinforcing for the individual, are preferable. Many clinicians profitably use the latter strategy along with pharmacotherapy to boost the self-esteem of the patient. In a more practical vein, there are clinical management strategies that are specifically useful for both the patients and their clinicians (Table 4.5.9.4).

Table 4.5.9.4 Psychotherapeutic principles in dysthymia*

Provide a believable dose of optimism
Optimize compliance to pharmacotherapy
Limit destructive expression of negative feelings
Address accumulated conflicts
Combat postdepressive resignation and inertia
Provide support for patient and significant others
Be aware of countertransference feelings
Consult experts with extensive experience in treating chronic depression
Gradually mobilize patient's skills and resources

*Updated from Akiskal.[6]

It is particularly important for the clinician not to be submerged by the negative thinking of the patient, and it is even more crucial for the therapist to recognize that a relative lack of progress can generate feelings of 'impotence' and countertransference; periodic consultation with more experienced clinicians in the treatment of chronic depression would be desirable.

Interpersonal psychotherapy has been used in medication failures. This is best viewed as a more practical abbreviation of psychodynamic psychotherapy, with a strong emphasis on support and encouragement for patients with dysthymia who seek help at a time of loss or role transition in their lives. Knowledge of the interpersonal context of depression is obviously important in formulating how the clinician would stage the psychological recovery process from dysthymia. Nonetheless, there are some suggestions that SSRIs often lead to improved coping behaviour, even without formal psychotherapy. Indeed, Canadian studies have shown that a successful response to SSRIs is often associated with decreased emotion-focused coping and decreased perception of daily hassles, as well as alleviation of the sense of loneliness one experiences in chronic depression.

No matter what the active ingredients in antidysthymic treatments, there is little doubt that for the first time in the history of psychiatry we have potent practical treatments to alleviate a major source of chronic human suffering, including what were once deemed depressive characterological attributes inseparable from the habitual self. Helping patients attain a new homeostasis of the self is an art unparalleled in the history of medical science. In our view, it does not constitute what Kraemer has erroneously labeled 'cosmetic psychopharmacology'.[32]

Prevention opportunities[20]

Community subjects with pure dysthymia have been found in two prospective studies to be at risk for major depressive episodes. Because dysthymia often makes its first appearance in juvenile years, identifying the disorder at this early stage represents a special opportunity for prevention in child psychiatry and paediatrics. 'Pure' dysthymia, even without major depression, responds better to pharmacotherapy better than to placebo in 8 out of 9 social domains. St. John's Wort, on the other hand, does not appear to be effective in dysthymia![33]

In still another group of patients, low-grade chronic depressive developments occur in the setting of disabling systematic medical and neurological disorders, and are best categorized as 'secondary dysthymias'. For instance, poliomyelitis may not only lead to deformities in musculoskeletal structures in children, but could permanently scar the sufferer's sense of enjoyment, fulfillment, and outlook of life. Likewise, low-grade chronic depressive development often complicates neurodegenerative cerebrovascular disease later in life. In both situations, psychological factors might be operative, yet the contribution of specific cerebral lesions to the subthreshold mood disturbance may be substantial. This group as a whole is not well captured by the conventional depressive categories in ICD-10 and DSM-IV. In these subacute dysthymic-like conditions, the affective state is often disabling, yet symptomatologically less severe than major depression; it is low grade, yet not as chronic as dysthymia. 'Minor depression' would not capture the clinical significance of their condition. Indeed, there is emerging data that treating these subacute dysthymias may improve the prognosis of the underlying neurological disorder.[22]

In concluding this review of the legitimacy of dysthymia from clinical, biologic and therapeutic standpoints, it is relevant to point out that dysthymia—properly defined—may well serve as a behavioral endophenotype for depressive illness.[34] Such a conceptualization highlights its potential as a target for preventing strategies for major depressive illness.

Cyclothymic disorder and labile-irritable variants[20]

History

Kraepelin included the cyclothymic disposition as one of the temperamental foundations from which manic-depressive illness arose. Kretschmer went one step further and proposed that this constitution represented the core characteristic of the illness: some patients were more likely to oscillate in a sad direction, while others would more readily resonate with cheerful situations; these were merely viewed as variations in the cyclothymic oscillation between these two extremes. Kurt Schneider, who did not endorse the concept of 'temperament', instead referred to 'labile psychopaths' whose moods constantly changed in a dysphoric direction, and who bore no relationship to patients with manic depression. To confuse matters further, Schneider used the term 'cyclothymia' as a synonym for all manic depressive illness, from the mildest to the most severe psychotic forms. Today, 'cyclothymia' is still sometimes used in this broader sense in Germanophone psychiatry. But in much of the rest of the world, cyclothymia (short for 'cyclothymic disorder') is reserved for a form of extreme temperament related to bipolar disorder.

Cyclothymia, which in ICD-9 and DSM-II was subsumed under the affective personalities, was first introduced into DSM-III and DSM-IV and subsequently into ICD-10 as a form of attenuated chronic mood disorder. The diagnosis is not commonly made in clinical practice, because it is almost always seen when a patient presents with major depressive episodes, warranting the designation of 'bipolar II'. Indeed, Hecker used cyclothymia as a synonym for what today we call bipolar II; his short monograph has recently been translated into English.[35] Nonetheless, systematic clinical and familial validation studies conducted in Memphis[36,37] have shown that the construct of cyclothymic temperament is of great theoretical, psychometric and practical significance as one of the possible substrates for major mood disorders. Moreover, it could shed light on social and occupational maladjustment and/or addictive behaviours that could otherwise be misattributed to personality disorder.

Clinical features and diagnostic considerations[20]

By definition, individuals with cyclothymia report short cycles of depression and hypomania that fail to meet the sustained duration criterion for major affective syndromes. At various times, they exhibit the entire range of manifestations required for the diagnosis of depression and hypomania, but only from a few days at a time up to 1 week, rarely longer. These cycles follow each other in an irregular fashion, often changing abruptly from one mood to another, with only rare interposition of 'even' periods. The unpredictability of mood swings is a major source of distress for cyclothymes, as they do not know from moment to moment, how they will feel. As one patient put it, 'my moods swing like a

Table 4.5.9.5 Discriminatory biphasic characteristics of cyclothymic disorder*

Lethargy alternating with eutonia
Shaky self-esteem alternating between low self-confidence and overconfidence
Decreased verbal output alternating with talkativeness
Mental confusion alternating with sharpened thinking
Introverted self-absorption alternating with uninhibited people-seeking

*Summarized from Akiskal *et al.*[36,37]

pendulum, from one extreme to another'. The rapid mood shifts, which undermine the patients' sense of self, may lead to the misleading diagnostic label of borderline personality. But unlike a personality disorder, the mood changes in cyclothymes have a circadian component. One patient described it as follows: 'I would go to bed in a cheerful mood and wake up down in the dumps'. This observation is in line with NIMH psychophysiological data on mood-switching occurring out of the rapid eye movement sleep phase, as reported in more typical cases of manic depression.

The mood swings of cyclothymes are biphasic: eutonia versus anergic periods; people-seeking versus self-absorption; sharpened thinking versus mental paralysis. Table 4.5.9.5 provides an empirically tested set of criteria. In addition, the following presentations characterize their roller-coaster biography.

Irritable periods. At one time or another, labile angry or irritable moods are observed in virtually all these patients. Cyclothymes, unlike patients with epilepsy, are aware of their 'fits of anger', which lead to considerable personal and social embarrassment after they subside. The patients often feel 'on edge, restless, and aimlessly driven'; family and friends report that during these periods patients seem inconsiderate and hostile toward people around them. The contribution of alcohol and sedative-hypnotic drugs to these moods cannot be denied, but the moods often occur in the absence of drugs. Electroencephalography typically reveals no seizure or subseizure activity. The interpersonal costs of such unpredictable interpersonal explosiveness can be quite damaging. One of our patients reported frequent periods where he would start unprovoked fights with very close friends, only to shift into periods of prolonged 'soul-searching, guilt, shame, and embarrassment'. In other patients, outbursts of anger are 'reactive' to minor interpersonal disputes—but once in full force, they are like emotional avalanches with the distinct potential to destroy relationships. Should they dominate the clinical picture, especially among young women who hurt themselves in response to interpersonal contexts, the problematic diagnosis of borderline personality disorder is often invoked (more so in North America than elsewhere). Although controversial, contemporary research suggests that many 'borderline' patients represent a severe labile-irritable variant of cyclothymia on the border of manic-depressive psychosis.[38] On the other hand, bizarre episodes of self-harm with features of post-traumatic stress are uncharacteristic of cyclothymia, and suggest other diagnoses.

Romantic–conjugal failure. It is easy to understand how individuals with mercurial moods would charm others when in an expansive people-seeking mode, and rapidly alienate them when dysphoric. In effect, the life of many of these patients is a tempestuous chain of intense but brief romantic liaisons, often with a series of unsuitable partners. Some rationalize their behaviour on the grounds that their spouse or partner is 'too conservative in sex, too unimaginative, too unaware of the intensity' needed to stimulate them. As expected, frequent marital separations, divorces, and remarriage to the same person occur.

Financial extravagance. One patient in our case series reported going to bars and buying people drinks because he wanted everybody to feel like him. Another patient intermittently showered his lovers with expensive jewellery. In general, however, the extravagance of the cyclothymic group reflects gregariousness and tends to occur on a smaller scale compared to the psychotic manner in which manic patients bring financial ruin to themselves and their families.

Uneven school and work record. Repeated and unpredictable shifts in work and study habits occur in most people with cyclothymia, giving rise to a dilettante biography. Although some do better during their 'high' periods—for example, one car salesman would sell cars only 'when up'—for others, the occasional 'even' periods were more conducive to meaningful work. It is sometimes unappreciated by clinicians inexperienced with bipolarity that the hypomanic period can be one of disorganized and unpatterned busyness that could easily lead to a serious drop in net productivity. For instance, one insurance salesman related that he was less successful when 'high', because he tended to enter into unproductive arguments with his clients. When 'down', productivity obviously abates, although two creative individuals in our case series—one inclined poetically, the other towards painting—produced their better work when coming out of mini-depressions.

Alcohol and drug abuse. An alternating pattern of the use of 'uppers' and 'downers' occurs in at least 50 per cent of patients. We have clinically evaluated at least five cyclothymes who 'sold dope' to maintain their habit: two went to prison. These and other observations suggest that a proportion of substance-abusing, especially stimulant-abusing, patients might be suffering from subtle or cryptic forms of bipolar disorder. The bipolar nature of mood swings in alcohol- or substance-abusing individuals can be documented by demonstrating mood swings well past the period of detoxification; in some cases, escalating mood instability makes its first appearance following abrupt drug or alcohol withdrawal. The DSM-IV criteria for drug-induced or drug withdrawal-induced mood disorder are, in our opinion, biased against the diagnosis of otherwise treatable bipolar spectrum disorders. New evidence suggests that the temperamental disorder might serve as the anlage for self-enhancement or augmentation with cocaine, other stimulants and heroin.[39,40] These features might raise differential diagnostic questions from adult attention deficit hyperactivity disorder (ADHD). The social warmth observed among most people with cyclothymia distinguishes them from ADHD. Also, elation and inflated self-confidence, which occur periodically in cyclothymia, are uncharacteristic of ADHD; the moodiness in the latter is largely depressive in nature. Finally, antidepressants and stimulants typically worsen the moods in cyclothymia; they treat ADHD. In rare cases, however, cyclothymia and ADHD can coexist.

Course patterns

In cyclothymia, hypomania and mini-depression alternate with each other from adolescence. For instance, the optimistic, over-

confident, people-seeking phase can give way to self-absorption, self-doubt, pessimism, and a sense of futility, emptiness, and suicidal ideation. More commonly, depressive periods dominate the clinical picture, interspersed by 'even', 'irritable', and occasional hypomanic periods. Indeed, most people with cyclothymia who present clinically do so because of depression. These depressions are typically short-lived, yet unrelenting in their cyclic course, creating much interpersonal havoc for the patient. The following vignette illustrates the cardinal clinical features of cyclothymia that has not yet progressed to major depression.

Case Study: This 24-year-old songwriter presented with the chief complaint of 'depression so bad that I become totally dysfunctional—I cannot even get out of bed'. Since her mid-teens she had experienced periods lasting from a few days up to a week, during which she would withdraw into herself, losing confidence and interest, feeling drained of energy, and crying when approached by anybody. These periods were particularly prevalent during the autumn and winter months, but they did not coincide with the premenstrual phase. All she needed sometimes was restful sleep to 'feel alive again'; at other times, she would have little sleep, and would wake up 'wired', 'ready to go', 'open to experience all the joy waiting for me out there'; she would exude confidence, 'sensuality and sexual aroma'. These occurred less frequently than the 'down' periods and usually lasted for 1 to 3 days, but were not associated with creative spurts. The latter came as she was descending from 'highs' into a more 'mellow depression'. Her success in music had been sporadic, paralleling the sporadic nature of her 'muses' that visited her on the descending limb of 'hypomania merging with tears'. However, the major toll of her mood swings had been in her personal life, the intensity of her moods had driven away most men she had loved, of whom she had lost count. She described periods of such intense sexual arousal, that sometimes she would go to bed 'with anybody, including women of all ages, shapes, and description'. But, she added: 'I am not a lesbian—oral love is just one way of relating to these women—why not?' She had also experimented with drugs, such as stimulants, which had made her moodiness worse. More recently, she had been prescribed at least two SSRIs, which after a period of 'success' for a few months, had made her depressive swings more frequent and lower in amplitude, leading to the present consultation in our clinic.

As documented in this case, sexual excesses with both sexes are often readily admitted by patients. Winter accentuation or clustering of depressive periods, as exemplified here, is not uncommon in cyclothymia. We also would like to point out the special relationship of the moods to artistic productivity which occur in up to 8 per cent of cyclothymic depressions.[41] The 4-day threshold for hypomania in the official diagnostic manuals is too conservative; as shown in this case, most patients with cyclothymia (and bipolar II disorder) report a threshold of 1 to 3 days (though on occasion, one would observe a hypomanic duration of 1 week or longer). It is also noteworthy that the episodes are short-lived and do not reach the duration threshold for rapid cycling. Sometimes, the term 'ultra rapid cycling' is used for these patients, but we prefer to reserve this for extremely severe cases who require hospitalization. The short cycle length in cyclothymia is, in part, a selection artifact: the universe of patients with bipolar disorder is composed of an extreme variety of overlapping patterns.

The relationship of a cyclothymic temperament to the bipolar spectrum is more complex than that of dysthymia to major depressive disorder. Although cyclothymia can be observed in some patients with full-blown manic-depressive illness (bipolar I with severe or hospitalized mania), it is more commonly associated with the bipolar II pattern (of recurrent major depression with self-limited hypomanias). In a recent French national study of patients with major depression, 88 per cent of those with a cyclothymic disposition belonged to the bipolar II subtype. (Mania, by contrast, has been reported to more likely represent either an extension of hyperthymic traits, or a reversal from a depressive temperamental baseline.)

One-third of patients with cyclothymia studied by us on a prospective basis, progressed to spontaneous affective episodes with more protracted hypomanias and clinical (major) depression.[36,37] Thus, 6 per cent of the original cyclothymic cohort could be reclassified as bipolar I, and 30 per cent as bipolar II. The tendency to switch to hypomania was further augmented by the administration of antidepressants. A larger National Institute of Mental Health study of patients with major depression who switched to bipolar II during a prospective observation period of up to 11 years, found that a temperamental mix of 'mood-labile', 'energetic-active', and 'daydreaming' traits (reminiscent of Kretschmer's concept of the cyclothymic temperament were the most specific predictors of such outcome. Actually such temperamental factors predict who among the offspring of bipolar probands will progress during prospective course to clinical episodes.[42] New data further indicate that cyclothymia might be one of the pathways to suicidality among adolescents.[42]

The foregoing clinical and course characteristics suggest that a cyclothymic temperament leading to major depressive recurrences represents a distinct longitudinal pattern of 'cyclothymic depression', and which appears to capture the core features of bipolar II disorder in contemporary clinical practice. Because hypomanic episodes cannot be easily ascertained by history, assessing cyclothymia in clinically depressed patients represents a more sensitive and specific approach to the diagnosis of bipolar II.

Epidemiology

An excess of interpersonal difficulties and psychiatric consultations distinguish people with cyclothymia in the community from controls; excessive use patterns of stimulants, caffeine, nicotine, and alcohol, have also been well documented. Explosive traits, probably representing the irritable component of cyclothymia, have been reported to be prevalent in the community in a British study. More recently, we found that 6.3 per cent of a national cohort of 1010 Italian students between the ages of 14 and 26 years of age scored above two standard deviations for cyclothymia; this was more prevalent in females, with a ratio of 3:2. Overall, the foregoing data testify to the fact that a cyclothymic and/or labile disposition can be accurately measured, is prevalent, and represents a population at risk for affective disorders. Table 4.5.9.6 summarizes the rates in different populations.

Aetiological aspects

The flamboyant behaviour and the restless pursuit of romantic opportunities in cyclothymia suggest the hypothesis that its

Table 4.5.9.6 Prevalence of cyclothymic and related mood-labile temperaments[20]*

Reference	Population (Country)	Rate (%)
Akiskal *et al.* (1977)	Mental health centre (USA)	10
Weissman and Myers (1978)	Community (USA)	4
Depue *et al.* (1981)	College students (USA)	6
Casey and Tyrer (1986)	Community (UK)	6
Wicki and Angst (1991)	Community (Zurich)	4
Placidi *et al.* (1988)	14–26–year old students (Italy)	6

*These data derive from interview-based studies. For more recent psychometric data based on cyclothymic and a broader range of sub-bipolar temperaments in a self-reported format can be found in a new monograph.[8]

constituent traits may have evolved as a mechanism in sexual selection.[23] Even their creative bent—in poetry, music, painting, or fashion design—may have evolved to subserve such a mechanism. Cyclothymic traits appear to lie on a polygenic continuum between excessive temperament and manic depression. Indeed, clinically identified cyclothymes have patterns of familial affective illness, as one would expect for a *forme fruste* disorder.

Cyclothymia has also been observed in the offspring of manic-depressive probands, with onset in the postpubertal period. Finally, family studies of patients with a bipolar disorder have revealed an excess of cyclothymia. Hypothetically, this temperament might represent one of, if not the most important, inherited trait diathesis for bipolar disorder. For instance, moody-temperamental individuals are over-represented in the 'discordant' monozygotic co-twins of Danish manic-depressives. Alternatively, and in a more theoretical vein, manic-depressive illness might be the genetic reservoir for the desirable cyclothymic traits in the population at large.[23,44] In line with these data and considerations, cyclothymia can be considered a behavioral endophenotype for bipolar disorder.[45] This is supported by recent findings from both Europe and the United States which have shown that it is present in the clinically well relatives of bipolar probands. Cyclothymia might also share familial-genetic relationship with alcohol and substance use,[40] bulimic,[46] panic,[47] obsessive-compulsive[48] as well as atypical depressive and borderline conditions.[49]

Clinical management

The proper pharmacological treatment for cyclothymic excesses is low doses of lithium (600–900 mg/day) or valproate (500–750 mg/day). These are based on open systematic studies.[50] There is some data from a controlled trial with lithium about the prevention of depression in cyclothymic individuals. Similarly controlled data exist for a related construct—'labile personality'. Generally speaking, patients with cyclothymia object to the 'overcontrol' that may come from mood stabilizers, and this is particularly the case with lithium. Lamotrigine is also being used on clinical grounds in the unstable dysthymic-cyclothymic spectrum. In those with 'borderline' features, lamotrigine is particularly promising.

Patients should be taught how to live with the extremes of their temperamental inclinations, and to seek professions where they determine the hours that they work. Marriage to a work-oriented or a rich older spouse might sustain them for a while, but eventually interpersonal friction and sexual jealousy terminate such marriages. The artistically inclined among them should be encouraged to live in those parts of a city inhabited by artists and other intellectuals, where temperamental excesses are better tolerated. ltimately, the decision to use mood stabilizers in such individuals should balance any benefits from decreased mood instability against the social and creative spurts that the cyclothymic disposition can bring to them. Their clinical management represents a challenging task for the psychiatrist who is willing to learn about the lifestyle of these individuals, not prejudging them by the more mundane norms of society. But the psychiatrist should also be there to help them during the multiple crises of their lives. Low-dose sedating neuroleptics, both classical (e.g. thioridazine 50 mg at bedtime) and atypical (e.g. quetiapine 25–50 mg at bedtime) may temporarily help to diffuse such crises. It is only when a clinician has earned therapeutic alliance with a patient that the latter will permit limit-setting on his or her extravagant or outrageous behaviour. Parents might also benefit from some counselling, because the dilettante life of their children is often a source of great sorrow for them. Rarely, parents or spouses are rewarded by great artistic or intellectual achievement, which does not necessarily reduce the pain that the volatile cyclothymes bring to their loved ones.

Kurt Schneider admonished the kin of labile individuals (who might approximate the contemporary concept of borderline personality disorder) 'on their bad days . . . to keep out of their way as far as possible'. Cyclothymes with some insight into their own temperament would give the same advice to their loved ones. A cautious trial of anticonvulsants will often prove effective in those distressed enough by their behaviour to comply with such treatment.

Prevention

The offspring of patients with bipolar disorder who exhibit a cyclothymic level of temperamental dysregulation represent a logical population for prevention studies. This is a challenge for the 21st century. Presumably molecular genetic testing will one day identify those moody individuals who carry the genes for bipolar disorder. At present, it would be useful to conservatively follow-up the cyclothymic offspring of people with bipolar disorders and provide them with psychoeducation about the necessity of avoiding stimulants and sleep deprivation. It may not be entirely possible to prohibit the use of occasional alcohol consumption, but benzodiazepines should not be used. It is also imperative, should they get depressed, to protect them from the indiscriminate prescription of antidepressants.

Mood-labile female prisoners, commonly given the diagnosis of antisocial or borderline personality disorder, may represent affective variants with irritable cyclothymic features. Formal studies are needed in prison populations, to assess more precisely the rates of preventable cryptic bipolarity among female and male offenders.

It is finally worthwhile to mention that affective temperaments with irritable, cyclothymic and irritable-hyperthymic traits might predispose to HIV infection. The public health dimensions of this question deserve further research focus.

The hyperthymic temperament[38]*

History and description

Although well described by classical German psychiatrists (e.g. Schneider), the hyperthymic type appears neither in DSM-IV nor in ICD-10. A lifelong disposition, hyperthymia must be distinguished from short-lived hypomanic episodes. Alternatively, this disposition can be characterized as trait hypomania. It derives from the ancient Graeco-Roman sanguine temperament, believed to represent the optimal mixture of behavioural traits. They are full of zest, fun-loving, and prone to lechery: their habitual disposition is one of buoyant action-orientation, extraverted people-seeking, overconfidence, and swift thinking. Typically short sleepers, they possess boundless energy to invest in sundry causes and projects, which often earn them leadership positions in the various professions and politics, yet their carefree attitudes and propensity for risk-taking can bring them to the brink of ruin; this is particularly true for their finances and sexual life, which can be marred by scandal. A hyperthymic lifestyle is so reinforcing that some resort to 'augmenting' it with stimulants such as amphetamines or cocaine. In brief, while a hyperthymic temperament *per se* does not constitute affective pathology—indeed, it represents a constellation of adaptive traits—in excess it could lead to undesirable complications. The latter will be our main focus here.

Epidemiology

The foregoing considerations partly explain why such a temperament has received scant research attention in the community. Extrapolating from studies on intermittent hypomania in the community, if strictly limited to hypomania with early onset and persistent course, the prevalence of hyperthymic temperament can be estimated to be slightly under 1 per cent. On the other hand, the traits that constitute the hyperthymic profile are so desirable that normal individuals tend to endorse them; in a recent Pisa-San Diego collaboration[16] involving 1010 students aged between 14 and 26, of whom 8.2 per cent met the full criteria for hyperthymic temperament, all participants scored between the first and second positive standard deviation. The TEMPS-A tends to validate these findings in Lebanon,[51] Argentina[52] and Hungary.[53] More work needs to be done on the psychometric standardization of this temperamental construct; on the other hand, all studies are consistent in showing marked male predominance.

Diagnostic aspects

On the positive side, hyperthymic individuals are enterprising, ambitious, and driven, often achieving considerable social and vocational prominence. Abuse of stimulants is not so much an attempt to ward off depression and fatigue as an effort to enhance their already high-level drive and, sometimes, to further curtail their already reduced need for sleep. Hyperthymic individuals typically marry three or more times. Others, without entering into legally sanctioned matrimony, form three or more families in different cities; these men are capable of maintaining such relationships for long periods, testifying to their financial and personal resourcefulness, as well as their generosity towards their lovers and the offspring from such unions. Unlike the antisocial psychopath who is predatory on others and neglects or abuses his women and children, these men care for their loved ones. But obviously, the 'arrangement' involving women of different generations is complex, and a fertile soil for jealousy, drama, scandal, and tragedy. Nonetheless, it is not uncommon to see more than three or four women crying profusely and expressing their common grief at the funerals of these men!

Although individuals with hyperthymia optimally enjoy the advantage of their reduced need for sleep (giving more time and energy to invest in work and pleasure), some present clinically because of insomnia. Thus, in a predominantly male sample of executives presenting to a sleep centre,[54] habitual sleep need was 4 to 5 h; however, they had been intermittently bothered by 'nervous energy' and difficulty falling asleep. Now, in late middle age, alcohol was no longer an effective hypnotic. Although they vigorously denied depressive and other mental symptoms—indeed, they had extremely low scores on self-rated depression—spouses or lovers provided collateral information about brief irritable-depressive dips, especially in the morning and, in some cases, more protracted 'fatigue states' of days to weeks during which the subject would vegetate. Despite these depressive dips, these patients were distinguished from the constantly shifting moods of cyclothymic patients by the fact that the depressions arose from a baseline of trait hypomania of a more or less stable course. Our most current diagnostic guidelines for a hyperthymic temperament consist of the following traits on a habitual basis since at least early adulthood: cheerful, overoptimistic, or exuberant; extraverted and people-seeking, often to the point of being overinvolved or meddlesome; overtalkative, eloquent, and jocular; uninhibited, stimulus-seeking, and sexually-driven; vigorous, full of plans, improvident; overconfident, self-assured, and boastful attitudes that may reach grandiose proportions.

A systematic retrospective review of the case records of people with manic depression, whose course was dominated by manic episodes, was recently undertaken in Munich, yielding attributes overlapping with our proposed list: active, vivid, extraverted, verbally aggressive, self-assured, strong willed, engaged in self-employed professions, risk-taking, sensation-seeking, breaking social norms, spendthrift, and generous. The fact that at least 10 per cent of patients with major depression in an Italian study could be characterized as premorbidly hyperthymic, suggests that this temperament has relevance to both major affective poles. This is an important diagnostic consideration, because rather than being considered narcissistic depressions, these should be recognized as a soft bipolar variant.

Aetiological aspects

It is of interest that Gardner's ethological analysis[55] of what constitutes 'leadership' led to a description that overlaps with a hyperthymic profile: cheerfulness, joking, irrepressible infectious quality, unusual warmth, expansive, strong sense of confidence in one's abilities, scheming, robust, tireless, pushy, and meddlesome. Hypothetically, this temperament evolved in primates whose social life required leadership to better face challenges to the group from within and without.

In a sleep electroencephalography study,[54] REM latency was found to be shortened; similar findings have been reported on the sleep of manic patients, thereby supporting the notion of a trait-state continuum at the neurophysiological level. (Although counter-intuitive, this neurophysiological marker appears to be shared between the depressive and hyperthymic poles.) Finally, family

* Unless otherwise specified, this review article contains most of the references to the concepts, history, and research on the hyperthymic type.

history for frank bipolar disorder characterizes many such individuals. The foregoing data, albeit limited, suggest that hyperthymic traits share several key biological underpinnings of affective disorder.

Course and treatment

Little is known about the natural course of the hyperthymic temperament, except what can be reconstructed retrospectively from biographical and clinical studies. Given their overoptimistic and self-assured style of thinking, these individuals feel perfectly fit in all areas of functioning and thus have no need to consult a psychiatrist. They do so only when forced by loved ones. There are no systematic treatment studies on hyperthymia. Anecdotally,[50] low doses of anticonvulsants such as valproate (e.g. 500–750 mg/day) can be useful in reducing drivenness, when deemed appropriate on clinical grounds, such as in the presence of cardiovascular disease, or when enormous sexual appetite places them at risk for social scandals and, in some cases, exposure to HIV infection.[56] Stimulant-abusing subjects with hyperthymia can be detoxified with valproate, carbamazepine, or gabapentin. Clinically depressed subjects with a hyperthymic temperament often respond poorly to antidepressants. In our opinion, the efficacy of lithium augmentation in resistant depression is partly based on the high prevalence of hyperthymia in resistant populations; it would be wise to keep the dose of lithium in augmentation in such individuals to a lower to middle range (i.e. 600–900 mg/day).

People with hyperthymia are action-oriented, and are not inclined to any type of self-examination. Furthermore, their hypertrophied sense of denial makes them poor candidates for psychotherapy. The physician must, nonetheless, attempt psychoeducation about the harm that can come to them and their loved ones because of their temperamental excesses. Alcohol consumption, which is common in these individuals, should not be abruptly interrupted because of the risk of the switch to a suicidal depression. If detoxification is necessary for health reasons, admission to a suitable inpatient facility should be arranged. The occasion might be profitably used for whatever counselling is deemed appropriate for life and health situations confronting them at the time.

Therapeutic and preventive aspects

In clinical practice, hyperthymic individuals are likely to be confused with narcissistic or antisocial types. Otherwise, they rarely present to psychiatrists, except as the premorbid adjustment of manic-depressive illness. Hyperthymic individuals are often the driving force of society in economic and political life and unless they are involved in scandals or suicidal behavior, rarely come to the attention of clinicians. In the rare circumstances they seek psychiatric advice, it is due to exasperated pressure from loved ones; even then, they tend to dictate rather than follow treatment recommendations. Their sense of entitlement derives in part from the fact these individuals have considerable leadership talent and often bequeath large sums of endowments for research, museums and other community projects. Some are performing artists. Others are famed for their erotic life in the tabloid and popular press. Rare biological investigations have been conducted on hyperthymia involving fascinating neurophysiologic and endophenotype studies.[57,58] The preventive potential of such investigations for manic-depressive illness remains uncertain at this time.

Further information

Akiskal, H.S., Akiskal, K.K. (eds.) (2005). TEMPS: Temperament Evaluation of Memphis, Pisa, Paris and San Diego. Special Issue, *Journal of Affective Disorders*, **85**, 1–242.

References

1. Akiskal, H.S., Judd, L.L., Gillin, J.C., *et al.* (1997). Subthreshold depressions: clinical and polysomnographic validation of dysthymic, residual and masked forms. *Journal of Affective Disorders*, **45**, 53–63.
2. Meier, W., Lichterman, D., Minges, J., *et al.* (1992). The risk of minor depression in families of probands with major depression: sex differences and familiality. *European Archives of Psychiatry and Clinical Neuroscience*, **242**, 89–92.
3. Kendler, K.S., Neale, M.C., and Kessler, R.C. (1992). A population-based twin study of major depression in women: the impact of varying definitions of illness. *Archives of General Psychiatry*, **49**, 257–66.
4. Akiskal, H.S., Bitar, A.H., Puzantian, V.R., *et al.* (1978). The nosological status of neurotic depression: a prospective three-to-four year examination in light of the primary-secondary and unipolar-bipolar dichotomies. *Archives of General Psychiatry*, **35**, 756–66.
5. Bronisch, T., Wittchen, H.-U., Krieg, C., *et al.* (1985). Depressive neurosis: a long-term prospective and retrospective follow-up study of former inpatients. *Acta Psychiatrica Scandinavica*, **71**, 237–48.
6. Akiskal, H.S. and Cassano, G.B. (ed.) (1997). *Dysthymia and the spectrum of chronic depressions*. Guilford Press, New York.
7. Akiskal, H.S., Akiskal, K.K., Lancrenon, S., *et al.* (2006). Validating the bipolar spectrum in the French National EPIDEP Study: Overview of the phenomenology and relative prevalence of its clinical prototypes. *Journal of Affective Disorders*, **96**, 197–205.
8. Akiskal, H.S. and Akiskal, K.K., (eds.) (2005). TEMPS: Temperament Evaluation of Memphis, Pisa, Paris and San Diego. Special Issue, *Journal of Affective Disorders*, **50**, 1–242.
9. Akiskal, H.S. and Weise, R.E. (1992). The clinical spectrum of so-called 'minor' depressions. *American Journal of Psychotherapy*, **46**, 9–22.
10. Lewis, A. (1936). Melancholia: a prognostic study. *Journal of Mental Science*, **82**, 488–558.
11. Akiskal, H.S. (1981). Subaffective disorders: dysthymic, cyclothymic and bipolar II disorders in the 'borderline' realm. *Psychiatric Clinics of North America*, **4**, 25–46.
12. Snaith, R.P. (1991). Measurement in psychiatry. *British Journal of Psychiatry*, **159**, 78–82.
13. Kendler, K.S., Neale, M.C., Kessler, R.C., *et al.* (1993). A longitudinal twin study of personality and major depression in women. *Archives of General Psychiatry*, **50**, 853–62.
14. Brieger, P. and Marneros, A. (1997). Dysthymia and cyclothymia: historical origins and contemporary development. *Journal of Affective Disorders*, **45**, 117–26.
15. Akiskal, H.S. (1983). Dysthymic disorder: psychopathology of proposed chronic depressive subtypes. *American Journal of Psychiatry*, **140**, 11–20.
16. Placidi, G.F., Signoretta, S., Liguori, A., *et al.* (1998). The Semi-Structured Affective Temperament Interview (TEMPS-I): reliability and psychometric properties in 1010 14–26 year students. *Journal of Affective Disorders*, **47**, 1–10.
17. Kasahara, Y. (1991). The practical diagnosis of depression in Japan. In *The diagnosis of depression* (eds. J.P. Feighner and W.F. Boyer), pp. 163–75. Wiley, Chichester.
18. Montassut, M. (1938). *La dépression constitutionnelle: l'ancienne neurasthénie dans ses rapports avec la médecine générale; clinique biologie, thérapeutique*. Masson, Paris.

19. Akiskal, H.S. (1996). Dysthymia as a temperamental variant of affective disorder. *European Psychiatry*, **11** (Supplement), 117s–22s.
20. Akiskal, H.S. (2001). Dysthymia and cyclothymia in psychiatric practice a century after Kraepelin. *Journal of Affective Disorders*, **62**, 17–31.
21. The WPA Dysthymia Working Group (1995). Dysthymia in clinical practice. *British Journal of Psychiatry*, **166**, 174–83.
22. Licinio, J., Prilpko, I., and Bolis, C.L. (eds.) (1997). Dysthymia in neurological disorders. In *Proceedings of the WHO Meeting*. World Health Organization, Geneva.
23. Akiskal, K., Akiskal, H.S. (2005). The theoretical underpinnings of affective temperaments: implications for evolutionary foundations of bipolarity and human nature. *Journal of Affective Disorders*, **85**, 231–9.
24. Markianos, M., Tripodianakis, J., Sarantidis, D., *et al*. (2006). Plasma testosterone and dehydroepiandrosterone sulfate in male and female patients with dysthymic disorder. *Journal of Affective Disorders*, **101**, 255–8.
25. Brambilla, F., Monteleone, P., Maj, M. (2004). Interleukin-1beta and tumor necrosis factor-alpha in children with major depressive disorder or dysthymia. *Journal of Affective Disorders*, **78**, 273–7.
26. Field, T., Diego, M.A., Hernandez-Reif, M., *et al.* (2008). Prenatal dysthymia versus major depression effects on maternal cortisol and fetal growth. *Depressive Anxiety*, **22**, in press.
27. Lyoo, I.K., Kwon, J.S., Lee, S.J., *et al.* (2002). Decrease in genus of the corpus callosum in medication-Naïve, early-onset dysthymia and depressive personality disorder. *Biological Psychiatry*, **52**, 1134–43.
28. Tebartz van Elst, L., Woermann, F.G., Lemieus, L., *et al.* (1999). Amygdala enlargement in dysthymia—a volumetric study of patients with temporal lobe epilepsy. *Biological Psychiatry*, **46**, 1614–23.
29. Oliveira, J.R., Carvalho, D.R., Pontual, D., *et al.* (2000*).* Analysis of the serotonin transporter polymorphism (5-HTTLPR) in Brazilian patients affected by dysthymia, major depression and bipolar disorder. *Molecular Psychiatry*, **5**, 348–9.
30. Gonda, X., Zsombok, T., Rihmer, Z., *et al*. (2006). The 5HTTLPR polymorphism of the serotonin transporter gene is associated with affective temperaments as measured by TEMPS-A. *Journal of Affective Disorders*, **91**, 125–31.
31. Burton, S.W. and Akiskal, H.S. (ed.) (1993). Dysthymic disorder. Gaskell and Royal College of Psychiatrists, London.
32. Kraemer, P.D. (1993). *Listening to Prozac*. Viking Press, New York.
33. Randlov, C., Mehlsen, J., Thomsen, C.F., *et al.* (2006). The efficacy of St. John's Wort in patients with minor depressive symptoms or dysthymia—a double-blind placebo-controlled study. *Phytomedicine*, **13**, 215–21.
34. Niculescu, A.B., Akiskal, H.S. (2001). Proposed endophenotypes of dysthymia: evolutionary, clinical and pharmacogenomic considerations. *Molecular Psychiatry*, **6**, 363–6.
35. Koukopoulos, A. (2003). Ewald Hecker's description of cyclothymia as a cyclical mood disorder: its relevance to the modern concept of bipolar II. *Journal of Affective Disorders*, **73**, 199–205.
36. Akiskal, H.S., Djenderedjian, A.H., Rosenthal, R.H., *et al*. (1977). Cyclothymic disorder: validating criteria for inclusion in the bipolar affective group. *American Journal of Psychiatry*, **134**, 1227–33.
37. Akiskal, H.S., Khani, M.K., and Scott-Strauss, A. (1979). Cyclothymic temperamental disorders. *Psychiatric Clinics of North America*, **2**, 527–54.
38. Akiskal, H.S. (1992). Delineating irritable-choleric and hyperthymic temperaments as variants of cyclothymia. *Journal of Personality Disorders*, **6**, 326–42.
39. Maremmani I, Perugi G, Pacini M, *et al*. (2007). Toward a unitary perspective on the bipolar spectrum and substance abuse: opiate addiction as a paradigm. *Journal of Affective Disorders*, **93**, 1–12.
40. Merikangas, K.R., Herrell, R., Swendsen, J., *et al*. (2008). Specificity of bipolar spectrum conditions in the comorbidity of mood and substance use disorders: results from the Zurich cohort study *Archives of General Psychiatry*, **65**, 47–52.
41. Akiskal, H.S. and Akiskal, K. (1988). Re-assessing the prevalence of bipolar disorders: Clinical significance and artistic creativity. [Psychiatric et Psychologic] *European Psychiatry*, **3**, 29s–36s.
42. Akiskal, H.S., Downs, J., Jordan, P., *et al.* (1985). Affective disorders in the referred children and younger siblings of manic-depressives: Mode of onset and prospective course. *Archives of General Psychiatry,* **42,** 996–1003.
43. Kochman, F.J., Hantouche, E.G., Ferrari, P., *et al*. (2005). Cyclothymic temperament as a prospective predictor of bipolarity and suicidality in children and adolescents with major depressive disorder. *Journal of Affective Disorders*, **85**,181–9.
44. Akiskal H.S., Akiskal, K.K. (2007) In search of Aristotle. *Journal of Affective Disorders*, **100**, 1–6.
45. Evans, L., Akiskal, H.S., Greenwood, T.A., *et al*. (2008). Suggestive linkage of a chromosomal locus on 18p11 to cyclothymic temperament in bipolar disorder families. *American Journal of Medical Genetics B Neuropsychiatric Genetics*, in press.
46. Perugi, G., Toni, C., Passino, M.C., *et al*. (2006). Bulimia nervosa in atypical depression: the mediating role of cyclothymic temperament. *Journal of Affective Disorders*, **82**, 91–7.
47. Akiskal, H.S., Akiskal, K.K., Perugi, G., *et al*. (2006). Bipolar II and anxious reactive 'comorbidity': toward better phenotypic characterization suitable for genotyping. *Journal of Affective Disorders*, **96**, 239–47.
48. Hantouche, E.G., Angst, J., Demonfaucon, *et al.* (2003). Cyclothymic OCD: a distinct form? *Journal of Affective Disorders*, **75**, 1–10.
49. Perugi, G., Toni, C., Travierso, M.C., *et al*. (2003). The role of cyclothymia in atypical depression: toward a data-based reconceptualization of the borderline-bipolar II connection. *Journal of Affective Disorders*, **73**, 87–98.
50. Akiskl, H.S. (1977). Chronic disturbances of temperament. *Bibliotheca Psychiatrica (Basel)*, **167**, 29–32.
51. Karam, E.G., Mneimneh, Z., Salamoun, M., *et al.* (2005). Psychometric properties of the Lebanese-Arabic TEMPS-A: a national epidemiologic study. *Journal of Affective Disorders*, **87**, 169–183.
52. Vázquez, G.H., Nasetta, S., Mercado, B., *et al.* (2007). Validation of the Temps-A Buenos Aires: Spanish psychometric validation of affective temperaments in a population study of Argentina. *Journal of Affective Disorders*, **100**, 23–9.
53. Rózsa, S., Rihmer, Z., Gonda, X., *et al.* (2007). A study of affective temperaments in Hungary: Internal consistency and concurrent validity of the TEMPS-A against the TCI and NEO PI-R. *Journal of Affective Disorders*, in press.
54. Akiskal, H. (1984). Characterologic manifestations of affective disorders: toward a new conceptualization. *Integrative Psychiatry*, **2**, 83–8.
55. Gardner, R. Jr. (1982). Mechanisms in manic-depressive disorder: an evolutionary model. *Archives of General Psychiatry*, **39**, 1436–41.
56. Moore, D., Atkinson, J.H., Gonzalez, B.A., *et al*. (2005). Temperament and risky behaviors: A pathway to HIV? *Journal of Affective Disorders*, **85**, 191–200.
57. Hensch, T., Herold, U., Brocke, B. (2007). An electrophysiological endophenotype of hypomanic and hyperthymic personality. *Journal of Affective Disorders*, **101**, 13–26.
58. Chiaroni, P., Hantouche, E.G., Gouvernet, *et al.* (2004). [Hyperthymic and depressive temperaments study in controls, as a function of their familial loading for mood disorders.] *Encephale*, **30**, 509–15.

4.6

Stress-related and adjustment disorders

Contents

4.6.1 Acute stress reactions

Anke Ehlers, Allison G. Harvey, and Richard A. Bryant

Introduction

Exceptionally stressful life events can cause severe psychological symptoms, including anxiety, feelings of derealization and depersonalization, and hyperarousal. In one of the first studies to comprehensively document acute reactions to extreme stress, Lindemann[1] observed that the symptoms reported by survivors of the Coconut Grove Fire included avoidance, re-experiencing scenes from the fire, reports of derealization, and the experience of anxiety when exposed to reminders of the event. Similarly, acute responses reported by soldiers who fought in the First and Second World Wars included re-experiencing symptoms and dissociative responses such as numbing, amnesia, and depersonalization.[2]

The *International Classification of Diseases* has recognized acute stress reactions since 1948 (ICD-6).[3] In the most recent edition (ICD-10),[4] early reactions to exceptionally stressful life events are diagnosed as acute stress reaction, one of the diagnoses in the section headed 'reactions to severe stress, and adjustment disorders'.

In contrast, the *Diagnostic and Statistical Manual of Mental Disorders* did not formally recognize that exceptionally stressful life events are a sufficient cause of psychological symptoms until 1980 when its third edition (DSM-III)[5] introduced the diagnosis of post-traumatic stress disorder (**PTSD**). DSM-III did not stipulate a duration for the symptoms, but the revised third version (DSM-III-R)[6] required that the symptoms of PTSD must be present for more than 1 month after the traumatic event. This stipulation precluded the inclusion of acutely traumatized individuals who instead were diagnosed with adjustment disorder.[7] In 1994 the fourth edition of DSM (DSM-IV)[8] formally recognized acute trauma reactions by introducing the new diagnosis of acute stress disorder into the anxiety disorders section.

The diagnoses of acute stress reactions in ICD-10 and of acute stress disorder in DSM-IV have similarities in that they are caused by extreme stress and have some overlap in symptom patterns. They can be considered as two separate points on a continuum from transient to more enduring symptoms. However, there are also differences in the underlying concepts, as we will discuss in this chapter.

Clinical features

Acute stress reactions, as defined in ICD-10, are transient reactions to exceptional physical and/or mental stress. There is an initial stage of a 'daze', including narrowing of attention, inability to comprehend stimuli, and disorientation. This is followed by a rapidly changing picture of symptoms that may include withdrawal from the surrounding situation, flight reactions, panic anxiety, and autonomic hyperarousal, depression, anger, or despair. Symptoms usually begin to diminish after 24 to 48 h and should be minimal after about 3 days.

In contrast, acute stress disorder, as defined in DSM-IV, is only diagnosed if the psychological symptoms persist for more than 2 days. Dissociative symptoms dominate the disorder. Dissociation refers to a disruption of the usually integrated feelings of consciousness, memory, identity, or perception of the environment. Symptoms include a subjective sense of numbing or detachment, reduced awareness of surroundings, derealization, depersonalization, or dissociative amnesia. In addition, patients with acute stress disorder experience symptoms that are typical of PTSD, namely re-experiencing aspects of the event, avoidance of reminders of

the event, and hyperarousal symptoms. Acute stress disorder is seen in DSM-IV as a precursor of PTSD. If the re-experiencing, avoidance, and hyperarousal symptoms persist for more than 4 weeks, PTSD is diagnosed.

Classification

ICD-10 classifies acute stress reactions (F43.0) among the reactions to severe stress and adjustment disorders (F43) that are primarily caused by stressful events. DSM-IV classifies acute stress disorder (308.3) among the anxiety disorders, like PTSD (see also Chapter 4.6.2).

Diagnosis and differential diagnosis

The main diagnostic criteria for acute stress reactions (ICD-10) and acute stress disorder (DSM-IV) are compared in Table 4.6.1.1.

Stressor criterion

Both ICD-10 and DSM-IV require that acute stress responses must occur in the immediate aftermath of an exceptionally stressful event. ICD-10 uses a broad concept of what qualifies as an 'exceptional mental or physical stressor'. This includes stressors that would be regarded as traumatic (e.g. rape, criminal assault, natural catastrophe) as well as unusually sudden changes in the social position and/or network of the individual (e.g. domestic fire or multiple bereavement). In contrast, DSM-IV uses a narrow definition of stressors that lead to acute stress disorder, which is identical to the stressor criterion of PTSD. It requires (i) that the traumatic event must have involved actual or threatened death or serious injury, or a threat to the physical integrity of self or others, and (ii) that the person's response to the traumatic event must have involved intense fear, helplessness, or horror (or disorganized or agitated behaviour in children) (see Chapter 4.6.2 for the rationale underlying this definition).

Symptom patterns

As shown in Table 4.6.1.1, the diagnostic criteria for acute stress reactions (ICD-10) and acute stress disorder (DSM-IV) overlap, in that they include symptoms of dissociation, anxiety, and hyperarousal. DSM-IV puts a much greater emphasis on dissociation, requiring a minimum of three of the dissociative symptoms specified in Table 4.6.1.1 (Criterion B). According to ICD-10, any combination of a minimum of four symptoms of generalized anxiety disorder (specified in Criterion C, Table 4.6.1.1) would be sufficient to establish the diagnosis of acute stress reaction. In addition, DSM-IV, but not ICD-10, requires the individual to have at least one re-experiencing symptom, to show marked avoidance of reminders of the trauma, and to experience significant distress or impairment of functioning.

In contrast to DSM-IV, ICD-10 distinguishes between mild, moderate, and severe forms of acute stress reactions on the basis

Table 4.6.1.1 Comparison of the criteria for acute stress reaction (ICD-10) and acute stress disorder (DSM-IV)

	Acute stress reaction (ICD-10 research diagnostic criteria)	Acute stress disorder (DSM-IV)
Stressor	Exposure to exceptional mental or physical stress	(1) Exposure to event involving actual or threatened death or serious injury to self or others (2) Experience of fear, helplessness, or horror
Symptoms	*Criterion C: Symptoms of generalized anxiety disorder (at least 4 symptoms)* Palpitations, sweating, trembling, dry mouth, difficulty breathing, choking, chest pain, nausea, dizziness, derealization or depersonalization, fear of losing control, fear of dying, hot flushes, numbness or tingling, muscle tension, restlessness, keyed up, difficulty swallowing, exaggerated startle response, difficulty concentrating, irritability, difficulty getting to sleep *Criterion C: Additional symptoms to determine severity* Social withdrawal, narrowed attention, disorientation, aggression, hopelessness, overactivity, excessive grief	*Criterion B: Dissociation (at least 3 symptoms)* Numbing, reduced awareness, derealization, depersonalization, dissociative amnesia *Criterion C: Re-experiencing* (at least one *symptom)* Recurrent images, thoughts, dreams, illusions, flashbacks, reliving, distress on exposure *Criterion D: Marked avoidance* Avoidance of thoughts, feelings, conversations, activities, places, people associated with the trauma *Criterion E: Marked anxiety or increased arousal* Difficulty sleeping, irritability, poor concentration, hypervigilance, exaggerated startle response, motor restlessness *Criterion F: Clinically significant distress or impairment in functioning*
Time from trauma	Onset within 1 h	Onset within 4 weeks; lasts for at least 2 days
Time course	Transient; symptoms begin to diminish within 48 h	May result in post-traumatic stress disorder
Relationship to post-traumatic stress disorder	Alternative diagnosis	Precursor
Diagnostic group	Reactions to severe stress	Anxiety disorder
Exclusion criteria	No other concurrent (within last 3 months) mental or behavioural disorder, except for generalized anxiety disorder or personality disorder	(1) Not due to effects of a substance or general medical condition (2) Not better accounted for by brief psychotic disorder (3) Not merely exacerbation of pre-existing Axis I or Axis II disorder

of additional symptoms (Criterion C, additional symptoms, Table 4.6.1.1) such as social withdrawal, hopelessness, or excessive grief. A mild severity is stipulated when none of these symptoms are present, moderate when two are reported, and severe when four are reported or when there is dissociative stupor.

Time course of symptoms

The two diagnoses cover distinct periods on a continuum from transient to more persistent symptoms. Specifically, to meet the criteria for an acute stress reaction (ICD-10), symptoms must be manifest within 1 h of the stressor (Criterion B) and begin to diminish after no more than 8 h for a transient stressor and after no more than 48 h for an enduring stressor (Criterion D).

The diagnostic criteria for acute stress disorder (DSM-IV) require that the disturbance must last for a minimum of 2 days and a maximum of 4 weeks post-trauma, after which a diagnosis of PTSD can be considered.

Assessment instruments

There are two recognized clinician-administered and two self-report measures of acute stress disorder (DSM-IV) available. As yet, there are no established standardized assessment instruments for transient acute stress reactions (ICD-10).

(a) Acute stress disorder interview

This structured clinical interview establishes the presence or absence of 19 symptoms of acute stress disorder.[9] The sum of the symptoms scored as being present indicates acute stress disorder severity. This measure has very good internal consistency ($r = 0.90$), and, with clinician-based diagnoses as the criterion, very good sensitivity (91 per cent) and specificity (93 per cent). Test–retest reliability is strong ($r = 0.88$).

(b) Structured clinical interview for DSM-IV (SCID[10])

The SCID interview indexes the presence, absence, or subthreshold presence of each acute stress disorder symptom specified in DSM-IV. An advantage of employing this interview is that it provides a comprehensive assessment of the differential diagnoses and comorbid disorders that can be present in trauma populations.

(c) Stanford acute stress reaction questionnaire[11]

This self-report inventory asks patients to rate the frequency of a range of dissociative, intrusive, somatic anxiety, hyperarousal, attention disturbance, and sleep disturbance symptoms. The questionnaire has very good internal consistency (Cronbach's alpha = 0.90 and 0.91 for dissociative and anxiety symptoms, respectively) and concurrent validity with scores on the Impact of Event Scale ($r = 0.52$ to 0.69).[12,13] It can be employed as a measure of the severity of symptoms, but does not allow the diagnosis of acute stress disorder to be established as it has not yet been validated against clinician diagnoses.

(d) Acute stress disorder scale[14]

This self-report scale is scored on a 5-point scale that reflects degree of severity of 19 acute stress disorder symptoms. The Acute Stress Disorder Scale possesses good sensitivity (95 per cent) and specificity (83 per cent) relative to the Acute Stress Disorder Interview. Test–retest reliability with a re-administration interval of 2 to 7 days is strong ($r = 0.94$).[14]

Differential diagnoses

Both ICD-10 and DSM-IV require that the symptoms are not merely an exacerbation of a pre-existing disorder. In addition, a number of alternative diagnoses need to be considered.

(a) Post-traumatic stress disorder

In ICD-10, PTSD is conceptualized as an alternative diagnosis of acute stress reactions. The definitions of acute stress reaction and PTSD differ in terms of the stressor criterion (exceptionally stressful life event versus exceptionally threatening or catastrophic event), the time course (symptoms start to diminish within 48 h versus no time limit), and symptom pattern (PTSD, but not acute stress reaction, includes involuntary re-experiencing the traumatic event).

In DSM-IV, acute stress disorder can be distinguished from PTSD by the time-frame covered by the diagnoses. Acute stress disorder refers to the period from 2 days to 1 month post-trauma, after which a diagnosis of PTSD can be considered. The primary difference between the symptom criteria for acute stress disorder and PTSD in DSM-IV is the former's emphasis on dissociative reactions.

(b) Adjustment disorder

This diagnosis covers a wide range of emotional or behavioural symptoms indicative of distress, which are judged to be out of proportion to the stressor experienced. This broad coverage can be contrasted with (i) the specific set of symptoms described by the acute stress disorder and acute stress reaction criteria, and (ii) the stipulation that the stressor involves both a threat to life and a subjective response of fear for the acute stress disorder and an exceptional stressor in the case of acute stress reaction.

(c) Brain injury

A number of acute stress disorder symptoms overlap with symptoms of brain injury including reduced awareness, depersonalization, derealization, irritability, and concentration difficulties.[15] While results from neuropsychological and neurological investigations may assist in the differential diagnosis, there appear to be a group of individuals with a mild head injury for whom there are no known tools to differentiate whether the disturbance is due to brain injury or acute stress disorder, or whether both are present.

(d) Brief psychotic disorder

When there is one or more psychotic symptoms present after experiencing an extreme stressor, the brief psychotic disorder diagnosis should be considered.

(e) Dissociative disorders

Given the emphasis on dissociative symptoms in acute stress disorder, it needs to be distinguished from dissociative amnesia and depersonalization disorder. The criteria for these diagnoses stipulate that if the amnesia or depersonalization can be accounted for by acute stress disorder then a dissociative disorder cannot be diagnosed (see Chapter 5.2.4).

Epidemiology

Incidence

There is little research into what proportion of people develop acute stress reactions to severe stress. In a study of accident survivors, 14 per cent experienced a response pattern characterized by

derealization, and a further 17 per cent exhibited strong anxiety or dysphoria.[16]

Estimates of the incidence of acute stress disorder range from about 14 per cent in motor vehicle accident survivors to 33 per cent in witnesses of a mass shooting.[17] Given the variable procedures and assessment tools employed across studies, it is difficult to determine whether the different rates of acute stress disorder detected are attributable to differences in method or in the type of trauma.

Comorbidity

Data on comorbidity are sparse. Given the similarities between acute stress disorder and PTSD it is likely that the conditions found to be comorbid with PTSD, in particular depression and substance abuse, will be applicable to acute stress disorder (see Chapter 4.6.2).

Aetiology

Both psychological and biological theories have attempted to explain the symptoms of acute stress disorder. They overlap largely with theories of PTSD (see Chapter 4.6.2). Given that acute stress reaction describes a transient disturbance, there are no specific theories of acute stress reactions as defined in ICD-10.

Psychological theories

The psychological mechanism that has received the most attention in relation to acute stress disorder is dissociation, as reflected in the DSM-IV criteria. It has been argued that dissociation minimizes the adverse emotional consequences of trauma by restricting awareness of the experience to avoid overwhelming fear and loss of control.[13] Dissociation is thought to prevent recovery because it prevents the integration of the traumatic experience into existing schemas[18] and it prevents the full activation of the trauma memory which is thought to be necessary for its modification.[19] In line with these hypotheses, dissociation during or immediately after a traumatic event predicts PTSD.[20] In contrast, an alternate view posits that dissociation at the time of trauma may serve a protective function because it may limit the encoding of aversive experiences and this may assist adaptation.[21] Consistent with this view, there is evidence that persisting dissociation (which is a form of cognitive avoidance) is more predictive of PTSD than dissociation that occurs at the time of trauma.[22,23]

Psychological theories of acute stress disorder and PTSD focus on the personal meaning of the trauma and its consequences, and characteristics of the trauma memory. Several hypotheses about the problems in memory that are responsible for the characteristic re-experiencing symptoms (i.e. unwanted memories of aspects of the trauma that occur in response to a wide range of stimuli) have been suggested (see also Chapter 4.6.2). Foa and colleagues[24,25] suggested that PTSD is characterized by a pathological network in memory that is particularly large and easily triggered. It contains many stimulus propositions that are erroneously linked to danger, causing fear responses to harmless stimuli that are associated with the traumatic event in memory. Brewin *et al.*[26] postulated that two different representations of the trauma are formed in memory. The first, termed verbally accessible memory, contains the conscious recollection of the trauma. The second memory representation, termed situationally accessible memory, comprises sensory, physiological, and motor aspects of the trauma in the form of codes that enable the re-experiencing of the original experience. Ehlers and Clark[27] suggested three memory processes to explain that a wide range of stimuli can trigger vivid memories and strong emotional responses, which are experienced as if the traumatic event was happening at present. First, trauma memory is thought to be inadequately linked to its context in memory, which leads to poor inhibition of stimulus-driven retrieval. Two other basic memory processes, perceptual priming and associative learning, are thought to further enhance the chances of stimulus-driven retrieval of memories. Consistent with these psychological theories, there is evidence that chronic PTSD is predicted by impaired access to autobiographical memories[28] and the perceived 'nowness' of trauma memories.[29] There is also evidence that maladaptive appraisals such as 'I am inadequate', 'My reactions since the trauma mean I am losing my mind', or 'I have permanently changed for the worse' in the acute phase after trauma exposure predict chronic PTSD.[30,31]

Psychological models also concur that successful adaptation to trauma involves integration of corrective information, and any strategies that minimize this process will contribute to subsequent PTSD. Excessive use of such avoidant strategies (e.g. trying not to think about the trauma, efforts to push intrusive memories out of one's mind, ruminating about how the trauma could have been avoided) prevent recovery.[24–27] There is preliminary empirical evidence supporting this hypothesis.[30–33]

Biological theories

Biological models have focused on fear conditioning and progressive neural sensitization in the weeks after trauma.[34] Specifically, when a traumatic event (unconditioned stimulus) occurs, people typically respond with fear (unconditioned response). It is argued that the strong fear elicited by the trauma will lead to strong associative conditioning between the fear and the stimuli surrounding the trauma. As reminders of the trauma occur (conditioned stimuli), people then respond with fear reactions (conditioned response). It has been hypothesized that extreme sympathetic arousal at the time of a traumatic event may result in the release of stress neurochemicals (including norepinephrine and epinephrine) into the cortex, mediating an overconsolidation of trauma memories. It is possible that sensitization occurs as a result of repetitive activation by trauma reminders and re-experiencing symptoms, elevating sensitivity of limbic networks, and that as time progresses these responses become increasingly conditioned to trauma-related stimuli.[35] In support of these proposals, there is evidence that people who eventually develop PTSD display elevated resting heart rates in the initial week after trauma.[36] There is also evidence that most people with acute stress disorder suffer panic attacks during the traumatic experience, and most of these people continue to suffer ongoing panic attacks in the subsequent month.[37]

Course and prognosis

Time course of symptoms

Whereas the ICD-10 criteria define an acute stress reaction as a disorder that remits within a few days, DSM-IV conceptualizes acute stress disorder as a marker of those vulnerable to the development of PTSD. Evidence relating to these different assumptions was sparse at the time the diagnoses were established. One explicit goal of the acute stress disorder diagnosis is to identify people who

will develop PTSD. This goal is difficult to achieve because most people who initially display acute stress reactions adapt in the following 3 months.[38] Across 12 studies that have assessed the relationship between acute stress disorder and PTSD, most studies have found that whereas most people with acute stress disorder do develop PTSD, the acute stress disorder diagnosis does not capture the majority of people who develop PTSD.[38] It appears that the requirement for dissociative symptoms in the acute phase to be present results in a failure to identify many people who are high risk for PTSD development. It is important to note that similar patterns have been noted in prospective studies of children after trauma.[39–41] Across studies, people who have more severe symptoms of PTSD in the weeks following trauma have a poorer prognosis than those with less severe symptoms.[42,43]

Predictors of acute stress disorder

Little is known about predictors of acute stress reactions. A history of psychiatric disorder, depressive and dissociative symptoms prior to the traumatic event, and previous trauma predict acute stress disorder.[22,44,45]

Treatment

Psychological treatments

(a) Debriefing

Critical incident stress debriefing is a widely practised intervention that has the goal of promoting adaptation to traumatic events. Debriefing is generally conducted in a group within 24 to 72 h of the trauma. However, these parameters have been modified to permit more flexible interventions. Mitchell[46] proposes that debriefing comprises seven phases:

1 an initial outline of the purpose and benefits of debriefing;

2 the fact phase, in which participants relate what happened to them;

3 a thought phase, in which participants relate their initial thoughts after the critical incident;

4 a feeling phase, which requires participants to focus on the worst aspects of the incident and engage in their emotional reactions to the incident;

5 an assessment phase, in which participants are trained to note their physical, cognitive, emotional, and behavioural symptoms;

6 an education phase, which provides information about stress responses and ways to manage them;

7 the re-entry phase, in which the information given is summarized and referral information offered.

These phases may take 1 to 5 h, and are usually coordinated by a trained mental health professional.

Anecdotal evidence and clinical reports attest to the efficacy of debriefing. However, despite its widespread use, few controlled trials have been conducted. These have mainly focused on single session individual debriefing. A Cochrane review of 15 randomized controlled studies of psychological debriefing[47] found that although participants usually found the intervention useful, debriefing did not prevent the onset of PTSD nor reduce psychological distress compared to control. The two studies with the longest follow-up actually found that the debriefing group had a worse long-term outcome than the control group.[48–50] These results suggest that single session individual debriefing is not effective.[47]

In line with these results on individual debriefing, the first two non-randomized controlled studies of the efficacy of group debriefing found that the intervention had no beneficial effects on post-trauma symptoms.[51,52] One of the studies found negative effects of the intervention after 18 months.[52] A large trial of 1050 soldiers on a peacekeeping mission found no differences on all outcomes between a full programme of critical incident stress debriefing, stress education, and no intervention.[53]

(b) Cognitive behaviour therapy

Cognitive behavioural interventions are effective in treating PTSD (see Chapter 4.6.2). The results of randomized controlled studies of rape victims and road traffic accident survivors suggest that a brief five-session version of this treatment is effective in acute stress disorder and prevents the development of chronic post-trauma reactions.[54–58] Treatment involved the following:

1 education about trauma reactions;

2 progressive muscle relaxation;

3 prolonged exposure;

4 cognitive restructuring of fear-related beliefs;

5 graded *in vivo* exposure.

Cognitive behaviour therapy was found to be superior to other psychological therapies such as non-directive therapy or relaxation in preventing chronic PTSD.[59,60]

Psychopharmacological treatment

Little is known about which pharmacological interventions are effective in acute stress reactions/disorder. Preliminary studies report the utility of tricyclic antidepressants, but no or even harmful effects of benzodiazepines.[59–61] Research on PTSD suggests that selective serotonin reuptake inhibitors are, to date, the best pharmacological treatment for persistent reactions to traumatic stress (see Chapter 4.6.2).

There is growing interest in pharmacological interventions in the acute phase following the trauma that may prevent the development of PTSD symptoms.[62] A pilot study attempted to prevent PTSD by administering propranolol (a β-adrenergic blocker) within 6 h of trauma exposure.[63] This approach is based on evidence that propranolol abolishes the epinephrine enhancement of fear conditioning.[64] Although propranolol did not result in reduced PTSD relative to a placebo condition, patients receiving propranolol displayed less reactivity to trauma reminders 3 months later. This outcome accords with an uncontrolled study that found that propranolol administered immediately after trauma resulted in reduced PTSD 3 months later.[65] These preliminary data suggest that propranolol administration shortly after trauma exposure may limit fear conditioning that may contribute to subsequent PTSD development. Two other pilot studies found that high doses of hydrocortisol given to medical patients in an intensive care environment had a beneficial effect on subsequent PTSD symptoms.[62,66] One possible pathway of action is that cortisol contains strong epinephrine responses during stress, and may thus indirectly influence the strength of fear conditioning.[62] However, these experimental studies are as yet too preliminary to suggest clinical application.[60]

Information and self-help booklets

Several studies have evaluated the effectiveness of information or self-help booklets as early interventions after trauma. They consistently found that such interventions are ineffective and do not decrease the risk of chronic PTSD symptoms, although patients report that they find the booklets helpful.[67,68]

Advice about management

As acute stress reactions (ICD-10) are transient, and trials showed that early psychological debriefing is not effective in reducing future symptoms, psychological interventions that focus on recounting the traumatic event and ventilation of feelings are not indicated in the initial days after trauma exposure.[59,60] Instead, clinicians should focus on ensuring trauma survivors' safety and security, providing support and practical assistance, and encouraging them to actively use their social support.[59,60] Although drug treatments are not recommended as a preventative intervention following traumatic exposure,[60] clinicians may consider short-term hypnotic medication or longer term use of antidepressants for the management of significant sleep disturbance in the acute phase after trauma.[59,61] Furthermore, patients may find information about common reactions to trauma and their course, and practical advice about issues such as hospital procedures, police questioning, insurance claims, legal procedures, and media pressure to tell one's story helpful. Clinicians should offer follow-up appointments and monitor trauma survivors for the development of PTSD.[59]

An important clinical task is the early identification of those trauma survivors who are likely to develop chronic and disabling post-trauma symptoms.[47,59] If symptoms persist for 2 weeks or longer, treatment may be indicated. However, patients with low symptom severity at 2 weeks have a good chance of recovering without intervention, and clinicians may want to initially monitor symptom progression for a month to determine whether the patient is recovering naturally.[59] A number of short symptom screening questionnaires have been developed that may provide useful predictive information for this purpose in the future, but still await cross-validation in other samples.[69]

For trauma survivors with more severe acute stress disorder symptoms, a course of cognitive behavioural treatment should be considered.[59,60] This form of intervention is typically not provided within 2 weeks of trauma exposure. If the patient is offered cognitive behavioural treatment, therapists need to be aware that avoidance is a hallmark symptom of acute stress disorder and may reduce the likelihood that an individual will attend treatment sessions regularly. Flexible treatment procedures (e.g. initial contacts by telephone, scheduling sessions around the patient's preferences) and discussions about the ambivalence towards treatment may be helpful. Therapists need to be knowledgeable about the conditions surrounding the traumatic event and be sensitive to the particular socio-cultural background of the patient, which will affect the personal meaning of the event. Other forms of psychological treatment that do not address traumatic memories (such as relaxation or non-directive therapy) should not be routinely offered.[59]

The lack of randomized controlled trials suggests that pharmacological treatment cannot be considered a front-line treatment for acute stress disorder,[59,60] but research on PTSD suggests that selective serotonin reuptake inhibitors and other antidepressants may be helpful (see Chapter 4.6.2).

Possibilities for prevention

Identifying highly symptomatic individuals with acute stress disorder and providing a cognitive behavioural intervention from 2 weeks post-trauma onwards may reduce the risk of chronic PTSD. Additional preventive methods have been explored that prepare individuals 'at risk' (e.g. emergency services and military personnel) for experiencing trauma so as to enhance their coping strategies and reduce the risk of them developing longer-term symptomatology. For those individuals at high risk of experiencing a trauma, providing them with training to remain calm, evaluate the situation objectively,[70] to not identify with victims, to utilize social supports, and to express emotional reactions[71] have all been found to be associated with better coping after the trauma. However, evidence remains preliminary and it remains unclear whether they affect the risk of PTSD.

Further information

McNally, R.J., Bryant, R.A., and Ehlers, A. (2003). Does early psychological intervention promote recovery from posttraumatic stress? *Psychological Science in the Public Interest*, **4**, 45–79.

National Institute of Clinical Excellence. (2005). *Clinical guideline 26: post-traumatic stress disorder: The management of PTSD in adults and children in primary and secondary care*. http://guidance.nice.org.uk/CG26

National Center for Posttraumatic Stress Disorder. http://www.ncptsd.va.gov/ncmain/providers/ and PILOTS database http://www.ncptsd.va.gov/ncmain/publications/pilots/

Australian Centre for Posttraumatic Mental Health. (2007). *Australian guidelines for the treatment of acute stress disorder and posttraumatic stress disorder*. http://www.acpmh.unimelb.edu.au/Resources/guidelines/ACMPH_PractionerGuideForASDandPTSD.pdf

References

1. Lindemann, E. (1944). Symptomatology and management of acute grief. *The American Journal of Psychiatry*, **101**, 141–8.
2. Sargent, W. and Slater, E. (1941). Amnesic syndromes in war. *Proceedings of the Royal Society for Social Medicine*, **34**, 757–74.
3. World Health Organization. (1948). *International statistical classification of diseases, injuries, and causes of death* (6th revision). World Health Organization, Geneva.
4. World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorder: diagnostic criteria for research* (10th revision). World Health Organization, Geneva.
5. American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd edn). American Psychiatric Association, Washington, DC.
6. American Psychiatric Association. (1989). *Diagnostic and statistical manual of mental disorders* (3rd edn, revised). American Psychiatric Association, Washington, DC.
7. Pincus, H.A., Frances, A., Davis, W.W., *et al.* (1992). DSM-IV and new diagnostic categories: holding the line on proliferation. *The American Journal of Psychiatry*, **149**, 112–17.
8. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
9. Bryant, R.A., Harvey, A.G., Dang, S., *et al.* (1998). Assessing acute stress disorder: psychometric properties of a structured clinical interview. *Psychological Assessment*, **10**, 215–20.
10. First, M.B., Spitzer, R.L., Gibbon, M., *et al.* (1995). *Structured clinical interview for DSM-IV Axis I disorders-patient version* (*SCID-I/P, Version 2.0*). Biometrics Research Department of the New York State Psychiatric Institute, New York.

11. Cardeña, E., Classen, C., and Spiegel, D. (1991). *Stanford acute stress reaction questionnaire*. Stanford University Medical School, Stanford, CA.
12. Koopman, C., Classen, C., and Spiegel, D. (1994). Predictors of post-traumatic stress symptoms among survivors of the Oakland/Berkeley, Calif. firestorm. *The American Journal of Psychiatry*, **151**, 888–94.
13. Spiegel, D. (1991). Dissociation and trauma. In *American psychiatric press review of psychiatry*, Vol. 10 (eds. A. Tasman and S.M. Goldfinger), pp. 261–75. American Psychiatric Press, Washington, DC.
14. Bryant, R.A., Moulds, M., and Guthrie, R. (2000). Acute Stress Disorder Scale: a self-report measure of acute stress disorder. *Psychological Assessment*, **12**, 61–8.
15. Harvey, A.G. and Bryant, R.A. (1998). Acute stress disorder following mild traumatic brain injury. *The Journal of Nervous and Mental Disease*, **186**, 333–7.
16. Schnyder, U. and Malt, U.F. (1998). Acute stress response patterns to accidental injuries. *Journal of Psychosomatic Research*, **45**, 419–24.
17. Harvey, A.G. and Bryant, R.A. (2002). Acute stress disorder: a synthesis and critique. *Psychological Bulletin*, **128**, 886–902.
18. Koopman, C., Classen, C., Cardena, E., *et al.* (1995). When disaster strikes, acute stress disorder may follow. *Journal of Traumatic Stress*, **8**, 29–46.
19. Foa, E.B. and Hearst-Ikeda, D. (1996). Emotional dissociation in response to trauma: an information-processing approach. In *Handbook of dissociation: theoretical and clinical perspectives* (eds. L.K. Michelson and W.J. Ray), pp. 207–22. Plenum Press, New York.
20. Ozer, E.J., Best, S., Lipzey, T.L., *et al.* (2003). Predictors of posttraumatic stress disorder and symptoms in adults: a meta-analysis. *Psychological Bulletin*, **129**, 52–73.
21. Bryant, R.B. (2007). Does dissociation further our understanding of PTSD? *Journal of Anxiety Disorders*, **21**, 183–91**.**
22. Murray, J., Ehlers, A., and Mayou, R.A. (2002). Dissociation and posttraumatic stress disorder: two prospective studies of road traffic accident victims. *The British Journal of Psychiatry*, **180**, 363–8.
23. Briere, J., Scott, C., and Weathers, F. (2005). Peritraumatic and persistent dissociation in the presumed etiology of PTSD. *The American Journal of Psychiatry*, **162**, 2295–301.
24. Foa, E.B. and Riggs, D.S. (1993). Post-traumatic stress disorder in rape victims. In *Annual review of psychiatry*, Vol. 12 (eds. J. Oldham, M.B. Riba, and A. Tasman), pp. 273–303. American Psychiatric Association, Washington, DC.
25. Foa, E.B., Steketee, G., and Rothbaum, B.O. (1989). Behavioral/cognitive conceptualizations of post-traumatic stress disorder. *Behavior Therapy*, **20**, 155–76.
26. Brewin, C.R., Dalgleish, T., and Joseph, S. (1996). A dual representation theory of post-traumatic stress disorder. *Psychological Review*, **103**, 670–86.
27. Ehlers, A. and Clark, D.M. (2000). A cognitive model of post-traumatic stress disorder. *Behaviour Research and Therapy*, **38**, 319–45.
28. Harvey, A.G., Bryant, R.A., and Dang, S.T. (1998). Autobiographical memory in acute stress disorder. *Journal of Consulting and Clinical Psychology*, **66**, 500–6.
29. Michael, T., Ehlers, A., Halligan, S., *et al.* (2005). Unwanted memories of assault: what intrusion characteristics predict PTSD? *Behaviour Research and Therapy*, **43**, 613–28.
30. Ehring, T., Ehlers, A., and Glucksman, E. (in press). Do cognitive models help in predicting posttraumatic stress disorder, phobia and depression after trauma? *Journal of Consulting and Clinical Psychology*.
31. Kleim, B., Ehlers, A., and Glucksman, E. (2007). Prediction of chronic posttraumatic stress disorder after assault. *Psychological Medicine*, **37**, 1457–68.
32. Ehlers, A., Mayou, R.A., and Bryant, B. (1998). Psychological predictors of chronic post-traumatic stress disorder after motor vehicle accidents. *Journal of Abnormal Psychology*, **107**, 508–19.
33. Harvey, A.G. and Bryant, R.A. (1998). The effect of attempted thought suppression in acute stress disorder. *Behaviour Research and Therapy*, **36**, 583–90.
34. Charney, D.S., Deutch, A.Y., Krystal, J.H., *et al.* (1993). Psychobiologic mechanisms of posttraumatic stress disorder. *Archives of General Psychiatry*, **50**, 294–305.
35. Post, R.M., Weiss, S.R.B., and Smith, M.A. (1995). Sensitization and kindling: implications for the evolving neural substrates of post-traumatic stress disorder. In *Neurobiological and clinical consequences of stress: from normal adaptation to post-traumatic stress disorder* (eds. M.J. Friedman, D.S. Charney, and A.Y. Deutch), pp. 203–24. Lippincott Williams and Wilkins, Philadelphia, PA.
36. Bryant, R.A. (2006). Longitudinal psychophysiological studies of heart rate: mediating effects and implications for treatment. *Annals of the New York Academy of Sciences*, **107**, 19–26.
37. Nixon, R. and Bryant, R.A. (2003). Peritraumatic and persistent panic attacks in acute stress disorder. *Behaviour Research and Therapy*, **41**, 1237–42.
38. Bryant, R.A. (2003). Early predictors of posttraumatic stress disorder. *Biological Psychiatry*, **53**, 789–95.
39. Kassam-Adams, N. and Winston, F.K. (2004). Predicting child PTSD: the relationship between acute stress disorder and PTSD in injured children. *Journal of the American Academy of Child and Adolescent Psychiatry*, **43**, 403–11.
40. Meiser-Stedman, R., Yule, W., Smith, P., *et al.* (2005). Acute stress disorder and posttraumatic stress disorder in children and adolescents involved in assaults or motor vehicle accidents. *The American Journal of Psychiatry*, **162**, 1381–3.
41. Bryant, R.A., Salmon, K., Sinclair, E., *et al.* (2007). The relationship between acute stress disorder and posttraumatic stress disorder in injured children. *Journal of Traumatic Stress,* **20**, 1075–9.
42. Rothbaum, B.O., Foa, E.B., Riggs, D.S., *et al.* (1992). A prospective examination of post-traumatic stress disorder in rape victims. *Journal of Traumatic Stress*, **5**, 455–75.
43. Blanchard, E.B., Hickling, E.J., Barton, K.A., *et al.* (1996). One-year prospective follow-up of motor vehicle accident victims. *Behaviour Research and Therapy*, **34**, 775–86.
44. Barton, K.A., Blanchard, E.B., and Hickling, E.J. (1996). Antecedents and consequences of acute stress disorder among motor vehicle accident victims. *Behaviour Research and Therapy*, **34**, 805–13.
45. Harvey, A.G. and Bryant, R.A. (1999). Predictors of acute stress following motor vehicle accidents. *Journal of Traumatic Stress*, **12**, 519–26.
46. Mitchell, J. (1983). When disaster strikes. The critical incident stress debriefing process. *Journal of Emergency Medical Services*, **8**, 36–9.
47. Rose, S., Bisson, J., Churchill, R., *et al.* (2002). Psychological debriefing for preventing posttraumatic stress disorder (PTSD). *Cochrane Database of Systematic Reviews*, (2), CD000560, DOI: 10.1002/14651858.CD000560.
48. Bisson, J.I., Jenkins, P.L., Alexander, J., *et al.* (1997). Randomised controlled trial of psychological debriefing for victims of acute burn trauma. *The British Journal of Psychiatry*, **171**, 78–81.
49. Hobbs, M., Mayou, R., Harrison, B., *et al.* (1996). A randomised controlled trial of psychological debriefing for victims of road traffic accidents. *British Medical Journal*, **7**, 1438–9.
50. Mayou, R.A., Ehlers, A., and Hobbs, M. A (2002) three year follow-up of a randomised controlled trial of psychological debriefing for road traffic accident victims. *The British Journal of Psychiatry*, **176**, 589–93.
51. Kenardy, J.A., Webster, R.A., Lewin, T.J., *et al.* (1996). Stress debriefing and patterns of recovery following a natural disaster. *Journal of Traumatic Stress*, **9**, 37–49.
52. Carlier, I.V.E., Lamberts, R.D., van Uchelen, A.J., *et al.* (1998). Disaster-related post-traumatic stress in police officers: a field study of the impact of debriefing. *Stress Medicine*, **14**, 143–8.

53. Litz, B. (2004). Cited in Rose, S., Bisson, J., Churchill, R., and Wessely, S. (2002). Psychological debriefing for preventing posttraumatic stress disorder (PTSD) (Review). *Cochrane Database of Systematic Reviews*, (2), CD000560, DOI: 10.1002/14651858. CD000560, p. 8.
54. Bryant, R.A., Harvey, A.G., Dang, S.T., *et al.* (1998). Treatment of acute stress disorder: a comparison of cognitive behaviour therapy and supportive counselling. *Journal of Consulting and Clinical Psychology*, **66**, 862–6.
55. Bryant, R.A., Sackville, T., Dang, S.T., *et al.* (1999). Treating acute stress disorder: an evaluation of cognitive behavior therapy and counseling techniques. *The American Journal of Psychiatry*, **156**, 1780–6.
56. Bryant, R.A., Moulds, M.L., Guthrie, R., *et al.* (2003). Treating acute stress disorder after mild brain injury. *The American Journal of Psychiatry*, **160**, 585–7.
57. Bryant, R.A., Moulds, M.L., Guthrie, R.M., *et al.* (2005). The additive benefit of hypnosis and cognitive behavior therapy in treating acute stress disorder. *Journal of Consulting and Clinical Psychology*, **73**, 334–40.
58. Bisson, J.I., Shepherd, J.P., Joy, D., *et al.* (2004). Early cognitive behavioural therapy for post-traumatic stress symptoms after physical injury. *The British Journal of Psychiatry*, **184**, 63–9.
59. National Institute of Clinical Excellence. (2005). *Clinical guideline 26: post-traumatic stress disorder: the management of PTSD in adults and children in primary and secondary care*. http://guidance.nice.org.uk/CG26
60. Australian Centre for Posttraumatic Mental Health. (2007). *Australian guidelines for the treatment of acute stress disorder and posttraumatic stress disorder*. http://www.acpmh.unimelb.edu.au/Resources/guidelines/ACMPH_PractionerGuideForASDandPTSD.pdf
61. Davidson, J.R.T. (2006). Pharmacologic treatment of acute and chronic stress following trauma. *The Journal of Clinical Psychiatry*, **67**(Suppl. 2), 34–9.
62. Pitman, R.K. and Delahanty, D.L. (2005). Conceptually driven pharmacologic approaches to acute trauma. *CNS Spectrums*, **10**, 99–106.
63. Pitman, R.K., Sanders, K.M., Zusman, R.M., *et al.* (2002). Pilot study of secondary prevention of posttraumatic stress disorder with propranolol. *Biological Psychiatry*, **51**, 189–92.
64. Cahill, L., Prins, B., Weber, M., *et al.* (1994). Beta-adrenergic activation and memory for emotional events. *Nature*, **371**, 702–4.
65. Vaiva, G., Ducrocq, F., Jezequel, K., *et al.* (2003). Immediate treatment with propranolol decreases posttraumatic stress disorder two months after trauma. *Biological Psychiatry*, **54**, 947–9.
66. Schelling, G., Kilger, E., and Roozendaal, B. (2004). Stress doses of hydrocortisone, traumatic stress, and symptoms of posttraumatic stress disorder in patients after cardiac surgery: a randomized trial. *Biological Psychiatry*, **55**, 627–33.
67. Ehlers, A., Clark, D.M., Hackmann, A., *et al.* (2003). A randomized controlled trial of cognitive therapy, self-help booklet, and repeated assessment as early interventions for PTSD. *Archives of General Psychiatry*, **60**, 1024–32.
68. Turpin, G., Downs, M., and Mason, S. (1994). Effectiveness of providing self-help information following acute traumatic injury: randomised controlled trial. *The British Journal of Psychiatry*, **187**, 76–82.
69. Brewin, C.R. (2005). Systematic review of screening instruments for adults at risk of PTSD. *Journal of Traumatic Stress*, **18**, 53–62.
70. Hytten, K. (1989). Helicopter crash in water: effects of simulator escape training. *Acta Psychiatrica Scandinavica*, **80**(Suppl. 355), 73–8.
71. Fullerton, C.S., McCarroll, J.E., Ursano, R.J., *et al.* (1992). Psychological responses of rescue workers: fire fighters and trauma. *The American Journal of Orthopsychiatry*, **62**, 371–8.

4.6.2 Post-traumatic stress disorder

Anke Ehlers

Introduction

Clinicians have long noted that traumatic events can lead to severe psychological disturbance. At the end of the nineteenth and the beginning of the twentieth centuries, railway disasters, the World Wars, and the Holocaust prompted systematic descriptions of the symptoms associated with traumatic stress reactions. These include the spontaneous re-experiencing of aspects of the traumatic events, startle responses, irritability, impairment in concentration and memory, disturbed sleep, distressing dreams, depression, phobias, guilt, psychic numbing, and multiple somatic symptoms. A variety of labels were used to describe these reactions including 'fright neurosis', 'combat/war neurosis', 'shell shock', 'survivor syndrome', and 'nuclearism'.[1–3]

Whether the traumatic event can be considered a major cause of these psychological symptoms, has been the subject of considerable debate. Charcot, Janet, Freud, and Breuer suggested that hysterical symptoms were caused by psychological trauma, but their views were not widely accepted. The dominant view was that a traumatic event in itself was not a sufficient cause of post-trauma symptoms, and experts searched for other causes. Many suspected an organic cause. For example, damage to the spinal cord was suggested as the cause of the 'railway spine syndrome', microsections of exploded bombs entering the brain as the cause of 'shell shock', and starvation and brain damage as causes of the chronic psychological difficulties of concentration camp survivors. Others doubted the validity of the symptom reports and suggested that malingering and compensation-seeking ('compensation neurosis') was the major cause in most cases. Finally, the psychological symptoms were attributed to pre-existing psychological dysfunction. The predominant view was that reactions to traumatic events are transient, and that therefore only people with unstable personalities, pre-existing neurotic conflicts, or mental illness would develop chronic symptoms.[1–3]

It was the recognition of the long-standing psychological problems of many war veterans, especially Vietnam veterans, that changed this view and convinced clinicians and researchers that even people with sound personalities can develop clinically significant psychological symptoms if they are exposed to horrific stressors. This prompted the introduction of post-traumatic stress disorder (PTSD) as a diagnostic category in DSM-III.[4] It was thus recognized that traumatic events such as combat, rape, man-made, or natural disasters give rise to a characteristic pattern of psychological symptoms. The diagnostic criteria specified the experience of a traumatic event as a necessary condition for the diagnosis. ICD-10[5] emphasized the causal role of traumatic stressors in producing psychological dysfunction even more clearly, in that a specific group of disorders, 'reaction to severe stress, and adjustment disorders', was created. These disorders are 'thought to arise always as a direct consequence of the acute severe stress or continued trauma. The stressful event . . . is the primary and overriding causal factor, and the disorder would not have occurred without its impact'.

What makes a stressor traumatic?

In everyday language, many upsetting situations are described as 'traumatic', for example, divorce, loss of job, or failing an examination. However, a field study designed to establish what kinds of stressors lead to the characteristic symptoms of PTSD, showed that only 0.4 per cent of a community sample developed the characteristic symptoms of PTSD in response to such 'low magnitude' stressors.[6] Thus, in diagnosing PTSD, it appeared necessary to employ a strict definition of what constitutes a traumatic stressor.

Few people would contest that horrific events such as rape or bombings are traumatic. In an attempt to capture the essence of these stressors, the authors of DSM-IIIR required a traumatic stressor to be 'outside the range of usual human experience' and that it 'would be markedly distressing to almost anyone'.[7] However, epidemiological studies showed that stressors that can lead to PTSD are actually quite common, for example road traffic accidents[8] or sexual assault.[9] Thus, the DSM-IIIR definition appeared too restrictive.

ICD-10 uses a broader definition and characterizes traumatic stressors by their exceptional severity and the distress they would cause for the average person 'a stressful event or situation . . . of an exceptionally threatening or catastrophic nature, which is likely to cause pervasive distress in almost anyone'.[5] Thus, the ICD-10 diagnosis refers to a common sense understanding of which situations are likely to be extremely distressing.

In contrast, the authors of DSM-IV[10] attempted a specific definition. On the basis of research findings that threat to life or physical integrity during the event is one of the most consistent predictors of PTSD,[11] DSM-IV requires that the person 'experienced, witnessed, or was confronted with an event or events that involved actual or threatened death or serious injury, or a threat to the physical integrity of self or others'. The authors of DSM-IV made a further important step, in that they moved away from a purely situational definition and included the person's subjective response to the situation as an additional criterion, requiring that the 'person's response involved intense fear, helplessness, or horror' (or disorganized or agitated behaviour in the case of children).[10] The latter criterion takes into account that there is a large inter-individual variability in the psychological response to the same situation.

The stressor criterion of DSM-IV is still under debate. Recent research suggests that both components of the definition may require extension. First, it may be necessary to include further possible emotional responses to traumatic stressors. There is accumulating evidence that emotional numbing during traumatic events is predictive of PTSD.[12] Furthermore, it has been established that perpetrators of violent crime sometimes develop PTSD. Witnessing or participating in war-related crimes such as torturing or killing prisoners of war and civilians and mutilation of corpses is more closely linked to PTSD in Vietnam veterans than the threat of death associated with combat.[13] The psychological state of the perpetrators during the events that later lead to PTSD has not been studied in detail, but it isdoubtful that they would meet the current DSM-IV definition. Feelings of shame or guilt that were experienced at the time or subsequently, may be predictive of PTSD in this group.[14]

Second, the emphasis on threat to life or physical integrity may omit important dimensions of subgroups of traumatic events. The threat to the perception of oneself as an autonomous human being may be a relevant dimension of traumatic events that involve intentional harm by other people.[15] Mental defeat, the perceived loss of all autonomy, was related to PTSD in political prisoners and assault victims,[15,16] independent of other indicators of trauma severity including threat to life and perceived helplessness.

Clinical features

The most characteristic symptoms of PTSD are the re-experiencing symptoms. Patients involuntarily re-experience aspects of the traumatic event in a very vivid and distressing way. This includes: flashbacks in which the person acts or feels as if the event were recurring; nightmares; and intrusive images or other sensory impressions from the event. For example, a woman who was assaulted kept seeing the eyes of the perpetrator looking through the letterbox before he broke into her house, and a man involved in a severe car crash at night kept hearing the sound of the impact. Despite these vivid memory fragments, intentional recall of the event is often disorganized, and some patients have amnesia for parts of the event (see also Chapter 4.6.3).

Reminders of the trauma arouse intense distress and/or physiological reactions and are consequently avoided, including conversations about the event. Patients try to push memories of the event out of their mind and avoid thinking about the event in detail, particularly about its worst moments. On the other hand, many ruminate excessively about questions that prevent them from coming to terms with the event, for example about why the event happened to them, about how it could have been prevented, or about how they could take revenge.

The patients' emotional state ranges from intense fear, anger, sadness, guilt, or shame to emotional numbness. They often describe feeling detached from other people and give up previously significant activities. Various symptoms of hyperarousal include hypervigilance, exaggerated startle responses, irritability, difficulty concentrating, and sleep problems.

Classification

ICD-10[5] classifies PTSD (F43.1) among the reactions to severe stress and adjustment disorders (F43) that are primarily caused by stressful events. DSM-IV[10] classifies PTSD (309.81) among the anxiety disorders because symptom patterns, psychophysiological responses, family studies, and the efficacy of exposure treatment and serotonergic drugs suggested a relationship with other anxiety disorders. However, some of the symptoms would also suggest a relationship with dissociative disorders (e.g. amnesia) or depression (e.g. loss of interest).[17,18]

Diagnosis and differential diagnosis

Diagnostic criteria in ICD-10 and DSM-IV

Table 4.6.2.1 compares the diagnostic criteria of ICD-10 and DSM-IV.[10] ICD-10 research diagnostic criteria,[19] as well as diagnostic guidelines,[5] are included. The diagnostic systems agree on the core symptoms of PTSD (re-experiencing, avoidance, emotional numbing, and hyperarousal), but differ in the weight assigned to them. DSM-IV criteria are stricter.

Table 4.6.2.1 Diagnostic criteria for PTSD in ICD-10 and DSM-IV

ICD-10 diagnostic guidelines	ICD-10 research diagnostic criteria	DSM-IV criteria
Stressor criterion		
1 Event or situation of exceptionally threatening or catastrophic nature	(a) 1 Event or situation of exceptionally threatening or catastrophic nature	(a) 1 The person experienced, witnessed,or was confronted with an event or events that involved actual or threatened death or serious injury, or a threat to the physical integrity of self or others
2 Likely to cause pervasive distress in almost anyone	2 Likely to cause pervasive distress in almost anyone	2 The person's response involved intense fear, helplessness, or horror (or disorganized or agitated behaviour in children)
Symptom criteria		
Necessary symptom	*Necessary symptoms*	*Necessary symptoms*
1 Repetitive intrusive recollection or re-enactment of the event in memories, daytime imagery, or dreams	(b) Persistent remembering or 'reliving' of the stressor in intrusive 'flashbacks', vivid memories, or recurring dreams, and in experiencing distress when exposed to circumstances resembling or associated with the stressor	(b) The traumatic event is persistently re-experienced in one (or more) of the following ways 1 Recurrent and intrusive distressing recollections of the event, including images, thoughts, or perceptions (or repetitive play in which the themes or aspects of the trauma are expressed in children)
Other typical symptoms		
2 Sense of 'numbness' and emotional blunting, detachment from others, unresponsiveness to surroundings, anhedonia	(c) Actual or preferred avoidance of circumstances resembling or associated with the stressor which was not present before exposure to the stressor	2 Recurrent distressing dreams of the event (or frightening dreams without recognizable content in children)
3 Avoidance of activities and situations reminiscent of trauma	(d) 1 Inability to recall, either partially or completely, some important aspects or the period of exposure to the stressor	3 Acting or feeling as if the traumatic event were recurring (or trauma-specific re-enactment in children)
Common symptoms		4 Intense psychological distress at exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event
4 Autonomic hyperarousal with insomnia		5 Physiological reactivity at exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event
5 Anxiety and depression	*or*	
Rare symptoms		
6 Dramatic acute bursts of fear, panic, or aggression triggered by reminders	2 Persistent symptoms of increased psychological sensitivity and arousal (not present before exposure to stressor), shown by any two of the following (a) Difficulty in falling or staying asleep (b) Irritability or outbursts of anger (c) Difficulty in concentrating (d) Hypervigilance (e) Exaggerated startle response	(c) Persistent avoidance of stimuli associated with the trauma and numbing of general responsiveness (not present before trauma), as indicated by three (or more) of the following 1 Efforts to avoid thoughts, feelings, or conversations associated with the trauma 2 Efforts to avoid activities, places, or people that arouse recollections of the trauma 3 Inability to recall an important aspect of the trauma 4 Markedly diminished interest or participation in significant activities 5 Feeling of detachment or estrangement from others 6 Restricted range of affect 7 Sense of foreshortened future (d) Persistent symptoms of increased arousal (not present before the trauma), as indicated by two (or more) of the following 1 Difficulty falling or staying asleep 2 Irritability or outbursts of anger 3 Difficulty concentrating 4 Hypervigilance 5 Exaggerated startle response
Time frame		
Symptoms should usually arise within 6 months of the traumatic event	Symptoms should usually arise within 6 months of the traumatic event	Symptoms present for at least 1 month
Disability criterion		
N/A	N/A	The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning

(continued)

Table 4.6.2.1 (Continued) Diagnostic criteria for PTSD in ICD-10 and DSM-IV

ICD-10 diagnostic guidelines	ICD-10 research diagnostic criteria	DSM-IV criteria
Differential diagnoses 1 Acute stress reaction F43.0 (immediate reaction in the first 3 days after event) 2 Enduring personality change after a catastrophic experience F62.0 (present for at least 2 years, only after extreme and prolonged stress) 3 Adjustment disorder (less severe stressor or different symptom pattern) 4 Other anxiety or depressive disorders (absence of traumatic stressor or symptoms precedes stressor)	Same as ICD-10 diagnostic guidelines	1 Acute stress disorder (duration of up to 4 weeks) 2 Adjustment disorder (less severe stressor or different symptom pattern) 3 Mood disorder or other anxiety disorder (symptoms of avoidance, numbing, or hyperarousal present before exposure to the stressor) 4 Other disorders with intrusive thoughts or perceptual disturbances (e.g. obsessive–compulsive disorder, schizophrenia, other psychotic disorders, substance-induced disorders)

- DSM-IV puts a stronger emphasis on the avoidance/numbing cluster of symptoms by requiring a minimum of three of these symptoms.
- DSM-IV states two additional criteria that are not included in ICD-10, namely a minimum symptom duration of 1 month and significant distress or impaired functioning.

A large-scale study[20] found a prevalence of ICD-10 PTSD of 6.9 per cent, and a prevalence of DSM-IV PTSD of 3 per cent.

Differential diagnoses

Differential diagnoses are summarized in Table 4.6.2.1. Distinguishing features include the following:

- the type of stressor (adjustment disorders, enduring personality change)
- the symptom pattern (adjustment disorders, enduring personality change)
- the duration of the symptoms (acute stress disorder, acute stress reaction)
- the question of whether the avoidance, numbing, and hyperarousal symptoms were present before the traumatic event occurred (other anxiety or depressive disorders)
- the nature of the intrusive cognitions and perceptual disturbances (obsessive-compulsive disorder, psychotic symptoms, substance-induced symptoms).

Prolonged repeated trauma, such as captivity or repeated childhood sexual abuse, may lead to a more complex pattern of symptoms, 'complex PTSD', that is characterized by somatization, dissociation, affect dysregulation, poor impulse control, self-destructive behaviour, and pathological patterns of relationships.[21] It was debated whether to include a category 'disorders of extreme stress not otherwise specified' (DESNOS) into DSM-IV to accommodate these cases, but the decision was not to include it.[17] In ICD-10, the diagnosis 'enduring personality changes after catastrophic experience' covers such long-standing consequences of enduring trauma.

Furthermore, it is currently being debated whether an additional diagnostic category 'traumatic grief' should be included into the psychiatric classification systems.[22]

Ongoing research on symptom criteria

Some research has questioned the symptom clusters of DSM-IV. In particular, it may be preferable to assess the emotional numbing symptoms separately from the avoidance symptoms, because these symptoms do not load on the same factor in factor analyses and may have different underlying mechanisms. Furthermore, it may be preferable to include severity criteria for the symptoms rather than relying on counting the presence of symptoms.[23]

Assessment instruments

Several semi-structured interviews assess the DSM-IV criteria for PTSD. The most commonly used diagnostic interviews are the Structured Clinical Interview for DSM-IV (SCID)[24] and the Clinician Administered PTSD scale (CAPS).[25]

The most widely used self-report measure of PTSD symptoms used to be the Impact of Event scale.[26] The original scale contained two scales, an intrusion and an avoidance scale. It has been expanded to include an additional hyperarousal scale (IES-R).[27] The IES-R does not cover all the symptoms of PTSD specified in DSM-IV. This is why new measures that are modelled on the DSM-IV criteria are now commonly used in research studies, for example the Post-traumatic Stress Diagnostic scale (PDS)[28] or the PTSD Checklist (PCL).[29]

Epidemiology

The available epidemiological data so far stem mainly from large-scale studies in industrialized societies such as the United States or Australia. It remains to be investigated whether these data replicate in other countries. One has to bear in mind that the society and natural environment set conditions for exposure to traumatic events. For example, in the last decades, people in developing countries have had a much greater exposure to war and natural disasters than people in industrialized western societies.[30]

How common are traumatic events in the population?

Traumatic events are common. In a large representative United States' sample, Kessler *et al.*[31] found that 60.7 per cent of the men and 51.2 per cent of the women had experienced at least one traumatic event meeting DSM-IIIR criteria in their lifetime. The most common types of trauma were witnessing someone being killed or severely injured, accidents, and being involved in a fire, flood, or natural disaster. Using DSM-IV criteria, Stein *et al.*[32] found a lifetime exposure to serious traumatic events of 81.3 per cent for men, and 74.2 per cent for women. Sudden death of a loved person was one of the most frequent traumatic stressors (DSM-IV criteria).[33]

What types of trauma are associated with high PTSD rates?

PTSD rates depend on the type of traumatic event. Rape was associated with the highest PTSD rates in several studies. For example, 65 per cent of the men and 46 per cent of the women who had been raped met PTSD criteria in the Kessler *et al.*[31] study. Other traumatic events associated with high PTSD rates included combat exposure, childhood neglect and physical abuse, sexual molestation; and for women only, physical attack and being threatened with a weapon, kidnapped, or held hostage. Accidents, witnessing death or injury, and fire or natural disasters were associated with relatively low-lifetime PTSD rates of less than 10 per cent.[31] Other research has shown high PTSD rates for torture victims,[34] survivors of the Holocaust,[35] and prisoners of war.[36] The emphasis in DSM-IV on threat to life or physical integrity has led to increasing awareness that medical illness and treatment (e.g. waking up during anaesthesia) can lead to PTSD.[37]

What proportion of people develop PTSD in response to a traumatic stressor?

Kessler *et al.*[31] found that the risk of developing PTSD after a traumatic event is 8.1 per cent for men, and 20.4 per cent for women. For young urban populations, higher risks have been reported; Breslau *et al.* found an overall risk of 23.6 per cent[38]; 13 per cent for men and 30.2 per cent for women.[39]

The figures reported in these studies may be influenced by two types of biases that have opposite effects on probability estimates. First, Breslau *et al.*[33] have pointed out that previous studies overestimated the PTSD-risk imposed by traumatic events because participants reported on the worst trauma that they had experienced. When assessment focused on the symptoms induced by a traumatic event that was randomly selected from the ones that a person had experienced, the conditional risk of PTSD following exposure to trauma was found to be 9.2 per cent, using DSM-IV criteria.

Second, the retrospective methodology used in the epidemiological studies may have led to underestimation of PTSD rates due to selective recall. For example, the prevalence of PTSD 3 months after road traffic accidents was found to be around 20 per cent in prospective longitudinal studies,[40,41] whereas the retrospective studies found prevalences below 10 per cent.

How prevalent is PTSD in the population?

Kessler *et al.*[31] estimated that the lifetime prevalence of PTSD is 7.8 per cent, using DSM-IIIR criteria. Women had a higher prevalence than men (10.4 versus 5.0 per cent). This was due to both a greater exposure to high-impact trauma (rape, sexual molestation, childhood neglect, and childhood physical abuse) and a greater likelihood of developing PTSD when exposed to a traumatic event. Other studies using DSM-IIIR criteria have yielded similarly high prevalence rates.[9,39] Estimates for the 12-month prevalence range between 1.3 per cent in an Australian study[42] and 3.6 per cent in an US study.[43] A study using DSM-IV criteria and found a past-month PTSD prevalence of 2.7 per cent for women and 1.2 per cent for men.[32]

Earlier studies using DSM-III criteria had reported lower lifetime prevalences of about 1 per cent. Besides differences in procedures and sampling methods, the low PTSD prevalence in these earlier studies may be due to the use of an interview schedule with low sensitivity in detecting PTSD.[44] In particular, the early interviews asked global questions about the occurrence of traumatic events and lacked the repeated probing for specific events or event categories that seems to be necessary in eliciting relevant experiences.

Partial PTSD

Several studies have found substantial levels of distress and disability for traumatized people who met some, but not all, of the PTSD criteria specified in DSM-IV.[32] These people may be at greater risk of developing the full PTSD syndrome than people with fewer symptoms.[40,41]

Comorbidity of PTSD with other disorders and symptoms

PTSD shows a substantial comorbidity with affective disorders, other anxiety disorders, substance-use disorders, and somatization. In the study by Kessler *et al.*,[31] 88.3 per cent of the men and 78.1 per cent of the women with PTSD had comorbid psychiatric diagnoses. Studies of veterans with PTSD have also indicated an enhanced level of problems in family and marital adjustment and violent behaviour,[45] and heavy smoking.[46] Furthermore, reports of poor health and increased rates of various diseases, in particular infectious and nervous system diseases, are associated with PTSD.[47]

Is PTSD primary or secondary to the comorbid diagnoses? There is, as yet, little research into this question. The retrospective accounts obtained by Kessler *et al.*[31] suggested that PTSD was primary to comorbid affective or substance-use disorders in the majority of cases, and PTSD was primary to comorbid anxiety disorders in about half of the cases. Similarly, Breslau *et al.*[39] found that PTSD increased the risks for first-onset major depression and alcohol-use disorder. Conversely, pre-existing major depression also increased vulnerability to the PTSD-inducing effects of traumatic events and risk for exposure to traumatic events. A prospective study confirmed that PTSD increased the risk of subsequent pain, conversion symptoms, and somatization symptoms.[48]

Most of the comorbidity research has concentrated on the nature of the relationship between PTSD and alcohol or drug abuse. The majority of studies found that PTSD precedes the development of alcohol-abuse problems. There are probably several mechanisms for this relationship. In the short-term, alcohol is used to self-medicate the symptoms of PTSD, but paradoxically intoxication and withdrawal symptoms may intensify the symptoms in the long-term.[49]

Summary of main findings from epidemiological studies

- The majority of people will experience at least one traumatic event in their lifetime.
- In assessing PTSD history, interviewers should probe for specific events.
- Assault, in particular sexual assault, and combat have a higher impact than accidents and disasters.[31,32]
- If the frequency and impact of traumatic events are considered together, sudden unexpected death of a loved one[33] and road traffic accidents[8] can be considered important causes of PTSD in western industrialized societies.
- Men tend to experience more traumatic events than women, but women experience higher impact events.[31,32]
- Women are at least twice as likely as men to develop PTSD in response to a traumatic event. This enhanced risk is not explained by differences in the type of traumatic event. The estimated lifetime prevalence for women is approximately 10 to 12 per cent, and for men 5 to 6 per cent.[9,31,38,39]
- Comorbid depression and substance-use disorders appear to be secondary to PTSD in the majority of cases.

Aetiology

There is no single accepted theory of PTSD. Theoretical explanations have focused on psychological and biological mechanisms that are not mutually exclusive.

Psychological processes

(a) Fear conditioning

Mowrer's two-factor conditioning theory of phobias has been applied to PTSD.[50–52] It is suggested that through classical (Pavlovian) conditioning, stimuli that were present at the time of the trauma (unconditioned stimulus) become associated with fear and arousal responses. Subsequently, the conditioned stimuli trigger similar (conditioned) responses when presented on their own. Through stimulus generalization and higher-order conditioning, a wide variety of stimuli become triggers of distress in the aftermath of trauma. Quite naturally, the person will try to avoid the conditioned stimuli and the associated distress. The avoidance behaviour is negatively reinforced (operant or instrumental conditioning) because it leads to a reduction in psychological and physical discomfort. In the long-term, however, avoidance prevents extinction of the conditioned fear responses to reminders of the traumatic event, and thus maintains the problem.

(b) Personal meanings of the traumatic event and its aftermath

The persistence of PTSD symptoms has been explained by individual differences in the appraisal of the traumatic event: that is to say, in what personal meaning it has for them.[53–55] Some people are able to see the trauma as a time-limited terrible experience that does not necessarily have negative global implications for their view of themselves, the world or the future. These people are likely to recover quickly. Individuals with persistent PTSD are characterized by *excessively* negative appraisals of the event that go beyond what everyone would find horrific about the event. The nature of predominant emotional responses in PTSD depends on the particular appraisals; for example, appraisals concerning danger lead to fear ('Nowhere is safe'), appraisals concerning others violating personal rules lead to anger ('Others have not treated me fairly'), appraisals concerning responsibility for the traumatic event lead to guilt or shame ('It was my fault', 'I did something despicable'), and appraisals concerning loss lead to sadness ('My life will never be the same again').[55] Such appraisals distinguish between traumatized individuals with and without PTSD, and predict chronic PTSD.[16,56]

Negative appraisals involved in maintaining PTSD do not only concern the traumatic event itself, but also its sequelae such as the initial PTSD symptoms or responses of other people in the aftermath of the traumatic event.[55,57,58] In line with this hypothesis, negative interpretations of intrusive recollections (e.g. 'I am going mad') after road traffic accidents were one of the most important predictors of PTSD at 1 year after the event.[41] Perceived negative responses from other people in the aftermath of trauma predicted PTSD in studies of assault and torture victims.[15,16]

(c) Nature of trauma memories

What exactly distinguishes trauma memories from other memories and what explains the distressing re-experiencing symptoms in PTSD is still under debate.[59,60] Phenomenological observations show that although a wide range of stimuli can trigger unwanted intrusive memories of parts of the traumatic event, people with PTSD show relatively poor intentional recall of details such as the order of events and their recall appears disjointed.[60] Several theories have been put forward to explain re-experiencing symptoms. Foa and colleagues[53,58] explain re-experiencing as spreading activation in a pathological network in memory. This network is thought to be particularly large and easily triggered. It contains many stimulus propositions that are erroneously linked to danger, causing fear responses to harmless stimuli associated with the traumatic event in memory. In addition, the person's reactions during the trauma are linked to the belief that the self is incompetent. Activation of components of the trauma memory (for instance, by confrontation to a reminder, by similar bodily sensations, or by thinking about the event) will activate the whole network, including the emotional responses that the person had during the traumatic event.

Brewin and colleagues[61] have proposed that the symptoms of PTSD can only be explained if one assumes several levels of representation of the traumatic event, namely a verbally accessible memory and a memory that is triggered by situation-specific cues. PTSD is thought to be characterized by easily accessible situationally specific memories that lead to re-experiencing symptoms.

Ehlers and Clark [55] suggested that re-experiencing results from three memory processes, namely, (i) poor inhibition of stimulus-driven retrieval of parts of the trauma memory (re-experiencing) due to insufficient elaboration of the trauma memory (insufficient links to other information that would give the worst moments of the trauma a context such as 'I survived the event' or 'I complied with the requests of the perpetrator because he had threatened me with a knife'), (ii) strong perceptual priming (low perceptual threshold for stimuli with similar sensory characteristics as those present during the trauma), and (iii) strong associations among stimuli present at the time of the event (e.g. footsteps behind me

associated with feeling knife in my back) and between stimuli and emotional responses (e.g. pressure on back associated with fear of death).

(d) Behaviours that maintain PTSD symptoms

Whereas many people will recover from initial PTSD symptoms, some do not get better. This has led researchers to specify possible maintaining behaviours. These include avoidance of reminders, suppression of thoughts and memories connected to the event, rumination, safety behaviours, dissociation, and the use of alcohol or drugs.[57] Ehlers and Clark[55] suggested that these behaviours and cognitive strategies maintain PTSD in three ways. First, some behaviours directly lead to increases in symptoms; for example, thought suppression leads to paradoxical increases in intrusion frequency. Second, other behaviours prevent changes in the problematic appraisals; for example, constantly checking one's rear mirror (a safety behaviour) after a car accident prevents change in the appraisal that another accident will happen if one does not check the mirror. Third, other behaviours prevent elaboration of the trauma memory and its link to other experiences. For example, avoiding thinking about the event may prevent people from incorporating the fact that they did not die into the worst moments of the trauma memory, and they thus continue to re-experience the fear of dying they originally experienced during the event. Several studies have found that avoidance, safety behaviours, thought suppression, and rumination predict maintenance of PTSD.[16,41]

Some of the cognitive processes that maintain PTSD symptoms are not intentional. Patients with PTSD have an unintentional attentional bias to stimuli that are reminiscent of the traumatic event.[6] Involuntary selective attention to reminders may be one of the reasons why these patients have frequent re-experiencing symptoms. Rumination is often described by the patient as unintentional. In particular, patients have problems stopping ruminating once they have started. Rumination may represent a cognitive habit that started as an intentional strategy employed to solve problems and that became more automatic with time.

Biological processes

A number of biological factors have been linked to PTSD symptoms. They have the effect that they make people with PTSD hyperresponsive to stressful stimuli, especially stimuli that are reminiscent of the trauma.[62]

(a) Chronic stress reaction

Patients with PTSD show several abnormalities that are consistent with a chronic stress reaction or an enhanced reactivity to minor stressors. There is evidence for an enhanced secretion of adrenaline (epinephrine) and noradrenaline (norepinephrine) to stress.[63] In psychophysiological studies, patients with PTSD showed enhanced startle responses and higher baseline heart rates and blood pressure than traumatized controls without PTSD. However, these responses may, in part, reflect anticipatory anxiety related to the expectation of trauma cues. Patients with PTSD exhibit greater physiological reactivity to trauma cues (e.g. sounds, pictures, or script-driven imagery) than control subjects without PTSD.[6]

(b) Hypothalamic-pituitary-adrenal axis abnormalities

Patients with PTSD show a different pattern of hypothalamic-pituitary-adrenal response than patients with major depression.[63] Like depressed patients, patients with PTSD hypersecrete corticotrophin-releasing factor (CRF). However, they show *lower* levels of cortisol compared to normal controls, traumatized individuals without current PTSD and depressed patients. When given a low dose of dexamethasone, PTSD patients exhibit *hyper*-suppression of cortisol. Some, but not all findings, suggest that the hypothalamic-pituitary-adrenal axis in PTSD may characterized by enhanced negative feedback.[64] Overall, the hypothalamic-pituitary-adrenal axis in PTSD appears to be set to produce large responses to further stressors.

(c) Hypothalamic-pituitary-thyroid axis

Studies suggest increased hypothalamic-pituitary-thyroid axis activity in PTSD.[63] Patients with PTSD showed increased levels of thyroid hormones, and an exaggerated thyrotropin-stimulating hormone (TSH) response to the standard TSH stimulation test.

(d) Neuroendocrinological abnormalities

Several neurotransmitter systems appear to be dysregulated in PTSD.[63,65] Research suggests exaggerated noradrenergic activity in PTSD in response to stressors.[65] People with PTSD have fewer plaletet α_2-adrenergic receptors, which has been interpreted as a response to chronic elevation of circulating catecholamines. Yohimbine (which blocks α_2-receptors) provokes flashbacks and panic attacks in a substantial subgroup of PTSD patients.[65]

There is also evidence for the involvement of the serotonergic system in PTSD.[63,65] Findings suggest decreased serotonin activity, including decreased serum concentrations, decreased sensitivity of platelet serotonin uptake sites, and blunted prolactin response to D-fenfluramine (indicative of central serotonin hypo-activity). Serotonin-reuptake inhibitors have therapeutic effects in PTSD.

Endogenous opiates have been suspected to mediate the symptoms of emotional numbing and amnesia. In the animal model, uncontrollable stress leads to the secretion of endogenous opiates that induce analgesia. There is some evidence for decreased baseline levels, but increased post-stimulation levels of ß-endorphin.[63]

The dopaminergic and γ-aminobutyric acid (GABA) systems have also been implicated in PTSD, but the evidence for these hypotheses is sparse at this stage.[63]

(e) Neuroimaging

Structural magnetic resonance imaging studies tend to show a reduced hippocampal volume in adults with PTSD, in particular in those with severe and very chronic PTSD, but not in children.[66] Disturbances of hippocampal function may lead to deficits in explicit memory and the appreciation of safe contexts. This line of research was prompted by animal studies showing that high levels of cortisol seen during stress are associated with damage to the hippocampus.[67] However, there is evidence from a study of monozygotic twins that the hippocampal volume of the *unexposed* twins correlated with the PTSD severity of the exposed twins. Thus, small hippocampal volume may be a vulnerability factor for PTSD, an effect of severe or chronic stress, or both. Preliminary evidence suggests that successful long-term treatment with paroxetine increases hippocampal volume in patients with PTSD.[67]

Functional neuroimaging studies have found differences between people with and without PTSD in neurocircuits that are involved in fear conditioning (amygdalae, medial prefrontal cortex, hippocampus). When exposed to trauma reminders or other anxiety cues,[68] people with PTSD tend to show heightened responsivity of the

amygdala and diminished responsivity in the medial prefrontal cortex. The latter plays a role in extinguishing fear reactions, as animal studies established that conditioned fear responses could only be extinguished if the cortex was intact.[69] Recent studies suggest different patterns of responses to trauma reminders for people with PTSD who show high arousal to trauma reminders (involving mainly the anterior cingulate, medial prefrontal cortex, and thalamus) versus those with dissociative reactions (involving mainly the parietal, occipital, and temporal cortex).[70]

(f) Animal models of PTSD

There are biological and psychological parallels between the animal model of inescapable shock and exposure to a traumatic event. The uncontrollability of an aversive event seems to make it particularly traumatic.[71] Inescapable shock leads to changes in the noradrenergic system, the HPA axis, and endogenous opiates that parallel findings in PTSD patients.[51]

However, these effects are usually only observed after repeated exposure to inescapable shock, whereas one traumatic event can be sufficient in inducing PTSD. This is why some authors have suggested that the animal model of kindling or behavioural sensitization is more appropriate in explaining PTSD. Kindling refers to a process whereby intermittent subconvulsive electrical stimulation of the limbic system eventually has the effect that the animal will respond with a seizure to a stimulus that previously was subthreshold. Post *et al.*[72] have suggested that the repeated re-experiencing of the traumatic event may constitute a kindling process, to the effect that PTSD symptoms become more easily triggered with time. Similarly, previous exposure to stressors may sensitize people to respond with PTSD symptoms to a traumatic event.

Animal models suggest that the massive secretion of neurohormones at the time of the trauma, in particular noradrenaline and vasopressin, leads to overconsolidation (long-term potentiation) of the trauma memory. This would have the effect that the conditioned fear responses are particularly difficult to extinguish and that stimuli that resemble those present during trauma are particularly likely to trigger intrusive memories, distress, and/or the corresponding physiological responses.[62]

(g) Genetic factors

Twin studies have found a higher concordance of PTSD among monozygotic than dizygotic twins. There is also an increased prevalence of psychiatric disorders, especially anxiety disorders, affective disorders, sociopathy, and/or substance abuse, among family members of people with PTSD.[73] A study found a higher proportion of the 5-HTTLPR s/s genotype in PTSD compared to controls.[74]

Course and prognosis

Time course of symptoms

For the vast majority of PTSD cases, symptoms begin immediately after the traumatic event. Delayed onset is found in a minority (11 per cent or less) of the cases.[6]

Prospective longitudinal studies suggest that a large proportion of people who initially develop PTSD after trauma will recover on their own. However, between one-third and 50 per cent of those who develop PTSD after a traumatic event will not recover for many years.[31,75] Long-term outcome depends on initial symptom severity and the experience of further traumatic events. People with high initial PTSD severity are more likely to remain symptomatic at follow-up than those with low initial symptom severity.[41,75]

Factors that influence the risk of developing PTSD

Meta-analyses[11,76] have identified several reliable predictors of PTSD. The results are summarized in Table 4.6.2.2. Overall, peri-traumatic factors such as perceived life threat and dissociation during the trauma and post-trauma factors such as low social support appear more predictive of PTSD than pre-trauma variables.

(a) Demographic and pre-trauma variables

Women have consistently shown to have a greater risk of developing PTSD than men, but the mechanisms remain unclear.[11,31,67,76,77] Similarly, it is as yet unclear why people with lower intelligence or education,[76] people with lower socioeconomic status,[76] and people from ethnic minorities[78] have an elevated PTSD risk. Other risk factors include previous trauma, childhood adversity, a personal or family history of anxiety or depression,[76] and neuroticism.[38]

(b) Stressor variables

PTSD risk depends on the severity of the stressor. Prolonged and repeated trauma, exposure to the grotesque aftermath of violence, events that involve intentional harm by another person and abusive violence, and events that involve harm to children are particularly likely to lead to PTSD.[79,80]

Injury severity is only a weak predictor of PTSD. Long-term health problems and loss of function may play a greater role in maintaining PTSD.[41,79] There are some reports that unconsciousness during a traumatic event may decrease the risk of PTSD,[81]

Table 4.6.2.2 Risk factors for post-traumatic stress disorder

	Weighted average effect size *r*
Demographic variables	
Female sex	0.13[76]
Race (minority status)	0.05[76]
Younger age	0.06[76]
Low socio-economic status	0.14[76]
Cognitive ability	
Low education	0.10[76]
Low intelligence	0.18[76]
Psychiatric and trauma history	
Psychiatric history	0.11[76]–0.17[11]
Family psychiatric history	0.13[76]–0.17[11]
Prior trauma	0.12[76]–0.17[11]
Childhood abuse	0.14[76]
Other adverse childhood	0.19[76]
Peri-traumatic factors	
Trauma severity	0.23[76]
Perceived life threat	0.26[11]
Peri-traumatic emotions	0.26[11]
Peri-traumatic dissociation	0.35[11]
Post-trauma factors	
Low social support	0.28[11]–0.40[76]
Further life stress	0.32[76]

(Modified from meta-analyses by Brewin *et al.* (2000) and Ozer *et al.* (2003))

but other studies have found small associations in the opposite direction.[41]

(c) Psychological responses during trauma

PTSD risk depends on the degree of psychological distress the traumatic event caused. The psychological impact of the trauma depends on the perceived threat to life,[11,76] the perceived loss of control (helplessness),[82] and the perceived threat to one's autonomy (mental defeat).[15,16] Among the psychological responses predicting PTSD are feelings of anger, guilt, or shame[83] and dissociation and numbing.[11]

Factors affecting recovery from trauma

(a) Recovery environment

Recovery is facilitated by social support, [11,76] and hindered by perceived negative responses from other people.[15,16] Further stressful or traumatic life events impede recovery from PTSD.[76] This includes the stress caused by long-lasting negative effects of the event on health and personal appearance, financial difficulties, disruptions in everyday life, and ongoing litigation.[41,81,84]

(b) Psychological processes

Excessively negative appraisals of the traumatic event impede recovery (e.g. 'Nowhere is safe', 'I cannot trust anyone', 'I am inadequate').[16,56] If individuals interpret their initial PTSD symptoms as signs that they are going mad or losing control, or as signs of a permanent change for the worse, they are less likely to recover.[16,41]

If individuals engage in behaviours or cognitive coping styles that prevent them from 'working through' and accepting the trauma, they are less likely to recover. Such maladaptive behaviours include avoidance, not talking about the experience, safety behaviours, denial, thought suppression, and rumination.[16,41,80,85]

Treatment of PTSD

Meta-analyses of randomized controlled trials have identified several effective psychological and pharmacological treatments for PTSD.[86–88]

Psychological treatments

Psychological treatments lead, on average, to large improvements in PTSD symptoms. The mean effect size (Cohen's *d* statistic) for the difference between the pre- and post-treatment scores was $d = 1.43$ across 26 studies of psychological treatments.[86] (An effect size $d = 1$ means that the treatment led to improvement by one standard deviation). In interpreting the effect sizes of treatments, one has to bear in mind that pill-placebo or waiting list conditions also lead to some improvement. Mean effect sizes for these conditions were $d = 0.77$ and $d = 0.75$ for observer-rated PTSD symptoms, and $d = 0.51$ and $d = 0.44$ for self-rated PTSD symptoms in 61 treatment-outcome trials.[88]

Not all psychological treatments are equally effective in treating PTSD. According to the meta-analyses, trauma-focused treatments are more effective than other treatments.[86,87] Trauma-focused cognitive behaviour therapy (CBT) and eye movement desensitization and reprocessing (EMDR) were superior to stress management and other therapies such as supportive therapy or hypnotherapy.[87] On average, 67 per cent of patients who complete the trauma-focused treatments (and 56 per cent of those who enter these treatments; intent-to-treat analysis) no longer met diagnostic criteria for PTSD.[86] For supportive therapy, the corresponding recovery rates were 39 per cent among treatment completers and 36 per cent in intent-to-treat analyses.[86] On the basis of these results, recent United Kingdom and Australian treatment guidelines recommend trauma-focused CBT and EMDR as the treatments of choice for PTSD.[89,90] Both treatments address the patient's troubling memories of the traumatic events and the personal meaning of the event and its consequences.

(a) Trauma-focused psychological treatments

(i) Cognitive behavioural therapy (CBT)

All effective CBT programmes for PTSD include an element of psycho-education about common reactions to trauma that normalizes the PTSD sufferer's symptoms, and a rationale for the interventions. Trauma-focused CBT programmes for PTSD include either exposure or cognitive therapy, or a combination of these interventions. Some also include elements of stress management training such as breathing training.

Exposure. Exposure treatment for PTSD comprises two components.[58] In imaginal exposure, patients are systematically exposed to the memory of the trauma. A commonly used procedure is imaginal reliving,[58] where patients relive the traumatic event in their imagination, including their thoughts and feelings at the time. This is repeated until the reliving no longer evokes high levels of distress. Writing a trauma narrative can also be used as a method of exposure to the trauma memory.[54] In vivo exposure[58] involves confronting (safe) situations that patients avoid because they remind them of the trauma (e.g. going to the site of the traumatic event, driving again after a road traffic accident). Exposure is repeated until the patient no longer responds with high levels of distress. The effects of exposure were originally explained as an effect of habituation, but there are probably several mechanisms for its efficacy.[55,58] For example, patients realize that exposure does not lead to a feared outcome (for example, going to the site of an accident will not mean that another accident will happen; thinking about the trauma will not make them go mad) and thus helps in correcting dysfunctional beliefs about danger of the world and the meaning of PTSD symptoms. Second, the repeated exposure helps patients to create an organized memory and facilitates the distinction that intrusive thoughts and images are memories rather than something happening right now.

Whereas exposure treatment is effective in the majority of cases, a minority of patients become worse.[91] In particular, exposure treatment does not appear suitable for patients whose traumatic memories are about being perpetrators rather than victims.[92] It may also have limits in treating survivors of complex and prolonged traumatic events such as torture, war, or captivity.[93]

Cognitive therapy. Cognitive therapy for anxiety disorders focuses on the identification and modification of misinterpretations that lead the patient to overestimate threat. In PTSD, the perceived threat stems from interpretations of the trauma and its aftermath.[54,55] For example, people with PTSD may feel strong guilt or shame related to the trauma: a rape victim may blame herself for the rape; a war veteran may feel it was his fault that his best friend was killed. Others overestimate the current danger they are encountering in everyday life. An accident survivor may become convinced that he is at great risk of having a further trauma.

Others may take the intrusive re-experiencing symptoms as a sign that they are about to go crazy. By discussing the evidence for and against the interpretations, and by testing out the predictions derived from the interpretations with the help of the therapist, the patient arrives at more adaptive conclusions. The patient is encouraged to drop behaviours and cognitive strategies that prevent a disconfirmation of the negative interpretations, e.g. excessive precautions to prevent further trauma or excessive rumination about what one could have done differently during the event. Recent studies have shown that cognitive therapy is effective on its own, without additional exposure treatment.[91,94] However, when verbal challenging of dysfunctional beliefs was used as an additional procedure after a session of imaginal exposure, it did not lead to additional treatment gains.[94] However, cognitive therapy may help reduce the amount of exposure necessary to achieve large treatment effects.[95,96]

(ii) Eye-movement desensitization and reprocessing (EMDR)

Like trauma-focused CBT, EMDR[97] aims to help patients process their traumatic memory and think more positively about their experience. It involves inducing a series of rapid and rhythmic eye movements while the patient focuses on a trauma-related image and related negative emotions, sensations, and thoughts. Patients are instructed to visually track the therapist's fingers as they move back and forth in front of the patient's eyes for sets of about 20 s. After each set, the patient discusses the images and emotions they experienced during the eye movements with the therapist. This process is repeated, and includes focusing on different memories that come up in connection with the trauma. Once distress to the target image is reduced, patients are instructed to focus on the image while rehearsing a positive thought connected to the image. The mechanism of treatment is not yet understood. Several empirical studies have suggested that the eye movements may not be necessary in producing the therapeutic effects observed with EMDR.[98]

(iii) Other psychological treatments

Most other psychological treatments have a small evidence-base. In general, non-trauma focused treatments were shown to be less effective than trauma-focused treatments in randomized controlled trials.[86,87]

Stress management (stress inoculation). The goal of this treatment is to teach the patient a set of skills that will help them cope with stress. Examples include relaxation training, training in slow abdominal breathing, thought stopping of unwanted thoughts, assertiveness training, and training in positive thinking. Stress management is more effective than supportive psychotherapy, but given on its own less effective than trauma-focused treatments.[87]

Psychodynamic therapy. Several different forms of psychodynamic treatments for PTSD have been described.[99] The focus lies on resolving unconscious conflicts provoked by the stressful event by making it conscious in tolerable doses. This is thought to help the patient reengage normal mechanisms of adaptation. The goal of treatment is to understand the meaning of the stressful event in the context of the individual's personality, attitudes, and early experiences. The psychological meaning of the event is explored by a range of methods such as 'sifting and sorting through wishes, fantasies, fears, and defences stirred up by the event'.[99] Treatment strategies include exploratory insight-oriented, supportive, or directive activity. It may also include working with transference, but with the therapist using a less strict technique than that used in psychoanalysis. To date there is only one controlled study of psychodynamic therapy, and the effect size observed in that study is below those observed with trauma-focused CBT or EMDR.[88]

Hypnotherapy. The patient is given instructions to induce a state of highly focused attention, a reduced awareness of peripheral stimuli, and a heightened suggestibility. The goal of this treatment is to enhance control over trauma-related emotional distress and hyperarousal symptoms and to facilitate the recollection of details of the traumatic event. To date there is only one controlled study of hypnotherapy, and the effect size observed in that study is below those observed with trauma-focused CBT or EMDR.[88] Shalev *et al.*[93] raised concerns about the use of hypnotherapy in the treatment of PTSD as it may induce dissociative states.

Supportive therapy. Supportive therapy is primarily non-directive and non-advisory. Through a supportive therapeutic relationship, clients are encouraged to explore their thoughts, feelings, and behaviour to reach clearer self-understanding; and to find and use their strengths so that they cope more effectively with their lives by making appropriate decisions, or by taking relevant action. Supportive therapy is less effective than trauma-focused psychological treatments.[86,87]

Pharmacological treatments

Although recent guidelines[89,90] recommend trauma-focused psychological treatments as the first line treatments for PTSD, they also acknowledge a role for medication. A Cochrane review[100] found an advantage of medication over placebo with response rates of 59.1 and 38.5 per cent, respectively. Indications for medication according to recent guidelines include patient choice, severe ongoing threat, the patient is too depressed or unstable to engage in psychological treatment, and failure to respond to psychological treatment.

(a) Selective serotonin-reuptake inhibitors (SSRIs)

Among the pharmacological treatments, the SSRIs have been most widely studied and are recommended as the first line medication choice for PTSD according to expert consensus[90,101] and a recent Cochrane review.[100] Paroxetine has received the most consistent support for its efficacy.[89,100] SSRIs reduce alcohol consumption; a relevant finding given the high comorbidity of PTSD with substance abuse or dependency.[102]

(b) Monoamine oxidase inhibitors (MAOIs)

There is some evidence that phenelzine is effective in PTSD,[89,100] particularly in reducing re-experiencing symptoms and insomnia.[102] Thus, phenelzine may be considered as one of the treatment options for PTSD by mental health specialists,[89] taking into consideration the risks and necessary dietary restrictions.

(c) Other antidepressants

There is some evidence for the efficacy of tricyclic antidepressants, particularly amitriptyline, in the treatment of PTSD.[89,100] They may be considered as a treatment option for PTSD by mental health specialists.[89,90] Mirtazepine, a noradrenergic and specific serotonergic antidepressant, has shown promise in initial trials and has been recommended as one of the treatment options for PTSD.[89,90]

(d) Benzodiazepines

Benzodiazepines do not appear to be effective in the treatment of PTSD. They do not affect the re-experiencing, avoidance, and numbing symptoms, although they may show some effects on insomnia, irritability, and general anxiety and arousal symptoms.[102]

Advice on management

Diagnosing PTSD

When assessing whether a patient has experienced a trauma, it is important to ask about specific examples of traumatic events. Patients may initially feel too overwhelmed or ashamed to report details of their traumatic experience, and may find it easier to reply to factual questions with 'yes' or 'no' answers. In diagnosing PTSD, clinicians need to ascertain that the patient involuntarily re-experiences parts of the event. In addition, patients will need to have some symptoms of hyperarousal, avoidance, and emotional numbing. Self-report instruments such as the Post-traumatic Stress Diagnostic scale[28] or semi-structured interviews such as the Clinician Administered PTSD scale[25] are useful in assessing the pattern and severity of symptoms. Clinicians should ask about the impact of the symptoms (e.g. distress, restrictions, effect on family and work) on the patient's life. The DSM-IV criterion of a minimum of three avoidance or numbing symptoms appears too strict for clinical purposes. It does not appear justified to withhold treatment if the patient is disabled by the PTSD symptoms but fails to meet this criterion.

Is PTSD the main problem?

As PTSD is often comorbid with other disorders, clinicians need to ascertain whether PTSD is the main problem for which the patient presently needs help at present. The time course of the onset of the comorbid disorders and changes in the severity of symptoms provide useful information. It is also helpful to ask patients whether they believe that they would need professional help for their other problems if the PTSD symptoms are resolved. In cases of comorbid depression, the PTSD should usually be treated first (as comorbid depression improves with successful PTSD treatment), unless the depression is so severe that the patient cannot engage in treatment. In contrast, significant substance dependence usually needs to be addressed before treating the PTSD.[89]

Assessments need to include a risk assessment. If there is a high risk of suicide or harm to others, clinicians will need to concentrate on managing the risk first.[89]

Clinicians will also need to establish whether there is any serious ongoing threat or health, social, and financial problems that may need to be addressed before the patient can engage in psychological treatment.[89]

(a) Choice of treatment

Patients should be informed about effective treatments for PTSD. Information material for patients and carers is available.[103] Trauma-focused psychological treatments are the best treatment option for PTSD to date, given the strong evidence for their effectiveness,[89,90] their established long-term effectiveness,[88] and the lower dropout rates compared to medication. Meta-analyses report that, across studies, 14–21 per cent of the patients can be expected to dropout of psychological treatments,[86,88] compared to 31–36 per cent for medication.[88,100] The advantage of SSRIs and other medications compared to psychological treatments is that they are more readily available.

(i) Psychological treatment

Trauma-focused psychological treatments usually last between 8 and 12 sessions (longer for patients with multiple traumas and comorbid personality disorders). When the trauma is discussed in the session, 90 min should be allowed for the session.[89,90] The treatments require a good therapeutic alliance.[90] Depending on the nature of the trauma and comorbid problems, additional sessions for establishing trust and emotional stabilization may be needed before the trauma-focused treatment commences.[89,90]

(ii) Pharmacological treatment

Clinicians should discuss the benefits and possible side effects of the prescribed medication with the patient, and address common concerns, such as fears of addiction. Patients also need to be informed that the medication needs to be discontinued gradually. Most antidepressants recommended for the use in PTSD need to be discontinued over at least 4 weeks.[89] The risk of self-harm needs to be considered when prescribing antidepressants. Those with high risk should be seen at least weekly until the risk is no longer considered significant.[89,90] Patients receiving SSRIs need to be monitored for akathisia, suicidal ideation, and increased anxiety and agitation. Patients receiving phenelzine require careful monitoring (including blood pressure measurement) and advice about interactions with other medicines and food.[89]

(b) Special problems in the management of PTSD patients

Avoidance is one of the main symptoms of PTSD, and it can thus take years for the patient to seek help for this condition. It is important for clinicians to bear in mind that even those who seek help may find it hard to talk about the traumatic experience, and may show signs of avoidance such as irregular attendance or failure to disclose the worst moments of the trauma initially. Therapeutic techniques to deal with this problem include empathy, gradual encouragement, and giving the patient control over the timing and mode of working through the experience (e.g. writing, talking into a tape recorder, reliving with the support of the therapist).

One of the requirements for change is that the patient feels safe. Therapists therefore have to make sure that they establish a good relationship with the patient, and that the therapeutic setting or their behaviour does not remind the patient of the traumatic event. Sometimes support in changing living circumstances may be necessary if they prevent the patient from being safe (e.g. moving house if assaulted by a neighbour).

Patients with PTSD often suffer from poor sleep and concentration, and find it painful to face reminders of the trauma. For these reasons, they have difficulty in dealing with the aftermath of traumatic events such as legal procedures and continuing treatment for physical injuries, including the long delays that this usually involves. Such ongoing stressors impede recovery, and patients may therefore benefit from problem-solving and practical advice.

Further information

National Institute of Clinical Excellence. (2005). Clinical guideline 26: post-traumatic stress disorder: the management of PTSD in adults and children in primary and secondary care. http://guidance.nice.org.uk/CG26

National Center for Posttraumatic Stress Disorder. http://www.ncptsd.va.gov/ncmain/providers/ and PILOTS database http://www.ncptsd.va.gov/ncmain/publications/pilots/

Nemeroof, C.B., Bremner, J.D., Foa, E.B., *et al.* (2006). Posttraumatic stress disorder: a state-of-the art review. *Journal of Psychiatric Research*, **40**, 1–21.

Taylor, S. (2006). *Clinician's guide to treating PTSD: a cognitive-behavioral approach*. Guilford Press, New York.

References

1. Gersons, B.P.R. and Carlier, I.V.E. (1992). Post-traumatic stress disorder: the history of a recent concept. *The British Journal of Psychiatry*, **161**, 742–8.
2. Kinzie, J.D. and Goetz, R.R. (1996). A century of controversy surrounding post-traumatic stress-spectrum syndromes: the impact of DSM-III and DSM-IV. *Journal of Traumatic Stress*, **9**, 159–79.
3. Van der Kolk, B.A., Weisaeth, L., and Van der Hart, O. (1996). History of trauma in psychiatry. In *Traumatic stress* (eds. B.A. van der Kolk, A.C. McFarlane, and L. Weisaeth), pp. 47–74. Guilford Press, New York.
4. American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd edn). American Psychiatric Association, Washington, DC.
5. World Health Organization. (1992). *International statistical classification of diseases and related health problems*, 10th revision. WHO, Geneva.
6. McNally, R.J. (2000). Post-traumatic stress disorder. In *Oxford textbook of psychopathology* (eds. T. Millon, P.H. Blaney, and R.D. Davis). Oxford University Press, Oxford.
7. American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd edn revised). American Psychiatric Association, Washington, DC.
8. Norris, F.H. (1992). Epidemiology of trauma: frequency and impact of different potentially traumatic events on different demographic groups. *Journal of Consulting and Clinical Psychology*, **60**, 409–18.
9. Resnick, H.S., Kilpatrick, D.G., Dansky, B.S., *et al.* (1993). Prevalence of civilian trauma and post-traumatic stress disorder in a representative national sample of women. *Journal of Consulting and Clinical Psychology*, **61**, 984–91.
10. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
11. Ozer, E.J., Best, S., Lipzey, T.L., *et al.* (2003). Predictors of posttraumatic stress disorder and symptoms in adults: a meta-analysis. *Psychological Bulletin*, **129**, 52–73.
12. Roemer, L., Orsillo, S.M., Borkovec, T.D., *et al.* (1998). Emotional response at the time of a potentially traumatizing event and PTSD symptomatology: a preliminary retrospective analysis of the DSM-IV criterion A-2. *Journal of Behaviour Therapy and Experimental Psychiatry*, **29**, 123–30.
13. Yehuda, R., Southwick, S.M., and Giller, E.L. (1992). Exposure to atrocities and severity of chronic post-traumatic stress disorder in Vietnam combat veterans. *The American Journal of Psychiatry*, **149**, 333–6.
14. Evans, C., Ehlers, A., Mezey, G., *et al.* (2007). Intrusive memories in perpetrators of violent crime: emotions and cognitions. *Journal of Consulting and Clinical Psychology*, **75**, 134–44.
15. Ehlers, A., Maercker, A., and Boos, A. (2000). PTSD following political imprisonment: the role of mental defeat, alienation, and perceived permanent change. *Journal of Abnormal Psychology*, **109**, 45–55.
16. Dunmore, E., Clark, D.M., and Ehlers, A. (2001). A prospective study of the role of cognitive factors in persistent posttraumatic stress disorder after physical or sexual assault. *Behaviour Research and Therapy*, **39**, 1063–84.
17. Brett, E.A. (1996). The classification of post-traumatic stress disorder. In *Traumatic stress* (eds. B.A. van der Kolk, A.C. McFarlane, and L. Weisaeth), pp. 117–28. Guilford Press, New York.
18. Davidson, J.R.T. and Foa, E.B. (1991). Diagnostic issues in post-traumatic stress disorder: consideration for the DSM-IV and beyond. *Journal of Abnormal Psychology*, **100**, 346–55.
19. World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*. World Health Organization, Geneva.
20. Andrews, G., Slade, T., and Peters, L. (1999). Classification in psychiatry: ICD-10 versus DSM-IV. *The British Journal of Psychiatry*, **174**, 3–5.
21. Herman, J.L. (1992). Complex PTSD: a syndrome in survivors of prolonged and repeated trauma. *Journal of Traumatic Stress*, **5**, 377–91.
22. Prigerson, H.G., Shear, M.K., Jacobs, S.C., *et al.* (1999). Consensus criteria for traumatic grief: a preliminary empirical test. *The British Journal of Psychiatry*, **174**, 67–73.
23. Foa, E.B., Riggs, D.S., and Gershuny, B.S. (1995). Arousal, numbing, and intrusion: symptom structure of PTSD following assault. *The American Journal of Psychiatry*, **152**, 116–20.
24. First, M.B., Spitzer, R.L., Gibbon, M., *et al.* (1995). *Structured clinical interview for DSM-IV Axis I disorders-patient version* (*SCID-I/P, Version 2.0*). Biometrics Research Department of the New York State Psychiatric Institute, New York.
25. Blake, D.D., Weathers, F.W., Nagy, L.M., *et al.* (1995). The development of a clinician-administered PTSD scale. *Journal of Traumatic Stress*, **8**, 75–90.
26. Horowitz, M.J., Wilner, N., and Alvarez, W. (1979). Impact of Event Scale: a measure of subjective stress. *Psychosomatic Medicine*, **41**, 209–18.
27. Weiss, D.S. and Marmar, C. R. (1997). The impact of Event Scale-revised. In *Assessing psychological trauma and PTSD* (eds. J.P Wilson and T.M. Keane), pp. 399–411. Guilford Press, New York.
28. Foa, E.B., Cashman, L., Jaycox, L., *et al.* (1997). The validation of a self-report measure of post-traumatic stress disorder: the post-traumatic diagnostic scale. *Psychological Assessment*, **9**, 445–51.
29. Weathers, F.W., Huska, J.A., and Keane, T.M. (1991). *The PTSD Checklist—Civilian version* (*PCL-C*) *for DSM-IV*. National Center for PTSD—Behavioral Science Division, Boston. http://www.oqp.med.va.gov/cpg/PTSD/PTSD_cpg/content/appendices/appendixC.htm
30. McFarlane, A.C. and de Girolamo, G. (1996). The nature of traumatic stressors and the epidemiology of post-traumatic reactions. In *Traumatic stress* (eds. B.A. van der Kolk, A.C. McFarlane, and L. Weisaeth), pp. 129–54. Guilford Press, New York.
31. Kessler, R.C., Sonnega, A., Bromet. E., *et al.* (1995). Post-traumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry*, **52**, 1048–60.
32. Stein, M.B., Walker, J.R., Hazen, A.L., *et al.* (1997). Full and partial post-traumatic stress disorder: findings from a community survey. *The American Journal of Psychiatry*, **154**, 1114–19.
33. Breslau, N., Kessler, R.C., Chilcoat, H.D., *et al.* (1998). Trauma and post-traumatic stress disorder in the community. *Archives of General Psychiatry*, **55**, 626.
34. Maercker, A. and Schützwohl, M. (1997). Longterm effects of political imprisonment: a group comparison study. *Social Psychiatry and Psychiatric Epidemiology*, **32**, 435–42.
35. Kuch, K. and Cox, B.J. (1992). Symptoms of PTSD in 124 survivors of the Holocaust. *The American Journal of Psychiatry*, **149**, 337–40.
36. Engdahl, B.E., Dikel, T.N., Eberly, R., *et al.* (1997). Post-traumatic stress disorder in a community group of former prisoners of war: a normative response to severe trauma. *The American Journal of Psychiatry*, **154**, 1576–81.
37. Mayou, R.A. and Smith, K.A. (1997). Post traumatic symptoms following medical illness and treatment. *Journal of Psychosomatic Research*, **43**, 121–3.

38. Breslau, N., Davis, G.C., Andreski, P., *et al.* (1991). Traumatic events and post-traumatic stress disorder in an urban population of young adults. *Archives of General Psychiatry*, **48**, 216–22.
39. Breslau, N., Davis, G.C., Andreski, P., *et al.* (1997). Sex differences in post-traumatic stress disorder. *Archives of General Psychiatry*, **54**, 1044–8.
40. Blanchard, E.B. and Hickling, E.J. (1997). *After the crash. Assessment and treatment of motor vehicle survivors*. American Psychological Association, Washington, DC.
41. Ehlers, A., Mayou, R.A., and Bryant, B. (1998). Psychological predictors of chronic post-traumatic stress disorder after motor vehicle accidents. *Journal of Abnormal Psychology*, **107**, 508–19.
42. Creamer, M., Burgess, P., and McFarlane, A.C. (2001). Post-traumatic stress disorder: findings from the Australian National Survey of Mental Health and Well-being. *Psychological Medicine*, **31**, 1237–47.
43. Narrow, W.E., Rae, D.S., Robins, L.N., *et al.* (2002). Revised prevalence estimates of mental disorders in the United States: using a clinical significance criterion to reconcile 2 surveys' estimates. *Archives of General Psychiatry*, **59**, 115–23.
44. Solomon, S.D. and Davidson, J.R.T. (1997). Trauma: prevalence, impairment, service use, and cost. *The Journal of Clinical Psychiatry*, **58**, 5–11.
45. Jordan, B.K., Marmar, C.R., Fairbank, J.A., *et al.* (1992). Problems in families of male Vietnam veterans with post-traumatic stress disorder. *Journal of Consulting and Clinical Psychology*, **60**, 916–26.
46. Beckham, J.C., Kirby, A.C., Feldman, M.E., *et al.* (1997). Prevalence and correlates of heavy smoking in Vietnam veterans with chronic post-traumatic stress disorder. *Addictive Behaviors*, **22**, 637–47.
47. Schnurr, P. and Green, B.L. (eds.) (2004). *Trauma and health. Physical consequences of exposure to extreme stress*. American Psychological Association, Washington, DC.
48. Andreski, P., Chilcoat, H., and Breslau, N. (1998). Post-traumatic stress disorder and somatization symptoms: a prospective study. *Psychiatry Research*, **79**, 131–8.
49. Stewart, S.H. (1996). Alcohol abuse in individuals exposed to trauma: a critical review. *Psychological Bulletin*, **120**, 83–112.
50. Mowrer, O.H. (1939). A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychological Review*, **46**, 553–65.
51. Charney, D., Deutch, A.Y., Krystal, J.H., *et al.* (1993). Psychobiological mechanisms of post-traumatic stress disorder. *Archives of General Psychiatry*, **50**, 294–305.
52. Keane, T.M., Zimering, R.T., and Caddell, J.M. (1985). A behavioral formulation of post-traumatic stress disorder. *Behavior Therapy*, **8**, 9–12.
53. Foa, E.B. and Riggs, D.S. (1993). Post-traumatic stress disorder in rape victims. In *Annual review of psychiatry*, Vol. 12 (eds. J. Oldham, M.B. Riba, and A. Tasman), pp. 273–303. American Psychiatric Association, Washington, DC.
54. Resick, P.A. and Schnicke, M.K. (1993). *Cognitive processing therapy for rape victims*. Sage, Newbury Park, CA.
55. Ehlers, A. and Clark, D.M. (in press). A cognitive model of persistent PTSD. *Behaviour Research and Therapy*.
56. Foa, E.B., Ehlers, A., Clark, D.M., *et al.* (1999). The post-traumatic cognitions inventory (PTCI): development and validation. *Psychological Assessment*, **11**, 303–14.
57. Ehlers, A. and Steil, R. (1995). Maintenance of intrusive memories in post-traumatic stress disorder: a cognitive approach. *Behavioural and Cognitive Psychotherapy*, **23**, 217–49.
58. Foa, E.B. and Rothbaum, B.O. (1998). *Treating the trauma of rape. Cognitive-behavior therapy for PTSD*. Guilford Press, New York.
59. McNally, R.J. (2003). *Remembering trauma*. Harvard University Press, Cambridge, MA.
60. Ehlers, A., Hackmann, A. and Michael, T. (2004). Intrusive reexperiencing in posttraumatic stress disorder: phenomenology, theory, and therapy. *Memory*, **12**, 403–15.
61. Brewin, C.R., Dalgleish, T., and Joseph, S. (1996). A dual representation theory of post-traumatic stress disorder. *Psychological Review*, **103**, 670–86.
62. Van der Kolk, B.A. (1996). The body keeps the score. Approaches to the psychobiology of post-traumatic stress disorder. In *Traumatic stress* (eds. B.A. van der Kolk, A.C. McFarlane, and L. Weisaeth), pp. 214–41. Guilford Press, New York.
63. Newport, D.J. and Nemoroff, C.B. (2000). Neurobiology of posttraumatic stress disorder. *Cognitive Neuroscience*, **10**, 211–8.
64. Kloet, C.S., Vermetten, E., Geuze, E., *et al.* (2006). Assessment of HPA-axis function in posttraumatic stress disorder: pharmacological and non-pharmacological challenge tests, a review. *Journal of Psychiatric Research*, **40**, 550–67.
65. Southwick, S.M., Rasmusson, A., Barron, J., *et al.* (2005). Neurobiological and neurocognitive alterations in PTSD. In *Neuropsychoology of PTSD: biological, cognitive and clinical perspectives* (eds. J. Vaterling and C. Brewin), pp. 27–58. Guilford Press, New York.
66. Karl, A., Schaefer, M., Malta, L.S., *et al.* (2006). A meta-analysis of structural brain abnormalities in PTSD. *Neuroscience and Biobehavioral Reviews*, **30**, 1004–31.
67. Nemeroof, C.B., Bremner, J.D., Foa, E.B., *et al.* (2006). Posttraumatic stress disorder: a state-of-the art review. *Journal of Psychiatric Research*, **40**, 1–21.
68. Shin, L.M., Rauch, S.L., and Pitman, R.K. (2006). Amygdala, medial prefrontal cortex, and hippocampal function in PTSD. *Annals of the New York Academy of Science*, **1071**, 67–79.
69. LeDoux, J.E. (1992). Emotion as memory: anatomical systems underlying indelible memory traces. In *Handbook of emotion and memory* (ed. S.A. Christianson), pp. 269–88. Erlbaum, Hillsdale, NJ.
70. Lanius, R.A., Bluhm, R., Lanius, U., *et al.* (2006). A review of neuroimaging studies in PTSD: heterogeneity of response to symptom provocation. *Journal of Psychiatric Research*, **40**, 709–29.
71. Foa, E.B., Zinbarg, R., and Rothbaum, B.O. (1992). Uncontrollability and unpredictability in post-traumatic stress disorder: an animal model. *Psychological Bulletin*, **112**, 218–38.
72. Post, R.M., Weiss, S.R.B., and Smith, M.A. (1995). Sensitization and kindling: implications for the evolving neural substrates of post-traumatic stress disorder. In *Neurobiological and clinical consequences of stress: from normal adaptation to post-traumatic stress disorder* (eds. M.J. Friedman, D.S. Charney, and A.Y. Deutch), pp. 203–24. Lippincott-Raven, Philadelphia, PA.
73. Connor, K.M. and Davidson, J.R.T. (1997). Familial risk factors in post-traumatic stress disorder. *Annals of the New York Academy of Sciences*, **821**, 35–51.
74. Lee, H.J., Lee, M.S., Kang, R.H., *et al.* (2005). Influence of the serotonin transporter promoter gene polymorphism on susceptibility to posttraumatic stress disorder. *Depression and Anxiety*, **21**, 135–9.
75. Perkonigg, A., Pfister, H., Stein, M.B., *et al.* (2005). Longitudinal course of posttraumatic stress disorder symptoms in a community sample of adolescents and young adults. *The American Journal of Psychiatry*, **162**, 1320–7.
76. Brewin, C.R., Andrews, B., and Valentine, J.D. (2000). Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. *Journal of Consulting and Clinical Psychology*, **68**, 748–66.
77. Olff, M., Langeland, W., Draijer, N., *et al.* (2007). Gender differences in posttraumatic stress disorder. *Psychological Bulletin*, **133**, 183–204.
78. Frueh, B.C., Brady, K.L., and Dearellano, M.A. (1998). Racial differences in combat-related PTSD: empirical findings and conceptual issues. *Clinical Psychology Review*, **18**, 287–305.
79. Green, B.L. (1994). Psychosocial research in traumatic stress: an update. *Journal of Traumatic Stress*, **7**, 341–62.
80. Clohessy, S. and Ehlers, A. (1999). PTSD symptoms, response to intrusive memories, and coping in ambulance service workers. *The British Journal of Clinical Psychology*, **38**, 251–66.

81. Mayou, R.A., Bryant, B., and Duthie, R. (1993). Psychiatric consequences of road traffic accidents. *British Medical Journal*, **307**, 647–51.
82. Baum, A., Cohen, L., and Hall, M. (1993). Control and intrusive memories as possible determinants of chronic stress. *Psychosomatic Medicine*, **55**, 274–86.
83. Andrews, B., Brewin, C.R., Rose, S., *et al.* (2000). Predicting PTSD in victims of violent crime. The role of shame, anger, and childhood abuse. *Journal of Abnormal Psychology*, **109**, 69–73.
84. Carr, J.V., Lewin, T.J., Webster, R.A., *et al.* (1997). Psychosocial sequelae of the 1989 Newcastle earthquake. II. Exposure and morbidity profiles during the first 2 years post-disaster. *Psychological Medicine*, **27**, 167–78.
85. Joseph, S., Dalgleish, T., Williams, R., *et al.* (1997). Attitudes towards emotional expression and post-traumatic stress in survivors of the Herald of Free Enterprise disaster. *The British Journal of Clinical Psychology*, **36**, 133–8.
86. Bradley, R., Greene, J., Russ, E., *et al.* (2005). A multidimensional meta-analysis of psychotherapy for PTSD. *The American Journal of Psychiatry*, **162**, 214–7.
87. Bisson, J., Ehlers, A., Matthews, R., *et al.* (2007). Systematic review and meta-analysis of psychological treatments for post-traumatic stress disorder. *The British Journal of Psychiatry*, **190**, 97–104.
88. Van Etten, M.L. and Taylor, S. (1998). Comparative efficacy of treatments for post-traumatic stress disorder: a meta-analysis. *Clinical Psychology and Psychotherapy*, **5**, 126–44.
89. National Institute of Clinical Excellence. (2005). Clinical guideline 26: post-traumatic stress disorder: The management of PTSD in adults and children in primary and secondary care. http://guidance.nice.org.uk/CG26
90. Australian Centre for Posttraumatic Mental Health. Australian Guidelines for the treatment of acute stress disorder and posttraumatic stress disorder. http://www.acpmh.unimelb.edu.au/Resources/guidelines/ACMPH_PractionerGuideForASDandPTSD.pdf
91. Tarrier, N., Pilgrim, H., Sommerfield, C., *et al.* (1999). A randomised trial of cognitive therapy and imaginal exposure in the treatment of chronic post traumatic stress disorder. *Journal of Consulting and Clinical Psychology*, **67**, 13–9.
92. Foa, E.B. and Meadows, E.A. (1997). Psychosocial treatments for post-traumatic stress disorder: a critical review. *Annual Review of Psychology*, **48**, 449–80.
93. Shalev, A.Y., Bonne, O., and Eth, S. (1996). Treatment of post-traumatic stress disorder: a review. *Psychosomatic Medicine*, **58**, 165–82.
94. Marks, I.M., Lovell, K., Norshirvani, H., *et al.* (1998). Treatment of post-traumatic stress disorder by exposure and/or cognitive restructuring. *Archives of General Psychiatry*, **55**, 317–25.
95. Ehlers, A., Clark, D.M., Hackmann, A., *et al.* (2003). A randomized controlled trial of cognitive therapy, self-help booklet, and repeated assessment as early interventions for PTSD. *Archives of General Psychiatry*, **60**, 1024–32.
96. Ehlers, A., Clark, D.M., Hackmann, A., *et al.* (2005). Cognitive therapy for PTSD: development and evaluation. *Behaviour Research and Therapy*, **43**, 413–31.
97. Shapiro, F. (1995). *Eye movement desensitization and reprocessing: basic principles, protocols, and procedures*. Guilford Press, New York.
98. Davidson, P.R. and Parker, K.C.H. (2001). Eye movement desensitization and reprocessing (EMDR): a meta-analysis. *Journal of Consulting and Clinical Psychology*, **69**, 305–16.
99. Kudler, H.S., Blank, A., and Krupnick, J.L. (2000). Psychodynamic therapy. In *Effective treatments for PTSD: practice guidelines from the international society for traumatic stress studies* (eds. E.B. Foa, T.M. Keane, and M.J. Friedman), pp. 176–98 and pp. 339–41. Guilford Press, New York.
100. Stein, D.J., Ipser, J.C., and Seedat, S. (2007). Pharmacotherapy for posttraumatic stress disorder (PTSD). *The Cochrane Database of Systematic Reviews. The Cochrane Library*, **2**.
101. Foa, E.B., Davidson, J.R.T., and Frances, A. (1999). The expert consensus guideline series. Treatment of post traumatic stress disorder. *The Journal of Clinical Psychiatry*, **60**(Suppl. 16).
102. Friedman, M.J. (1998). Current and future drug treatment for post-traumatic stress disorder. *Psychiatric Annals*, **28**, 461–8.
103. National Institute for Clinical Excellence. (2005). *Posttraumatic stress disorder (PTSD): the treatment of PTSD in adults and children*. Information about NICE Clinical Guidelines 26. http://guidance.nice.org.uk/CG26

4.6.3 Recovered memories and false memories

Chris R. Brewin

Clinicians working with survivors of traumatic experiences have frequently noted the existence of memory loss with no obvious physical cause and the recovery of additional memories during clinical sessions. Indeed, amnesia is described in diagnostic manuals as a feature of post-traumatic stress disorder, although its presence is not necessary for this diagnosis. In the majority of these cases, people forget details of the traumatic event or events, or forget how they reacted at the time, although they remember that the event happened. They typically report that they have endeavoured not to think about the event, but have never forgotten that it occurred. Controversy is centred on memories of traumatic events, particularly concerning child abuse, that appear to be recovered after a long period of time in which there was complete forgetting that they had ever happened. It has sometimes been suggested that many, if not all, of these apparent recovered memories are the product of inappropriate therapeutic suggestion. This argument has been promulgated in particular by the False Memory Syndrome Foundation in the United States, by its counterpart, the British False Memory Society, and by their scientific advisors.

The 'false memory' position

Loftus[1] suggested that at least some of the memories of child sexual abuse recovered in therapy after apparent total amnesia may not be veridical, but may be false memories encouraged or 'implanted' by therapists who have prematurely decided that the patient is an abuse victim and who use inappropriate therapeutic techniques to persuade him or her to recover corresponding 'memories'. The false memory societies have claimed that there are many cases known to them in which previously happy families have been disrupted by accusations of abuse that were only triggered when an adult child entered therapy. Particular scepticism has been levelled at reports of repeated abuse, all of which has apparently been forgotten, and it has been claimed that such reports are contradicted by what is known scientifically about memory. Reports of 'repressed' memories of childhood abuse are generally regarded as clinical speculations and the psychoanalytical concept of repression as one that has no credible scientific support.

Several reviewers claim that there is no empirical support for repression or dissociative amnesia in trauma victims.[2–4]

Lindsay and Read[5] have marshalled evidence to suggest that the creation of false memories within therapy is a possibility that must be taken seriously. For example, they review experimental studies conducted with non-clinical subjects concerning the fallibility and malleability of memory, and note the potential for inaccurate recall involved in techniques such as hypnosis. Experiments have demonstrated that people are sometimes confused about whether a recent event in the laboratory actually happened, or whether they only imagined it happening. Other experiments have repeatedly succeeded in implanting apparent childhood memories of single non-abusive events in approximately 25 to 30 per cent of subjects, particularly in those who score highly on measures of hypnotizability or suggestibility.[6] Further evidence comes from individuals who claim to remember impossible events such as being kidnapped by aliens.[4]

Critics have argued that these experiments are a long way from being evidence that therapists could implant false memories of child abuse, and even the experimental studies have shown that successful suggestion depends on the plausibility of the event subjects are asked to believe in. Nevertheless, although no one has performed experiments in an attempt to implant the notion that abuse occurred, it is reasonable to argue that some patients may be highly suggestible and inclined to go along with the beliefs of therapists who may be their only source of support. If their therapist was convinced that abuse had occurred, put overt or covert pressure on their patient to 'remember' this abuse, and was insufficiently alert to the unreliability of memory, there would be a greatly increased risk of false memories occurring.

In conclusion, the recently developed 'false memory' position goes beyond previous concerns of a general nature about errors in memory, and specifically identifies a process whereby errors arise after a person has been subjected to repeated suggestive influences that the explanation for their symptoms lies in forgotten child sexual abuse. These influences are usually thought to occur in therapy, although it has been proposed that exposure to certain books or broadcast media may have the same effect. This position relies partly on information from the false memory societies about their members, and partly on experimental evidence from non-traumatic procedures in the laboratory. There has been little independent scrutiny of the data from members of false memory societies, and many of their claims, for example that parents have been falsely accused, that accusations only follow entry into therapy, or that there is a 'false memory syndrome', are anecdotal and have not been empirically verified.[7]

Evidence for genuine 'recovered memories'

Over 20 longitudinal and retrospective studies have now found that a substantial proportion of people reporting child sexual abuse (somewhere between 20 and 60 per cent) report periods in their lives (often lasting for several years) when they could not remember that the abuse had taken place.[8,9] Although the rates vary between studies, broadly similar findings have been obtained by clinical psychologists, psychiatrists, and cognitive psychologists in both clinical and community samples. As has been pointed out by critics of these studies, this evidence supports the forgetting of trauma, but does not yet have much to say about the mechanism (for example 'repression') by which it occurs. Thus it would be true to say that while there is evidence for forgetting, there is little evidence for 'repression' as such.

Three main factors support the argument that these apparently forgotten memories are not necessarily false.

1 Surveys have also found recovered memories of other traumatic experiences such as witnessing accidents, experiencing medical procedures, and physical abuse in childhood.[10] It is unclear how these could have been brought about by suggestion.

2 A number of studies have found that apparent recovered memories occur prior to any therapy, and in the absence of any obvious prolonged suggestive influence.[11] Again, it is unclear how these could have been brought about by suggestion.

3 Surveys of psychologists and therapists found that approximately 40 per cent of those with apparent recovered memories reported corroborative evidence for the content of the memories, such as abusers' confessions, testimony from other victims, and court records. Although the quality of this corroboration has been criticized, it seems unlikely that all these cases can be summarily dismissed. There are also substantial numbers of case studies reporting more detailed corroborative evidence for apparent recovered memories, some of this evidence of reasonably high quality.[10–12]

The quality of the research evidence supporting genuine recovered memories is mixed, and almost all the studies can be argued to have some flaws, but taken together the evidence for genuine memories of major traumatic events is far more extensive than the evidence for false memories of such events. Moreover, these observations need not, as has sometimes been claimed, contradict what we know about memory. Cognitive psychology recognizes that ordinary memory relies as much for its efficiency on the ability to inhibit unwanted material as on the ability to gain rapid access to relevant material. Experimental studies clearly demonstrate the inhibition of memory retrieval and the existence of a subgroup of individuals with poor memories for negative experiences.[11]

The intimate neuroanatomical connections between brain circuits involved in emotion and those involved in memory provide a good reason for believing that memory may not behave in the same way under conditions of extreme real-world stress as it does in ordinary laboratory experiments. Whereas high levels of arousal often make events more difficult to forget, it has been argued by several well-known neuroscientists that extraordinarily high levels of catecholamines or other neuropeptides at the time of the trauma, perhaps in combination with a failure to release sufficient cortisol, may produce amnesia. More specifically, it has been suggested that extreme stress produces at the same time both enhanced fear conditioning and impaired autobiographical memory.[13] Much of the evidence is indirect and not yet compelling, but it illustrates that claims concerning recovered memories of trauma need not violate current knowledge concerning the cognitive psychology and neurobiology of memory.

Why the debate?

From a purely scientific point of view, it should be evident that the quality of the available evidence is insufficient to justify any extreme position at present. The questions are extremely difficult to study

empirically, and there has been little new research since the 1990s to suggest that this state of affairs is likely to change. However, scientific considerations have sometimes been secondary to the passionate advocacy practised by parents who claim to be falsely accused, and by accusers who claim that their memories of abuse are being ignored. Psychiatrists and psychologists have in the past become caught up in the debate and in some cases abandoned any pretence at neutrality. In the face of these compelling and intensely painful personal concerns, the quality of much of the argument became debased. Thus, supposedly scientific contributions on both sides of this debate questioned the motives and integrity of people with whom they disagreed and attempted to disparage opponents' professional abilities. Some of these same authors made exclusive claims for the scientific legitimacy of their own perspective, subjecting opposing data to fierce scrutiny while being relatively uncritical of studies that supported their point of view. Much of the literature was obfuscatory and confusing. Logical errors abounded, seen for example in the conclusion that because a memory has been recovered in therapy, the practitioner must have been using 'recovered memory therapy'.

A good example of the debate in action is the article by Pope *et al.*[4] on the evidence for dissociative amnesia in trauma victims and the commentary that followed it.[14] These articles demonstrate how widely differing conclusions can be drawn from the same set of studies, depending on the way terms are defined, on assumptions about what evidence should be given the most weight, and on the rigour with which alternative explanations are evaluated.

An emerging scientific and professional consensus

What should be clear by now is that extreme views, claiming that either false memories or genuine recovered memories are rare or impossible, cannot be supported by the available data. Nevertheless, the dispute continues about whether traumatic events, and particularly repeated traumas, can be forgotten and then remembered with essential accuracy. In my view it is safe to conclude from the evidence reviewed that the hypothesized implantation of false memories by practitioners cannot account for more than a subset of recovered memories (and at present it is entirely unclear how large or small this subset might be). False memories may certainly arise in other circumstances, but as yet there is little pertinent evidence. On the other hand, there is a great deal of plausible evidence supporting the existence of genuine recovered memories.

Most commentators, including some members of the advisory boards of false memory societies,[6] now accept that traumatic events can be forgotten and then remembered. Cognitive psychologists Lindsay and Read[5] summed it up well: 'In our reading, scientific evidence has clear implications . . . memories recovered via suggestive memory work by people who initially denied any such history should be viewed with scepticism, but there are few grounds to doubt spontaneously recovered memories of common forms of child sexual abuse or recovered memories of details of never-forgotten abuse. Between these extremes lies a grey area within which the implications of existing scientific evidence are less clear and experts are likely to disagree'. Similarly, the consensus view among independent commentators, repeated in the 1995 report of the British Psychological Society's Working Party on Recovered Memories and the 1995 interim statement of the American Psychological Association's Working Group on Investigation of Memories of Childhood Abuse, is that memories may be recovered from total amnesia and they may sometimes be essentially accurate. Equally, such 'memories' may sometimes be inaccurate in whole or in part.

In practical terms, the debate has had two major effects. First, proponents of 'recovered memory therapy' are now almost impossible to find within the ranks of leading psychiatrists and psychologists. Despite the small amount of empirical support, there is widespread agreement that situations in which there is sustained suggestive influence, such as therapy, do have the potential to induce false memories. Active attempts to recover suspected forgotten memories may sometimes be appropriate in unusual or extreme cases, but both the client and the therapist must be aware of the risk of false memories. Techniques such as hypnosis and guided imagery should not be used without safeguards against potential suggestive influence. Second, good practice now requires both the therapist and the client to adopt a critical attitude towards any apparent memory that is recovered after a period of amnesia, whether or not this is within a therapeutic context, and not to assume that it necessarily corresponds to a true event. Even highly vivid traumatic memories (sometimes known as 'flashbacks') may be misleading or inaccurate in some cases. Clinical guidelines are now available to help the practitioner avoid the twin perils of uncritically accepting false memories as true or summarily dismissing genuine recovered memories.[9,15]

Further information

Brewin, C.R. (2003). *Posttraumatic stress disorder: malady or myth?* Yale University Press, New Haven, CT (see Chapters 7 and 8).

Davies, G.M. and Dalgleish, T. (eds.) (2001). *Recovered memories: seeking the middle ground*. Wiley, Chichester.

References

1. Loftus, E.F. (1993). The reality of repressed memories. *The American Psychologist*, **48**, 518–37.
2. Pope, H.G. and Hudson, J.I. (1995). Can memories of childhood abuse be repressed? *Psychological Medicine*, **25**, 121–6.
3. Pope, H.G., Hudson, J.I., Bodkin, J.A., *et al.* (1997). Can trauma victims develop 'dissociative amnesia'? the evidence of prospective studies. *The British Journal of Psychiatry*, **172**, 210–15.
4. McNally, R.J. (2003). *Remembering trauma.* Harvard University Press, Cambridge, Mass.
5. Lindsay, D.S. and Read, J.D. (1995). 'Memory work' and recovered memories of childhood sexual abuse: scientific evidence and public, professional and personal issues. *Psychology, Public Policy and the Law*, **1**, 846–908.
6. Wright, D., Ost, J., and French, C.C. (2006). Recovered and false memories. *The Psychologist*, **19**, 352–5.
7. Pope, K.S. (1996). Memory, abuse and science: questioning claims about the False Memory Syndrome epidemic. *The American Psychologist*, **51**, 957–74.
8. Gleaves, D.H., Smith, S.M., Butler, L.D., *et al.* (2004). False and recovered memories in the laboratory and clinic: a review of experimental and clinical evidence. *Clinical Psychology: Science and Practice*, **11**, 3–28.
9. Mollon, P. (1998). *Remembering trauma: a psychotherapist's guide to memory and illusion*. Wiley, Chichester.

10. Andrews, B., Brewin, C.R., Ochera, J., *et al.* (1999). The characteristics, context, and consequences of memory recovery among adults in therapy. *The British Journal of Psychiatry*, **175**, 141–6.
11. Brewin, C.R. (2003). *Posttraumatic stress disorder: malady or myth?* Yale University Press, New Haven, Conn.
12. Cheit, R. (2005). *The recovered memory project*. Internet posting http://www.brown.edu/Departments/Taubman_Center/Recovmem/Archive.html
13. Elzinga, B.M. and Bremner, J.D. (2002). Are the neural substrates of memory the final common pathway in posttraumatic stress disorder (PTSD)? *Journal of Affective Disorders*, **70**, 1–17.
14. Brewin, C.R. (1998). Commentary: questionable validity of 'dissociative amnesia' in trauma victims. *The British Journal of Psychiatry*, **172**, 216–17.
15. Pope, K.S. and Brown, L.S. (1996). *Recovered memories of abuse: assessment, therapy, forensics*. American Psychological Association, Washington, DC.

4.6.4 Adjustment disorders

James J. Strain, Kimberly Klipstein, and Jeffrey Newcorm

Introduction

The psychiatric diagnoses that arise between normal behaviour and major psychiatric morbidities constitute the problematic subthreshold disorders. These subthreshold entities are also juxtaposed between problem-level diagnoses and more clearly defined disorders. Adjustment disorder (AD) would 'trump' problem-level disorders, but would be 'trumped' by a specific diagnosis even if it were in the NOS category. The subthreshold disorders present major taxonomical and diagnostic dilemmas in that they are often poorly defined, overlap with other diagnostic groupings, and have indefinite symptomatology. It is therefore not surprising that issues of reliability and validity prevail. One of the most commonly employed subthreshold diagnosis that has undergone a major evolution since 1952 is AD (Table 4.6.4.1). The advantage of the indefiniteness of these subthreshold disorders is that they permit the classification of early or prodromal states when the clinical picture is vague and indistinct, and yet the morbid state is in excess of that expected in a normal reaction and this morbidity needs to be identified and often treated. Therefore, AD has an essential place in the psychiatric taxonomy.

Many questions prevail with regard to the concept of the AD diagnosis: (1) the role of stressors and the place of specific stressors; (2) the importance of age; (3) the role of concurrent medical morbidity, for example comorbidity of Axis I and/or Axis III disorders; (4) the lack of specificity of the diagnostic criteria; (5) the absence of a symptom checklist; (6) uncertainty as to optimal treatment protocols; and (7) undocumented prognosis or outcomes. Research data regarding these questions will be examined.

The DSM was conceptually designed with an atheoretical framework to encourage psychiatric diagnoses to be derived on phenomenological grounds with an avowed dismissal of pathogenesis or aetiology as diagnostic imperatives. In frank contradiction to this atheoretical conceptual framework, the stress-induced disorders require the inclusion of an aetiological significance to a life event—a stressor—and the need to relate the stressor's effect on the patient in clinical terms. However, the stress-related disorders are unique in that they are psychiatric diagnoses with a known aetiology—the stressor—and thus aetiology is essential for the diagnosis. Four other diagnostic categories also invoke aetiology in their diagnostic criteria: (1) organic mental disorders (aetiology-physical abnormality); (2) substance abuse disorders (aetiology-ingestion of substances); (3) post-traumatic; and (4) acute stress disorders

Table 4.6.4.1 DSM-IV Evolution of the diagnosis for adjustment disorder

(a) The development of emotional or behavioural symptoms in response to an identifiable psychosocial stressor(s), which occurs within 3 months of the onset of the stressor(s)

(b) These symptoms or behaviours are clinically significant as evidenced by either of the following

1 Marked distress that is in excess of what would be expected from exposure to the stressor

2 Significant impairment in social or occupational (academic) functioning

(c) The stress-related disturbance does not meet the criteria for any specific Axis I disorder and is not merely an exacerbation of a pre-existing Axis I or Axis II disorder

(d) The symptoms do not represent bereavement

(e) Once the stressor (or its consequences) has terminated, the symptoms do not persist for more than an additional 6 moths

Specify if:

Acute: if the disturbance lasts less than 6 months.

Persistent/chronic: if the disturbance lasts for 6 months or longer

AD is a stress-related phenomenon in which the stressor precipitates maladaptation and symptoms that are time limited until either the stressor is diminished or eliminated, or a new state of adaptation to the stressor occurs (Table 4.6.4.2). At the same time that AD was evolving, other stress-related disorders, for example, post-traumatic stress disorder and acute stress disorder were described. (Acute stress disorder was formulated by Spiegel during the development of the DSM-IV.[1,2]) Acute stress reactions could result from involvement in a natural disaster such as a flood, or an avalanche, or a cataclysmic personal event, for example, loss of a body part (aetiology-an identifiable stressor).

The diagnosis of AD also requires a careful titration of the timing of the stressor in relation to the adverse psychological sequelae that ensue. Maladaptation and disturbance of mood should occur within 3 months of the patient experiencing the stressor. Until the DSM-IV criteria, the ADs were regarded as transitory diagnoses that should not exceed 6 months in duration. Thereafter, that diagnostic appellation could not be employed and had to be changed to a major psychiatric disorder or discontinued.

Definition and history

With the opportunity in 1994 to develop another evolutionary step of the DSM, i.e. DSM-IV,[3] the authors were asked to re-examine the subthreshold diagnostic category of AD with the goal of

Table 4.6.4.2 ICD-10 definition of adjustment disorder

(a) Onset of symptoms must occur within 1 month of exposure to an identifiable psychosocial stressor, not of an unusual or catastrophic type
(b) The individual manifests symptoms or behavioural disturbances of the types found in any of the affective disorders (except for delusions and hallucinations), any disorders in F40–F48 (neurotic, stress-related, and somatoform disorders) and conduct disorders, but the criteria for an individual disorder are not fulfilled. Symptoms may be variable in both form and severity
The predominant feature of the symptoms may be further specified by use of a fifth character:
Brief depressive reaction Prolonged depressive reaction Mixed anxiety and depressive reaction With predominant disturbance of other emotions With predominant disturbance of conduct With mixed disturbance of emotions and conduct With other specified predominant symptoms
(c) Except in prolonged depressive reaction, the symptoms do not persist for more than 6 months after the cessation of the stress or its consequences. However, this should not prevent a provisional diagnosis being made if this criterion is not yet fulfilled

improving its acknowledged 'shortcomings'. The research included: review of the literature, reanalysis of existing studies of AD and their data sets, and examination of field studies (e.g. minor depression, minor anxiety) to observe if there was sufficient differentiation among these minor disorders from the ADs (e.g. how often was a stressor identified in those patients assigned the diagnosis; minor depression or minor anxiety?). From these three sources and consultations, modifications for DSM-IV and their rationale were formulated based on the best evidence extant.

Changes in the criteria for adjustment disorder in DSM-IV

The review of the literature, the reanalysis of the Western Psychiatric Institute and Clinic data (University of Pittsburgh), and consultations with experts supported the following changes in DSM-IV.

1 Enhance the understanding of the language.

2 Describe the timing of the reaction to reflect the duration of the AD: **acute** (less than 6 months) or **chronic** (6 months or greater).

3 Allow for the continuation of the stressor for an indefinite period; psychological reactions to chronic stress states (e.g. chronic arthritis, HIV, abuse by an alcoholic spouse) do not necessarily terminate at 6 months, nor do they necessarily lead to a major psychiatric disorder.

4 Eliminate the subtypes of mixed emotional features, work (or academic) inhibition, withdrawal, and physical complaints (as they were rarely employed by diagnosticians).

Although it might be argued that ADs could be placed in a new category of 'stress response syndromes', the literature and research reports did not support such a taxonomical organization. Another possibility was that AD could be eliminated altogether, with the advantage of maintaining the atheoretical approach of the DSM conceptual framework, and substitute in its place the appropriate minor or NOS categories as established by the accompanying mood states or behaviours. However, these solutions do not seem appropriate with recent findings that demonstrate AD to be a valid and frequently employed diagnosis.[4] AD was diagnosed in over 60 per cent of burned inpatients,[5] over 20 per cent of patients in early stages of multiple sclerosis,[6] and over 40 per cent of poststroke patients.[7] Furthermore, evaluations of patients in a psychiatric walk-in clinic showed a significant difference in the symptom profile of those assigned AD and the other diagnosis, including minor diagnoses.[8] (The McArthur field trials on the prospective assessment of minor depressive and anxiety disorders which collected data on the occurrence of stressors immediately preceding the outbreak of symptoms are important databases that need further study to establish whether stress per se is a distinguishing characteristic between AD and the other minor mood disorders.)

Problems with the adjustment disorder diagnosis

(a) The symptom profile

Critics of the AD diagnosis argue that the symptom complex is too subjective or 'depends structurally on clinical judgement' as opposed to sound, operational criteria.[9,10] Because of the lack of any quantitative behavioural or operational criteria, the problem of reliability and validity are obvious. Criterion reference was evaluated by Aoki *et al.*[11] who reported that three psychological tests, Zung's Self-Rating Anxiety Scale,[12] Zung's Self-Rating Depression Scale,[13] and the Profile of Mood States,[14] were useful tools for the diagnosis of AD in physical rehabilitation patients. While these measures succeeded in reliably differentiating AD patients from normal patients, they were not able to distinguish them from patients with major depression or post-traumatic stress disorders. Thus, the construct of AD is designed as a means for classifying psychiatric conditions having a symptom profile that is at the time of its application insufficient to meet the more specifically operationalized criteria for the major syndromes but is:

1 clinically significant and deemed to be in excess of a normal reaction to the stressor in question;

2 associated with impaired vocational or interpersonal functioning;

3 not solely the result of a psychosocial problem (V Code) requiring medical attention (e.g. non-compliance, phase of life problem, etc.).

However, field studies are being performed[15] to assess whether a reliable checklist from an elaborate list of symptoms associated with AD can be constructed (Table 4.6.4.3). (The V Codes—a problem level of diagnoses—are understandably devoid of a symptom-based diagnostic schema.)

Table 4.6.4.3 DSM-IV subtypes of adjustment disorders

Adjustment disorder with depressed mood
Adjustment disorder with anxious mood
Adjustment disorder with mixed anxiety and depressed mood
Adjustment disorder with disturbance of conduct
Adjustment disorder with mixed disturbance of emotions and conduct
Adjustment disorder unspecified

(b) The meaning of 'maladaptive'

The imprecision of the diagnostic criteria for AD is immediately apparent in the DSM-IV description of this disorder as a maladaptive reaction to an identifiable psychosocial stressor, or stressors that occurs within 3 months after onset of the stressor. It is assumed that the disturbance will remit soon after the stressor ceases or, if the stressor persists, when a new level of adaptation is achieved.[1] In addition to the problem of no symptom checklist, difficulties are inherent within each of these diagnostic elements.

First, with regard to the 'maladaptive reaction', it is unclear how this concept can or should be operationalized. Is the assessment of maladaptation subjective or objective? Who makes the decision—a third party, a mental health professional, the patients themselves, or an admixture of these? Is this decision 'culture bound'? Succinctly when does an individual cross the threshold into 'patienthood', and who will make the decision? Powell and McCone (2004) make this point in their case report of the treatment of a patient with AD secondary to the stressors of the 11 September terrorist attacks. Since there has never before been a large-scale terrorist attack in United States, how are clinicians to know what a 'normal' response to such an event would be?[16]

(c) The stressor

Most recently, in the DSM-IV text revision (DSM-IV-TR; American Psychiatric Association, 2000),[17] the term psychosocial stressor was changed to the broader concept of stressor. Emotional reactions to physical stress, such as the Chernobyl reactor incident[18] or cardiac surgery[19] are well documented in literature and suggest that psychosocial stressor as a criterion is too restrictive. Moreover, the concept of 'psychosocial' versus 'physical' stressors has led to confusion.[20]

Obviously, the stressor and its effect are central to the AD diagnosis. The second major confound emanates from the fact that the DSM-IV presents no criteria or 'guidelines' to quantify stressors or to assess their effect or meaning for a particular individual at a given time. Furthermore, the assessment of stress is not linked by an algorithm to Axis IV—a statement of stress—during the previous year and so internal consistency or reinforcement within the diagnostic lexicon is not mandated (D. Schafer, personal communication, 1990). Mezzich *et al.*[8] attempted to classify and quantify the psychosocial stressors in 13 specific domains: health, bereavement, love and marriage, parental, family stressors for children and adolescents, other familial relationships, other relationships outside the family, work, school, financial, legal, housing, and miscellaneous. Such specificity has not been defined in DSM-IV and the construct is vague and generic with minimal opportunity to achieve quantification. Despland *et al.*[21] observed that the type of stressor may indeed be of help in diagnosing AD. His study demonstrated that AD with depressed mood and mixed mood was associated with more marital problems than major depressive disorders. AD with anxiety could be distinguished from the major anxiety disorders by the quantity of family and marital problems.

(d) The time course

The time course and chronicity of both stressors and their consequent symptoms were left vague in DSM-IIIR and were not consistent with the clinical situation. The modifications introduced in DSM-IV, which differentiate between *acute* and *chronic* forms of AD, solved the problem of the 6-month limitation of the AD diagnosis in DSM-IIIR and is more in keeping with what is observed in the clinical situation. This change was validated by Despland *et al.*[21] who observed that 16 per cent of patients with AD required treatment longer than 1 year—the mean exceeded the prior limitation of 6 months.

Other problems of definition

Even serious symptomatology (e.g. suicidal behaviour) that is not regarded as part of a major mental disorder requires treatment and a 'diagnosis' under which it can be placed, for example a V Code, 'Phase of Life Problem', AD, acute stress response, etc. De Leo *et al.*[22,23] reported on AD and suicidality. Recent life events, which would constitute an acute stress, were commonly found to correlate with suicidal behaviour in a patient cohort which included those with AD.[24] Spalletta *et al.*[25] observed the assessment of suicidal behaviour to be an important tool in the differentiation among major depression, dysthymia, and AD. AD patients were found to be among the most common recipients of a deliberate self-harm diagnosis, with the majority involving self-poisoning.[26] Thus deliberate self-harm is more common in AD patients,[26] while the percentage of suicidal behaviour was found to be higher in AD patients with depressed mood.[25]

The AD DSM-IV Work Group suggested that suicide and deliberate self-harm could be subtypes of AD. However, there were concerns that patients with other diagnoses, for example major affective disorder, borderline personality disorder, etc. and suicide behaviour, would be assigned the AD diagnosis since there was a specific placement for suicidal ideation and behaviour and that would be a predominant reason to use AD. The final decision was to place the problem of suicidal symptomatology without a psychiatric diagnosis in the DSM-IV F Code section for other problems 'that may be a focus of clinical attention'. Obviously a subthreshold diagnosis, AD, does not necessarily imply the presence of subthreshold symptomatology!

Recognizing some of the limitations of the diagnosis including the aforementioned lack of specificity of symptoms and the lack of clarity of the role of the stressor, the authors of a recent article proposed adding an additional 'A-Criterion' to the DSM IV diagnosis of AD. They studied 328 young conscripts diagnosed by DSM IV with AD secondary to non-combat military stress. The diagnosis was closely associated with undisturbed psychosocial function outside of military life but with marked symptoms within military life.[9] Thus, location-specific stress was associated with location-specific symptoms, a phenomenon that the authors found helpful in distinguishing AD from other psychiatric diagnoses. Whether or not this finding would be consistent in non-military populations requires further evaluation.

'Splitting' and 'lumping' continue, for example, the subthreshold diagnosis of mixed anxiety-depressive disorder is a new category included in the DSM-IV. This disorder is very similar to AD with mixed mood; a boundary between the two is difficult to demarcate. The main difference between the two diagnoses was the chronicity of the mixed anxiety-depressive disorder (as was noted in the mixed anxiety-depression field trial).[27] The change in criterion C for AD—allowing a chronic or recurrent disturbance—confounds the differentiation of these two subthreshold diagnoses. This uncertainty is further complicated by the question of treatment. Is this

an anxiety accompanied by depression, which should be treated with anxiolytics, such as benzodiazepines, or is this a depression accompanied by anxiety, which should be treated with an antidepressant, such as a selective serotonin reuptake inhibitor (SSRI)? Furthermore, it is commonly viewed that the majority of patients with AD should be treated with psychotherapy or counselling as the initial approach.

Another potential mood disorder, subsyndromal symptomatic depression (**SSD**), has been suggested.[28] It joins AD in the grey area of subthreshold diagnoses. However, there are two critical differences between SSD and AD: SSD employs a symptom checklist, and is not associated with a stressor. By definition, SSD is the simultaneous presence of any two or more symptoms of depression, persistent for most or all of the time for a duratio of at least 2 weeks, associated with social dysfunction, and occurring in patients who do not meet the criteria for minor depression (which also requires two symptoms), major depression, and/or dysthymia.[28] In some cases, the SSD diagnosis is the same as the DSM-IV diagnosis for minor depression, termed by the authors 'SSD with mood disturbance', and has to be documented as such. In other cases, the disorder is 'SSD without mood disturbance'.

In a recent study, Casey *et al.* (2006) examined variables that might distinguish AD from other depressive episodes. The patients were screened for depression severity with the Beck Depression Inventory (BDI) and then interviewed with the Schedule for Clinical Assessment in Neuropsychiatry (SCAN) which includes questions assessing the presence of AD. The authors were unable to find any independent variables that distinguished AD from other depressive episodes, including the severity of the BDI score at the outset.[29] Maercker *et al.* conceptualize AD as a stress response syndrome in which intrusions, avoidance of reminders, and failure to adapt are the central processes and symptoms.[30]

Age and medical comorbidity

In contrast with DSM-IIIR, DSM-IV has tried to accommodate the presence of comorbid medical illness. DSM-IIIR was regarded as 'medical illness and age unfair' (i.e. inadequate consideration of age and/or medical illness) (L. George, personal communication, 1981).[31] To enhance reliability and validity, there will need to be a psychiatric taxonomy that takes into account medical illness and symptomatology and developmental epochs (e.g. children and adolescents, adults, 'young' elderly, and 'old' elderly). (Actually, the original DSM did divide the AD by developmental epochs.) It is clear from the Western Psychiatric Institute studies that the symptom profile for children and adolescents is very different from that for adults with regard to the entire spectrum of diagnoses. With regard to age, recent studies report AD patients to be significantly younger compared with those with major psychiatric diagnosis.[21,32] Zarb's study[33] suggests that in cognitively impaired elderly, using individual items of the Geriatric Depression Scale, AD could be differentiated from major depression. In addition, Despland *et al.*[21] showed that patients labelled AD with depressive or mixed symptoms included more women: a sex ratio resembling that seen in major depression or dysthymia. The future evolution of the DSM needs to consider the effect of developmental epochs, gender, and medical comorbidity on symptom profiles in the various diagnostic categories.

Grassi *et al.*[34] investigated psychosomatic symptoms in patients with AD in a hospital setting in order to further characterize the diagnosis of AD in the medically ill. Results showed a considerable overlap between AD and abnormal illness behaviour including health anxiety, somatization, conversion symptoms, and demoralization among others. Only 13 out of 100 AD patients interviewed did not present with psychosomatic symptoms.

Epidemiology

The Epidemiologic Catchment Area Study (ECA) did not include AD in its historic survey of patients in the population of five major settings in the United States. Most studies are of smaller or more discrete samples and have the problem of generalization. Andreasen and Wasek[35] reported that 5 per cent of an inpatient and outpatient sample at the university hospital and clinics in Iowa were labelled as having AD. Fabrega *et al.*[36] reported that 2.3 per cent of a sample of patients presenting to a walk-in clinic (diagnostic and evaluation centre) met criteria for AD, with no other diagnoses on Axis I or Axis II; when patients with other Axis I diagnoses (i.e. Axis I and II comorbidities) were included, 20 per cent had the diagnosis of AD. In general hospital inpatient psychiatric consultation populations, AD was diagnosed as 21.5 per cent, 18.5 per cent, and 11.5 per cent, respectively.[37–39] D. Schafer (personal communication, 1990) noted that up to 70 per cent of children in the psychiatric setting may be given the diagnosis of AD in a variety of mental health care environments. Faulstich *et al.*[40] reported the prevalence of AD for adolescent psychiatric inpatients. Andreasen and Wasek,[35] utilizing a chart review, reported that more adolescents than adults experienced acting out and behavioural symptoms, but adults had significantly more depressive symptomatology (87.2 per cent versus 63.8 per cent). Anxiety symptoms were frequent at all ages.

Mezzich and coworkers[8] evaluated 64 symptoms currently present in three cohorts: subjects with specific diagnoses, those with AD, and those who were not ill. Vegetative, substance use, and characterological symptoms were greatest in the specific diagnosis group, intermediate in the AD group, and least in the group with non illness. The symptoms of mood and affect, general appearance, behaviour, disturbance in speech and thought pattern, and cognitive functioning had a similar distribution. The AD group was significantly different from the non-illness group with regard to more 'depressed mood' and 'low self-esteem' ($p \leq 0.0001$). The AD and non-illness groups had minimal pathology of thought content and perception. A positive response on the suicide indicators was obtained in 29 per cent of AD compared with 9 per cent of the non-illness group. The three cohorts did not differ on the frequency of Axis III disorders.

Associated features of adjustment disorder

Andreasen and Wasek[35] observed that in their AD cohorts 21.6 per cent of the adolescents' fathers and 11.8 per cent of the adults' fathers had problems with alcohol. Greenberg *et al.*[41] report more substance abuse in adults with diagnosed AD compared with all those with other diagnoses. Breslow *et al.*[42] comparing patients with AD and other psychiatric diagnoses, observed that alcohol or substance use/abuse did not help to differentiate between diagnostic groups. Thus, currently the higher rate of substance use does not serve as an incontrovertible predictive factor for the diagnosis of an AD diagnosis.

Aetiology—the role of stress

(a) Nature of the stressor

Andreasen and Wasek[35] described the differences between the chronicity of stressors found in adolescents compared with those in adults: present for a year or more, 59 per cent and 35 per cent; present for 3 months or less, 9 per cent and 39 per cent. Fabrega *et al.*[36] reported that their AD group had greater registration of stressors compared with other diagnoses and the non-illness cohorts. Compared with other diagnoses and the non-illness patients, AD was over-represented in the 'higher stress category'. In their consultation cohort, Popkin *et al.*[37] found that in 68.6 per cent of the cases the medical illness itself was judged to be the primary psychosocial stressor. Snyder and Strain[39] observed that stressors as assessed on Axis IV were significantly higher (p = 0.0001) for consultation patients with AD than for patients with other diagnostic disorders.

(b) Modifiers of stress

Stress has been described as the aetiological agent for AD. Vulnerability to stress is another risk factor. Diverse variables and modifiers are involved regarding who will experience AD following a stress. Cohen[43] argues as follows:

1 acute stresses are different from chronic stresses in both psychological and physiological terms;

2 the meaning of the stress is affected by 'modifiers' (e.g. ego strengths, support systems, prior mastery);

3 manifest and latent meanings of the stressor(s) may be associated with differential impact (e.g. loss of job may be a relief or a catastrophe).

An objectively overwhelming stress may have little impact on one individual, whereas a minor stress could be regarded as cataclysmic by another. A recent minor stress superimposed on a previous underlying (major) stress that has no observable effect on its own may have a significant additive impact (i.e. concatenation of events) (B. Hamburg, personal communication, 1990). Despland *et al.*[21] reported that stressors were present on Axis IV in 100 per cent of those assigned AD with depressed mood, while it was present in 83 per cent of those with major depression, 80 per cent of those with dysthymia, and 67 per cent of those with non-specific depression, which emphasizes the importance of stressors in the AD diagnosis.

Clinical features

Nine different types of AD are listed in DSM-IV.[1] As in DSM-III, AD is classified in DSM-IIIR according to the predominant symptom picture. In DSM-IV, AD has been reduced to six types that, again, are classified according to their clinical features:

1 AD with depressed mood;

2 AD with anxious mood;

3 AD with mixed anxiety and depressed mood;

4 AD with disturbance of conduct;

5 AD with mixed disturbance of emotions and conduct; and

6 AD not otherwise specified.

In their study, Despland *et al.*[21] suggested reducing the subtypes even further, demonstrating identical profiles for AD with depressed mood and AD with mixed mood, and proposing assimilation of mixed mood into the depressed mood category. Fifty-seven per cent of their sample was represented by these two groups; the remainder was accounted for by AD with 'anxiety' and 'other' categories.

Treatment

(a) Evidence regarding treatment

In terms of randomized controlled trials (RCTs), a search of the Cochrane Database revealed only two RCTs on psychotherapeutic treatment of AD. Gonzales-Jaimes and Turnbull-Plaza (2002) showed that 'mirror psychotherapy' for patients suffering from AD with depressed mood secondary to a myocardial infarction was both an efficient and effective treatment. Mirror therapy is described as a type of therapy with psychocorporal, cognitive, and neurolinguistic components with a holistic focus. As part of the treatment, a mirror is used to encourage patient acceptance of his/her physical condition that resulted from past self-care behaviours. In this study, mirror therapy was compared to two other treatments, Gestalt psychotherapy or medical conversation in addition to a control group. Depressive symptoms improved in all treatment groups compared with the control group, but mirror therapy appeared significantly more effective than the other treatments in decreasing symptoms of AD at post-test evaluation.[44]

In a second RCT, an 'activating intervention' was carried out for the treatment of AD which had resulted in occupational dysfunction. A total of 192 employees were randomized to receive either the intervention or care as usual.[45] The intervention consisted of an individual cognitive behavioural approach to a graded activity, similar to stress inoculation training. Goals of treatment emphasized the acquisition of coping skills and the regaining of control. The treatment proved to be effective in decreasing sick leave duration and shortening long-term absenteeism when compared to the control group; both intervention and control groups, however, showed similar amounts of symptom reduction. This study formed the basis for the Dutch Practice Guidelines for the treatment of AD in primary and occupational health care.[46] These guidelines were prepared by a team of 21 occupational health physicians and one psychologist and subsequently reviewed and tested by 15 experts, including several psychiatrists and psychologists and 21 practicing occupational health physicians.

Though no other RCTs involving the psychotherapeutic treatment of pure cohorts of patients with AD could be found, many RCTs exist that studied an array of depressive and anxiety disorders and that included AD in their cohorts. For example, a recent trial comparing brief dynamic therapy (BDT) with brief supportive therapy (BSP) in patients with minor depressive disorders, including AD, was found in the Cochrane Database. Though both therapies proved efficacious in reducing symptoms, BDT was more effective as demonstrated in a 6 months follow-up.[47]

Another therapeutic modality, eye movement desensitization and reprocessing (EMDR) has been recently studied in patients with AD.[48] EMDR, a psychotherapeutic technique shown to be effective in the treatment of post-traumatic stress disorder, was carried out on nine patients suffering from AD. Results showed significant improvement in patients with anxious or mixed features but not in those with depressed mood. Additionally, those with ongoing stressors did not show improvement.

Hameed *et al*. in a retrospective chart review sought to determine if there was a difference in antidepressant efficacy in the treatment of major depressive disorder versus AD in a primary care setting. Patients had been prescribed mostly SSRIs. DSM-IV symptoms, Patient Health Questionnaire-9 (PHQ-9) depression rating scale scores, and functional disability reports were systematically used to assess patients' response. Results showed that neither depressed, nor AD patients demonstrated a difference in clinical response to any particular antidepressant. Patients with a diagnosis of AD, however, were twice as likely to respond to standard antidepressant treatment as depressed patients. This study suggests that antidepressants are very effective in treating depression in the primary care setting and may be even more effective in the treatment for AD with depressed mood.[49]

In another recent double-blind RCT, the efficacies of etifoxine, a non-benzodiazepine anxiolytic drug, and lorazepam, a benzodiazepine, were compared in the treatment of AD with anxiety in a primary care setting.[50] Efficacy was evaluated on days 7 and 28 using the Hamilton Rating Scale for Anxiety (HAM-A). The two drugs were found to be equivalent in anxiolytic efficacy on day 28. However, more etifoxine recipients responded to the treatment. Moreover, 1 week after stopping treatment, fewer patients taking etifoxine experienced rebound anxiety, compared to those given lorazepam.[50]

Management

(a) Psychotherapy and counselling

Though brief therapeutic interventions are usually all that are needed, ongoing stressors or enduring character pathology that may make a patient vulnerable to stress intolerance may signal the need for lengthier treatments.[51]

Treatment of AD initially focuses on psychotherapeutic and counselling interventions to reduce the stressor, enhance the capacity to cope with a stressor that cannot be reduced or removed, and establish a system of support to maximize adaptation. The patient needs to be made aware of the significant dysfunction that the stressor has caused and consider strategies to manage the disability. Some stressors, for example taking on more responsibility than can be managed by the individual or putting oneself at risk (e.g. unprotected sex with an unknown partner), can be avoided or minimized. Other stressors may elicit an overreaction on the part of the patient (e.g. abandonment by a lover). The patient may attempt suicide or become reclusive, damaging his or her source of income. In this situation, the therapist would assist the patient to verbalize his or her disappointed feelings and rage rather than behaving destructively. The role of verbalization in minimizing the discomfort of the stressor and enhancing coping cannot be overestimated. It is necessary to clarify and interpret the meaning/reality of the stressor for the patient. For example, if a mastectomy has devastated a patient's feelings about her body and herself, it is mandatory to articulate that the patient is still a woman, capable of having a fulfilling relationship, including a sexual one and that recurrence of the cancer may not occur. Without the correction of distortions, the patient's pernicious fantasies—'all is lost'—may occur as sequelae to the stressor (i.e. the mastectomy) and intensify incapacitation at work and/or sex, as well as contribute to a profound disturbance of mood.

Counselling, psychotherapies, crisis intervention, family therapy, and group treatment are utilized to encourage the verbalization of fears, anxiety, rage, helplessness, and hopelessness to the stressors imposed upon a patient. As mentioned above, the goal of treatment is to expose the concerns and conflicts that the patient is experiencing, identify means to reduce the stressor(s), enhance the patient's coping skills, clarify the patient's perspective on the adversity, and enable the establishment of supporting relationships. The primary treatment for AD is talking.

(b) Psychopharmacotherapy

Should drugs be used in the treatment of AD? The pharmacological studies are not conclusive. The diagnostic dilemmas of the AD present sufficient difficulty in and of themselves.[52–54] It would be preferred that cautious psychotropic drug administration be employed, to avoid subjecting the patient to the risk of unfavourable medical drug-psychotrophic drug interaction(s). Psychotrophic drug treatment will not be necessary if the condition resolves. If it evolves into a major psychiatric illness then drug treatment needs to be actively entertained. And, for a refractory AD treatment with psychopharmacological agents should be considered. Small doses of antidepressants and anxiolytics may sometimes be appropriate for AD patients when dysphoria remains profound despite several sessions of psychological treatment.

Although formal psychotherapy is presently the treatment of choice, psychotherapy combined with benzodiazapines are utilized, especially for patients with severe life stress(es) and an unrelenting anxious component. Tricyclic antidepressants or buspirone were recommended in place of benzodiazapines for AD patients with current or past excessive alcohol use because of their greater risk of dependence.[55] The use of antidepressants may assist some patients if their maladaptation is debilitating and the accompanying mood is pervasive, especially if a trial of psychotherapy has been shown to be ineffective.

Adults and adolescents

Andreasen and Hoenk[56] report that the long-term outcome of AD has a good prognosis for adults, but that a majority of adolescents eventually have major psychiatric disorders. Follow-up at 5 years after original diagnosis of AD revealed that 71 per cent of adults were completely well, 8 per cent had an intervening problem, and 21 per cent had developed a major depressive disorder or alcoholism. However, in adolescents at 5-year follow-up, only 44 per cent were without a psychiatric diagnosis, 13 per cent had an intervening psychiatric illness, and 43 per cent went on to develop major psychiatric morbidity (e.g. schizophrenia, schizoaffective disorders, major depression, bipolar disorder, substance abuse, and personality disorders). In contrast with the adults, the chronicity of the illness and the presence of behavioural symptoms in the adolescents were the strongest predictors for major psychopathology 5 years after the initial AD diagnosis. The number and type of symptoms were less useful than the length of treatment and chronicity of symptoms as predictors of future outcome.

As Chess and Thomas[57] have reported, it is important to note that AD with disturbance of conduct, regardless of age, has a more guarded outcome. In agreement with the findings of Andreasen and Wasek,[35] Chess and Thomas[57] emphasize that:

> a significant number [of AD adolescents] did not improve or even grew worse in adolescence and early adult life. It was not possible to predict the developmental course of the disorder in the early period after its first identification. Hence, we would suggest active appropriate

therapeutic intervention in all cases but especially adolescents [and adequate follow-up].

Spalletta *et al.*[25] report that suicidal behaviour and deliberate self-harm are important predictors in the diagnosis of AD. As mentioned before, these are obviously not subthreshold symptoms; they can lead to the most dire consequence—death. This outcome, when reached, can be neither corrected nor resolved. These behaviours mandate immediate and protective interventions. The diagnosis of AD may suggest that the patient has minor symptomatology. Such erroneous assessment may be life-threatening. There needs to be a definite split from viewing a diagnosis as subthreshold, and therefore assuming the attendant symptoms to be subthreshold. It is similar to labelling a patient with hypochondriasis, which in some settings can influence a more casual physical assessment, when such a patient could have serious physical morbidity concomitant with their hypochondriacal Axis I pathology.

Although most studies do point to a more benign prognosis for the AD, it is important to realize that the risk of serious morbidity and mortality still exists. Several recent studies investigating the association between suicide and AD underscore the importance of monitoring patients closely for suicidality, especially in younger populations. Portzky *et al.* conducted psychological autopsies on adolescents with AD who had committed suicide and found that suicidal thinking in these patients was brief and evolved rapidly without warning.[60] A slightly different profile was found in two other studies that looked at suicide attempters with a diagnosis of AD. These patients were more likely to have poor overall psychosocial functioning, prior psychiatric treatment, personality disorders, substance abuse histories and a current 'mixed' symptom profile of depressed mood and behavioural disturbances.[61,62] A study of the neurochemical variables of AD patients of all ages who had attempted suicide revealed biologic correlates consistent with the more major psychiatric disorders. Attempters were found to have lower platelet MAO activity, higher MHPG activity, and higher cortisol levels than controls. Though these findings differ from the lower MHPG and cortisol levels found in patients with major depression and suicidality, they are similar to findings in other major stress-related conditions.[63]

As mentioned earlier, the diagnosis of an AD may be in the early phase of an evolving disorder that has not yet developed to the extent that full-blown symptoms are evident to reach threshold for a major psychiatric disorder. If a patient continues to worsen, becomes more symptomatic, and does not respond to treatment, it is critical to review the diagnosis for the presence of a major disorder.

A recent report by Jones *et al.* (2002) looked at 10 years of readmission data for various psychiatric diagnoses including the AD. They found that admission diagnosis was a significant predictor of readmission and that AD had the lowest readmission rates.[58] Furthermore, initial psychological recovery from an AD may in large part be attributable to removal of the stressor. This was found to be the case in prisoners who developed AD after being placed in solitary confinement and whose symptoms resolved shortly after their release.[59]

The domains of diagnostic rigour and clinical utility seem at odds for AD. Studies that employ reliable and valid instruments (e.g. depression or anxiety rating scales, stress assessments, length of disability, treatment outcome, family patterns, etc.) would enhance more exact specification of the parameters of the AD diagnosis. Identification of the time course, remission or evolution to another diagnosis, and the evaluation of stressors (characteristics, duration, and the nature of adaptation to stress) would enhance the understanding of the aetiology, mechanisms, and mediators of a stress-response illness.

Studies with adequate symptom checklists rated independently from the establishment of the diagnosis would clarify the threshold between major and minor depression and anxiety, as well as guide an entry threshold to employ the AD diagnosis. Although the upper threshold is established by the specified criteria for major syndromes, the entry threshold between an AD and problem-level diagnoses and normality is undesignated with operational criteria. The careful examination of associated demographic and treatment outcome variables would also enable clinicians to describe more specifically the boundaries among subthreshold diagnoses, problem-level diagnosis, and normal behaviour. Associated features such as family history, biological correlates, treatment response, long-term course, and so forth, are all critical to establishing the validity of a diagnosis. The theory and practice of medicine have demonstrated the need for a comprehensive multidimensional formulation of all these physiological and functional variables to describe an illness and develop the most appropriate working diagnosis.

Subthreshold syndromes can encompass significant psychopathology that must not only be identified but treated (e.g. suicidal ideation/behaviour). Cross-sectionally, AD may appear to be the incipient phase of an emerging major syndrome. Consequently, AD, despite its questionable reliability and validity, serves an important diagnostic function in the practice of psychiatry. Problem- and subthreshold-level diagnoses are critical to the function of any medical discipline. Because this may be the initial phase, or a mild form, of a dysfunction that is not yet fully developed, there is a need to describe the relationship of this incipient state to other potential diagnoses. This lack of specificity and questionable reliability and validity are the hallmark of interface disorders and subthreshold phenomena, whether they are in diabetes mellitus, hypertension, or depression.

As mentioned earlier, the characteristics of a mental disorder vary over the life cycle, and this is clearly illustrated by the AD. Certain developmental epochs may be associated with a particular symptom profile, as seen with acute myocardial infarction or appendicitis. The effect of the stressor may vary, and the assessment of functioning must be 'measured' according to the demands of the developmental stage (e.g. school [adolescents], work [adults], self-care and maintenance [elderly]). The symptom characteristics and functional assessment of other diagnoses may also vary along the developmental schema from birth to senescence; illnesses such as major depressive disorders, organic mental disorders, sexual dysfunctions, and eating disorders need to be 'recast' in another hierarchy to incorporate the stage of the life cycle extant at the time of the assessment, and symptom profiles adjusted accordingly. The normal variations across developmental epochs would make AD and the other psychiatric disorders more reliable and valid across the life cycle. Similarly, there needs to be a consideration of a possible concomitant state of medical illness. The result would be a taxonomy tempered by the vicissitudes of development and medical illness.

A taxonomy which considers the development epoch and the presence of medical illness would be more useful to child

psychiatrists, paediatricians, geriatricians, geriatric psychiatrists, and primary care specialists, who are often convinced that a patient does not conform with today's psychiatry lexicon. A significant number of their patients remain at the problem level of diagnoses with their somatic complaints as well. It is not uncommon for a fever of unknown origin to not be diagnosed, or for a chest pain to remain unspecified. It is the art of medicine that makes it a profession, and a most difficult one, at the interface of medicine and psychiatry, or at the interface of normality and pathology. Freud[64] has emphasized the difficulty of understanding normality and pathology in her assessments of childhood. This important advice would obtain across the life cycle and be an important challenge to the developers of the subthreshold diagnoses (e.g. AD) and the future evolution of the DSM.

Further information

Grassi, L., Mangelli, L., Fava, G.A., *et al.* (2007). Psychosomatic characterization of adjustment disorders in the medical setting; some suggestions for DSM-V. *Journal of Affective Disorders*, **101**, 251–4.

Kumana, H., Ida, I., Oshima, A., *et al.* (2007). Brain metabolic changes associated with predisposition to onset of major depressive disorder and adjustment disorder in cancer patients—a preliminary PET study. *Journal of Psychiatric Research*, **41**, 591–9.

Casey, P., Maracy, M., Kelly, B.D., *et al.* (2006). Can adjustment disorder and depressive episode be distinguished? Results from ODIN. *Journal of Affective Disorders*, **92**, 291–7.

Bruinvels, D.J., Rebergen, D.S., Nieuwenhuijsen, K., *et al.* (2007). *Return to work interventions for adjustment disorders. Cochrane database of systematic reviews: protocols*, Issue 1. John Wiley & Sons, Chichester, UK.

References

1. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
2. Spiegel, D. (1994). *DSM-IV options book*. American Psychiatric Association, Washington, DC.
3. American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd edn). American Psychiatric Association, Washington, DC.
4. Kovacs, M., Ho, V., and Pollock, M.H. (1995). Criterion and predictive validity of the diagnosis of adjustment disorder: a prospective study of youths with new-onset insulin-dependent diabetes mellitus. *The American Journal of Psychiatry*, **152**, 523–8.
5. Perez-Jimenez, J.P., Gomez-Bajo, G.J., Lopez-Catillo, J.J., *et al.* (1994). Psychiatric consultation and post-traumatic stress disorder in burned patients. *Burns*, **20**, 532–6.
6. Sullivan, M.J., Winshenker, B., and Mikail, S. (1995). Screening for major depression in the early stages of multiple sclerosis. *The Canadian Journal of Neurological Science*, **22**, 228–31.
7. Shima, S., Kitagawa, Y., Kitamura, T., *et al.* (1994). Poststroke depression. *General Hospital Psychiatry*, **16**, 286–9.
8. Mezzich, J.E., Dow, J.T., Rich, C.L., *et al.* (1981). Developing an efficient clinical information system for a comprehensive psychiatric institute. II. Initial evaluation form. *Behavioral Research Methods and Instrumentation*, **13**, 464–78.
9. Bonelli, R.M. and Bugram, R. (2000). Additional A-criterion for adjustment disorders. *Canadian Journal of Psychiatry*, **45**, 763.
10. Casey, P. (2001). Adult adjustment disorder: a review of its current diagnostic status. *Journal of Psychiatric Practice*, **7**, 32–40.
11. Aoki, T., Hosaka, T., and Ishida, A. (1995). Psychiatric evaluation of physical rehabilitation patients. *General Hospital Psychiatry*, **17**, 440–3.
12. Zung, W. (1971). A rating instrument for anxiety disorders. *Psychosomatics*, **12**, 371–9.
13. Zung, W. (1965). A self-rating depression scale. *Archives of General Psychiatry*, **12**, 63–70.
14. McNair, D.M., Lorr, M., and Doppelman, L.F. (eds.) (1971). *Manual for the profile of mood states*. Educational and Industrial Testing Service, San Diego, CA.
15. Strain, J.J., Newcorn, J., Mezzich, J., *et al.* (1998). Adjustment disorder: the McArthur reanalysis. In *DSM-IV source book*, Vol. 4, pp. 404–24. American Psychiatric Association, Washington, DC.
16. Powell, S. and McCone, D. (2004). Treatment of adjustment disorder with anxiety: a September 11, 2001, case study with a 1-year follow up. *Cognitive and Behavioral Practice*, **11**, 331–6.
17. American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th edn, text revision). American Psychiatric Association, Washington, DC.
18. Havenaar, J.M., Van den Brink, W., Van den Bout, J., *et al.* (1996). Mental health problems in the Gomel region (Belarus): an analysis of risk factors in an area affected by the Chernobyl disaster. *Psychological Medicine*, **26**, 845–55.
19. Oxman, T.E., Barrett, J.E., Freeman, D.H., *et al.* (1994). Frequency and correlates of adjustment disorder related to cardiac surgery in older patients. *Psychosomatics*, **35**, 557–68.
20. Leigh, H. (1993). Physical factors affecting psychiatric condition: a proposal for DSM-IV. *General Hospital Psychiatry*, **15**, 155–9.
21. Despland, J.N., Monod, L., and Ferrero, F. (1995). Clinical relevance of adjustment disorder in DSM-III-R and DSM-IV. *Comprehensive Psychiatry*, **36**, 456–60.
22. De Leo, D., Pellegrini, C., and Serraiotto, L. (1986). Adjustment disorders and suicidality. *Psychology Reports*, **59**, 355–8.
23. De Leo, D., Pellegrini, C., Serraiotto, L., *et al.* (1986). Assessment of severity of suicide attempts: a trial with the dexamethasone suppression test and two rating scales. *Psychopathology*, **19**, 186–91.
24. Isometsa, E., Heikkinen, M., Henriksson, M., *et al.* (1996). Suicide in non-major depressions. *Journal of Affective Disorders*, **36**, 117–27.
25. Spalletta, G., Troisi, A., Saracco, H., *et al.* (1996). Symptom profile: axis II comorbidity and suicidal behaviour in young males with DSM-III-R depressive illnesses. *Journal of Affective Disorders*, **39**, 141–8.
26. Vlachos, I.O., Bouras, N., Watson, J.P., *et al.* (1994). Deliberate self-harm referrals. *European Journal of Psychiatry*, **8**, 25–8.
27. Zinbarg, R.E., Barlow, D.H., Liebowitz, M., *et al.* (1994). The DSM-IV field trial for mixed anxiety-depression. *The American Journal of Psychiatry*, **151**, 1153–62.
28. Judd, L.L., Rapaport, M.H., Paulus, M.P., *et al.* (1994). Subsyndromal symptomatic depression: a new mood disorder? *The Journal of Clinical Psychiatry*, **55**(4 Suppl), 18–28.
29. Casey, P., Maracy, M., Kelly, B.D., *et al.* (2006). Can adjustment disorder and depressive episode be distinguished? *Journal of Affective Disorders*, **92**, 291–7.
30. Maercker, A., Einsle, F., and Kollner, V. (2006). Adjustment disorders as stress response syndromes: a new diagnostic concept and its exploration in a medical sample. *Psychopathology*, **627**.
31. Strain, J.J. (1981). Diagnostic considerations in the medical setting. *The Psychiatric Clinics of North America*, **4**, 287–300.
32. Mok, H. and Walter, C. (1995). Brief psychiatric hospitalization: preliminary experience with an urban sort-stay unit. *Canadian Journal of Psychiatry*, **40**, 415–17.
33. Zarb, J. (1996). Correlates of depression in cognitively impaired hospitalized elderly referred for neuropsychological assessment. *Journal of Clinical and Experimental Neuropsychology*, **18**, 713–23.

34. Grassi, L., Gritti, P., Rigatelli, M., *et al.* (2000). Psychosocial problems secondary to cancer: an Italian multicenter survey of consultation-liaison psychiatry in oncology. Italian Consultation-Liaison Group. *European Journal of Cancer*, **36**, 579–85.
35. Andreasen, N.C. and Wasek, P. (1980). Adjustment disorders in adolescents and adults. *Archives of General Psychiatry*, **37**, 1166–70.
36. Fabrega, H. Jr, Mezzich, J.E., and Mezzich, A.C. (1987). Adjustment disorder as a marginal or transitional illness category in DSM-III. *Archives of General Psychiatry*, **44**, 567–72.
37. Popkin, M.K., Callies, A.L., Colón, E.A., *et al.* (1990). Adjustment disorders in medically ill patients referred for consultation in a university hospital. *Psychosomatics*, **31**, 410–14.
38. Foster, P. and Oxman, T. (1994). A descriptive study of adjustment disorder diagnoses in general hospital patients. *Irish Journal of Psychological Medicine*, **11**, 153–7.
39. Snyder, S. and Strain, J.J. (1989). Differentiation of major depression and adjustment disorder with depressed mood in the medical setting. *General Hospital Psychiatry*, **12**, 159–65.
40. Faulstich, M.E., Moore, J.R., Carey, M.P., *et al.* (1986). Prevalence of DSM-III conduct and adjustment disorders for adolescent psychiatric inpatients. *Adolescence*, **21**, 333–7.
41. Greenberg, W.M., Rosenfeld, D.N., and Ortega, E.A. (1995). Adjustment disorder as an admission diagnosis. *The American Journal of Psychiatry*, **152**, 459–61.
42. Breslow, R.E., Klinger, B.I., and Erickson, B.J. (1996). Acute intoxication and substance abuse among patients presenting to a psychiatric emergency service. *General Hospital Psychiatry*, **18**, 183–91.
43. Cohen, F. (1981). Stress and bodily illness. *The Psychiatric Clinics of North America*, **4**, 269–86.
44. Gonzales-Jaimes, E.I. and Turnbull-Plaza, B. (2003). Selection of psychotherapeutic treatment for adjustment disorder with depressive mood due to acute myocardial infarction. *Archives of Medical Research*, **34**, 298–304.
45. van der Klink, J.J., Blonk, R.W., Schene, A.H., *et al.* (2003). Reducing long term sickness absence by an activating intervention in adjustment disorders: a cluster randomized controlled design. *Occupational and Environmental Medicine*, **60**, 429–37.
46. van der Klink, J.J.L. and van Dijk, F.J.H. (2003). Dutch practice guidelines for managing adjustment disorders in occupational and primary health care. *Scandinavian Journal of Work, Environment & Health*, **29**, 478–87.
47. Maina, G., Forner, F., and Bogetto, F. (2005). Randomized controlled trial comparing brief dynamic and supportive therapy with waiting list condition in minor depressive disorders. *Psychotherapy and Psychosomatics*, **74**, 43–50.
48. Mihelich, M.L. (2000). Eye movement desensitization and reprocessing treatment of adjustment disorder. *Dissertation Abstracts International*, **61**, 1091.
49. Hameed, U., Schwartz, T.L., Malhotra, K., *et al.* (2005). Antidepressant treatment in the primary care office: outcomes for adjustment disorder versus major depression. *Annals of Clinical Psychiatry*, **17**, 77–81.
50. Nguyen, N., Fakra, E., Pradel, V., *et al.* (2006). Efficacy of etifoxine compared to lorazepam monotherapy in the treatment of patients with adjustment disorders with anxiety: a double-blind controlled study in general practice. *Human Psychopharmacology*, **21**, 139–49.
51. Newcorn, J.H., Strain, J.J., and Mezzich, J.E. (2000). Adjustment disorders. In *Comprehensive Textbook of Psychiatry* (7th edn) (eds. B. Saddock and V. Saddock). Lippincott Williams and Wilkins, Philadelphia, PA.
52. Hosaka, T., Aoki, T., and Ichikawa, Y. (1994). Emotional states of patients with hematological malignancies: preliminary study. *Japanese Journal of Clinical Oncology*, **24**, 186–90.
53. Oxman, T.E., Barrett, J.E., Freeman, D.H., *et al.* (1994). Frequency and correlates of adjustment disorder relates to cardiac surgery in older patients. *Psychosomatics*, **35**, 557–68.
54. Hugo, F.J., Halland, A.M., Spangenberg, J.J., *et al.* (1996). DSM-III-R classification of psychiatric symptoms in systemic lupus erythematosus. *Psychosomatics*, **37**, 262–9.
55. Uhlenhuth, E.H., Balter, M.B., Ban, T.A., *et al.* (1995). International study of expert judgment on therapeutic use of benzodiazepines and other psychotherapeutic medications. III. Clinical features affecting experts' therapeutic recommendations in anxiety disorders. *Psychopharmacology Bulletin*, **31**, 289–96.
56. Andreasen, N.C. and Hoenk, P.R. (1982). The predictive value of adjustment disorders: a follow-up study. *The American Journal of Psychiatry*, **139**, 584–90.
57. Chess, S. and Thomas, A. (1984). *Origins and evolution of behavior disorders: from infancy to early adult life*. Brunner-Mazel, New York.
58. Jones, R., Yates, W.R., and Zhou, M.D. (2002). Readmission rates for adjustment disorders: comparison with other mood disorders. *Journal of Affective Disorders*, **71**, 199–203.
59. Andersen, H.S., Sestoft, D., Lillebaek, T., *et al.* (2000). A longitudinal study of prisoners on remand: psychiatric prevalence, incidence and psychopathology in solitary vs. non-solitary confinement. *Acta Psychiatrica Scandinavica*, **102**, 19–25.
60. Portzky, G., Audenaert, K., and van Heeringen, K. (2005). Adjustment disorder and the course of the suicidal process in adolescents. *Journal of Affective Disorders*, **87**, 265–70.
61. Pelkonen, M., Marttunen, M., *et al.* (2005). Suicidality in adjustment disorder, clinical characteristics of adolescent outpatients. *European Child & Adolescent Psychiatry*, **14**, 174–80.
62. Kryzhanovskaya, L. and Canterbury, R. (2001). Suicidal behavior in patients with adjustment disorders. *Crisis*, **22**, 125–31.
63. Tripodianakis, J., Markianos, M., Sarantidis, D., *et al.* (2000). Neurochemical variables in subjects with adjustment disorder after suicide attempts. *European Psychiatry*, **15**, 190–5.
64. Freud, A. (1968). *Normality and pathology: assessment of childhood*. International Universities Press, New York.

4.6.5 Bereavement

Beverley Raphael, Sally Wooding, and Julie Dunsmore

Bereavement is the complex set reactions that occurs with the death of a loved one: the emotions of grief with yearning, angry protest, and sadness; the cognitive processes of understanding and making meaning of the finality and nature of death; and the social, cultural, spiritual, and religious contexts of adaptation. Grief may also result from other losses such as health, home, country, and safe worlds. There have been investigations into potential neurobiological substrates, without, as yet consensus about the explanatory model.

In 'Mourning and Melancholia', Freud[(1)] described the psychological processes of mourning which involved the gradual relinquishment of bonds with the deceased, and how mourning differed from melancholia. Lindemann[(2)] described the 'Symptomatology and Management of Acute Grief' in his classic paper on his experiences assessing and treating the survivors of a nightclub fire. Engel[(3)] asked 'Is Grief a Disease?', and concluded in the negative.

Bowlby's work on attachment, separation, and loss[(4–6)] has been the most influential in informing research and clinical practice, with many studies of both adults and children utilizing such concepts. Early research focused chiefly on bereavement following the

death of a spouse, describing normal, high risk, and pathological patterns of grief.[7–9] There is also a number of excellent reviews of theory and research, including those of Stroebe's group.[10,11]

Phenomenology of 'normal grief'

Common phenomena of the grief experience of adults, identified through many research studies,[12–14] relate to similar domains influenced by developmental trajectories, through childhood and adolescence. Adult studies indicate consistent patterns: numbness, disbelief; yearning, angry protest, and 'searching' behaviours representing separation distress; and sadness with reviewing of memories of the lost relationship, with a range of associated emotions; progressive acceptance of the death and changed circumstances, sometimes referred to as resolution.

Bonanno[14] has shown that resilient trajectories, defined by low overall distress, are common. Other transient phenomena described by clinicians working with bereaved people[15] include: identificatory symptoms, reflecting the deceased's illness; a sense of the deceased's ongoing at presence, at times as though seeing the face, hearing the voice, or feeling the touch of the dead person. 'Yearning' is considered to be the most pathognomic of these grief phenomena, which usually settle over the first year, but may continue, triggered by anniversaries, or specific memories. Older people who have had a long relationship with a spouse may continue this relationship in their minds for the comfort of 'talking' with the person, and a need for the ongoing closeness.[16]

Recent research[17] has modelled sequential peaks of the reactive phenomena: disbelief, yearning, anger, and depression, which bereaved people more usually describe as sadness. Grief may be a precipitant of depression in those with pre-existing or bereavement-related vulnerabilities and the differentiation of normal and more pathological forms of grief from depression is important clinically.[15,18] Intense grief and the peaks of distress identified above do not usually continue beyond the first 4–6 months.[12,13] Continuing 'acute' grief beyond this time suggests the possibility of pathological response, as do other risk indicators, although some phenomena may continue intermittently for many years. Comparative studies have demonstrated that the intensity of adult grief is likely to be greatest for the death of a child, then spouse, or partner, then parent.[12,19]

Neurobiology of bereavement

Recent research has examined the neurobiology of grief through studies using functional magnetic resonance imaging of grief[20] and brain activity in womºen grieving the break-up of a romantic relationship.[21] Workers have attempted to develop a theoretical model based on a wide range of relevant data, encompassing a 'neurobiopsychosocial' framework for sadness and loss.[22] Stress hormones[23] and psychoimmune function is a further area of research. A comprehensive model integrating the relevant research findings is yet to be established.

Risk and protective factors influencing course and outcome

Pre-existing vulnerabilities that may influence the course and outcome have been reviewed alongside other risk factors.[24] These include personality vulnerabilities related to relationship styles such as avoidant and insecure attachments. Genetic factors do not appear to have been directly studied, but it is likely that the short allele of the serotonin gene promoter polymorphism of 5HTTLPR which influences response to adversity may contribute, through gene–environment interactions.[25] Prior loss and adverse experiences may add vulnerability, for instance multiple losses faced by indigenous peoples, with loss of culture, land, and loved ones, with multiple premature deaths and separations.[26] Separation anxiety in childhood, as well as pre-existing psychiatric disorder, family psychiatric disorder, and substance abuse may add to vulnerability. Successful negotiation of earlier losses, mature defence styles, and optimism may be protective.

The nature of the lost relationship has been identified in a number of studies as being a significant factor.[15,27] The special relationship between parent and child is associated with greater vulnerabilities, including increased risk of psychiatric hospitalization and even death by suicide. Patterns of distress differ by gender with stillbirth, neonatal deaths, and sudden infant death syndromes, perhaps suggesting different attachment patterns.[28] The death of an adolescent child is not infrequently by accident, suicide, or risk-taking with illicit drugs, bringing the extra complexities of adaptation for the grieving parents.

A great deal of research has explored the grief associated with the death of a partner or spouse both young and old. High levels of dependence and ambivalence have been shown to complicate grief and to be associated with more difficult bereavement,[15,27] and prolonged or complicated grief may be more likely.

Family members may have different relationships with the deceased, and thus varying patterns and trajectories of grief, which may cloud the recognition of children's and others needs.

Adults' loss of an older parent appears to be the least distressing, although there is still sadness plus the recognition of one's own mortality. Here, as at any age, intense fantasies of reunion with the deceased may indicate risk of suicide, especially for older widowed men.

Circumstances of the death may influence outcome. When dying is *prolonged*, as in the later stages of a terminal illness, the dying person may experience grief over his or her own life, and the loved ones who will be lost to him, alongside the anticipatory grieving of family members. While palliative care systems may provide bereavement programmes, families have complex dynamics, and may require family focused interventions.[29]

Sudden unexpected deaths bring an extra level of emotional shock,[30] especially if also untimely as with children's death. When violent death occurs, as with homicide or the mass violence of terrorism or war, those bereaved may experience a complex mixture of traumatic stress reactions and grief reactions, sometimes called traumatic grief.[31] The specific issues facing those bereaved by violent deaths of loved ones have been reported in a recent volume by Rynearson,[32] which deals with homicide, terrorism, and other violent deaths. The prolonged and difficult grief in such circumstances is highlighted by findings from September 11, Oklahoma, and Bali bombings.

When people are *missing*, believed dead, the uncertainty, other stressors including complex legal and evidentiary processes (e.g. Disaster Victim Identification requirements), may lead to alternating hope and dread. When there are no remains, it will be more difficult for those bereaved to accept the reality of the death.

Seeing and 'saying goodbye' to the dead person, have been shown to help those bereaved in disasters. If remains are much disfigured, as with burns, it is important that those bereaved are supported in their choice about this.

Social support, particularly the perceptions of the supportiveness of family and social network are likely to be protective and assist the bereaved psychologically,[33] while perceptions of unhelpfulness may be associated with more negative outcomes.[9] Cultural requirements for social support may differ, as may the delineation of the period of mourning, the roles of the bereaved, and associated spiritual and religious needs.[34]

Multiple other adversities may occur, either coincidentally or as a consequence of the circumstances and the loss of the person, for instance financial difficulties, loss of resources, changed status, loss of meaning and identity, or other profound stressors of illness, injury, or other bereavements. Such additional stressors may increase vulnerability.[15]

In terms of Prolonged Grief Disorder (previously known as complicated grief disorder, and initially traumatic grief) Prigerson *et al.*[35] have carried out extensive research to refine this syndrome. Bringing together the views of international researchers, they have developed consensus criteria for a distinct psychiatric disorder to be considered for inclusion in DSM-V. This definition requires that the reaction to loss encompasses one of three symptoms of separation distress (e.g. yearning) and a minimum of five from a total of nine other symptoms, experienced at least daily, to a distressing or disruptive level. These include: shock; emotional numbing; avoidance of the reality of the loss; difficulty accepting the loss; feelings of meaninglessness; difficulty moving on with life; bitterness over the loss; mistrust; and a diminished sense of self. Such symptoms would need to last at least 6 months, and be associated with significant levels of functional impairment. These findings fit well with earlier research identifying more chronic patterns of grief.[12,13]

Physical and mental health consequences of bereavement

A recent valuable review[36] outlines the evidence of increased *mortality* for bereaved spouses, particularly males which is most pronounced in the first 6 months, and includes a range of conditions such as heart disease, leading to death from 'a broken heart'. It is also more pronounced for those younger. Death of a child is associated with even greater mortality risk, particularly for mothers. Suicide is one of the heightened risks, especially for mothers and older males.

Physical health impairments are also found,[36] with a variety of physical symptoms as well as greater use of medical services, and medications. Further research is needed to clarify the nature of any increased rate of specific diseases. Changed health behaviours, the impact of loss of a health supporting partner, functional or social changes, or shared environments of risk may contribute.

With regard to *mental health* there may be an increased level of anxiety and depressive symptoms. There may be a heightened risk for some bereaved individuals for exacerbations of pre-existing conditions, or the precipitation of new illnesses, including Post-Traumatic Stress Disorder when there is a violent death.[32] Other anxiety disorders, major depression, substance use disorder, and bipolar disorder may be precipitated by bereavement. Complicated or prolonged grief disorder may also represent a psychiatric consequence.[35]

Assessment and management

Most bereaved people do not require counselling so assessment must be a basis for intervention. Assessment should be simple, synchronous with need, and do no harm, addressing the death and its circumstances, the relationship, and the bereaved's experience since, including social support.[31] A more structured format, potentially including grief measures,[37] can clarify the presence of PTSD, Depression or Prolonged Grief, or other health needs, including physical health changes or problems, establishing the basis for intervention.

Initial management of acutely bereaved persons requires empathic, compassionate support; and responding to any acute needs in ways that are protective of their mental health, recognizing the 'roller-coaster' of emotions that may occur, and facilitating natural resilience. Concepts of Psychological First Aid are also valuable in the immediate period after the death.[38] Dealing with concerns about the deceased's suffering, and support to view the dead person's remains should the bereaved choose to, are likely to be helpful.

Most evidence-based interventions focus on psychotherapeutic methods, ranging from preventive counselling of those with demonstrated heightened risk[39] to self-help guided interventions,[40] interpersonal psychotherapy modifications for traumatic grief,[41] integrated cognitive behaviour therapy models,[42] or psychodynamically informed models[43] and web-based treatments.[44] Counselling models[45] and psychoeducation have also focused on those bereaved through specific deaths such as those of infants.[46] Other models deal with grief work and tasks,[47] as well as specific treatment for morbidity of complicated grief[48] and depressive or anxiety disorders including the use of pharmacotherapy for such conditions when indicated.[49] Rynearson's[32] work with 'restorative retelling' following violent deaths emphasizes the narrative story which is central to much bereavement counselling and testimony.

A practical approach to assessment and counselling may be initiated with some gentle queries, such as 'Can you tell me a little about your loss?', 'What happened with 'John's' death?', 'Can you tell me about 'John' and your relationship?', 'What's been happening since?'

If there is: intense continuing distress; circumstances of death which are untimely or traumatic; a complex relationship with the deceased; disruption of family functioning which is impacting on the needs of children; inadequacies of social support; 'unresolved' earlier losses; multiple additional stressors, the bereaved may be at heightened risk of adverse outcomes. Preventive counselling, which facilitates grieving, is attuned to the bereaved person's readiness, vulnerabilities, and strengths, and helps them to tell the story of their experiences and the person they have lost, is likely to improve outcomes.[39]

If assessment indicates that psychiatric disorders have arisen, for instance depression, post-traumatic stress disorder, anxiety, or substance use disorders, these conditions should be managed appropriately alongside counselling for the bereavement should this be required. The use of antidepressants or other medication is not indicated for bereavement itself but may be appropriate for such complications.[49] Monitoring for suicide risk should accompany clinical management.

Prolonged or complicated grief may benefit from cognitive behaviour therapy interventions, as well as relevant rehabilitation. Those who are both traumatized and bereaved may need the traumatic stress issues, to be dealt with first, and facilitation of grieving following this.[43]

There is a need for more comprehensive research and evaluation of prevention, early intervention and treatment modalities, and their appropriate provision, to individuals, families, or groups.[50] Culturally appropriate models of support and intervention also need to be further developed. Many bereaved people present first within primary care settings, and to their general practitioners who will need the skills and knowledge to deal with their distress, assess their needs, counsel them as appropriate, and refer when necessary.

Much support also comes from community and non-government agencies including bereavement focused groups for specific losses or for grief generally, and from specialized services in public or private mental health sectors.

Telling the bereaved person how to grieve, that they should 'forget about the past', that 'time heals all', is usually perceived as unhelpful. Treating grief as a disease, for example antidepressant medication for normal sadness, is seen by many bereaved people as interfering with their capacity to grieve for their loved one.

Counselling bereaved people requires hopeful, compassionate psychotherapeutic intervention which recognizes the human suffering involved, validates the person's strengths, and respects their spiritual needs. Loss is a central issue for all of us, both our fears of it, and its reality. Counselling requires those involved to recognize their own sensitivities in this regard, and to assist the 'journey' of those affected in dealing with their loss. Most people grieve, remember with love those whom they have lost, and continue to love, and love anew.

Further information

Kellehear, A. (2007). *A social history of dying*. Cambridge University Press, Cambridge.

http://www.childhoodgrief.org.au/

References

1. Freud, S. (1915). Mourning and melancholia. In *Sigmund Freud: collected papers*, (ed. Strachey, J.) Vol. 4. Basic Books, New York.
2. Lindemann, E. (1944). Symptomatology and management of acute grief. *The American Journal of Psychiatry*, **101**, 141–8.
3. Engel, G.L. (1961). Is grief a disease? *Psychosomatic Medicine*, **23**, 18–22.
4. Bowlby, J. (1969). *Attachment and loss, Vol. 1. Attachment*. Hogarth, London.
5. Bowlby, J. (1973). *Attachment and loss. Vol. 2. Separation, anxiety and anger*. Hogarth, London.
6. Bowlby, J. (1980). *Attachment and loss. Vol. 3. Loss: sadness and depression*. Hogarth, London.
7. Parkes, C.M. (1996). *Bereavement: studies of grief in adult life* (3rd edn). Routledge, London.
8. Maddison, D.C. and Walker, W.L. (1967). Factors affecting the outcome of conjugal bereavement. *The British Journal of Psychiatry*, **113**, 1057–67.
9. Jacobs, S. (1993). *Pathologic grief: maladaptation to loss*. American Psychiatric Press, Washington, DC.
10. Stroebe, M.S., Hansson, R.O., Stroebe, W., *et al.* (eds.) (2001). *Handbook of bereavement research: consequences, coping, and care*. American Psychological Association, Washington, DC.
11. Stroebe, M.S., Hansson, R.O., Stroebe, W., *et al.* (eds.) (2007). *Handbook of bereavement research and practice: 21st century perspectives*. American Psychological Association Press, Washington, DC.
12. Middleton, W., Raphael, B., Burnett, P., *et al.* (1998). A longitudinal study comparing bereavement phenomena in recently bereaved spouses, adult children and parents. *The Australian and New Zealand Journal of Psychiatry*, **32**, 235–41.
13. Byrne, G.J. and Raphael, B. (1994). A longitudinal study of bereavement phenomena in recently widowed elderly men. *Psychological Medicine*, **24**, 411–21.
14. Bonanno, G.A., Wortman, C.B., Lehman, D.R., *et al.* (2002). Resilience to loss and chronic grief: a prospective study from preloss to 18-months postloss. *Journal of Personality and Social Psychology*, **83**, 1150–64.
15. Raphael, B. (1983). *The anatomy of bereavement: a handbook for the caring professions*. Basic Books, New York.
16. Moss, M.S., Moss, S.Z., and Hansson, R.O. (2001). Bereavement and old age. In *Handbook of bereavement research: consequences, coping, and care* (eds. M.S. Stroebe, R.O. Hansson, W. Stroebe, and H. Schut). American Psychological Association, Washington, DC.
17. Maciejewski, P.K., Zhang, B., Block, S.D., *et al.* (2007). An empirical examination of the stage theory of grief. *The Journal of the American Medical Association*, **297**, 716–23.
18. Prigerson, H.G., Bierhals, A.J., Kasl, S.V., *et al.* (1996). Complicated grief as a distinct disorder from bereavement-related depression and anxiety: a replication study. *The American Journal of Psychiatry*, **153**, 84–6.
19. Sanders, C.M. (1980). A comparison of adult bereavement in the death of a spouse, child, and parent. *Omega*, **10**, 303–22.
20. Gündel, H., O'Connor, M., Littrell, L., *et al.* (2003). Functional neuroanatomy of grief: an fMRI study. *The American Journal of Psychiatry*, **160**, 1946–53.
21. Najib, A., Lorberbaum, J., Kose, S., *et al.* (2004). Regional brain activity in women grieving a romantic relationship break up. *The American Journal of Psychiatry*, **161**, 2245–56.
22. Freed, P.J. and Mann, J.J. (2007). Sadness and loss: toward a neurobiopsychosocial model. *The American Journal of Psychiatry*, **164**, 28–34.
23. McCleery, J.M., Bhagwagar, Z., Smith, K.A., *et al.* (2000). Modelling a loss event: effect of imagined bereavement on the hypothalamic-pituitary-adrenal axis. *Psychological Medicine*, **30**, 219–23.
24. Stroebe, M., Folkman, S., Hansson, R.O., *et al.* (2006). The prediction of bereavement outcome: development of an integrative risk factor framework. *Social Science & Medicine*, **63**, 2446–51.
25. Kaufman, J., Yang, B.-Z., Douglas-Palumberi, H., *et al.* (2004). Social supports and serotonin transporter gene moderate depression in maltreated children. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 17316–21.
26. Raphael, B., Swan, P., and Martinek, N. (1998). Intergeneration aspects of trauma for Australian aboriginal people. In *An international handbook of multigenerational legacies of trauma* (ed. Y. Danieli), pp. 327–39. Plenum Press, New York.
27. Parkes, C.M. (2006). *Love and loss: the root of grief and its complications*. Routledge, Hove, UK.
28. Vance, J.C., Boyle, F.M., Najman, J.M., *et al.* (2002). Couple distress after sudden infant or perinatal death: a 30-month follow-up. *Journal of Paediatrics and Child Health*, **38**, 368–72.
29. Kissane, D.W., McKenzie, M., Bloch, S., *et al.* (2006). Family focused grief therapy: a randomized controlled trial in palliative care and bereavement. *The American Journal Psychiatry*, **163**, 1208–18.
30. Lundin, T. (1984). Morbidity following sudden and unexpected bereavement. *The British Journal of Psychiatry*, **144**, 84–8.
31. Raphael, B., Martinek, N., and Wooding, S. (2004). Assessing loss, psychological trauma and traumatic bereavement. In *Assessing psychological trauma and PTSD* (2nd edn). The Guilford Press, New York.

32. Rynearson, E.K. (ed.) (2006). *Violent death: resilience and intervention beyond the crisis*. Routledge Psychosocial Stress Series, Taylor & Francis, New York.
33. Vanderwerker, L.C. and Prigerson, H.G. (2004). Social support and technological connectedness as protective factors in bereavement. *Journal of Loss and Trauma*, **9**, 45–57.
34. Parkes, C.M., Laungani, P., and Young, B. (1997). *Death and bereavement across cultures*. Routledge, London.
35. Prigerson, H., Horowitz, M.J., Selby, C.J., *et al*. (manuscript submitted for publication). Field trial of consensus criteria for prolonged grief disorder proposed for DSM-V. *Archives of General Psychiatry*.
36. Stroebe, M., Schut, H., and Stroebe, W. (2007). The physical and mental health consequences of bereavement: a review. Seminar for *Lancet,* **8**, 1960–73.
37. Neimeyer, R.A. and Hogan, N.S. (2001). Quantitative or qualitative? Measurement issues in the study of grief. In *Handbook of bereavement research: consequences, coping, and care* (eds. M.S. Strobe, R.O. Hansson, W. Stroebe, and H. Schut), pp. 89–118. American Psychological Association, Washington, DC.
38. National Center for PTSD Psychological First Aid Manual http://www.ncptsd.va.gov/ncmain/ncdocs/manuals/nc_manual_psyfirstaid.html
39. Raphael, B. (1977). Preventive intervention with the recently bereaved. *Archives of General Psychiatry*, **34**, 1450–4.
40. Vachon, M.L.S., Lyall, W.A.L., Rogers, J., *et al*. (1980). A controlled study of self-help interventions for widows. *The American Journal of Psychiatry*, **137**, 1380–4.
41. Shear, K., Frank, E., Houck, P.R., *et al*. (2005). Treatment of complicated grief: a randomized controlled trial. *The Journal of the American Medical Association*, **293**, 2601–8.
42. Shear, K. and Frank, E. (2006). Treatment of complicated grief: integrating cognitive–behavioral methods with other treatment approaches. In *Cognitive behavioural therapies for trauma* (eds. V. Follette and J. Ruzek), pp. 290–320. Guilford Press, New York.
43. Raphael, B., Dunsmore, J., and Wooding, S. (2004). Early mental health interventions for traumatic loss in adults. In *Early intervention for trauma and traumatic loss* (ed. B. Litz), pp. 147–78. Guilford Press, New York.
44. Wagner, B., Knaevelsrud, C., and Maercker, A. (2006). Internet-based cognitive–behaviorial therapy (INTERAPY) for complicated grief: a randomized controlled trial. *Death Studies*, **30**, 429–53.
45. Rando, T.A. (1993). *The treatment of complicated mourning*. Research Press, Champaign, IL.
46. Murray, J.A., Terry, D.J., Vance, J.C., *et al*. (2000). Effects of a program of intervention on parental distress following infant death. *Death Studies*, **24**, 275–305.
47. Worden, J.W. (1991). *Grief counselling and grief therapy*. Springer Publishing Company, New York.
48. Boelen, P.A. (2005). *Complicated grief: assessment, theory, and treatment*. Ipskamp, Enscede/Amsterdam.
49. Raphael, B., Dobson, M., and Minkov, C. (2001). Psychotherapeutic and pharmacological intervention for bereaved people chapter in *Handbook of bereavement research: consequences, coping, and care* (eds. M.S. Stroebe, R.O. Hansson, W. Stroebe, and H. Schut). American Psychological Association, Washington, DC.
50. Schut, H., Stroebe, M.S., Van Den Bout, J., *et al*. (2001). The efficacy of bereavement interventions: determining who benefits. In *Handbook of bereavement research: consequences, coping, and care* (eds. M.S. Stroebe, R.O. Hansson, W. Stroebe, and H. Schut). American Psychological Association, Washington, DC.

4.7

Anxiety disorders

Contents

4.7.1 Generalized anxiety disorders

Stella Bitran, David H. Barlow, and David A. Spiegel

Anxious apprehension and overconcern are common to many anxiety and mood disorders. Prior to 1980 in the American DSM diagnostic system, and 1992 in the international ICD system, individuals who experienced those symptoms in the absence of a realistic focus of concern were classified as having an 'anxiety neurosis' (DSM-II) or 'anxiety state' (ICD-9). In DSM-III, panic disorder was split off from that classification, and the residual category was renamed generalized anxiety disorder (**GAD**). A similar nomenclature was adopted in ICD-10.

Since its inception, GAD as a nosological entity has been troubled by problems of poor reliability and high comorbidity.[1] Those concerns have prompted several revisions of the DSM criteria and also have raised more basic questions regarding the validity of GAD as a disorder distinct from other anxiety and mood states. The question of what is the nature of GAD is still being debated and it remains one of the least reliably diagnosed anxiety or mood disorders.[2] This diagnostic unreliability has led to various suggestions for revisions to the diagnostic criteria and criticisms of the current definition of GAD.

Clinical features

Individuals with GAD experience persistent anxiety and worry that is out of proportion to actual events or circumstances. Typically, the anxiety and worry involve minor or everyday matters, such as work, finances, relationships, the health or safety of loved ones, and routine tasks. Often, the focus of worry shifts from one concern to another. Although people with GAD do not always consider their worries to be unrealistic or excessive, they do find them difficult to control. Consequently, the worries often interfere with concentration and performance.

Associated with the anxiety and worry, individuals with GAD have a variety of cognitive and somatic symptoms, including trembling, feeling shaky, aching in the back and shoulders, tension headaches, chest tightness, restlessness, exaggerated startle, irritability, insomnia, fatigue, dry mouth, sweating, urinary frequency, trouble swallowing, nausea, and diarrhoea. In addition, GAD may be accompanied by other conditions typically associated with stress, such as irritable bowel syndrome or atypical chest pain.

Classification

Diagnosis

(a) DSM criteria

In DSM-III, GAD was essentially a residual category for individuals with somatic symptoms of anxiety who did not meet diagnostic criteria for another, more specific, anxiety disorder. Diagnosis required the presence, for at least a month, of symptoms from three of four symptom clusters: motor tension, autonomic hyperactivity, apprehensive expectation, and vigilance and scanning. Unfortunately, clinicians had difficulty applying those criteria. In addition, its diagnosis depended on the application of the criteria for other diagnoses since GAD was not diagnosed if another anxiety disorder was present.

In DSM-III-R, apprehensive expectation was removed from the diagnostic symptom clusters, was redefined as unrealistic or excessive anxiety and worry about two or more life circumstances, and was made the essential feature of GAD. In addition, the duration criterion was changed from 1 to 6 months, and the hierarchical exclusion rule was dropped, allowing GAD to be diagnosed in addition to other disorders.

Table 4.7.1.1 DSM-IV inclusion criteria for GAD

(a) Excessive anxiety and worry, occurring more days than not for at least 6 months, about a number of events or activities
(b) The person finds it difficult to control the worry
(c) The anxiety and worry are accompanied by at least three of the following six symptoms (one in children): restlessness or feeling keyed up or on edge; being easily fatigued; difficulty concentrating or mind going blank; irritability; muscle tension; sleep disturbance
(d) The anxiety, worry or physical symptoms cause significant distress or functional impairment

(American Psychiatric Association (2000), *Diagnostic and statistical manual of mental disorders* (4th edn, text revision). APA, Washington, DC)

Despite those changes, the diagnostic reliability of GAD remained essentially unchanged.[1] Investigations revealed that the new worry criterion was problematic. Interviewers commonly disagreed as to whether two distinct spheres of worry were present, whether the worry was unrealistic or excessive, or whether the focus of the worry could be construed to be part of the symptomatology of another disorder. Moreover, studies indicated that patients with GAD did not differ substantially from control subjects in the content of their worries.[3,4] The main difference between patients and controls was that the former experienced their worrying to be uncontrollable while the latter did not.

Based on those and other findings, the GAD criteria were revised again in DSM-IV. The 'unrealistic' descriptor and the requirement for anxiety or worry to involve at least two spheres of life circumstances were deleted, and a new criterion was added that the worry must be experienced as difficult to control. In addition, the associated symptom criterion was modified to require only three of six symptoms from the previous motor tension and vigilance and scanning clusters (Table 4.7.1.1). For additional information about the evolution of the DSM criteria for GAD, see Barlow or Wincze.[5]

(b) ICD-10 criteria

Like DSM-IV, ICD-10 requires a period of 6 months of generalized anxiety and worry accompanied by certain somatic symptoms (Table 4.7.1.2). The 6 months of 'prominent' tension and worry needs to be accompanied by at least 4 of 22 associated symptoms. The ICD-10 differs from DSM-IV in that it does not require that worry be 'uncontrollable', that the symptoms of GAD occur exclusively outside the context of a mood disorder, or that they meet a 'clinical significance' criterion.

(c) Differential diagnosis

Everyone experiences anxiety and worry sometimes, and some people describe themselves as born worriers. GAD differs from these non-pathological anxiety experiences in that it is both persistent and severe enough to cause significant distress or interference. Typically also, the worries are more pervasive and difficult to control than normal worries and are associated with physical symptoms of anxiety and tension.

A number of general medical conditions can present with signs and symptoms resembling GAD (Table 4.7.1.3). In addition, substances such as caffeine, alcohol, other drugs of abuse, toxins, and some medications (Table 4.7.1.4) can cause anxiety-like

Table 4.7.1.2 ICD-10 inclusion criteria for GAD

(a) At least 6 months of prominent tension, worry, and feelings of apprehension about everyday events and problems
(b) At least four of the following 22 symptoms must be present, at least one of which must be from the autonomic arousal cluster
Autonomic arousal symptoms: palpitations or pounding heart or accelerated heart rate; sweating; trembling or shaking; dry mouth (not due to medication or dehydration)
Symptoms involving the chest and abdomen: difficulty in breathing; feeling of choking; chest pain or discomfort; nausea or abdominal distress
Symptoms involving mental state: feeling dizzy, unsteady, faint or lightheaded; derealization or depersonalization; fear of losing control, 'going crazy', or passing out; fear of dying
General symptoms: hot flushes or cold chills; numbness or tingling sensations
Symptoms of tension: muscle tension or aches and pains; restlessness and inability to relax; feeling keyed up, or on edge, or mentally tense; a sensation of a lump in the throat, or difficulty in swallowing
Other non-specific symptoms: exaggerated response to minor surprises or being startled; difficulty in concentrating, or mind 'going blank' because of worrying or anxiety; persistent irritability; difficulty getting to sleep because of worrying

(World Health Organization (2004), *International statistical classification of diseases and health related problems* (2nd edn). WHO, Geneva, Switzerland)

Table 4.7.1.3 General medical conditions that can cause symptoms resembling anxiety

Cardiac conditions: arrhythmias, coronary insufficiency, mitral valve prolapse, heart failure
Endocrine conditions: hyperthyroidism, hypoparathyroidism, hypoglycaemia
Neurological conditions: temporal lobe epilepsy, vestibular nerve disease
Respiratory conditions: asthma, hypoxia, hyperventilation, obstructive lung disease, pulmonary embolism
Other conditions: porphyria, carcinoid tumour, systemic lupus erythemacosus, pellagra

Table 4.7.1.4 Medications that can cause symptoms resembling anxiety

Psychotropics: antidepressants, neuroleptics (akathisia), sedative hypnotics (withdrawal syndrome), disulfiram
Respiratory drugs: B-adrenergic stimulants, bronchodilators
Cardiovascular drugs: antiarrhythmics, antihypertensives
Neurological disorder medications: anticonvulsants, anticholinergic agents, L-dopa
Anaesthetic drugs: pre-anaesthetics, general anaesthetics (post-anaesthetic syndrome)
Other drugs: thyroid hormone, antibiotics, non-steroidal anti-inflammatory drugs, anticancer drugs

symptoms either as a direct effect or as part of a withdrawal syndrome. These causes may be established on the basis of a medical and substance use history, physical examination, or laboratory tests.

GAD is distinguished from other psychiatric disorders, in part, by the focus of the anxiety and worry, which is not limited to a feature of another disorder. For example, the worry is not only about the possible occurrence or implications of panic attacks (as in panic disorder), or about negative evaluations by others (social phobia), gaining weight (anorexia nervosa), or having a serious illness (hypochondriasis). In obsessive–compulsive disorder, the anxiety and worry are associated with intrusive thoughts, images, or impulses that are distressing.

Generalized anxiety commonly occurs in depression, and GAD and depression also share associated symptoms such as sleep disturbance, fatigue, restlessness, and poor concentration. When the associated symptoms could fit with either disorder, the distinction is made on the basis of the presence and time course of depressed mood relative to anxiety. In DSM-IV, GAD is not diagnosed if its features occur exclusively during a mood disorder.

(d) Epidemiology

Prevalence estimates of GAD vary considerably with the diagnostic criteria used. One large-scale study found a 12-month prevalence rate of 2.07 per cent and a lifetime prevalence rate to be 4.1 per cent.[6] Socio-demographic factors associated with increased risk for GAD included being female, middle-aged, and with low income. However, being African American, Asian, or Hispanic was associated with a decreased risk.[7]

The National Comorbidity Study Replication (NCS-R), which used DSM-IV criteria and included structured interviews of over 9000 individuals in the United States, found a 12-month prevalence of GAD of 3.1 per cent and a lifetime prevalence rate at 5.7 per cent.[8,9] Lifetime prevalence rates were lowest among 18- to 29-year-olds (4.1 per cent) and those 60 or older (3.65 per cent), with the highest rates found among 45- to 59-year-olds (7.7 per cent).

(e) Comorbidity

GAD usually coexists with other anxiety and mood disorders. One large-scale study found that 68 per cent of individuals with a principal diagnosis of GAD met criteria for another Axis I disorder (Table 4.7.1.5).[10] The most frequently comorbid disorders were MDD, social phobia, or panic disorder with or without agoraphobia. Ninety-two per cent of individuals from this study with a principal diagnosis of GAD met criteria for another lifetime disorder, with 64 per cent meeting criteria for MDD. Similarly, in the major epidemiological surveys, nearly two-thirds of individuals with GAD had additional DSM Axis I diagnoses.[11] Most common among these were specific (21–59 per cent) and social (16–59 per cent) phobias, followed by panic disorder (3–27 per cent) and depression (8–39 per cent). In addition, GAD was found to be approximately twice as common among women as men. There is less information about the prevalence of personality disorders among patients with GAD.

Table 4.7.1.5 Prevalence of comorbid disorders in 279 patients with GAD

Any additional lifetime disorder	96%
Any anxiety/mood	94%
Any anxiety disorder	85%
Any mood disorder	74%
Anxiety disorders	
Panic disorder with/or without agoraphobia	47%
Social phobia	46%
Specific phobia	22%
Mood disorders	
Major depressive disorder	67%
Dysthymia	11%

(f) Is GAD a valid disorder?

The findings of only fair diagnostic reliability and high comorbidity for GAD have been interpreted as indicating poor discriminant validity of the disorder, suggesting that differentiating GAD from other anxiety and mood disorders may be artifactual. In considering those arguments, it is important to distinguish the diagnostic criteria sets specified in the DSM and ICD classification systems from the clinical syndromes they are intended to identify. Low discriminant validity for a disorder may be due to problems with the former rather than the latter. To establish the construct validity of a syndrome, one must demonstrate that it has a consistent set of features, the pattern of which separates it from other related syndromes. One approach to doing that is to compare the profiles of patients with different diagnoses across various illness dimensions.

In one such study, data from patients who took part in the DSM-IV mixed anxiety-depression field trial were examined.[12] Using factor analyses of patients' scores on 73 items from the Hamilton anxiety and depression rating scales, four clusters were identified that corresponded to the dimensions of anxiety, depression, physiological arousal, and general negative affect (containing items that loaded on both the anxiety and depression factors). Patients with a principal diagnosis of GAD had a unique profile (high on negative affect and anxiety, low on physiological arousal and depression) that distinguished them from individuals with panic disorder, major depression, anxiety or depressive disorder not otherwise specified, or no mental disorder.

A subsequent study, using an anxiety clinic sample and an expanded array of measures, yielded similar results.[13] In this case, five primary factors (corresponding to panic, agoraphobia, social anxiety, obsessions/compulsions, and general anxiety) and a higher order factor (negative affect) were identified. Again, patients with GAD had a unique factor profile.

The findings from the preceding studies were replicated and extended in an independent sample of patients.[14] As in the earlier studies, GAD was found to be distinct from other anxiety syndromes and depression, although it had the highest degree of overlap with other syndromes, especially depression. In addition, GAD was strongly associated with the non-specific dimension of negative affect, which is common to anxiety and depression. The authors suggest that GAD may represent a 'basic emotional disorder', because it consists of features that are present to varying degrees in all anxiety and mood disorders.

Finally, all three of the preceding studies (and a variety of others, e.g. Barlow *et al.*[15] support the differentiation of symptoms of autonomic arousal, which are characteristic of panic attacks, from somatic symptoms related to central nervous system tension, which form the DSM-IV-associated symptom cluster of GAD.

(g) Aetiology

Findings from genetics, neurobiology, and psychology infer a multifactorial aetiology for GAD, which has been organized into a triple vulnerabilities model.[16,17] This model suggests that anxiety disorders result from the combination of a generalized biological vulnerability, a general psychological vulnerability, and a specific psychological vulnerability.

(i) Generalized biological vulnerability

Genetic contributions. Several studies investigating genetic vulnerabilities for mental disorders have supported the notion that a shared vulnerability underlies anxiety disorders.[18] It was shown through a meta-analysis of genetic epidemiological studies that many anxiety disorders (including GAD, panic disorder, phobias, and OCD) aggregate in families and that genetics has the most influence when examining familial risk.[19] In a family study that used DSM-III criteria, GAD (but not other anxiety disorders) was five times more prevalent (19.5 per cent versus 3.5 per cent) among first-degree relatives of patients with GAD than among relatives of controls.[20] However, two twin studies using the same criteria found concordance rates for GAD were no higher among monozygotic than dizygotic twins.[21] Two subsequent studies that used DSM-III-R criteria found a shared heritability for GAD and mood disorders.[22] At present, it appears that genetic factors play a modest role in the aetiology of GAD, and one that is more closely related to vulnerability for depression than for other anxiety disorders.

Neurobiological mechanisms. A variety of neuroanatomical, neurochemical, neuroendocrine, and neurophysiological systems have been implicated in the pathogenesis of anxiety states. Much of this information has come from animal models and research on the effects of stress. Studies of neurobiological functioning in humans with GAD are limited. Some of the physical systems that may be involved in the emotion of anxiety are summarized below. Additional information may be found in reviews by Davidson,[23] and Gray and McNaughton.[24]

The noradrenergic nervous system. Noradrenergic pathway (the **locus coeruleus-noradrenaline-sympathetic nervous system**) have long been associated with fear and arousal and play an important role in the body's response to threat. However, their role in persistent anxiety states is not clear. Resting catecholamine levels in patients with GAD appear to be normal. On the other hand, GAD patients exhibit subnormal responses to both stimulation[25] and blockade[26] of α_2-adrenergic receptors and a reduced density of α_2-receptors in platelets.[27] Those findings could reflect downregulation of the α_2-receptors due to initially high levels of noradrenaline (norepinephrine).

Consistent with those neurochemical findings, somatic measures of autonomic nervous system function (e.g. skin conductance, respiratory rate, heart-rate variability, blood pressure) in patients with GAD tend to show normal resting values with blunted and sometimes prolonged responses to stressful stimuli.[27] Psychophysiological studies have found that worry is associated with restricted sympathetic arousal and low vagal tone.[27] In contrast, it has been shown that compared to controls, individuals with GAD show greater muscle tension at baseline in response to psychological challenge.[28] In addition, structural analyses suggest that GAD, unlike other anxiety disorders, is not associated with autonomic hyperarousal when levels of negative affect are held constant.[14] These findings indicate diminished autonomic nervous system responsiveness in individuals with GAD.

The hypothalamic-pituitary-adrenal axis. The **hypothalamic-pituitary-adrenal axis** and its end-product, cortisol, are also involved in reactions to stress. Activity in the hypothalamic-pituitary-adrenal axis is subject to a variety of influences. Primary control is by means of hypothalamic secretion of corticotrophin-releasing factor, which stimulates pituitary secretion of ACTH, which in turn stimulates adrenal secretion of cortisol. Circulating cortisol, and analogues such as dexamethasone, exert inhibitory feedback at the level of the pituitary gland and apparently also by means of receptors on the hippocampus.

In rats, chronic exposure to stress or exogenous steroids results in a reduction of corticosteroid receptors in the hippocampus and a consequent decrease in feedback inhibition by cortisol.[29] These animals exhibit reduced dexamethasone suppression of cortisol secretion and greater or more prolonged adrenocortical responses to stress. Reduced dexamethasone suppression also has been observed in approximately one-third of patients with DSM-III-diagnosed GAD.[30] This reduction in the normal regulatory control of cortisol secretion may be one mechanism through which chronic or repeated stress can lead to persistent anxiety.

The amygdala and the bed nucleus of the stria terminalis. LeDoux[31] and others have demonstrated the central role played by the **amygdala** in the mediation of fear reactions. The amygdala is thought to be responsible for the detection of potential threats to the organism and the mobilization of a range of defensive responses (Fig. 4.7.1.1). Through connections with the hypothalamus, it can activate the sympathetic nervous system and hypothalamic-pituitary-adrenal axis. Through efferent fibres to the central grey area of the midbrain, it can mediate behavioural defence responses such as the fight-or-flight response and behavioural 'freezing'. Through connections to the nucleus reticularis pontis caudalis, it can enhance the defensive startle reflex.

A structural magnetic resonance imaging (MRI) study of children and adolescents found that those individuals with GAD had increased total and right amygdala volume compared to non-anxious controls.[32] In addition, abnormalities in fear circuitry, especially hyperactivation in the right amygdala, have been found in adolescents with GAD.[33]

The extent to which these pathways are involved in the neurobiology of anxiety (as opposed to fear) is unclear. However, a structure

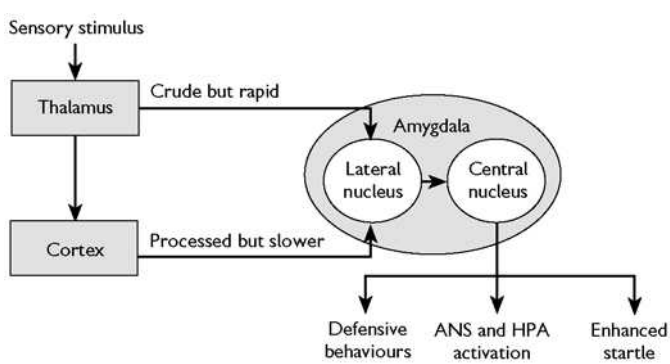


Fig 4.7.1.1 Fear pathways (based on descriptions by LeDoux[31] and Davis[34]). ANS, autonomic nervous system; HPA, hypothalamic-pituitary-adrenal axis.

closely related to the amygdala, the bed nucleus of the stria terminalis, may be involved in this emotion. The bed nucleus resembles the amygdala in its neurotransmitter content, cell morphology, and hypothalamic and brainstem connections and, like the amygdala, it exerts a modulating effect on the startle reflex.[34] Studies of this latter effect implicate it in the experience of anxiety.

Administration of corticotrophin-releasing factor into the cerebral ventricles of rats produces a state of generalized arousal resembling anxiety. Under those conditions, the startle reflex also is enhanced. Exposing rats to bright light for 5 to 20 min has similar effects. These effects are not blocked by damage to amygdala but are by lesions to the bed nucleus of the stria terminalis and by treatment with benzodiazepines or buspirone. Conversely, infusion of corticotrophin-releasing factor directly into the bed nucleus of the stria terminalis, but not the amygdala, produces a rapid increase in startle. Based on these observations, Davis[34] has suggested that the stria terminalis may play a role in anxiety analogous to that of the amygdala in fear reactions and, further, that prolonged or repeated stimulation of the stria terminalis by corticotrophin-releasing factor during periods of stress might lead to sustained activation and thus to persistent anxiety. A recent study has confirmed the differential association of these structures with fear and anxiety, respectively.[35]

The septohippocampal system (behavioural inhibition system). The bed nucleus of the stria terminalis is part of the larger septohippocampal system.[36] In 1982, based on data from several lines of research, Gray hypothesized that the septohippocampal system, together with the Papez circuit (a neural loop connecting the subicular area in the hippocampal formation to the mammillary bodies, anterior thalamus, cingulate cortex, and back to the subiculum), is responsible for mediating the emotion of anxiety as well as the major effects of anxiolytic drugs.[36] Gray called this network the **behavioural inhibition system**, because he believed that, when activated, it interrupts ongoing behaviour and redirects the organism's attention to signs of possible danger.

According to Gray's model,[24,36] the behavioural inhibition system receives information about the environment from the sensory cortex via the temporal lobe and hippocampal formation. The system checks the information for consistency with predictions, which are updated continuously by the Papez circuit based on preceding information and stored patterns, as well as for consistency with the immediate goals of the organism. When a mismatch is found, or if a predicted event is aversive, the outputs of the behavioural inhibition system are activated, resulting in a constellation of emotional and behavioural effects consistent with anxiety (Fig. 4.7.1.2).

The activation of the behavioural inhibition system appears to be moderated by ascending noradrenergic and serotonergic projections to the septohippocampal complex, providing a possible mechanism for the anxiolytic actions of some drugs. The amygdala also provides inputs to the behavioural inhibition system and may relay its outputs to the hypothalamus and autonomic nervous system, thereby mediating anxious arousal. Sustained activation of the behavioural inhibition system might therefore account for many of the features of GAD.

The benzodiazepine-γ-aminobutyric acid system. The powerful anxiolytic and sedative effects of benzodiazepines are believed to be mediated by benzodiazepine recognition sites located on γ-aminobutyric acid (**GABA**) type A receptor complexes in the central nervous system. When bound to those complexes,

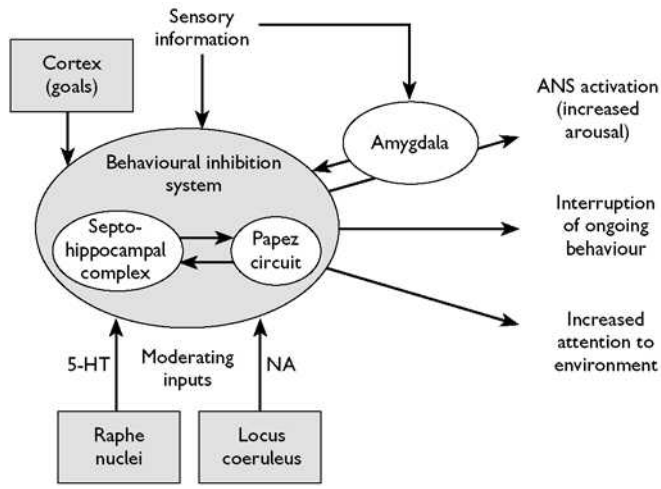


Fig 4.7.1.2 Behavioural inhibition system (based on descriptions by Gray and McNaughton[24]). ANS, autonomic nervous system; 5-HT, serotonin; NA, noradrenaline.

benzodiazepines allosterically modulate the GABA receptors to enhance the normal inhibitory effects of GABA on neurotransmission. Activation of **central benzodiazepine-GABA receptor complexes** also suppresses hypothalamic-pituitary-adrenal axis activity and, consequently, cortisol levels.

In addition to these central receptor complexes, benzodiazepine recognition sites of a different type are present widely in cells outside the central nervous system. These so-called **peripheral benzodiazepine receptors** are believed to be instrumental in controlling the synthesis of regulatory steroids. Their role in the anxiolytic actions of benzodiazepines is unknown; although they bind some drugs (e.g. diazepam), they have low affinity for others (e.g. clonazepam). Interestingly, peripheral benzodiazepine receptors are decreased in blood cells of individuals with untreated GAD but return to normal levels after successful treatment with benzodiazepines. Their numbers also vary in response to stress, being elevated following acute stressors and reduced during chronic stress.

A possible explanation for those changes has been suggested by Rocca *et al.*[37] The investigators note that peripheral benzodiazepine receptors in brain glial cells control the production of neurosteroids that act as modulators of $GABA_A$ receptor sensitivity. Their effect on GABA functioning appears to be opposite to that of clinically effective benzodiazepines, that is, they 'decrease' rather than increase the inhibitory effects of GABA. It is hypothesized that an endogenous ligand of these glial cell receptors (possibly diazepam-binding inhibitor) is released during stress, initiating the cycle of events depicted in Fig. 4.7.1.3.

The immediate effect of these events would be to enhance the stress-induced release of cortisol. However, prolonged cortisol excess is hypothesized to downregulate peripheral benzodiazepine receptors, resulting in the reduced receptor densities found in GAD. Administration of a clinically effective benzodiazepine drug would interrupt the proposed pathway at the point of the central GABA receptor, lowering cortisol levels and restoring synthesis of peripheral benzodiazepine receptors.

Other neurotransmitter systems. Individuals with GAD have been reported to have reduced serotonin levels in the cerebral spinal fluid[35] and decreased platelet binding of paroxetine, a selective serotonin reuptake inhibitor.[38] In addition, drugs that affect

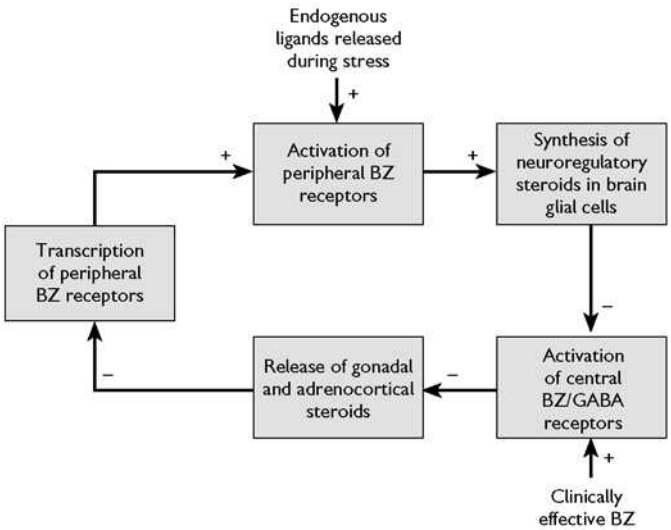


Fig 4.7.1.3 Possible involvement of peripheral benzodiazepine receptors in acute and chronic stress reactions (based on descriptions by Rocca *et al.*[37]). BZ, benzodiazepine.

serotonergic transmission (e.g. buspirone and venlafaxine) are effective in the treatment of GAD. These findings suggest that serotonin regulation may be abnormal in GAD.

Cholecystokinin neuropeptides (CCK-4 and CCK-8S) have been implicated in the genesis of arousal and fear responses.[40] It is unclear how those effects are mediated; however, cholecystokinin interacts with several neurotransmitters and systems believed to be involved in anxiety responses, including the noradrenergic nervous system, the hypothalamic-pituitary-adrenal axis, the benzodiazepine-GABA system, and serotonin.

(ii) Generalized psychological vulnerability

A diminished sense of control. Early experiences of uncontrollability may serve as a psychological vulnerability for emotional disorders.[16] Individuals in clinical populations, including people diagnosed with anxiety disorders, sexual dysfunctions, and depression, often perceive themselves as having little control over their experiences.[41,42] Patients with GAD are more likely than controls to perceive a lack of control over threatening events and to regard ambiguous information as threatening.[43] This perceived lack of control may result from a variety of events, including trauma and insecure attachment to primary caregivers.[44] In patients with GAD, worry may be an ineffective attempt to assert control on uncertain future events. Intolerance of uncertainty, a construct related to perceived lack of control, has emerged as an important variable in the study of anxiety disorders.[45] Defined as the inability to accept that future negative events may occur, intolerance of uncertainty has been associated with symptoms of several anxiety disorders, but is greater in individuals with GAD than in patients from a mixed anxiety disorder sample.[41]

Parenting. There is an extensive literature on the influences of early environmental factors on the development of anxiety and other negative emotions in children (for an integrative review, see Chorpita and Barlow[46]). Attachment theory holds that parents or other consistent caregivers serve an important function in a child's development by providing a protective and secure base from which the child can operate. Disruption of this base is hypothesized to lead initially to anxious apprehension and dependency and, if the disruption is severe, subsequently to withdrawal and depression.

An important aspect of a healthy parent–child relationship is its ability to foster in the child a sense of control over events. According to Chorpita and Barlow,[46] an individual who lacks sufficient early experiences of control may develop a general perception of personal inefficacy, which may predispose him or her to chronic negative emotional states such as GAD. Two aspects of parenting appear to be important in providing a child with opportunities to experience control: responsiveness to the child's efforts at engagement and encouragement of the child to explore and manipulate the environment. A parenting style characterized by excessive control of the child's environment (overprotection) coupled with a lack of warmth and responsiveness toward the child would deprive the child of such opportunities and thus, theoretically, could contribute to the development of anxiety.

Consistent with this theory, mothers of anxious preschool children were found to be more critical and intrusive and less responsive to their children than mothers of non-anxious children.[47] In addition, adults who rated their parenting as more protective and less caring had higher trait anxiety scores than other individuals surveyed. A similar pattern was found to distinguish patients meeting DSM-III-R criteria for GAD or panic disorder from controls.[48] One hypothesis is that the relationship of these early parenting experiences to the subsequent development of anxiety (or depression) is mediated by the early formation of cognitive vulnerability best described as a sense of uncontrollability regarding future events in one's life (Fig. 4.7.1.4).[46]

Specific psychological vulnerability. Data from twin studies suggest that environmental influences contribute more to the variance in aetiology of GAD than heredity.[19] In addition, the triple vulnerability model of the aetiology of anxiety disorders suggests that psychological and biological vulnerabilities interact with

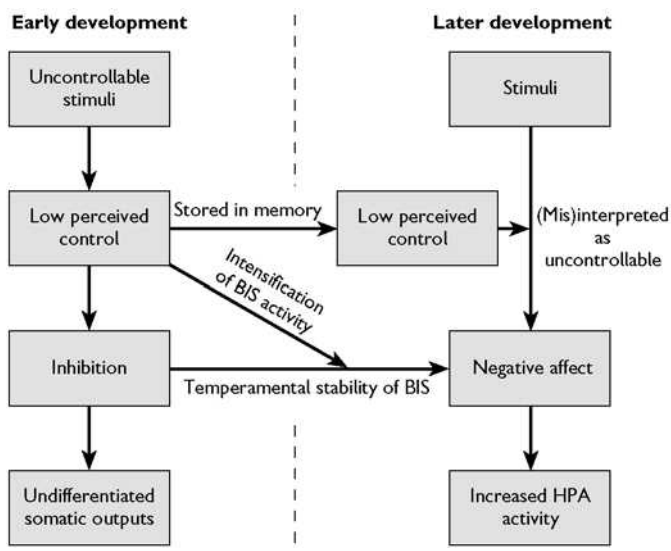


Fig 4.7.1.4 Model of the development of vulnerability for anxiety and depression. BIS, behavioural inhibition system[24]; HPA, hypothalamic-pituitary-adrenal axis (Copyright © 1998 by the American Psychological Association. Reproduced with permission. Chorpita, B.F. and Barlow, D.H. The Development of anxeity: the role of control in the early environment, *Psychological Bulletin*, **124**, 3–2. The use of APA information does not imply endorsement by APA).

environmental factors (i.e. stressful life events), to produce anxiety symptoms that reach a clinical level.[16,17] Several environmental factors have been implicated.

Stressful life events. Several studies have found an association between stressful or traumatic life events and GAD. The natures of the stressors that precede the onset of GAD remain unclear. However, in one study, 52 per cent of individuals diagnosed with GAD reported experiencing at least one past traumatic event (i.e. an event that would meet for Criterion A of the DSM-IV definition of PTSD), compared to only 21 per cent of non-anxious controls. It was unclear in this study if the events occurred prior to the onset of the GAD.[49] In addition, a variety of stressors have been associated with increased risk of GAD, including early parental death,[50] rape or combat,[51] and chronically dysfunctional marital and family relationships.[52]

Course and prognosis. Although there is evidence to the contrary, GAD has often been considered a chronic and disabling condition. The Harvard/Brown Anxiety Research Program study provides information on the course and impact of GAD among patients treated naturalistically over a 3-year period.[53] The mean age at onset of GAD was 21 years (range 2–61 years), and the average duration of illness at initial evaluation was 20 years. Excluding patients with comorbid panic disorder, one-third of subjects had never married and another 17 per cent were separated, widowed, or divorced. Unemployment was higher than average, and 37 per cent of subjects had received public financial assistance. Despite the fact that more than 80 per cent of patients received treatment during the study period, remission from GAD was uncommon (15 per cent at 1 year, 27 per cent by 3 years). Among patients with comorbid psychiatric disorders, the proportions achieving remission from GAD and coexisting anxiety disorders were only 8 per cent and 17 per cent at 1 and 3 years, respectively.

However, there is evidence that the perception of GAD as a chronic, unremitting condition may not be completely correct. A longitudinal study of individuals with GAD found that many (46 per cent of women and 56 per cent of men) experienced episodes of full remission and that the periods with no symptoms last longer in women.[54]

Treatment

(a) Pharmacotherapy

Several pharmacological agents have been shown to be effective for the treatment of GAD (for a review, see Davidson[23]). Chief among these are the benzodiazepines, azapirones, and antidepressants. Evidence shows that several types of medications may be effective for at least short-term relief of anxiety, however, many are not effective in the long term unless taken indefinitely.[55]

(i) Benzodiazepines

Benzodiazepines have for decades been prescribed for short-term relief of anxiety.[55] Although, evidence demonstrates that these drugs can be effective in relieving anxiety for a short period, there is little or no evidence that they work over a long period. However, compared with other agents used to treat anxiety disorders, they are safe, fast-acting, and have relatively few side-effects. All currently available benzodiazepines probably are efficacious for GAD. Approximately two-thirds of patients experience moderate to marked improvement, with effects being evident within the first 1 or 2 weeks of treatment.[56]

Benzodiazepines appear to be more effective for the somatic symptoms of GAD than for psychic symptoms such as apprehensive worry and irritability, possibly because of their sedative and myorelaxant properties. In some studies, irritability actually has increased during treatment with benzodiazepines. Consequently, these drugs may be better for patients whose complaints are more somatic than psychic, whereas other agents may be better when the reverse is true.

(ii) Azapirones

In recent years, azapirone drugs have become a popular alternative to benzodiazepines for the treatment of GAD. These drugs lack the sedative and muscle relaxant properties of the benzodiazepines as well as their ability to potentiate the effects of alcohol. However, improvement is somewhat slower (2–4 weeks) than with benzodiazepines. The most widely used of these agents is buspirone, whose efficacy and safety have been demonstrated in several well-controlled trials (see Rickels[57] for a review) and became, in 1996, the only non-benzodiazepine approved by the U.S. Food and Drug Administration for the treatment of GAD.[58] It is as effective as benzodiazepines for general anxiety and may reduce some of the associated features of GAD as well, including depression and agitation. Buspirone also does not seem to be associated with as much dependence and withdrawal as found with the benzodiazepines.[55] Ipsapirone, an azapirone with somewhat greater affinity and selectivity for the 5-hydroxytryptamine-1A receptor and fewer side-effects than buspirone also appears to be effective.[59] Other drugs in this class include gepirone, tandospirone, and flesinoxan.

(iii) Antidepressants

Both imipramine and trazodone have been superior to placebo for the treatment of GAD in controlled trials. In one trial, the two drugs were comparable to each other and to diazepam in reducing anxiety after the first 2 weeks of treatment.[56] In another study, imipramine was as effective as chlordiazepoxide overall and produced greater reductions in associated depression. In addition, nefazodone, an agent related to trazodone but less sedating, was effective for GAD in a small open trial.[60]

Many studies have examined the use of newer antidepressants, including selective reuptake inhibitors (SSRIs) and serotonin and norepinephrine reuptake inhibitors (SRNIs), in the treatment of GAD and have had favourable results. One placebo-controlled study of the SSRI paroxetine found that 62 per cent and 68 per cent of patients receiving 20 and 40 mg of paroxetine had significant reduction in symptoms, compared to only 46 per cent in the placebo group.[61] In addition, due to results from several large, placebo-controlled trials, venlafaxine was the first antidepressant approved by the Federal Drug Administration for the treatment of GAD.[62,63]

(b) Psychosocial treatment

Early psychological treatments for GAD consisted mostly of non-specific interventions such as supportive psychotherapy, relaxation training, and biofeedback. In general, those treatments were not very effective. However, cognitive behavioural treatment (CBT) has been found in several randomized controlled trials to be associated with clinically significant improvement. Cognitive behavioural treatment consists of psychoeducation about the nature of anxiety,

symptom monitoring, relaxation training, exposure (imaginal and *in vivo*), and cognitive restructuring. Psychoeducation is a common CBT component in which the nature of anxiety is discussed in order to normalize the patient's experience. It also serves to teach patients to differentiate between adaptive and maladaptive anxiety. Symptom monitoring allows records of mood and anxiety to be recorded and allows patients to learn to observe their symptoms. Progressive muscle relaxation (PMR) is a common technique used to teach patients relaxation strategies to use in anxiety-provoking situations. Exposures are tailored to the nature of GAD and are either *in vivo* or imaginal and are often combined with rehearsal of coping skills (i.e. PMR) until anxiety subsides. Cognitive restructuring targets distorted cognitions and information-processing biases associated with GAD. Strategies to challenge catastrophizing or overestimating the probability of a negative event are often employed during this component of treatment. CBT is typically administered in a dozen or so sessions, and can be conducted in group or individual formats. In controlled trials, cognitive behavioural treatments have been more effective than no treatment or a psychological or drug placebo treatment and at least as effective as benzodiazepines (for a review, see Barlow *et al.*).[64] Attrition is low (10–15 per cent), and reductions in anxiety average about 50 per cent, with gains being maintained at follow-up. Currently, the most successful treatments combine relaxation training with cognitive interventions focused on making the worry process more controllable.

Consistent with the findings of individual studies, meta-analyses for GAD have found that CBT is more efficacious than control conditions, resulting in medium to large effect sizes when compared to pill or psychological placebo.[65] However, in studies directly comparing CBT with pharmacotherapy (most often benzodiazepines), there was no difference in effect sizes between the two treatments. Currently, neither mode of treatment has been shown to be consistently superior to the other.

Recent developments in psychosocial treatments for GAD have integrated acceptance and mindfulness approaches into traditional CBT.[66] These treatments are based on the notion that individuals' attempts to control their internal experience often backfire, resulting in increased anxiety. Components to these treatments may include experiential exercises, mindfulness training, identification of overriding values, and encouragement towards taking action in ways that are consistent with these values.[66] Further exploration with controlled trials needs to be conducted to examine the efficacy of these newer approaches with cognitive behavioural treatments.

(c) Combined pharmacotherapy and psychotherapy

Although common in clinical practice, little is known about the effects of combining pharmacotherapy and psychotherapy for GAD. In the only published study to date, Power *et al.*[67] compared cognitive behavioural therapy, diazepam, a pill placebo, cognitive behavioural therapy plus diazepam, and cognitive behavioural therapy plus a pill placebo in a sample of DSM-III-diagnosed GAD patients. At post-treatment and follow-up, patients in all three cognitive behavioural therapy conditions were more improved than those who received diazepam or placebo alone. Although the cognitive behavioural therapy groups did not differ significantly from each other on any measure, the cognitive behavioural therapy plus diazepam group improved earliest and had the largest percentage of patients achieving a criterion of clinically significant change on all measures. Unfortunately, the use of DSM-III criteria in this study makes its relevance to GAD as it is currently defined uncertain. In addition, the cognitive behavioural therapy used was briefer (seven sessions) and less specific than the currently recommended forms.

(d) Effect of comorbidity on treatment outcome and vice versa

Many treatment trials investigating GAD have excluded patients with comorbid Axis I disorders. A review of 48 GAD studies published between 1980 and 1991 found that only eight reported including patients with other psychiatric disorders.[68] When comorbid disorders have been permitted, their effect on treatment outcome generally has not been evaluated. In the Harvard-Brown Anxiety Research Program study, the presence of a comorbid psychiatric disorder reduced response rates at 1 and 3 years by nearly 50 per cent. In addition, it has been found that concurrent personality disorders impair outcome of treatment for GAD. In one study improvement was comparable among treatment completers with or without Axis II disorders, but attrition was greater in the former (44 per cent) than the latter (19 per cent) group.[69]

The high rates of comorbidity associated with GAD have important implications for treatment. Higher rates of comorbidity are associated with lower rates of remission and greater likelihood of relapse over 12-year follow-up.[70] On the other hand, successful treatment of GAD in patients with comorbid disorders often reduces the severity of the other disorders as well. Borkovec *et al.*[68] examined the effect of various psychosocial treatments for GAD on coexisting anxiety and mood disorders (except panic disorder or severe depression, which were excluded from the study). Across treatments, patients whose GAD improved exhibited reductions as well in other anxiety disorders (mostly social and simple phobia) and dysthymia. Only 4 of 13 successfully treated patients who had additional disorders at pre-treatment continued to have them at post-treatment.

Clinical management

Many anxious patients do not meet diagnostic criteria for GAD. These patients often respond to conservative measures. If the symptoms are minor or are related to a situational stressor, brief psychotherapy and support is the treatment of first choice. In one study, patients who initially reported physical or minor emotional complaints responded better to counselling than to diazepam, even when counselling was limited to only 3 h.[72] Often, an explanation of the relationship of physical symptoms to stress is reassuring to patients and can interrupt a spiral of symptoms leading to anxiety and worry about health, leading to increased symptoms, and so on. Simple behavioural interventions such as relaxation training for patients with prominent muscle tension or breathing exercises for those with dyspnoea or hyperventilation may be helpful as well.

For patients with marked adrenergic symptoms or insomnia, the temporary (a few days to a few weeks) use of a benzodiazepine may be helpful as an adjunct to psychotherapy. In some cases, a hypnotic drug alone is sufficient. In general, as-needed use of benzodiazepines should be discouraged, because it is more likely than scheduled use to foster reliance on drugs as the principal means of coping with anxiety. For the same reason, the drug dose should be kept as low as possible and should be tapered as therapy proceeds.

For patients meeting diagnostic criteria for GAD, treatment with an empirically validated form of psychosocial therapy for GAD is strongly recommended. Such treatments are available in manualized form with accompanying patient workbooks.[55]

When medication treatment is preferred, a trial of buspirone is a good initial choice. Exceptions are patients with comorbid panic disorder or marked adrenergic symptoms, for which benzodiazepines may be better if they are not contraindicated (see below). The typical starting dose of buspirone is 15 mg/day in divided doses (5 mg thrice daily or 7.5 mg twice a day), which is increased by 5 mg/day every few days to a target dose of 30 mg/day. If the response is insufficient after 2 to 4 weeks at that amount, or if the patient is experiencing significant depressive symptoms, the dose may be advanced gradually to a maximum of 60 mg/day. Improvement may continue for up to 3 months.

It is important to inform patients of the typical side-effects and response time of buspirone. Patients who have taken benzodiazepines previously may expect prompt relief and sedative side-effects from the medication and may become discouraged when these are absent. When switching from a benzodiazepine to buspirone, it may be helpful to continue the benzodiazepine during the first month of buspirone therapy before initiating a gradual taper. Remember that buspirone will not prevent benzodiazepine withdrawal symptoms.

Failing a course of buspirone, or in patients with comorbid major depression, trials of antidepressant medications (e.g. imipramine, venlafaxine, trazodone) would be a reasonable second choice. Venlafaxine may be effective at doses as low as 75 mg/day (the usual starting dose). Because of its short metabolic half-life, the extended release form is preferred, which allows once per day dosing and thus may improve compliance. The dose typically is advanced by 75 mg/day every 2 weeks to a maximum of 225 mg/day. Dosing for other antidepressants is the same as for the treatment of depression.

Because of the risk of dependence and (uncommonly) abuse, long-term use of benzodiazepines generally is reserved for patients who do not respond sufficiently to other options. Relative contraindications include a need to be alert (e.g. drivers and machinery operators), a personal or family history of alcoholism or drug abuse, and prominent aggressiveness or irritability. Generally, longer half-life benzodiazepines (e.g. diazepam, clonazepam), which can be taken once or twice daily, are preferred. A typical starting dose is 5 to 10 mg/day of diazepam or equivalent, which is advanced every few days to a maximum of 40 mg/day.

In instances where pharmacotherapy is the primary treatment, responders should be continued on medication for at least 6 months before a gradual drug taper is attempted. Even so, a substantial proportion will relapse after drug discontinuance and will require further treatment. Discontinuing pharmacotherapy in the context of effective psychosocial treatment may improve success rates.

Conclusions

Based on current knowledge, GAD seems to be the exaggerated expression of the human potential to apprehensively anticipate and prepare for future misfortune. As such, it may represent a 'basic' disorder, a better understanding of which may shed light on other anxiety and mood disorders. Although the definition of GAD has been revised considerably since being given status of a full disorder in DSM-III-R, much more information needs to be gathered regarding psychosocial treatments for GAD. In addition, due to its poor diagnostic reliability, many revisions to the diagnostic criteria of GAD have been suggested. Despite the progress in elucidating the nature of GAD, further research needs to be conducted to continue to enhance our understanding and development of effective and generalizable treatments.

Further information

Barlow, D.H. (1987). The classification of anxiety disorders. In *Diagnoses and classification in psychiatry: a critical appraisal of DSM-III* (ed. G.L. Tischler), pp. 223–42. Cambridge University Press, New York.

Brown, T.A., Barlow, D.H., and Liebowitz, M.R. (1994). The empirical basis of generalized anxiety disorder. *The American Journal of Psychiatry*, **151**, 1272–80.

Kessler, R.C., McGonagle, K.A., Zhao, S., *et al.* (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. *Archives of General Psychiatry*, **51**, 8–19.

Barlow, D.H., Chorpita, B.F., and Turovsky, J. (1996). Fear, panic, anxiety, and disorders of emotion. In *Nebraska symposium on motivation. Perspectives on anxiety, panic, and fear*, Vol. 43 (ed. D.A. Hope), pp. 251–328. University of Nebraska Press, Lincoln, NE.

Barlow, D.H. (2000). Unraveling the mysteries of anxiety and its disorders from the perspective of emotion theory. *The American Psychologist*, **55**, 1247–63.

Suarez, L., Bennett, S.M., Goldstein, C.M., *et al.* (2008). Understanding anxiety disorders from a "triple vulnerability" framework. In *Handbook of anxiety and the anxiety disorders* (eds. M.M. Antony and M.B. Stein). Oxford University Press, New York.

LeDoux, J. (1996). *The emotional brain: the mysterious underpinnings of emotional life*. Simon and Schuster, New York.

Craske, M.G. and Barlow, D.H. (2006). *Mastery of your anxiety and worry* (2nd edn). Oxford University Press, New York.

Barlow, D.H., Allen, L.B., and Basden, S.L. (2007). Psychosocial treatments for panic disorders, phobias, and generalized anxiety disorder. In *A guide to treatments that work* (3rd edn) (ed. P.E. Nathan and J.M. Gorman), Oxford University Press, New York.

References

1. DiNardo, P.A., Moras, K., Barlow, D.H., *et al.* (1993). Reliability of DSM-III-R anxiety disorder categories: using the Anxiety Disorders Interview Schedule-Revised (ADIS-R). *Archives of General Psychiatry*, **50**, 251–6.
2. Brown, T.A., DiNardo, P.A., Lehman C.L., *et al.* (2001). Reliability of DSM-IV anxiety and mood disorders: implications for classification of emotional disorders. *Journal of Abnormal Psychology*, **110**, 49–58.
3. Abel, J.W. and Borkovec, T.D. (1995). Generalizability of DSM-III-R GAD to proposed DSM-IV criteria and cross validation of proposed changes. *Journal of Anxiety Disorders*, **9**, 303–15.
4. Craske, M.G., Rapee, R.M., Jackel, L., *et al.* (1989). Qualitative dimensions of worry in DSM-III-R generalized anxiety disorder subjects and nonanxious controls. *Behavior Research and Therapy*, **27**, 189–98.
5. Barlow, D.H. and Wincze, J. (1998). DSM-IV and beyond: what is generalized anxiety disorder? *Acta Psychiatrica Scandinavica*, **98** (Suppl. 393), 23–9.
6. Grant, B.F., Stinson, F.S., Dawson, D.A., *et al.* (2004). Prevalence and co-occurrence of substance abuse disorders and independent mood and anxiety disorders. *Archives of General Psychiatry*, **61**, 807–16.
7. Grant, B.F., Hasin, D.S., Stinson, F.S., *et al.* (2005). Prevalence, correlates, co-morbidity, and comparative disability of DSM-IV

generalized anxiety disorder in the US: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Psychological Medicine*, **35**, 1747–59.
8. Kessler, R.C., Berglund, P., Demler, O., *et al.* (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **62**, 593–602.
9. Kessler, R.C., Chiu, W.T., Demler, O., *et al.* (2005). Prevalence, severity, comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **62**, 617–27.
10. Brown, T.A., Campbell, L.A., Lehman, C.L., *et al.* (2001). Current and lifetime comorbidity of the DSM-IV anxiety and mood disorders in a large clinical sample. *Journal of Abnormal Psychology*, **110**, 585–99.
11. Wittchen, H.-U., Zhao, S., Kessler, R.C., *et al.* (1994). DSM-III-R generalized anxiety disorder in the National Comorbidity Survey. *Archives of General Psychiatry*, **51**, 355–64.
12. Zinbarg, R.E., Barlow, D.H., Liebowitz, M., *et al.* (1994). The DSM-IV field trial for mixed anxiety-depression. *American Journal of Psychiatry*, **151**, 1153–62.
13. Zinbarg, R. and Barlow, D.H. (1996). Structure of anxiety and the anxiety disorders: a hierarchial model. *Journal of Abnormal Psychology*, **105**, 181–93.
14. Brown, T.A., Chorpita, B.F., and Barlow, D.H. (1998). Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology*, **107**, 179–92.
15. Barlow, D.H., Chorpita, B.F., and Turovsky, J. (1996). Fear, panic, anxiety, and disorders of emotion. In *Nebraska Symposium on Motivation.* Vol. 43, *Perspectives on anxiety, panic, and fear* (ed. D.A. Hope), pp. 251–328. University of Nebraska Press, Lincoln, NE.
16. Barlow, D.H., (2002). *Anxiety and its disorders* (2nd edn). Guilford Press, New York.
17. Suarez, L., Bennett, S.M., Goldstein, C.M., *et al.* (2008). Understanding anxiety disorders from a "triple vulnerability" framework. In *Handbook of anxiety and the anxiety disorders* (eds. M.M. Antony and M.B. Stein). Oxford University Press, New York.
18. Hettema, J.M., Prescott, C.A., Myers, J.M., *et al.* (2005). The structure of genetic and environmental risk factors for anxiety disorders in men and women. *Archives of General Psychiatry*, **62**, 182–8.
19. Hettema, J.M., Neale, M.C., and Kendler, K.S. (2001). A review and meta-analysis of the genetic epidemiology of anxiety disorders. *Archives of General Psychiatry*, **158**, 1568–78.
20. Noyes, R., Clarkson, C., Crowe, R.R., *et al.* (1987). A family study of generalized anxiety. *The American Journal of Psychiatry*, **144**, 1019–24.
21. Andrews, G., Stewart, S., Allen, R., *et al.* (1990). The genetics of six neurotic disorders: a twin study. *Journal of Affective Disorders*, **19**, 23–9.
22. Kendler, K.S., Neale, M.C., Kessler, R.C., *et al.* (1992). Major depression and generalized anxiety disorder: same genes, (partly) different environments? *Archives of General Psychiatry*, **49**, 716–22.
23. Davidson R.T. (2001). Pharmacotherapy of generalized anxiety disorder. *The Journal of Clinical Psychiatry*, **62**(Suppl. 11), 46–50.
24. Gray, J.A. and McNaughton, N. (1996). The neuropsychology of anxiety: a reprise. In *Nebraska symposium on motivation. Perspectives on anxiety, panic, and fear*, Vol. 43 (ed. D.A. Hope), pp. 61–134. University of Nebraska Press, Lincoln, NE.
25. Charney, D.S., Woods, S.W., Heninger, G.R., *et al.* (1989). Noradrenergic function in generalized anxiety disorder: effects of yohimbine in healthy subjects and patients with generalized anxiety disorder. *Psychiatry Research*, **27**, 173–82.
26. Abelson, J.L., Glitz, D., Cameron, O.G., *et al.* (1991). Blunted growth hormone response to clonidine in patients with generalized anxiety disorder. *Archives of General Psychiatry*, **48**, 157–62.
27. Cameron, O.G., Smith, C.B., Lee, M.A., *et al.* (1990). Adrenergic status in anxiety disorders: platelet alpha two-adrenergic receptor binding, blood pressure, pulse, and plasma catecholamines in panic and generalized anxiety disorder patients and in normal subjects. *Biological Psychiatry*, **28**, 3–20.
28. Hazlett, R.L., McLeod, D.R., and Hoehn-Saric, R. (1994). Muscle tension in generalized anxiety disorder: elevated muscle tonus or agitated movement? *Psychophysiology*, **31**, 189–95.
29. Sapolsky, R.M. (2007). Stress, stress-related disease, and emotional regulation. In *Handbook of emotional regulation* (ed. J.J. Gross). Guilford Press, New York.
30. Tiller, J.W.G., Biddle, N., Maguire, K.P., *et al.* (1988). The dexamethasone suppression test and plasma dexamethasone in generalized anxiety disorder. *Biological Psychiatry*, **23**, 261–70.
31. LeDoux, J. (1998). Fear and the brain: where have we been, and where are we going? *Biological Psychiatry*, **44**, 1229–38.
32. DeBellis, M.D., Casey, B.J., Dahl, R.E., *et al.* (2000). A pilot study of amygdala volumes in pediatric generalized anxiety disorder. *Biological Psychiatry*, **48**, 51–7.
33. McClure, E.B., Monk, C.S., Nelson, E.E., *et al.* (2007). Abnormal attentional modulation of fear circuit function in pediatric generalized anxiety disorder. *Archives of General Psychiatry*, **64**, 97–106.
34. Davis, M. (1998). Are different parts of the extended amygdala involved in fear versus anxiety? *Biological Psychiatry*, **44**, 1239–47.
35. Waddell, J., Morris, R.W., and Bouton, M.E. (2006). Effect of bed nucleus of the stria terminalis lesions on conditioned anxiety: aversive conditioning with long-duration conditional stimuli and reinstatement of extinguished fear. *Behavioral Neuroscience*, **120**, 324–36.
36. Gray, J.A. (1982). *The neuropsychology of anxiety*. Oxford University Press, New York.
37. Rocca, P., Beoni, A.M., Eva, C., *et al.* (1998). Peripheral benzodiazepine receptor messenger RNA is decreased in lymphocytes of generalized anxiety disorder patients. *Biological Psychiatry*, **43**, 767–73.
38. Brewerton, T.D., Lydiard, R.B., Johnson, M.R., *et al.* (1995). CSF serotonin: diagnostic and seasonal differences. In *New research abstracts of the 148th meeting of the American Psychiatric Association*, Abstract NR385:151. American Psychiatric Press, Washington, DC.
39. Iny, L.J., Pecknold, J., Suranyi-Cadotte, B.E., *et al.* (1994). Studies of neurochemical link between depression, anxiety, and stress from [^{3}H]imipramine and [^{3}H]paroxetine binding on human platelets. *Biological Psychiatry*, **36**, 281–91.
40. Bradwejn, J., Koszycki, D., Couetoux du Tertre, A., *et al.* (1992). The cholecystokinin hypothesis of panic and anxiety disorders: a review. *Journal of Psychopharmacology*, **6**, 345–51.
41. Ladouceur, R., Dugas, M.J., Freeston, M.H., *et al.* (1999) Specificity of generalized anxiety disorder symptoms and processes. *Behavior Therapy*, **30**, 191–207.
42. Weisberg, R.B., Brown, T.A., Wincze, J., *et al.* (2001). Causal attributions and male sexual arousal: the impact of attributions for a bogus erectile difficulty on sexual arousal, cognitions, and affect. *Journal of Abnormal Psychology*, **110**, 324–34.
43. Rapee, R.M. (1991). Generalized anxiety disorder: a review of clinical features and theoretical concepts. *Clinical Psychology Review*, **11**, 419–40.
44. Borkovec, T.D. (1994). The nature, functions, and origins of worry. In *Worrying: perspectives on theory, assessment, and treatment* (eds. G.C.L. Davey and F. Tallis). Wiley, New York.
45. Dugas, M.J., Gagnon, F., Ladouceur, R., *et al.* (1998) Generalized anxiety disorder: a preliminary test of a conceptual model. *Behavior Research Therapy*, **36**, 215–26.
46. Chorpita, B.F. and Barlow, D.H. (1998). The development of anxiety: the role of control in the early environment. *Psychological Bulletin*, **124**, 3–21.

47. Dumas, J.E., LaFreniere, P.J., and Serketich, W.J. (1995). 'Balance of power': a transactional analysis of control in mother-child dyads involving socially competent, aggressive, and anxious children. *Journal of Abnormal Psychology*, **104**, 104–13.
48. Silove, D., Parker, G., Hadzi-Pavlovic, D., *et al.* (1991). Parental representations of patients with panic disorder and generalized anxiety disorder. *The British Journal of Psychiatry*, **159**, 835–41.
49. Roemer, L., Molina, S., Litz., B.T., *et al.* (1996). Preliminary investigation of the role of previous exposure to potentially traumatizing events in generalized anxiety disorder. *Depression and Anxiety*, **4**, 134–8.
50. Torgersen, S. (1986). Childhood and family characteristics in panic and generalized anxiety disorders. *The American Journal of Psychiatry*, **143**, 630–2.
51. Steketee, G. and Foa, E.B. (1987). Rape victims: post traumatic stress responses and their treatment: a review of the literature. *Journal of Anxiety Disorders*, **1**, 69–86.
52. Ben-Noun, L. (1998). Generalized anxiety disorder in dysfunctional families. *Journal of Behavior Therapy and Experimental Psychiatry*, **29**, 115–22.
53. Yonkers, K.A., Warshaw, M.G., Massion, A.O., *et al.* (1996). Phenomenology and course of generalised anxiety disorder. *The British Journal of Psychiatry*, **168**, 308–13.
54. Yonkers, K.A., Bruce, S.E., Dyck, I.R., *et al.* (2003). Chronicity, relapse, and illness-course of panic disorder, social phobia, and generalized anxiety disorder: findings in men and women from 8 years of follow-up. *Depression and Anxiety*, **17**, 173–9.
55. Craske, M.G. and Barlow, D.H. (2006). *Mastery of your anxiety and worry* (2nd edn). Oxford University Press, New York.
56. Rickels, K., Downing, R., Schweizer, E., *et al.* (1993). Antidepressants for the treatment of generalized anxiety disorder: a placebo-controlled comparison of imipramine, trazodone, and diazepam. *Archives of General Psychiatry*, **50**, 884–95.
57. Rickels, K. (1990). Buspirone in clinical practice. *The Journal of Clinical Psychiatry*, **51**(Suppl. 9), 51–4.
58. Apter, J.T. and Allen, L.A. (1999). Buspirone: future directions. *Journal of Clinical Psychopharmacology*, **19**, 86–93.
59. Cutler, N.R., Sramek, J.J., Keppel-Hesselink, J.M., *et al.* (1993). A double-blind, placebo-controlled study comparing the efficacy and safety of ipsapirone versus lorazepam in patients with generalized anxiety disorder: a prospective multicenter trial. *Journal of Clinical Psychopharmacology*, **13**, 429–37.
60. Hedges, D.W., Reimherr, F.W., Strong, R.E., *et al.* (1996). An open trial of nefazodone in adult patients with generalized anxiety disorder. *Psychopharmacology Bulletin*, **32**, 671–6.
61. Rickels, K., Zanielli, R., McCafferty, J., *et al.* (2003). Paroxetine treatment of generalized anxiety disorder: a double-blind, placebo-controlled study. *The American Journal of Psychiatry*, **160**, 749–56.
62. Sheehan, D.V. (1999). Venlafaxine extended release (XR) in the treatment of generalized anxiety disorder. *The Journal of Clinical Psychiatry*, **60**(Suppl. 22), 23–8.
63. Davidson, J.R., DuPont, R.L., Hedges, D., *et al.* (1999). Efficacy, safety, and tolerability of venlafaxine extended release and buspirone in outpatients with generalized anxiety disorder. *The Journal of Clinical Psychiatry*, **60**, 528–35.
64. Barlow, D.H., Allen, L.B., and Basden, S.L. (2007). Psychosocial treatments for panic disorders, phobias, and generalized anxiety disorder. In *A guide to treatments that work* (3rd edn) (ed. P.E. Nathan and J.M. Gorman). Oxford University Press, New York.
65. Mitte, K. (2005). Meta-analysis of cognitive-behavioral treatments for generalized anxiety disorder: A comparison with pharmacotherapy. *Psychological Assessment*, **4**, 224–7.
66. Orsillo, S.M., Roemer, L., and Barlow, D.H. (2003). Integrating acceptance and mindfulness into existing cognitive-behavioral treatments for GAD. *Cognitive and Behavioral Practice*, **10**, 222–30.
67. Power, K.G., Simpson, R.J., Swanson, V., *et al.* (1990). A controlled comparison of cognitive-behaviour therapy, diazepam, and placebo, alone and in combination, for the treatment of generalized anxiety disorder. *Journal of Anxiety Disorders*, **4**, 267–92.
68. Swinson, R.P., Cox, B.J., and Fergus, K.D. (1993). Diagnostic criteria in generalized anxiety disorder treatment studies. *Journal of Clinical Psychopharmacology*, **13**, 455.
69. Sanderson, W.C., Beck, A.T., and McGinn, L.K. (1994). Cognitive therapy for generalized anxiety disorder: significance of comorbid personality disorders. *Journal of Cognitive Psychotherapy*, **8**, 13–18.
70. Bruce, S.E., Yonkers, K.A., Otto, M.W., *et al.* (2005). Influence of psychiatric comorbidity on recovery and recurrence in generalized anxiety disorder, social phobia, and panic disorder: a 12-year prospective study. *The American Journal of Psychiatry*, **162**, 1179–87.
71. Borkovec, T.D., Abel, J.L., and Newman, H. (1995). Effects of psychotherapy on comorbid conditions in generalized anxiety disorder. *Journal of Consulting and Clinical Psychology*, **63**, 479–83.
72. Boulenger, J., Fournier, M., Rosales, D., *et al.* (1997). Mixed anxiety and depression: from theory to practice. *The Journal of Clinical Psychiatry*, **58**(Suppl. 8), 27–34.

4.7.2 Social anxiety disorder and specific phobias

Michelle A. Blackmore, Brigette A. Erwin, Richard G. Heimberg, Leanne Magee, and David M. Fresco

Introduction

As our classification systems have been refined, we have come to view social anxiety disorder (social phobia) and specific phobias as distinct disorders, with divergent patterns of prevalence, aetiology, and course. Moreover, treatments for these disorders have become increasingly sophisticated. This chapter presents an overview of the current state of the field with regard to social anxiety disorder and specific phobias.

Social anxiety disorder

In the first and second editions of the *Diagnostic and Statistical Manual of Mental Disorders* (**DSM**),[1,2] all phobias were grouped together. However, in 1966 Marks and Gelder[3] observed that various phobias had different ages of onset and symptom presentations, providing the initial impetus for the inclusion of *social phobia* as a distinct disorder in DSM-III.[4] At first, research into the nature and treatment of social phobia lagged behind that of other anxiety disorders, leading to its description in 1985 as the neglected anxiety disorder.[5] Over the past two decades, however, attention to the conceptualization, definition, and classification of social phobia has increased dramatically. To acknowledge the significant impairment now known to be associated with social phobia and its differentiation from specific phobia, the alternative (and increasingly utilized) label, *social anxiety disorder*, was added in DSM-IV.[6]

Clinical presentation

Anxiety in situations involving potential evaluation by others (e.g. job interviews, public speaking engagements, first dates) falls within the realm of 'normal' social anxiety. For individuals with social anxiety disorder, however, situations such as these are typically associated with incapacitating levels of anxiety and a desire for escape or avoidance. Socially anxious individuals are often self-critical and perfectionistic and go to great lengths to avoid the negative evaluation of others that they may perceive as epidemic. Commonly, persons with social anxiety disorder experience somatic symptoms such as blushing, trembling, dry mouth, or perspiring, which they believe will be noticed by others and provide further evidence of their incompetence. Children may manifest their anxiety differently than adults; they may cry, throw tantrums, freeze, shrink from interactions with strangers, and they may not acknowledge that their fears are irrational.[6] By leaving anxiety-provoking situations (escape) or foregoing them entirely (avoidance), individuals with social anxiety disorder may reduce or prevent the immediate experience of anxiety, but this behaviour reinforces beliefs in their inadequacies and serves to maintain anxiety in the absence of objective threat.[7,8]

Functional impairment

Individuals with social anxiety disorder experience significant impairment in social, educational, and occupational functioning.[5,9] They are less likely to marry and are more likely to divorce than those without the disorder.[10] They also have fewer friends and more trouble getting along with the friends they have than persons without the disorder.[11] Individuals with social anxiety disorder assessed in a primary care setting reported missing an average of 3 days of work and having an average of 6 days of reduced productivity in the last month because of their emotional problems.[12] Comparatively, mentally healthy individuals reported less than 1 day of lost work and reduced productivity combined. Unemployment, underemployment (working at a level below the individual's abilities), and financial dependency are also characteristic of individuals with social anxiety disorder.[10]

Classification

(a) DSM and ICD

Whereas the DSM is widely used in North America, the *International Classification of Mental and Behavioural Disorders* (**ICD**) is commonly used in other parts of the world. Social anxiety disorder, termed social phobia in ICD-10,[13] first appeared in ICD-10 12 years after its appearance in DSM-III. The ICD-10 criteria for social phobia are less detailed and more circumscribed than those in DSM-IV. Specifically, DSM-IV requires excessive fear of humiliation or embarrassment in social or performance situations, anxiety provoked by exposure to feared situations, recognition that the fear is excessive, avoidance of situations or endurance with distress, and significant distress or impairment. Further, the fear and avoidance cannot be better accounted for by another psychiatric disorder, a general medical condition, or the effects of a substance. The ICD-10, in contrast, requires only that the symptoms be representative of anxiety and not another psychiatric disorder, that the anxiety occurs in relation to social situations, and that avoidance of anxiety-provoking situations be present. Because most published research on social anxiety disorder relies on DSM rather than ICD criteria, this chapter will do so also.

(b) Diagnostic issues

Individuals presenting for treatment of social anxiety disorder endorse multiple fears and significant impairment. The *generalized* subtype is specified when 'most social situations' are feared, whereas the *non-generalized* subtype describes persons who fear a more limited set of social situations. Individuals with generalized and non-generalized social anxiety disorder differ on several dimensions, including symptom severity, functional impairment, and physiological symptoms when exposed to feared situations.[14] Conclusive differences between subtypes in course and response to treatment remain to be demonstrated.[15,16]

Like social anxiety disorder, avoidant personality disorder is regarded as an extreme fear of negative evaluation, leading researchers to view the two conditions on a continuum that is artificially divided at the boundary between Axes I and II. Many investigators conclude that the co-occurrence of generalized social anxiety disorder and avoidant personality disorder represent persons with the most severe social anxiety and the poorest global functioning.[17]

Social anxiety may also develop as a result of medical conditions, such as becoming excessively anxious or avoiding social situations because of obesity, acne, benign essential tremor, stuttering, or the disability associated with Parkinson's disease. These conditions are not considered exemplars of social anxiety disorder because anxiety developed secondary to the medical condition. Rather, they are assigned to the category 'anxiety disorder not otherwise specified'. However, persons who experience secondary social anxiety are often responsive to pharmacological or cognitive behavioural treatments with demonstrated efficacy for social anxiety disorder.[18]

(c) Comorbidity and differential diagnosis

Approximately 81 per cent of persons with primary social anxiety disorder meet criteria for at least one other lifetime psychiatric disorder.[19] Social anxiety disorder most commonly co-occurs with other anxiety disorders,[20] although comorbid diagnoses of depression and alcohol use disorders are also common. Differential diagnosis is complicated by the fact that certain Axis I disorders both resemble and co-occur with social anxiety disorder.

(d) Social anxiety disorder versus panic disorder with agoraphobia (PDA)

PDA can be differentiated from social anxiety disorder in several ways. Although many individuals with social anxiety disorder experience panic attacks, the attacks occur in anticipation of negative evaluation by others. For persons with panic disorder, panic attacks are often unexpected, may not be associated with specific cognitions, and can be nocturnal.[5] Persons with social anxiety disorder are more likely to experience blushing and muscle twitches, whereas individuals with PDA are more likely to experience symptoms such as blurred vision, headaches, chest pain, ringing in the ears, and fear that they will die or go crazy.[21] The age of onset for social anxiety disorder tends to be earlier than that for PDA.[22] Individuals presenting for social anxiety disorder treatment either show an equal gender distribution or are slightly more likely to be male,[23] whereas those presenting for PDA treatment are substantially more likely to be female.[21] Finally, persons with social

anxiety disorder report feeling more comfortable when alone, whereas persons with PDA may feel more at ease in the presence of others.[22]

(e) Social anxiety disorder versus generalized anxiety disorder (GAD)

Individuals with GAD endorse higher levels of social anxiety than other persons with non-social anxiety disorders.[23] Although individuals with either social anxiety disorder or GAD may devote excessive amounts of time to worrying and ruminating, the focus of worry in social anxiety disorder is on fear of evaluation in social or performance situations, whereas the hallmark feature of worry in GAD is heightened focus on possible catastrophic consequences across several domains of life. Persons with social anxiety disorder are more likely to experience sweating, flushing, and breathing problems; those with GAD more commonly experience headaches, insomnia, and fear of dying.[24]

(f) Social anxiety disorder versus depression

Social anxiety disorder and depression may have withdrawal from social situations in common.[21] In differentiating between the two disorders, one must consider the reason for this withdrawal. Persons with depression withdraw because they fail to experience pleasure or lack the energy for social engagement. Individuals with social anxiety disorder fear the negative evaluation they believe to be associated with such situations. Persons with depression may be indifferent about engaging in social situations, whereas individuals with social anxiety disorder often have a strong desire to affiliate with others that is hampered by anxiety.

Epidemiology

(a) Prevalence

The National Comorbidity Survey Replication (**NCS-R**) reported a lifetime prevalence rate of 12.1 per cent[25] and a 12-month prevalence rate of 6.8 per cent[20] for social anxiety disorder. NCS-R lifetime prevalence rates render social anxiety disorder the fourth most common psychiatric disorder behind major depression (16.6 per cent), alcohol abuse (13.2 per cent), and specific phobia (12.5 per cent).

(b) Age at onset/age of treatment seeking

Social anxiety disorder often begins early in life. Mean age of onset for the disorder ranges from 13 to 20 years old, although patients often report having experienced symptoms for as long as they can remember.[26] Despite early onset, persons with social anxiety disorder often do not seek treatment for approximately 16 years after onset,[27] and many never do.[28]

(c) Gender differences

Although men and women with social anxiety disorder who seek treatment do so in relatively equal numbers,[23] epidemiological studies suggest that women are more likely than men to have the disorder.[19,25] In a clinical sample, women reported fear of more social situations and scored higher on several social anxiety disorder assessment measures.[29] Thus, although women are more likely to experience social anxiety, men are more likely to seek treatment and may do so when troubled with less severe symptoms. It may be that social anxiety disorder impairs the expected role functioning of men to a greater extent that it does for women.[29]

Aetiology of social anxiety disorder

Genetic factors appear to contribute to the emergence of social anxiety disorder. Higher rates of social anxiety disorder have been found in relatives of individuals with the disorder compared to relatives of persons without the disorder.[30] Further, rates of social anxiety disorder in first-degree relatives of probands with the generalized subtype are higher than in relatives of probands with the non-generalized subtype or with no psychiatric history.[31] Kendler *et al.*[32] report concordance rates for social anxiety disorder among monozygotic twins (24.4 per cent) to be greater than the rates for dizygotic twins (15.3 per cent). A study conducted with the same cohort 8 years later found the heritability of social anxiety disorder to be approximately 50 per cent in female twins[33] and 25 per cent in male twins.[34]

Neurobiological factors may also be associated with social anxiety disorder. Imaging studies of individuals with social anxiety disorder have demonstrated increased activity in regions associated with fear and anxiety (i.e. prefrontal cortex, amygdala, hippocampus) during anxiety-provoking tasks.[35] Given the efficacy of serotonin reuptake inhibitors and monoamine oxidase inhibitors in treating social anxiety disorder,[36] dysregulation of the serotonin[37] and dopamine[38] systems have been investigated as potential correlates of the disorder.

Several studies also suggest the importance of parental influences and significant life events in the development of social anxiety disorder. Persons with social anxiety recall observing their mothers act more fearful and avoidant of social interactions[39] and describe their parents as overprotective.[40] Stressful social and performance situations early in life (e.g. public ridicule, being bullied, mind going blank during a presentation) are also commonly reported by persons with social anxiety disorder.[41]

Course of social anxiety disorder

Social anxiety disorder is chronic and unlikely to remit without treatment. The disorder persists throughout adulthood[42] and its course is unrelated to gender, age of onset, duration of illness, level of functioning at intake, lifetime history of anxiety disorders, or current comorbidity of anxiety or depressive disorders.[42,43] Two conditions related to social anxiety disorder emerge in childhood and are relatively stable into adulthood—shyness and behavioural inhibition. Individuals who had been shy as children exhibited overall lower levels of functioning when assessed 30 years later.[44] Similarly, children described as behaviourally inhibited, or having the tendency to withdraw from novel people, settings, or objects, have demonstrated increased risk for the development of social anxiety disorder in adolescence.[45] Behavioural inhibition was also more prevalent in children of individuals with anxiety disorders and remained relatively stable for over 7 years in children initially assessed between the ages of 21 to 31 months.[46] These findings suggest that extreme shyness and behavioural inhibition may be early manifestations of social anxiety disorder.

Empirically evaluated treatments

(a) Cognitive behavioural interventions

Cognitive behavioural treatments have been subjected to the most thorough evaluation in the empirical literature. Treatments that utilize exposure alone or combined with cognitive restructuring have received the greatest empirical support and are the focus of

our review. Because of space limitations, other cognitive behavioural treatments, such as social skills training and applied relaxation, will not be reviewed, and the reader is referred to other sources.[47]

(b) Exposure

Exposure requires individuals to imagine (imaginal exposure) or directly confront (*in vivo* exposure) feared stimuli. Research examining the efficacy of imaginal exposure for social anxiety disorder is limited; however, *in vivo* exposure has repeatedly demonstrated short- and long-term efficacy in therapist-directed and self-directed formats.[47] Exposure requires patients to progressively confront anxiety-provoking situations beginning with situations that elicit moderate fear. Patients turn to the next most feared situation after repeated and prolonged exposure to the previous situation no longer elicits a distressing level of fear. Individuals with social anxiety disorder treated with exposure alone experienced greater improvement than individuals receiving relaxation training,[48] pill placebo,[49] or delayed treatment.[50]

(c) Exposure combined with cognitive restructuring

Contemporary cognitive behavioural models of social anxiety disorder propose that anxiety is largely maintained by dysfunctional beliefs and information-processing biases, and that successful treatment will be associated with modification of cognition.[7,8] Accordingly, exposure is typically combined with techniques designed to modify dysfunctional thinking patterns.[51,52]

Cognitive restructuring is an intervention based on the theory that one's thoughts about a situation, not the situation itself, generate anxiety.[53] The intervention is designed to help patients challenge maladaptive beliefs by identifying irrational thoughts, evaluating the dysfunction inherent in these thoughts, and deriving rational alternatives to these thoughts. By engaging in this process and utilizing alternative thoughts during exposure to feared situations, patients acquire new, adaptive learning that competes with their previously-learned maladaptive views, and, in turn, lessens the anxiety they experience.[54]

Efficacy for the combination of cognitive restructuring and exposure has been demonstrated in comparison to wait-list control conditions,[55,56] pill placebo,[57] and psychological placebo conditions.[57,58] Several studies demonstrate equivalent outcomes for exposure alone and exposure plus cognitive restructuring, whereas others indicate the combination shows superior efficacy and additional gains at follow-up.[47] In one meta-analysis,[59] only the combination of exposure and cognitive restructuring was superior to placebo treatments, but this difference has not been reliably demonstrated.[60] Nevertheless, patients treated with exposure alone tend to show deterioration of gains after treatment, suggesting durability of gains may be enhanced with the addition of cognitive restructuring techniques.

(d) Cognitive behavioural group therapy (CBGT)

CBGT, originally developed by Heimberg and Becker,[51] is one of the most thoroughly examined cognitive behavioural treatments for social anxiety disorder. It integrates cognitive techniques and exposure and is typically conducted in 12 weekly, 2.5 h sessions, with approximately six patients and two therapists. In sessions 1–2, patients receive psychoeducation, rationale and instructions for exposure, training in cognitive restructuring, and homework assignments. Thereafter, therapists lead patients through individualized exposures preceded and followed by therapist-directed cognitive restructuring exercises. For homework, patients practice cognitive restructuring before and after exposure to real-life anxiety-provoking situations.

One study evaluating the efficacy of CBGT compared it to educational-supportive group therapy (ES), a credible placebo treatment consisting of lectures, discussions, and social support. Seventy-five per cent of CBGT patients made significant improvement compared to 40 per cent of ES patients.[58] At follow-up (4.5–6.25 years), CBGT produced durable treatment gains.[61] A comparison of CBGT to the monoamine oxidase inhibitor phenelzine, pill placebo, and ES demonstrated equivalent response and attrition rates after 12 weeks of treatment for CBGT and phenelzine, both of which were superior to placebo and ES.[57] Although the phenelzine group evidenced superior improvement on a subset of measures after 12 weeks, CBGT demonstrated more durable treatment gains, with only 17 per cent relapse compared to 50 per cent relapse in the phenelzine group after a 6-month maintenance phase and 6-month follow-up phase.[62] An intensive version of CBGT, based on a hybrid of the treatment approaches developed by Heimberg and by Clark, involving 2 weeks of daily treatment sessions separated by 1 week of homework assignments, also proved superior to a wait-list control.[63]

(e) Individual cognitive behavioural therapy

Group CBT may not always be feasible, particularly in clinical settings where it may be difficult to obtain an adequate number of patients to form a group. However, CBGT has been adapted to an individual format and proven superior to a wait-list control (Heimberg, unpublished observations). Clark[52] also developed a cognitively-focused individual treatment for social anxiety disorder that has demonstrated substantial efficacy. The treatment instructs patients on how to shift their attention externally (rather than on the self) and to reduce reliance on safety behaviours. Video feedback and exposure to feared situations aimed at restructuring distorted cognitions are also incorporated.

Individual cognitive therapy demonstrated superior efficacy to wait-list control, with clinically significant gains observed in 76 per cent of patients receiving cognitive therapy, compared to 38 per cent of patients receiving an applied relaxation treatment and 0 per cent of patients in the wait-list control group.[64] Cognitive therapy was also more efficacious than fluoxetine plus self-exposure instructions and placebo plus self-exposure instructions, with gains maintained at 1-year follow-up.[65] Although meta-analytic studies suggest that individual and group CBT are similar in efficacy,[60] individual cognitive therapy was superior to a group therapy based on Clark's model on several measures.[66] Similarly, individual cognitive therapy proved superior to a 3-week intensive group therapy based on Clark's model and treatment with psychiatrist-selected SSRIs.[67]

Pharmacotherapy

The efficacy of selective serotonin uptake inhibitors (**SSRI**s), serotonin–norepinephrine reuptake inhibitor venlafaxine (**SNRI**s), benzodiazepines, reversible and irreversible monoamine oxidase inhibitors (**MAOI**s), ß-adrenergic blockers, buspirone, gabapentin, and pregabalin have been evaluated in placebo-controlled studies. To date, most controlled trials have included patients with generalized social anxiety disorder, so it is unclear if similar efficacy would be observed among patients with the non-generalized subtype of the disorder.

(a) Selective serotonin reuptake inhibitors and venlafaxine

The SSRIs (e.g. paroxetine, sertraline, fluoxetine, fluvoxamine, and escitalopram) and the SNRI venlafaxine are regarded as first-line medication treatments for generalized social anxiety disorder, given demonstrated efficacy in placebo-controlled trials for the disorder as well as for other anxiety and affective disorders. Further, potential adverse side effects of SSRIs (e.g. sweating, sexual dysfunction) pose a lower risk than potential adverse side effects of the MAOIs (e.g. high blood pressure, seizure) or benzodiazepines (e.g. dependence, overdose).

The SSRIs have demonstrated superior efficacy to placebo as well as CBT in the short-term,[68] with the possible exception of fluoxetine.[65,69] There have been few direct comparisons of the various SSRIs or SNRIs. One study comparing escitalopram (5, 10, 20 mg), paroxetine (20 mg), and placebo, demonstrated superior efficacy for 5 mg and 20 mg of escitalopram and 20 mg of paroxetine compared to placebo at week 12.[70] At week 24, all doses of escitalopram proved superior to placebo and 20 mg of escitalopram proved superior to paroxetine on a subset of social anxiety measures. The SNRI venlafaxine produced results comparable to paroxetine, including superior efficacy relative to placebo.[71]

A fixed-dose trial of paroxetine showed 20 mg/day to be an efficacious and safe dose for the treatment of social anxiety disorder, with superior efficacy to placebo.[72] In a flexible-dose trial, paroxetine proved more efficacious than placebo after 8–12 weeks of treatment, with 55 per cent of paroxetine patients compared to 24 per cent of placebo patients identified as much or very much improved.[73] A double-blind discontinuation study demonstrated a 14 per cent relapse rate in the paroxetine group compared to a 39 per cent relapse rate in the placebo group at week 24.[74]

Van Ameringen and colleagues[75] found sertraline to be more efficacious than placebo, and a 20-week double-blind discontinuation trial demonstrated significantly lower relapse rates for patients continuing on sertraline than patients switched to placebo (4 per cent versus 36 per cent, respectively).[76] A study of fluvoxamine demonstrated superior outcome, beginning at week 8, for patients receiving fluoxamine (20 mg mean daily dose) than patients receiving placebo.[77]

Fluoxetine has yielded less promising results and is the only SSRI to date that has failed to separate from placebo in controlled trials.[65,69] However, in one study fluoxetine evidenced superior efficacy to placebo, although no differences were found between fluoxetine, CBGT, or their combination.[78]

(b) Benzodiazepines

Research suggests benzodiazapines have performed better than MAOIs and CBT in the short-term[68] and offer rapid onset of effect, good tolerability, and minimal side effects (e.g. sedation, withdrawal effects with abrupt discontinuation). However, benzodiazepines are contraindicated in the presence of substance abuse and lack efficacy for comorbid mood disorders. Clonazepam was shown to be efficacious for social anxiety disorder (2–4 mg/day), with patients showing improvement after 6 weeks of treatment and superior response rates compared to placebo patients at week 10.[79] Clonazepam also produced equivalent gains to CBGT after 12 weeks of treatment.[80] One study of alprazolam demonstrated that the medication was no more efficacious than placebo.[81]

(c) Monoamine oxidase inhibitors

MAOIs have demonstrated efficacy in the treatment of social anxiety disorder[36] similar to that of SSRIs. However, MAOI therapy requires strict dietary restrictions as ingesting foods rich in tyramine (e.g. most aged cheeses, red wine, beer) can increase the risk of rapidly escalating blood pressure. MAOIs are also associated with more adverse side effects (e.g. hypotension, sedation, weight gain, sexual dysfunction) than SSRIs. However, MAOI therapy (usually 45–90 mg/day) is justified when other treatments with more benign side effect profiles prove ineffective.

As discussed earlier, phenelzine proved superior to placebo and demonstrated equivalent response rates to CBGT after 12 weeks of treatment,[57], although CBGT evidenced more durable treatment gains.[62] Liebowitz and colleagues[82] found that after 8 weeks of acute treatment and 8 weeks of maintenance, phenelzine produced results superior to those of placebo and the ß-adreneregic blocker atenolol. Versiani and colleagues[83] found patients who received phenelzine endorsed significantly greater reductions in social anxiety than placebo patients at 8 weeks. In a comparison of phenelzine, alprazolam, CBGT, and placebo with instructions for self-exposure,[81] all groups produced relatively equivalent response. However, phenelzine demonstrated maintenance of gains at a 2-month follow-up, whereas alprazolam did not.

(d) Reversible inhibitors of monoamine oxidase

Moclobemide and brofaramine have been evaluated for the treatment of social anxiety disorder and offer fewer side effects and dietary restrictions than MAOIs. However, moclobemide has shown modest efficacy in controlled trials and is less |efficacious than the SSRIs and MAOIs.[36] Although brofaramine was found to be efficacious for social anxiety, it was never marketed.

(e) ß-adreneregic blockers

ß-adreneregic blockers (e.g. atenolol) have failed to surpass placebo in controlled trials.[49] Propranolol appears to be efficacious on an as-needed basis (10–40 mg) for anxiety related to performance situations.

(f) Buspirone

Buspirone appears no more efficacious than placebo at low dosages.[84] One study demonstrated greater efficacy of buspirone in doses greater than or equal to 45 mg/day or when used as an augmentation to an SSRI.[85]

(g) Gabapentin and pregabalin

A single study shows some efficacy of high doses of the anticonvulsant gabapentin (maximum of 3600 mg/day) in the treatment of social anxiety disorder relative to placebo.[86] Pregabalin was also superior to placebo in one study,[87] but a high dose was required as well.

(h) Other agents

D-cycloserine, a partial adrenergic agonist associated with the facilitation of learning and memory consolidation, appears to enhance the efficacy of exposure. Several animal trials demonstrated enhanced fear extinction when d-cycloserine was administered prior to and after exposure trials.[88] Similar results were shown when individuals receiving d-cycloserine 1 h prior to exposure demonstrated significantly less social anxiety posttreatment than those receiving placebo.[89]

Integrating pharmacotherapy and cognitive behavioural interventions

Several studies have examined the utility of combining pharmacotherapy with cognitive behavioural interventions. Blomhoff and colleagues[90] compared sertraline and pill placebo alone and in combination with physician-directed exposure or general medical care, which included non-directive encouragement. After 12 weeks, all active treatments surpassed pill placebo and non-directive encouragement, with sertraline and exposure proving equally efficacious. Only sertraline alone or in combination with exposure proved superior to placebo at 24 weeks. However, only patients who received exposure alone demonstrated further improvement at 1-year follow-up, whereas patients receiving sertraline with or without exposure showed some degree of deterioration.[91]

Preliminary data from a study comparing CBGT, phenelzine, their combination, and pill placebo, suggests phenelzine plus CBGT may be more likely to surpass placebo than either treatment alone.[92] A similar study by Davidson and colleagues[78] found no advantage in combining fluoxetine and CBGT. Preliminary data suggests the combination of d-cycloserine with exposure may enhance the efficacy of exposure treatment.[89]

Evidence for the utility of the combination of these treatments is mixed. To summarize, different studies suggest that combining some medication treatments with CBT is no more efficacious than either treatment alone;[78] combining medication with CBT may diminish the efficacy of CBT;[91] and combining treatments may show only modest benefits over the administration of either alone.[92] The hypothesis that sequential treatments might capitalize on rapid medication response for faster symptom relief and cognitive behavioural treatment for superior relapse protection needs evaluation. Cognitive behavioural treatments may be utilized to help patients discontinue medical regimens after initial drug response.

Management of social anxiety disorder

(a) Treatment selection

Given the chronic nature of social anxiety disorder, CBT offers more enduring treatment gains compared to medication and lacks the adverse side effects. When considering cognitive behavioural interventions, it is important to assess the patient's ability and willingness to endure exposure to anxiety-provoking situations, as exposure necessitates short-term increases in anxiety over and above what the individual may typically experience. Additionally, it is helpful to assess level of compliance with previous therapy experiences (e.g. homework compliance). Managing a patient's treatment expectancy may further improve outcome as individuals reporting higher expectancy regarding the efficacy of treatment have demonstrated greater improvement and maintenance of treatment gains.[93]

Medication may be a preferred choice for individuals requiring rapid response, CBT non-responders, those with a preference for medication, or those with severe social anxiety disorder and/or comorbid depression. When selecting medication treatment, it is important to consider possible side effects, the patient's physical health, and prior medication compliance. Benzodiazepines offer rapid onset of effect, particularly beneficial for short-term performance situations, but their use may be contraindicated among patients with a history of alcohol or substance abuse. Given these concerns and their more benign side effect profile, SSRIs and venlafaxine are best regarded as the first-line medications, with the benzodiazepines and MAOIs held in reserve for non-responding patients. To date, we know little of the appropriate duration of medication treatments in social anxiety disorder, although two studies evidenced significant relapse rates upon discontinuation of SSRIs after a total treatment period of 36–40 weeks.[74,76]

(b) Other issues in management

Comorbid conditions should also be considered when treating social anxiety disorder with cognitive behavioural or pharmacological treatments. Comorbid anxiety disorders do not appear to degrade a patient's response to CBT; however, patients with comorbid mood disorders may present with more severe social anxiety symptoms and, although they improve at a similar rate, require extended treatment.[94] In general, if any comorbid condition prevents the patient from engaging in exposure exercises, taking medications in the prescribed manner, or complying with any other therapeutic activities, it may be necessary to treat the comorbid condition before the social anxiety disorder. In other cases, a comorbid condition might become the focus after social anxiety disorder treatment. Patients who abuse substances to 'treat' their social anxieties are often reticent to give up their substances before other coping strategies are made available. Ultimately, the management of comorbid conditions in individuals with social anxiety disorder depends on the analysis of the relationships between the disorders and varies from case to case.

(c) Prevention

Few studies to date have specifically examined ways of preventing social anxiety disorder; however, evidence for familial aggregation and environmental influence is strong. Parents may reinforce anxious children for making avoidant choices.[95] Thus, it is important to consider including parents of socially anxious children in treatment to provide parents with strategies to help their child manage the social anxiety. Parents with social anxiety disorder may also benefit from treatment themselves.

Since social anxiety disorder has an early age of onset, treatment of children and adolescents should help prevent social anxiety from becoming a chronic condition. Prevention programmes have been integrated into school settings to teach children and parents coping skills (e.g. cognitive restructuring, relaxation) and instruct them on the proper conduct of exposures. Individuals in the intervention group have shown significant improvement compared to non-intervention controls.[96] Similar prevention programmes have targeted at-risk youth and resulted in decreased anxiety symptoms and lower rates of anxiety disorders compared to control groups.[97]

Informational Websites

- http://www.adaa.org/GettingHelp/AnxietyDisorders/SocialPhobia.asp (Anxiety Disorders Association of America)
- http://www.nimh.nih.gov/HealthInformation/socialphobia-menu.cfm (National Institute of Mental Health)

Specific phobia

Clinical features and functional impairment

The hallmark feature of specific phobia, prior to DSM-IV called simple phobia, is a 'marked and persistent fear that is excessive or

unreasonable, cued by the presence or anticipation of a specific object or situation' (6, p. 405). Commonly feared/avoided objects include animals, aspects of nature, or blood. Many individuals endorse some fear of these stimuli; however, in specific phobia, fear and avoidance cause significant interference with one's normal routine, career, academic pursuits, or social activities. Some individuals with specific phobias maintain a relatively normal routine by pursuing a lifestyle that minimizes exposure to the phobic stimulus. Often, specific phobias accompany a more impairing primary disorder that is the stimulus for seeking treatment.[98]

Classification

(a) DSM and ICD

Criteria for specific phobia are similar between DSM-IV and ICD-10. Both view fear as arising from exposure to a specific object or situation, which leads to acute autonomic and psychological symptoms of anxiety. However, ICD-10 does not stipulate that anxiety may be cued by the anticipation of the feared object or situation. Our review of specific phobia follows DSM conventions.

(b) Specific phobia subtypes

DSM-IV classifies specific phobias into five subtypes:

- animal
- natural environment
- blood-injection-injury
- situational
- other (e.g. dental/medical procedures, choking, etc.).

With the exception of blood-injection-injury phobias, exposure to the phobic stimulus evokes intense anxiety that may meet criteria for a situationally-bound panic attack. Additionally, there is extreme apprehension and desire to escape or avoid the phobic stimulus.[99] By contrast, individuals with blood-injection-injury phobias exhibit a biphasic anxiety reaction (vasovagal syncope) characterized by initial, short-lived sympathetic arousal, followed by parasympathetic arousal that may result in fainting.[100] The subjective experiences of these individuals tend to be characterized by disgust and repulsion rather than pure apprehension.[99]

(c) Comorbidity

The vast majority (83.4 per cent) of individuals with specific phobia experience at least one other lifetime psychiatric disorder.[19] In the original National Comorbidity Survey, individuals with specific phobia were 5 times more likely to have at least one additional disorder than individuals who had never met criteria for a phobic disorder. In most cases, the onset of the specific phobia preceded the onset of the other disorder.

Epidemiology

(a) Prevalence

The NCS-R reported a lifetime prevalence rate of 12.5 per cent[25] and a 12-month prevalence rate of 8.7 per cent for specific phobia.[20] One study found situational/environmental phobias to be the most common (13.2 per cent), followed by animal phobias (7.9 per cent) and blood-injection-injury phobias (3.0 per cent).[101]

(b) Age at onset/age of treatment seeking

Age at onset of specific phobia tends to be earlier than other anxiety disorders[19,102] and varies as a function of phobia subtype. For example, animal phobias onset earliest (age 7), followed by blood-injection-injury phobias (age 8), and most situational phobias (early 20s).[103] Individuals with phobic disorders often do not seek treatment for 20 years after onset.[27]

(c) Gender distribution

Women receive diagnoses of specific phobia more often than men. Lifetime rates for specific phobia in the NCS-R were 15.7 per cent for women but only 6.7 per cent for men.[25] In one study, women reported higher rates of animal and situational/environmental phobias, but rates of blood-injection-injury phobia did not differ.[101]

Aetiology of specific phobia

There is considerable evidence for familial/genetic transmission of specific phobia.[32,104] In one study, 31 per cent of first-degree relatives of persons with specific phobia also met criteria for specific phobia.[104] Rates of specific phobia were higher among first-degree relatives of persons with specific phobia and no other anxiety disorder than among first-degree relatives of persons who were never mentally ill.[105] In the Virginia Twin Study,[32] concordance rates for animal phobia were 25.9 and 11.0 per cent among monozygotic and dizygotic twins, respectively. Concordance rates for situational phobia were similar in monozygotic and dizygotic twins. Moreover, children classified as behaviourally inhibited have shown higher risk for the development of multiple specific phobias.[106]

Classical conditioning theory[107] holds that phobias are learned through the association of negative experience with an object or situation. However, individuals with more previous non-traumatic experiences with the object or situation are less likely to develop a phobia upon a negative encounter than those with less prior experience when traumatized. Two-factor learning theory[108] introduced avoidance as a critical component to the maintenance of anxiety. That is, responses of avoidance or escape are learned and serve to decrease the discomfort arising from exposure to conditioned stimuli. Repeated negative reinforcement of avoidance behaviour (i.e. reduction of arousal on removal of oneself from proximity to the phobic object or situation) maintains the fear and makes it resistant to extinction.

Some conditioning theorists assert that feared stimuli are not randomly determined; rather, humans have inherited a predisposition to fear specific stimuli through natural selection. Marks'[109] 'preparedness' theory maintains that commonly feared objects are those that historically threatened the survival of the individual or the species. In this model, phobias are viewed as instances of 'prepared learning', which is selective, easily acquired, difficult to extinguish, and non-cognitive.[110] However, a large number of studies also suggest that phobias may be acquired via observational and informational learning (e.g. hearing that the situation is dangerous).

Course of specific phobia

Individuals with specific phobias acquire their fear(s) early in life, and the disorder tends to be chronic or recurrent without treatment.[111] Often individuals with specific phobias adapt their lifestyle to avoid contact with the feared stimuli, such that only

persons with the most severe specific phobias seek treatment. Events that commonly precipitate treatment seeking include a change in lifestyle that increases exposure to the feared stimulus (e.g. change in occupation that requires frequent air travel) or the experience of a panic attack in anticipation or in the presence of the feared stimulus.

Empirically evaluated treatments

(a) Cognitive behavioural interventions

(i) Exposure

Prolonged and repeated *in vivo* exposure to feared stimuli is by far the most studied and efficacious intervention for specific phobia[112] and should be considered the first-line treatment. Although *in vivo* exposure is generally believed to be more efficacious than imaginal exposure for specific phobia, some studies found *in vivo* and imaginal exposure techniques to be similarly efficacious.[113] Modelling in the form of observing another patient receive treatment has been shown to enhance the effects of exposure and to increase the speed at which positive outcomes are attained.[114] Multiple exposures sessions are considered more efficacious than one session, although some studies report positive outcomes for one-session treatments.[115] Further, one-session group *in vivo* exposure treatment has produced gains similar to those of one-session individual *in vivo* exposure treatment.[115] Variations in spacing of exposure sessions (e.g. 10 daily versus 10 weekly) have shown equivalent outcomes.[116] Finally, therapist-directed treatments have generally been more efficacious than self-directed treatments,[117] with gains enduring up to 8 years.[114] When possible, exposures should be conducted in a variety of settings to enhance generalization outside the therapeutic setting.[54]

In vivo exposure situations can sometimes be difficult to arrange and imaginal exposure may not achieve the reality or intensity needed to elicit an anxiety response. In such instances, virtual reality exposure (**VRE**), in which feared situations are presented in three dimensional simulations, may greatly enhance the efficacy of exposures. The salience of virtual environments can be augmented by instructing patients to touch real objects (e.g. toy spiders) which correspond with the virtual environment. Recent case studies demonstrated the efficacy of VRE either alone[118,119] or in combination with anxiety management training.[120] In a controlled trial for fear of flying, VRE training showed equivalent gains to *in vivo* exposure at 6 and 12-month follow-up[121]. Another study indicated that VRE produced superior gains to an attention placebo therapy, although gains were not maintained at 6-month follow-up.[122] D-cycloserine in combination with VRE has also shown enhanced treatment efficacy over VRE with placebo.[123]

(b) Applied relaxation and applied tension

Applied relaxation combines focused attention on different muscle groups while tensing and relaxing muscles,[124] with instruction to practice these skills first in non-anxiety-provoking situations and then in anxiety-provoking situations.[125] Although research is limited, applied relaxation has demonstrated efficacy,[126] especially among patients with higher levels of physiological reactivity than behavioural avoidance. *Applied tension*, designed specifically for blood-injection-injury phobia to treat parasympathetic arousal, requires the patient to tense different muscle groups in the presence of phobic stimuli to elevate blood pressure. Persons with phobias for blood, wounds, and injuries responded equally well to applied tension, applied relaxation, or their combination[127] Individuals treated with applied tension also evidenced greater treatment gains posttreatment and at 1-year follow-up than those treated with *in vivo* exposure alone.[128] In one dismantling study, individuals treated with applied tension and tension only (without the exposure component) evidenced equivalent gains and maintained superior outcomes at posttreatment and 1-year follow-up compared to patients who received *in vivo* exposure.[129]

(c) Cognitive restructuring

Phobia-specific irrational thoughts may contribute to the development of the phobia, maintain avoidance behaviour, and contribute to physiological symptoms.[130] When combined with exposure to feared stimuli, cognitive restructuring has proven efficacious,[131] although there are relatively few studies of cognitively-oriented treatments for specific phobia.

Pharmacotherapy

Drug treatments for specific phobia have consistently been shown to be less efficacious than behavioural treatments and may hinder maintenance of treatment gains. ß-Adrenergic blockers reduce some symptoms of sympathetic arousal during exposure to feared stimuli but fail to decrease subjective fear.[132] Although benzodiazepines (e.g. diazepam) may facilitate approach to feared stimuli, they may also reduce the efficacy of behaviour therapies by inhibiting the experience of anxiety during exposure.[133] Recent studies suggest d-cycloserine may facilitate exposure and extinction to specific feared stimuli. In a study of acrophobia patients, d-cycloserine was administered prior to exposure sessions and proved superior to placebo in reducing anxiety symptoms.[123]

Management of specific phobia

(a) Treatment selection

Exposure is clearly the treatment of choice for specific phobia, although facing feared stimuli may be particularly challenging for some patients, and their willingness and ability to participate in exposures should be assessed prior to treatment. Tailoring treatment to individual response patterns may improve outcome. For instance, patients with heightened physiological reactivity may respond preferentially to applied relaxation, whereas patients showing avoidance behaviour may respond better to *in vivo* exposure.[126] Further, individuals who experience anxiety primarily in the form of anxious thoughts may respond better to cognitive techniques.[134]

For patients unwilling or unable to engage in cognitive behavioural interventions, medication may be an appropriate alternative. However, it is first important to educate patients about the possibility of dependence with regular use and the side effects of high doses (e.g. sedation). If the patient is participating in exposure therapy, it is also necessary to explain that medication may interfere with the efficacy of exposure treatment.

(b) Prevention

Children with specific phobia were included in the Queensland Early Intervention and Prevention of Anxiety Project, but no other preventive efforts have been mounted. It is tempting to speculate that children could be 'inoculated' against a variety of the more common specific phobias by systematic exposure to potentially feared objects or situations.

Further information

http://www.adaa.org/GettingHelp/AnxietyDisorders/SpecificPhobia.asp (Anxiety Disorders Association of America)
http://www.mentalhealthamerica.net/go/phobias (Mental Health America)

References

1. American Psychiatric Association. (1952). *Diagnostic and statistical manual of mental disorders* (1st edn). American Psychiatric Press, Washington, DC.
2. American Psychiatric Association. (1968). *Diagnostic and statistical manual of mental disorders* (2nd edn). American Psychiatric Press, Washington, DC.
3. Marks, I.M. and Gelder, M.G. (1966). Different ages of onset in varieties of social phobia. *The American Journal of Psychiatry*, **123**, 218–21.
4. American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd edn). American Psychiatric Press, Washington, DC.
5. Liebowitz, M.R., Gorman, J.M., Fyer, A.J., *et al.* (1985). Social phobia: review of a neglected anxiety disorder. *Archives of General Psychiatry*, **42**, 729–36.
6. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Press, Washington, DC.
7. Rapee, R.M. and Heimberg, R.G. (1997). A cognitive-behavioral model of anxiety in social phobia. *Behaviour Research and Therapy*, **35**, 741–56.
8. Clark, D.M. and Wells, A. (1995). A cognitive model of social phobia. In *Social phobia: diagnosis, assessment, and treatment* (eds. R.G. Heimberg, M.R. Liebowitz, D.A. Hope, and F.R. Schneier), pp. 69–93. Guilford Press, New York.
9. Schneier, F.R., Heckelman, L.R., Garfinkel, R., *et al.* (1994). Functional impairment in social phobia. *The Journal of Clinical Psychiatry*, **55**, 322–31.
10. Wittchen, H., Fuetsch, M., Sonntag, H., *et al.* (1999). Disability and quality of life in pure and comorbid social phobia: findings from a controlled study. *European Psychiatry*, **14**, 118–31.
11. Whisman, M., Sheldon, C., and Goering, P. (2000). Psychiatric disorders and dissatisfaction with social relationships: does type of relationship matter? *Journal of Abnormal Psychology*, **109**, 803–8.
12. Stein, M., McQuaid, J., Laffaye, C., *et al.* (1999). Social phobia in the primary medical care setting. *Journal of Family Practice*, **49**, 514–9.
13. World Health Organization. (1992). The ICD-10 classification of mental and behavioural disorders: clinical description and diagnostic guidelines. World Health Organization, Geneva.
14. Heimberg, R.G., Hope, D.A., Dodge, C.S., *et al.* (1990). DSM-III-R subtypes of social phobia: comparison of generalized social phobics and public speaking phobics. *The Journal of Nervous and Mental Disease*, **173**, 172–9.
15. Heimberg, R.G., Holt, C.S., Schneier, F.R., *et al.* (1993). The issue of subtypes in the diagnosis of social phobia. *Journal of Anxiety Disorders*, **7**, 249–69.
16. Vriends, N., Becker, E.S., Meyer, A., *et al.* (2007). Subtypes of social phobia: are they of any use? *Journal of Anxiety Disorders*, **21**, 59–75.
17. Heimberg, R.G. (1996). Social phobia, avoidant personality disorder and the multiaxial conceptualization of interpersonal anxiety. In *Trends in cognitive and behavioural therapies*, Vol. 1 (ed. P. Salkovskis), pp. 43–61. John Wiley & Sons, Chichester, England.
18. Schneier, F.R., Liebowitz, M.R., Beidel, D.C., *et al.* (1998). MacArthur data reanalysis for *DSM-IV*: social phobia. In *DSM-IV sourcebook*, Vol. 4 (eds. T.A. Widiger, A.H. Frances, H.A. Pincus, R. Ross, M.J. First, W. Davis, and M. Kline), pp. 307–28. American Psychiatric Press, Washington, DC.
19. Magee, W.J., Eaton, W.W., Wittchen, H.U., *et al.* (1996). Agoraphobia, simple phobia, and social phobia in the National Comorbidity Survey. *Archives of General Psychiatry*, **53**, 159–68.
20. Kessler, R.C., Chiu, W.T., Demler, O., *et al.* (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **62**, 617–27.
21. Heckelman, L.R. and Schneier, F.R. (1995). Diagnostic issues. In *Social phobia: diagnosis, assessment, and treatment* (eds. R.G. Heimberg, M.R. Liebowitz, D.A. Hope, and F.R. Schneier), pp. 3–20. Guilford Press, New York.
22. Mannuzza, S., Fyer, A.J., Liebowitz, M.R., *et al.* (1990). Delineating the boundary of social phobia: its relationship to panic disorder and agoraphobia. *Journal of Anxiety Disorders*, **4**, 41–59.
23. Rapee, R.M., Sanderson, W.C., and Barlow, D.H. (1988). Social phobia features across the DSM-III-R anxiety disorders. *Journal of Psychopathology and Behavioral Assessment*, **10**, 287–99.
24. Reich, J., Noyes, R., and Yates, W. (1988). Anxiety symptoms distinguishing social phobia from panic and generalized anxiety disorders. *The Journal of Nervous and Mental Disease*, **176**, 510–3.
25. Kessler, R.C., Berglund, P.D., Demler, O., *et al.* (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **62**, 593–602.
26. Hazen, A.L. and Stein, M.B. (1995). Clinical phenomenology and comorbidity. In *Social phobia: clinical and research perspectives* (ed. M.B. Stein), pp. 3–41. American Psychiatric Press, Washington, DC.
27. Wang, P.S., Berglund, P., Olfson, M., *et al.* (2005). Failure and delay in initial treatment contact after first onset of mental disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **62**, 603–13.
28. Olfson, M., Guardino, M., Struening, E., *et al.* (2000). Barriers to treatment of social anxiety. *The American Journal of Psychiatry*, **157**, 521–7.
29. Turk, C.L., Heimberg, R.G., Orsillo, S.M., *et al.* (1998). An investigation of gender differences in social phobia. *Journal of Anxiety Disorders*, **12**, 209–23.
30. Tillfors, M. (2004). Why do some individuals develop social phobia? A review with emphasis on the neurobiological influences. *Nordic Journal of Psychiatry*, **58**, 267–76.
31. Stein, M.B., Chartier, M.J., Hazen, A.L., *et al.* (1998). A direct-interview family study of generalized social phobia. *The American Journal of Psychiatry*, **155**, 90–7.
32. Kendler, K.S., Neale, M.C., Kessler, R.C., *et al.* (1992). The genetic epidemiology of phobias in women: the interrelationship of agoraphobia, social phobia, situational phobia, and simple phobia. *Archives of General Psychiatry*, **49**, 273–81.
33. Kendler, K.S., Karkowski, L.M., and Prescott, C.A. (1999). Fears and phobias: reliability and heritability. *Psychological Medicine*, **29**, 539–53.
34. Kendler, K.S., Myers, J., Prescott, C.A., *et al.* (2001). The genetic epidemiology of irrational fears and phobias in men. *Archives of General Psychiatry*, **58**, 257–65.
35. Stein, M.B., Goldin, P.R., Sareen, J., *et al.* (2002). Increased amygdala activation to angry and contemptuous faces in generalized social phobia. *Archives of General Psychiatry*, **59**, 1027–34.
36. Blanco, C., Schneier, F.R., Schmidt, A., *et al.* (2003). Pharmacological treatment of social anxiety disorder: a meta-analysis. *Depression and Anxiety*, **18**, 29–40.
37. Stein, D.J. and Stahl, S. (2000). Serotonin and anxiety: current models. *International Clinical Psychopharmacology*, **15**(Suppl. 2), S1–6.
38. Schneier, F.R., Liebowitz, M.R., Abi-Dargham, A., *et al.* (2000). Low dopamine D2 receptor binding potential in social phobia. *The American Journal of Psychiatry*, **157**, 457–9.
39. Bruch, M.A., Heimberg, R.G., Berger, P., *et al.* (1989). Social phobia and perceptions of early parental and personal characteristics. *Anxiety Research*, **2**, 57–65.

40. Lieb, R., Wittchen, H.U., Hofler, M., *et al.* (2000). Parental psychopathology, parenting styles, and the risk of social phobia in offspring. *Archives of General Psychiatry*, **57**, 859–66.
41. Erwin, B.A., Heimberg, R.G., Marx, B.P., *et al.* (2006). Traumatic and socially stressful life events among persons with social anxiety disorder. *Journal of Anxiety Disorders*, **20**, 896–914.
42. Reich, J., Goldenberg, I., Vasile, R., *et al.* (1994). A prospective follow-along study of the course of social phobia. *Psychiatry Research*, **54**, 249–58.
43. Reich, J., Goldenberg, I., Goisman, R., *et al.* (1994). A prospective, follow-along study of the course of social phobia: II. testing for basic predictors of course. *The Journal of Nervous and Mental Disease*, **182**, 297–301.
44. Caspi, A., Edler, G.H., and Bem, D.J. (1988). Moving away from the world: life-course patterns of shy children. *Developmental Psychology*, **24**, 824–31.
45. Schwartz, C.E., Snidman, N., and Kagan, J. (1999). Adolescent social anxiety as an outcome of inhibited temperament in childhood. *Journal of the American Academy of Child and Adolescent Psychiatry*, **38**, 1008–15.
46. Rosenbaum, J.F., Biederman, J., Hirshfeld, D.R., *et al.* (1991). Behavioral inhibition in childhood: a possible precursor to panic disorder or social phobia. *The Journal of Clinical Psychiatry*, **52**(Suppl.), 5–9.
47. Rodebaugh, T.L., Holaway, R.M., and Heimberg, R.G. (2004). The treatment of social anxiety disorder. *Clinical Psychology Review*, **24**, 883–908.
48. Al-Kubaisy, T., Marks, I.M., Logsdail, S., *et al.* (1992). Role of exposure homework in phobia reduction: a controlled study. *Behavior Therapy*, **23**, 599–621.
49. Turner, S.M., Beidel, D.C., and Jacob, R.G, (1994). Social phobia: a comparison of behavior therapy and atenolol. *Journal of Consulting and Clinical Psychology*, **62**, 350–8.
50. Newman, M.G., Hofmann, S.G., Trabert, W., *et al.* (1994). Does behavioral treatment of social phobia lead to cognitive changes? *Behavior Therapy*, **25**, 503–17.
51. Heimberg, R.G. and Becker, R.E. (2002). Cognitive-behavioral group therapy for social phobia: basic mechanisms and clinical strategies. Guilford Press, New York.
52. Clark, D.M. (2001). A cognitive perspective on social phobia. In *International handbook of social anxiety: concepts, research and interventions relating to the self and shyness* (eds. W.R. Crozier and L.E. Alden LE), pp. 405–30. John Wiley & Sons, Chichester, United Kingdom.
53. Beck, A.T. and Emery, G. (1985). *Anxiety disorders and phobias: a cognitive perspective*. Basic Books, New York.
54. Bouton, M.E. (2002). Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biological Psychiatry*, **52**, 976–86.
55. Hope, D.A., Heimberg, R.G., and Bruch, M.A. (1995). Dismantling cognitive-behavioral group therapy for social phobia. *Behaviour Research and Therapy*, **33**, 637–50.
56. Butler, G., Cullington, A., Munby, M., *et al.* (1984). Exposure and anxiety management in the treatment of social phobia. *Journal of Consulting and Clinical Psychology*, **52**, 642–50.
57. Heimberg, R.G., Liebowitz, M.R., Hope, D.A., *et al.* (1998). Cognitive-behavioral group therapy versus phenelzine in social phobia: 12-week outcome. *Archives of General Psychiatry*, **55**, 1133–41.
58. Heimberg, R.G., Dodge, C.S., Hope, D.A., *et al.* (1990). Cognitive behavioral group treatment for social phobia: comparison with a credible placebo control. *Cognitive Therapy and Research*, **14**, 1–23.
59. Taylor, S. (1996). Meta-analysis of cognitive-behavioral treatments for social phobia. *Journal of Behavior Therapy and Experimental Psychiatry*, **27**, 1–9.
60. Gould, R.A., Buckminster, S., Pollack, M.H., *et al.* (1997). Cognitive-behavioral and pharmacological treatment for social phobia: a meta-analysis. *Clinical Psychology: Science and Practice*, **4**, 291–306.
61. Heimberg, R.G., Salzman, D.G., Holt, C.S., *et al.* (1993). Cognitive-behavioral group treatment for social phobia: effectiveness at five-year followup. *Cognitive Therapy and Research*, **17**, 325–39.
62. Liebowitz, M.R., Heimberg, R.G., Schneier, F.R., *et al.* (1999). Cognitive-behavioral group therapy versus phenelzine in social phobia: long-term outcome. *Depression and Anxiety*, **10**, 89–98.
63. Mörtberg, E., Karlsson, A., Fyring, C., *et al.* (2006). Intensive cognitive-behavioral group treatment (CBGT) of social phobia: a randomized controlled study. *Journal of Anxiety Disorders*, **20**, 646–60.
64. Clark, D.M., Ehlers, A., Hackmann, A., *et al.* (2006). Cognitive therapy versus exposure and applied relaxation in social phobia: a randomized controlled trial. *Journal of Consulting and Clinical Psychology*, **74**, 568–78.
65. Clark, D.M., Ehlers, A., McManus, F., *et al.* (2003). Cognitive therapy vs fluoxetine in generalized social phobia: a randomized placebo controlled trial. *Journal of Consulting and Clinical Psychology*, **71**, 1058–67.
66. Stangier, U., Heidenreich, T., Peitz, M., *et al.* (2003). Cognitive therapy for social phobia: individual versus group treatment. *Behaviour Research and Therapy*, **41**, 991–1007.
67. Mörtberg, E., Clark, D.M., Sundin, Ö., *et al.* (2007). Intensive group cognitive treatment and individual cognitive therapy vs treatment as usual in social phobia: a randomized controlled trial. *Acta Psychiatrica Scandinavica*, **115**, 142–54.
68. Fedoroff, I.C. and Taylor, S. (2001). Psychological and pharmacological treatments for social anxiety disorder: a meta-analysis. *Journal of Clinical Psychopharmacology*, **21**, 311–24.
69. Kobak, K.A., Greist, J.H., Jefferson, J.W., *et al.* (2002). Fluoxetine in social phobia: a double-blind, placebo-controlled pilot study. *Journal of Clinical Psychopharmacology*, **22**, 257–62.
70. Lader, M., Stender, K., Burger, V., *et al.* (2004). The efficacy and tolerability of escitalopram in the short-and long-term treatment of social anxiety disorder: a randomised, double blind, placebo-controlled, fixed-dose study. *Depression and Anxiety*, **19**, 241–48.
71. Liebowitz, M.R., Gelenberg, A.J., and Munjack, D, (2005). Venlafaxine extended release vs placebo and paroxetine in social anxiety disorder. *Archives of General Psychiatry*, **62**, 190–8.
72. Liebowitz, M.R., Stein, M.B., Tancer, M., *et al.* (2002). A randomized, double-blind, fixed-dose comparison of paroxetine and placebo in the treatment of generalized social anxiety disorder. *The Journal of Clinical Psychiatry*, **63**, 66–74.
73. Stein, M.B., Liebowitz, M.R., Lydiard, R.B., *et al.* (1998). Paroxetine treatment of generalized social phobia (social anxiety disorder): a randomized controlled trial. *The Journal of the American Medical Association*, **280**, 708–13.
74. Stein, D.J., Versiani, M., Hair, T., *et al.* (2002). Efficacy of paroxetine for relapse prevention in social anxiety disorder: a 24-week study. *Archives of General Psychiatry*, **59**, 1111–18.
75. Van Ameringen, M.A., Lane, K.M., Walker, J.R., *et al.* (2001). Sertraline treatment for generalized social phobia: a 20-week, double-blind, placebo controlled study. *The American Journal of Psychiatry*, **158**, 275–81.
76. Walker, J.R., van Ameringen, M.A., Swinson, R., *et al.* (2000). Prevention of relapse in generalized social phobia: results of a 24-week study in responders to 20 weeks of sertraline treatment. *Journal of Clinical Psychopharmacology*, **20**, 636–44.
77. Stein, M.B., Fyer, A.J., Davidson, J.R.T., *et al.* (1999). Fluvoxamine treatment of social phobia (social anxiety disorder): a double-blind, placebo-controlled study. *The American Journal of Psychiatry*, **156**, 756–60.
78. Davidson, J.R., Foa, E.B., Huppert, J.D., *et al.* (2004). Fluoxetine, comprehensive cognitive behavioral therapy, and placebo in generalized social phobia. *Archives of General Psychiatry*, **61**, 1005–13.
79. Davidson, J.R.T., Potts, N., Richichi, E., *et al.* (1993). Treatment of social phobia with clonazepam and placebo. *Journal of Clinical Psychopharmacology*, **13**, 423–8.

80. Otto, M.W., Pollack, M.H., Gould, R.A., *et al.* (2000). A comparison of the efficacy of clonazepam and cognitive-behavioral group therapy for the treatment of social phobia. *Journal of Anxiety Disorders*, **14**, 345–58.
81. Gelernter, C.S., Uhde, T.W., Cimbolic, P., *et al.* (1991). Cognitive-behavioral and pharmacological treatments of social phobia: a controlled study. *Archives of General Psychiatry*, **48**, 938–45.
82. Liebowitz, M.R., Schneier, F.R., Campeas, R., *et al.* (1992). Phenelzine vs atenolol in social phobia: a placebo-controlled comparison. *Archives of General Psychiatry*, **49**, 290–300.
83. Versiani, M., Nardi, A.E., Mundim, F.D., *et al.* (1992). Pharmacotherapy of social phobia: a controlled study with moclobemide and phenelzine. *The British Journal of Psychiatry*, **161**, 353–60.
84. Clark, D.B. and Agras, W.S. (1991). The assessment and treatment of performance anxiety in musicians. *The American Journal of Psychiatry*, **148**, 598–605.
85. Schneier, F.R., Saoud, J., Campeas, R., *et al.* (1993). Buspirone in social phobia. *Journal of Clinical Psychopharmacology*, **13**, 251–6.
86. Pande, A.C., Davidson, J.R.T., Jefferson, J.W., *et al.* (1999). Treatment of social phobia with gabapentin: a placebo-controlled study. *Journal of Clinical Psychopharmacology*, **19**, 341–8.
87. Feltner, D.E., Pollack, M.H., Davidson, J.R.T., *et al.* (2000). A placebo-controlled, double-blind study of pregabalin treatment of social phobia: outcome and predictors of response. *European Neuropsychopharmacology*, **10**, 345–55.
88. Walker, D.L., Ressler, K.J., Lu, K.T., *et al.* (2002). Facilitation of conditioned fear extinction by systemic administration or intra-amygdala infusions of d-cycloserine as assessed with fear-potentiated startle in rats. *The Journal of Neuroscience*, **22**, 2343–51.
89. Hofmann, S.G., Meuret, A.E., Smits, J.A.J., *et al.* (2006). Augmentation of exposure therapy with d-cycloserine for social anxiety disorder. *Archives of General Psychiatry*, **63**, 298–304.
90. Blomhoff, S., Haug, T.T., Hellstrøm, K., *et al.* (2001). Randomized controlled general practice trial of setraline, exposure therapy, and combined treatment in generalized social phobia. *The British Journal of Psychiatry*, **179**, 23–30.
91. Haug, T.T., Blomhoff, S., Hellstrøm, K., *et al.* (2003). Exposure therapy and sertraline in social phobia: 1-year follow-up of a randomized controlled trial. *The British Journal of Psychiatry*, **182**, 312–8.
92. Heimberg, R.G. (2002). Cognitive behavioral therapy for social anxiety disorder: current status and future directions. *Biological Psychiatry*, **51**, 101–8.
93. Safren, S.A., Heimberg, R.G., and Juster, H.R. (1997). Clients' expectancies and their relationship to pretreatment symptomatology and outcome of cognitive-behavioral group treatment for social phobia. *Journal of Consulting and Clinical Psychology*, **65**, 694–8.
94. Erwin, B.A., Heimberg, R.G., Juster, H., *et al.* (2002). Comorbid anxiety and mood disorders among persons with social anxiety disorder. *Behaviour Research and Therapy*, **40**, 19–35.
95. Barrett, P.M., Rapee, R.M., Dadds, M.R., *et al.* (1996). Family enhancement of cognitive style in anxious and aggressive children. *Journal of Abnormal Child Psychology*, **24**, 187–203.
96. Barrett, P. and Turner, C. (2001). Prevention of anxiety symptoms in primary school children: preliminary results from a universal school-based trial. *The British Journal of Clinical Psychology*, **40**, 399–410.
97. Dadds, M.R., Holland, D.E., Laurens, K.R., *et al.* (1999). Early intervention and prevention of anxiety disorders in children: results at 2-year follow-up. *Journal of Consulting and Clinical Psychology*, **67**,145–50.
98. Barlow, D.H., DiNardo, P.A., Vermilyea, B.B., *et al.* (1986). Co-morbidity and depression among the anxiety disorders: issues in diagnosis and classification. *The Journal of Nervous and Mental Disease*, **174**, 63–72.
99. Merckelbach, H., de Jong, P.J., Muris, P., *et al.* (1996). The etiology of specific phobias: a review. *Clinical Psychology Review*, **16**, 337–61.
100. Thyer, B.A., Himle, J., and Curtis, G.C. (1985). Blood-injury-illness phobia: a review. *Journal of Clinical Psychology*, **41**, 451–9.
101. Fredrikson, M., Annas, P., Fischer, H., *et al.* (1996). Gender and age differences in the prevalence of specific fears and phobias. *Behaviour Research and Therapy*, **34**, 33–9.
102. Scheibe, G. and Albus, M. (1992). Age at onset, precipitating events, sex distribution, and co-occurrence of anxiety disorders. *Psychopathology*, **25**, 11–8.
103. Öst, L.G. and Treffers, P.D.A. (2001). Onset, course, and outcome for anxiety disorders in children. In *Anxiety disorders in children and adolescents: research, assessment and intervention* (eds. W.K. Silverman and P.D.A. Treffers), pp. 293–312. Cambridge University Press, Cambridge.
104. Fyer, A.J., Mannuzza, S., Chapman, T.F., *et al.* (1995). Specificity in familial aggregation of phobic disorders. *Archives of General Psychiatry*, **52**, 564–73.
105. Fyer, A.J., Mannuzza, S., Gallops, M.S., *et al.* (1990). Familial transmission of simple phobias and fears: a preliminary report. *Archives of General Psychiatry*, **47**, 252–6.
106. Biederman, J., Rosenbaum, J.F., Hirshfeld, D.R., *et al.* (1990). Psychiatric correlates of behavioral inhibition in young children of parents with and without psychiatric disorders. *Archives of General Psychiatry*, **47**, 21–6.
107. Pavlov, I. (1927). *Conditioned reflexes*. Oxford University Press, London.
108. Mowrer, O.H. (1939). Stimulus response theory of anxiety. *Psychological Review*, **46**, 553–65.
109. Marks, I.M. (1969). *Fears and phobias*. Academic Press, New York.
110. Seligman, M.E. (1971). Phobias and preparedness. *Behavior Therapy*, **2**, 307–20.
111. Boyd, J.H., Rae, D.S., Thompson, J.W., *et al.* (1990). Phobia: prevalence and risk factors. *Social Psychiatry and Psychiatric Epidemiology*, **25**, 314–23.
112. Marks, I.M. (1987). Fears, phobias, and rituals: panic, anxiety, and their disorders. Oxford University Press, London.
113. Hecker, J.E. (1990). Emotional processing in the treatment of simple phobia: a comparison of imaginal and in vivo exposure. *Behavioural Psychotherapy*, **18**, 21–34.
114. Gotestam, K.G. and Berntzen, D. (1997). Use of the modeling effect in one-session exposure. *Scandinavian Journal of Behaviour Therapy*, **26**, 97–101.
115. Öst, L.G. (1996). One-session group treatment of spider phobia. *Behaviour Research and Therapy*, **34**, 707–15.
116. Chambless, D.L. (1990). Spacing of exposure sessions in treatment of agoraphobia and simple phobia. *Behavior Therapy*, **21**, 217–29.
117. Öst, L.G., Salkovskis, P.M., and Hellstrom, K. (1991). One-session therapist-directed exposure vs self-exposure in the treatment of spider phobia. *Behavior Therapy*, **22**, 407–22.
118. Carlin, A.S., Hoffman, H.G., and Weghorst, S. (1997). Virtual reality and tactile augmentation in the treatment of spider phobia: a case report. *Behaviour Research and Therapy*, **35**, 153–8.
119. North, M.M., North, S.M., and Coble, J.R. (1997). Virtual reality therapy for fear of flying. *The American Journal of Psychiatry*, **154**, 130.
120. Rothbaum, B.O., Hodges, L., Watson, B.A., *et al.* (1996). Virtual reality exposure therapy in the treatment of fear of flying: a case report. *Behaviour Research and Therapy*, **34**, 477–81.
121. Rothbaum, B.O., Hodges, L., Anderson, P.L., *et al.* (2002). Twelve-month follow-up of virtual reality and standard exposure therapies for the fear of flying. *Journal of Consulting and Clinical Psychology*, **70**, 428–32.
122. Maltby, N., Kirsch, I., Mayers, M., *et al.* (2002). Virtual reality exposure therapy for the treatment of fear of flying: a controlled investigation. *Journal of Consulting and Clinical Psychology*, **70**, 1112–8.
123. Ressler, K.J., Rothbaum, B.O., Tannenbaum, L., *et al.* (2004). Cognitive enhancers as adjuncts to psychotherapy: use of d-cycloserine in phobic individuals to facilitate extinction of fear. *Archives of General Psychiatry*, **61**, 1136–44.

124. Bernstein, D.A., Borkovec, T.D., and Hazlett-Stevens, H. (2000). *New directions in progressive relaxation training: a guidebook for helping professionals*. Prager/Greenwood, Westport, CT.
125. Öst, L.G. (1987). Applied relaxation: description of a coping technique and review of controlled studies. *Behaviour Research and Therapy*, **25**, 397–409.
126. Öst, L.G., Johansson, J., and Jerremalm, A. (1982). Individual response patterns and the effects of different behavioral methods in the treatment of claustrophobia. *Behaviour Research and Therapy*, **20**, 445–60.
127. Öst, L.G., Sterner, U., and Fellenius, J. (1989). Applied tension, applied relaxation, and the combination in the treatment of blood phobia. *Behaviour Research and Therapy*, **27**, 109–21.
128. Öst, L.G., Fellenius, J., and Sterner, U. (1991). Applied tension, exposure *in vivo*, and tension-only in the treatment of blood phobia. *Behaviour Research and Therapy*, **29**, 561–74.
129. Hellstrøm, K., Fellenius, J., and Öst. L.G. (1996). One versus five sessions of applied tension in the treatment of blood phobia. *Behaviour Research and Therapy*, **34**, 101–12.
130. Thorpe, S.J. and Salkovskis, P.M. (1995). Phobia beliefs: do cognitive factors play a role in specific phobias? *Behavior Research and Therapy*, **33**, 805–16.
131. Greco, T.S. (1989). A cognitive-behavioral approach to fear of flying: a practitioner's guide. *Phobia Practice and Research Journal*, **2**, 3–15.
132. Campos, P.E., Solyom, L., and Koelink, A. (1984). The effects of timolol maleate on subjective and physiological components of air travel phobia. *Canadian Journal of Psychiatry*, **29**, 570–4.
133. Sartory, G. (1983). Benzodiazepines and behavioral treatment of phobic anxiety. *Behavioural Psychotherapy*, **11**, 204–17.
134. Norton, G.R. and Johnson, W.E. (1983). A comparison of two relaxation procedures for reducing cognitive and somatic anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, **14**, 209–14.

4.7.3 Panic disorder and agoraphobia

James C. Ballenger

Introduction

Panic disorder draws its name from the Greek god Pan, god of flocks. Pan was known for suddenly frightening animals and humans 'out of the blue'. The spontaneous 'out of the blue' character of panic attacks is the principal identifying characteristic of panic disorder and central to its recognition and diagnosis.

We know the syndrome that we currently call panic disorder with and without agoraphobia has probably existed since the beginning of recorded history. Hippocrates presented cases of obvious phobic avoidance around 400 BC.[1] One of the first modern descriptions was by Benedikt around 1870, describing individuals who developed sudden anxiety and dizziness in public places.[2]

Certainly, our current modern ideas of panic disorder evolved essentially simultaneously in the United States and Europe in the early to mid-1960s. Donald Klein in the United States described in 1964 the panic syndrome and reported that it was responsive to imipramine.[3] Isaac Marks in the United Kingdom also described panic attacks and agoraphobic avoidance, and treating the syndrome effectively with behaviour therapy.[4]

Until the last several decades, panic disorder and agoraphobia were actually thought to be rare syndromes. It is now clear that individuals with these difficulties are anything but rare. In fact, panic disorder is one of the most common presenting problems in individuals seeking mental health treatment and the fifth most common problem seen in primary care settings.[5] It was thought to be a mild problem, but we now know that it is associated with significant dysfunction. The disability in social, occupational, and family life is in fact comparable to major depression.

Although there are differences in the understandings of panic disorder and its treatments across the world, this chapter will review the current understanding about panic disorder, its characteristics, diagnosis, aetiology, and treatments.

Clinical features

One of the earliest and most accurate descriptions of panic attacks was provided by Charles Darwin in 1872 as he described his own episodes:

> The heartbeats quickly and violently so that it palpitates and knocks against the ribs . . . the skin instantly becomes pale as during incipient faintness . . . under a sense of great fear . . . in connection with the disturbed action of the heart, the breathing is hurried . . . one of the best marked symptoms is the trembling of all the muscles of the body.[6]

The most characteristic type of panic attack is the spontaneous 'out of the blue' episode of extreme anxiety. Other 'situational panic attacks' occur immediately upon exposure, or in anticipation of exposure to particular situations, usually where panic attacks have occurred previously. Some individuals have panic attacks in certain situations some of the time, but not always, and these are labelled 'situationally predisposed panic attacks'.

Panic attacks also occur in other anxiety syndromes and are more or less the same in whatever syndrome where they occur. However, spontaneous panic attacks in panic disorder tend to have more dizziness, paraesthesia, shaking, chest pain, and fears of going mad. Shortness of breath is more common in panic attacks in agoraphobia. Certainly blushing is particularly characteristic of panic attacks in social phobia.

The symptoms of panic attacks in order of their frequency include palpitations, pounding heart, tachycardia, sweating, trembling or shaking, shortness of breath or smothering, feeling of choking, chest pain or discomfort, nausea or abdominal distress, feeling dizzy, unsteady, lightheaded or faint, derealization or depersonalization, fear of losing control or going mad, fear of dying, paraesthesia, and chills or hot flushes.

Panic attacks by definition generally involve four or more of the above symptoms to meet diagnostic criteria for panic disorder in the DSM-IV. The anxiety is crescendo in nature, building to a peak in 10 min in most cases. Panic attacks usually last for several minutes, but in some patients they can last for hours. The frequency and severity of panic attacks varies greatly between individuals and, at times, in individuals. Some have only one to three panic attacks per year, whereas others may have multiple panic attacks each day. Some individuals have bursts of panic attacks and then an absence of all attacks for a period of time. Across a large panic disorder clinical trial, the typical patient described one to two panic attacks per week.[7]

Panic attacks with fewer than four symptoms have been labelled 'limited-symptom attacks' or 'little panic attacks', and most individuals with panic disorder have these, as well as panic attacks with four or more symptoms. The threshold of four symptoms was chosen for DSM-IV because individuals with panic attacks with four symptoms or more had more disability than patients with one- to three-symptom attacks. This threshold is clearly arbitrary, and patients having panic attacks with fewer symptoms do have significant morbidity.

Panic attacks are extremely frightening and patients develop an essentially logical fear of having more panic attacks. Patients develop worry and anxiety about the possibility of panic attacks recurring. This anxiety between attacks has been called 'anticipatory anxiety' and can be almost constant, and characteristically increases prior to exposure to situations previously associated with panic attacks (e.g. having to shop in the supermarket where a panic attack has occurred).

A significant number of people with panic attacks go on to develop fear and avoidance of situations associated with previous panic attacks. They also fear situations where escape would be difficult or embarrassing, or where help might not be available. Most patients mistakenly believe they become incapacitated and incapable of taking care of themselves during a panic attack and therefore, many go on to develop avoidance of a variety of situations where they could not easily get help. Factor analytical studies demonstrate that there are clusters of situations associated with avoidance. These typically include public transportation (e.g. buses, trains, planes), riding in or driving a car, especially on heavily travelled roads, crowds (e.g. the cinema, a football match, large shopping centres), shopping (especially in supermarkets), particularly where one must stand in queues, and bridges, tunnels, elevators, and other enclosed spaces.[8] In the event of a panic attack, people often have an overwhelming need to escape or return to a place of safety such as home. Therefore they fear situations where escape is difficult or impossible, e.g. airplanes, traffic jams on a bridge, dental appointments, etc. On closer examination, it is clear that patients do not actually fear the situation itself but rather reason 'what if' the panic feelings occur while in that situation. This has led to the syndrome being called the 'what if' syndrome, emphasizing that there is actually a 'fear of the fear'.[9]

Patients tend to avoid such situations or force themselves to endure them in distress or take a companion along 'to help'. Others limit their travel to short distances from home or take longer routes where they perceive help would be available (e.g. police, doctors' offices, fire stations, etc).

Some patients develop agoraphobic avoidance following their first attack, some only after frequent and severe attacks, and some never develop agoraphobic avoidance. In community samples, one-third to half of patients who meet criteria for panic disorder also has significant agoraphobic avoidance. The rate is higher (75 per cent) in most clinical samples. A minority (5 per cent) ultimately become unable to leave their homes and are housebound.

Many patients have panic attacks that awaken them from sleep (nocturnal panic attacks), as well as during the day. These are in fact quite common and the majority of panic disorder patients experience them. These occur during slow wave sleep early in sleep. These panic attacks are essentially identical to traditional panic attacks that occur during the day. There is a group of patients who have what are called non-fearful panic attacks. These involve the sudden onset of physiological symptoms without the cognitive components of fear or anxiety. These primarily are medical patients, usually cardiac, who might have episodes of sudden tachycardia and palpitations but no fear.[10]

Classification

The earliest modern accounts of what is almost certainly the panic disorder syndrome began appearing in the mid-1900s. There were accounts beginning in 1866 of paroxysmal anxiety episodes that did not use the term 'panic attack'. During the American Civil War, patients were diagnosed with 'irritable heart syndrome' or 'Da Costa syndrome' (1871) with clear descriptions of what we now know as panic disorder. Westphal in Germany in 1872 clearly described patients having panic attacks and agoraphobic avoidance of wide open spaces.[11] Again, in the First World War (1918) the syndrome 'neurocirculatory asthenia' was described which had most of the features of panic disorder.

It was, in fact, Freud in Case IV of Katharina, published in 1895, who set the background for the modern classification of panic disorder. However, it was the Feighner criteria published in the United States in 1972 that give the first formal diagnostic recognition to the syndrome.[12] The Research Diagnostic Criteria (RDC) which followed in 1978 first split panic disorder away from what we now call generalized anxiety disorder (GAD). In the RDC, panic disorder had panic attacks while GAD did not. These diagnoses were made part of the DSM-III diagnostic scheme in 1980.

It was the conceptualization of panic disorder by Donald Klein in 1964 in the United States that led to the predominant view of the syndrome, certainly in the United States.[13] Klein argued that panic attacks were the core of the syndrome, and the remaining clinical phenomena were consequences of the panic attacks. He conceptualized that anticipatory anxiety was the fear of the possible recurrence of panic attacks, and similarly that agoraphobia was the subsequent fear and avoidance of situations where panic attacks had occurred. The bringing together of these three phenomena into one concept was accepted in the DSM-IIIR, and more recently in the DSM-IV in the United States.[14]

The biological findings that typical panic attacks could be elicited in panic disorder patients with infusions of lactate, doses of caffeine, or breathing 35 per cent CO_2 supported this conceptualization of the syndrome as primarily centered around panic attacks. This hypothesis was further supported by epidemiological findings of essentially the same illness around the world.

However, this idea remains controversial across the different sides of the Atlantic. The American DSM-IIIR and DSM-IV diagnostic schema continue to utilize the idea that panic attacks are pre-eminent and created two diagnoses: panic disorder and panic disorder with agoraphobia. In Europe and in the ICD-10, agoraphobia is conceptualized as dominant over panic attacks. Therefore, when agoraphobia and panic attacks are both present, the diagnosis is conceptualized as a phobia and that condition is diagnosed as agoraphobia with panic attacks. Beyond this theoretical debate is the clinical question whether the treatment should be aimed first at panic attacks (in the United States concept) or at agoraphobic avoidance, for example with exposure therapy (in the European schema).

(a) Diagnosis

The most recently revised diagnostic schema for this syndrome is the DSM-IV. Changes from the DSM-IIIR were made based on two principles:

1 any new empirical data that required changes be made;

2 an attempt to make the DSM-IV and ICD-10 more compatible.

For the diagnoses of panic disorder and agoraphobia, an attempt was also made to more nearly describe the prototypic patient and to move away from the pseudoquantification of using number of panic attacks per week.[14]

The DSM-IV clarified that panic attacks occurred in multiple syndromes including social phobia, obsessive–compulsive disorder, depression, and others. However, DSM-IV utilized the distinction that only in panic disorder were there recurrent spontaneous panic attacks not bound to any particular situation. The diagnosis of panic disorder has several requirements including the following:

- Recurrent, unexpected panic attacks (situational panic attacks could also occur but there would need to be at least two unexpected panic attacks by history).
- Panic attacks needed to be followed by at least 1 month of persistent anxiety about potential recurrence of further panic attacks, implications of these attacks (e.g. going mad, something wrong medically), or a significant behavioural change because of these attacks. This was necessary because some patients had panic attacks and completely changed their lives but denied that they were worried about experiencing more panic attacks or the implications of the panic attacks.[14]

The agoraphobia criteria remained largely unchanged, but it was made more clear that the diagnosis was based on persistent fear and avoidance of certain clusters of situations and listed the most common.

The controversial diagnosis of agoraphobia without a history of panic disorder was retained until further clarification is obtained through research. Our current understanding is that these patients generally have never fully met criteria for panic disorder because their panic-like symptoms have not met the diagnostic criterion requiring four full symptoms for a panic attack. Available research suggests that these patients are otherwise very much like typical patients with panic disorder and agoraphobia. Some patients actually have only one or two symptoms (e.g. fear of loss of bladder or bowel control or only tachycardia).

Perhaps the most difficult diagnostic issue is the frequent comorbidity. The Epidemiologic Catchment Area study documented that approximately 50 per cent of panic disorder patients over their lifetime have another anxiety disorder.[15]

In actuality, depression is more commonly comorbid with panic disorder than even agoraphobia and suggests a close relationship between these syndromes. Comorbid depression ranged from 22.5 to 68.2 per cent in various samples. Lifetime rates vary from 35 to 91 per cent.[16] Although approximately half the patients developed panic disorder and depression at essentially the same time, one-quarter develop depression before panic disorder and one-quarter panic disorder before depression.[17] Surprisingly, bipolar illness has been reported to be as high as 20.8 per cent.

Easily one-third of panic disorder patients abuse alcohol. The percentage in clinical samples is much higher with 13 to 43 per cent of panic disorder patients also meeting criteria for alcoholism.[18]

(b) Differential diagnosis

It is particularly important to determine whether agoraphobic avoidance is present, because its treatment usually requires some sort of exposure therapy. Patients will often not volunteer that they are avoiding certain situations out of embarrassment. As mentioned earlier, depression and panic disorder often occur together and again patients often do not describe the other syndrome, but rather describe the syndrome which is most painful to them at that time. However, proper recognition of comorbid depression is especially important because of the marked increase (four-fold) in suicide attempts in patients with panic disorder and depression.

The difference between panic disorder and GAD depends on whether patients have panic attacks and whether they have multiple, unrealistic, and excessive worries about most aspects of life, not just panic attacks. These worries in GAD often concern money, health, children, work problems, etc. The differential with social phobia centres on whether the anxiety is confined entirely towards social situations where the individual fears embarrassment and humiliation. Specific phobias involve panic attacks, but they occur in very specific situations (e.g. high places, thunderstorms) or in the presence of specific objects (e.g. animals, snakes). The post-traumatic stress disorder patient may have many panic-like symptoms, but their illness begins quite specifically after a traumatic experience and anxiety is associated with reminders of that trauma. Finally, obsessive–compulsive disorder can involve panic attacks but only in the specific context of obsessional concerns (e.g. contamination, etc.). In these patients, panic attacks are also dwarfed in importance by typical obsessions and compulsions/rituals concerning contamination, symmetry, bad events, etc.

(c) Medical conditions

Panic-like symptoms do occur in various medical conditions (hyperthyroidism, phaeochromocytoma, hyperparathyroidism, seizures, cardiac arrhythmias, especially supraventricular tachycardia, inner ear difficulties, chronic obstructive pulmonary disease, use of marijuana, withdrawal from drugs of abuse, and over the counter drugs containing caffeine or pseudoephedrine) (Table 4.7.3.1). Also, the typical panic disorder patient does report a large number of physical symptoms and usually to a non-psychiatric physician.

As mentioned, there are medical conditions that can mimic panic disorder (Table 4.7.3.1). There is also evidence that there are slightly increased rates of certain illnesses, for example hyperthyroidism (11–13 per cent) and perhaps mitral valve prolapse. However, these are uncommon in panic disorder patients. Most experts recommend a relatively conservative diagnostic medical work-up. Generally, the most valuable part of a medical examination is a careful history with follow-up of any strongly suggested possibilities, supplemented by a few laboratory tests (complete blood count, thyroid function tests, metabolic screen, and ECG, especially if the patient is over 40 years of age).

(d) Panic disorder in the general medical setting

Conservative estimates of panic disorder in primary care have ranged from 3 to 8 per cent with at least 50 per cent going unrecognized.[19] The average panic disorder patient in general medicine

Table 4.7.3.1 Medical conditions that produce panic-like symptoms

Endocrine	*Respiratory*
Hyperthyroidism	Chronic obstructive
Hypoparathyroidism	pulmonary disease
Hypoglycaemia	Asthma
Phaeochromacytoma	*Substance-induced*
Carcinoid syndrome	Caffeine
Cushing's disease	Cocaine
Cardiovascular	Marijuana
Arrhythmias (supraventricular)	Theophylline
Atypical chest pain	Amphetamines
Mitral valve prolapse	Steroids
Neurological	Alcohol/sedative
Seizures	withdrawal
Vestibular disease	*Haematological*
	Anaemia

takes 10 years or more for a correct diagnosis to be made with an escalation of the use of health care services over this period. In general, the presence of panic disorder leads to a three-fold increase in utilization of general medical services.

The percentage of panic disorder patients is also markedly higher in certain medical groups. These include the very prevalent but most difficult to diagnose patients with vague symptoms such as fatigue, back pain, headache, dizziness, chest pain, etc.[20] or multiple symptoms (more than five).[21]

In a classic study of unrecognized panic disorder patients who were referred for a psychiatric consultation from primary care, 39 per cent had cardiovascular symptoms, 33 per cent gastrointestinal symptoms, and 44 per cent neurological.[22] It is now clear that 16 to 25 per cent of patients presenting to the emergency room with chest pain have panic disorder. Fully 25 per cent of cardiology practice involves panic disorder, usually unrecognized with 80 per cent of patients with chest pain and normal angiograms ultimately diagnosed with panic disorder. We could also increase our yield of recognizing panic disorder patients in certain procedure-oriented situations. For instance, 28 per cent of patients referred for Holter monitoring for palpitations have panic disorder, as do 66 per cent of patients undergoing a work-up to rule out phaeochromocytoma.[23] Also, 44 per cent of irritable bowel syndrome and 15 per cent of patients with headache symptoms seeing a physician have panic disorder, and these are both very prevalent disorders.

Recent studies document that treatment of panic disorder in the medical setting when diagnosed there is most cost-effective.

(e) Comment

A recent large study sponsored by the World Health Organization (WHO) studied primary care patients in 14 different countries.[19] Of patients in that study who had a single panic attack in the previous month, 99 per cent had an anxiety disorder or depression, or a subthreshold anxiety disorder or depressive disorder. The occurrence of a single panic attack also predicted the onset of panic disorder in two-thirds of the patients studied in the next year, a four-fold increase in depression (51 per cent) in the next year, and marked increases in alcoholism, social phobia, and obsessive–compulsive disorder. It would appear that the occurrence of even a single panic attack may well represent the 'tip of the iceberg' and should serve as a signal for increased scrutiny for anxiety and depressive syndromes. This has been recently replicated in a large ($N = 3021$) European sample.[24]

Epidemiology

Surveys largely utilizing DSM diagnoses have found wide agreement and generally equal prevalence's of panic disorder across many countries.[16,25] Utilizing specific criteria for panic attacks, prevalence for panic attacks has generally averaged 7 to 9 per cent of the population (range 1.8–22.7 per cent). However, if criteria for panic attacks are liberalized somewhat ('fearful spells') in terms of the number of times and severity, the prevalence doubles.

There is a striking uniformity worldwide for the observed prevalence of panic disorder. In 10 community studies involving over 40 000 subjects, the majority of studies found lifetime prevalence rates for panic disorder averaging 1.5 to 3.7 per cent, with a yearly prevalence of around 1 and 1.1 per cent of panic disorder with agoraphobia (lifetime).[25] In clinical samples there is greater variability. In the previously mentioned WHO survey of 14 countries, the prevalence for panic disorder in primary care ranged from a low of 1.4 per cent to a high of 16.5 per cent for panic attacks and 0 to 3.5 per cent for panic disorder itself.[19] The average was 1.1 per cent (currently) and 3.5 per cent (lifetime), which is surprisingly similar to the community samples. As mentioned, rates are much higher in specialized medical clinics and range from 15 per cent in dizziness clinics, to 16 to 65 per cent in cardiology practices, to 35 per cent in hyperventilation clinics, etc.

Risk factors

Panic disorder has been uniformly observed to be at least two times more prevalent in females than males.[25] The Epidemiologic Catchment Area study demonstrated a prevalence of 3:2. In clinical samples it is generally 3:1. The onset of panic disorder appears to fall into two peaks. The first occurs in the early to mid-twenties (15–24 years old) with a second peak at 45–54 years of age. The onset of panic disorder after the age of 65 is rare (0.1 per cent).

The highest rates of panic disorder and agoraphobia occur in widowed, divorced, or separated individuals living in cities. Limited education, early parental loss, and physical or sexual abuse are also risk factors. Agoraphobia is clearly more prevalent in females, and females make up three-quarters of the sample with extensive avoidance. Males tend to have longer duration of illness but less agoraphobia and depression, and less frequent help seeking. Perhaps the greater necessity to perform in the workplace retards avoidance in males.

Aetiology

Genetic predisposition

Certainly the preponderance of evidence suggests there is a genetic contribution to the predisposition to develop panic attacks and agoraphobia. There are increased rates of panic disorder in first-degree relatives ranging from 2- to 20-fold with the median

seven- to eight-fold. Overall, studies suggest that another affected relative can be found 25 to 50 per cent of the time, two times as often in female relatives. The increased familial aggregation is specific for panic disorder. These findings are certainly consistent with a modest genetic transmission with relatively high specificity.

Although twin studies are limited, Torgersen[26] did find increased concordance in monozygotic twins (31 versus 0 per cent). In the largest sample of interviewed female twins, a several-fold increase was again found (23.9 versus 10.9 per cent).[27] Skre *et al.*[28] found a two-fold increase of panic disorder in monozygotic twins, while other studies fail to find an increased incidence.

Overall, evidence from family and twin studies suggests that panic disorder involves modest inheritability of around 30 to 40 per cent. The best model suggests 50 per cent genetic and 50 per cent environmental influences. Recent linkage studies to confirm these hypotheses have been contradictory (e.g. with angiotensin, brain-derived neurotropic factor) but do suggest that single-gene transmission is unlikely. However, research is active in this area with positive replicated linkages with chromosomes 13q, 22q, 7p, and 9q31. Identified candidate genes include the ADOR2A, 10832/T, CCK genes, and genes coding for the 5HT1A, 5HT2A, and COMT genes and there is evidence of a link to the corticotrophin releasing hormone gene.[29] This leaves the possibilities of either heterogeneity across families and/or a polygenic inheritance.[30]

Several converging lines do link childhood anxiety with adult anxiety, consistent with a genetic predisposition. This is particularly true for separation anxiety in children. Kagan *et al.*[31] have prescribed that 10 per cent of Caucasian children are born with heightened anxiety which they call behavioural inhibition. Behavioural inhibition is higher in children of anxiety-disordered parents, and there are high rates of anxiety disorders in children of adults with panic disorder. As behaviourally inhibited children have matured, they have been found to show higher rates of anxiety and phobic disorders.[32] Currently, this is probably the best model of an inherited anxiety predisposition. A variant of this type model proposes that there is an aethological factor involving an evolutionarily determined vulnerability to unfamiliar territory. This might explain why the seemingly inherited anxiety is to specific situations. This is also consistent with the high rate of precipitating events prior to the onset of clinical difficulties. In this model an evolutionarily/genetically determined vulnerability would be clinically 'activated' by critical stressors.

Precipitating events have been reported in 60 to 96 per cent of cases. These have often centred on separation or loss, relationship difficulties, taking on new responsibility, and physiological stressors (e.g. childbirth, surgery, hyperthyroidism).[33] This is certainly consistent with a diathesis/vulnerability model with the illness being precipitated in a predisposed individual in adulthood.

There are also many studies suggesting that traumatic early events may figure in the vulnerability leading to panic disorder. The majority of children in some studies have experienced early parental separation. A traumatic event in childhood has been retrospectively reported in at least two-thirds of individuals, a three-fold increase.[34] The most common adult disorder following sexual abuse before the age of five is in fact a 44 per cent incidence of agoraphobia.[35]

There is some evidence in a prospective study involving over 3000 individuals that dependent personality traits were later associated with the development of anxiety disorders. There are also retrospective data that adult panic disorder patients describe their parents as being overly protective and less caring. It is difficult to separate the effects on individuals of the anxiety disorders themselves which create dependent behaviour, or overprotectiveness in the parents producing dependent personality traits.

Biological models

(a) Noradrenaline

There is considerable evidence implicating the brain noradrenaline (norepinephrine) brain systems and panic disorder. The noradrenergic agents yohimbine and isoproterenol stimulate panic attacks in panic disorder patients, suggesting a possible subsensitivity of pre-synaptic alpha 2 inhibitory adrenoreceptors. Both these drugs increase the firing rate of the locus ceruleus, generally thought to be part of the brain anxiety circuit. It is also true that most effective medications in the treatment of panic disorder in fact decrease locus ceruleus firing rate and most panicogenic stimuli increase the locus ceruleus firing rate.

(b) Serotonin (5HT)

Findings with 5HT brain systems in panic disorder are contradictory, probably because of the different 5HT circuits and receptors in different areas of the brain. Most investigators believe, however, that an increase in 5HT transmission decreases panic disorder perhaps because 5HT neurones in ventrolateral periaqueductal grey appear to inhibit sympathoexcitation and the fight or flight response in the rat.[36] The principal human evidence for 5HT being central in panic is that the selective serotonin reuptake inhibitors are effective and that they increase 5HT transmission after long-term use. Also, rapid depletion of 5HT has been shown to result in an increase in panic responses to flumazenil. Whether this increased neurotramsmission in fact leads to downregulation of a supersensitive post-synaptic receptor is one logical possibility, but is as yet unproven. These theories received recent support from PET scan studies demonstrating reductions in brain 5HT1A receptors[37] and 5HT transporter binding.[38]

(c) γ-Aminobutyric acid

The γ-aminobutyric acid (GABA) system is almost certainly involved in panic disorder with perhaps the strongest evidence being that benzodiazepine agonists such as alprazolam and clonazepam are clearly effective treatments for panic disorder. Also GABA antagonists such as flumezanil have panicogenic effects in panic disorder patients, and reverse benzodiazepine agonists such as β-carbolines can cause panic attacks. There is an impaired GABA neuronal response to benzodiazepines (BZs) on brain magnetic spectroscopy and decreased GABA levels in the cingulate and basal ganglia, also on magnetic spectroscopy. Also, positron emission tomography data have demonstrated decreased benzodiazepine binding in the inferior brain areas, including the inferior parieto-temporo-occipital areas.

(d) Cholecystokinin–pentagastrin

Cholecystokinin is clearly involved in anxiety in animals. Also, panic disorder patients develop panic attacks in a dose-dependent fashion with administration of pentagastrin. However, cholecystokinin antagonists have not yet been shown to be effective in humans.

Recent genetic studies implicate CCK gene polymorphisms in panic disorder.

(e) Brain imaging

The explosion of brain imaging data demonstrating a brain circuit for fear and anxiety involving the extended amygdala circuit (amygdala, hippocampus, periaqueductal grey, locus coeruleus, thalamus, cingulate, and orbitofrontal areas) has lead to the hypotheses that it is this circuit which is abnormally active in panic disorder.[39]

(f) Psychological factors

Many critics disagree with the importance of biological findings in panic disorder, principally, various European workers and the cognitive theorists. They argue that panic attacks are not 'biological' and that a phobic attitude is required, and/or certain temperamental factors. Others attempt an integrated model utilizing both findings of biological differences and psychological factors of temperament and child-rearing practices.

Course and prognosis

There is limited evidence with appropriate population-based samples to clearly delineate the course of panic disorder. Available evidence suggests that panic disorder is a stable but chronic condition once criteria for the disorder are met. Most patients seeking treatment have experienced chronic, frequently chronically worsening, illness generally for 10 to 15 years prior to diagnosis.[7] However, other evidence does suggest that there is heterogeneity in terms of course.

As previously mentioned, the Klein model suggests that spontaneous panic attacks are the first manifestation of the illness, followed by anticipatory anxiety, and agoraphobia in some individuals. However, for most panic disorder patients examined closely, a panic attack is rarely the first symptom. In some studies, over 90 per cent of patients have had mild phobic or hypochondriacal, milder symptoms prior to the onset of their first panic attack.

The first panic attack is usually in a 'phobogenic' situation such as a public place, street, store, public transportation, crowd, elevator, tunnel, bridge, or open space. As mentioned, these are often preceded by stressful life events.

The earliest studies indicated low-recovery rates with chronic waxing and waning in most patients. Some individuals have outbreaks of symptomatology with less difficulty in between, but the majority of untreated individuals seem to have a more or less continuous symptom picture which ranges from mild to severe.

Recovery rates vary from 25 to 75 per cent for 1 to 2 years follow-up. Over a 5-year follow-up, only 10 to 12 per cent had fully recovered in one study and 30 per cent in another. The most common picture is about 50 per cent of patients are neither well nor very sick with mild symptoms most of the time.[40] After diagnosis and some sort of treatment, functional recovery occurs in the majority of patients.[41]

In acute pharmacological trials, 50 to 70 per cent of patients have excellent acute responses with another 20 per cent having a moderate response. With behavioural therapy, again the majority of patients recover and in some trials over 75 per cent of patients are much improved 1 to 9 years following therapy, with an average decrease in symptoms of 50 per cent.[42]

Poor responses were most consistently associated with initial high symptom severity and high agoraphobic avoidance at baseline. Poor response is also associated with low socio-economic status, less education, longer duration, limited social networks, death of a parent, divorce or unmarried status, and personality disorders.

Treatment

Introduction

Multiple effective treatments have been developed since the early 1960s and include both psychological and pharmacological treatments. Both exposure-based treatments and imipramine were shown to be effective in treating panic disorder and agoraphobia in the early 1960s.[3,4,43] Psychological-based treatments have moved increasingly towards cognitive behavioural therapy with efficacy roughly comparable to pharmacological treatments.

Imipramine and monoamine oxidase inhibitors (MAOIs) were the first medications shown to effectively treat panic disorder in the 1960s.[3,44] The high-potency benzodiazepines (alprazolam and clonazepam) were also shown to be effective in the 1980s.[7] Most now agree that the selective serotonin reuptake inhibitors (SSRIs) are the medication of first choice.[45–47]

Factors that influence the choice of initial treatment include patient preference and past history of treatment response, costs, and often availability. Medication treatment is usually easier to obtain but does involve significant costs and side effects. Although many patients prefer psychological treatments, as many as 10–30 per cent refuse treatments that involve exposure to frightening situations or resist the time and effort required. Where available, CBT can be more cost-effective than medications.

CBT has been modified in various ways to try to make it more easily available. This has included delivery in groups by computer or telephone or in shorter amounts or in a high intensity strategy with multiple hours of therapy over just a few days. All of these approaches show promise. There is clear evidence that bibliotherapy with and without phone contacts is also effective.

Eye movement desensitization and reprocessing (EMDR) was developed for treatment of post-traumatic stress disorder. The evidence available suggests that it is not effective in panic disorder.

Panic disorder can have an onset prior to adolescence, and it does occur frequently during adolescence. Although empirical data are very limited, it is generally assumed that treatments effective in adults are also effective in children. (see American Academy of Child and Adolescent Psychiatry's *Practice Parameters for the Assessment and Treatment of Children and Adolescents with Anxiety Disorders*.[48]

Medication treatments

(a) Selection of initial pharmacotherapy

Evidence indicates that medication from five classes—the SSRIs, SNRIs, benzodiazepines, tricyclics (TCAs), and monoamine oxidase inhibitors (MAOIs) are all roughly equal in their efficacy and therefore the choice of initial therapy should be made on other factors such as tolerability, cost, prior treatments, etc. (see Table 4.7.3.2). For patients with a history or concurrent depression (usually 25 per cent), the antidepressants would be preferable to the benzodiazepines. The antidepressants generally take 4 to 6 weeks to become effective, whereas the benzodiazepines begin working within the first week. There is evidence that adding benzodiazepines to the antidepressants speeds the therapeutic response.

Table 4.7.3.2 Advantages and disadvantages of various antipanic agents

	Advantages	Disadvantages
SSRIs	Well-tolerated antidepressant Safe in overdose Little weight gain Once-daily dosing	Initial activation Nausea, headache, asthenia, insomnia initially Sexual side effects
SNRIs	Very similar to SSRIs	*Hypertension*
Benzodiazepines	Rapid efficacy Reduce anticipatory anxiety Well tolerated No initial activation Safe in overdose	Sedation Some memory problems Withdrawal Abuse potential Rare sexual dysfunction
Tricyclic antidepressants	Single daily dose Less expensive Long experience Antidepressant	Initial activation Anticholinergic side effects Weight gain Orthostatic hypotension Dangerous in overdose Sexual dysfunction
MAOIs	More effective (against comorbid depression)? Antidepressant	Dietary restrictions Hypertensive crises (rare) Initial activation, insomnia Onset delayed Anticholinergic side effects Orthostatic hypotension Dangerous in overdose

A hyperstimulation reaction has been observed to all of the antidepressants and has lead to the widespread use of very low doses initially with gradual escalation. Benzodiazepines do not appear to produce the initial hyperstimulation response and are therefore preferred by many patients. However, difficulties with tapering and discontinuing benzodiazepines are the principal negative consideration for their use.

(b) Selective serotonin reuptake inhibitors (SSRIs)

In the United States, there are now six SSRIs available and three have FDA approval for panic disorder (fluoxetine, sertraline, and paroxetine—IR and CR formulations). There is no scientific or even clinical evidence to suggest significant differences in efficacy between the SSRIs in this indication. However, there are differences in side effects, principally, weight gain and discontinuation symptoms, different potential drug interactions, and availability of generic formulations.[49]

Initial jitteriness or increased anxiety are observed with the SSRIs. Therefore, treatment is often begun with the lowest doses available (see Table 4.7.3.3). Some patients do respond at lower doses, although most require higher doses (again see Table 4.7.3.3). The reason the SSRIs are generally regarded as the first choice for pharmacotherapy include their better tolerability and absence of anticholinnergic effects, compared to the TCAs. Patients often have mild difficulties with nausea, insomnia, headache. Certainly, the most problematic side effect is sexual dysfunction in both men and women, most frequently delay in orgasm. There are rare reports of extrapyramidal side effects and gastrointestinal bleeding. Discontinued too rapidly, withdrawal symptoms of headache, irritability, dizziness, can appear in the first 1 to 5 days after withdrawal, and generally clear within 1 to 2 weeks. These can generally be avoided by taper over 1 to 3 weeks.

(c) SNRIs

The SNRI venlafaxine has recently been demonstrated in large multicenter trials to be effective in the range of 75 to 225 mg/day. The side effect profile was similar in severity and symptomatology to the SSRIs, although a small number may develop sustained hypertension. There is some evidence that venlafaxine can result in higher rates of death from overdose than the SSRIs.

(d) Tricyclics (TCAs)

The first trial demonstrating that imipramine was significantly better than placebo was published in 1964[3] and was followed by multiple controlled trials demonstrating its effectiveness. There is also a significant number demonstrating effectiveness of clomipramine and some even demonstrating greater efficacy than imipramine.[50,51] Some data are supportive of desipramine and nortriptyline. Consistent problems with poorer tolerability compared to the SSRIs has lead to the TCAs being used only infrequently.

(e) Benzodiazepines (BZs)

The most widely studied and utilized benzodiazepine has been alprazolam. The largest trial was the Cross National Collaborative

Table 4.7.3.3 Medication doses in treatment of panic disorder

	Starting dose mg/day	Therapeutic range mg/day
SSRIs		
Paroxetine	10–12.5 CR	10–40*
Fluoxetine	2.5–10	10–20
Sertraline	25	50–200**
Fluovoxamine	50	100–300
Citalopram	10	20–30***
Escitalopram	5	5–10
SNRIs		
Venlafaxine	37.5	75–225
TCAs		
Imipramine	10	50–200
Clomipramine	25	25–150
BZs	TID or QID	Acute total daily dose
Alprazolam	0.25–0.5	2–10†
Clonazepam	0.25–0.5	1–4
Lorazepam	0.5	1–7
Diazepam	5	5–40
MAOIs	BID	
Phenelzine	15	15–45 (or 90)
Tranylcypromine	10	10–40 (or 70)

*40 mg demonstrated as target dose in RCT.

**All doses equivalent in one trial 50, 100, 200.

***In one trial, 20–30 was more effective than 40–60 mg/day.

†Mean dose 5.4 mg/day in largest trial.

Panic study involving more than 1000 patients and 11 trials, the majority of which were double-blind.[7] Alprazolam was effective against all of the symptoms of panic disorder, was comparable to imipramine and better tolerated. A sustained release form of alprazolam which can be taken once-daily is currently available. This formulation's long half-life appears to have solved the inter-dose rebound symptoms and 'clock watching' which was problematic with the immediate release form.

Clonazepam has also been demonstrated to be effective and is FDA approved for panic disorder in the United States.[52] Diazepam and lorazepam have also been shown to be clinically effective. Although generally well tolerated, the benzodiazepines principal side effects include sedation, occasional ataxia, slurred speech, and small increases in memory complaints. The largest concern and controversy has been with the possibility of dependency and the possibility of recreational abuse. However, the *American Psychiatric Association Task Force on Benzodiazepine Dependence, Toxicity and Abuse*, based on multiple large trials stated 'there are no data to suggest that long-term therapeutic use of benzodiazepines by patients commonly leads to dose escalation or to recreational use'.[53] Doses in long-term treatment are either similar to short-term or lower. Discontinuation symptomatology is perhaps the largest problem with the benzodiazepines. Panic patients have more difficulty discontinuing benzodiazepines than patients with generalized anxiety disorder. Symptoms are often seen during taper and are greatest during the last part of taper and the first week after taper. It is inconclusive whether these symptoms represent withdrawal, rebound or relapse, or the combination. However, it is clear that abrupt discontinuation results in greater symptomatology than a gradual taper. Most experts suggest a taper over several months (2–4 months).[54] Gradual taper and personality issues (more symptoms with a higher anxiety sensitivity and avoidance) are more critical than half-life of the medication.

Daily doses of alprazolam have varied from 2 to 10 mg. Greater efficacy is generally seen with higher doses, and the largest trial averaged 5.4 mg/day[7] (see Table 4.7.3.3).

Doses of clonazepam are often 50 per cent or less than doses with alprazolam.[52] Some clinicians prefer clonazepam over alprazolam because its longer half-life allows it to be used less frequently each day.[52] However the recent availability of an extended release once daily alprazolam (alprazolam ER) has perhaps reversed that issue. Although the BZs are generally safe in overdose, there is recent evidence that alprazolam is associated with more morbidity than the other BZs. There are studies demonstrating that lorazepam is effective averaging 7 mg/day and 5 to 40 mg/day is effective in diazepam trials (Table 4.7.3.3). The most important aspects of benzodiazepine treatments are the rapid response and increased tolerability and perhaps the greater reduction in everyday anticipatory anxiety (Table 4.7.3.2). Certainly, the largest drawbacks are in general the lack of efficacy against depression and difficulties with the tapering and discontinuation of treatment.

(f) Monoamine oxidase inhibitors (MAOIs)

Many experienced clinicians feel the MAO inhibitors may be the most effective medication class in treatment of panic disorder. However, there is only one scientifically rigorous trial supporting its use.[44] Further, concerns about its safety and the requirement of a tyramine diet limit their use. Modern studies aimed at developing reversible inhibitors of the MAOI enzyme that could eliminate the dietary restrictions have unfortunately been disappointing. The MAOI-B inhibitor selegiline has recently been made available in the United States but there are no known studies of its use in panic disorder.

Patients must also avoid sympathomimetic agents frequently found in decongestants, the antibiotic linezolid, meperidine, fentanyl, and tramadol, serotonergic agents like fenfluramine and the migraine triptan medications. Other significant side effects include weight gain, sexual dysfunctions, postual hypotension, anticholinergic side effects, and sleepiness.

(g) Other antidepressants

There are two uncontrolled trials of buproprion and buproprion SR in panic disorder, one positive and one negative. Mirtazepine has limited evidence supporting its use, as does inositol. Reboxetine has positive and negative evidence, as does with buspirone.

(h) Other agents

Although there is no evidence that conventional antipsychotic medications are effective, there is growing clinical use of the atypical antipsychotics, particularly in treatment resistant patients. There are single trials suggesting the anticonvulsants valproate and levetiracetam may be effective.

(i) Antihypertensives

Although widely used by non-psychiatric physicians in panic patients, the evidence suggests that propranolol, although it does reduce heart rate, is ineffective in treating panic disorder.

(j) Second-line medication treatments

There is limited scientific evidence to guide the clinician in the choice of the second course of treatment if the first is ineffective. Clinically, if the patient has experienced some benefit from the first medication, most clinicians would either add a benzodiazepine or CBT (which has been shown to work in SSRI failures). If the first treatment is completely ineffective, certainly switching treatment to a different SSRI or a different class of medications would be reasonable. If the second-line treatment is also ineffective, there is preliminary evidence that use of the atypical antipsychotics, olanzapine, or resperidone might be appropriate in severe non-responsive patients, and the other atypicals are also likely to work.

(k) Length of treatment

If patients do respond to an antidepressant, continuing treatment for 6 months or longer generally results in continued improvement and a decreased risk for relapse and recurrence, especially if symptoms remit.[55] Response is generally retained as long as medications are maintained. Most studies, clinical experience, and consensus opinion suggests continuation of effective medications for 12 to 18 months or longer.

(l) Discontinuing treatment after an effective response

Although antidepressants can be tapered over 10 days, clinical evidence recommends a much slower taper involving weeks to months. With benzodiazepines, it appears critical to taper even more slowly. Discontinuation symptoms are common but are significantly minimized if taper is accomplished over 2 to 4 months.[56] CBT has been demonstrated to be helpful in decreasing discontinuation difficulties when focused on the sensations, bodily symptoms, and catastrophic misinterpretations that are often seen.

Psychological treatments

(a) Psychodynamic psychotherapy

Although psychodynamic psychotherapy remains a popular treatment for all psychiatric disorders including panic disorder, there has been very little research demonstrating its efficacy in panic disorder. There is one large case-report study of patients with panic disorder reporting that most patients did respond well. One trial compared clomipramine with clomipramine plus 15 weekly sessions of brief dynamic psychotherapy. Patients in both groups responded well with 75 per cent of the patients in the clomipramine group being panic free and all of the patients in the combination group.

In an extension of this work, an emotion-focused treatment for panic disorder has been developed which explores typical fears of being abandoned or trapped as stimuli for panic attacks. This often involves a 12-session acute treatment with six sessions of monthly maintenance in which patients are encouraged to identify, reflect upon, and attempt to change problematic feelings and their responses. A specific form of dynamic psychotherapy 'panic-focused psychodynamic psychotherapy' which is generally applied in 3-month increments, was recently shown to be effective in a RCT.[57]†

Behavioural treatments

(a) Exposure treatments

Behavioural treatments which utilize *in vivo* exposure to phobic situations have been the mainstay of the behavioural treatments of panic disorder. They are based on the theory and evidence that patients who enter a feared situation experience habituation of their anxiety whether they are exposed slowly or suddenly and extensively (flooding). The critical nature of exposure for improvement was made clear in one study in which patients being treated with imipramine received no improvement from imipramine when given anti-exposure instruments.[58] However, they significantly improved if it were simply suggested that they re-expose themselves to their previously phobic situations when ready. Exposure treatment is consistently associated with long-lasting continuation of acute improvements even without formal follow-up treatment.[59]

Studies have compared use of exposure therapy to cognitive behavioural therapy, and have found them to be essentially equal in efficacy.[60] At this point, exposure has become an integral part of most CBT protocols.

(b) Cognitive behavioural therapy (see Chapter 6.3.2.1)

Cognitive behavioural therapy of panic disorder evolved from early work of Aaron Beck, but has been applied to panic disorder primarily by Barlow and colleagues working in the United States[42,61] and by Clark in the United Kingdom (see Chapter 6.3.2.1).

CBT for panic disorder usually begins with education (e.g. symptoms are part of body's fear response and aren't dangerous) about panic and the cognitive model of panic attacks, use of diaries for self-monitoring of symptoms, cognitive restructuring, habituation to fearful cues including internal cues (e.g. dizziness, tachycardia, etc.) and external situations (e.g. public places, elevators, etc.), anxiety management techniques (e.g. diaphragmatic breathing), and education to prevent relapse. Although evidence suggests that breathing retraining is not an essential component of effective CBT, it is widely utilized as an anxiety management technique.

The most informative trial to date was a multicentre (N 312) 11-week acute trial with a 6-month follow-up for responders and a 6-month follow-up after discontinuation of treatment.[62] This trial compared imipramine to cognitive behavioural therapy, their combination, and placebo. Improvement in all the active treatment cells was approximately equal at the end of the acute trial and significantly greater than placebo on most measures. This was also true at the 6-month follow-up. Interestingly, responders to imipramine had a more robust response than responders to cognitive behavioural therapy alone. The combination of imipramine plus cognitive behavioural therapy was significantly better than CBT or imipramine treatment alone at the 6-month point where all treatments were still present. However, at follow-up 6 months after treatment, none of the treatment cells were statistically different from placebo. There are many well-controlled trials demonstrating that CBT delivered in various forms and formats is effective. Its effects are robust, and at least comparable to medications.[62–64] Benefits are generally long-lasting, with or without booster sessions in follow-ups to 5 years.[63,65] CBT delivered in a group format is also effective.

Continuation/maintenance treatments

There are now a series of studies utilizing antidepressants or benzodiazepines as maintenance treatment for 6 and 12 months for panic disorder and agoraphobia. In most trials, treatment gains from acute treatment are almost always maintained, and generally are extended while the medication is continued. In a 12-month trial comparing clomipramine and placebo, the clomipramine group continued to improve and tolerated the medication well. Placebo patients who were switched to active medication matched the good responses of the clomipramine group.

In a 6-month continuation study of alprazolam patients following an 8-week initial trial, the group maintained their efficacy with a dose at the end of the 8-week trial of 5.1 ± 2.3 mg/day. This decreased to 4.7 ± 2.1 mg/day at week 32 and subsequent follow-up 1 to 2 years later found most patients' doses had drifted down to 1 to 2 mg/day.

In the large follow-up to the Phase II Cross-National Panic Trial, there was a 32-week double-blind comparison of alprazolam, imipramine, and placebo in 181 patients. Again, efficacy was maintained with both medications with no escalation of dose. Patients on both active treatments generally extended their improvement, although the placebo patients tended to lose some efficacy and certainly had a higher drop-out rate. A long-term extension continued paroxetine, clomipramine, and placebo in 176 patients following an acute trial. During the 1-year extension, both the paroxetine and clomipramine patients continued to improve and again, placebo patients tended to lose some of their initial response.

As evidence has accumulated of the high relapse rate with discontinuation of effective medication treatments for panic disorder, longer-term treatment, generally 6 to 18 months, has become routine. Although not well documented, perhaps one of the more important issues is that it appears that patients not only continue to improve for the first 6 months but that improvements continue to be extended the longer patients are on treatment, perhaps even throughout the first 2 years of treatment.

Prevention of recurrence

Available evidence remains inconclusive about the percentage of patients who will relapse if effective pharmacotherapy of panic disorder is discontinued. Early estimates suggested that most patients relapse. It remained the prevailing opinion that 35 to 85 per cent of patients relapse after antidepressants or benzodiazepines were discontinued. However, one trial reported almost no relapse after discontinuation of clomipramine patients, perhaps because they used a gradual taper.

The early trial by Zitrin *et al*. reported only 26 per cent relapse.[65] In a modern trial comparing imipramine and cognitive behavioural therapy, the imipramine relapse rate was 40 per cent. One of the most carefully performed relapse prevention trials followed a fixed-dose study of paroxetine. After acute treatment, patients were re-randomized in double-blind fashion to receive either paroxetine at their prior dose, or to placebo for an additional 3 months of treatment. Interestingly, only 30 per cent of the patients randomized to placebo relapsed, compared with 5 per cent relapse if paroxetine was continued.[47] The relapse rate after medication discontinuation in a recent trial was only 14 per cent. Although these studies certainly need replication, it suggests that the relapse rate may be lower than previously estimated if patients are slowly tapered and carefully followed.

There is one small study that suggests that the relapse rate is lower if treatment is longer. Mavissikalian followed a small group of patients who responded to imipramine, discontinuing some after 6 months of treatment and the other group after 18 months of treatment.[66] In these patients, there was an 80 per cent relapse rate in the 6-month treatment group, but only 20 per cent in the 18-month treatment group. This suggestive finding is consistent with clinical experience but certainly needs replication; however, it is certainly supportive of the general recommendation of continued treatment for 12 to 18 months if effective.

There is certainly a strong suggestion that rapid taper of benzodiazepines produces significant withdrawal symptoms which probably stimulates relapse.

Management

The suggestions for management in this section are based on evidence of the empirically based treatments in panic disorder and agoraphobia, much of which has been reviewed above. However, as in treatment of all patients, there are suggestions that also involve the 'art' of treating these patients which have evolved, but have not been empirically studied or confirmed.

Management of the uncomplicated patient

As reviewed above, it appears clear that the average patient with panic disorder can be treated with a variety of medications or exposure-based and/or cognitive behavioural treatments designed for panic disorder with approximately equal efficacy. There are some patients who have strong feelings or prejudices for or against both medication and cognitive behavioural treatments. Given that situation, as well as the lack of any clear reason to choose one treatment over the other, ethical practice would dictate offering patients a choice of treatment. There is also some evidence that patients will respond better to the treatment they 'believe in'. Unfortunately, the types of treatments are not equally available in all settings or all countries. Psychiatrists tend to use medication treatment with education, exposure, and cognitive based work of a less systematic nature than psychologists and other non-physician caregivers. Although many behaviourally and cognitively oriented psychologists do affiliate with psychiatrists and other physicians to provide medications, for many this ease of combination treatment is not readily available.

As mentioned, most psychiatrists utilize one of the medications mentioned above, supplemented by clear educational efforts with the patient and pertinent family members. This generally includes use of some written material which the patient and spouse read and discuss with the psychiatrist. (see Appendix for suggestions) Education is also a critical part of the initial treatment of patients in exposure-based treatments and cognitive behavioural therapy. These educational efforts are almost always very helpful and in the more mildly symptomatic patients may suffice. Certainly a central issue to increase the therapeutic alliance is to make clear that the therapist does understand that panic attacks involve marked 'real' physical symptoms which are extremely frightening, even though they are not dangerous and are short-lived generally.

For the patient who will be prescribed medication, it is most reasonable to offer a discussion of which medications might be appropriate, and the pros and cons of each. As outlined in Table 4.7.3.2, each medication is different, and depending upon the individual patient's needs and previous experience, any of the classes of medications might be appropriate. As mentioned, current opinion would suggest that the medication of first choice, would probably be an SSRI. In a recent meta-analysis of all the effective medications utilized in the treatment of panic disorder, the SSRIs were shown to be more effective than the other classes.[67] Coupled with their greater tolerability, lack of weight gain, and safety in overdose, they would appear to be the logical first choice.

For clinical and other practical issues, all patients should be told initially that whatever medication is the initial treatment, there are multiple other effective medications. It is important to emphasize that there is little way of knowing which specific medication is most appropriate for which patient and that the initial choice may not be effective, but subsequent choices are likely to be effective.

There are few data to direct the choice of the medications beyond those favouring the SSRIs mentioned above (see Table 4.7.3.2). There is only one trial documenting a difference in patient type leading to a choice of medication. In the large cross-national comparison of imipramine, alprazolam, and placebo, patients with predominantly respiratory symptoms responded better to imipramine.[68] Similarly, patients with a predominantly cardiovascular symptom picture responded better to alprazolam than imipramine. Otherwise, there are no data suggesting a particular medication for a specific patient beyond the various advantages and disadvantages listed in Table 4.7.3.2.

Once medication is chosen, it is prudent to begin at the lowest dose possible (see Table 4.7.3.3). This beginning low dose also extends to the benzodiazepines but is less critical since they are not associated with an initial hyperstimulation reaction. This is one of the reasons why benzodiazepines are easier to utilize and usually more popular with patients who somehow realize that from previous experience or feedback from other patients that they are not associated with an initial worsening of symptoms and are better tolerated overall. At a practical level, management of the worries about the initial hyperstimulation reaction is one of the most important issues in the psychopharmacological management of

panic disorder patients. If handled incorrectly, this issue can lead to a drop-out rate that reaches 25 to 50 per cent. With proper reassurance and close follow-up of patients, this drop-out rate can be reduced to almost zero. Patients need to be told that hyperstimulation can occur in one-third of patients but is transient and not dangerous. Because of the inherent anxiety and even phobia about taking medications, this reassurance is not usually sufficient and patients need to be invited to contact the treating physician with any anxiety or questions they might have about taking medications coupled with a quick response to their concerns.

After initial tolerance of medication is established, the dose can be raised over several weeks to a target level. Obviously, if a patient does not show a response at lower doses, the medication should be raised to maximum doses (Table 4.7.3.3).

It is important for the patient and physician alike to keep in mind that effectiveness of medications often requires a significant amount of time. The antidepressants as a class routinely take 2 to 6 weeks and with certain medications and patients as much as 6 to 12 weeks before significant effectiveness is established. This is less an issue with benzodiazepines, where initial effectiveness is generally seen in the first week or two, but there too the appropriate doses must be obtained, which often takes several weeks. Higher doses of all medicines are needed to reduce agoraphobic avoidance.

As mentioned, if agoraphobic patients do not gradually re-expose themselves to situations they fear, their avoidance fears will not be decreased. This exposure to their actual phobic situations can be accomplished in many ways. Some patients are capable of gradually re-exposing themselves after they understand the principles of exposure and the need to remain in the situation until their fears diminish. They may need help in establishing a hierarchy of their fears, although it is not actually necessary that they in fact do work-up the hierarchy in a gradually increasing fashion from 'least feared' to 'most feared'. However, it is often easier for most patients to conceptualize and accomplish it in this fashion.

Many therapists develop a hierarchical list and then monitor the patient's progress on a regular basis. Use of a standard scale which monitors the various symptom domains can be very helpful, such as *The Panic Disorder Severity scale* (PDSS).[69] There is evidence that the exposure must be regular, and extensive, often on a daily basis. Also, encouragement from the therapist and partner have both been shown to be important. Some patients can re-enter their phobic situations better if supported by their partner, or other phobics from a support group. If they are particularly afraid, an *in vivo* therapist (often recovered phobics) can be very helpful. The critical issues appear to be approaching their fears in a consistent and systematic basis in the real phobic situations accompanied by encouragement and support.

In a similar fashion, many psychiatrists combine principles of cognitive behavioural therapy without embarking on a formal cognitive behavioural therapy programme. Certainly, this should always involve education about the illness and its treatments. Other elements of identification and challenge of catastrophic thinking are widely applied by psychiatrists, but in a less systematic fashion than in formal cognitive behavioural therapy protocols.

It is important that from the beginning most patients be told that if medication treatment is effective, the expectation is to continue the medication for 12 to 18 months. An important issue to negotiate with the patient is how, when, and if effective pharmacotherapy should be tapered and discontinued.[70] Most evidence and experience suggests that patients be continued long enough to receive maximum benefit from medication treatment. In that context, patients should have experienced symptomatic and functional recovery to a maximum extent possible before discontinuation is considered. Patients should have regained a sense of confidence and control of their symptoms and lives. This might be conceptualized as a 'period of normal living' after attainment of symptomatic control before consideration is given to discontinuing an effective treatment. Relapse rates after such a remission are lower. Because the principal danger of discontinuation is relapse, the time should be carefully chosen. This should be a time when potential disruption from discontinuation symptoms and/or relapse would be least problematic.

There are strong suggestions in the literature, some of which is reviewed above, that all medications, and certainly the benzodiazepines, should be tapered very slowly, probably over 2 to 6 months, if possible.[54] This is both to minimize withdrawal symptoms which are especially frightening to panic patients and to observe for relapse symptoms as medications are slowly tapered.

The strongest reasons for discontinuing effective pharmacotherapy are the problematic side effects and expense.[70] Because this is a syndrome frequently seen in young women, the most important reason may be the wish to conceive a child or the onset of pregnancy. Certainly, routine practice is to try to taper and discontinue all medications before or during pregnancy, but sometimes this is not possible. There are now a series of women who have delivered normal children after having tried unsuccessfully to discontinue medication during pregnancy.

Many patients want to manage their own symptoms without the use of medications, and this is also a reasonable reason to taper and discontinue medications, if strongly felt by the patient.[70] The therapist should explore, however, unreasonable prejudices against the use of medications stimulated by reading, television shows, relatives, or even well-meaning physicians. Most in the field now believe that panic disorder and agoraphobia are conditions similar to hypertension and diabetes in the sense that most patients do not like the thought that they are ill and resist compliance with medication treatment. However, treatment is beneficial and not harmful, and patients often need encouragement and education in order to agree to a programme where they continue medications rather than press to discontinue them.

If medications are discontinued, the patient should be followed closely, at least by telephone, for difficulties that could include withdrawal symptoms, especially with benzodiazepines, or incipient relapse. If relapse does occur, evidence suggests that patients will respond to reinstitution of the same medication treatment regimen. If symptoms and/or functional disability associated with relapse are problematic, patients should be offered retreatment with the same medication or offered other effective non-medication treatments.

Psychological treatments

Management of patients with predominantly psychological treatments also begins with the use of educational materials, as is frequently the case with medication treatment of panic disorder. Almost all psychological treatments involve some sort of exposure-based treatment. In some, this is the predominant modality with considerable variation on how exposure to feared situations is

accomplished. Although some initial exposure therapy in particularly frightened patients may be accomplished in imagination prior to *in vivo* exposure, most exposure treatments are usually attempted *in vivo* from the outset. Most treatments have been shown to be effective if they involve *in vivo* exposure. Gradual exposure is the norm, although some programmes use very rapid exposures often called 'flooding', which involves exposure to multiple phobic situations rapidly over several days. Most programmes involve therapist-assisted exposure, sometimes utilizing professional *in vivo* therapists or volunteers who accompany phobics into their feared situations. Partners of patients are often enlisted as assistants, and there is some evidence that this more effective than non-partner exposure aides.

Most exposure-based treatments involve frequent, often daily practices involving several hours. Often the critical issue is adequate support of the patient to accomplish this much exposure 'homework', as well as encouragement and praise.

Almost anything that can help the patient accomplish the actual exposure appears to be useful and helpful. For instance, manuals and computer programs, as well as telephone-based supervision and encouragement have been shown to be effective. Although many therapists utilize relaxation techniques, applied relaxation has been the most widely utilized and effective. Many therapists also employ breathing retraining, encouraging people who hyperventilate to slow their breathing by utilizing their diaphragm. Although both have been shown to be effective and are widely utilized, other studies suggest they are not essential components of treatment, and their use does vary. Although use of benzodiazepines do decrease patient anxiety about exposure, evidence suggests that the benzodiazepines interfere with the cognitive benefits and habituation effects of exposure.

Cognitive restructuring involves the patient and therapist identifying the so-called 'automatic thoughts' they have with and after each panic attack. These are the misinterpretations that patients make about what these symptoms mean. For instance, the patient and therapist together identify that these symptoms often trigger thoughts that they are very ill, having a heart attack, or perhaps even dying. Over several sessions, these are identified and it is made clear to the patient the power these thoughts have to frighten them. At that point a number of strategies can be tried to try to correct these cognitions. Some involve attempts to compute the actual probability of the catastrophic consequences that patients fear. Others involve 'decatastrophizing' in which the ultimate consequences that patients fear are exposed, and generally can then be disavowed by the patient as extremely unrealistic. Patients can be taught to correct these thoughts or substitute more positive self-statements in their place. These skills are then worked on as 'homework', including actual exposure in which negative thoughts are identified and challenged *in vivo*.

The other usual aspect of cognitive behavioural therapy for panic disorder is interoceptive exposure, in which physical symptoms that frighten patients are identified and then they are taught to habituate to those symptoms and challenge the negative cognitions that arise with them. This can easily be accomplished in an office setting. For instance, if the patients are afraid of dizzy feelings, they can be spun in a chair. If they have fears of fast heartbeat, they can run up the stairs and challenge the negative cognitions that arise.

Both exposure-based treatments and cognitive behavioural therapy often are delivered in an 8- to 16-week treatment format, with varying frequencies of follow-up and attempts are underway to shorten these treatments. There is evidence that patients failing to respond to CBT often respond to subsequent medication treatment.

Treatment of comorbid patients

Treatment of the panic disorder patient comorbid with substance abuse is probably the most difficult challenge. In general, the substance abuse problem tends to be predominant even if it were temporally secondary to panic disorder symptoms. Therefore, treatment of substance abuse generally has to be initiated and completed first, although as soon as possible treatment of panic disorder needs be initiated.

Treatment of comorbid social phobia, obsessive–compulsive disorder, or GAD has recently been made somewhat simpler with the demonstration that the SSRIs are effective in these other conditions as well. Although not yet empirically demonstrated, it is reasonable to expect that an SSRI would effectively treat the panic disorder as well as the other comorbid anxiety disorders. This is an important area for future research. Behavioural treatments specific to obsessive–compulsive disorder and to social phobia may well be needed in addition to medication treatment.

The most common comorbidity is with depression and again one of the advantages of antidepressants is probable dual treatment of panic disorder and depression.[71] Perhaps the most important management issue is recognition that comorbid depression carries with it a marked increase in suicide risk. As mentioned, there is some evidence that depressed panic disorder patients respond better to MAOIs.

Resistance to treatment

There are relatively few systematic data about treatment options for patients resistant to initial medication or to exposure- or cognitive behavioural-based treatments. Generally, however, most patients can be tried on another medication, often with success. If they are on medicine and have not tried exposure or cognitive behavioural therapy that should definitely be added. The converse is also true. Non-response can often be traced to inadequate doses or blood levels of the medication or an inadequate length of trial. Comorbid psychiatric and particularly comorbid medical conditions need to be ruled out. Apparent resistance is often related to concomitant personality disorders or failure of agoraphobics to actually attempt exposure treatments. True resistance to one medication is sometimes overcome by a switch to another medication or to two medications at a time. If the combination of an antidepressant and benzodiazepine has not been tried, that is often the first attempted combination. In highly resistant patients, sometimes a combination of tricyclic and SSRI antidepressants or an atypical antipsychotic can be utilized.

Pregnancy

Panic disorder occurs disproportionately in young women making the issue of pregnancy a critical one. The course of panic disorder through pregnancy is highly variable. Use of SSRIs may be associated with low birth weight and a higher rate of spontaneous abortions, cardiac abnormalities, pulmonary hypertension, and withdrawal symptoms in the newborn if used late in pregnancy.

In general however, increases in congenital abnormalities have not been observed with the SSRIs. Whether benzodiazepines are associated with major malformations like cleft palate is unclear. Use of benzodiazepines near delivery is associated with sedation in the newborn. Also, both antidepressants and benzodiazepines are secreted in breast milk. For all these concerns, CBT is strongly recommended for pregnant women and should be considered in women planning pregnancy.

Ethical issues

The principal ethical issues concern the availability of treatment. Because the two types of effective treatments (medications and exposure or cognitive behavioural therapy) are not widely or equally distributed in all practices or locations, sometimes caregivers face a difficult ethical choice of having only one type of treatment available. In these instances, patients should be informed of the limitations and participate in the choices made.

Most of the treatment experience and certainly the empirical evidence has been in Caucasian patients. We do know that symptoms are different across ethnic groups and that response is often less positive in non-Caucasian groups. Treatments need to be tested and developed for all ethnic and national groups as part of the ethical development of the field.

Possibilities for prevention

The best evidence now suggests that panic disorder is often preceded by an anxiety pattern in childhood. In Kagan's model of behaviourally inhibited children, there is certainly a tendency for the pattern to persist throughout life, but some children appear to lose this trait during their development. This may well be related to parental child-rearing practices in which children are encouraged to face issues they fear rather than be withdrawn and fearful. Research is needed to explore whether different parental rearing practices or educational efforts or early treatment with these children can reduce later development of anxiety disorders. If so, these efforts at the public health and school level need to developed.

The other major preventable aetiological consideration for panic disorder and agoraphobia has been the evidence of negative traumatic events occurring in the childhood of adults with panic disorder. Preventative efforts need to be aimed at these issues through public education and education of caregivers of children. Also, one of the intervention goals in helping a traumatized child should be to prevent future development of anxiety disorders and other problems.

Further information

Barlow, D.H. and Craske, M.G. (2000). *Mastery of your anxiety and panic (MAP-3): client workbook for anxiety and panic* (3rd edn). Oxford University Press, New York.

Barlow, D.H. and Crawke, M.G. (2000). *Mastery of your anxiety and panic (MAP-3): client workbook for agoraphobia* (3rd edn). Oxford University Press, New York.

Pollard, C.A. and Zuercher-White, E. (2003). *The agoraphobia workbook: a comprehensive program to end your fear of symptom attacks.* New Harbinger, Oakland, CA.

Wilson, R.R. (2003). *Facing panic: self help for people with panic attacks*. Anxiety Disorders Association of America, Silver Spring, MD.

Zuercher-White, E. (1999). *Overcoming panic disorder and agoraphobia: client manual*. New Harbinger Publication, Oakland, CA.

Bourne, E.J. (2005). *The anxiety and phobia workbook* (4th edn). New Harbinger Publications, Oakland, CA.

Brantley, J. and Kabat-Zinn, J. (2003). *Calming your anxious mind*. New Harbinger, Oakland, CA.

Foa, E.B. and Andrews, L.W. (2006). *If your adolescent has an anxiety disorder: an essential resource for parents*. Oxford University Press, New York.

Marks, I.M. (2002). *Living with fear: understanding and coping with anxiety* (2nd edn). McGraw-Hill, New York.

Anxiety Disorders Association of America, 3730 Georgia Ave., Suite 600, Silver Spring, MD 20910, www.adaa.org

References

1. Hippocrates (1870). *On epidemics V*. Section 82 (trans. S. Farrar). Cadel, London.
2. Benedikt, M. (1870). Uber Platzschwindel. *Allgemeine Wiener Medizinische Zeitung*, **15**, 488.
3. Klein, D.F. (1964). Delineation of two drug responsive anxiety syndromes. *Psychopharmacology*, **5**, 397–408.
4. Marks, I.M. (1969). *Fears and phobias*. Heinemann, London.
5. Klerman, G.L., Weissman, M., Ovellete, R., *et al.* (1991). Panic attacks in the community: social morbidity and health care utilization. *The Journal of the American Medical Association*, **265**, 742–6.
6. Noyes, R., Jr. and Barloon, T.J. (1997). Charles Darwin and panic disorder. *The Journal of the American Medical Association*, **277**, 138–41.
7. Ballenger, J.C., Burrows, G.D., Dupont, R.L., *et al.* (1988). Aprazolam in panic disorder and agoraphobia: results from a multicenter trial. I. Efficacy in short-term treatment. *Archives of General Psychiatry*, **455**, 413–22.
8. Burns, L.E. and Thorpe, G.L. (1977). The epidemiology of fears and phobias with particular reference to the national survey of agoraphobics. *The Journal of International Medical Research*, **5**, 1–7.
9. Goldstein, A.J. and Chambless, D.L. (1978). A reanalysis of agoraphobia. *Behavior Therapy*, **9**, 47–59.
10. Beitman, B.D., Kushner, M.G., Lamerti, J.W., *et al.* (1990). Panic disorder without fear in patients with angiographically normal coronary arteries. *The Journal of Nervous and Mental Disease*, **178**, 307–12.
11. Westphal, C. (1872). Agoraphobie, eine neuropahtische Erscheinung. *Archiv für Psychiatrie und Nervenkrankheiten*, **3**, 138–61.
12. Feighner, J.P., Robins, E., Guze, S.B., *et al.* (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, **38**, 57–63.
13. Klein, D.F. (1964). Delineation of two drug responsive anxiety syndromes. *Psychopharmacology*, **5**, 397–408.
14. Ballenger, J.C. and Fyer, A.J. (1993). Examining criteria for panic disorder. *Hospital & Community Psychiatry*, **44**, 226–8.
15. Weissman, M.M. (1990). Epidemiology of panic disorder and agoraphobia. In *Frontiers of clinical neuroscience. Clinical aspects of panic disorder*, Vol. 9 (ed. J.C. Ballenger), pp. 57–65. Wiley-Liss, New York.
16. Weissman, M.M., Bland, R.C., Canino, G.J., *et al.* (1997). The cross-national epidemiology of panic disorder. *Archives of General Psychiatry*, **54**, 305–9.
17. Lépine, J.P., Wittchen, H.U., Essau, C.A., and participants of the WHO-ADAMHA CIDI Field Trials. (1993). Lifetime and current comorbidity of anxiety and affective disorders: results from the International WHO-ADAMHA CIDI Field Trials. *International Journal of Methods in Psychiatric Research*, **3**, 67–77.
18. Lydiard, R.B. and Brady, K. (1993). Association of anxiety and alcoholism. *The Psychiatric Quarterly*, **64**, 135–49.

19. Sartorius, N., Uestuen, B., Costa e Silva, J.A., *et al.* (1993). An international study of psychological problems in primary care: preliminary report from the World Health Organization collaborative project on psychological problems in general health care. *Archives of General Psychiatry*, **50**, 819–24.
20. Kroenke, K. and Mangelsdorff, A.D. (1989). Common symptoms in ambulatory care: incidence, evaluation, therapy and outcome. *The American Journal of Medicine*, **86**, 262–6.
21. Simon, G.E. and Von Korff, M. (1991). Somatization and psychiatric disorder in the Epidemiologic Catchment Area study. *The American Journal of Psychiatry*, **148**, 1494–500.
22. Katon, W. (1984). Panic disorder and somatization. Review of 55 cases. *The American Journal of Medicine*, **77**, 101–8.
23. Ballenger, J.C. (1998). Panic disorder in primary care and general medicine. In *Panic disorder and its treatment* (eds. J. Rosenbaum and M. Pollack), pp. 1–36. Dekker, New York.
24. Goodwin, R.D., Lieb, R., Hoefler, M., *et al.* (2004). Panic attack as a risk factor for severe psychopathology. *The American Journal of Psychiatry*, **161**, 2207–14.
25. Kessler, R.C., Chiu, W.T., Jim, R., *et al.* (2006). The epidemiology of panic attacks, panic disorder, and agoraphobia in the national comorbidity survery replication. *Archives of General Psychiatry*, **63**, 415–24.
26. Torgersen, S. (1983). Genetic factors in anxiety disorders. *Archives of General Psychiatry*, **40**, 1085–90.
27. Kendler, K.S., Neale, M.C., Kessler, R.C., *et al.* (1993). Panic disorder in women: a population-based twin study. *Psychological Medicine*, **40**, 397–406.
28. Skre, I., Onstad, S., Torgersen, S., *et al.* (1993). A twin study of DSM-III-R anxiety disorders. *Acta Psychiatrica Scandinavica*, **88**, 85–92.
29. Arnold, P.D., Zai, G., and Richter, M.A. (2004). Genetics of anxiety disorders. *Current Psychiatry Reports*, **6**, 243–54.
30. Roy-Byrne, P.P., Craske, M.G., and Stein, M.G. (2006). Panic disorder. *Lancet*, **368**, 1023–32.
31. Kagan, J., Reznick, J.S., Clarke, C., *et al.* (1984). Behavioral inhibition to the unfamiliar. *Child Development*, **55**, 2212–25.
32. Rosenbaum, J.F., Biederman, J., Gersten, M., *et al.* (1988). Behavioral inhibition in children of parents with panic disorder and agoraphobia: a controlled study. *Archives of General Psychiatry*, **45**, 463–70.
33. Scocco, P., Barbieri, I., and Frank, E. (2007). Interpersonal problem areas and onset of panic disorder. *Psychopathology*, **40**, 8–13.
34. Laraia, M.T., Stuart, G.W., Frye, L., *et al.* (1994). Childhood environment of women with panic disorder and agoraphobia. *Journal of Anxiety Disorders*, **8**, 1–17.
35. Saunders, B.E., Villeponteaux, L.A., Lipovsky, J.A., *et al.* (1992). Child sexual assault as a risk factor for mental disorders among women: a community survey. *Journal of Interpersonal Violence*, **7**, 189–204.
36. Johnson, P.L., Lightman, S.L., and Lowry, C.A. (2004). A functional subset of sertonergic neurons in the rat ventrolateral periaqueductal gray implicated in the inhibition of sympathoexcitation and panic. *Annals of the New York Academy of Sciences*, **1018**, 58–64.
37. Neumeister, A., Bain, E., Nugent, A.C., *et al.* (2004). Reduced serotonin type 1_A receptor binding in panic disorder. *The Journal of Neuroscience*, **24**, 589–91.
38. Maron, E., Kuikka, J.T., Shlik, J., *et al.* (2004). Reduced brain serotonin transporter binding in patients with panic disorder. *Psychiatry Research*, **132**, 173–81.
39. Bremner, J.D. (2004). Brain imaging in anxiety disorders. *Expert Review of Neurotherapeutics*, **4**, 275–84.
40. Katschnig, H., Amering, M., Stolk, J.M., *et al.* (1996). Predictors of quality of life in a long-term followup study in panic disorder patients after a clinical drug trial. *Psychopharmacology Bulletin*, **32**, 149–55.
41. Mavissakalian, M.R. and Prien, R.F. (eds.) (1996). *Long-term treatments of anxiety disorders*. American Psychiatric Press, Washington, DC.
42. Landon, T.M. and Barlow, D.H. (2004). Cognitive-behavioral treatment for panic disorder: current status. *Journal of Psychiatric Practice*, **10**, 211–26.
43. Marks, I.M. (1987). *Fears, phobias, and rituals.* Oxford University Press, New York.
44. Sheehan, D., Ballenger, J.C., and Jacobsen, G. (1980). Treatment of endogenous anxiety with phobic hysterical and hypochondriacal symptoms. *Archives of General Psychiatry*, **37**, 51–9.
45. Jobson, K.O. and Poter, W.Z. (1995). International psychopharmacology algorithim project report. *Psychopharmacology Bulletin*, **31**, 457–507.
46. Ballenger, J.C., Davidson, J.R., Lécrubier, Y., *et al.* (1998). Consensus statement on panic disorder from the International Consensus Group on depression and anxiety. *The Journal of Clinical Psychiatry*, **59**, 47–54.
47. Ballenger, J. (1999). Selective serotonin reuptake inhibitors (SSRIs) in panic disorder. In *Panic disorder: clinical diagnosis, management and mechanisms* (eds. D. Nutt, J. Ballenger, and J.P. Lépine), pp. 159–78. Dunitz, London.
48. Connolly, S.D., Bernstein, G.A., and Work Group on Quality Issues (2007). Practice parameters for the assessment and treatment of children and adolescents with anxiety disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, **46**(2), 267–83.
49. Fava, M. (2006). Prospective studies of adverse events related to antidepressant discontinuation. *The Journal of Clinical Psychiatry*, **67**(Suppl. 4), 14–21.
50. Cassano, G.B., Petracci, A., Perugi, G., *et al.* (1998). Clomipramine for panic disorder. I. The first 10 weeks of a long-term comparison with imipramine. *Journal of Affective Disorders*, **14**, 123–7.
51. Modigh, L., Westberg, P., and Eriksson, E. (1992). Superiority of clomipramine over imipramine in the treatment of panic disorder: a placebo-controlled trial. *Journal of Clinical Psychopharmacology*, **51**, 53–8.
52. Herman, J.B., Rosenbaum, J.F., and Brotman, A.W. (1987). The alprazolam to clonazepam switch for the treatment of panic disorder. *Journal of Clinical Psychopharmacology*, **7**, 175–8.
53. American Psychiatric Association. (1990). Benzodiazepines dependence, toxicity, and abuse: a task force report of the American Psychiatric Association. APA, Washington, DC.
54. Dupont, R.L., Swinson, R.P., Ballenger, J.C., *et al.* (1992). Discontinuation of alprazolam after long-term treatment of panic-related disorders. *Journal of Clinical Psychopharmacology*, **12**, 352–4.
55. Lecrubier, Y. and Judge, R. (1997). Long-term evaluation of paroxetine, clomipramine and placebo in panic disorder. Collaborative Paroxetine Panic Study Investigators. *Acta Psychiatrica Scandinavica*, **95**, 153–60.
56. Ballenger, J.C., Pecknold, J., Rickels, K., *et al.* (1993). Medication discontinuation in panic disorder. *The Journal of Clinical Psychiatry*, **54**, 15–21.
57. Milrod, B., Leon, A.C., Busch, F., *et al.* (2007). A randomized controlled clinical trial of psychoanalyltic psychotherapy for panic disorder. *The American Journal of Psychiatry*, **164**, 265–72.
58. Mavissakalian, M. (1990). Sequential combination of imipramine and behavioral instructions in the treatment of panic disorder with agoraphobia. *Archives of General Psychiatry*, **46**, 127–31.
59. Fava, G.A., Rafanelli, C., Grandi, S., *et al.* (2001). Long-term outcome of panic disorder with agoraphobia treated by exposure. *Psychological Medicine*, **31**, 891–8.
60. Ost, L.G., Thulin, U., and Tamnero, J. (2004). Cognitive behavior therapy vs. exposure *in vivo* in the treatment of panic disorder with agoraphobia (corrected from agrophobia). *Behaviour Research and Therapy*, **42**, 1105–27.
61. Barlow, D.H., Craske, M.G., Cerny, J.A., *et al.* (1989). Behavioral treatment of panic disorder. *Behavior Therapy*, **20**, 261–82.

62. Barlow, D.H., Gorman, J.M., Shear, M., *et al.* (2000). Cognitive-behavioral therapy, imipramine, or their combination for panic disorder: a randomized controlled trial. *The Journal of the American Medical Association*. **283**, 2529–36.
63. Craske, M.G., Brown, T.A., and Barlow, D.H. (1991). Behavioral treatment of panic disorder: a two-year follow-up. *Behavior Therapy*, **22**, 289–304.
64. Butler, A.C., Chapman, J.E., Forman, E.M., *et al.* (2006). The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clinical Psychology Review*, **26**, 17–31.
65. Zitrin, C.M., Klein, D.F, Woerner, M.G., *et al.* (1983). Treatment of phobias. I. Comparison of imipramine hydrochloride and placebo. *Archives of General Psychiatry*, **46**, 127–31.
66. Mavissakalian, M. (1990). Sequential combination of imipramine and behavioral instructions in the treatment of panic disorder with agoraphobia. *Archives of General Psychiatry*, **46**, 127–31.
67. Boyer, W. (1995). Serotonin uptake inhibitors are superior to imipramine and alprazolam in alleviating panic attacks: a meta-analysis. *International Clinical Psychopharmacology*, **10**, 45–9.
68. Briggs, A.C., Stretch, D.D., and Brandon, S. (1993). Subtyping of panic disorder by symptom profile. *The British Journal of Psychiatry*, **163**, 201–9.
69. Shear, M.K., Brown, T.A., Barlow, D.H., *et al.* (1997). Multicenter collaborative panic disorder severity scale. *The American Journal of Psychiatry*, **154**, 1571–5.
70. Ballenger, J.C. (1992). Medication discontinuation in panic disorder. *The Journal of Clinical Psychiatry*, **53**, 26–31.
71. Ballenger, J.C. (1988). Comorbidity of panic and depression: implications for clinical management. *International Journal of Clinical Psychopharmacology*, **13**(Suppl. 4), S13–17.

4.8

Obsessive–compulsive disorder

Joseph Zohar, Leah Fostick, and Elizabeth Juven-Wetzler

Introduction

Obsessive–compulsive disorder (OCD) is a common, chronic, and disabling disorder marked by obsessions and/or compulsions that are egodystonic and cause significant distress to the patients and their families. During the last 25 years, there has been a resurgence of studies into various aspects of OCD, including epidemiological, pathophysiological, and pharmacological investigations. With the progress in finding effective treatments for OCD, different algorithms for the management of these patients have been developed. The progress in OCD includes advanced methodologies of imaging studies (both before and after treatment), along with insight into the neurological aspects of OCD and OCD-related conditions, leading to selective treatments.

Up to the early 1980s, OCD was considered a rather rare, treatment-refractory, and chronic condition of psychological origin. Dynamic psychotherapy was of little benefit and several pharmacological treatments were attempted without much success.[1] Since then, several researchers have reported that the prevalence of OCD is around 2 per cent in the general population.[2,3] In addition, numerous studies have reported on the efficacy of various serotonin reuptake inhibitors, and consequently an understanding of the biological basis of OCD has begun to unfold.

The observation that clomipramine, a tricyclic antidepressant with a serotonergic profile, is effective in treating symptoms of OCD[4,5] has increased interest in OCD in general and in the relationship between serotonin and OCD in particular. Substantial evidence currently suggests that OCD is almost unique among psychiatric disorders, as only serotonergic medications appear to be effective in this disorder.[6] For example, non-serotonergic drugs, such as desipramine, a potent antidepressant and antipanic agent, are entirely ineffective in OCD.[7–9] This specific response to serotonergic drugs has paved the way for further research on the role of serotonin in the pathogenesis of OCD in particular, and in OCD-related disorders in general.

Epidemiology

The lifetime prevalence of OCD in the general population is between 2 and 3 per cent (i.e. it is more prevalent than schizophrenia).[2,10] This rate has been confirmed across different cultures.[3] The prevalence of OCD among children and adolescents appears to be as high as among adults.[11] However, Nelson and Rice[12] and Stein *et al.*[13] have suggested that the diagnosis of OCD by the Diagnostic Interview Schedule administered by lay people leads to overdiagnosis, and so have proposed lower prevalence rates of 1 to 2 per cent.

Men and women are equally likely to be affected, although some reports have suggested a slight female predominance.[3] During adolescence, boys are more commonly affected than girls. The mean age of onset is about 20 years of age. Single people are more commonly affected, probably representing the difficulty for people with OCD to maintain a relationship.

Patients with OCD are commonly afflicted by other mental disorders; for instance, the lifetime prevalence for a major depressive episode in these patients is around 67 per cent.[3,14] Other common comorbid psychiatric diagnoses include alcohol-use disorders, social phobia, specific phobia, panic disorder, eating disorders, and post-traumatic stress disorder (PTSD).[15] The comorbidity with schizophrenia and with tic disorders raises interesting pathophysiological and therapeutic implications. The rate of tic disorders approaches 40 per cent in juvenile OCD, and there is an increase in the prevalence of Tourette's syndrome among the relatives of OCD patients.[16]

The relationship between OCD and obsessive–compulsive personality disorder (OCPD) has been a focus of debate. Although prospective research is lacking, it appears that OCPD is not a prominent risk factor for developing OCD, as the prevalence of OCPD among patients with OCD is not far from its prevalence in other psychiatric disorders.

Clinical features and diagnosis

The diagnosis of OCD according to DSM-IV criteria is based on the presence of either obsessions or compulsions, which cause marked distress, are time-consuming (more than an hour per day), or significantly interfere with the person's normal routine and social and occupational activities. It stipulates that, at some point during the course of the disorder, but not necessarily during the current episode, the person has recognized that the obsessions or compulsions are excessive or unreasonable. However, if the patient does not recognize for most of the time during the current episode that the obsessions and compulsions are excessive or unreasonable, the diagnosis is OCD with poor insight.

If another Axis I disorder is present, it is mandatory that the content of the obsessions or compulsions is not restricted to it (e.g. a preoccupation with food or weight in eating disorders, or guilt feelings in the presence of a major depressive episode). The disturbance should not be due to the direct effects of a substance (e.g. of a drug abuse or a medication) or a general medical condition.

The obsessions are recurrent, intrusive, and distressing thoughts, images, or impulses, whereas the compulsions are repetitive, seemingly purposeful, behaviours that a person feels driven to perform. Obsessions are usually unpleasant and increase a person's anxiety, whereas carrying out compulsions reduces anxiety. Resistance to carrying out a compulsion results in increased anxiety. The patient usually realizes that the obsessions are irrational and experiences both the obsession and the compulsion as egodystonic.

Patients with both obsessions and compulsions constitute at least 75 per cent of the affected patients, with most patients presenting with multiple obsessions and compulsions. The symptoms may shift, for example a patient who had washing rituals during childhood may present with checking rituals as an adult.

OCD can express itself in many different symptoms, but the classical presentations include washing, checking, aggressive, religious, or sexual obsessions, and ordering, counting, hoarding, and symmetry compulsions. Dimensional approaches have been used to analyze these characteristic subtypes, and present the different symptoms in an innovative way.[17]

The most common pattern is an obsession with dirt or germs, followed by washing or avoiding presumably contaminated objects (doorknobs, electrical switches, newspapers, people's hands, telephones). Because it is hard to avoid the feared object is (e.g. faeces, urine, dust, or germs), patients wash their hands excessively and sometimes avoid leaving home because of their fear of germs. A second common pattern is an obsession of doubt, followed by a compulsion of checking. The person checks whether the oven is turned off or the front and back doors are closed—the checking may involve many trips back home to recheck what had already been checked. In OCD, the checking, instead of resolving uncertainty, often contributes to even greater doubt, which leads to further checking. The patients exhibit obsessional self-doubt, and feel guilty for having committed some damage (for instance, a fear of hurting someone while driving, leading to driving back over the same spot again and again). Other patterns include hoarding and religious obsessions. More recently, a dimensional approach to OCD has been launched by Leckman and colleagues, stating four symptom dimension of OCD: obsessions/checking, symmetry/ordering, contamination/cleaning, and hoarding.[17,18] Another pattern of OCD involves intrusive obsessional thoughts without a compulsion. Such obsessions are usually repetitious thoughts of some sexual or aggressive act that is reprehensible to the patient. Still another pattern is the need for symmetry or precision, which leads to a compulsion of slowness. Patients can take hours to eat a meal or shave, in an attempt to do things 'just right'. Unlike other patients with OCD, these patients usually do not resist their symptoms.

The gap between the knowledge that the symptoms are irrational on one hand and the overwhelming urge to perform them on the other hand contributes to the immense suffering associated with OCD.

OCD and schizophrenia

About 25 per cent of patients with chronic schizophrenia may also present with OCD symptoms (range 5 to 45 per cent)[19]; and 15 per cent of the patients with schizophrenia may fully qualify for the diagnosis of OCD. As in OCD, the OC symptoms in these patients will not necessarily surface unless specific questions are asked. Many patients with schizophrenia can distinguish the egodystonic, obsessive–compulsive symptoms, perceived as coming from within, from the egosyntonic delusions perceived as introduced from the outside. Follow-up studies demonstrate a diagnostic stability over the years, and it seems that the presence of OCD in schizophrenia predicts a poor prognosis.[19] Several studies among patients with schizophrenia and OCD reported an improvement in OCD symptomatology after the addition of serotonin reuptake inhibitors.[19]

The poor prognosis of patients with schizophrenia and OCD, preliminary data regarding their response to the unique combination of antipsychotic and anti-obsessive medications, along with the high prevalence of this presentation has led several researchers to suggest that a 'schizo-obsessive' category may be of value.[20]

Differential diagnosis

Personal distress and functional impairment, which are required for the diagnosis, differentiate OCD from ordinary or mildly excessive worries, thoughts, and habits. The medical differential diagnosis includes tic disorders (especially Tourette's syndrome), temporal-lobe epilepsy, trauma, and postencephalitic complications.

Psychiatric diagnoses that should be ruled out include depressive disorder, schizophrenia, OCPD, PTSD, phobias, delusions, hypochondriasis, and paraphilias. OCD can usually be differentiated from schizophrenia by the absence of other schizophrenic symptoms and by the patients' insight into their disorder. Moreover, patients with OCD usually attempt to resist the obsessions. OCPD does not have the degree of functional impairment characteristic of OCD and it is egosyntonic.

Phobias are distinguished by the absence of a relationship between the obsessive thoughts and the compulsions. The fears in OCD usually involve harm to others rather than harm to oneself. In addition, in OCD, when patients are 'phobic' they are usually afraid of an unavoidable stimulus (for instance, viruses, germs, or dirt) as opposed to the classic phobic objects like tunnels, bridges, or crowds.

Major depressive disorder (MDD) can sometimes be associated with obsessive ideas, but patients with OCD usually fail to meet all the criteria of MDD. Other psychiatric diagnoses closely related to OCD are hypochondriasis, body dysmorphic disorder, and trichotillomania. As these patients have repetitive worries or behaviours, although they are focal, they are still related to the 'OCD Spectrum'.

Course and prognosis

Many patients with OCD may have an onset of symptoms after a stressful event (e.g. pregnancy, a loss, or a sexual problem). Owing to the secretive nature of the disorder, there is often a delay of 5 to 10 years before patients come to psychiatric attention. However, the delay may shorten due to increased public awareness to the disorder through articles, books, and movies. The course of OCD is

usually long, but variable; some patients experience a fluctuating course, while others experience a chronic course.[21]

About 20 to 30 per cent of the patients show a significant improvement in their symptoms, and 40 to 50 per cent a moderate improvement. The remaining 20 to 40 per cent become chronic or their symptoms worsen.

OCD patients are prone to depression and sometimes even to suicide. A poor prognosis is indicated by yielding (rather than resisting) to compulsions, a early onset, male gender, tic related forms of OCD with associated to hoarding/symmetry compulsions, the need for hospitalization, psychotic features, a coexisting major depressive disorder, delusional beliefs, the presence of overvalued ideas (i.e. some acceptance of the obsessions and compulsions), and the presence of personality disorder (especially schizotypal personality disorder).[22–24] A good prognosis is indicated by good social and occupational adjustment and less avoidance.[21] The obsessional content does not seem to be related to the prognosis, except for hoarding, which is usually considered to have a less favourable outcome.

Aetiology

Neurotransmitters

Many clinical trials of various serotonergic drugs lend support to the hypothesis that a dysregulation of serotonin is involved in the beneficial therapeutic effect in OCD. However, this does not necessarily reflect on pathogenesis. Abnormality of the serotonergic system, and particularly the hypersensitivity of postsynaptic 5-HT receptors, constitutes the leading hypothesis for the underlying pathophysiology of OCD.[7,25–38] However, a potential role for dopamine has been emerging as well.[39]

Clinical studies have assayed cerebrospinal levels of serotonin metabolites (e.g. 5-hydroxyindoleacetic acid [5-HIAA] a 5-HT metabolite that serves as an index of 5-HT turnover)[25,26] and affinities of imipramine and paroxetine[40] binding sites on platelets show that it binds to serotonin reuptake sites,[27–30] in some studies of OCD patients. A study supporting the relationship between a decreased function of the serotonergic system and a positive response to selective serotonin reuptake inhibitors (SSRIs), demonstrated normalization of the number of platelet 5-HT transporters following treatment with different SSRIs.[31] In an earlier study, patients who responded to clomipramine had higher pretreatment levels of 5-HIAA than the non-responders.[25] Moreover, the clinical improvement was positively correlated with a decrease in the concentration of 5-HIAA in cerebrospinal fluid.[25]

Another approach is to examine peripheral measures of serotonergic and noradrenergic function in patients with OCD. In one study, clinical improvement during clomipramine therapy closely correlated with pretreatment platelet serotonin concentration and monoamine oxidase activity, as well as with the decrease in both measures during clomipramine administration.[32] Moreover, only the plasma levels of clomipramine (a potent 5-HT reuptake inhibitor), but not the plasma levels of its primary metabolite, desmethyl clomipramine (which has noradrenergic properties), correlated significantly with an improvement in OCD symptoms. These findings suggest that the effects of anti-obsessive medications, clomipramine in this study, on serotonin function are pertinent to the anti-obsessional action observed.

Additional support for the importance of serotonin in the therapeutic response to serotonin reuptake inhibitors (SRIs) in OCD came from a study by Benkelfat *et al.*[38] in which the investigators administered the serotonin receptor antagonist metergoline and placebo to 10 patients with OCD in a double-blind crossover study. Patients receiving clomipramine on a long-term basis responded with greater anxiety to a 4-day administration of metergoline when compared with the placebo phase of the study.

Additional evidence for disturbances of the serotonergic system in OCD was provided by challenge studies. Challenges with L-tryptophan,[33] *m*-chlorophenylpiperazine (mCPP),[7,34] sumatriptan (a 5-HT1D agonist[6]), ipsapirone (a 5-HT1A receptor ligand[35]), and MK-212 (a 5-HT1A and 5-HT2C agonist[36]), among others, were used to evaluate whether they worsen obsessive–compulsive symptoms or whether they elicit different physiological responses (thermal or neuroendocrine) in patients with OCD compared with controls. Only two compounds (*m*-chlorophenylpiperazine and sumatriptan) have shown behavioural hypersensitivity and neuroendocrine hyposensitivity to be characteristic of serotonergic challenges in patients with OCD. These studies may have the potential to pinpoint the receptor subtype involved in OCD, raising the possibility that 5-HT_{1B} (but not 5-HT_{1A}) could be involved in OCD.[37]

Dopamine

The most compelling evidence for dopaminergic involvement in OCD comes from the abundance of OCD symptoms in basal ganglia disorders, such as Tourette's syndrome, Sydenham's chorea, and postencephalitic parkinsonism. The therapeutic benefits obtained with the coadministration of dopamine blockers and SRIs in a subset of patients with both OCD and tic disorder[41] has also suggested a role for dopamine dysfunction. A study evaluating levels of platelet sulphotransferase, an enzyme involved in the catabolism of catecholamines (providing a marker of presynaptic dopamine function), reported a decreased level of platelet [^{3}H]imipramine binding and a parallel increase in the level of sulphotransferase activity in OCD compared with controls. This provides further support for the hypothesis of reduced 5-HT activity and increased dopamine transmission in OCD.[28,39]

Immune factors

Study of autoimmune factors has been prompted by the association of OCD and the autoimmune disease of the basal ganglia, Sydenham's chorea. This complication of rheumatic fever is accompanied by obsessive–compulsive symptoms in over 70 per cent of cases[42]: 10 out of 11 children had antibodies directed against the caudate.[42] These children had a history of obsessive–compulsive symptoms, which started prior to the onset of the chorea, reached a peak in line with the motor symptoms, and declined with their resolution. This is consistent with the hypothesis of basal ganglia dysfunction in OCD.

Antibodies against two peptides of the basal ganglia have also been found.[43] A strong connection was reported between OCD/Tourette's syndrome and the B-cell antibody D8/17, which is another antibrain antibody.[44] The specificity of these antibodies to OCD, as well as the generalizability of these rare cases, is as yet unclear.

Brain imaging studies

The use of positron emission tomography has demonstrated the presence of increased activity (i.e. metabolism and blood flow) in the frontal lobes, the basal ganglia (especially the caudate nucleus), and the cingulum of patients with OCD.[45] Pharmacological and behavioural treatments reportedly reverse those abnormalities.[46] The data from functional imaging studies are consistent with the data from structural brain imaging studies. Both CT and magnetic resonance imaging studies have found decreased sizes of caudate bilaterally. Both functional and structural imaging procedures are also consistent with the observation that neurological procedures involving the cingulum are sometimes effective in the treatment of patients with OCD.

Overall, the brain imaging research suggests a role for the prefrontal cortex-basal ganglia thalamic circuitry. Dysfunction of these circuits can be explored by neuropsychological testing and recording evoked potentials. Indeed, a study of patients with OCD demonstrated that they are slower in performing tasks involving frontocortical systems, suggesting alterations at this level.[47] An evoked potential study showed enhanced processing negativity in the frontal cortex consistent with the prefrontal hyperactivity shown in brain imaging studies.[48] Moreover, the reflection of behavioural challenge on brain activity (brain responsivity) may be a potential tool for predicting a response to successful intervention with SSRI.[49]

Genetics

A significantly higher concordance rate was found for monozygotic twins than for dizygotic twins.[50] Of the first-degree relatives of patients with childhood-onset OCD, 35 per cent are also afflicted with the disorder.[51] Although this high rate is possibly related to the early-onset subtype, it nevertheless suggests a genetic component in OCD. Genetic research has yet to find abnormalities at the 5-HT transporter gene level. A study exploring the polymorphism of the promoter region of the gene for the 5-HT transporter failed to identify any differences between patients with OCD and controls.[52] However, several studies found polymorphism of $5HT_{1B\beta}$ in OCD,[53,54] hence providing further support for the $5HT_{1B}$ involvement in OCD.

Other biological data

Sleep electroencephalography and neuroendocrine studies have found abnormalities similar to those seen in depression, such as decreased rapid eye movement latency, non-suppression on the dexamethasone suppression test, and decreased growth hormone secretion with clonidine infusions.[55,56]

Behavioural factors

According to the learning theory, obsessions can be considered conditioned stimuli. When a relatively neutral stimulus is coupled with an anxiety-provoking stimulus, through conditioning, it will produce anxiety even when presented alone. In this regards, even the thought of the anxiety-provoking stimulus can cause anxiety, similarly to Pavlov's dog, which salivated even before he actually had food. Consequently, avoidant behaviour is being adopted in order to avoid the anxiety-provoking stimulus and any other stimuli, which remind it. The compulsions are learnt as a way to reduce anxiety. Once producing a relief of the anxiety, the relief serves as reinforce to the compulsion, which are then being repeated by the patient. Through the process of conditioning, reward and reinforcement, rituals, and avoidant strategies are become fixed.

Psychological factors

The dynamic aspects of OCD were first described by Sigmund Freud, who coined the term 'obsessional neurosis'. The disorder was thought to result from a regression from the Oedipal phase to the anal phase, with its characteristic ambivalence. The coexistence of hatred and love towards the same person leaves the patient paralyzed with doubt and indecision. Freud originally suggested that obsessive symptoms result from unconscious impulses of an aggressive or sexual nature. These impulses cause extreme anxiety, which is avoided by the defence mechanisms. One of the striking features of patients with OCD is the degree to which they are preoccupied with aggression or cleanliness (anal phase), either overtly in the content of their symptoms or in the underlying associations.

Freud described three major psychological defence mechanisms that are important in OCD: isolation, undoing, and reaction formation. According to the psychoanalytical formulation, OCD develops when these defences fail to contain the anxiety. Isolation is the separation of the idea and the affect that it arouses. Undoing is a secondary defence to combat the impulse and quit the anxiety that its imminent eruption into consciousness arouses. Undoing is a compulsive act, performed to prevent or undo the results that the patient irrationally anticipates from a frightening obsessional thought or impulse. Reaction formation is related to the production of character traits rather than symptom formation (characteristic of the above defences). The trait seems highly exaggerated and inappropriate (i.e. the switch of anger and hate into exaggerated love and dedication).

Summary

The efficacy of the SRIs for OCD, together with the lack of efficacy of adrenergic antidepressants, has suggested that serotonin is involved in the pathophysiology of OCD. This relationship was validated by research on serotonergic markers in OCD and by the challenge paradigm.[6] Which type of serotonergic receptor is involved in the pathogenesis and/or the mechanism of action of anti-obsessional drugs, is still unclear. However, the possible role of $5HT_{1B}$ has emerged. Further studies are crucial for elucidating the role of serotonin and other neurotransmitters (i.e. dopamine) in the pathophysiology and management of OCD.

The pharmacological treatment of OCD

Since the early 1980s, several potent SRIs have been studied extensively in OCD. Aggregate statistics for all SRIs suggest that 70 per cent of treatment-naive patients will improve at least moderately.[57]

Efficacy of serotonergic versus adrenergic antidepressants

Whilst anecdotal reports have suggested that clinical benefit can be obtained with a range of reuptake blockers, effectiveness has only been demonstrated consistently for the SRIs. Several studies have directly compared clomipramine with other antidepressants with a consistent finding: antidepressant drugs that are less potent SRIs than clomipramine are generally ineffective in OCD.[7–9,25]

In the late 1960s, clomipramine was the first reported effective medication for OCD.[4,5] Since then, numerous placebo-controlled studies have clearly shown clomipramine's effectiveness, and this has been confirmed in a United States multicentre controlled trial ($n = 520$).[58] In this study, after 10 weeks of treatment, 58 per cent of patients treated with clomipramine rated themselves much or very much improved versus 3 per cent of placebo-treated patients.

Besides the SRI clomipramine, the newer non-tricyclic SSRIs, such as fluoxetine, fluvoxamine, paroxetine, sertraline, citalopram, and escitalopram are gaining acceptance as effective alternatives for the treatment of OCD in controlled studies. Actually, they were found to be as effective as clomipramine.[59–62] Since SSRIs are less toxic in case of overdose, and as they have less cholinergic side-effects, they are considered as a first-line treatment for OCD.

(a) Onset of treatment response

It has been suggested that a relatively long period, up to 8 or even 12 weeks, is needed before one can consider a serotonin reuptake inhibitor to be ineffective. Several months' treatment is often needed to achieve a maximum response.

(b) Long-term treatment

Most patients relapse after prematurely discontinuing treatment, but, as stated above, it may take many months for a maximum response to be seen. Pato *et al.*[63] reported that 16 out of 18 patients with OCD relapsed within 7 weeks after stopping clomipramine, although some had been treated for more than a year (mean, 10.7 months). All patients regained the therapeutic effects when clomipramine was reintroduced. Leonard *et al.*[9] examined the effect of clomipramine substitution during a long-term clomipramine treatment in 26 children and adolescents with OCD (mean duration of treatment was 17 months). Half the patients were blindly assigned to 2 months of desipramine treatment, and then clomipramine was reintroduced. Almost 90 per cent relapsed during the 2 months' substitution period compared with only 18 per cent of those kept on clomipramine throughout the study. Therefore, it seems advisable that patients with OCD should be maintained on anti-obsessive medications for a long period, and certainly for more than a year before a very gradual attempt is made to discontinue the treatment.

The maintenance dose needed in OCD is also unclear. In a study that examined this issue, Mundo *et al.*[64] investigated the effect of dose reduction in patients previously treated successfully with fluoxetine. Patients were randomized to receive the same drug dosage or to receive a reduced dose. It appears that 'the dose that makes you well keeps you well'; i.e. that medium to high doses of SRIs are required.

(c) Drug dosage

Higher doses of SSRIs have been used in the treatment of OCD as compared to treatment of depression. Two fixed-dose studies using fluoxetine and one pan-European study with paroxetine have found some advantage with using higher doses, and those effects were found with citalopram and escitalopram studies.[6,37,62,65–67] A theoretical basis for this clinical finding, which is related to the 'stickiness' of the $5HT_{1B}$ receptor has been reported.[67,68]

Comparative studies of clomipramine versus SSRIs

The introduction of SSRIs has raised the question regarding the comparative efficacy of clomipramine versus that of the SSRIs. SSRIs are important alternatives to clomipramine, since their range of side-effects is different (absence of anticholinergic side-effects, sedation, safety with overdose, etc.). Although SSRIs may be associated with sexual side-effects, headaches, and appetite disturbances, these side-effects are usually less troublesome as compared to clomipramine's side-effects.

Fluoxetine was compared with clomipramine in 11 patients with OCD in a 10-week crossover study.[69] Although no significant differences were noted regarding clinical efficacy, the proportion of fluoxetine non-responders who later responded to clomipramine tended to be higher compared with the clomipramine non-responders who were switched to fluoxetine. However, patients reported significantly fewer side-effects while on fluoxetine. Freeman *et al.*[70] compared the efficacy of fluvoxamine and clomipramine in a multicentre randomized double-blind parallel-group comparison in 66 patients. Both drugs were equally effective and well tolerated, but fluvoxamine produced fewer anticholinergic side-effects and caused less sexual dysfunction than clomipramine, but more reports of headache and insomnia.

Paroxetine was of comparable efficacy to clomipramine and both were significantly more effective than placebo in a multinational double-blind placebo-controlled parallel group study of 399 patients with OCD.[71] Bisserbe *et al.*[72] reported that sertraline (50–200 mg/day) was significantly more effective than clomipramine (50–200 mg/day) in a double-blind study ($n = 160$).

Other pharmacological approaches and neurosurgery

Considered as one of the anxiety disorders according to DSM-IV (but not according to the ICD-10), it is not surprising that anxiolytics have been suggested for the treatment of patients with OCD. Thus, clonazepam has been reported as efficient in several uncontrolled studies and case series, and even in a small double-blind randomized multiple crossover study. However, since OCD is a chronic disorder, the long-term use of anxiolytics raises questions of dependency.

Despite reports in open studies regarding the efficacy of trazodone, buspirone, and lithium, results from double-blind studies proved negative. Adding drugs that affect dopamine function, especially the atypical antipsychotics (i.e. risperidone or quetiapine), to SRI therapy in patients with treatment-resistant OCD, may result in improvement for patients with a personal or family history of tics.[53] A combination of SSRI and small doses of high potent dopamine blocker (haloperidol or pimozide) was found to be useful for both the tics and the OC symptoms.[73]

Neurosurgery and Deep Brain Stimulation (DBS) have been reported to be effective in some patients with OCD. Neurosurgery involves procedures that disconnect the outflow pathways originating from the orbitofrontal cortex. Cingulotomy can help some intractable patients, but although the immediate results may be striking, the long-term prognosis is more reserved.[74] As for DBS, initial reports are optimistic[75] but as the total number of patients who underwent this procedure is very small, its efficacy needs to be further elucidated.

Summary of drug treatment for OCD

The first-line treatment consists of either an SSRI or clomipramine. Any one of the six SSRIs (fluoxetine, fluvoxamine, paroxetine, sertraline, citalopram, escitalopram) in current use constitutes an effective and safe choice, but choosing which SSRI depends on the

drug's pharmacokinetic profile, as well as the physician's familiarity with the drug. The dose should be higher than that used for treating depression (e.g. 40–60 mg of fluoxetine) and the trial should last at least 12 weeks. If clomipramine is chosen, cardiovascular problems and closed-angle glaucoma should first be ruled out. Doses of 200 to 300 mg of clomipramine are needed, but titration should last for 1 to 3 weeks and the optimal dose should be continued for at least 12 weeks before determining a lack of response.

If the patient cannot tolerate the first drug (an SSRI) or did not respond, a trial of another SSRI or CMI is advised or augmentation (i.e. risperidone) is recommended.

The third stage in non-responders, and in cases of only a partial response, includes small doses of antipsychotics (especially in Tourette's syndrome), or the addition of lithium or trazodone, buspirone, or tryptophan. The fourth stage consists of atypical neuroleptics, thyroid supplementation, clonidine, a monoamine oxidase inhibitor, and intravenous clomipramine. In truly severe and resistant cases, neurosurgery or DBS could be tried.

Psychological approaches

The effect of a psychodynamic approach in OCD is limited, whereas modern interventions like cognitive and behavioural therapy show promising results.[76] Behavioural therapy was found to be effective in OCD,[76,77] and some data indicate that the beneficial effects of behavioural therapy may be longer lasting.[78] About two-thirds of patients with moderately severe rituals can be expected to improve substantially, but not completely. A combination of behavioural therapy and pharmacotherapy may constitute the optimal treatment for OCD. Recently, two neuroimaging studies found that patients with OCD who are successfully treated with behavioural therapy show changes in cerebral metabolism similar to those produced by successful treatment with SRIs.[47,79]

Behavioural therapy can be conducted in in- and outpatient settings. The principal behavioural approaches in OCD are exposure for obsessions and response prevention for rituals (see Chapter 6.3.2.1). Desensitization, thought stopping, flooding, implosion therapy, and aversion conditioning have also been used in patients with OCD. In behavioural therapy the patient must collaborate and perform assignments. In a study of 18 patients with OCD, those who received exposure and response prevention therapy showed significant improvement, whereas patients on a general anxiety management intervention (control) showed no improvement from baseline.[80] Direct comparisons of behavioural therapy and pharmacotherapy are few and are limited by methodological issues.

In thought stopping, the patient (or initially the therapist) shouts 'stop' or applies an aversive stimulus to counteract the obsessional preoccupation. The patient may also imagine a stop sign with a police officer nearby or another image that evokes inhibition at the same time that he or she recognizes the presence of the obsession. Another technique is to 'postpone' the thought until a specified time (e.g. an hour later) and not to think about it until then.

Psychological factors might be of considerable benefit in understanding what precipitates exacerbations of the disorder and in treating various forms of resistances to treatment, such as non-compliance to medications or to homework assignments. It is important to remember that the symptoms may have important psychological meanings that make patients reluctant to give them up.

In the absence of controlled studies of insight-oriented psychotherapy for OCD, the anecdotal reports reporting lasting change do not allow generalizations to be made regarding efficacy. Also, the efficacy of medications in producing quick improvement has rendered slow and long-term psychotherapy out of favour.

Supportive psychotherapy has a non-specific place in managing patients with OCD, and may help patients improve their functioning and adjustment. The management plan should also include attention to the family members through the provision of emotional support, reassurance, explanation, and advice on how to manage and respond to the patient. Family therapy may reduce marital discord and build a treatment alliance, as well as helping in the resistance to compulsions. Group therapy is useful as a support system for some patients.

Summary

The treatment of OCD was characterized by pessimism until 25 years ago when effective treatments including behavioural therapy and the serotonin reuptake inhibitors were developed. Although introduced for OCD in 1967, it was only in the 1980s that double-blind studies confirmed the efficacy of clomipramine, an SRI. This was followed by the introduction of the selective serotonin reuptake inhibitors, which also proved effective for OCD. The anti-obsessive activity of these drugs was found to be independent from the drugs' antidepressant effect, as established by their efficacy both in depressed and non-depressed patients. Overall, serotonergic therapies have provided a better outlook for these patients and have contributed to our understanding of the pathophysiology of OCD.[6,7] Previously thought to be a rare and untreatable disorder, OCD is now recognized as common, and there is good reason to expect that patients with OCD will benefit substantially from potent SRIs and behaviour therapy.

Many patients with OCD do not seek treatment and the disease tends to be chronic. There is about a 10-year lag between the onset of symptoms and the seeking of professional help due to feelings of embarrassment. Further delay ensues until the diagnosis and correct treatment are given.[81] Census data suggest that over $8 billion are spent in the United States each year on the management of OCD, one-fifth of that spent on cardiac disease.[82] Because patients with OCD often attempt to conceal their symptoms, it is incumbent on clinicians to screen for OCD in every mental status examination, since appropriate treatment can result in improved quality of life, reduced OCD chronicity, and a decrease in cost to the individual and society.

Further information

Kasper, S., Zohar, J., and Stein, D.J. (2002). *Decision making in psychopharmacology*. Martin Dunitz, London.

Montgomery, S. and Zohar, J. (1999). *Obsessive compulsive disorder*. Martin Dunitz, London.

Zohar, J., Hollander, E., Stein, D.J., and Westenberg, H.G. (2007). The Cape Town consensus group. From obsessive-compulsive spectrum to obsessive-compulsive disorders: the Cape Town consensus statement. *CNS Spectrums*, **12**(2 Suppl. 3).

http://www.ocfoundation.org

References

1. Salzman, L. and Thaler, F.H. (1981). Obsessive compulsive disorder: a review of the literature. *The American Journal of Psychiatry*, **138**, 286–96.

2. Robins, L.N., Helzer, J.E., Weissman, M.M., *et al.* (1984). Lifetime prevalence of specific psychiatric disorders in three sites. *Archives of General Psychiatry*, **41**, 949–58.
3. Weissman, M.M., Bland, R.C., Canino, G.J., *et al.* (1994). The cross national epidemiology of obsessive compulsive disorder. *The Journal of Clinical Psychiatry*, **55**, (Suppl. 3), 5–10.
4. Renynghe de Voxrie, G.V. (1968). Anafranil (G34586) in obsessive compulsive neurosis. *Archives Belges de Neurologie*, **68**, 787–92.
5. Fernandez-Cordoba, E. and Lopez-Ibor, A.J. (1967). La monoclorimipramina en enfermos psiquiatricos resistenses a otros tratamientos. *Actas Lusoespañolas de Neurologia, Psiquiatria y Ciencias Afines*, **26**, 119–47.
6. Dolberg, O.T., Iancu, I., Sasson, Y., *et al.* (1996). The pathogenesis and treatment of obsessive-compulsive disorder. *Clinical Neuropharmacology*, **19**, 129–47.
7. Zohar, J. and Insel, T. (1987). Obsessive-compulsive disorder: psychobiological approaches to diagnosis, treatment, and pathophysiology. *Biological Psychiatry*, **22**, 667–87.
8. Goodman, W.K., Price, L.H., Delgado, P.L., *et al.* (1990). Specificity of serotonin reuptake inhibitors in the treatment of obsessive compulsive disorder: comparison of fluvoxamine and desipramine. *Archives of General Psychiatry*, **47**, 577–85.
9. Leonard, H., Swedo, S.E., Lenane, M.C., *et al.* (1991). A double-blind desipramine substitution during long-term clomipramine treatment in children and adolescents with obsessive-compulsive disorder. *Archives of General Psychiatry*, **48**, 922–7.
10. Karno, M., Golding, J.M., Sorenson, S.B., *et al.* (1988). The epidemiology of obsessive-compulsive disorder in five US communities. *Archives of General Psychiatry*, **45**, 1094–9.
11. Flament, M.F., Whitaker, A., Rapoport, J.L., *et al.* (1988). Obsessive compulsive disorder in adolescence: an epidemiological study. *Journal of the American Academy of Child and Adolescent Psychiatry*, **27**, 764–71.
12. Nelson, E. and Rice, J. (1997). Stability of diagnosis of obsessive compulsive disorder in the epidemiologic catchment area study. *The American Journal of Psychiatry*, **154**, 826–31.
13. Stein, M.B., Forde, D.R., Anderson, G., *et al.* (1997). Obsessive compulsive disorder in the community: an epidemiologic survey with clinical reappraisal. *The American Journal of Psychiatry*, **154**, 1120–6.
14. Rasmussen, S.A. and Eisen, J.L. (1992). Epidemiology and clinical features of obsessive-compulsive disorder. In *Obsessive compulsive disorders. Theory and management* (eds. M.A. Jenike, L. Baer, and W.E. Minichiello), pp. 10–27. Year Book, Chicago, IL.
15. Sasson, Y., Dekel, S., Chopra, M., *et al.* (2005). Posttraumatic obsessive compulsive disorder—a case series. *Psychiatry Research*, **135**, 145–52.
16. Pauls, D. (1992). The genetics of OCD and Gilles de la Tourette's syndrome. *The Psychiatric Clinics of North America*, **15**, 759–66.
17. Mataix-Cols, D., Rosario-Campos, M.C., and Leckman, J.F. (2005). A multidimensional model of obsessive-compulsive disorder. *The American Journal of Psychiatry*, **162**, 228–38.
18. Leckman, J.F., Grice, D.E., Boardman, J., *et al.* (1997). Symptoms of obsessive-compulsive disorder. *The American Journal of Psychiatry*, **154**, 911–17.
19. Berman, I., Sapers, B.L., Chang, H.H.J., *et al.* (1995). Treatment of obsessive-compulsive symptoms in schizophrenic patients with clomipramine. *Journal of Clinical Psychopharmacology*, **15**, 206–10.
20. Zohar, J. (1997). Is there room for a new diagnostic subtype—the schizo-obsessive subtype? *CNS Spectrums*, **2**, 49–50.
21. Ravizza, L., Maina, G., and Bogetto, F. (1997). Episodic and chronic OCD. *Depression and Anxiety*, **6**, 154–8.
22. Mataix-Cols, D., Wooderson, S., Lawrence, N., *et al.* (2004). Distinct neural correlates of washing, checking, and hoarding symptom dimensions in obsessive-compulsive disorder. *Archives of General Psychiatry*, **61**, 564–76.
23. Mataix-Cols, D. (2006). Deconstructing obsessive-compulsive disorder: a multidimensional perspective. *Current Opinion in Psychiatry*, **19**, 84–9.
24. Iraqi, Z., El Yazaji, M., Hjiej, H., *et al.* (2005). Obsessive and compulsive symptoms in schizophrenia. Presented in the XIII WPA congress, Cairo-Egypt.
25. Thoren, P., Asberg, M., Gronholm, B., *et al.* (1980). Clomipramine treatment of obsessive compulsive disorder. II. Biochemical aspects. *Archives of General Psychiatry*, **27**, 1289–94.
26. Insel, T.R., Mueller, E.A., Alterman, I., *et al.* (1985). Obsessive compulsive disorder and serotonin: is there a connection? *Biological Psychiatry*, **20**, 1174–88.
27. Weizman, A., Carmi, M., Hermesh, H., *et al.* (1986). High affinity imipramine binding and serotonin uptake in platelets of eight adolescent and ten adult obsessive-compulsive patients. *The American Journal of Psychiatry*, **143**, 335–9.
28. Marazziti, D., Hollander, E., Lensi, P., *et al.* (1992). Peripheral markers of serotonin and dopamine function in obsessive-compulsive disorder. *Psychiatry Research*, **42**, 41–51.
29. Vitiello, B., Shimon, H., Behar, D., *et al.* (1991). Platelet imipramine binding and serotonin uptake in obsessive-compulsive patients. *Acta Psychiatrica Scandinavica*, **84**, 29–32.
30. Kim, S.W., Dysken, M.W., Pandey, G.N., *et al.* (1991). Platelet 3H-imipramine binding sites in obsessive compulsive behavior. *Biological Psychiatry*, **30**, 467–74.
31. Marazziti, D., Pfanner, C., Palego, L., *et al.* (1997). Changes in platelet markers of obsessive compulsive patients during a double-blind trial of fluvoxamine versus clomipramine. *Pharmacopsychiatry*, **30**, 245–9.
32. Flament, M.F., Rapoport, J.L., Murphy, D.L., *et al.* (1987). Biochemical changes during clomipramine treatment of childhood obsessive-compulsive disorder. *Archives of General Psychiatry*, **44**, 219–25.
33. Charney, D.S., Goodman, W.K., Price, L.H., *et al.* (1988). Serotonin function in obsessive-compulsive disorder: a comparison of the effects of tryptophan and *m*-chlorophenylpiperazine in patients and healthy subjects. *Archives of General Psychiatry*, **45**, 177–85.
34. Hollander, E., DeCaria, C.M., Nitescu, A., *et al.* (1992). Serotonergic function in obsessive-compulsive disorder: behavioral and neuroendocrine responses to oral *m*-chlorophenylpiperazine and fenfluramine in patients and healthy volunteers. *Archives of General Psychiatry*, **49**, 21–8.
35. Lesch, K.P., Hoh, A., Disselkamp-Tietze, J., *et al.* (1991). 5-Hydroxytryptamine 1A receptor responsivity in obsessive compulsive disorder. Comparison of patients and controls. *Archives of General Psychiatry*, **48**, 540–7.
36. Bastani, B., Nash, J.F., and Meltzer, H.Y. (1990). Prolactin and cortisol responses to MK-212, a serotonin agonist, in obsessive-compulsive disorder. *Archives of General Psychiatry*, **47**, 833–9.
37. Sasson, Y. and Zohar, J. (1996). New developments in obsessive-compulsive disorder research: implications for clinical management. *International Clinical Psychopharmacology*, **11**(Suppl. 5), 3–12.
38. Benkelfat, C., Murphy, D.L., Zohar, J., *et al.* (1989). Clomipramine in obsessive compulsive disorder: further evidence for a serotonergic mechanism of action. *Archives of General Psychiatry*, **46**, 23–8.
39. Denys, D., Zohar, J., and Westenberg, H.G. (2004). The role of dopamine in obsessive-compulsive disorder: preclinical and clinical evidence. *The Journal of Clinical Psychiatry*, **65**(Suppl. 14), 11–7.
40. Marazziti, D., Dell'Osso, L., Presta, S., *et al.* (1999). Platelet [3H]paroxetine binding in patients with OCD-related disorders. *Psychiatry Research*, **89**, 223–8.

41. McDougle, C.J., Goodman, W.K., Leckman, J.F., *et al.* (1994). Haloperidol addition in fluvoxamine-refractory obsessive-compulsive disorder. A double-blind, placebo-controlled study in patients with and without tics. *Archives of General Psychiatry*, **51**, 302–8.
42. Swedo, S.E., Leonard, H.L., and Kiessling, L.S. (1994). Speculations on antineuronal antibody-mediated neuropsychiatric disorders of childhood. *Pediatrics*, **93**, 323–6.
43. Roy, B.F., Benkelphat, C., Hill, J.L., *et al.* (1994). Serum antibody for somatostatin, 14 and prodynorphin 209–240 in patients with obsessive-compulsive disorder, schizophrenia, Alzheimer's disease, multiple sclerosis and advanced HIV infection. *Biological Psychiatry*, **35**, 335–44.
44. Swedo, S.E., Leonard, H.L., Mittelman, B.B., *et al.* (1997). Identification of children with pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections by a marker associated with rheumatic fever. *The American Journal of Psychiatry*, **154**, 110–12.
45. Rauch, S.L. (1998). Neuroimaging in OCD: clinical implications. *CNS Spectrum*, **3**, (Suppl. 1), 26–9.
46. Baxter, L.R. Jr, Schwartz, J.M., Bergman, K.S., *et al.* (1992). Caudate glucose metabolic rate changes with both drug and behaviour therapy for OCD. *Archives of General Psychiatry*, **49**, 681–9.
47. Galderisi, S., Mucci, A., and Catapano, F. (1995). Neuropsychological slowness in obsessive-compulsive patients: is it confined to tests involving the fronto-subcortical systems? *The British Journal of Psychiatry*, **167**, 394–8.
48. Towey, J.P., Tenke, C.E., Bruder, G.E., *et al.* (1994). Brain event-related potential correlates of over focused attention in obsessive-compulsive disorder. *Psychophysiology*, **31**, 535–43.
49. Hendler, T., Lustig, M., Goshen, E., *et al.* (2003). Brain reactivity to specific symptom provocation indicates prospective therapeutic outcome in OCD. *Psychiatry research*, **124**, 87–103.
50. Rasmussen, S.A. and Tsuang, M.T. (1986). Clinical characteristics and family history in DSM-III obsessive-compulsive disorder. *The American Journal of Psychiatry*, **143**, 317–22.
51. Lenane, M.C., Swedo, S.E., Leonard, H., *et al.* (1990). Psychiatric disorders in first degree relatives of children and adolescents with obsessive-compulsive disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, **29**, 407–12.
52. Billet, E.A., Richter, M.A., King, N., *et al.* (1997). Obsessive compulsive disorder, response to serotonin reuptake inhibitors and the serotonin transporter gene. *Molecular Psychiatry*, **2**, 403–6.
53. Mundo, E., Richter, M.A., Zai, G., *et al.* (2002). 5HT1Dbeta receptor gene implicated in the pathogenesis of obsessive-compulsive disorder: further evidence from a family-based association study. *Molecular Psychiatry*, **7**, 805–9.
54. Camarena, B., Aguilar, A., Loyzaga, C., *et al.* (2004). A family-based association study of the 5-HT-1Dbeta receptor gene in obsessive-compulsive disorder. *The International Journal of Neuropsychopharmacology*, **7**, 49–53.
55. Insel, T.R., Gillin, J.C., Moore, A., *et al.* (1982). The sleep of patients with OCD. *Archives of General Psychiatry*, **39**, 1372–7.
56. Insel, T.R., Kalin, N.H., Guttmacher, L.B., *et al.* (1982). The dexamethasone suppression test in patients with primary OCD. *Psychiatry Research*, **6**, 153–8.
57. Rasmussen, S.A., Eisen, J.L., and Pato, M.T. (1993). Current issues in the pharmacological management of obsessive-compulsive disorder. *The Journal of Clinical Psychiatry*, **54**, 4s–9s.
58. Clomipramine Collaborative Study Group. (1991). Clomipramine in the treatment of patients with obsessive-compulsive disorder. *Archives of General Psychiatry*, **48**, 730–8.
59. Geller, D.A., Biederman, J., Stewart, S.E., *et al.* (2003). Which SSRI? A meta-analysis of pharmacotherapy trials in pediatric obsessive-compulsive disorder. *The American Journal of Psychiatry*, **160**, 1919–28.
60. Greist, J.H., Bandelow, B., Hollander, E., *et al.* (2003). World council of anxiety. WCA recommendations for the long-term treatment of obsessive-compulsive disorder in adults. *CNS Spectrums*, **8**, 7–16.
61. Kasper, S., Zohar, J., and Stein, D.J. (2002). *Decision making in psychopharmacology*. Martin Dunitz, London.
62. Blier, P., Habib, R., and Flament, M.F. (2006). Pharmacotherapies in the management of obsessive-compulsive disorder. *Canadian Journal of Psychiatry*, **51**, 417–30.
63. Pato, M.T., Zohar-Kadouch, R., Zohar, J., *et al.* (1988). Return of symptoms after discontinuation of clomipramine in patients with obsessive compulsive disorder. *The American Journal of Psychiatry*, **145**, 1521–5.
64. Mundo, E., Barregi, S.R., Pirola, R., *et al.* (1997). Long-term pharmacotherapy of obsessive-compulsive disorder: a double-blind controlled study. *Journal of Clinical Psychopharmacology*, **17**, 4–10.
65. March, J.S., Frances, A., Carpenter, D., *et al.* (1997). The expert consensus guideline series: treatment of obsessive-compulsive disorder. *The Journal of Clinical Psychiatry*, **58**(Suppl. 4), 1–72
66. Montgomery, S.A., Kasper, S., Stein, D.J., *et al.* (2001). Citalopram 20 mg, 40 mg and 60 mg are all effective and well tolerated compared with placebo in obsessive-compulsive disorder. *International Clinical Psychopharmacology*, **16**, 75–86.
67. El Mansari, M. and Blier, P. (2006). Mechanisms of action of current and potential pharmacotherapies of obsessive-compulsive disorder. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, **30**, 362–73.
68. Blier, P. and Szabo, S.T. (2005). Potential mechanisms of action of atypical antipsychotic medications in treatment-resistant depression and anxiety. *The Journal of Clinical Psychiatry*, **66**(Suppl. 8), 30–40.
69. Pigott, T.A., Pato, M.T., Bernstein, S.E., *et al.* (1990). Controlled comparisons of clomipramine and fluoxetine in the treatment of obsessive-compulsive disorder: behavioral and biological results. *Archives of General Psychiatry*, **47**, 926–32.
70. Freeman, C.P.L., Trimble, M.R., Deakin, J.F.W., *et al.* (1994). Fluvoxamine versus clomipramine in the treatment of obsessive compulsive disorder: a multi-center, randomized, double-blind, parallel group comparison. *The Journal of Clinical Psychiatry*, **55**, 301–5.
71. Zohar, J. and Judge, R. (1996). Paroxetine versus clomipramine in the treatment of obsessive-compulsive disorder. *The British Journal of Psychiatry*, **169**, 468–74.
72. Bisserbe, J.C., Lane, R.M., Flament, M.F., *et al.* (1997). A double-blind comparison of sertraline and clomipramine in outpatients with obsessive-compulsive disorder. *European Psychiatry*, **153**, 1450–4.
73. McDougle, J., Goodman, W.K., Price, L.H., *et al.* (1990). Neuroleptic addition in fluvoxamine-refractory obsessive-compulsive disorder. *The American Journal of Psychiatry*, **147**, 652–4.
74. Jenike, M.A., Baer, L., Ballantine, T., *et al.* (1991). Cingulotomy for refractory obsessive-compulsive disorder. *Archives of General Psychiatry*, **48**, 548–55.
75. Greenberg, B.D., Malone, D.A., Friehs, G.M., *et al.* (2006). Three-year outcomes in deep brain stimulation for highly resistant obsessive-compulsive disorder. *Neuropsychopharmacology*. **31**, 2384–93.
76. Van Balkom, A.J.L.M., De Haan, E., Van Oppen, P., *et al.* (1998). Cognitive and behavioral therapies alone versus in combination with fluvoxamine in the treatment of obsessive compulsive disorder. *The Journal of Nervous and Mental Disease*, **186**, 492–9.
77. Marks, I.M., Hodgson, R., Rachman, S., *et al.* (1975). Treatment of chronic obsessive-compulsive neurosis *in vivo* exposure: a 2-year follow-up and issues in treatment. *The British Journal of Psychiatry*, **127**, 349–64.

78. Greist, J.H. (1996). New developments in behaviour therapy for obsessive-compulsive disorder. *International Clinical Psychopharmacology*, **11**(Suppl.), 63–73.
79. Schwartz, J.M., Stoessel, P.W., Baxter, L.R. Jr., *et al.* (1996). Systematic changes in cerebral glucose metabolic rate after successful behavior modification treatment of OCD. *Archives of General Psychiatry*, **53**, 109–13.
80. Lindsay, M., Craig, R., and Andrews, G. (1997). Controlled trial of exposure and response prevention in obsessive-compulsive disorder. *The British Journal of Psychiatry*, **171**, 135–9.
81. Hollander, E. (1997). Obsessive compulsive disorder: the hidden epidemic. *The Journal of Clinical Psychiatry*, **12**(Suppl.), 3–6.
82. DuPont, R.L., Rice, D.P., Shiraki, S., *et al.* (1995). Economic costs of obsessive compulsive disorder. *Medical Interface*, **8**, 102–9.

4.9

Depersonalization disorder

Nick Medford, Mauricio Sierra,
and Anthony S. David

Introduction

Depersonalization, a term coined by Dugas in 1898,[1] is defined in DSM-IV as 'an alteration in the experience of self so that one feels detached from and as if one is an outside observer of one's outside mental processes or body'. Brief, self-limiting experiences of depersonalization commonly occur in healthy people in the context of fatigue, intense stress, or during/after intoxication with alcohol or illicit drugs. However, some people experience chronic depersonalization of a disturbing intensity, causing significant distress and impacting on quality-of-life and daily functioning. This may occur as a **primary depersonalization disorder (DPD)**, or in the context of other psychiatric or neurological conditions. In this chapter, we consider the primary disorder, although some sections are also relevant to secondary depersonalization.

The depersonalization experience is one of feeling strangely altered and unreal, in a way that sufferers often find very hard to convey. It is often accompanied by the related phenomenon of **derealization**, in which the person's surroundings are experienced as somehow remote and lacking immediacy and vibrancy, as if the world itself has become oddly unreal. Patients with persistent depersonalization and derealization often use the analogy of feeling as if they are on the set of a play or film, where nothing is real and they are acting out a role rather than living a real life.

Clinical features

The diagnosis requires the presence of persistent, distressing depersonalization and/or derealization, occurring in clear consciousness, and not due to another disorder or substance. Some patients find it impossible to divide their symptoms into depersonalization and derealization, seeing them as essentially two ways of describing the same experience. Nevertheless, one may encounter patients who describe one without the other. 'Pure' derealization is, however, uncommon.

In addition, there are a number of other symptoms that occur with sufficient frequency to be considered as part of the depersonalization syndrome, although their presence is not essential for making the diagnosis. These are as follows:

Desomatization—a loss or diminution of bodily sensation, sometimes accompanied by a feeling of disembodiment.

De-affectualization—a loss or diminution of emotional reactivity-the feeling that life has somehow been drained of emotional content, or that the sufferer feels little emotion in response to people or events that would normally be expected to elicit an emotional response. This may have significance for intimate relationships. It should be noted that de-affectualization is not usually accompanied by blunted affect of the type commonly seen in schizophrenia.

De-ideation—a feeling of mental emptiness which may cause difficulty in concentrating, a distorted experience of time, and a sense of detachment from memories. Often accompanied by a feeling of 'stuffiness in the head' or 'as if my brain has turned to cotton wool'.

While 'depersonalization' and 'derealization' are well-established in the psychiatric lexicon, the three terms listed above are not widely used or discussed. However, a recent analysis of symptoms reported by patients with DPD gave strong support to the idea that the condition should be considered as a syndrome, with symptoms occurring in domains corresponding to the terms used above.[2]

Classification

In DSM-IV, DPD is classified as a dissociative disorder, while in ICD-10 it falls under the vague heading of 'other neurotic disorders', and is not linked to any other category of disorder.

It has been argued that DPD is not truly a dissociative disorder, as dissociation is generally characterized by a lack of subjective awareness of change, whereas in DPD the experience of feeling changed is central. However, this apparent contradiction can be resolved if dissociation is conceptualized as a category incorporating both types of phenomenon.[3]

Diagnosis and differential diagnosis

The diagnosis should be established by a careful clinical assessment. Because DPD remains a somewhat obscure disorder, patients with this condition may have had unproductive consultations with other professionals and formed the impression that their symptoms are baffling, perhaps even unique. Being given the correct diagnosis and the opportunity to discuss it in depth with an informed psychiatrist may come as a great relief. The reassurance derived from this may in itself have a powerful therapeutic effect.

Where there are other psychiatric symptoms (e.g. anxiety, panic attacks, depressive features), the distinction between primary and secondary depersonalization may be difficult. The best way to approach this is simply to establish what the dominant symptoms are at the time of presentation. In a patient with a history of panic disorder who has developed severe unremitting depersonalization, and now has very infrequent panic attacks, the most pragmatic approach is to diagnose DPD. The fact that the panic symptoms preceded the onset of DPD is less important than the fact that it is now the DPD symptoms that dominate the clinical picture.

This issue aside, the main psychiatric differential hinges on the possibility that when patients describe feeling altered or unreal, they are articulating delusional beliefs. It is important to establish that patients have no psychotic symptoms; in particular, that they do not literally believe themselves to be unreal or dead, as this is suggestive of psychotic depression and the Cotard delusion, rather than DPD.

Depersonalization can also occur in neurological disorders, principally temporal lobe epilepsy and migraine.[4] Here the history is usually of brief, stereotyped episodes, with associated features that should provide sufficient clues to the underlying diagnosis.

Epidemiology

Until recently, DPD was considered rare, but contemporary epidemiological work suggests that it affects 1–2 per cent of the general adult population,[5] with a gender ratio of 1:1.

Symptom surveys suggest that depersonalization is perhaps the third commonest symptom (after anxiety and low mood) in psychiatric populations. It should be noted that these studies do not distinguish between primary and secondary depersonalization.

Aetiology

Various factors have been implicated in the genesis and maintenance of the condition. Biological and psychological issues are considered separately here, but should not be seen as mutually exclusive.

Psychological factors: Many patients with primary DPD have concurrent anxiety or mood symptoms, or a previous history of anxiety and/or panic attacks. The clinical impression is often that feelings of detachment and unreality have arisen as a defence against feeling anxious and threatened—a way of keeping a stressful world at a safer psychological distance. There is sometimes a history of DPD symptoms first occurring in the context of some particularly stressful event or period. Often, however, specific precipitants are not identifiable.

Once DPD develops, further anxiety may follow—patients may worry that, for example, the peculiar feeling of unreality is a sign that they are on the verge of mental breakdown. These concerns often manifest as obsessional rumination and self-monitoring, characterized by a compulsive checking of the inner state and a comparison of this state with some idealized standard of normality. This further anxiety can lead to reinforcement and perpetuation of depersonalization and derealization, so that symptoms feed each other in a 'vicious circle'.

Biological factors: There is objective biological evidence relating to the loss of emotional reactivity that patients with DPD commonly report. Patients with DPD show attenuated skin conductance responses to emotional stimuli,[6] while fMRI work suggests that the brain's emotional response circuitry is inhibited in DPD.[7,8]

Some 10–20 per cent of patients with DPD describe symptoms beginning during or after an episode of illicit drug use, cannabis being the drug most commonly implicated. Many illicit drugs are known to cause depersonalization phenomena acutely, but it is unclear how any drug might produce chronic symptoms persisting long after the drug is cleared from the system. It seems likely that the initial symptoms are due to the drug, but become chronic and unremitting through the kind of "vicious circle" outlined above.[9]

Course and prognosis

Two large case series[10,11] suggest that age of onset is usually in late adolescence or early adulthood, although up to a third of patients describe symptoms originating in childhood. Onset may be sudden or gradual, and symptoms thereafter may be episodic or continuous. Patients often report little or no fluctuation in symptom nature or severity, although with close questioning it is often possible to establish that certain factors (e.g. stress, fatigue) worsen the symptoms.

Symptoms are often unremitting for many years. In the largest case series to date,[10] patients had been symptomatic for a mean of 13.9 years at the time of initial presentation to a specialist DPD clinic. This striking statistic may, in part, reflect the widespread lack of familiarity with the condition and consequent delays in diagnosis.

While there is little available data on which to base predictions about prognosis, symptoms that have been continuous for many years tend to be more refractory than those of more recent origin.

Evaluation of treatments

Until recently, the treatment literature consisted of small case series or single case reports, but a few larger studies have now been performed. Key findings are:

Pharmacological treatments

Lamotrigine: Despite encouraging results from a pilot study, a placebo-controlled crossover trial did not show evidence of efficacy. However, in a larger, more recent open-label study of lamotrigine-antidepressant combination therapy, significant improvements were seen in a majority of patients.[12]

Opioid antagonists: Transient reductions in symptoms have been reported in response to naloxone infusion, and a recent open-label study of oral naltrexone showed some evidence of efficacy.[13]

Clomipramine: One small open-label study, results inconclusive.[14]

There remains a paucity of data from rigorous controlled trials. To date, the only large double-blind randomised controlled trial is a study of *fluoxetine*, which found no evidence of efficacy.[15]

Psychological treatments

Despite a number of reports of successful treatment with a range of psychotherapeutic techniques, the only treatment trial is an open study of cognitive behavioural therapy (CBT), which showed significant clinical benefits.[16]

Management

It will be appreciated from the above that there is insufficient evidence on which to base definitive treatment guidelines, and as yet

no drugs are licensed for the treatment of DPD in the UK. Current treatment strategies are based on encouraging results from exploratory studies, rather than on any overwhelming weight of evidence, and this limitation should not be concealed from patients.

The combination of lamotrigine (up to 500 mg per day) and an SSRI (usually citalopram or escitalopram) is often used as first-line treatment (see Medford *et al.* in Further Reading below). An alternative is clonazepam (0.5–4 mg per day).There have been no clinical trials of clonazepam in DPD, but many patients find it helpful, particularly when there is co-morbid anxiety. The risk of dependency must be carefully weighed against possible clinical benefits.

Naltrexone (see above) may also have a role, while clomipramine may be helpful when obsessional ruminations are prominent. There is anecdotal evidence to support the use of bupropion in treatment-refractory cases.

CBT may be beneficial, either alone or in combination with pharmacotherapy. Depersonalization experiences do not readily lend themselves to a cognitive behavioural analysis, but CBT techniques may be helpful in addressing associated anxieties, ruminations, and avoidance behaviours, and use of CBT should be considered whenever any of these features are prominent.

The use of standardized rating scales can assist in diagnosis and monitoring response to treatment. The Cambridge Depersonalization Scale[17] is particularly recommended.

In secondary depersonalization, it is usually appropriate to simply pursue conventional treatment of the primary condition, but if depersonalization becomes severe and disabling it may be necessary to treat it more specifically. Treatment of depersonalization in the context of substance misuse is particularly difficult: since most drugs of abuse can cause or exacerbate the symptoms, there is generally little point in attempting specific treatment unless abstinence has been established.

Further information

Mayer-Gross, W. (1935). On depersonalization. *British Journal of Medical Psychology*, **15**, 103–22. Classic early monograph, still of relevance.

Medford, N., Sierra, M., Baker, D., *et al.* (2005). Understanding and treating depersonalisation disorder. *Advances in Psychiatric Treatment*, **11**, 92–100. Conceptual and clinical overview by clinicians from the UK's only specialized DPD clinic.

Schilder, P. (1950). The Image and Appearance of the Human Body. New York: International Universities Press. Contains a fascinating psychodynamic account of DPD.

Simeon, D. andAbugel, J. (2006). Feeling Unreal: Depersonalization Disorder and the Loss of the Self. Oxford University Press, USA. Readable and informative book on the disorder.

References

1. Sierra, M. and Berrios, G.E. (1996) Un cas de depersonnalisation, by L.Dugas. Translation and introduction. *History of Psychiatry*, **7**, 451–61.
2. Sierra, M., Baker, D., Medford, N., *et al.* (2005). Unpacking the depersonalization syndrome: an exploratory factor analysis on the Cambridge Depersonalization Scale. *Psychological Medicine*, **35**, 1523–32.
3. Holmes, E.A., Brown, R.J., Mansell, W., *et al.* (2005). Are there two qualitatively distinct forms of dissociation? A review and some clinical implications. *Clinical Psychology Review*, **25**, 1–23.
4. Lambert, M.V., Sierra, M., Phillips, M.L., *et al.* (2002). The spectrum of organic depersonalisation: a review plus four new cases. *Journal of Neuropsychiatry and Clinical Neurosciences*, **14**, 141–54.
5. Hunter, E.C.M., Sierra, M. and David, A.S. (2004). The epidemiology of depersonalization and derealisation: a systematic review. *Social Psychiatry Psychiatric Epidemiology*, **39**, 9–18.
6. Sierra, M., Senior, C., Dalton, J., *et al.* (2002). Autonomic response in depersonalization disorder. *Archives of General Psychiatry*, **59**, 833–38.
7. Phillips, M., Medford, N., Senior, C., *et al.* (2001). Depersonalization disorder: thinking without feeling. *Psychiatry Res Neuroimaging*, **108**, 145–60.
8. Medford, N., Phillips, M., Brierley, B., *et al.* (2006) Emotional memory in depersonalization disorder. *Psychiatry Res Neuroimaging*, **148**, 93–102.
9. Medford, N., Baker, D., Hunter, E., *et al.* (2003). Depersonalization following illicit drug use: a controlled analysis of 40 cases. *Addiction*, **12**, 1731–6.
10. Baker, D., Hunter, E., Lawrence, E., *et al.* (2003). Depersonalization disorder: clinical features of 204 cases. *British Journal of Psychiatry*, **182**, 428–33.
11. Simeon, D., Knutelska, M., Nelson, D., *et al.* (2003). Feeling unreal: a depersonalization disorder update of 117 cases. *Journal of Clinical Psychology*, **64**, 990–7.
12. Sierra, M., Baker, D., Medford, N., *et al.* (2006). Lamotrigine as an add-on treatment for depersonalization disorder: a retrospective study of 32 cases. *Clinical Neuropharmacology*, **29**, 253–8.
13. Simeon, D. and Knutelska, M. (2005). An open trial of naltrexone in the treatment of depersonalization disorder. *Journal of Clinical Psychopharmacology*, **25**, 267–270.
14. Simeon, D., Stein, D.J., Hollander, E. (1998). Treatment of depersonalization disorder with clomipramine. *Biological Psychiatry*, **44**, 302–03.
15. Simeon, D., Guralnik, O., Schmeidler, J., *et al.* (2004). Fluoxetine therapy in depersonalization disorder: randomized controlled trial. *British Journal of Psychiatry*, **185**, 31–6.
16. Hunter, E., Baker, D., Phillips, M., *et al.* (2005). Cognitive-behaviour therapy for depersonalisation disorder: an open study. *Behaviour Research and Therapy*, **43**, 1121–30.
17. Sierra, M.and Berrios, G.E. (2000). The Cambridge Depersonalization Scale: a new instrument for the measurement of depersonalization. *Psychiatry Research*, **93**, 153–64.

4.10

Disorders of eating

Contents

4.10.1 Anorexia nervosa

Gerald Russell

Introduction: history of ideas

Two different approaches may be discerned in the conceptualization of anorexia nervosa.

1 The medicoclinical approach defines the illness in terms of its clinical manifestations; the main landmarks were the descriptions by William Gull in 1874[1] and Charles Lasègue in 1873.[2]

2 The sociocultural approach is unlike the more empirical clinical approach and takes causation into account by viewing the illness as a response to prevailing social and cultural systems. This was well argued by the social historian Joan Jacobs Brumberg who considers anorexia nervosa simply as a control of appetite in women responding to widely differing forces which may change during historical times.[3]

There is a strong argument for accepting the original descriptions by Gull and Lasègue as containing the essence of anorexia nervosa. They both recognized a disorder associated with severe emaciation and loss of menstrual periods, inexplicable in terms of recognized physical causes of wasting. They were both extremely cautious about the nature of the mental disorder. Gull spoke of a morbid mental state or 'mental perversity', and adopted the more general term anorexia 'nervosa' which has persisted until today. Lasègue also referred to 'mental perversity' but was bold enough to call the condition 'anorexie hystérique, faute de mieux'.

It is probably best to seek a balance between the diagnostic rectitude of the medicoclinical approach and the malleability of anorexia nervosa in different sociocultural settings. Looking back in historical times, it may well be that the self-starvation and asceticism of St Catherine of Siena corresponded to modern anorexia nervosa.[4] In more recent times the preoccupations of the patients have altered so that their disturbed experience with their own body[5] or their 'morbid fear of fatness'[6] has become one of the diagnostic criteria. Yet this concern with body size was not remarked upon by Gull or Lasègue. This is an argument for concluding that at least the psychological content, and perhaps also the form, of anorexia nervosa are changeable in response to historical times and sociocultural influences.

Epidemiology

Screening instruments

The most commonly used screening test in the detection of anorexia nervosa is the Eating Attitudes Test (**EAT**).[7] Doubt has, however, been expressed about the predictive value of the EAT in the very populations where its use was introduced, as only a small percentage of the EAT-screened positive scores will have an actual eating disorder. Thus, the EAT has limited usefulness in surveys for detecting anorexia nervosa unless it is supplemented by detailed clinical assessments. There is also a risk of failing to detect cases of anorexia nervosa as it was found that patients currently receiving active treatment were among the non-respondents, presumably because they wished to conceal their disorder.[8]

Populations surveyed

(a) General population surveys

These are often impracticable when the aim is to detect anorexia nervosa, a relatively uncommon disorder.

(b) Surveys of primary care populations

A useful compromise is that of surveying populations of patients who consult their general practitioners. A Netherlands study was successful because a large population was surveyed (over 150 000) and the general practitioners themselves, after suitable training, were responsible for making the diagnoses.[9,10]

(c) Populations thought to be more at risk

Surveys of ballet and modelling students were conducted because it was thought likely that there would be a high prevalence of anorexia nervosa among them as a result of pressures exerted to sustain a slim figure in keeping with their professional image.[11,12]

Populations of adolescent schoolgirls have also been surveyed as their susceptibility might be raised by virtue of their age, sex, and the frequency of dieting among the school population.[13] The most thorough survey of 15-year-old schoolchildren was that conducted in Göteborg, Sweden.[14]

(d) Surveys based on case registers and hospital records

Data have been obtained on patients referred to inpatient and outpatient psychiatric services, or with the addition of patients who had consulted paediatricians, general medical services, or gynaecologists.[15,16]

Results of epidemiological surveys

(a) Incidence of anorexia nervosa

The studies which counted only hospitalized patients tended to yield low estimates of the annual incidence of anorexia nervosa expressed per 100 000 population (e.g. 0.45 in Sweden[15]). Estimates based on case registers of psychiatric patients similarly yielded fairly low incidence rates (e.g. 0.64 in Monroe County, New York[15]). The incidence found in community-based studies was by far the highest (7.7 in The Netherlands[10] and 8.2 in Rochester, Minnesota[17]), presumably because they included the less severe cases.

(b) Prevalence of anorexia nervosa in vulnerable populations

A high prevalence rate was found among Canadian ballet students (6.5 per cent) and modelling students (7 per cent).[11] A similar survey in an English ballet school also showed a high prevalence of 'possible' cases of anorexia nervosa (7.0 per cent).[12]

Surveys among schoolgirls have shown a fairly wide variation in prevalence rates, ranging from 0 per cent to 1.1 per cent. In the English studies a consistent difference in prevalence rate was found between private schools (1 per cent) and state schools (0–0.2 per cent).[13] This social class distinction was not so definite in the Swedish study where the overall prevalence of 0.84 per cent of schoolgirls, up to and including 15 years of age, represents a high rate for anorexia nervosa.[14]

(c) Age and sex

Epidemiological surveys have confirmed clinical opinion that anorexia nervosa commences most frequently in the young, especially within a few years of puberty. The peak age of onset is 18 years.[13] The illness is less common before puberty, but in a series of patients admitted to a children's hospital the age of onset ranged from 7.75 to 14.33 years.[18] A prevalence of 2.0 per cent in females aged 18 to 25 was found in a community survey in Padua, Italy.[19]

A marked predominance of females over males is usually reported in surveys, for example 92 per cent in North-east Scotland[15] and 90 per cent among the children in Göteborg.[14] On the other hand a community survey in the province of Ontario gave rise to a more balanced female to male ratio (2:1) when cases of partial anorexia nervosa were included.[20]

(d) Social class and socio-economic status

The view has been widely held that anorexia nervosa occurs predominantly in patients with middle-class backgrounds, since Fenwick's classical observation that anorexia nervosa 'is much more common in wealthier classes of society than amongst those who have to procure their bread by daily labour'.[15]

Epidemiological surveys aimed at wider populations leave the question of social class distribution somewhat equivocal. Whereas a high percentage of combined social classes 1 and 2 (Registrar General's categories) were found in clinical studies,[21] a high social class predominance was not found in studies utilizing case registers.[15] On the other hand, the schoolgirl studies mentioned above tended to confirm a high social class predominance.

Has the incidence of anorexia nervosa increased since the 1950s?

From the 1970s experienced clinicians have expressed their view that anorexia nervosa had increased in frequency. This view gained support from surveys repeated on the same population after intervals of 10 years or more in Sweden, Scotland, Switzerland, and Monroe County, New York.[15] Most investigations found a clear trend for an increased incidence over time although it appears that a plateau was reached in the 1980s.[16] In the most thorough study, from Rochester, Minnesota[17] the increase was confined to female patients aged 15 to 24 years. There were similar findings from the Netherlands with a rise from 56.4 to 109.2 per 100 000 among 15- to 19-year-old females from 1985–1989 to 1995–1999.[10] Although there is support for an increased incidence of anorexia nervosa, there remain dissenting voices.

It is better to pose a different question which renders any controversy unnecessary. We should ask instead whether there has been an increase in the incidence of eating disorders including anorexia nervosa. This is especially relevant in view of the fact that bulimia nervosa is a variant of anorexia nervosa.[22] There is strong evidence that bulimia nervosa is a new disorder and has not simply appeared because of improved medical recognition.[23] Moreover, the incidence of bulimia nervosa has risen sharply at least until the mid 1990s, so that it is now about double that of anorexia nervosa[24] (see also Chapter 4.10.2). In conclusion, it is clear that since the 1960s there has been a significant increase in eating disorders, of which the two clearest syndromes are anorexia nervosa and bulimia nervosa.

Aetiology

Aetiological concepts

According to one robust opinion, it is essential to pursue the search for a specific and necessary cause of anorexia nervosa because the currently popular 'multifactorial' approach has little explanatory power. Accordingly the failure to identify a necessary causal element is regrettable. Many of the factors within a wide range of psychological, social, and physical causes so far studied may therefore only be relevant in predisposing to anorexia nervosa, whose causes still elude clarification'.[25]

The multidimensional approach to anorexia nervosa

It is precisely because we do not know the fundamental (necessary) cause of anorexia nervosa that recourse has to be had to a multi-

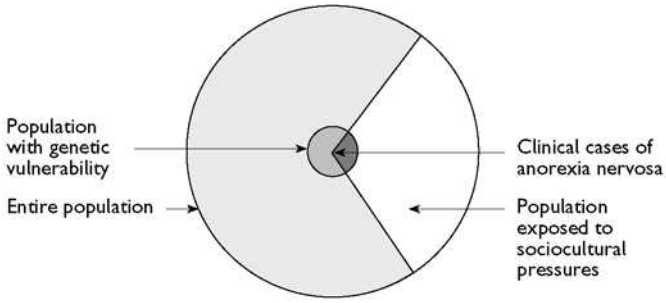


Fig. 4.10.1.1 Diagrammatic illustration of the way that genetic predisposition interacts with sociocultural pressures to cause anorexia nervosa.

dimensional approach, faute de mieux. Although it has its limitations, a multidimensional approach permits one to consider a range of possible causal factors which not only act in an additive manner but may combine in a specific way to bring about the illness: 'It is the interaction and timing of these phenomena in a given individual which are necessary for the person to become ill'.[26]

It is useful to provide a simple model of the way that two broad sets of factors may interact and augment each other (Fig. 4.10.1.1). The outer circle represents the entire population in a developed 'westernized' country. Within the circle there is a large sector representing females within an age range of 10 to 50 years who experience prevailing social pressures to acquire a slender body shape through dieting. Evidently only a small proportion of these women develop the illness. It is likely that for anorexia nervosa to develop it is also necessary to possess a genetic predisposition, represented by the small inner circle. The intersection of the inner circle and the large sector produces a small sector of females who have the genetic predisposition and also experience sociocultural pressures to lose weight, interacting to cause clinical anorexia nervosa.

Causal factors may not only interact as explained above, but they can also influence the content of the illness, its 'colouring', and its form. This modelling function is described by the term 'pathoplastic' which was introduced by Birnbaum.[27] Pathoplastic features are to be distinguished from the more fundamental causes of psychiatric illness, but they do exert a predisposing tendency as well as a modelling role.

Sociocultural causes

(a) The cult of thinness

The pathoplastic sociocultural causes of eating disorders have been subsumed under the term 'the modern cult of thinness' prevalent in westernized societies.[15] Vulnerable patients are likely to respond to this pressure by experimenting with weight-reducing diets which carry a degree of risk, and anorexia nervosa is arguably but an extension of determined dieting.[28]

It is proposed that the cult of thinness is responsible for the increase in the incidence of eating disorders since the 1950s. The social pressures which lead to dietary restraint include the publication of books and magazines advising weight-reducing diets, the fashion industry which caters mainly for the slimmer figure, television attaching sexual allure and professional success to the possession of a svelte figure, and the emphasis on physical fitness and athleticism.[3]

(b) Changes in the psychopathology of anorexia nervosa

At the beginning of this chapter it was pointed out that the psychological content and the form of anorexia nervosa have changed over historical time and in response to sociocultural influences. Whereas Gull and Lasègue spoke only of 'anorexia nervosa' and 'anorexie hystérique' respectively, more recent descriptions of the psychopathology of anorexia nervosa have stressed a disturbed experience of one's own body,[5] a weight 'phobia',[29] or a morbid dread of fatness.[6] It is precisely the modern anorectic's dread of fatness that is most in keeping with today's cult of thinness. It is arguable therefore that modern societal pressures have determined the patients' preoccupations and contributed to their food avoidance. These beliefs are held obstinately and amount to overvalued ideas.

(c) Anorexia nervosa as a culture-bound syndrome

The proposition that anorexia nervosa is a culture-bound syndrome has much support.[30,31] An intense fear of becoming obese is as culture bound as the disorder koro (a fear of shrinkage of the genitals) in Malaysia and South China.

A culture-bound syndrome may be defined as a collection of signs and symptoms which is not to be found universally in human populations, but is restricted to a particular culture or group of cultures. Implicit is the view that culture factors play an important role in the genesis of the symptom cluster . . . [30]

Anorexia nervosa meets these criteria, first because it is limited to westernized or industrialized nations, and secondly because it is clear that the psychosocial pressures on women to become thin constitute a powerful cultural force leading to anorexia nervosa.

In order to allow for exceptions to the rule, when anorexia nervosa occurs in non-westernized countries, the illness may be understood as arising from cultures undergoing rapid cultural change.[31] Anorexia nervosa is thus a 'culture-change syndrome', explaining its increased incidence in Japan and Israel. Anorexia nervosa remains rare in Asia (particularly India), the Middle East (with the exception of Israel), and generally in poorly developed countries.

Young female immigrants who move to a new culture may suffer from an increased prevalence of eating disorders. For example, children with anorexia nervosa were found among Asian families living in Britain. This was linked with an exposure to a conflicting set of sociocultural norms in comparison with their parents and grandparents.[32]

Adverse life events

Anorexia nervosa may be precipitated by adverse life events. Early clinical studies depended solely on the patient's reports of an adverse life event preceding the onset of the illness. These varied widely in severity and included the death of grandparents, a father's remarriage, a severe physical illness, stress or failure of school examinations, or being teased about being overweight.[21]

Recent studies have relied on more objective measurements of adverse life events and comparisons with control groups. In one such study,[33] life events rated included the death of a close relative, a poor relationship with parents, or an unhappy marriage. Fairly severe life events and lasting difficulties were found in the majority of patients with a late onset (after 25 years), whereas they were only found in a minority of patients with an early onset.

Anorexia nervosa can also be precipitated by sexual experiences and conflicts. In a series of 31 adolescent and young women it was

observed that in 14 first sexual intercourse had occurred before the illness.[34] Sexual problems were seen by 13 patients as major precipitants of their illness; 10 of them had experienced intercourse. The authors concluded that a specific aetiological role of sexual factors seemed unlikely, but there might be a direct relationship between the onset of eating difficulties and concurrent sexual problems.

In a series of 15 patients, anorexia nervosa developed during pregnancy or, more often, during the post-natal period when it was accompanied by depression.[35] Risk factors included ambivalence about motherhood, a large weight gain during the pregnancy, physical complications during pregnancy, post-natal depression, and past psychiatric illness.

Childhood sexual abuse

Since it was reported that a high proportion of patients in a treatment programme for anorexia nervosa gave histories of sexual abuse in childhood, it has been supposed that this trauma would be a contributory causal factor.[36] It would be better if this history could be corroborated, but for obvious reasons it is often difficult to do so. Hence this subject raises unusual difficulties in judging the reliability of the data. Child sexual abuse is also discussed in the chapter on bulimia nervosa (Chapter 4.10.2).

In a careful study of childhood sexual experiences reported by women with anorexia nervosa, the authors classified the events according to the seriousness of the sexual act in childhood and concentrated on sexual experiences with someone at least 5 years older.[37] They found surprisingly high rates (about one-third) of adverse sexual experiences in women with eating disorders, and, unlike other investigators, they did not find that the frequency of these reports was less in anorexia nervosa than bulimia nervosa. They concluded that it was plausible that childhood sexual contact with an adult may in some cases contribute to a later eating disorder.

The complexity of this subject has been increased by a study on the relationship between sexual abuse, disordered personality, and eating disorders.[38] The authors found that 30 per cent of patients referred to an eating disorders clinic gave a history of childhood sexual abuse. In addition, they found that 52 per cent of their patients had a personality disorder. A significantly higher proportion of women with a personality disorder had a history of childhood sexual abuse, compared with those without a personality disorder. Surprisingly they still concluded that in patients with eating disorders it was not possible to show a clear causal link between child sexual abuse and personality disorder.

In a review of the subject a number of hypotheses were examined on the relationship between childhood sexual abuse and eating disorders.[39] One hypothesis is that child sexual abuse is more common in bulimia nervosa than in anorexia nervosa. The authors had to concede that the findings remain inconclusive. They also examined the question of whether child sexual abuse is a specific risk factor for eating disorders. They concluded that this was not the case, as the rates of child sexual abuse in eating disorder patients were similar to or less than those in various other psychiatric comparison groups. Finally, they found strong evidence that patients with eating disorders reporting child sexual abuse were more likely to show general psychiatric symptoms including depression, alcohol problems, or suicidal gestures, as well as an association with personality disorders.

Family factors

Two influential groups of family therapists (Minuchin at the Philadelphia Child Guidance Clinic and Selvini Palazzoli in Milan) have devised family models to explain the genesis of anorexia nervosa.

Minuchin *et al.* identified faulty patterns of interaction between members of the anorexic patient's family; they in turn were thought to lead to the child's attempt to solve the family problems by starving herself:[40] 'The sick child plays an important role in the family's pattern of conflict avoidance, and this role is an important source of reinforcement for his symptoms'. Five main characteristics of family interaction were identified as detrimental to the function of the family: enmeshment (a tight web of family relationships with the members appearing to read each other's minds), overprotectiveness, rigidity, involvement of the sick child in parental conflicts, lack of resolution of conflicts.

Selvini Palazzoli[41] also identified abnormal patterns of communication within these families and in addition described abnormal relationships between the family members. She assumed that anorexia nervosa amounted to a logical adjustment to an illogical interpersonal system.

Bruch[42] described girls who developed anorexia nervosa as 'good girls', who previously had a profound desire to please their families to the point of becoming unaware of their own needs. The frequency of broken families in anorexia nervosa is thought to be fairly low. Anorexic families have been found to be closer: the patients more often perceived themselves as having happy relationships within the family.

It remains uncertain whether these abnormal interactions are to blame for the illness or develop as a response by parents faced with a starving child. Careful therapists take pains to reassure parents at the commencement of family therapy: ' . . . we always find it useful to spend some time discussing the nature of the illness, stressing in particular that we do not see the family as the origin of the problem'.[43]

Personality disorders

A sizeable proportion of patients (30 per cent[15] and 32 per cent[21]) were said to have had a 'normal' personality during childhood before their illness. Nevertheless there is general agreement of a close relationship between obsessional personalities and the later development of anorexia nervosa. In fact Janet, who carefully described obsessions and psychasthenia, was dubious about the validity of the diagnostic concept of anorexia nervosa. He thought that the patient's fear of fatness was an elaborate obsessional idea.[44]

In a study of patients admitted for treatment they were classified into anorexia nervosa, bulimia nervosa, or a combination of the two disorders.[45] Personality disorders were identified through the Structured Clinical Interview for DSM-IIIR personality disorders (SCID-II). Seventy-two per cent of the patients met the criteria for at least one personality disorder. Anorectics were found to have a high rate of obsessive–compulsive personality disorder.

There have been attempts to disentangle the features of premorbid personality and illness in order to clarify the personality characteristics predisposing to anorexia nervosa. Women who had recovered from restricting anorexia nervosa were tested at an 8- to 10-year follow-up, using a number of self-report instruments.[46] They were compared with two control groups: normal women and the sisters of the recovered anorexic patients. The women who had

recovered rated higher on risk avoidance and conforming to authority. They also showed a greater degree of self-control and impulse control, and less enterprise and spontaneity.

Biomedical factors and pathogenesis

(a) Historical notes

Since the early part of the twentieth century a recurring theme has been the possibility that anorexia nervosa is primarily caused by an endocrine or cerebral disturbance. From 1916 there was much preoccupation with the concept of Simmonds' cachexia,[47] the assumed result of latent disease of the pituitary gland. There was diagnostic confusion between anorexia nervosa and hypopituitarism, which was only clarified much later when it became known that in true hypopituitarism weight loss and emaciation are uncommon. Hormonal deficits indicative of impaired pituitary function are indeed common in anorexia nervosa, but are merely a secondary manifestation of prolonged malnutrition.

Interest in the neuroendocrinology of anorexia nervosa led to the formulation of the hypothalamic model.[6,48] From the beginning the model was aimed at explaining pathogenesis rather than aetiology; it was not considered an alternative to the psychological origin for anorexia nervosa, but a means of explaining a constellation of disturbed neural mechanisms, as follows:

1 a disordered regulation of food intake;

2 a neuroendocrine disorder affecting mainly the hypothalamic-anterior pituitary-gonadal axis;

3 a disturbance in the regulation of body temperature.

It was known that these functions all reside within the complex of hypothalamic physiology. A 'feeding centre' had been described in the lateral hypothalamus because bilateral lesions there induced self-starvation and death in rats.[49] Over the years it has become clearer that many of the disturbances could be attributed to the patients' malnutrition, as demonstrated by experimental studies in healthy young women who were asked to follow a weight-reducing diet. It was found that ovarian function is extremely sensitive to even small restrictions of caloric intake which often lead to impaired menstrual function.[50]

(b) Cerebral lesions and disturbances

Interest in the hypothalamic model was fuelled early on by clinical reports of patients diagnosed as suffering from anorexia nervosa who were later found to have cerebral lesions, especially tumours of the hypothalamus. More recently, occult intracranial tumours have been detected, masquerading as anorexia nervosa in young children.[51]

Neuroimaging studies in anorexia nervosa have led to findings suggestive of an atrophy of the brain. CT has disclosed a widening of the cerebral sulci and less frequently an enlargement of the ventricles.[52] The outer cerebrospinal fluid spaces were enlarged markedly in 36 per cent of the patients. When the CT examination was repeated after weight gain 3 months later, the widening of the sulci had disappeared in 42 per cent of the patients who had previously shown this finding. In other patients, however, the widening remained unaltered for 1 year after body weight had returned to normal.

Functional neuroimaging techniques have also been applied to research in anorexia nervosa. Regional cerebral blood flow was measured in three series of children.[53] In the majority of the children there was an above-critical difference in the regional cerebral blood flow most often between the temporal lobes but sometimes affecting other brain regions. Hypoperfusion was found on the left side in about two-thirds of the children. Follow-up scans were undertaken in children after they had returned to normal weight; the reduced regional cerebral blood flow in the temporal lobe persisted on the same side as the initial scan. The authors found that there was a significant association between reduced blood flow and impaired visuospatial ability, impaired complex visual memory, and enhanced information processing.

(c) Genetic causes

As with general psychiatric disorders, genetic and environmental factors are no longer viewed as opposing causes of anorexia nervosa, but rather as interacting with one another. Thus there may be groups of genes that determine risk of the illness and specific environmental situations that elicit or prevent the expression of these genes. There is evidence for a family aggregation of anorexia nervosa. This does not necessarily mean that the origin of the disorder is genetic because environmental factors common to the family must also be considered. In a series of 387 first-degree relatives of 97 probands with anorexia nervosa, it was found that the illness occurred in 4.1 per cent of the first-degree relatives of the anorexic probands, whereas no case was found among relatives of the controls who were probands with a primary major affective disorder, or with various non-affective disorders.[54] The authors concluded that anorexia nervosa was familial with intergenerational transmission. It was roughly eight times more common in female first-degree relatives of anorexic probands than in the general population. The absence in the relatives of probands with major affective disorder indicated the specificity in the risk of transmission of anorexia nervosa and the absence of shared familial liability with affective disorders.

The strongest argument favouring a part-genetic causation of anorexia nervosa is derived from the classical method of comparing the concordance rate of the illness in monozygotic and dyzygotic twins. Underlying the method of twin studies is the supposition that the environment for MZ twins and DZ twins is the same, or at least it does not differ in such a way as to cause greater concordance for MZ twins. Although it is known that the environments shared by MZ twins are more often similar than those shared by DZ twins, these similarities are thought not to be of aetiological relevance to anorexia nervosa.[55] The first sizeable series of twins came from the Maudsley Hospital and St George's Hospital in London and showed higher concordance rate for MZ than for DZ twins.[56] Since then the findings have been replicated and the analysis of the twin data has suggested that the liability can be broken down into three sources of variability:

a^2 Additive genetic effects	88 per cent
c^2 Common environmental effects, found to be	0
e^2 Individual-specific environmental effects.[55]	12 per cent

The individual-specific environmental effects are those which contribute to differences between the members of the twin pair. This would happen if only one member of the twin pair had suffered from an adverse effect. It is perhaps surprising that the unique environmental effects contribute to the disorder but not the twins' common environment.

Bulik has found the concept of heritability and its estimation prone to misinterpretation. There is not just one heritability estimate because this statistic varies across the population sampled and the time of the study. One estimate of heritability of anorexia nervosa gave it as 0.74 in 17-year-old female twins.[55]

It remains uncertain how the genetic vulnerability to anorexia nervosa expresses itself in terms of the pathogenesis of this disorder. One view is that this vulnerability confers instability of the homeostatic mechanisms, which normally ensure the restoration of weight after a period of weight loss. This hypothesis would explain why in western societies, where dieting behaviour is common, those who are genetically vulnerable would be more likely to develop anorexia nervosa.[57] It has also been found that MZ twins have a higher correlation for the trait of 'body dissatisfaction' than DZ twins.[55]

(d) Neurotransmitters

Since the early 1970s, the hypothalamic model of anorexia nervosa has been transformed from a consideration of anatomical 'centres' to 'systems' involving neurotransmitters. Much evidence has been presented to show that a wide range of neurotransmitters modulate feeding behaviour, and it was only a small conceptual step to suggest that some were involved in the pathophysiology of eating disorders. At first the neurotransmitters considered were mainly the monoamine systems-noradrenaline (norepinephrine), dopamine, and serotonin. In addition, opioids, the peptide cholecystokinin, and the hormones corticotrophin-releasing factor and vasopressin have also been thought to play a part in the pathogenesis of eating disorders. During recent years the main interest has been focused on the role of serotonin (5-hydroxytryptamine, 5-HT) in the control of natural appetite, especially those aspects concerning the phenomenon of satiety, mediated through a range of processes called the 'satiety cascade'.[58] There is now strong evidence that pharmacological activation of serotonin leads to an inhibition of food consumption. It was also postulated that a defect in serotonin metabolism confers a vulnerability to the development of an eating disorder.[59]

A boost to the concept of altered serotonin activity in anorexia nervosa has come from research showing that these patients while still underweight had significant reductions in cerebrospinal fluid 5-hydroxyindoleacetic acid (**5-HIAA**). The levels became normal when the patients were retested 2 months after they reached their target weight.[60] In order to test whether these findings were secondary to malnutrition, the researchers resorted to the ingenious step of studying patients after 'recovery' when they had reached normal weight. They found elevated levels of cerebrospinal fluid 5-HIAA, possibly indicating increased serotonin activity contributing to the abnormal eating behaviour which often persists in patients who have otherwise recovered.[61] The arguments against this simple model of enhanced serotonin activity as a vulnerability trait in anorexia nervosa should be briefly presented. Serotonin function was again assessed in long-term weight-restored anorectics.[62] The investigators used a dynamic neuroendocrine challenge with d-fenfluramine as a specific probe of serotonin function, which mediates the release of prolactin. If there were any persistent abnormality in serotonin function, the response to this challenge test should differ from that in normal controls. In fact, the rise in prolactin levels was very similar in former patients and normal controls. Accordingly, this study failed to support the notion of increased central serotonin function as a vulnerability trait in anorexia nervosa.[62]

Clinical features: classical anorexia nervosa (postpubertal)

The illness usually occurs in girls within a few years of the menarche so that the most common age of onset is between 14 and 18. Sometimes the onset is later in a woman who has married and had children.

By the time the patient has been referred for psychiatric treatment she is likely to have reduced her food intake and lost weight over the course of several months, and her menstrual periods will have ceased. A regular feature of the illness is its concealment and the avoidance of treatment. Even after having lost 5 to 10 kg in weight and missed several periods, the patient's opening remark is often 'there is nothing wrong with me, my parents are unduly worried'. It is only when the clinician asks direct questions that she will admit to insomnia, irritability, sensitivity to cold, and a withdrawal from contacts with her friends, including her boyfriend if she has one.

Because of this denial, it is important to enquire from a close relative, as well as the patient, about the most relevant behavioural changes.

History

History of food intake. A **food intake history** is obtained by asking the patient to recall what she has eaten on the previous day, commencing with breakfast, which is often missed. An avoidance of carbohydrate and fat-containing foods is the rule. What remains is an often stereotyped selection of vegetables and fruit. 'Diet' drinks and unsweetened fruit juice are preferred, although some patients are partial to black coffee. It is the mother who will indicate that her daughter finds ways of avoiding meals, preferring to prepare her own food and withdrawing into her bedroom to eat it.

Weight history. The patient is usually willing to provide an accurate **weight history**. She may try to conceal her optimum weight before her decision to 'diet', but she is likely to be objective about her current weight, if only to express pride in the degree of 'self-control' she has exerted. The clinician then has an opportunity to enquire into her 'desired' weight by simply asking what weight she would be willing to return to. Her answer will betray a determination to maintain a suboptimal weight.

History of exercising. A **history of exercising** should be taken. Again, the patient is likely to conceal the fact that she walks long distances to school or to work rather than use public transport. She may also cycle vigorously or attend aerobic classes. A parent may report that his or her daughter is running on the spot or performing press-ups in the privacy of her bedroom, judging from the noise that can be heard. The amount of exercising may be grossly excessive, with the patient indulging in brisk walks or jogging even in the presence of painful knees or ankles due to soft tissue injuries.

Additional harmful behaviours, which should be enquired into include self-induced vomiting, purgative abuse, and self-injury. Vomiting and purgative abuse are similar to the behaviours that occur in bulimia nervosa (see Chapter 4.10.2). In anorexia

nervosa they may occur without the prelude of overeating and the patient's motive is simply to accelerate weight loss. The laxative abuse is often at the end of the day. The favourite laxatives in the United Kingdom are Senokot and Dulcolax, and the patient is likely to take them in increasing quantities to achieve the wanted effect as tolerance develops. Self-injury should also be enquired into, and the skin of the wrists and forearms inspected for scratches or cuts with sharp instruments.

Menstrual history. The patient may not volunteer that her periods ceased soon after commencing the weight-reducing diet. On the other hand she may admit that she is relieved that her periods have stopped as she found them to be a nuisance or unpleasant.

(a) The patient's mental state

(i) Specific psychopathology

Several near-synonyms have been used to describe the specific attitude detectable in the patient who systematically avoids fatness: a 'disturbance of body image',[5] a 'weight phobia',[29] or a 'fear of fatness'.[6] Magersucht, or seeking after thinness, was a term applied in the older German literature. The patient will express a sensitivity about certain parts of her body, especially her stomach, thighs, and hips. Not only is she likely to assert that fatness makes her unattractive, but she may add that it is a shameful condition betraying greed and social failure. These distorted attitudes generally amount to overvalued ideas rather than delusions. Occasionally, however, a patient may be frankly deluded, such as one young woman who believed that her low weight was due to thin bones and that fatness was still evident on the surface of her body.

Studies have demonstrated that wasted patients, when asked to estimate their body size, see themselves as wider and fatter than they actually are.[63] Since these early observations, numerous perceptual studies have been undertaken and the conclusion drawn that anorexic patients overestimate their body width more often than normal controls. These distorted attitudes are often associated with a negative affect, so that the disturbance might be viewed as one of 'body disparagement'.[64]

The patient's dread of fatness is so common that it is pathognomonic of anorexia nervosa. There are, however, exceptions. Sometimes a patient may simply deny these faulty attitudes. Another exception is the occurrence of anorexia nervosa in eastern countries where thinness is not generally admired (e.g. Hong Kong and India). The imposition of fear of fatness as a diagnostic criterion on patients from a different culture, where slimness is not valued, amounts to a failure to understand the illness in the context of its culture.[65] An appropriate solution, proposed by Blake Woodside[66] is to accept diagnostic criteria where the psychopathology includes culture-specific symptoms which will differ for western, Chinese, and Indian ethnic groups.

(ii) Denial

Denial in anorexia nervosa is sometimes included within diagnostic criteria of the disorder, for example DSM-IV. Vandereycken[67] has written a scholarly treatize on the subject. He has pointed out that the term 'denial' is used with apparently simplicity in clinical practice, but this stands in contrast to its intriguing complexity. It is a multi-layered concept, which is difficult to measure. Denial is not a static condition but fluctuates over time. A crucial element in its assessment is the inherent conflict of perception between the patient and the clinician. The patient carries out distortions by omissions, concealments, or misrepresentations. She often denies hunger and fatigue and appears not to accept that she is thin. She shows a lack of concern for the physical and psychosocial sequelae of being underweight and may even deny the danger of her condition.

Vandereycken has proposed two categories of denial:

1 Unintentional denial (e.g. the patient's way to improve her self-esteem and preserve her sense of identity).

2 A deliberate denial including the pretence of being healthy and avoiding the treatment others want her to accept. Strong denial is often accompanied by the avoidance of treatment, but the refusal of help is not entirely due to the patient's lack of recognition of her problems.

(iii) Depression

Depression of varying severity, including a major depressive disorder, is common. The patients express guilt after eating, adding that they do not deserve food. A high rate of depression (42 per cent) was found at presentation in one study[21] and the lifetime history of depression in follow-up studies may be as high as 68 per cent.

(iv) Obsessive–compulsive features

The patients frequently eat in a ritualistic way, for example restricting their food intake to a narrow range of foods which experience tells them are 'safe' because they will not lead to weight gain. There is often a compulsive need to count the daily caloric intake. One patient rejected prescribed vitamin tablets in case they contained 'calories'. The frequency of obsessive–compulsive disorder in anorexia nervosa was found to be 22 per cent in a clinical series.[21] In studies using structured clinical interviews the frequency ranged from 25 to 70 per cent.

(v) Neuropsychological deficits

These deficits are seldom detected clinically, and an emaciated patient may pursue studies and obtain surprisingly good examination marks. On the other hand, neuropsychological testing has revealed deficits in attention, and impairment of memory and visuospatial function.[59]

(b) The endocrine disorder

(i) The impairment of hypothalamic-pituitary-gonadal function

Amenorrhoea is an early symptom of anorexia nervosa and in a minority of patients may even precede the onset of weight loss. Amenorrhoea is an almost necessary criterion for the diagnosis of anorexia nervosa. An exception is when a patient takes a contraceptive pill, which replaces the hormonal deficit and may lead her to say she still has her periods.

Generally, when the patient is undernourished, levels of gonadotrophins and oestrogens in the blood are found to be low or undetectable. Not only do malnourished patients show low blood levels of luteinizing hormone and follicle-stimulating hormone, but the secretion patterns of these pituitary hormones regress to a phase of earlier development. For example, severely wasted patients display an infantile luteinizing hormone secretory pattern with a lack of major fluctuations over the course of 24 h. With some degree of weight gain a pubertal secretory pattern appears, consisting first of a sleep-dependent increase of luteinizing hormone at night, and later displaying more frequent fluctuations.[68]

When a patient is still malnourished the ultrasound pelvic examination will reveal that ovarian volume is much smaller than in normal women.[69] Three stages can be discerned in the appearance of the ovaries as the patient gains weight:

1 small amorphous ovaries;

2 multifollicular ovaries (with cysts 3–9 mm in diameter);

3 dominant follicle (10 mm or more in diameter).

At the same time there is a corresponding return of hormonal secretion; follicle-stimulating hormone appears first, followed by luteinizing hormone and finally oestradiol which leads to enlargement of the uterus.[70]

These abnormalities signify that the patient is infertile and remains so until endocrine function recovers. Pregnancy occasionally occurs as the patient is still underweight and improving but before the appearance of the first menstrual period.[71] The pregnancy carries a risk of poor foetal growth during the first trimester, albeit with some 'catch-up' growth during the neonatal period.[72] Occasionally an underweight anorexic patient may seek treatment at an infertility clinic. Treatment with gonadotrophin-releasing hormone may restore fertility, but this practice has been severely criticized.[73]

(ii) Hypothalamic-pituitary-adrenal and hypothalamic-pituitary-thyroid axes

The emaciated anorexic patient shows an increased 24-h plasma cortisol level which returns to normal with a minimal increase in weight, as it is the nutritional intake that is critical.[74]

Reduced tri-iodothyronine (T_3) levels are linked to a reduction in energy expenditure during starvation and are adaptive in nature, so that treatment with L-thyroxin is not appropriate in anorexia nervosa.[68]

(c) Weight loss and malnutrition

(i) Body weight

The severity of weight loss may be recorded as follows:

1 As a percentage of an 'average' body weight to be found in tables for normal populations according to age and height (e.g. the *Metropolitan Life Insurance Tables*).

2 As a percentage drop from the patient's 'healthy' weight before the onset of her illness.

3 Using the Quetelet body mass index

$$\text{BMI} = \frac{\text{weight (kg)}}{(\text{height (m)})^2}$$

A BMI between 20 and 25 is regarded as healthy. A BMI of 14 or less indicates a need for hospitalization. A BMI between 10 and 12 represents the lower limit of human survival.

(ii) Physical examination

Wasting is variable but may be extremely severe, resulting in a skull-like appearance of the head, stick-like limbs, and flat breasts, buttocks, and abdomen. The hands and feet feel cold and readily turn blue in winter. The skin is dry with an excess of downy hair (lanugo) covering the cheeks, the nape of the neck, the forearms, and the legs. Heartbeat is slow (50–60 beats/min) and the blood pressure is low (e.g. 90/60 mmHg) with orthostatic lability. During the routine physical examination muscle power should be tested to detect proximal myopathy, as explained below.

Differential diagnosis of classical anorexia nervosa

The diagnosis of anorexia nervosa is usually straightforward, especially as the modern diagnostic criteria are objective. Wasting diseases such as inflammatory bowel disease (Crohn's disease or ulcerative colitis), thyrotoxicosis, and diabetes mellitus may sometimes be mistaken for anorexia nervosa, but they can be identified through specific investigations. Occasionally there is an interaction between such a medical illness and anorexia nervosa, when a patient wishes to perpetuate the weight loss caused by the former. Rarely, anorexia nervosa may be mimicked by a cerebral tumour altering the function of the hypothalamus.

In older patients there may be diagnostic difficulty with a major depressive or schizophrenic illness. The weight loss in a severe depressive illness results from loss of appetite and the patient's belief that she does not deserve food. A schizophrenic patient may avoid food because of paranoid delusions of being poisoned.

Complications of malnutrition

The complications, which are part of the illness, such as amenorrhoea due to failure of the hypothalamic-pituitary-gonadal axis, have already been discussed. The range of medical complications has been extensively reviewed,[74] with recommendations for the investigations needed on presentation of the patient.

(a) Fluid and electrolyte disturbances and cardiovascular complications

Patients who induce vomiting, abuse laxatives, or take diuretics are likely to experience dehydration and various electrolyte disturbances.

(i) Self-induced vomiting

Loss of gastric acid leads to a metabolic alkalosis and hypokalaemia. A low serum level of potassium may lead to cardiac conduction defects and arrhythmias, skeletal muscle weakness, and renal tubular dysfunction.

(ii) Laxative abuse

The abuse of laxatives is likely to cause dehydration, metabolic acidosis, hyponatraemia, and hypokalaemia.

(iii) The use of diuretics

The use of diuretics such as the thiazide group gives rise to low serum sodium levels which in turn cause fatigue and general weakness. A level below 120 mmol/l may lead to coma. The patient often justifies the use of a diuretic as a treatment for swollen ankles or even 'bloating' of the stomach, which she misinterprets as being due to fluid retention.

(iv) Impaired renal function

This may be caused by different mechanisms: pre-renal failure due to dehydration or hypokalaemic nephropathy.

(b) Peripheral oedema

Fluid retention leading to oedema occurs frequently in patients with anorexia nervosa. It is important to distinguish between 'benign' oedema, which often occurs during the course of effective refeeding, and oedema from other causes, which may have serious consequences.

(i) 'Benign' oedema

'Benign' oedema may occur during inpatient treatment. If the patient accepts a high-caloric intake, fluid retention is the rule as shown by a steep upward rise in the weight curve.

(ii) Famine oedema

If oedema is detected when the patient first presents clinically, or if it develops without a preceding improvement in food intake, the underlying mechanisms should be carefully investigated in order to avoid the risks of congestive cardiac failure and pulmonary oedema. By the time peripheral oedema is detectable, the amount of fluid retained in the body contributes several kilograms to the patient's weight, thereby concealing the true loss of weight. It is a misconception that peripheral oedema is usually due to a lowering of plasma albumen, as this is not necessarily so in anorexia nervosa. Therefore the clinician must look for other reasons.

1 A fall in interstitial fluid pressure has been proposed whereby water seeps from the blood into the interstitial spaces[75] whereas the total exchangeable sodium is likely to remain within the normal range.

2 'Rebound' oedema following hyponatraemia due to abuse of laxatives or diuretics.

3 Wet beri–beri: vitamin deficiency in anorexia nervosa is uncommon because many patients continue to eat vegetables and fruit. Nevertheless a lack of vitamin B_1 (thiamine) may occur in patients who eat a stereotyped diet, and Wernicke's encephalopathy may follow.[76] It may also lead to congestive cardiac failure with severe oedema from a nutritional cardiomyopathy, precipitated by refeeding without thiamine.

4 Congestive heart failure and pulmonary oedema may occur as a result of general undernutrition leading to a decreased cardiac mass, cardiac output, and volume. **Death may be caused by an injudicious intravenous infusion of fluids**.

(c) Metabolic disturbances

(i) Hypoglycaemia

Severe hypoglycaemia with plasma glucose levels as low as 1.0 mmol/l is recognized as a cause of death. Hypoglycaemia may go unrecognized, as a lack of sympathetic nervous response may mask the classical symptoms and signs.

(ii) Hypercarotenaemia

This is a benign condition causing an orange–yellow colouration of the skin of the palms, soles, forearms, and the region of the nose. It is partly due to the consumption of large amounts of foods rich in carotenoids, especially carrots, tomatoes, and the green outer leaves of vegetables.

(d) Myopathy

Weakness of specific muscle groups is common and is due to severe protein-energy malnutrition. There is a 'proximal' myopathy, affecting the musculature of the pelvic girdle and the shoulder girdle. The patient first notices an increasing difficulty in climbing stairs. Weakness may also affect the muscles of the head and neck and the face. When the patient is asked to rise from a sitting position, she will tend to push herself up using her hands and arms. She also has difficulty in rising unaided from a squatting position.[78]

There are no characteristic abnormalities in blood chemistry, although creatine kinase and liver enzymes may be elevated and the activity of the enzyme carnosinase may be reduced. Myopathic changes are consistently present in muscle biopsy specimens. Histology reveals the 'chequerboard' distribution of muscle fibres with a selective type 2 fibre atrophy. Electron microscopy reveals the presence of strikingly abundant glycogen granules between the myofibrils and under the sarcolemma.[78]

The detection of myopathy is a clear indication for the patient's admission to hospital and a refeeding programme. After a weight gain of a few kilograms her muscle strength will begin to return and a complete recovery is the rule.

(e) Osteoporosis

A high proportion of patients with anorexia nervosa risk developing osteoporosis and consequent pathological fractures. Significant reduction in bone mineral density of the femoral neck was found in all 20 patients with anorexia nervosa of 6 years or more duration.[79] The favourite method of measuring bone density in the lumbar spine and hip is by dual-energy X-ray absorptiometry. A measurement for all patients with anorexia nervosa of 2 years duration or more is recommended.[74, 79] It is difficult to disentangle the harmful effect of the nutritional deficiency itself from the associated oestrogen deficiency. The pathogenesis of osteoporosis in anorexia nervosa differs from that in postmenopausal women. In anorexia nervosa the nutritional deficiency (including a lack of calcium and vitamin D) leads to a low rate of the recycling of bone through bone formation and resorption, but the balance is disturbed with a relative increase in bone resorption.

The evidence favours the likelihood of improvement of bone density with nutritional recovery and resumption of menstruation.[79] There is much uncertainty about the best treatment. Although patients are often automatically prescribed oestrogen replacement, the only controlled trial undertaken so far indicated that oestrogen was only effective in severely underweight anorexic patients.[81] Indeed, it has been argued that oestrogen replacement could be harmful in some patients (children with premenarchal onset) and unnecessary in others with an illness of less than 3 years duration.[82] Instead the emphasis should be on encouraging weight gain, with the possible addition of calcium and vitamin D. Older patients with a poor prognosis (e.g. who have been ill over 10 years) might benefit from oestrogen replacement. There is little evidence for the use of other medication such as biphosphonates.

(f) Disturbed temperature regulation

Disturbances in the control of temperature are evident from clinical observations; the patients frequently complain of feeling cold, and in the winter they have cold and blue extremities and suffer from chilblains. In the severely malnourished patient hypothermia may be a cause of death. Severe malnutrition is accompanied by a low central body temperature, presumably because of a low metabolic rate. Ingestion of a high-calorie meal can cause a significant increase in the central body temperature[83] with complaints of heat in the periphery and sweating after food.

(g) Haematological changes

Anorexic patients may develop a significantly lowered haemoglobin level and a reduced haematocrit and white cell count, with a relative decrease in neutrophil leucocytes and an increase in lymphocytes.[84] The anaemia is usually moderate and is normocytic normochromic in type. The mechanism is that of starvation-induced bone marrow hypoplasia. Only occasionally is there an iron deficiency. The anaemia gradually becomes corrected with weight gain. A low platelet count in the blood has also been observed and there may be an associated thrombocytopenic purpura.

(h) Complications arising during rapid refeeding

(i) Acute dilatation of the stomach

This complication has been described in anorexia nervosa during the course of refeeding.[85] The patient develops copious vomiting, upper abdominal pain, distension of the upper abdomen, and rapid dehydration. Treatment is by continuous gastric aspiration, and this complication is one of the rare indications for intravenous infusions of glucose and saline. Gastric dilation is best prevented by avoiding a food intake above 1500 cal daily during the first week of refeeding.

(ii) Hypophosphataemia

When the illness has been protracted and has led to severe emaciation, the body stores of phosphate become depleted. The fall in serum phosphate is aggravated during refeeding, especially parenteral feeding.[86] Levels of phosphate may fall as low as 0.2 mmol/l. Clinical features are cardiac irregularities with a prolonged QT interval, impairment of consciousness, and delirium. Treatment is with oral phosphates rather than by intravenous administration.

Anorexic mothers as parents

A patient who is improving may conceive despite having a suboptimal weight and still not menstruating.[71] A mother may also develop the illness after having borne children. In a series of eight mothers, 9 out of 13 of their children suffered from food deprivation, identified by reductions in weight for age and in height for age as shown on Tanner–Whitehouse charts.[87] The anorexic mothers had no intention of abusing the children and indeed were affectionate towards them. **They simply extended their abnormal concern with body size to their own children**. They adopted different ways to ration their children's food intake according to their age. They might prolong breast feeding, dilute the bottle feeds, reduce the amount of food available in the home, confine eating to meal times, forbid the consumption of sweets, and prevent others giving them food. An important part of the management is to anticipate the risk to any children and conduct tactful enquiries. A whole-family approach should be adopted, focusing on the children's needs for food. The children should be followed up to ensure that they gain weight and catch-up growth is established.[87]

Clinical features: anorexia nervosa of early onset (premenarchal)

It would be too arbitrary to define an early onset by age limits such as an onset from 8 to 16 years. A more meaningful frame of reference is the onset in relation to the stage of puberty reached by the child.[88] Because puberty is a complex developmental process spanning 2 to 3 years, it is best to name as 'premenarchal' the illness which commences some time after the first signs of puberty and before its completion as shown by the first menstrual period. In true prepubertal anorexia nervosa the illness begins even earlier, before the very first signs of puberty. Postpubertal anorexia nervosa is when the illness commences after menstruation has been established.

There is much similarity between the clinical features of an illness of early onset and one which is postpubertal. However, there are two important differences. The first is the potential for the illness to interfere with the child's pubertal development. The second is the heart-rendering impact on the child's family. It follows that the management of the family is of supreme importance, and the clinician should be prepared for parental reactions, which may detract from a rational plan of treatment.

The clinical presentation is variable. Often there are precipitating events such as a family bereavement or a physical illness leading to weight loss. The onset is likely to be insidious,[18] with the parents noticing nothing amiss except for non-specific features. Symptoms of depression and irritability are common.[14,77] Some children cannot describe feelings of depression, but tearfulness may be obvious. They withdraw from friends and may refuse to go to school. Others express ideas of being underserving of love or food. Increasingly these youngsters have been found to injure themselves, especially by scratching with their nails or cutting the skin of the wrists and forearms with sharp objects, and occasionally by knocking and bruising the head. In a severe depression the child may say she hears voices calling her 'bad', but further questioning indicates that these are not true hallucinations but vivid expressions of her own thoughts. Another common presentation is with bodily symptoms, especially headaches, abdominal discomfort, and a wide range of gastrointestinal symptoms, which inevitably elicit physical investigations.

At some stage, however, the parents observe that their child is avoiding food and is reluctant to eat at normal meal times. She resorts to deviousness and secrecy. The omission of school meals often goes undetected. Eventually it is noticed that she has become thinner and may have lost a great deal of weight. Resistance is met when attempts are made to reverse the loss of weight. Even a young child may become preoccupied with the caloric values of foods and adopt additional methods of inducing weight loss. The child is likely to exercise excessively—jogging, walking, or cycling long distances. An attempt to reduce the excess activity may lead to solitary exercising in the bedroom, including press-ups or running on the spot. Other patients may induce vomiting or take laxatives even after small meals, but overeating typical of bulimia nervosa is rare in young children.[77]

(a) Weight loss

Because of the early onset while the child should still be growing in stature, there is a failure to gain the weight, which normally accompanies the growth spurt. Later there occurs an actual loss of weight and, because the child has not yet reached her optimum weight, a very low weight indeed may result—25 kg or even less. Symptoms of malnutrition ensue including tiredness, constipation, and sensitivity to cold with cold extremities.

(b) The psychological disorder

Even younger children are likely to disclose that they are fearful of becoming fat, a disturbance similar to the overvalued idea held by older patients. A minority of patients will disclose their reluctance to develop personal, sexual, or social maturity, in a manner, which fits into Crisp's model (see p. 796 of this chapter). A few may express reluctance to have menstrual periods. A girl may say she is indifferent whether she menstruates or not, but would like to develop breasts like other girls in her class. The reluctance to 'grow up' may be expressed in social terms, with the patient saying that she could not imagine herself ever leaving home or her mother. On the other hand most girls are keen to reach a normal stature.

Severe depression was found in 69 per cent of youngsters in the Göteborg study.[14] In the same series one-third of the patients had an obsessive–compulsive personality disorder, and 8 per cent developed hand-washing and other compulsions.

(c) Physical examination

Physical examination will reveal a varying degree of wasting, affecting the limbs, the abdomen, the buttocks, and the facial appearance. The extremities are blue and cold, and ischaemic changes may lead to gangrene affecting the toes. Other physical changes are similar to those in the adult, with the addition of a delay in puberty.

(d) Delay or arrest of puberty

The illness may adversely affect pubertal development depending on its time of onset. If the onset is truly prepubertal the child will not yet have shown the first signs of puberty, such as the appearance of pubic hair and breast buds. When the illness begins during the course of puberty these early signs may have appeared, but the breasts will show early growth only (Tanner stages 1 or 2), and the child will not yet have menstruated. The effects on pubertal development can be divided into those affecting growth in stature, breast development, and menstrual function.[88]

(i) Growth in stature

Growth in stature may have become arrested. In a series of 20 patients with a premenarchal onset, only two of them had reached the 50th centile in height. With successful treatment and weight gain, catch-up growth of 2 to 5 cm may be achieved, but only in patients aged 17 or less.[88]

(ii) Breast development

Breast development: in the same series only six patients had normal breasts and as many as 10 had infantile breasts. After prolonged weight gain, eight of the 14 patients showed a considerable response in breast growth.[88]

(iii) Menstrual function

Menstrual function: Primary amenorrhoea is the rule. In the series already referred to only four of the 20 patients had menstruated by the age of 16 years. A further three began their periods between 16 and 18 years of age, and four at ages ranging from 18 to 25. The remaining patients had prolonged amenorrhoea.

The above series consisted of selected patients in whom pubertal delay was severe, whereas marked pubertal delay has seldom been reported in other series in which only delayed menstruation has been remarked upon.

A young boy who develops anorexia nervosa is also likely to become preoccupied with fatness and accordingly avoids food in order to lose weight. The endocrine disorder in the male similarly consists of a disturbance of the hypothalamic-pituitary-gonadal axis. With a prepubertal onset, the penis and scrotum remain infantile, there is only a scanty growth of pubic and facial hair, and the voice may not break.

Special investigations

Pelvic ultrasound monitoring of the ovaries and uterus is a useful method of ascertaining regression and recovery in children with anorexia nervosa.[89] On first testing, and in the presence of low weight and amenorrhoea, ovarian volume and uterine volume were found to be reduced in comparison with normal pubertal girls. In the latter the normal range of ovarian volume is 3.95 ± 1.7 cm^3 and uterine volume is 14.8 ± 7.6 cm^3. On retesting the patients 18 months after the first scan those who were menstruating showed significantly larger ovarian and uterine volumes than those with amenorrhoea. The authors concluded that for ovarian and uterine maturity to occur it is necessary to achieve a mean weight of 48 kg and a mean weight-to-height ratio of 96.5 per cent. They found that pelvic ultrasound scanning helped to motivate the children to accept a higher body weight.

Differential diagnosis of early-onset anorexia nervosa

Frequent bodily complaints, loss of weight, and abnormalities of growth lead these children to be referred to paediatricians for special investigations. It has been proposed that young anorexic boys should be investigated by means of neuroimaging, so as not to miss occult intracranial tumours.[32] The diagnosis of anorexia nervosa should be distinguished from atypical eating disorders in childhood.[77]

Pervasive refusal syndrome is characterized by a child refusing to eat, drink, walk, talk, or take care of herself. Anxiety, phobic responses, and depression are also present.

Selective eating is the term applied to a child who restricts food intake to two or three different foods, such as biscuits, crisps, or potatoes, but usually remains in good health.

Food avoidance emotional disorder. this condition is one in which food avoidance is attributable to an emotional disturbance in the absence of a dread of fatness, a necessary criterion for the diagnosis of anorexia nervosa.

Food fads and food refusal. they are usually intermittent and physical health is not compromised.

Classification

ICD-10[80] (including patients with premenarchal onset[88])

1 Body weight is maintained at least 15 per cent below that expected for health. Prepubertal children may fail to gain the weight expected during the growth spurt. Weight loss is caused by avoidance of 'fattening' foods, possibly with the addition of self-induced vomiting, purgative abuse, excessive exercise, or the use of appetite suppressants.

2 The patient holds the overvalued idea of a dread of fatness and keeps her weight below a self-imposed threshold.

3 There is an endocrine failure manifest in women as amenorrhoea and in men as a loss of sexual interest and potency.

4 If the onset is premenarchal, the sequence of pubertal events is delayed: growth is arrested; in girls the breasts do not develop and there is a primary amenorrhoea; in boys the genitals remain juvenile.

DSM-IV[90]

1 Refusal to maintain body weight above the minimally normal weight for age and height (e.g. weight less than 85 per cent of that expected).

2 Intense fear of gaining weight or becoming fat, even though underweight.

3 Disturbance in the way in which one's body weight or shape is experienced.

4 In postmenarchal females, there is amenorrhoea of at least three menstrual cycles.

DSM-IV subdivides anorexia nervosa as follows.

1 Restricting type, without regular binge-eating or 'purging' behaviour (in the United States this term includes self-induced vomiting and the misuse of diuretics as well as laxatives).

2 Binge-eating/purging type.

Anorexia nervosa in males

The reduced frequency of anorexia nervosa in the male might lead one to surmise that the disorder is likely to differ between the sexes in its aetiology, clinical manifestations, and prognosis. Yet, there are remarkable similarities between the sexes as regards the age of onset and the specific features of the psychopathology.[91–93] For example, the male patients tended to select a diet which was low in fattening foods and resorted to subterfuges to dispose of food, such as self-induced vomiting and purging, and strenuous exercising. They expressed a fear of fatness and considered themselves overweight, even when they were thin.

Of course there are fundamental biological differences which inevitably alter the manifestations of the endocrine disorder in the male and, to a lesser extent, the nutritional disorder. Testicular function, as gauged by the urinary output of testosterone, is disturbed in male patients when they are emaciated. Refeeding leads to a partial correction of this abnormality.[91] The body composition of the mature male differs from that in the female; he has a lower reserve of adipose tissue so that protein depletion occurs more rapidly when he loses weight.

The relative resistance of the male against developing anorexia nervosa remains a mystery. It is even unclear whether the sex difference is likely to be due to biomedical factors or psychosocial differences. It has been suggested that young females often become preoccupied with 'fatness' because of its reproductive, biological, and social significance, whereas young males are more concerned with their musculature and its significance for strength, dominance, and masculinity.[92] These differences are linked with the frequency of dieting among adolescent girls and its rarity in boys.[94]

In a series of male patients with 'primary' anorexia nervosa most of the patients reported problems with sexuality.[95] Sexual anxieties had been present with respect to heterosexual as well as homosexual behaviour. One quarter admitted homosexual tendencies. Almost all were relieved by the loss of libido following weight loss. The authors concluded that males with atypical gender role behaviour had an increased risk of developing anorexia nervosa in adolescence.

There are only a few follow-up studies of anorexia nervosa in males. In one impressive study, 27 patients were followed up for a minimum of 2 years and a mean of 8 years.[96] Expressed in terms of the Morgan–Russell categories of general outcome, a good outcome was found in 44 per cent, an intermediate outcome in 26 per cent, a poor outcome in 30 per cent, and no deaths. Only a few predictors of outcome were identified: disturbed relationships with a parent in childhood led to a poor outcome, and the occurrence of previous sexual activity was associated with a good outcome. The outcome in males was remarkably similar to that in females.[94]

Course and prognosis

The natural outcome is defined as the long-term results of the illness or disease process. The clinical prognosis is the difficult task of forecasting the future course and final outcome of the illness in an individual patient.

Outcomes from follow-up studies in anorexia nervosa

There have been comprehensive appraisals of follow-up studies in anorexia nervosa[97] which have put forward criteria for the near-perfect follow-up study, which in practice are seldom fully met. Among these criteria are precision in the diagnostic features, the use of standardized interviews, 100 per cent success in tracing the patients, and a sufficiently long follow-up to determine eventual outcome. An arbitrary interval of at least 4 years was previously set[21] and most recent studies have adhered to this recommendation. Several groups of investigators have adopted measures of outcome based on the Morgan–Russell scales.[98] Their use gives rise to three possible categories of general outcome based on body weight and menstrual function: 'good', 'medium', and 'poor'.

In a review comparing three British studies and one Swedish study,[99] each with a mean follow-up of 5 to 6 years, it was found that the patients treated in Bristol had a better outcome than those treated at the Maudsley Hospital in London. The explanation is one of selection bias already mentioned: the Maudsley patients had all required inpatient treatment, whereas in the Bristol series the majority were outpatients. The third British study, from St George's Hospital, London, showed a quality of outcome intermediate between the other two.

The Swedish study was extended by two later follow-ups at 15 and 33 years. On the one hand, the percentage of good outcomes gradually increased and the percentage of poor outcomes diminished. On the other hand, the death rates increased with time; after 33 years the total mortality from anorexia nervosa or suicide was 18 per cent. Slow recovery was the rule: only 29 per cent of patients recovered in less than 3 years of illness, another 35 per cent within 3 to 6 years, and the remainder took much longer with recovery after 12 years being rare.[99] The Maudsley series of patients was also extended to a mean follow-up of 20 years. There was a reduction in the percentage of patients with a poor outcome, but an increase in the mortality rate close to that of the Swedish study.[100]

Prognostic predictors of outcome

In the long-term Maudsley study a poorer outcome was predicted by an older age of onset, a longer duration of illness, the presence of neurotic traits in childhood, personality problems, and the occurrence of relationship difficulties within the family.[100]

Comparison of mortality rates

In a review of 42 studies the aggregate annual mortality rate from anorexia nervosa was found to reach 0.56 per cent on average.[101] Complications of anorexia nervosa accounted for 54 per cent of deaths, suicide for 27 per cent, and other causes for 19 per cent.

A fair measure of consistency has been found in different parts of the world, especially when allowance is made for selection biases: in Denmark 0.5 per cent per annum (younger patients), in Sweden 0.75 per cent per annum, in the United States 0.66 per cent per annum, and in the United Kingdom 0.75 per cent per annum.[100]

Follow-up studies in early-onset anorexia nervosa

In a series of 60 patients in Berlin with a mean age of onset of 14.7, followed up for at least 4 years, recovery occurred in 68 per cent of patients but there was a 6.6 per cent mortality over a mean of 4.8 years,[102] a higher rate than was found in the Danish study already mentioned. A study of teenage-onset anorexia nervosa had the advantage of community screening and the use of a comparison group. At a 10-year follow-up the outcome was good in 43 per cent, intermediate in 29 per cent and poor in 20 per cent. There were no deaths. Strikingly there was a poor psychosocial outcome in half the anorexic patients with high dependence on state benefits.[103]

Conclusions

Theander[104] has provided good advice and useful definitions:

> Clinical experience has repeatedly shown that patients with anorexia nervosa can recover after a period of illness amounting to more than 10 years.

A distinction should be made between a 'chronic' illness and a long-standing illness. The word 'chronic' should be restricted to a continuous illness of more than 15 years' duration because up to then the patient is still capable of a full recovery. The term 'protracted' is preferable for an illness of prolonged duration but not as long as 15 years.

Treatment: review of evidence

In this review the evidence underpinning treatments for anorexia nervosa will be presented. Whenever possible the clinician should use a treatment whose efficacy has been established, but he will seldom be able to confine himself to such treatments. For example, in the case of the severely malnourished patient who presents as an emergency an admission to the specialized eating disorder unit may be the first necessary step, even though it has sometimes been argued that the evidence underpinning inpatient treatment is limited. In the majority of patients presenting with only moderate weight loss it is possible to select a treatment with a higher level of evidence for its efficacy. Such a treatment is family therapy, the treatment of choice in an adolescent patient.

Some cautionary words are called for on the subject of evidence-based psychiatry. For anorexia nervosa, there is no truly specific treatment in the sense of effecting a long-term 'cure'. The choice of treatment should be determined by the needs of the patient at the time when she presents for treatment and will vary according to the severity of the illness and the stage it has reached. It is wise to take into account the patient's age, her maturity, and her personality. A given treatment may be indicated for a short period only at the end of which the therapeutic aims need to be reviewed. Such a selective approach has sometimes been called the 'stepped care' approach but this concept can be misleading. The steps may not be in the same direction. They may go down as well as up. For example, a patient may be in urgent need of refeeding and if she makes good progress the next stage would consist of consolidating her improvement. If she is unfortunate, however, earlier priorities will reassert themselves.

A tendency to decry the benefits of treatment in anorexia nervosa first appeared in the early 1990s, and in more recent years cries of despair have become ever shriller. To some extent this is a side effect of the treatment evaluation industry which has unintentionally caused confusion between the usefulness of treatments on the one hand and the solid scientific evidence for their efficacy on the other. In an ideal world the two would be the same. In an illness such as anorexia nervosa they often are not, as will be illustrated further on.

A key member of the treatment evaluation industry in the United Kingdom is **the National Institute for Clinical Excellence** (NICE).[105] NICE has developed a hierarchy for the strength of the evidence underpinning various treatments, as described in Chapter 1.10. I is the strongest level of evidence, followed by Ia, IIa, IIb, III, and IV, the last being the lowest level. Anorexia nervosa is an area of care, which particularly requires Clinical Practice guidelines because of substantial dangers to life and physical health and some clinical uncertainties. So far NICE has failed to provide the required 'statements to assist practitioner decisions about appropriate health care for specific clinical circumstances'. NICE has not recognized the solid evidence favouring family therapy in adolescents (according it only Grade B in an ill-defined set of Grades A, B, and C for clinical guidelines). Family therapy should be regraded as A. NICE has also failed to identify the specific value of a specialized eating disorder service in patients with a risk to physical health and life. This should have been graded at least as B, or as (I) using the system of clinical recommendations described in Chapter 1.10. The correction of severe malnutrition requires a specialized eating disorder service rather than the involvement of a physician and medical unit (proposed by NICE) whose treatment is usually restricted to tube-feeding. NICE has also failed to recognize the value of other treatments, according them only a Grade C, implying there have been no good clinical studies.

The randomized controlled trial (RCT) has been idealized as the 'gold standard' for the level of evidence. The absence of such evidence for a given treatment does *not* signify that this treatment lacks efficacy. Some colleagues write and speak as if treatments other than those confirmed by RCT are not worthy of consideration. In truth they may well be effective, but the evidence for this is hard to come by. An effect of concentrating on RCTs and ignoring sound clinical evidence, is the conclusion reached by one authority that evidence-based treatment of anorexia nervosa is 'barely' possible. This is to ignore the fact that there are essential treatments for which evaluation by RCT has not yet been possible, because of insuperable ethical obstacles or limited patient compliance.

RCTs are particularly problematic in the assessment of treatments for patients with anorexia nervosa. For example, the NICE guidelines do not mention the role of a specialized eating disorder service, although there are oblique references to 'professionals with specialist experience of eating disorders' and 'a place for inpatient management'. Specialized eating disorder units (EDUs) long ago graduated to providing a broad service not confined to inpatient treatment but including daycare and outpatient treatment. It is important to evaluate the benefits of a specialist eating disorder service, and this is not mentioned in the NICE Guidelines. There has been only one study randomizing low weight patients to inpatient or outpatient treatments,[96] as the difficulties are obvious.

A balanced view needs to be taken towards the role of RCTs in the evaluation of treatments in anorexia nervosa. There has recently been a swing of the pendulum in the reverse direction. Some researchers have reached the surprising conclusion that RCTs in anorexia nervosa are so difficult that they should be considered as 'premature' in view of their cost and the failure of some of hem.[106]

An important reason why researchers may fail to complete a satisfactory RCT is the patient's behaviour in the course of the trial. The anorexic patient's resistance to treatment may extend to sabotaging the basic requirements of the treatment trial itself, e.g. compliance with the treatment and the follow-up assessments. These problems are difficult to eliminate entirely. Nevertheless RCTs are still possible if the researchers are assiduous in following-up their patients. Evaluators of research trials should see that the perfect trial in anorexia nervosa is a contradiction in terms. The pessimists favouring the abandonment of RCTs should remember that new ideas can emerge unexpectedly during the course of a trial, thus pointing the researcher towards fresh treatment approaches capable of being retested at a later date.

In the remainder of this section levels of evidence for the efficacy of treatments will be discussed, even when they are not considered by NICE.

Family therapy

Family therapy is the main treatment to have been carefully evaluated in anorexia nervosa, having now been tested in a number of randomized controlled trials. Any one of these trials has its imperfections, but taken as a body of evolving evidence, family therapy should be seen as the method of choice for treating anorexic patients with an early age of onset. Some of these trials will be summarized in order to illustrate ethical and technical difficulties, as well as the strength of the evidence.

(a) The first controlled trial of family therapy

The ***first controlled trial of family therapy*** was begun at the Maudsley Hospital in the early 1980s on a series of 57 patients.[107] The principles of the family therapy have been incorporated in the Maudsley model, which has been adopted by researchers in the United States who have reached broadly similar conclusions.[108] One assumption of the Maudsley model of family therapy is that the family is ineffective in helping the patient eliminate her symptoms and might indeed contribute to their maintenance, because family life has become so organized around a potentially life-threatening problem. It is considered essential to correct the patient's starvation by assisting the parents to take control of their child's eating until her weight has returned to normal.

The first Maudsley trial had to overcome ethical objections to a randomized allocation bearing in mind the severity of the illness. Risks were minimized by first ensuring that the patients' weight loss had been corrected during an admission to an eating disorders unit. Family therapy itself commenced on discharge and aimed at the prevention of relapse. The outpatient family therapy was administered for 1 year and the patients were assessed on the completion of 1 year's treatment and 5 years later.

The 57 anorexic patients were subdivided into three groups according to recognized prognostic criteria. This showed that the benefits of family therapy were confined to one of the three subgroups, namely patients with the age of onset less than 19 years and a duration of illness less than 3 years.

Patients in each subgroup were randomly allocated to family therapy or the control treatment which consisted of a supportive and problem-centred individual therapy. Most important in assessing progress was body weight. The Morgan-Russell scales were utilized to obtain categories of general outcome and measures of adjustment along five dimensions: nutritional status, menstrual function, mental state, psychosexual adjustment and socio-economic status, as well as their mean, the 'average outcome score'.

(i) Group 1: early age of onset (short history)

A tendency for the patients to lose weight on discharge from hospital was reversed more readily in patients in receipt of family therapy. The superiority of family therapy was demonstrated by a higher weight at the end of 1 year's treatment (Fig. 4.10.1.2). Family therapy was also demonstrated to be more beneficial on the Morgan-Russell scales in terms of more good outcome categories and a higher average outcome score.

(ii) Group 2: patients with late age of onset

In this group the effect of the two therapies was reversed (Fig. 4.10.1.3). Individual therapy appeared to result in a greater weight gain than family therapy but significant only at a 6-months follow-up.

It was concluded that family therapy was more effective than individual therapy in patients whose illness began before the age of 19 years and had lasted less than 3 years.

(b) Beneficial components of family therapy in adolescents: the second Maudsley RCT[109]

The finding that family therapy is effective in younger patients with anorexia nervosa has led to a search for the effective components of this therapy. Many family therapists consider that it is important to understand the way the family functions. Others, however, prefer a symptom oriented approach with an emphasis on helping the parents to manage their child's problem.[43] The relative importance of these two components of family therapy has been investigated by comparing conjoint family therapy with separated family therapy.[109] In conjoint family therapy (CFT) the whole family is seen together in the treatment sessions. In separated family therapy (SFT) the parents are seen together and the same therapist also sees

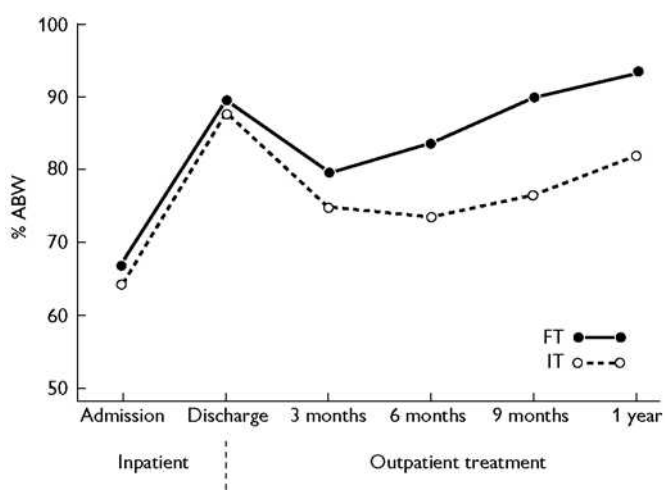


Fig. 4.10.1.2 Group of patients whose illness had an early onset and was of short duration. FT, family therapy; IT, individual therapy; ABW, average body weight.

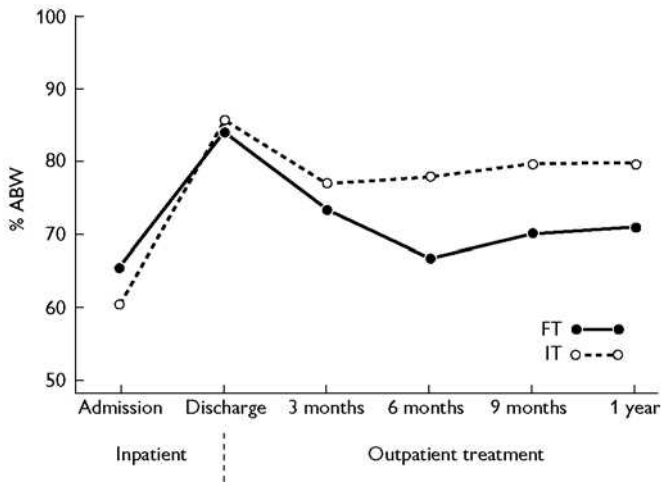


Fig. 4.10.1.3 Group of patients with relatively late onset of anorexia nervosa. FT, family therapy; IT, individual therapy; ABW, average body weight.

the patient separately for therapy and support. In CFT it is possible to observe and interpret family transactions, whereas this is not possible in SFT. The two treatments share the therapeutic advice for the parents to sustain a united approach aimed at improving their child's eating pattern.

The RCT was between CFT and SFT, given for 1 year on an outpatient basis. Forty adolescent patients were enrolled, most of them severely underweight (mean BMI 15.4).

In the course of the year the benefits of CFT and SFT were similar with considerable gains in weight (mean 8.2 kg) and marked improvements in the Morgan/Russell average outcome scores. Significant differences in the patients' responses to CFT and SFT were only obtained when treatment outcome was analysed separately in patients of high and low expressed emotion (EE) families. The patients within the high EE group allocated to SFT reached a significantly higher weight at follow-up (99.9 per cent ABW) than those in the CFT group (85.8 per cent ABW).

(c) RCT undertaken in Stanford, USA[108]

This trial was also on adolescent patients, this time comparing two courses of family therapy: one lasting for 6 months (10 sessions) and the other for 1 year (20 sessions). The aim of the study was to ascertain whether the duration of the family therapy exerted a critical effect on outcome.

The therapy was given according to the Maudsley Model, but the US study had a number of advantages, including a larger patient population (86 adolescents) and hence a greater statistical power. The therapy was given by therapists trained on a manual-based form of family therapy.

In comparison with the 2000 Maudsley Study[109] the patients were not so ill, (mean BMI 17.1). The principal measures of outcome were body weight and the eating disorder examination (EDE). There were no statistically significant differences between the two treatments.

The authors concluded that the short-term course of family was as effective as the long-term one for adolescents with a short duration anorexia nervosa. However, patients with more severe obsessive compulsive thinking, or coming from non-intact families, benefited from the longer treatment.

Strictly speaking, this RCT as such did not provide evidence of benefit from family therapy as there was no alternative form of treatment as a control. Nevertheless the outcomes at the end of 1 year's treatment were highly satisfactory with improvements not reported previously in natural outcome follow-up studies.

(d) Enduring benefits of family therapy

(i) 5-year follow-up with the first Maudsley RCT[110]

Patients who took part in the original Maudsley RCT of family therapy were followed up 5 years after the end of treatment in order to look for evidence of long-term benefit. Follow-up information was obtained on all 57 patients.[110] Apart from three deaths, there was a further overall improvement mainly attributable to the natural outcome of anorexia nervosa.

Within the subgroup of patients with early onset/short history, the mean weight achieved was 103 per cent ABW in patients who had received family therapy. On the Morgan/Russell scale, significantly greater improvement after family therapy was still discernable with a higher proportion of patients achieving a 'good' outcome and a higher average outcome score.

(ii) 5-year follow-up with the Maudsley adolescent RCT[112]

Of the original 40 patients 38 agreed to be reassessed. Compared with the end of treatment there were more patients in the good outcome category and fewer in the poor outcome category, and a further increase in weight. There were no deaths. The main purpose of this study had been to compare CFT and SFT. There was a significant difference between the two groups in the number of patients who resumed normal menstruation, higher in the SFT than in the CFT group. The findings at 1 year in patients from families with raised levels of maternal criticism (high EE) were still evident at 5 years. This group had done less well at the end of treatment in terms of weight and resumption of menstruation if they had been offered CFT.

(iii) 4-year follow-up with the US (Stanford) adolescent RCT[113]

Of the original 86 adolescent patients 71 were followed up. The long-term clinical outcome was good. Again there was little difference between patients treated with short-term or long-term family therapy. The BMI at follow-up for the short-term was 20.6, and the group as a whole showed good psychosocial functioning.

(e) Family therapy conclusions

There is now a strong body of evidence from RCTs that family therapy benefits younger patients with anorexia nervosa. The strength of the evidence in favour of family therapy has grown because of the increasing size of the patient populations tested: 21 in the 1987 Maudsley study, 40 in the 2000 Maudsley study, and 86 in the 2005 Stanford study.

1 From the search for specific components of the family therapy in the Maudsley studies, it can be concluded that the therapy should not be aimed at 'changing the family' but rather at helping them manage a sick family member.

2 Differences in the patients' responses to CFT and SFT were detected when there was a separate analysis for treatment outcome in patients in high and low expressed emotion (EE) families. From this it was concluded that SFT is the more appropriate treatment in families with high expressed emotion.

3 The Stanford study points towards the family therapy of shorter duration (6 months) being acceptable, except in the case of patients with obsessive compulsive thinking or in split families.

4 Finally, it is impressive that the quality of the clinical outcome increases in the long-term follow-ups. The patients in the 1987 Maudsley study showed the best outcome at the end of 5 years. This is not simply due to an expected good natural outcome, as the benefits of family therapy were still evident after 5 years.

Cognitive behavioural therapy

A theory for faulty cognitions maintaining the illness has been put forward by Fairburn *et al.*[114] The argument for examining the role of faulty cognitions in anorexia nervosa is inescapable. The original description of perceptual and conceptual disturbances in anorexia nervosa was put forward by Bruch in 1962.[5] It was appreciated that faulty attitudes to body size contributed in part to the patient's determination to reduce her food intake and lose weight. These observations led to the development of cognitive behavioural therapy (CBT) for anorexia nervosa. At first the evidence of its benefit relied on clinical impressions and case reports.[115] In recent years there have been valiant attempts to assess CBT in adult anorexic patients. It must be said, however, that these studies have met with limited success.

(a) CBT as a post-hospitalization treatment

The most fruitful RCT was conducted at the New York State Psychiatric Institute.[116] The design depended on the random allocation to CBT or the control treatment nutritional counselling (NC) of outpatients after discharge from an inpatient programme. Patients were eligible to enter the outpatient trial when they reached at least 90 per cent of ideal body weight. 33 who were eligible were randomly assigned to one or other treatment. The value of each treatment was assessed by its ability to prevent a relapse, defined as a fall in weight below a BMI of 17.5, or the development of medical or psychiatric complications.

Overall treatment failure was counted as the number of patients who had relapsed plus those who had dropped out from the treatment. The main finding is that the CBT group had a lower relapse and dropout rate and a better clinical outcome than the NC group.

(b) Comparison of CBT with a non-specific supportive clinical management (NSCM)

This study was conducted in Christchurch, New Zealand.[117] In this randomized controlled trial three therapies were compared:

1 CBT

2 Interpersonal psychotherapy

3 Non-specific supportive clinical management (NSCM).

The hypothesis was that the two specialized psychotherapies would be more effective than NSCM.

56 women were enrolled in the study. In fact the mean weights at baseline were not all that low (mean BMI 17.3). Twenty sessions were provided over a course of 20 weeks. The main assessments were a Global Assessment of Functioning (GAF) scale designed by the authors, body weight and the EDE.

The improvement in the group as a whole was perhaps disappointing with mean weight gains of 2.2 kg for CBT to 4.0 kg for NSCM. The hypothesis was not upheld and the specific therapies were less good than the non-specific supportive clinical management. The main lesson from this study is that good general clinical care combined with supportive psychotherapy can be at least as good as CBT.

(c) Multi-centred US study of CBT on its own or combined with medication[106]

An RCT was conducted in three well-established centres in the United States (White Plains New York, Minneapolis, and Stanford). In the original design there were three treatment groups:

1 Medication only

2 CBT only

3 Combined CBT and medication

Altogether 122 patients were randomized to one or other of these treatments. The patients were mainly adult. Their mean BMI (17.7–17.9) suggested the patients were not extremely thin. Intensive treatment was offered for up to 1 year. The patients were withdrawn from the study if they were 'treatment failures'. There were only low rates of completion: 27 per cent of the patients allocated medication, 43 per cent in the CBT group and 41 per cent in the combination group.

The methodological problems prevented an evaluation of the relative effectiveness of the treatments. The overall conclusions were somewhat pessimistic:

> 'It appears premature to conduct randomized controlled trials for *adults with anorexia nervosa* until the reasons for poor acceptance and high dropout rates . . . have been identified, and methods to remedy these serious problems have been devised'.

In fact the investigators were unlucky with the high rate of non-compliance; other studies offering psychotherapy to adult anorexic patients did not encounter such high dropout rates.[118]

The specialist eating disorder service

There is evidence that the treatment of anorexia nervosa is superior in specialized eating disorder units (EDUs) to that of general psychiatric units (adult or adolescent) or general medical units. The EDU comprises a team of professionals with training in the treatment of patients with anorexia nervosa and in maintaining the particular therapeutic ethos that is required. An EDU used to provide mainly inpatient care, but in recent years this has been extended to day and outpatient care. The clinical staff consist of a wide range of personnel including a psychiatric leadership and a trained nursing staff. The psychotherapeutic skills may vary but should include family therapy. The ethos of the EDU needs to combine a mix of therapeutic empathy for the patient with the ability to persuade her to return to a normal weight and eating habits. It is essential that the EDU nurse should be able to combine these skills.

This ethos was established in EDUs set-up in London in the 1970s and 1980s, particularly at St George's Hospital, the Royal Free Hospital and the Maudsley Hospital. It is from there that evidence has arisen for the effectiveness of a specialized eating disorder service.

An ambitious study from St George's Hospital, London, initially aimed at evaluating the advantages of three treatment settings.[96] Fortunately the study incorporated a control group and the

comparison with the active treatment groups yielded the most important results. The four groups were as follows:

1 Inpatient treatment: 4 months on average followed by outpatient psychotherapy.

2 Outpatient treatment combining individual and family therapy.

3 Outpatient group therapy.

4 The control group provided no treatment at the EDU after a 'one-off' evaluation.

90 patients were randomly allocated to one or other of the four options. The study began in the mid-1980s when uncertainty in the subject still warranted from an ethical point of view a random allocation to four groups, including one in which no treatment was provided but the patients returned to the care of their general practitioners. There were methodological difficulties. The patients sometimes dropped out of the treatment when it was not what they wished: out of 30 patients offered inpatient treatment, only 18 accepted it.

At 1 year follow-up there were few clear-cut differences between the patients in the three treatment groups: the inpatients' weights were similar to the outpatients. The clearest finding was that patients allocated to any one of the three treatment groups fared much better than those allocated to the one-off evaluation session only.

The main lesson from this valiant study is that care within an ED service, whether inpatient or outpatient, and irrespective of the specific treatment modality, is superior to treatment received outside an ED service.

(a) Day care and community treatment

There is a paucity of controlled trials of day programmes. An exception is the comparison of inpatient and day treatment carried out in Edinburgh.[119] A traditional inpatient programme, aimed strictly at weight gain, was compared with a more permissive day programme stressing open communication and patient autonomy. The day programme consisted of intensive psychological treatments and was available on 4 days each week. 32 patients who would have merited admission to hospital were randomly allocated to the day programme or the inpatient programme. Although statistically significant differences were not found between the two methods of treatment, the author reported interesting trends in improved outcome using the Morgan/Russell scales. The only advantage conferred by inpatient treatment was a slightly greater weight gain. In contrast, the day programme was more popular with patients who preferred to have a greater say in their rate of weight gain. There was also a greater return of personal autonomy.

(b) Inpatient treatment

Inpatient treatment in recent years has attracted a bad press. It has been argued that there are few RCTs, which have thrown any light on the benefits of inpatient treatments. This is a blinkered approach because RCTs addressed to some of the questions, are either inappropriate or ethically impossible to carry-out. An example of a largely inappropriate question from a clinical point of view is the comparison of inpatient and outpatient treatment in anorexia nervosa because the indications and the patient's suitability for each are very different. It is justified to state this categorically because there now exists solid clinical experience to show what can be achieved by inpatient treatment. The limitations of inpatient treatment are that the benefits may last only a number of months when there is a variable likelihood of relapse. A useful research question would be to identify patients who are most likely to relapse after inpatient treatment. The firm statement made above is warranted because of the valiant study by Crisp and his colleagues[120] previously discussed. Today this study would probably not be possible from an ethical point of view.

The most obvious benefit from inpatient treatment in an EDU is that this is the surest and safest method of improving the patient's nutrition and correcting her weight loss, thereby reversing physical complications, and sometimes saving her life. In various cohort studies the weight gain has varied around 12 kg in 12 weeks.[120–122] There are no comparable rapid improvements with outpatient treatments, not even family therapy in adolescents.

There have been a small number of RCTs in anorexia nervosa in addition to the Crisp *et al* study, and they have addressed narrower questions. For example, a comparison of two different inpatient treatment programmes has yielded the helpful finding that a strict operant conditioning programme offers no advantages over a 'lenient' programme.[123] The value of a specialized eating disorder unit in inducing weight gain was also demonstrated in a cohort of patients admitted to the Maudsley Hospital in 1990. They had previously been admitted to a general psychiatric or medical unit, and information on their previous weights was obtained. In the case of the Maudsley admissions the mean weight rose from 65.8 per cent average body weight (abw) to 90.4 per cent abw; in comparison the admissions elsewhere only led to a mean weight gain from 64.4 to 75 per cent abw.[122]

(c) Compulsory treatments

A study of the use of compulsory treatment in patients admitted to the Eating Disorders Unit at the Maudsley Hospital, comprised 81 patients or16 per cent of admissions. Section 3 of the Mental Health Act, valid for up to 6 months, was the most frequently applied section.[121] The compulsorily admitted patients were compared with a group of voluntary patients. The need for a compulsory admission was found to have two dimensions—the presence of a severe illness and a rejection of treatment. The compulsory patients gained at least as much weight as the voluntary patients but required a longer admission for them to return to a near-normal weight. The compulsory patients represented a selected group by virtue of a more entrenched reluctance to accept treatment. Accordingly it was predicted that in the long term they would fare less well than voluntary patients. The mortality rate of these patients was determined by the National Register, which provided the data at a mean of 5.7 years after the index admission. Ten out of 79 detained patients had died in comparison with two out of 78 voluntary patients, a statistically significant difference. The deaths among the compulsory patients were all due to anorexia nervosa or one of its nutritional complications. Thus the mortality rate among compulsory patients was extremely high at 2.17 per cent per annum, presumably skewed because of the selection factors. Compulsory treatment is an obvious example of the inappropriateness of an RCT.

Treatment: advice on management

It is not possible for the clinician to confine his treatment of anorexia nervosa to those methods that have been subjected to

randomized control trials. This does not mean that his treatment is ineffective or unsupported by evidence. The evidence has to be derived from other kinds of studies. They should not be dismissed as mere clinical experience if they can be supported by evidence-based clinical or experimental observations.

The main obstacle to treatment

The avoidance of treatment by the patient is part and parcel of anorexia nervosa and accentuated by her capacity for denial (see p. 783 of this chapter). There have been attempts to predict the likely level of the patient's compliance with treatment. For example, a 'transtheoretical model of change' is aimed at improving the patient's motivation by overcoming ambivalence while at the same time avoiding confrontation.[124] Different stages in the patient's approach to treatment have been recognized: precontemplation, contemplation, preparation, action, maintenance, and relapse prevention. Motivational enhancement therapy is at an early stage of development with only preliminary information on its impact on treatment outcome.

At our present level of knowledge it is simplest to ascertain the patient's attitude to treatment through the mental state examination. At the initial interview she should be asked what weight she would be willing to reach. At this stage it is best to refrain from challenging the patient's weight threshold, even though it will be well below a reasonable therapeutic goal. Having ascertained the limited degree of compliance, the clinician needs to develop a strategy to improve it gradually as treatment proceeds. It is poor clinical practice to place all the onus on the patient herself for accepting or rejecting a package of treatment at the first interview, or even at a later stage. 'Engaging' the patient in treatment is an essential part of most psychotherapeutic methods including CBT.[115] Another tactic for improving the patient's co-operation is to enlist the help of close members of the family. Young patients are likely in any case to require a family treatment.

Inpatient treatment

Although inpatient treatment is less often employed nowadays it will be described first because it is most important in patients who present as emergencies with an urgent need to preserve life and correct serious physical and psychiatric complications.

(a) Indications for admission

- It is tempting to specify a BMI threshold for admission, but it should be remembered that the BMI can only provide a very rough guide with many exceptions. If the BMI is below 14 admission should always be considered, all the more if there has been rapid weight loss during recent weeks. Indeed a BMI of 16 may cause anxiety if it represents a drop in weight from a BMI of 21 in the course of 2 or 3 months. Close attention should be given to the physical manifestations of malnutrition, especially hypoglycaemia and electrolyte disturbances.
- **The BMI may be totally unreliable in patients who have developed oedema as the result of malnutrition, causing a vicarious weight gain, which may deceive the patient and even the clinician into a false sense of security.**
- The BMI is also less reliable in children and adolescents, especially if their growth in height has been retarded thus distorting the calculation for the BMI towards a falsely reassuring value.
- Dangerous or disabling physical complications:
- Emergencies may result from low blood glucose levels, hypokalaemia or hyponatraemia. The electrolyte disturbances are most likely to occur in the presence of self-induced vomiting or laxative abuse. Hyponatraemia may be induced by polydipsia. Signs of proximal myopathy also indicate a need for admission, as do the occurrence of anaemia, low platelet counts (sometimes accompanied by purpura) and impaired liver function tests.
- Severe depression or obsessional behaviour may also arise in part as complications of malnutrition. Suicidal ideas may require admission, but persistent depression of lesser severity may also be an indication, because it renders the patient less amenable to outpatient psychological treatments, as does severe obsessional behaviour affecting the patient's eating pattern.
- Intractable self-induced vomiting or laxative abuse or determined fasting may require a greater supervision than is possible in an outpatient setting. These behaviours are less disconcerting in patients with bulimia and at a reasonable weight, but a severe weight loss makes it probable that vomiting will result in electrolyte disturbances or other physical complications.
- The decision to admit is often determined by the patient not engaging in the recommended outpatient psychological treatment when the malnutrition persists several months to 1 year.
- Young patients with a retarded puberty (premenarchal anorexia nervosa) should elicit an urgent correction of malnutrition, especially if there is evidence of retarded growth and short stature. There is a narrow window of opportunity for regaining stature, which may depend partly on bone age but it is safest to assume that after the age of 16 the probability of resuming growth becomes increasingly limited, especially in girls.

The great advantage of treating a patient in a specialized eating unit is the certainty that considerable benefit will accrue, including a substantial gain in weight, so long as the patient can be persuaded to remain in hospital.[120]

(b) Nursing and dietary care

Inpatient treatment will include a wide range of psychotherapeutic interventions (individual, group, and family) as well as occupational therapy and an educational programme. But the sheet anchor of successful treatment is a well-trained nursing staff working as a team. The role of the medical staff is one of maintaining a high level of expectation that the patient's weight will be restored to a normal (or healthy) level. It is necessary to give the nursing staff moral support so that they can develop their confidence and skills.

There are two main components to the nurses' treatment: their psychotherapeutic input and their supervisory role. The latter should never be draconian. The nursing team establish a relationship of trust with the patient and get to know her personal needs and concerns. The nurses should also be acutely aware of the anorexic patient's tendency to avoid food and exercise excessively. The treatment programme should stress the supportive aspects of the nurse's relationship with her patient rather than the undoubted need for careful supervision. The nurse will come to rely on the daily weight record to monitor the success of her treatment as the weight chart should show a smoothly rising curve. All meals are taken in the ward: thus the anorexic patients constitute a group in which peer interactions take place. The meal is taken in the presence

of one or more nurses also seated at the table. The patient learns that she is expected to consume all the food placed before her.

A detailed protocol for the refeeding programme has been produced by Andersen and his group.[120] It is not only the patient who tends to underestimate the food requirements to restore her weight to normal. Metabolic studies have demonstrated that for each kilogram of weight gain a surplus calorie balance of 7500 cal is needed.[111] It is prudent to begin with a modest calorie intake of 1200 to 1500 cal daily during the first 7 days in order to avoid the complications of hypophosphataemia and acute gastric dilation. Thereafter the caloric intake is gradually increased and may rise to 4000 cal. daily. The best diet is that consisting of a wide range of foods including carbohydrate and fat-containing foods. Concentrated foods (e.g. Build-up, Complan, Scandishake, Fortisip, Ensure Plus) may be used to achieve a high caloric intake. The aim is to achieve a positive energy balance of at least 1500 cal daily, leading to a weekly weight gain of 1 to 1.5 kg.

(c) Assessment of progress

Weighing should be a standardized daily procedure before breakfast after the patient has emptied her bladder and while wearing light night clothes. A paradoxical psychological improvement, with a diminution in concern with body size and shape, occurs with weight gain. The improvement is partly through the correction of malnutrition and partly the result of the 'exposure treatment' whereby the patient gradually accepts a higher body weight.

(d) Medication

Exceptionally the patient's tension and depression do not improve and there is a continued resistance to food. It may then be helpful to prescribe moderate doses of olanzapine (not more than 10 mg daily), carefully avoiding a fall in blood pressure, which is a risk in the emaciated patient. In the case of persistent depression, treatment with an antidepressant may be indicated. However, antidepressants are often ineffective in the presence of malnutrition, and by themselves do not assist the patient to gain weight.

(e) General measures

Inevitably the patient will find it irksome to forego home visits during the early period of weight gain. Therefore interesting and therapeutic activities should be provided through group meetings, occupational therapy, and social interactions. Visiting is generally encouraged unless the visiting parents are subjected to emotional appeals to take her home. They may then be asked to postpone their visits or reduce their duration. The monotony can be relieved when the patient's weight gain is on course by providing accompanied outings avoiding mealtimes.

The aim is to restore body weight to a healthy level within 8 to 10 weeks; a further period (usually 2 weeks) in hospital is needed to allow the patient to test her ability to maintain her weight by eating in a general dining room or going on home leave for two or more days at a time. The aim is to effect a smooth transition to day care or outpatient treatment.

Compulsory treatment in anorexia nervosa

The management of patients reluctant to accept essential therapeutic goals requires that they should be gradually engaged in a genuine alliance. However, there remain a minority of patients with whom this strategy fails and whose health and life become endangered. For them, compulsory treatment should be considered.[121] A compulsory admission to hospital is indicated not only when patients threaten suicide or suffer from a life-threatening malnutrition, but also when they fail to respond to simpler measures such as outpatient treatment or day care, or in the event of avoiding treatment altogether. Ill health persisting over the course of several months or the development of serious physical complications (e.g. water and electrolyte imbalance, hypoglycaemia, or myopathy) should also elicit compulsory admission if the patient cannot be persuaded to accept inpatient treatment voluntarily.

The evidence points to the usefulness of a compulsory admission in appropriate circumstances in so far as the patients respond well in the short term. Nevertheless, a patient who has required a compulsory admission carries a higher risk, so that it is essential to safeguard her through a long follow-up.

Day and community treatment

The Edinburgh trial of a day patient programme has already been discussed. At the Toronto Hospital a day hospital programme for eating disorders has been established since 1985.[125] and is now offered on 4 days a week.

The goals are as follows:

1 A normalization of disturbed eating behaviour and weight gain.

2 The identification of psychological and familial processes that serve to perpetuate the eating disorder.

Two meals and a snack are provided during the treatment hours. The staff take turns to supervise the patients during meal times. The psychological treatment consists of intensive group therapy addressing disturbed behaviours around eating and weight, and more general conflicts.

The clinical advantages of day treatment are a reduction in the dependence of patients who need to be able to function outside the hospital. The group treatment provides an atmosphere of mutual support while permitting interventions through group pressures. A wide range of patients can be admitted but those with medical or suicidal risks will elicit inpatient care instead. When patients succeed in reducing their disturbed eating behaviours they may 'act out' by self-harm. The clinical staff may find their skills severely taxed by the continuous staff/patient interaction.

A home oriented service extended to outpatient and day care, has been devised by Robinson (2006) for an area of North London (1.2 million). A wide range of treatments were devised including family interventions and carer support. The educational needs of members of the multi-disciplinary team are well described as are administrative and financial issues. Robinson's book includes an audit on the use of hospital and hostel beds requiring only 4–5 total beds per million population per year.

Outpatient psychotherapies

In the event of a patient's weight loss being less than 20 per cent of her healthy weight, it may be possible to obtain a therapeutic response by outpatient treatment, including attendance at non-specialist general psychiatric clinics or child/adolescent psychiatric services. The feasibility of this approach will depend on the availability of appropriate psychotherapeutic resources. The clinician should guard himself against imaginary conflicts between a psychotherapeutic approach and recording the weight of the patient. It is never justified to accept apparent compliance with psychotherapy if the patient's weight continues to decline.

(a) Family treatments

(i) Conjoint family therapy

The frequency of the sessions is determined by clinical need: it averaged 10.5 over the course of 1 year in the Maudsley trial. There are three stages to the therapy:

1 In the first phase the parents are urged to identify their future joint attitude to the feeding pattern that should be adopted by their daughter. With a younger patient the therapist assumes that the parents would initially need to take control of their child's eating.

2 During the second phase the parents are urged to be present together at each meal so that they can reinforce each other's efforts in the practical task of ensuring an improved food intake and a steady weight gain.

3 When the patient's weight is under control, responsibility for continued weight gain is handed back to her. Discussions can then commence on more normal family concerns. With an adolescent patient the main focus is on achieving increased autonomy. The eventual aim is to establish healthy relationships with her parents without the eating disorder as a medium for communication.

(ii) Separated family therapy

As in conjoint family therapy the parents are given direct advice on how to manage their daughter's eating problem. The patient herself is provided with individual psychotherapy. The therapist provides counselling about abnormal attitudes to weight and emphasizes the weight issue until steady progress has been made. This method is often preferred by the patient and her parents, largely because it avoids confrontation. It is also easier to gain access to the patient's fears and conflicts.

(b) CBT

CBT has much in common with other methods of treatment including the refeeding programme described under inpatient treatments. It relies on building a positive therapeutic alliance between therapist and patient.

The patient's weight and food intake is monitored at each session and she is told her weight. She is encouraged to think of food as medication and to follow a meal plan. The patient is educated in the disturbances of bodily and psychological function consequent on the state of starvation. The content of the therapy may be divided into two 'tracks'. The first track includes an examination of the behaviours adopted by the patient in order to reduce her weight. The second track is more concerned with psychological themes such as self-esteem, perfectionism, interpersonal functioning, and family conflicts. By asking the patient to give her reasons for specific behaviours, the therapist discovers faulty beliefs and assumptions on her part. For example, the 'anorexic wish' is the patient's wish to recover from her disorder without gaining weight. She is gently persuaded that this is an impossible aim because her real psychological difficulties will remain inaccessible so long as her experiences are clouded by starvation and dieting. The patient also expresses a fear of 'losing control' meaning that she will run the risk of overeating and become fat. It is explained to the patient that her rigid 'control' overeating deprives her completely of choice, and that far from being in control the reality is the converse. It is also useful for the therapist to analyse the pros and cons of maintaining the disorder of anorexia nervosa. The patient often feels more uncomfortable at confronting the hidden rewards of remaining thin.

Having ascertained the particular meanings of attitudes and behaviours for the patient, she is helped to find more adaptive ways of achieving healthier goals, including more relaxed normal eating and weight gain.

(c) Dynamic psychotherapy: therapeutic model

(i) Crisp's model of anorexia nervosa as a flight from growth

Crisp has explained how his programme of treatment is based on his model of anorexia nervosa as a refuge from puberty, which the patient has found overwhelming. The youngster reverses her pubertal development by limiting her intake of food.[126] His treatment initially involves an intensive inpatient programme lasting 10 to 12 weeks followed by outpatient treatment for up to 6 years. The advantage of this extensive programme is that it enables the patient to accept gradually an increase in weight while facing up to the feelings of panic and helplessness that originally led her to arrest her puberty through self-starvation. This interpretation is presented to the patient and to her family so that they come to see the psychotherapy as a way of solving the problems.

The model is a useful one even within a more limited outpatient psychotherapeutic setting. Some patients will readily identify their distress when overwhelmed by powerful sexual feelings or when confronted with personal and social responsibilities perceived as the result of growing up. The aims of the psychotherapy are to support the patient while she is beginning to abandon the psychobiological regression of anorexia nervosa. In addition she is encouraged to broaden her perception of herself in ways that are no longer wholly dependent on her physical appearance but include an improved sense of competence and self-esteem. She is helped to tackle personal and social problems from which she had previously escaped so that she can address her own and her parents' concerns about sexuality.

Ethical and medico-legal issues

In the United Kingdom the Mental Health Act Commission[127] has clarified many of the doubts in the minds of clinicians and social workers called upon to consider a compulsory admission under the Mental Health Act 1983. It recognized that anorexia nervosa is a mental disorder within the meaning of the Act and that in some patients their ability to consent may be compromised by fears of obesity or denial of the consequences of their actions. The Mental Health Act Commission concluded that when the patient's health is seriously threatened by food refusal she may be detained in hospital so as to treat the self-imposed starvation. The Commission went as far as to state that nasogastric feeding can be a medical process forming an integral part of the treatment for anorexia nervosa, notwithstanding that nasogastric feeding is seldom required.

Prevention

Preventive measures have been aimed at eating disorders generally rather than just anorexia nervosa. The main approach has been school-based intervention programmes educating adolescent girls into the risks of dieting and other methods of weight control. The commonest outcome has been an increased knowledge about

eating disorders without a change in the behaviours, such as dieting, likely to cause them. One controlled study, six weekly sessions were provided by teachers on a wide range of topics covering attitudes and behaviours relevant to eating disorders including a 'non-dieting approach'. Positive changes were detected at a 6-month follow-up but they were modest in size and poorly sustained.[128]

Currently there are active campaigns by the Academy of Eating Disorders and Beat Eating Disorders (UK) to encourage the fashion industry not to employ thin models. Unilever has adopted a policy not to employ models with a BMI less than 18.5. John Lewis requires models to be not less than size 12.

Further information

Robinson, P.H. (2006). *Community treatment of eating disorders*. John Wiley, Chichester.

Russell, G.F.M. (1993). Social psychiatry of eating disorders. In *Principles of social psychiatry* (eds, D. Bhugra and J. Leff). pp. 273–97. Blackwell Science, Oxford.

References

1. Gull, W.W. (1874). Anorexia nervosa (apepsia hysterica, anorexia hysterica). *Transactions of the Clinical Society of London*, **7**, 22–8.
2. Lasègue, C. (1873). De l'anorexie hystérique. *Archives Générales de Médicine*, **21**, 385–403.
3. Brumberg, J.J. (1988). *Fasting girls: the emergence of anorexia nervosa as a modern disease*. Harvard University Press, Cambridge, MA.
4. Bell, R.M. (1985). *Holy anorexia*. University of Chicago Press, Chicago and London.
5. Bruch, H. (1962). Perceptual and conceptual disturbances in anorexia nervosa. *Psychosomatic Medicine*, **24**, 187–94.
6. Russell, G.F.M. (1970). Anorexia nervosa: its identity as an illness and its treatment. In *Modern trends in psychological medicine* (ed. J. Harding Price), pp. 131–64. Butterworths, Norwich.
7. Garner, D.M. and Garfinkel, P.E. (1979). The eating attitudes test: an index of the symptoms of anorexia nervosa. *Psychological Medicine*, **9**, 273–9.
8. Johnson-Sabine, E., Wood, K., Patton, G., *et al.* (1988). Abnormal eating attitudes in London schoolgirls—a prospective epidemiological study: factors associated with abnormal response on screening questionnaires. *Psychological Medicine*, **18**, 615–22.
9. Hoek, H.W. (1991). The incidence and prevalence of anorexia nervosa and bulimia nervosa in primary care. *Psychological Medicine*, **21**, 455–60.
10. van Son, G.E., van Hoeken, D., Bartelds, A.I., *et al.* (2006). Time trends in the incidence of eating disorders: a primary care study in the Netherlands. *The International Journal of Eating Disorders*, **39**, 565–9.
11. Garner, D.M. and Garfinkel, P.E. (1980). Socio-cultural factors in the development of anorexia nervosa. *Psychological Medicine*, **10**, 647–56.
12. Szmukler, G.I., Eisler, I., Gillies, C., *et al.* (1985). The implications of anorexia nervosa in a ballet school. *Journal of Psychiatric Research*, **19**, 177–82.
13. Crisp, A.H., Palmer, R.R.L., *et al.* (1976). How common is anorexia nervosa? A prevalence study. *The British Journal of Psychiatry*, **128**, 549–54.
14. Råstam, M., Gillberg, C., and Garton, M. (1989). Anorexia nervosa in a Swedish urban region. A population based study. *The British Journal of Psychiatry*, **155**, 642–6.
15. Russell, G.F.M. (1993). Social psychiatry of eating disorders. In *Principles of social psychiatry* (eds. D. Bhugra and J. Leff), pp. 273–97. Blackwell Science, Oxford.
16. Milos, G., Spindler, A., Schnyder, U., *et al.* (2004). Incidence of severe anorexia nervosa in Switzerland: 40 years of development. *The International Journal of Eating Disorders*, **35**, 250–61.
17. Lucas, A.R., Beard, C.M., O'Fallen, W.M., *et al.* (1991). 50-year trends in the incidence of anorexia nervosa in Rochester, Minn: a population-based survey. *The American Journal of Psychiatry*, **148**, 917–22.
18. Fosson, A., Knibbs, J., Bryant-Waugh, R., *et al.* (1987). Early onset anorexia nervosa. *Archives of Disease in Childhood*, **62**, 114–18.
19. Favaro, A., Ferrara, S., and Santonastaso, P. (2003). The spectrum of eating disorders in young women: a prevalence study in a general population sample. *Psychosomatic Medicine*, **65**, 701–8.
20. Blake Woodside, D., Garfinkel, P.E., Lin, E., *et al.* (2001). Comparisons of men with full or partial eating disorders, and women with eating disorders in the community. *The American Journal of Psychiatry*, **158**, 570–4.
21. Morgan, H.G. and Russell, G.F.M. (1975). Value of family background and clinical features as predictors of long-term outcome in anorexia nervosa: four-year follow-up study of 41 patients. *Psychological Medicine*, **5**, 355–71.
22. Russell, G.F.M. (1979). Bulimia nervosa: an ominous variant of anorexia nervosa. *Psychological Medicine*, **9**, 429–48.
23. Russell, G.F.M. (1995). Anorexia nervosa through time. In *Handbook of eating disorders: theory, treatment and research* (eds. G. Szmukler, C. Dare, and J. Treasure), pp. 5–17. Wiley, Chichester.
24. Currin, L., Schmidt, U., Treasure, J., *et al.* (2005). Time trends in eating disorder incidence. *The British Journal of Psychiatry*, **186**, 132–5.
25. Campbell, P.G. (1995). What would a causal explanation of the eating disorders look like? In *Handbook of eating disorders: theory, treatment and research* (eds. G. Szmukler, C. Dare, and J. Treasure), pp. 49–64. Wiley, Chichester.
26. Garfinkel, P.E. and Garner, D.M. (1982). *Anorexia nervosa: a multidimensional perspective*. Brunner-Mazel, New York.
27. Birnbaum, K. (1923). *Der Aufbau der Psychose*, pp. 6–7. Springer, Berlin.
28. Patton, G.C., Johnson-Sabine, E., Wood, K., *et al.* (1990). Abnormal eating attitudes in London schoolgirls-a prospective epidemiological study: outcome at twelve month follow-up. *Psychological Medicine*, **20**, 383–94.
29. Crisp, A.H. (1970). Premorbid factors in adult disorders of weight, with particular reference to primary anorexia nervosa (weight phobia). A literature review. *Journal of Psychosomatic Research*, **14**, 1–22.
30. Prince, R. (1985). The concept of culture-bound syndromes: anorexia nervosa and brain-fag. *Transcultural Psychiatric Research Review*, **22**, 117–21.
31. DiNicola, V.F. (1990). Anorexia multiforme: self-starvation in historical and cultural context. *Transcultural Psychiatric Research Review*, **27**, 165–96, 245–86.
32. Bryant-Waugh, R. and Lask, B. (1991). Anorexia nervosa in a group of Asian children living in Britain. *The British Journal of Psychiatry*, **158**, 229–33.
33. Mynors-Wallis, L., Treasure, J., and Chee, D. (1992). Life events and anorexia nervosa: differences between early and late onset cases. *The International Journal of Eating Disorders*, **4**, 369–75.
34. Beumont, P.J.V., Abraham, S.F., and Simson, K.G. (1981). The psychosexual histories of adolescent girls and young women with anorexia nervosa. *Psychological Medicine*, **11**, 131–40.
35. Tiller, J. and Treasure, J. (1998). Eating disorders precipitated by pregnancy. *European Eating Disorders Review*, **6**, 178–87.
36. Sloan, G. and Leichner, P. (1986). Is there a relationship between sexual abuse or incest and eating disorders? *Canadian Journal of Psychiatry*, **31**, 656–60.
37. Palmer, R.L., Oppenheimer, R., Dignon, A., *et al.* (1990). Childhood sexual experiences with adults reported by women with eating disorders: an extended series. *The British Journal of Psychiatry*, **156**, 699–703.

38. McClelland, L., Mynors-Wallis, L., Fahy, T., *et al.* (1991). Sexual abuse, disordered personality and eating disorders. *The British Journal of Psychiatry*, **158**(Suppl. 10), 63–8.
39. Wonderlich, S.A., Brewerton, T.D., Jocic, Z., *et al.* (1997). Relationship of childhood sexual abuse and eating disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, **36**, 1107–15.
40. Minuchin, S., Baker, L., Rosman, B.L., *et al.* (1975). A conceptual model of psychosomatic illness in children: family organization and family therapy. *Archives of General Psychiatry*, **32**, 1031–8.
41. Selvini Palazzoli, M. (1974). *Self-starvation. From the intrapsychic to the transpersonal approach to anorexia nervosa* (trans. A. Pomerans), pp. 202–16. Human Context Books, London.
42. Bruch, H. (1978). *The golden cage: the enigma of anorexia nervosa*. Open Books, London.
43. Dare, C. and Eisler, I. (1997). Family therapy for anorexia nervosa. In *Handbook of treatment for eating disorders* (2nd edn) (eds. D.M. Garner and P.E. Garfinkel), pp. 307–24. Guilford Press, New York.
44. Janet, P. (1903). L'obsession de la honte du corps. In *Les obsessions et la psychasthénie*, Vol. 1, Section 5. Germer Gaillière, Paris.
45. Wonderlich, S.A., Swift, W.H., Slotnick, H.B., *et al.* (1990). DSM-III-R personality disorders in eating disorder subtypes. *The International Journal of Eating Disorders*, **9**, 607–16.
46. Casper, R.C. (1990). Personality features of women with good outcome from restricting anorexia nervosa. *Psychosomatic Medicine*, **52**, 156–70.
47. Simmonds, K. (1916). Über kachexie hypophysaren ursprungs. *Deutsche Medizinische. Wochenschrift*, **42**, 190–1.
48. Russell, G.F.M. (1977). The present status of anorexia nervosa. *Psychological Medicine*, **7**, 363–7.
49. Anand, B.K. and Brobeck, J.R. (1951). Localization of a 'feeding center' in the hypothalamus of the rat. *Proceedings of the Society for Experimental Biology and Medicine*, **77**, 323–4.
50. Schweiger, U., Tuschl, R.J., Laessle, R.G., *et al.* (1989). Consequences of dieting and exercise on menstrual function in normal weight young women. In *The menstrual cycle and its disorders* (eds. K.M. Pirke, W. Wuttke, and U. Schweiger), pp. 142–9. Springer Verlag, Berlin.
51. DeVile, C.H., Sufraz, R., Lask, B., *et al.* (1995). Occult intracranial tumours masquerading as early onset anorexia nervosa. *British Medical Journal*, **311**, 1359–60.
52. Ploog, D.W. and Pirke, K.M. (1987). Psychobiology of anorexia nervosa. *Psychological Medicine*, **17**, 843–59.
53. Lask, B., Gordon, I., Christie, D., *et al.* (2005). Functional neuroimaging in early onset anorexia nervosa. *The International Journal of Eating Disorders*, **37**, S49–51.
54. Strober, M., Lampert, C., Morrell, W., *et al.* (1990). A controlled family study of anorexia nervosa: evidence of familial aggregation and lack of shared transmission with affective disorders. *The International Journal of Eating Disorders*, **9**, 239–53.
55. Bulik, C.M. (2004). Role of genetics in anorexia nervosa, bulimia nervosa and binge eating disorder. In *Clinical handbook of eating disorders* (ed. T.D. Brewerton), pp. 165–82. Marcel Dekker, New York.
56. Holland, A.J., Hall, A., Murray, R., *et al.* (1984). Anorexia nervosa: a study of 34 twin pairs and one set of triplets. *The British Journal of Psychiatry*, **145**, 414–19.
57. Holland, A.J., Sicotte, N., and Treasure, J. (1988). Anorexia nervosa: evidence for a genetic basis. *Journal of Psychosomatic Research*, **32**, 561–71.
58. Blundell, J.E. and Bill, A.J. (1991). Serotonin, eating disorders and the satiety cascade. In *Serotonin-related psychiatric syndromes: clinical and therapeutic links* (eds. G.B. Cassano and H.S. Asikal), pp. 125–9. Royal Society of Medicine Services, London.
59. Szmukler, G.I., Andrewes, D., Kingston, K., *et al.* (1992). Neuropsychological impairment in anorexia nervosa before and after refeeding. *Journal of Clinical and Experimental Neuropsychology*, **14**, 347–52.
60. Kaye, W.H., Gwirtsman, J.E., George, D.T., *et al.* (1988). CSF-5HIAA concentrations in anorexia nervosa: reduced values in underweight subjects normalize after weight gain. *Biological Psychiatry*, **23**, 102–5.
61. Kaye, W.H., Gwirtsman, H.E., George, D.T., *et al.* (1991). Altered serotonin activity in anorexia nervosa after long-term weight restoration: does elevated CSF-5HIAA correlate with rigid and obsessive behavior? *Archives of General Psychiatry*, **48**, 556–62.
62. O'Dwyer, A.-M., Lucey, J.V., and Russell, G.F.M. (1996). Serotonin activity in anorexia nervosa after long-term weight restoration: response to d-fenfluramine challenge. *Psychological Medicine*, **26**, 353–9.
63. Slade, P.D. and Russell, G.F.M. (1973). Awareness of body dimensions in anorexia nervosa: cross-sectional and longitudinal studies. *Psychological Medicine*, **3**, 188–99.
64. Hsu, L.K.G. and Sobkiewicz, T.A. (1989). Body image disturbance: time to abandon the concept for eating disorders? *The International Journal of Eating Disorders*, **10**, 15–30.
65. Hsu, L.K.G. and Lee, S. (1993). Is weight phobia always necessary for a diagnosis of anorexia nervosa? *The American Journal of Psychiatry*, **150**, 1466–71.
66. Blake Woodside, D. and Twose, R. (2004). Diagnostic issues in eating disorders: historical perspectives and thoughts for the future. In *Clinical handbook of eating disorders* (ed. T.D. Brewerton), pp. 1–19. Marcel Dekker, New York.
67. Vandereycken, W. (2007). Denial of illness in anorexia nervosa—a conceptual review: part I diagnostic significance and assessment. *European Eating Disorder Review*, **14**, 352–68.
68. Fichter, M.M. and Pirke, K.M. (1995). Starvation models and eating disorders. In *Handbook of eating disorders: theory, treatment and research* (eds. G. Szmukler, C. Dare, and J. Treasure), pp. 83–107. Wiley, Chichester.
69. Treasure, J.L., Gordon, P.A.L., King, E.A., *et al.* (1985). Cystic ovaries: a phase of anorexia nervosa. *Lancet*, **2**, 1379–82.
70. Treasure, J.L., Wheeler, M., King, E.A., *et al.* (1988). Weight gain and reproductive function: ultrasonographic and endocrine features in anorexia nervosa. *Clinical Endocrinology*, **29**, 607–16.
71. Namir, S., Melman, K.N., and Yager, J. (1986). Pregnancy in restrictor-type anorexia nervosa: a study of six women. *The International Journal of Eating Disorders*, **5**, 837–45.
72. Treasure, J.L. and Russell, G.F.M. (1988). Intrauterine growth and neonatal weight gain in babies of women with anorexia nervosa. *British Medical Journal*, **296**, 1038.
73. Van Wezel-Meijler, G. and Wit, J.M. (1989). The offspring of mothers with anorexia nervosa: a high-risk group for undernutrition and stunting? *European Journal of Pediatrics*, **49**, 130–5.
74. Treasure, J. and Szmukler, G. (1995). Medical complications of chronic anorexia nervosa. In *Handbook of eating disorders: theory, treatment and research* (eds. G. Szmukler, C. Dare, and J. Treasure), pp. 197–220. Wiley, Chichester.
75. Passmore, R. and Eastwood, M.A. (1986). *Human nutrition and dietetics*. Churchill Livingstone, Edinburgh.
76. Handler, C.E. and Pirkin, G.D. (1982). Anorexia nervosa and Wernicke's encephalopathy: an undiagnosed association. *Lancet*, **2**, 771–2.
77. Lask, B. and Bryant-Waugh, R. (1992). Early onset anorexia nervosa and related eating disorders. *Journal of Child Psychology and Psychiatry*, **33**, 281–300.
78. McLoughlin, D.M., Spargo, E., Wassif, W.S., *et al.* (1998). Structural and functional changes in skeletal muscle in anorexia nervosa. *Acta Neuropathologica*, **95**, 632–40.
79. Treasure, J.L., Fogelman, I., Russell, G.F.M., *et al.* (1987). Reversible bone loss in anorexia nervosa. *British Medical Journal*, **295**, 474–5.
80. World Health Organization. (1992). *International statistical classification of diseases and related health problems* (10th revision). WHO, Geneva.

81. Klibanski, A., Biller, B.M.K., Schoenfeld, D.A., *et al.* (1995). The effects of oestrogen administration on trabecular bone loss in young women with anorexia nervosa. *Journal of Clinical Endocrinology and Metabolism*, **80**, 898–904.
82. Serpell, L. and Treasure, J. (1997). Osteoporosis—a serious health risk in chronic anorexia nervosa. *European Eating Disorders Review*, **5**, 149–57.
83. Wakeling, A. and Russell, G.F.M. (1970). Disturbances in the regulation of body temperature in anorexia nervosa. *Psychological Medicine*, **1**, 30–9.
84. Rieger, W., Brady, J.P., and Weisberg, E. (1978). Hematologic changes in anorexia nervosa. *The American Journal of Psychiatry*, **135**, 984–5.
85. Russell, G.F.M. (1966). Acute dilation of the stomach in a patient with anorexia nervosa. *The British Journal of Psychiatry*, **112**, 203–7.
86. Beumont, P.J.V. and Large, M. (1991). Hypophosphataemia, delirium and cardiac arrhythmia in anorexia nervosa. *Medical Journal of Australia*, **155**, 519–22.
87. Russell, G.F.M., Treasure, J., and Eisler, I. (1998). Mothers with anorexia nervosa who underfeed their children: their recognition and management. *Psychological Medicine*, **28**, 93–108.
88. Russell, G.F.M. (1985). Premenarchal anorexia nervosa and its sequelae. *Journal of Psychiatric Research*, **19**, 363–9.
89. Lai, K.Y.C., de Bruyn, R., Lask, B., *et al.* (1994). Use of pelvic ultrasound to monitor ovarian and uterine maturity in childhood onset anorexia nervosa. *Archives of Disease in Childhood*, **71**, 228–31.
90. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
91. Beumont, P.J.V., Beardwood, C.J., and Russell, G.F.M. (1972). The occurrence of the syndrome of anorexia nervosa in male subjects. *Psychological Medicine*, **2**, 216–31.
92. Crisp, A.H. and Burns, T. (1990). Primary anorexia nervosa in the male and female: a comparison of clinical features and prognosis. In *Males with eating disorders* (ed. A.E. Andersen), pp. 77–99. Brunner–Mazel, New York.
93. Blake Woodside, D., Garfinkel, P.E., Lin, E., *et al.* (2001). Comparisons of men with full or partial eating disorders, and women with eating disorders in the community. *The American Journal of Psychiatry*, **158**, 570–4.
94. Burns, T. and Crisp, A.H. (1990). Outcome of anorexia nervosa in males. In *Males with eating disorders* (ed. A.E. Andersen), pp. 163–86. Brunner-Mazel, New York.
95. Fichter, M.M. and Daser, C. (1987). Symptomatology, psychosexual development and gender identity in 42 anorexic males. *Psychological Medicine*, **17**, 409–18.
96. Crisp, A.H., Norton, K., Gowers, S., *et al.* (1991). A controlled study of the effect of therapies aimed at adolescent and family psychopathology in anorexia nervosa. *The British Journal of Psychiatry*, **159**, 325–33.
97. Steinhausen, H.-C., Rauss-Mason, C., and Seidel, R. (1991). Follow-up studies of anorexia nervosa. A review of four decades of outcome research. *Psychological Medicine*, **21**, 447–54.
98. Morgan, H.G. and Hayward, A.E. (1988). Clinical assessment of anorexia nervosa. *The British Journal of Psychiatry*, **152**, 367–71.
99. Theander, S. (1985). Outcome and prognosis in anorexia nervosa and bulimia: some results of previous investigations, compared with a Swedish long-term study. *Journal of Psychosomatic Research*, **19**, 493–508.
100. Ratnasuriya, R.H., Eisler, I., Szmukler, G.I., *et al.* (1991). Anorexia nervosa: outcome and prognostic factors after 20 years. *The British Journal of Psychiatry*, **158**, 495–502.
101. Sullivan, P.F. (1995). Mortality in anorexia nervosa. *The American Journal of Psychiatry*, **152**, 1073–4.
102. Steinhausen, H.-C. and Seidel, R. (1993). Outcome in adolescent eating disorders. *The International Journal of Eating Disorders*, **14**, 487–96.
103. Wentz, E., Gillberg, C., Gillberg, I.C., *et al.* (2001). Ten-year follow-up of adolescent-onset anorexia nervosa: psychiatric disorders and overall functioning scales. *Journal of Child Psychology and Psychiatry*, **42**, 613–22.
104. Theander, S. (1992). Chronicity in anorexia nervosa: results from the Swedish long-term study. In *The course of eating disorders* (eds. W. Herzog, H.-C. Deter, and W. Vandereycken), pp. 198–213. Springer-Verlag, Berlin.
105. National Institute of Clinical Excellence. (2004). *Eating disorders: core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders*. Clinical Guideline, 9. National Collaborating Centre for Mental Health, London.
106. Halmi, K.A., Agras, S., Crow, S., *et al.* (2005). Predictors of treatment acceptance and completion in anorexia nervosa: implications for future study designs. *Archives of General Psychiatry*, **62**, 776–81.
107. Russell, G.F.M., Szmukler, G.I., Dare, C., *et al.* (1987). An evaluation of family therapy in anorexia nervosa and bulimia nervosa. *Archives of General Psychiatry*, **44**, 1047–56.
108. Lock, J., Agras, W.S., Bryson, S., *et al.* (2005). A comparison of short—and long-term family therapy for adolescent anorexia nervosa. *Journal of the American Academy of Child and Adolescent Psychiatry*, **44**, 632–9.
109. Eisler, I., Dare, C., Hodes, M., *et al.* (2000). Family therapy for adolescent anorexia nervosa: the results of a controlled comparison of two family interventions. *Journal of Child Psychology and Psychiatry*, **41**, 727–36.
110. Eisler, I., Dare, C., Russell, G.F.M., *et al.* (1997). Family and individual therapy in anorexia nervosa—a 5-year follow-up. *Archives of General Psychiatry*, **54**, 1025–30.
111. Russell, G.F.M. and Mezey, A.G. (1962). An analysis of weight gain in patients with anorexia nervosa treated with high calorie diets. *Clinical Science*, **23**, 449–61.
112. Eisler, I., Simic, M., Russell, G.F.M., *et al.* (2007). A randomized controlled treatment trial of two forms of family therapy in adolescent anorexia nervosa: a five-year follow-up. *Journal of Child Psychology and Psychiatry*.
113. Lock, J., Couturier, J., and Agras, W.S. (2006). Comparison of long-term outcomes in adolescents with anorexia nervosa treated with family therapy. *Journal of the American Academy of Child and Adolescent Psychiatry*, **45**, 666–72.
114. Fairburn, C.G., Shafran, R., and Cooper, Z. (1999). A cognitive-behavioural theory of anorexia nervosa. *Behaviour Research and Therapy*, **37**, 1–13.
115. Garner, D.M., Vitousek, K.M., and Pike, K.M. (1997). Cognitive-behavioural therapy for anorexia nervosa. In *Handbook of treatment for eating disorders* (2nd edn) (eds. D.M. Garner and P.E. Garfinkel), pp. 94–144. Guilford Press, New York.
116. Pike, K.M., Walsh, B.T., Vitousek, K., *et al.* (2003). Cognitive behaviour therapy in the posthospitalisation of anorexia nervosa. *The American Journal of Psychiatry*, **160**, 2046–9.
117. McIntosh, V.V.W., Jordan, J., Carter, F.A., *et al.* (2005). Three psychotherapies for anorexia nervosa: a randomized controlled trial. *The American Journal of Psychiatry*, **162**, 741–7.
118. Dare, C., Eisler, I., Russell, G., *et al.* (2001). Psychological therapies for adults with anorexia nervosa. *The British Journal of Psychiatry*, **178**, 216–21.
119. Freeman, C. (1992). Day patient treatment for anorexia nervosa. *British Review of Bulimia and Anorexia Nervosa*, **6**, 3–9.
120. Andersen, A.E., Bowers, W., and Evans, K. (1997). In-patient treatment of anorexia nervosa. In *Handbook of treatment for eating disorders* (2nd edn) (eds. D.M. Garner and P.E. Garfinkel), pp. 327–53. Guilford Press, New York.

121. Ramsay, R., Ward, A., Treasure, J., *et al.* (1999). Compulsory treatment in anorexia nervosa: short-term benefits and long-term mortality. *The British Journal of Psychiatry*, **175**, 147–53.
122. The Royal College of Psychiatrists. Eating Disorders. (1992). Council Report CR14, p. 13.
123. Touyz, S.O., Beumont, P.J.V., Glaun, D., *et al.* (1984). A comparison of lenient and strict operant conditioning programmes in refeeding patients with anorexia nervosa. *The British Journal of Psychiatry*, **144**, 517–20.
124. Prochaska, J.O. and Di Clemente, C.C. (1992). The transtheoretical approach. In *Handbook of psychotherapy integration* (eds. J.C. Norcross and M.R. Goldfried), pp. 300–34. Basic Books, New York.
125. Kaplan, A.S. and Olmsted, M.P. (1997). Partial hospitalisation. In *Handbook of treatment for eating disorders* (2nd edn) (eds. D.M. Garner and P.E. Garfinkel), pp. 354–60. Guilford Press, New York.
126. Crisp, A.H. (1997). Anorexia nervosa as flight from growth: assessment and treatment based on model. In *Handbook of treatment for eating disorders* (2nd edn) (eds. D.M. Garner and P.E. Garfinkel), pp. 248–77. Guilford Press, New York.
127. Mental Health Act Commission. (1997). *Guidance on the treatment of anorexia nervosa under the Mental Health Act 1983*, pp. 1–6. Mental Health Act Commission, Nottingham.
128. Stewart, D.A., Carter, J.C., Drinkwater, J., *et al.* (2001). Modification of eating attitudes and behaviour in adolescent girls: a controlled study. *The International Journal of Eating Disorders*, **29**, 107–18.

4.10.2 Bulimia nervosa

Christopher G. Fairburn,
Zafra Cooper, and Rebecca Murphy

Introduction

Origins of the concept

The history of the diagnosis bulimia nervosa begins as recently as 1979. It was in this year that Russell published his now seminal paper 'Bulimia nervosa: An ominous variant of anorexia nervosa'[1] in which he described 30 patients (28 women and 2 men), seen between 1972 and 1978, who had three major features in common. First, they had recurrent episodes of uncontrolled overeating; second, they regularly used self-induced vomiting or laxatives as means of weight control; and third, they had a morbid fear of becoming fat. Russell described many other features shared by these patients, including a history of anorexia nervosa (present in 80 per cent), the presence of severe depressive symptoms, and the fact that in most cases their body weight was in the healthy range. He noted that the disorder tended to run a chronic course and that it was 'extremely difficult to treat'. Finally, he suggested that this clinical picture should be viewed as a syndrome, distinct from anorexia nervosa, and he proposed the term 'bulimia nervosa'.

It is difficult to exaggerate the importance of Russell's paper. Its greatest contribution was perhaps its prescience in that it crystallized out from among the range of eating disorders seen in clinical practice a subgroup of patients that was just starting to become more common; it identified its central features; and it gave it an appropriate name.

Events since 1980

Events gathered pace in the 1980s. In 1980 a syndrome termed 'bulimia' was included in DSM-III.[2] This was intended to denote the type of patient that Russell had described, although its diagnostic criteria proved to be overly inclusive. In 1987, the criteria were refined and brought more in line with Russell's original concept. The syndrome was also renamed bulimia nervosa.[3] Also in the early 1980s evidence mounted that bulimia nervosa might be common and this led to a spate of studies of its prevalence. At the same time reports were published describing the successful treatment of these patients, the two most promising approaches being a specific form of cognitive behaviour therapy and the use of antidepressant drugs. By the mid-1980s, both treatments had been tested in the first of what has become a large series of controlled trials.

Now, three decades later, the diagnosis bulimia nervosa is included in both DSM-IV[4] and ICD-10,[5] its prevalence is established, aspects of its aetiology are beginning to be understood, and much has been learned about its treatment.

Classification and diagnosis

The classification of the eating disorders and their principal diagnostic criteria are shown in Table 4.10.2.1. Bulimia nervosa is one of the two main eating disorders recognized in DSM-IV and ICD-10, the other being anorexia nervosa (discussed in Chapter 4.10.1). In addition, in DSM-IV there is a residual category termed 'eating disorder not otherwise specified'. This is reserved for eating disorders of clinical severity that do not meet the diagnostic criteria for anorexia nervosa or bulimia nervosa.[6] In ICD-10, various eating disorder categories other than anorexia nervosa and bulimia nervosa are recognized (for example, atypical anorexia nervosa, atypical bulimia nervosa, overeating associated with other psychological disturbances), although these concepts have never been adequately defined or differentiated.

The relationship between the three DSM-IV diagnoses is represented diagrammatically in Fig. 4.10.2.1. The two overlapping inner circles represent anorexia nervosa (the smaller circle) and bulimia nervosa (the larger circle) respectively, the area of potential overlap being that occupied by those people who would meet the diagnostic criteria for both disorders but for the rule that the diagnosis of anorexia nervosa trumps that of bulimia nervosa (see Table 4.10.2.1). Surrounding these two circles is an outer circle which defines the boundary between having an eating disorder, a state of clinical significance, and having a lesser, non-clinical, eating problem. It is this boundary that demarcates what is, and is not, an eating disorder. Within the outer circle, but outside the two inner circles, lies eating disorder not otherwise specified (eating disorder NOS).

In DSM-IV a new eating disorder diagnosis was proposed termed 'binge eating disorder'. It is designed to denote an eating problem characterized by recurrent binge eating in the absence of the extreme weight-control behaviour seen in bulimia nervosa. Since binge eating disorder is a provisional diagnostic concept, it is currently an example of eating disorder NOS.

The two schemes for classifying eating disorders encourage the view that anorexia nervosa and bulimia nervosa are distinct clinical states. Consideration of their clinical features and course over time does not support this.[7] Binge eating disorder aside, patients with anorexia nervosa, bulimia nervosa, and eating disorder NOS have

Table 4.10.2.1 Classification of eating disorders and their principal diagnostic criteria.

Classification of eating disorders
◆ Anorexia nervosa
◆ Bulimia nervosa
◆ Eating disorder not otherwise specified (eating disorder NOS)
Binge eating disorder (a provisional new diagnosis, currently subsumed under eating disorder NOS)
Principal diagnostic criteria
Anorexia nervosa
1 Over-evaluation of shape and weight (i.e. judging self-worth largely, or exclusively, in terms of shape and weight)
2 Active maintenance of an unduly low body weight (e.g. body mass index < 17.5)
3 Amenorrhoea (in post-menarcheal females who are not taking an oral contraceptive). This criterion is often omitted.
Bulimia nervosa
1 Over-evaluation of shape and weight (i.e. judging self-worth largely, or exclusively, in terms of shape and weight)
2 Recurrent binge eating (i.e. recurrent episodes of uncontrolled overeating)
3 Extreme weight-control behaviour (e.g. strict dieting, frequent self- induced vomiting, or laxative misuse)
4 Diagnostic criteria for anorexia nervosa are not met
Eating disorder not otherwise specified
Eating disorders of clinical severity that do not conform to the diagnostic criteria for anorexia nervosa or bulimia nervosa
Binge eating disorder
Recurrent binge eating in the absence of the extreme weight-control behaviour seen in bulimia nervosa

(Reproduced from Fairburn, C.G. Cognitive Behaviour Therapy and Eating Disorders, copyright 2008, Guildford Press, NY.)

many features in common, most of which are not seen in other psychiatric disorders, and studies of their course indicate that patients migrate between these diagnoses over time: indeed, temporal migration is the norm rather than the exception. This temporal movement, together with the fact that the disorders share the same distinctive psychopathology, has led to the suggestion that the current diagnostic scheme is a poor reflection of clinical reality and that common 'transdiagnostic' mechanisms may be involved in the maintenance of eating disorder psychopathology.[8]

Clinical features

The great majority of patients with bulimia nervosa are female and most are in their 20s (although the age range is between 10 and 60 years). In considering the psychopathology of the disorder, a distinction may be drawn between its 'specific' and 'general' features. The former comprises features that are largely peculiar to eating disorders (for example, self-induced vomiting, the over-evaluation of shape and weight), whereas the latter consists of features seen in other psychiatric conditions (for example, depressive symptoms). The clinical features of bulimia nervosa are similar in men and women and in those with and without a history of anorexia nervosa.

Specific psychopathology

(a) Dieting and binge eating

The eating habits of patients with bulimia nervosa are characterized by strict dieting punctuated by repeated episodes of binge eating (see Fig. 4.10.2.2). The dieting is extreme and it is governed by multiple self-imposed dietary rules. These rules tend to be applied to all aspects of eating, including when to eat, what to eat, and how much to eat. As a result, the food eaten (when not binge eating) is restricted in quantity and range.

Recurrent episodes of 'binge eating' interrupt this dieting. (The term binge eating denotes discrete episodes of eating that have two characteristics: first, an unusually large amount of food is eaten,

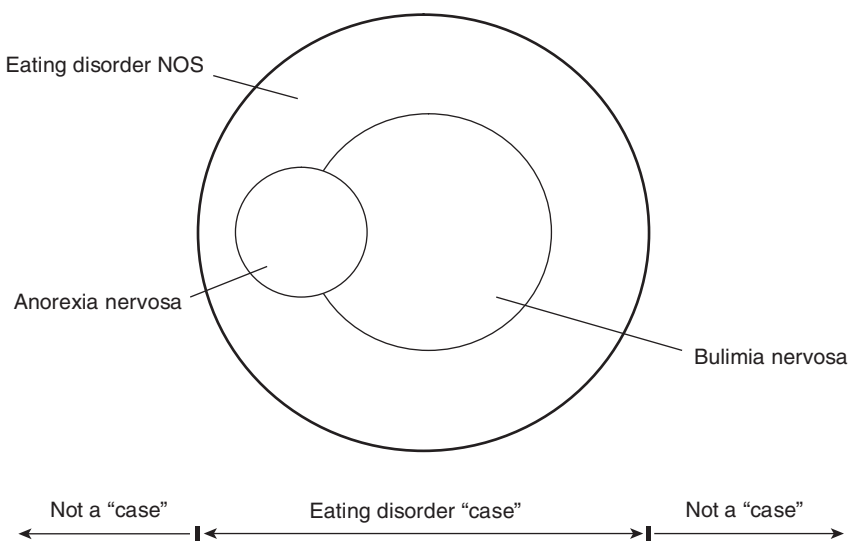


Fig. 4.10.2.1 A schematic representation of the relationship between anorexia nervosa, bulimia nervosa, and eating disorder NOS (Reprinted from Fairburn, C.G. and Bohn, K. Eating disorder NOS (EDNOS): an example of the troublesome "not otherwise specified" (NOS) category in DSM-IV, *Behaviour Research and Therapy*, **43**, 691–701, copyright (2005), with permission from Elsevier).

Day.....Thursday..... Date .....March 21 .....

Time	Food and drink consumed	Place	*	V/L	Context and comments
7.30	Glass water	Kitchen			[8stone 9lbs - really gross] Thirsty after yesterday
8:10	1 bowl muesli with skimmed milk Black coffee	Kitchen	*		Should have had less muesli. Must not binge today.
10:35	Half banana Black coffee	Work - at desk			Better - on track
11:45	Ham and lettuce slimline sandwich Diet coke	In canteen			Usual lunch
6.40 to 7.30	Slice of chocolate cake 1/2 tub ice cream 4 slices of toast with jam Diet coke Another slide of cake 2 slices of toast with jam Diet coke Jam from jar Two kit kats Mars bar Diet coke - large	Kitchen	*	V	Help - I can't stop eating. I'm completely out of control. I hate myself. I am disgusting. Why do I do this? I started as soon as I got in. I've ruined another day.
9:30	Rice cake with fat-free cheese Diet coke	Kitchen			Really lonely. Feel fat and unattractive. Feel like giving up.

Fig. 4.10.2.2 A monitoring record illustrating the eating habits of a patient with bulimia nervosa. Asterisks are used to signify episodes of eating that were viewed by the patient as excessive. The column headed 'V/L' is for recording episodes of self-induced vomiting or laxative misuse.

given the circumstances; and second, there is a sense of loss of control at the time. Some patients with eating disorders have similar episodes of uncontrolled overeating that do not involve the consumption of objectively large amounts of food. These episodes are sometimes referred to as 'subjective binges' although technically speaking they do not meet the definition of a 'binge'.) The frequency and regularity of the binge eating varies. Some patients have episodes almost every day, whereas in others the episodes are intermittent. In DSM-IV, it is specified that the binges should occur on average at least twice a week, but this is an arbitrary figure that has little discriminatory value. Among those patients in whom the binge eating is frequent, the binges have few, if any, obvious triggers, although there may be circumstances under which binge eating is more likely (for example, when alone at home). Among patients in whom the binge eating is less frequent, the binges often have clear precipitants. These tend to be of three overlapping types: first, there is breaking a personal dietary rule (for example, exceeding a daily calorie-limit or eating a banned food); second, there are situations which intensify concerns about shape and weight (for example, receiving an adverse comment about appearance); and third, there is the occurrence of negative moods (often as a result of interpersonal events). All three undermine the maintenance of strict dietary control.

The amount of food eaten during binges varies, both from patient to patient and from episode to episode. Typical episodes involve the consumption of 1000 to 2000 kcal.[9] The food eaten generally comprises items that are otherwise being avoided. Thus binges tend to be composed of energy-dense, high-fat items such as chocolate, ice cream, and pastries. Binges come to an end as a result of the combined influence of exhaustion, extreme fullness, a diminution of the drive to eat, and the running out of food supplies. In about three-quarters of patients they are immediately followed by measures designed to counteract the effects of the overeating, the most common being self-induced vomiting and the taking of laxatives or diuretics.

The binges are a source of considerable distress. They magnify these patients' fears of weight gain and fatness, and they may result in shame and self-disgust. For this reason most binges occur in private and are kept secret from others. It is the binge eating that eventually drives these people to seek help.

(b) Purging and other forms of weight control

In DSM-IV bulimia nervosa is subdivided into two types, a purging and non-purging type. In the purging type there is regular self-induced vomiting or the misuse of laxatives or diuretics, or both, whereas in the non-purging type such behaviour ('purging') is either not present or it is infrequent. The majority of patients seen in clinical practice have the purging form of the disorder and it has been the focus of most research.

Self-induced vomiting is the most common form of purging. In most patients it only takes place after binge eating. It is generally achieved by stimulating the gag reflex, using the fingers or some

other long object, although in more established cases it can be accomplished with no mechanical aid. The vomiting is repeated until patients think that they have retrieved all the food that they can. Patients get extremely distressed if they are unable to vomit after binge eating: indeed, if they foresee that they may not have the opportunity to vomit, they tend not to binge. A minority of patients also induce vomiting at other times, for example, following smaller episodes of overeating (subjective binges) or ordinary meals or snacks.

The misuse of laxatives or diuretics is somewhat less common than self-induced vomiting. It takes two forms: one is to compensate for specific episodes of binge eating, like self-induced vomiting; and the other is as a general method of weight control (like dieting), in which case it is not tied to particular episodes of overeating. The number of laxatives or diuretics taken varies considerably, sometimes far exceeding the recommended dose.

None of these methods of purging is an effective method of weight control. Self-induced vomiting results in the retrieval of only about half to two-thirds of what has been eaten, the taking of laxatives has a minimal effect on food absorption, and diuretic-taking has none. As a result, a significant proportion of each binge is absorbed.

The weight of most of these patients is in the healthy range (BMI between 20 and 25) due to the effects of the under-eating and over-eating cancelling each other out. As a result they do not experience the secondary psychosocial and physical effects associated with maintaining a very low weight seen in anorexia nervosa.

Other forms of weight-control behaviour are practised by some patients, including over-exercising, the spitting out of food, and the taking of repeated enemas or saunas. Over-exercising is the most common of these, but it is not nearly as prominent or as obviously pathological as in anorexia nervosa. A minority of patients ruminate, that is, repeatedly regurgitate and re-chew food that has been eaten. They may then either re-swallow the food or spit it out. This behaviour is not well-understood.

In the non-purging type of bulimia nervosa there is no vomiting or misuse of laxatives or diuretics, or they occur infrequently. Instead, there is sustained and marked dietary restriction outside the binges. This is both a response to the binge eating and contributor to it, in that this type of eating increases the risk of further episodes. In all other respects the two subtypes of the disorder are similar.

(c) Attitudes to shape and weight

A characteristic set of attitudes to shape and weight is the other distinctive element of the specific psychopathology of bulimia nervosa. Equivalent attitudes are found in anorexia nervosa and most cases of eating disorder NOS. These attitudes are often described as the 'core psychopathology' of eating disorders. They are characterized by an overconcern with shape and weight in which there is a fear of weight gain and fatness that is generally accompanied by a pursuit of weight loss and thinness. Underlying this psychopathology is the tendency to judge self-worth largely, or even exclusively, in terms of shape and weight. Whereas it is usual to evaluate self-worth on the basis of perceived performance in a variety of domains of life (such as interpersonal relationships, work, sport, artistic ability, etc.), people with anorexia nervosa or bulimia nervosa evaluate themselves primarily in terms of their shape and weight. These attitudes and values constitute a good example of an overvalued idea.

Most features of bulimia nervosa can be understood as being secondary to these attitudes to shape and weight. The dieting, purging, and over-exercising are obvious secondary features. In addition, there are direct behavioural expressions of these concerns. For example, many patients repeatedly weigh themselves and scrutinize their appearance in mirrors. Others avoid any knowledge of their weight while being acutely sensitive about their appearance. Some avoid others seeing their body and some even avoid seeing it themselves. This can have a major impact on social and sexual relationships.

The concerns about shape and weight, and eating, have a major effect on others in the patient's immediate environment. Meals are often times of tension and social events which involve eating may be avoided. The feeding of children may be affected[10] and their growth may be impaired[11] (see Chapter 9.3.6).

General psychopathology

General psychiatric symptoms are prominent in bulimia nervosa; more so than in anorexia nervosa. The nature of the comorbid symptoms also differs. Depressive features are particularly striking: indeed, the level of depressive symptoms in bulimia nervosa is equivalent to that seen in major depressive disorder. Anxiety symptoms are also encountered, many of which are directly related to the eating disorder; for example, there is often pronounced anxiety about eating in public. Obsessive-compulsive features are sometimes present, although they are less common than in anorexia nervosa. Similarly, social functioning is less impaired.

The depressive features of bulimia nervosa deserve special mention. In most patients the depressive features can be attributed to the presence of the eating disorder but in a subgroup there appears to be an independent coexisting, but interacting, clinical depression. Features suggestive of such coexisting clinical depressions include the following: recent intensification of depressive features (in the absence of any change in the eating disorder or the patient's circumstances); pervasive and extreme negative thinking (i.e. broader in content than concerns about eating, shape, and weight); hopelessness in general (i.e. seeing the future as totally bleak, seeing no future, resignation); recurrent thoughts about death and dying; suicidal thoughts; guilt over events in the far past; a decrease in involvement with others over and above any impairment that already accompanied the eating disorder (e.g. ceasing to see friends); loss of interest in activities that had been pursued despite the eating disorder (e.g. ceasing to listen to music; ceasing to read newspapers or follow the news); and a decrease in drive and initiative.

These coexisting clinical depressions often go undetected since they are viewed as characteristic of bulimia nervosa. This is unfortunate for two reasons: first, they interfere with the treatment of the eating disorder; and second, they are readily treated with antidepressant drugs (unlike the secondary depressive features).

A minority of patients with bulimia nervosa have 'impulse-control' problems, such as the overconsumption of alcohol or drugs, or repeated self-harm (e.g. cutting). Some of these patients also meet diagnostic criteria for borderline personality disorder (see Chapter 4.12.2). The prevalence of such features varies according to treatment setting: they are unusual in community samples, whereas they are more frequent among patients seen in specialist treatment centres.

Much more common than frank personality disorders are two traits which are also seen in anorexia nervosa. The first is low self-esteem. This generally antedates the eating disorder, although it is often exaggerated by it. Many patients with bulimia nervosa describe longstanding doubts about their worth and ability, irrespective of their accomplishments. The second is perfectionism, that is, imposing on oneself inordinately high personal standards in a range of domains (for example, work, sport, personal conduct). Since many of these standards are unachievable, it is common for these patients to give long histories of viewing themselves as perpetually failing.

Physical features

There are few physical abnormalities in bulimia nervosa. Body weight is unremarkable in the majority of patients and thus the physical effects of starvation are rarely seen. Nevertheless, menstrual abnormalities or amenorrhoea are present in about a quarter of patients. These are likely to be secondary to the disturbed eating since they generally respond to the successful correction of the eating disorder. On laboratory testing endocrine abnormalities are sometimes encountered and these tend to be mild versions of those found in anorexia nervosa. Fertility appears not to be affected.

Frequent purging, and especially the combination of vomiting and laxative misuse, results in fluid and electrolyte abnormalities in some patients. These abnormalities are varied in nature but most often consist of some combination of hypokalemia, hyponatremia, and hypochloremia. The patients appear to accommodate to these abnormalities since medically serious complications (for example, cardiac arrhythmias) are much less common than might otherwise be expected given the laboratory findings. Some patients experience intermittent oedema particularly if there is a sudden decrease in the extent of their purging.

Localized physical abnormalities include erosion of the dental enamel (especially from the lingual surfaces of the front teeth) among those who have vomited for many years; traumatic calluses on the knuckles of the hand of those who use their fingers to induce the gag reflex (Russell's sign); and enlargement of the salivary glands, especially the parotids, probably as a result of chronic inflammation. A small proportion of patients have raised serum amylase levels usually due to an increase in the salivary isoenzyme.

Relationship to other disorders

Anorexia nervosa and eating disorder NOS

Bulimia nervosa has many features in common with anorexia nervosa and eating disorder NOS, particularly the characteristic attitudes to shape and weight and the behaviour that arises directly as a result.[12] In most cases, bulimia nervosa is preceded either by frank anorexia nervosa (in about a quarter of cases) or an anorexia nervosa-like form of eating disorder NOS. While movement from bulimia nervosa to anorexia nervosa is unusual, progression on to some form of eating disorder NOS is common. Whether it is appropriate to view such patients as having recovered from one psychiatric disorder and developed another is a moot point: rather, it would seem more appropriate to view them as having a single evolving eating disorder.

There is some evidence of co-aggregation between bulimia nervosa, anorexia nervosa, and eating disorder NOS with there being increased rates of all three diagnoses among the relatives of probands with either condition.[13]

Obesity

Few patients with bulimia nervosa are overweight or have obesity. On the other hand there is evidence of raised rates of parental and premorbid obesity.[14] Obesity is an unusual sequel of the disorder although this may be because those at most risk of obesity are less likely to recover and so continue to suppress their weight.

Other psychiatric disorders

As noted above, depressive features are common in bulimia nervosa and they may antedate the eating disorder. The same is true of anxiety and anxiety disorders. Most family studies have found a raised rate of affective disorder among these patients' relatives whereas little is known about the familial relationship between bulimia nervosa and the anxiety disorders.[15] There is a raised rate of alcohol and drug abuse among patients with bulimia nervosa and a raised rate among these patients' relatives.[15] Substance abuse rarely antedates the eating disorder but this is to be expected given the age of onset of substance abuse disorders.

It is hazardous making personality disorder diagnoses among those with eating disorders. This is because eating disorders have their onset in adolescence and they directly affect many of the characteristics upon which personality is judged. Thus there is a risk of overestimating the presence of personality disturbance. Nevertheless, some patients with bulimia nervosa do seem to have a coexisting personality disorder, the most common form being borderline personality disorder. Little is known about the rate of personality disturbance among these patients' relatives although there is evidence of familial co-aggregation of anorexia nervosa and obsessional and perfectionist traits.[15]

Diabetes mellitus

It was thought that eating disorders were over-represented among those with Type I diabetes mellitus. This is now not clear. Controlled studies in which eating disorder features have been assessed by interview rather than self-report questionnaire (the preferred method and one which minimizes the risk of false positive diagnoses) have found little evidence of an elevated rate of anorexia nervosa although the rate of bulimia nervosa may be increased.[16]

Distribution

The fact that it took Russell more than 6 years (1972–1978) to collect 30 cases of bulimia nervosa suggested that the disorder was not common. It is therefore remarkable that within a few years of the publication of Russell's paper it was evident that bulimia nervosa was an important source of psychiatric morbidity among young women.

In the early 1980s large numbers of previously undetected cases were identified using the media.[17,18] These cases were remarkably similar to Russell's, except that almost all were female and a small proportion had a history of anorexia nervosa. Most had kept their eating disorder secret for many years, and because of shame and hopelessness few had sought help. Many thought that they were the only person with this type of eating disorder. Simultaneously however, and doubtless partly as a result of the media attention,

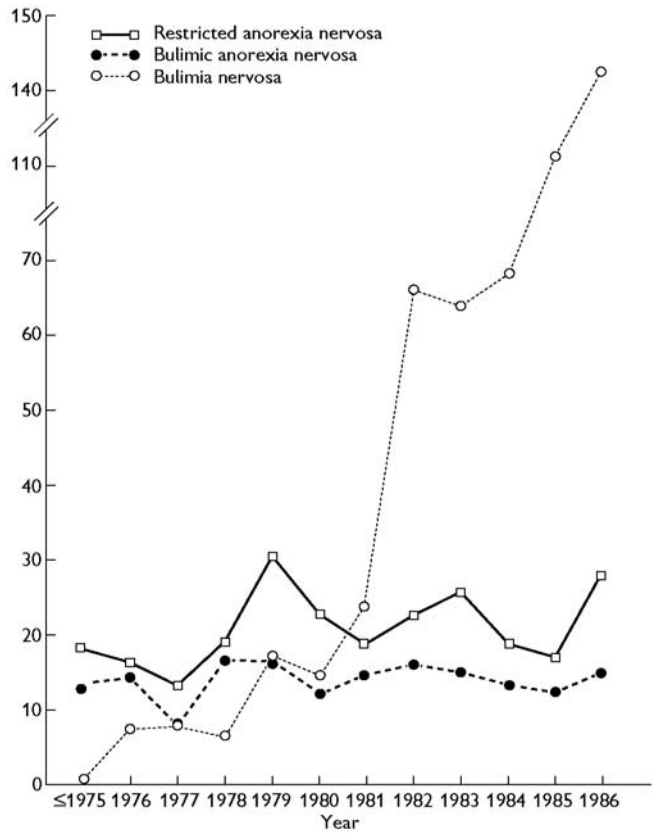


Fig. 4.10.2.3 Rates of referral to a major eating disorder centre in Toronto (1975–1986) (Reproduced from Garner, D.M. and Fairburn, C.G Relationship between anorexia nervosa and bulimia nervosa: diagnostic implications. In *Diagnostic issues in anorexia nervosa and bulimia nervosa* (eds. D.M. Garner and P.E. Garfinkel), copyright 1998, Brunner/Mazel, New York).

there was also a sharp increase in the number of people requesting treatment for bulimia nervosa (see Fig. 4.10.2.3).

The marked increase in the number of patients with bulimia nervosa stimulated interest in the prevalence of the disorder. By 1989 over 50 prevalence studies had been conducted, many of which yielded unrealistically high prevalence figures as a result of using weak assessment and sampling procedures. Gradually methods improved with the result that estimates of the prevalence of bulimia nervosa decreased to more modest and consistent levels with the point prevalence among young adult women (aged 16 to 35 years) being in the region of 1 per cent,[19,20] a similar figure being obtained for lifetime prevalence.[20] The prevalence of bulimia nervosa among men is not known. Among patient samples, male cases are unusual. Bulimia nervosa is thought to be uncommon in non-Western societies although few prevalence studies have been conducted and most have had significant methodological shortcomings.[21]

There have been few estimates of the incidence of bulimia nervosa, and since these have been based upon clinic rather than community samples, they are likely to underestimate the true figures. Even today, many people with bulimia nervosa do not seek help. The lack of reliable community-based incidence figures also makes it impossible to know whether the disorder has become more common since the 1970s or whether the upsurge in cases in the early 1980s was as a result of undetected cases being more likely to seek treatment. Data from the assessment of women in different birth cohorts suggest that the disorder has become much more common[22] although other explanations for the apparent increase cannot be ruled out.

Aetiology

Development of the disorder

As noted in Chapter 4.10.1, anorexia nervosa generally starts in mid-adolescence with a period of voluntary dietary restriction which proceeds to get out of control. As a result body weight falls and a state of starvation develops. Shape and weight concerns may pre-date the onset of the dieting or develop as weight is lost.

Bulimia nervosa starts in a similar way although the age of onset is typically some years later and shape and weight concerns usually antedate the dieting. The dietary restriction resembles that seen in anorexia nervosa and it leads to weight loss sufficient to result in anorexia nervosa in about a quarter of cases. (As a result of referral bias, this proportion is higher in cases seen in specialist centres.) In the remaining cases there is also weight loss but it is less extreme. After a variable length of time (generally within 3 years) dietary control breaks down with the patient's dieting becoming punctuated by episodes of overeating. At first, the episodes of overeating may be modest in size and intermittent, but gradually they become larger and more frequent. As a result, the lost weight is regained and body weight returns to near its original level. By this point the disorder tends to be self-perpetuating. At some stage in this sequence of events, self-induced vomiting and laxative misuse may be adopted to compensate for the overeating. In practice, however, both forms of behaviour have the opposite effect since belief in their effectiveness encourages a relaxation of control over eating. In those who vomit this phenomenon is exaggerated by the discovery that the process is easier after eating large amounts of food.

Predisposing factors and processes

There are many risk factors for the development of bulimia nervosa[14] and these overlap with those for anorexia nervosa[23] (see Table 4.10.2.2).

The risk factors may be usefully divided into a number of categories.

- Demographic factors—these are being female, adolescent and living in a Western society.
- Exposure to an immediate social environment that encourages dieting—this includes being brought up in a family in which there is intense interest in shape, weight, and eating as a result of one or more members either having some degree of eating disorder or having a medical condition that affects eating or weight (such as diabetes mellitus). Extreme occupational or recreational pressures to diet also appear to be associated with increased risk (for example, ballet dancing), although there may also be an element of self-selection. Another important influence is parental and childhood obesity, the rates of which are substa ntially increased in bulimia nervosa. Both are likely to sensitize individuals to their appearance and weight, and thereby make them prone to diet. There is also some evidence that puberty occurs comparatively early which may also magnify concerns about shape.

Table 4.10.2.2 Principal risk factors for anorexia nervosa and bulimia nervosa

General factors
Female
Adolescence and early adulthood
Living in a Western society
Individual-specific factors
Family history
Eating disorder of any type
Depression
Substance abuse, especially alcoholism (bulimia nervosa)
Obesity (bulimia nervosa)
Premorbid experiences
Obstetric complications
Adverse parenting (especially low contact, high expectations, parental discord)
Sexual abuse
Family dieting
Critical comments about eating, shape, or weight from family and others
Occupational and recreational pressure to be slim
Premorbid characteristics
Low self-esteem
Perfectionism (anorexia nervosa and to a lesser extent bulimia nervosa)
Neuroticism
Anxiety and anxiety disorders
Obesity (bulimia nervosa)
Early menarche (bulimia nervosa)
Type I diabetes (bulimia nervosa)

- Exposure to factors that increase the risk of psychiatric disturbance in general and depression in particular—these include a family history of psychiatric disorder, especially depression, and a range of adverse childhood experiences including parenting deficits and sexual and physical abuse. It was thought that sexual abuse was especially common among those who develop bulimia nervosa, but the balance of evidence suggests that the rate is no higher than that among those who develop other psychiatric disorders.
- Perfectionism and low self-esteem—both traits are common antecedents of anorexia nervosa and bulimia nervosa. Typically they interact resulting in feelings of incompetence and ineffectiveness.
- Family history of substance abuse—there is a raised rate of substance abuse in the families of patients with bulimia nervosa. It is not clear how this increases the risk of bulimia nervosa. Clinical observations suggest that some of those who develop bulimia nervosa learn to modulate their mood by engaging in self-harm (e.g. by cutting themselves) or by consuming large quantities of food, alcohol, or psychoactive drugs.

An important question is how those with anorexia nervosa are able to maintain strict control over their eating whereas this is not true of those with bulimia nervosa. The explanation is unclear but several processes may be of relevance. First, perfectionism is even more pronounced in anorexia nervosa and it may enhance self-control. Second, the vulnerability to obesity found in bulimia nervosa may somehow undermine dietary restraint. Third, the mood lability of bulimia nervosa may also disrupt restraint.

Contribution of genetic factors

The fact that eating disorders run in families suggests that there may be a genetic contribution. In the absence of adoption studies, twin designs have been used to establish its extent.[24] Clinic samples show concordance for anorexia nervosa of around 55 per cent in monozygotic twins and 5 per cent in dizygotic twins, with the corresponding figures for bulimia nervosa being 35 and 30 per cent, respectively.[7,25] These findings suggest a significant heritability to anorexia nervosa but not to bulimia nervosa. Despite this, there is uncertainty as to the size of the genetic contribution to both disorders with there being differing point estimates and wide confidence intervals. The same applies to the contributions of individual-specific and shared (common) environmental factors. A number of issues affect the interpretation of the data. For example, there has been insufficient power to detect shared environmental effects, and established diagnostic criteria have been broadened considerably to increase the number of 'affected' twins available for analysis.

Given the clear and possibly substantial genetic contribution to both disorders, molecular genetic studies have been conducted to identify the underlying loci and genes. Genetic association studies have focussed in particular on polymorphisms in 5-HT (serotonin)-related genes because this neurotransmitter system is important in regulation of eating and mood, but a range of other polymorphisms have also been investigated. Despite this, no associations with eating disorders have been clearly replicated or confirmed in a family study or by meta-analysis. There has been one multi-centre genome-wide linkage study. It found linkage peaks for anorexia nervosa and bulimia nervosa on chromosomes 1, 4, 10, and 14. A further analysis, which covaried for related behavioural traits, identified a different locus on chromosome 1, as well as loci on chromosomes 2 and 13. All these findings await replication.

Neurobiological findings

There has been extensive research into the neurobiology of eating disorders. This has focussed on neuropeptide and monoamine (especially 5-HT) systems thought to be central to the physiology of eating and weight regulation.[26] Of the various central and peripheral abnormalities reported, many are likely to be secondary to the disturbed eating and associated weight loss. However, some aspects of 5-HT function and its receptors remain abnormal after recovery, leading to speculation that there is a trait monoamine abnormality which may predispose to the development of eating disorders or to associated characteristics such as perfectionism. Furthermore, normal dieting in healthy women has been shown to alter central 5-HT function, providing a potential mechanism by which eating disorders might be precipitated in women vulnerable for other reasons.

Brain functional imaging studies have identified altered activity in the frontal, cingulate, temporal, and parietal cortical regions in both anorexia nervosa and bulimia nervosa, and there is some evidence that these alterations persist after recovery.[27] Whether they are a consequence of the eating disorder or have somehow contributed to it is not known.

Maintaining factors and processes

Once established, bulimia nervosa tends to run a chronic course although it commonly evolves into eating disorder NOS. There are a number of processes which account for its self-perpetuating character. These are discussed in Chapter 6.3.2.2 (CBT for eating disorders). They include the ongoing influence of the extreme concerns about shape and weight; the form of these patients' dieting, which encourages binge eating; the mood-modulating effect of binge eating; and the fact that the loss of control overeating perpetuates fears of weight gain.

Assessment

The identification of patients with bulimia nervosa is not difficult so long as the diagnosis is considered. This is important because the shame that characterizes the disorder leads many people to delay seeking help—the average delay between onset and presentation is about 5 years—and when they do present for treatment, some do so indirectly. Thus they may complain of depression, substance abuse, menstrual disturbance, or gastrointestinal symptoms, rather than the eating disorder itself. The best policy is for psychiatrists to always keep in mind the possibility of bulimia nervosa when assessing female patients aged between 16 and 35 years. Negative responses to the following questions should suffice to exclude the disorder:

- 'Do you have any problems with your eating?'
- 'Do you have any problems controlling your eating; that is, problems with binge eating?'
- 'Do you ever make yourself sick or take laxatives to control your weight or shape?'

Patients who present directly complain of having lost control over eating and their assessment is generally straightforward. It should always include an assessment of the extent to which the disorder is interfering with everyday functioning and an evaluation of general psychiatric features and especially those of depression.

The best established measure of eating disorder features is the Eating Disorder Examination.[28] This interview is widely regarded as the 'gold standard' measure of eating disorders, but it is possibly too exhaustive to use on a routine clinical basis. Various self-report questionnaires are available but they provide a more basic level of assessment and they cannot be used to make a clinical diagnosis. The leading self-report measures are the Eating Disorder Inventory[29] and the self-report version of the Eating Disorder Examination.[30]

No assessment is complete without weighing the patient and checking their height. Weighing needs to be done with considerable sensitivity because of these patients' concerns about their weight. A physical examination is not essential unless the patient is underweight (or there are other medical indications), nor are laboratory tests required except in those cases in which there is reason to suspect that there might be fluid or electrolyte disturbance.

Treatment

Given that bulimia nervosa has only recently been described, there has been a remarkable amount of research on its treatment. Over 70 randomized controlled trials have been completed. An authoritative meta-analysis has been conducted by the United Kingdom National Institute for Health and Clinical Excellence or 'NICE'.[31] The majority of the trials have focussed on adults with bulimia nervosa, the treatment of adolescents having received little attention, and almost all these studies have been 'efficacy' rather than 'effectiveness' studies. However, there are reasons to think that their findings are of direct relevance to routine patient care not least because the patients studied have been similar to those seen in clinical practice. Nevertheless, there is a definite need for effectiveness studies particularly now that the main treatment options are clear.

Studies of pharmacological treatment

A variety of drugs have been tested as possible treatments for bulimia nervosa including antidepressants, appetite suppressants, anticonvulsants, and lithium. Only antidepressants have shown promise.

(a) Antidepressant medication

All the major classes of antidepressant drug have been evaluated, including tricyclic antidepressants, monoamine oxidase inhibitors, selective serotonin uptake inhibitors, and atypical antidepressants. The findings have been relatively consistent and may be summarized as follows (adapted from Wilson and Fairburn[32]):

- Antidepressant drugs are more effective than placebo at reducing the frequency of binge eating and purging. On average, among treatment completers there is about a 50 per cent reduction in the frequency of binge eating and a cessation rate of about 20 per cent. The therapeutic effect is more rapid than that seen in depression. There is generally little change in the placebo group. The dropout rate varies but averages about 30 per cent.
- The longer-term effects of antidepressant drugs remain largely untested. Almost all the studies to date have been of their short-term use (16 weeks or less). The findings of the few longer-term studies suggest that outcome is poor and compliance low.[33]
- Few studies have evaluated the effects of antidepressant drugs on features other than binge eating and purging. Mood improves as the frequency of binge eating declines but this effect is common to all treatments for bulimia nervosa. Antidepressant drugs do not appear to modify these patients' extreme dieting which may account for the poor maintenance of change.
- Different antidepressant drugs seem to be equally effective, although there have been no direct comparisons of different drugs.
- With one exception, there have been no systematic dose-response studies. The exception showed that fluoxetine at a dose of 60 mg/day, but not 20 mg/day, was more effective than placebo.[34]
- Patients who fail to respond to one antidepressant drug may respond to another. There have been no drug augmentation studies.
- No consistent predictors of response have been identified. Pretreatment levels of depression appear not to be related to outcome.

- The mechanism(s) whereby antidepressant drugs exert their 'antibulimic' effects is not known. The apparent comparability of different classes of drug implicates a common mechanism but this is unlikely to be their antidepressant action since the response is too rapid and the level of depression does not predict outcome.

Studies of psychological treatment

(a) Cognitive behaviour therapy (CBT)

The most intensively studied psychological treatment is a specific form of CBT.[35,36] This was the first promising treatment described and it remains the leading treatment for the disorder. The treatment and its rationale are described in Chapter 6.3.2.2.

CBT is conducted on an outpatient basis and involves 15 to 20 sessions over about 5 months. It is suitable for all patients bar the small minority (less than 5 per cent) who require hospitalization.

The findings of the studies of CBT (over 20 controlled trials) are summarized below (adapted from Wilson and Fairburn[32]):

- The drop-out rate with CBT (about 20 per cent) is less than that seen with antidepressant drugs. The treatment is also more acceptable to these patients than treatment with medication.
- CBT has a substantial effect on the frequency of binge eating and purging. On average, among treatment completers there is about an 80 per cent reduction in the frequency of binge eating, and a cessation rate of about 60 per cent.
- The effects of CBT appear to be well-maintained. Most of the recent studies have included a 6 to 12-month follow-up period. The relapse rates are low.
- CBT affects most aspects of the psychopathology of bulimia nervosa including the binge eating, purging, dietary restraint, and the over-evaluation of shape and weight. In common with other treatments, the level of depression decreases as the frequency of binge eating declines. Social functioning and self-esteem also improve.
- CBT is more effective than delayed treatment (i.e. a waiting list control group), other psychological treatments (other than possibly interpersonal psychotherapy—see below) and antidepressant drugs at reducing the frequency of binge eating and purging. Among the other psychological treatments studied have been supportive psychotherapy, focal psychotherapy, supportive-expressive psychotherapy, interpersonal psychotherapy, hypnobehavioural treatment, stress management, nutritional counselling, behavioural versions of cognitive behaviour therapy, and exposure with response prevention.
- No consistent predictors of response to CBT have been identified. Severity of symptoms at presentation, a history of anorexia nervosa, low self-esteem, and the presence of borderline personality disorder have been associated with worse outcome in some studies but not others. However, the extent of initial response (over the first 4 weeks of treatment) is a potent and potentially valuable predictor of outcome.[37,38]
- The mechanism(s) of action of CBT have yet to be established although they appear to be mediated at least in part by a reduction in dietary restraint.[39] It also seems that the cognitive procedures are required for progress to be maintained since behavioural versions of the treatment are associated with a greater risk of relapse.[40]
- There is some evidence that the combination of CBT and antidepressant drugs may be more effective than CBT alone in reducing accompanying anxiety and depressive symptoms.
- Simpler forms of CBT may help a small subset of patients although the findings are not consistent. These include brief versions of the treatment and cognitive behavioural self-help in which patients follow a cognitive behavioural self-help programme either on their own (pure self-help) or with the guidance of a therapist (guided self-help). The programme may be delivered via a book, CD-ROM or the internet. This type of treatment is still in its infancy and it has yet to be rigorously evaluated.

(b) Interpersonal psychotherapy (IPT)

IPT is the leading alternative to CBT. This treatment was originally devised by Klerman and colleagues as a treatment for depression (see Chapter 6.3.3).[41] It is a focal psychotherapy, the main emphasis of which is to help patients identify and modify current interpersonal problems. The treatment is both non-directive and non-interpretative and, as adapted for bulimia nervosa,[42] it pays little attention to the patient's eating disorder. It is therefore very different to CBT. There have been two comparisons of CBT and IPT in the treatment of bulimia nervosa and both have found that they are about comparably effective but that IPT takes 4 to 8 months longer to achieve its effects.[40,43] The second of these studies was also designed to identify variables that might allow patients to be matched to CBT or IPT but none emerged.

Management of bulimia nervosa

Given that the leading treatment for bulimia nervosa is CBT, the ideal form of management is the provision of CBT by a therapist trained in its implementation. Unfortunately there is a shortage of the necessary expertise with the consequence that a 'stepped care' approach has been advocated on pragmatic grounds. With such an approach a simple treatment is used first and only if this proves insufficient a more complex and specialized intervention is provided. Three steps may be distinguished.

Step 1—Having established the diagnosis, the first decision is whether the patient may be managed on an outpatient basis. The great majority (over 95 per cent of referrals to non-specialist centres) may be managed this way. Exceptions are patients at significant risk of suicide and the presence of physical complications necessitating inpatient or day patient care. Severe substance abuse requires treatment in its own right, although this can sometimes be integrated with the treatment of the eating disorder. For example, it is possible to adapt CBT for bulimia nervosa so that it addresses the patient's substance abuse at the same time.

Step 2—If the patient is suitable for outpatient-based treatment, guided cognitive behavioural self-help, and/or antidepressant medication is the next step. The former involves following a cognitive behavioural self-help programme under the guidance of a 'facilitator' (a non-specialist therapist). Three cognitive behavioural self-help books are available,[44–46] two of which are direct translations of CBT for bulimia nervosa. There is evidence to support the use of all three and it is a matter of preference which is chosen. Each provides information about bulimia nervosa together with a self-help programme. The role of the facilitator is not to provide

treatment as such, as in a conventional 'therapist-led' treatment, but rather to support and encourage the patient to follow the programme. Thus, this is a 'programme-led' form of treatment, and it is this characteristic that makes it suitable for widespread dissemination. Treatment generally takes about 4 months and involves eight to ten meetings with the facilitator, each lasting up to 30 min. It is best if the first few appointments are weekly.

Guided self-help may take place in a variety of settings including primary care. Unfortunately only a small proportion of patients show substantial change and there are no reliable predictors of outcome. Patients who have obtained little benefit after 4 to 6-weeks are unlikely to do so and should be moved on to Step 3.

Antidepressant medication is an alternative to guided self-help. It is important to note that the medication is being used for its antibulimic effect not its antidepressant action. This affects the choice of drug and the dose chosen. The drug of choice is fluoxetine at a dose of 60 mg in the morning. It is usually well tolerated. As with guided self-help, if there has been little benefit after 4 to 6-weeks it is best to move on to Step 3.

A third alternative would be to combine antidepressant medication and guided self-help to see if the two augment each other. This would be a reasonable strategy although there are few data to support it.

Step 3—Patients who do not benefit from cognitive behavioural self-help or antidepressant medication should receive full CBT on a one-to-one basis (see Chapter 6.3.2.2). Ideally this should be delivered by a well-trained therapist who is used to following the protocol.[36]

Step 4—The fourth step is for those who are still symptomatic after having received well-delivered CBT. This step is pragmatic since there are few research findings of relevance. To guide the choice of treatment, the reasons for the poor response need to be carefully considered. Explanations include the presence of an undetected clinical depression; failure of CBT (which itself needs to be explained); poorly delivered CBT; poor patient compliance (which also needs to be explained); and disruption by outside events.

There are a number of different treatment options under these circumstances, the choice depending upon the outcome of the reassessment and the resources available. They include the following:

- Stop treatment and arrange to re-evaluate the patient after an interval of some months. Some patients and therapists 'burn out' after a sustained period of therapeutic work. A break can often be helpful. Deciding to stop treatment should be a joint decision and it is not appropriate with patients who are distressed or with those whose physical or psychological well-being is a cause for concern.
- Embark upon a new psychological treatment. While the obvious choice is IPT, there are no grounds for supposing that patients who fail to respond to CBT will respond to IPT. Indeed, there is evidence that this is not the case.[47] An alternative strategy would be to change the form of the CBT. The re-evaluation of the patient may have resulted in the identification of problems that might be amenable to cognitive behavioural procedures outside the realm of mainstream CBT for bulimia nervosa. Indeed, the fact that this occurs has led to the development of a new 'enhanced' form of CBT for eating disorders that is not only designed to be more potent that the earlier treatment but it also addresses common obstacles to change external to the eating disorder psychopathology (i.e. mood intolerance, clinical perfectionism, low self-esteem, and interpersonal problems).[8,48] It is also of note that this treatment is designed to be suitable for all forms of eating disorder not just bulimia nervosa.
- Arrange for day patient or inpatient treatment. In a small minority of cases outpatient treatment proves not to be sufficient, either because the disorder is resistant to outpatient-based forms of treatment or because the patient's life circumstances are interfering with progress. In such cases day patient or inpatient treatment can be useful. Generally this involves a combination of therapeutic approaches including elements of CBT. It is essential that both day patient treatment and inpatient treatment are followed by outpatient treatment designed to ensure that progress is maintained.

Course and outcome

Much remains to be learned about the course and outcome of bulimia nervosa. It is clear from epidemiological studies that many people do not present for treatment. The course of their disorder is completely unknown. Those who do present tend to do so after a considerable period of time indicating that among this subgroup the disorder has a tendency to run a protracted course. On the other hand, the findings of the treatment studies indicate that the outcome is considerably better than Russell originally suggested, although it must be stressed that even with CBT, the most effective treatment, only about half the patients make a full and lasting recovery.

There have been few studies of long-term course or outcome. A 5-year prospective study of a community sample found that at each assessment point between a half and two-thirds of the cases had an eating disorder of clinical severity, the majority being cases of eating disorder NOS.[49] A 10-year follow-up study found that about 10 per cent met diagnostic criteria for bulimia nervosa and a further 20 per cent had a form of eating disorder NOS.[50] There is no evidence that bulimia nervosa evolves into any other psychiatric disorder, and anorexia nervosa is a very unusual outcome. Body weight changes little over time and, in contrast with anorexia nervosa, the mortality rate appears not to be raised. Robust predictors of long-term course or outcome have been identified.

Acknowledgements

We are grateful to the Wellcome Trust for its support. CGF holds a Principal Research Fellowship (046386). ZC and RM are supported by a programme grant (046386).

Further information

Fairburn, C.G. (ed.). (2008). *Cognitive behavior therapy and eating disorders.* Guilford Press, New York.

International Journal of Eating Disorders.

References

1. Russell, G.F.M. (1979). Bulimia nervosa: an ominous variant of anorexia nervosa. *Psychological Medicine*, **9**, 429–48.

2. American Psychiatric Association. (1980). *DSM-III: diagnostic and statistical manual of mental disorders* (3rd edn). American Psychiatric Association, Washington, DC.
3. American Psychiatric Association. (1987). *DSM-III-R*. American Psychiatric Association, Washington, DC.
4. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
5. World Health Organization. (1992). The ICD-10 classification of mental and behavioural disorders. Author, Geneva.
6. Fairburn, C.G. and Bohn, K. (2005). Eating disorder NOS (EDNOS): an example of the troubl esome 'not otherwise specified' (NOS) category in DSM-IV. *Behaviour Research and Therapy*, **43**, 691–701.
7. Fairburn, C.G. and Harrison, P.J. (2003). Eating disorders. *Lancet*, **361**, 407–16.
8. Fairburn, C.G., Cooper, Z., and Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: a 'transdiagnostic' theory and treatment. *Behaviour Research and Therapy*, **41**, 509–28.
9. Walsh, B.T. (1993). Binge eating in bulimia nervosa. In *Binge eating: nature, assessment and treatment* (eds. C.G. Fairburn and G.T. Wilson), pp. 37–49. Guilford Press, New York.
10. Stein, A., Woolley, H., Cooper, S.D., *et al.* (1994). An observational study of mothers with eating disorders and their infants. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, **35**, 733–48.
11. Stein, A., Murray, L., Cooper, P., *et al.* (1996). Infant growth in the context of maternal eating disorders and maternal depression: a comparative study. *Psychological Medicine*, **26**, 569–74.
12. Fairburn, C.G., Cooper, Z., Bohn, K., *et al.* (2007). The severity and status of eating disorder NOS: implications for DSM-V. *Behaviour Research and Therapy*, **45**, 1705–15.
13. Strober, M., Freeman, R., Lampert, C., *et al.* (2000). Controlled family study of anorexia nervosa and bulimia nervosa: evidence of shared liability and transmission of partial syndromes. *American Journal of Psychiatry*, **157**, 393–401.
14. Fairburn, C.G., Welch, S.L., Doll, H.A., *et al.* (1997). Risk factors for bulimia nervosa—a community-based case-control study. *Archives of General Psychiatry*, **54**, 509–17.
15. Lilenfeld, L.R., Kaye, W.H., Greeno, C.G., *et al.* (1998). A controlled family study of anorexia nervosa and bulimia nervosa—psychiatric disorders in first-degree relatives and effects of proband comorbidity. *Archives of General Psychiatry*, **55**, 603–10.
16. Mannucci, E., Rotella, F., Ricca, V., *et al.* (2005). Eating disorders in patients with Type 1 diabetes: a meta-analysis. *Journal of Endocrinological Investigation*, **28**, 417–9.
17. Fairburn, C.G. and Cooper, P.J. (1982). Self-induced vomiting and bulimia nervosa: an undetected problem. *British Medical Journal*, **284**, 1153–5.
18. Fairburn, C.G. and Cooper, P.J. (1984). Binge-eating, self-induced vomiting and laxative misuse—a community study. *Psychological Medicine*, **14**, 401–10.
19. Fairburn, C.G. and Beglin, S.J. (1990). Studies of the epidemiology of bulimia nervosa. *American Journal of Psychiatry*, **147**, 401–8.
20. Hoek, H.W. and van Hoeken, D. (2006). Review of the prevalence and incidence of eating disorders. *International Journal of Eating Disorders*, **34**, 383–96.
21. Keel, P.K. and Klump, K.L. (2003). Are eating disorders culture-bound syndromes? Implications for conceptualizing their etiology. *Psychological Bulletin*, **129**, 747–69.
22. Kendler, K.S., MacLean, C., Neale, M., *et al.* (1991). The genetic epidemiology of bulimia nervosa. *American Journal of Psychiatry*, **148**, 1627–37.
23. Fairburn, C.G., Cooper, Z., Doll, H.A., *et al.* (1999). Risk factors for anorexia nervosa—three integrated case-control comparisons. *Archives of General Psychiatry*, **56**, 468–76.
24. Slof-Op't Landt, M., van Furth, E.F., Meulenbelt, I., *et al.* (2005). Eating disorders: from twin studies to candidate genes and beyond. *Twin Research and Human Genetics*, **8**, 467–82.
25. Bulik, C.M., Sullivan, P.F., Wade, T.D., *et al.* (2000). Twin studies of eating disorders: a review. *The International Journal of Eating Disorders*, **27**, 1–20.
26. Kaye, W.H., Frank, G.K., Bailer, U.F., *et al.* (2005). Serotonin alterations, in anorexia and bulimia nervosa: new insights from imaging studies. *Physiology & Behavior*, **85**, 73–81.
27. Kaye, W.H., Wagner, A., Frank, G., *et al.* (2006). Review of brain imaging in anorexia and bulimia nervosa. In *Annual review of eating disorders Part 2* (eds. S. Wonderlich, J.E. Mitchel, M. de Zwaan, and H. Steiger), pp. 113–29. Radcliffe, Oxford.
28. Fairburn, C.G. and Cooper, Z. (1993). The eating disorder examination (12th edn). In *Binge eating: nature, assessment and treatment* (eds. C.G. Fairburn and G.T. Wilson), pp. 317–60. Guilford Press, New York.
29. Garner, D.M. (1991). *Eating disorder inventory-2*. Psychological Assessment Resources, Odessa, FL.
30. Fairburn, C.G. and Beglin, S.J. (1994). Assessment of eating disorders: interview or self-report questionnaire? *The International Journal of Eating Disorders*, **16**, 363–70.
31. National Collaborating Centre for Mental Health. (2004). Eating disorders: core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders. British Psychological Society and Royal College of Psychiatrists, London.
32. Wilson, G.T. and Fairburn, C.G. (2007). Treatments for eating disorders. In *A guide to treatments that work* (3rd edn) (eds. P.E. Nathan and J.M. Gorman). Oxford University Press, New York.
33. Romano, S.J., Halmi, K.A., Sarkar, N.P., *et al.* (2002). A placebo-controlled study of fluoxetine in continued treatment of bulimia nervosa after successful acute fluoxetine treatment. *American Journal of Psychiatry*, **159**, 96–102.
34. Fluoxetine Bulimia Nervosa Collaborative Study Group. (1992). Fluoxetine in the treatment of bulimia nervosa. A multicenter, placebo-controlled, double blind trial. *Archives of General Psychiatry*, **49**, 139–47.
35. Fairburn, C. (1981). A cognitive behavioural approach to the treatment of bulimia. *Psychological Medicine*, **11**, 707–11.
36. Fairburn, C.G., Marcus, M.D., and Wilson, G.T. (1993). Cognitive-behavioral therapy for binge eating and bulimia nervosa: a comprehensive treatment manual. In *Binge eating: nature, assessment and treatment* (eds. C.G. Fairburn and G.T. Wilson), pp. 361–404. Guilford Press, New York.
37. Agras, W.S., Crow, S.J., Halmi, K.A., *et al.* (2000). Outcome predictors for the cognitive behavior treatment of bulimia nervosa: data from a multisite study. *American Journal of Psychiatry*, **157**, 1302–8.
38. Fairburn, C.G., Agras, W.S., Walsh, B.T., *et al.* (2004). Prediction of outcome in bulimia nervosa by early change in treatment. *American Journal of Psychiatry*, **161**, 2322–4.
39. Wilson, G.T., Fairburn, C.C., Agras, W.S., *et al.* (2002). Cognitive-behavioral therapy for bulimia nervosa: time course and mechanisms of change. *Journal of Consulting and Clinical Psychology*, **70**, 267–74.
40. Fairburn, C.G., Jones, R., Peveler, R.C., *et al.* (1993). Psychotherapy and bulimia nervosa: longer-term effects of interpersonal psychotherapy, behavior therapy, and cognitive-behavior therapy. *Archives of General Psychiatry*, **50**, 419–28.
41. Klerman, G.L., Weissman, M.M., Rounsaville, B.J., *et al.* (1984). *Interpersonal psychotherapy of depression*. Basic Books, New York.
42. Fairburn, C.G. (1997). Interpersonal psychotherapy for bulimia nervosa. In *Handbook of treatment for eating disorders* (eds. D.M. Garner and P.E. Garfinkel), pp. 278–94. Guilford Press, New York.

43. Agras, W.S., Walsh, B.T., Fairburn, C.G., *et al.* (2000). A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa. *Archives of General Psychiatry*, **57**, 459–66.
44. Fairburn, C.G. (1995). *Overcoming binge eating*. Guilford Press, New York.
45. Cooper, P.J. (1995). Bulimia nervosa and binge eating: a guide to recovery. Robinson, London.
46. Schmidt, U.H. and Treasure, J.L. (1993). *Getting better bit(e) by bit(e)*. Erlbaum, London.
47. Mitchell, J.E., Halmi, K., Wilson, G.T., *et al.* (2002). A randomized secondary treatment study of women with bulimia nervosa who fail to respond to CBT. *The International Journal of Eating Disorders*, **32**, 271–81.
48. Fairburn, C.G. (ed.) (2008). *Cognitive behavior therapy and eating disorders*. Guilford Press, New York.
49. Fairburn, C.G., Cooper, Z., Doll, H.A., *et al.* (2000). The natural course of bulimia nervosa and binge eating disorder in young women. *Archives of General Psychiatry*, **57**, 659–65.
50. Keel, P.K., Mitchell, J.E., Miller, K.B., *et al.* (1999). Long-term outcome of bulimia nervosa. *Archives of General Psychiatry*, **56**, 63–9.

4.11

Sexuality, gender identity, and their disorders

Contents

4.11.1 Normal sexual function

Roy J. Levin

Introduction

Normal sexual function means different things to different people. It is studied by a variety of disciplines: biology, physiology, psychology, medicine (in the domains of endocrinology, gynaecology, neurology, psychiatry, urology, and venereology), sociology, ethology, culture, philosophy, psychoanalysis, and history. There is often little liaison or cross-fertilization between these disciplines and each has its own literature and terminology. Some are regarded as 'hard science', suggesting hypotheses that can be supported or rejected by experiment, observation, or measurement (evidence-based). Others are looked on as 'soft science', where individual and anecdotal evidence are the norm and are encouraged.

As space is limited, this chapter will characterize 'normal sexual activity' in the Western world mainly from biological, physiological, and psychological aspects but will occasionally utilize other disciplines when they yield insights not available from the 'harder sciences'.

Biological determinants of normal sexual function

Humans are the highest evolved primates. A number of our anatomical/biological features unrelated to reproduction have been described as strongly enhancing our sexual behaviour when compared with other primates,[1] although recent studies have shown that the bonobos (pygmy chimpanzees) also use sex for reasons unconnected with reproduction.[2]

In brief, these features are as follows:

1 The relative hairlessness of our bodies allows well-defined visual displays (see point 6 below) and enhanced tactile skin sensitivity.

2 The clitoris, which is an organ whose sole function is for inducing female sexual arousal/pleasure.

3 Orgasms, in both male and female, provide intense euphoric rewards for undertaking sexual arousal to completion. The female is able to have multiple serial orgasms.

4 The largest penis among primates, whether flaccid or erect, the latter acting as a good sexual stimulator of the female genitalia.

5 Concealed (cryptic) ovulation which could influence males to undertake coitus more frequently to create pregnancy and prevent cuckolding.

6 Well-defined visual sexual displays in the female that are not linked to season or fertility (i.e. breasts, pubic hair, buttocks, and lips). The everted mucous membranes of the lips serve both as a surface display (red, moist, and shiny), and for haptic stimulation during kissing and sucking.

7 Ability of the female to undertake sexual arousal and coitus independent of season, hormonal status, or ovulation. Human females (unlike other primates) can and often do willingly partake of sexual activity and coitus when they are menstruating, pregnant, or menopausal.

8 Development of large mammary glands during puberty which act as visual sexual signals in most cultures.

These biological determinants are augmented by socio-cultural factors:

1 Language, art, and music for erotic stimulation.

2 Facial adornment with make-up to heighten appearance and sex displays (viz., lipstick).

3 Clothing, especially of the female, such as brassières to redefine the shape of breasts, corsets to redefine the shape of the body, and high-heeled shoes to elongate the legs and thrust out the

buttocks. Young adult males use tight trousers to create a genital 'bulge' and to emphasize firm rounded buttocks, the latter being a highly sexually attractive feature to young women.

4 Perfumes and scents to enhance body aroma.

The last three features (2, 3, and 4) use artificial means to enhance normal sexual signals.[1] These biological and socio-cultural factors give human sexual activity an increased appetitiveness and make it more rewarding.

Sexuality as a social construct and the concept of sexual scripting

Laqueur[3] suggests that while the sexual biology remained unchanged, its expression has been influenced over the centuries by culture, social class, ethnic group, and religion. This concept, that human sexuality is a social construct, has been strongly argued by Foucault[4] and promoted by other social constructionist authors.[5] Gagnon and Simon[6] introduced the concept of 'sexual scripting'. Scripts organize and determine the circumstances under which sexual activity occurs, they are involved in 'learning the meaning of internal states, organizing the sequences of specific sexual acts, decoding novel situations, setting limits on sexual responses and linking meanings from non-sexual aspects of life to specifically sexual experience'. Money[7] employed a similar construction in his development of 'love maps' for the individual. While patterns of behaviour are influenced by society and social forces, there is a dearth of evidence to show that sexual identity, orientation, or sexual mechanisms are also influenced.

Modelling normal sexual function—(i) the sex survey

One obvious way of describing normal sexual function is to ask people what they do. Two classic sex surveys were conducted by Kinsey and his coworkers who reported the results of interviews with 12 000 males in 1948[8] and 8000 females in 1953.[9] Their technique of sampling was to interview everyone in specific cooperating groups (clubs, hospital staff, universities, police force, school teachers, etc.). This gave samples of convenience but not a valid statistical sampling of the population. Despite their age and faulty sampling, however, there are still useful data in these surveys. In the sexual climate of the 1950s many of the findings were regarded as highly controversial. Clement[10] has reviewed the subsequent studies of human heterosexual behaviour up to 1990.

Surveys give a selective picture of sexual function. Results depend on the formulation of the questions, they rely on self-reports, and they represent only those prepared to describe their sexual behaviour. It is known, for example, that females tend to under-report their premarital sexual experiences[11] while males tend to over-report their lifetime partners.[11] Berk *et al.*[12] studied the recall by 217 university students of their sexual activity over a 2-week period assessed by questionnaires answered 2 weeks after the recording period, and by daily diaries kept over the same 2 weeks. Subjects reported more sexual activity in the questionnaires than in their diaries. Women reported giving and having more oral sex than the men. Clearly, data from questionnaire surveys should be treated cautiously.

A survey tells only what is frequent and not necessarily what is normal, but the most frequent practices often become identified with normal sexual behaviour. Surveys also vary in the range of behaviours that are asked about, for example coitus without condoms is important in the age of AIDS. Surveys have one great disadvantage, the facts that they produce are often 'perishable'; many aspects of the sex surveys of the pre-pill era, or more recently the pre-AIDS era, are now of use only in a historical or comparative basis.

Two recent well-organized surveys based on samples of the whole population have been undertaken, one in the United States and the other in the United Kingdom. Interestingly, in both surveys, questions about masturbation were disliked by the respondents. In the American survey these questions were asked in a separate self-administered questionnaire, while in the British survey they were abandoned.

The American survey[13] was conducted face to face with 3159 selected individuals who spoke English in representative households by 220 trained interviewers (mainly women). Nearly 80 per cent of the individuals chosen agreed to be interviewed. Men thought about sex often, more than 50 per cent having erotic thoughts several times a day, while females thought about sex from a few times a week to a few times a month. The frequency of partnered sex had little to do with race, religion, or education. Only three factors mattered: age, whether married or cohabiting, and how long the couple had been together. Fourteen per cent of males reported having no sex in the previous year, 16 per cent had sex a few times in the year, 40 per cent a few times a month, 26 per cent two to three times a week, and 8 per cent four times a week. The percentages were similar for women. The youngest and the oldest people had the least sex with a partner; those in their 20s had the most. Of the women aged 18 to 59 years, approximately one in three said they were uninterested in sex, and one woman in five said sex gave her no pleasure. Unlike frequency, reported sexual practices do depend on race and social class. Most practices other than vaginal coitus were not very attractive to the vast majority. In women aged 18 to 44 years of age, 80 per cent rated vaginal coitus as 'very appealing' and an additional 18 per cent rated it as 'somewhat appealing'. Among men 85 per cent regarded vaginal coitus as 'very appealing'. The most appealing activity second after coitus was watching the partner undress, and this was appealing to more men (50 per cent) than women (30 per cent). This reflects the greater voyeuristic nature of men and their willingness to pay to look at women undressing or undressed.

In regard to oral sex, both men and women liked receiving more than giving. This practice varied markedly with race and education, with higher reported rates among better educated white people than among less educated and black people. Some 68 per cent of all women had given oral sex to their partner and 19 per cent experienced active oral sex the last time they had intercourse. Seventy-three per cent of all women had received oral sex from the partner, and 20 per cent had received it the last time they had had intercourse. Corresponding experiences were reported by men.

This survey, unlike many earlier ones, asked about anal sex. Of females aged 18 to 44, 87 per cent thought it not at all appealing, and only 1 to 4 per cent thought it very or somewhat appealing. In males of the same age 73 per cent thought it not at all appealing and rather more than women thought it very or somewhat appealing. Similar reports were obtained from women and men aged 44 to 59.

Regarding masturbation, older people (over 54 years old) had lower rates than at any other age, indicating that they do not use masturbation to compensate for an overall decrease in sexual activity with their partners.

In the United Kingdom survey,[14] 18 876 people were interviewed by 488 interviewers (of whom 421 were women). The sampling used one person per address and the acceptance rate was 71.5 per cent. Questions were asked about the frequency of vaginal coitus, oral sex, and anal sex, but not masturbation. The median number of occasions of sex with a man or woman was five times a month for females aged 20 to 29 and males aged 25 to 34, but declined to a median of two per month for males aged 55 to 59. More than 50 per cent of the females in the 55 to 59 age group reported no sex in the last month, but in this age group females are more likely than men to have no regular partner because they are widowed, separated, or divorced.

Vaginal coitus was reported by nearly all females and males by the age of 25. Fifty-six per cent of males and 57 per cent of females reported vaginal coitus in the previous week, and non-penetrative sex was practiced by 75 per cent men and 82 per cent of women. Twenty-five per cent of males had genital stimulation in the previous 7 days. Cunnilingus and fellatio were common but less practiced than vaginal coitus. Of men and women aged 18 to 44, 60 per cent had oral sex in the previous year but in the 45- to 59-year-old group this fell to 30 per cent for women and 42 per cent for men. This and other sex surveys suggest that the practice of oral sex has increased since the 1950s and 1960s. Anal coitus was infrequent; approximately 14 per cent of the males and 13 per cent of the females had ever undertaken it, and only 7 per cent of males or females had practiced it in the previous year.

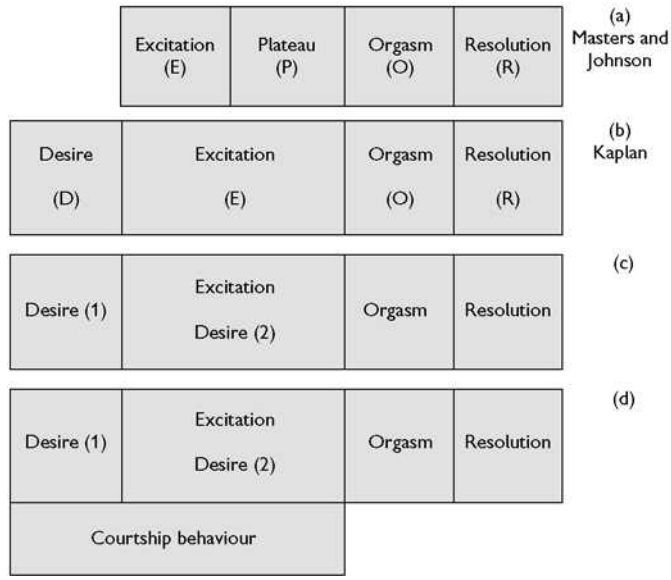


Fig. 4.11.1.1 The development of the human sexual response model from (a) the original **EPOR** model of Masters and Johnson[15] through (b) the desire, excitation, orgasmic, and resolution (**DEOR**) model of Kaplan[17] to (c) the proposed modification with desire phase 1 (before initiation of the excitation phase and desire phase 2 during excitation phase) and finally (d) with added courtship behaviour.

Modelling normal sexual function—(ii) the sexual response cycle

A direct way of investigating normal sexual function is to observe and measure the body changes that take place when men and women become sexually aroused. From these data, models have been constructed of the normal sequence of changes during sexual arousal, coitus, and orgasm. The first models described a simple sequence of increasing arousal and excitement culminating in rapid discharge by orgasm, displayed graphically as an ascent, peak, and then descent. As the investigations became more sophisticated, understanding of the body responses grew and the models became more detailed and complex.[5,15,16]

The EPOR model—a sexual response cycle model

A most successful human sexual response model was that formulated by Masters and Johnson.[15] In the laboratory, they observed the changes that took place in the male and female body and especially the genitals during sexual arousal to orgasm either by masturbation or by natural or artificial coitus with a plastic penis that allowed internal filming of the female genitalia. After studying approximately 7500 female and 2500 male arousals to orgasm in some 382 female and 312 male volunteers over 11 years, they proposed a four-phase, sequential, and incremental model of the human sexual response cycle (Fig. 4.11.1.1). The phases were described as the excitation (**E**) phase (stimuli from somatogenic or psychogenic sources raise sexual tensions), the plateau (**P**) phase (sexual tensions intensified), the orgasmic (**O**) phase (involuntary pleasurable climax), and finally the resolution (**R**) phase (dissipation of sexual tensions). The great success of this **EPOR** model was its wide compass; it could characterize the sexual responses of women and men, both heterosexual and homosexual, ranging from simple petting to vaginal or anal coitus with or without orgasm. However, it had several weaknesses.

Modifying the EPOR model into the DEOR model

The first weakness of the EPOR model is that it was derived from the study of a highly selected group of American men and women volunteers who could arouse themselves to orgasm in a laboratory, on demand, and allow themselves to be watched/filmed or measured for scientific and altruistic (or perhaps exhibitionistic) purposes. The second weakness was the lack of interobserver agreement about the changes observed and of confirmation of their sequential reliability. Robinson[16] examined the **E** phase and **P** phase, and concluded convincingly that the **P** phase was simply the final stage of the **E** phase. Helen Kaplan,[17] a New York sex therapist, proposed that before the **E** phase there should be a 'desire phase' (**D** phase). This proposal came from her work with women who professed to have no desire to be sexually aroused, even by their usual partners. She suggested that the desire must occur before sexual arousal can begin. Kaplan's subjects were attending a clinic and remarkably no studies were ever conducted with a control normal population (either women or men) to investigate whether this 'self-evident' fact was true. Despite this, the **EPOR** model gradually became replaced by the desire, excitation, orgasmic, and resolution phase (**DEOR**) modification. While this is the currently accepted model, the centrality of the desire phase in women remains uncertain (Fig. 4.11.1.1). In a survey of non-clinic sexually experienced women in Denmark, about a third reported that they never experienced spontaneous sexual desire[18] and in an American survey women reported periods of several months when they lacked interest in sex.[13] The other problem with the desire phase is its location in the sequential **DEOR** model.

Sexual desire (D1-proceptive desire) that appears to be spontaneous (but presumably must still be activated by a trigger) should obviously be placed at the beginning of the model (Fig. 4.11.1.1b), while sexual desire (D2-receptive desire) created when the person is sexually aroused by another occurs during the E-phase (Fig. 4.11.1.1c).

It has been proposed[19] that while the DEOR model fits for females in younger couples' relationships for longer maintained ones sexual activity is undertaken often for factors such as intimacy, security, and acceptance and becomes more influenced by cognitive and emotional processes and the possible outcomes of the experience (e.g. mutual pleasure, confirming commitment and trust, enhancing emotional intimacy) rather than proceptive/receptive desires.

Courtship (mating) behaviour—activity initiating normal sexual behaviour

With the possible exception of rape, the pre-initiation of sexual activity normally starts with flirting/courtship behaviours. Sometimes this activity can precede the desire phases; sometimes it occurs during the desire phases, and sometimes in the excitation phase (Fig. 4.11.1.1d).

Evolutionary psychology attempts to explain the strategies of mating.[20] Its message is not always palatable to modern sensitivities about sexual equality. It is argued that women invest more in their offspring than men,[21] that this investment is a scarce resource that men compete for, and that men can enhance their reproductive strategy by mating frequently. Most men are first visually attracted to a possible female sexual partner. They look for youthfulness and physical attractiveness in the form of regular features (symmetry), smooth complexion, optimum stature, and good physique, and they value virginity and chastity. Partner variety is highly desired. Women, however, need to obtain high-quality mates with abundant resources and look for emotional and financial status and security. Clearly the strategies conflict giving rise to different preferences in mate choice and casual sex, and different levels of investment or commitment to relationships.[22]

Once the chosen female (or male) accepts the initiation of flirting/courtship behaviour, the subsequent stages form a stereotyped sequence which is found in many different cultures. The stages are look, approach, talk, touch, synchronize, kiss (caress), sex play, coitus. It is a sequence that the poet Ovid knew in the first century BCE. Morris[1] characterized human courtship behaviour further into 12 basic stages: eye to body, eye to eye, voice to voice, hand to hand, arm to shoulder, arm to waist, mouth to mouth, hand to head, hand to body, mouth to breast, hand to genitals, genitals to genitals. Similar hierarchies have been constructed extending the behaviour to oral-genital contacts.

Although kissing has been described as 'an inhibited rehearsal for intercourse and other sexual practices'[23] and is usually undertaken in the courtship behaviour well before genital activity occurs, it is sometimes thought of as more intimate than coitus. Prostitutes, for example, traditionally do not kiss their clients on the mouth, reserving the activity for their private sexual behaviour. Nicholson[24] has speculated that kissing may be a mechanism by which semiochemicals (similar to pheromones) are exchanged between humans to induce bonding.

Common extragenital changes during sexual arousal

In both males and females, effective sexually arousing stimuli cause a number of physiological changes.[15,25] There is increased respiration, heart rate, and blood pressure, nipple erection, often sweating, and a sex flush (maculopapular skin rash). Muscle tension increases (myotonia) and the pupils dilate. All these changes become more intense as arousal increases. Following orgasm the changes dissipate rapidly (R phase); without orgasm they dissipate more slowly.

The endocrinology of normal sexual function

(a) Peptide hormones and neuropeptides

Prolactin, a peptide hormone, is secreted by cells of the anterior pituitary in both males and females. In females it is a key lactogenic hormone involved in stimulating the manufacture of breast milk but it has no proven physiological reproductive function in males. Pathologically high plasma levels of prolactin (hyperprolactinaemia) are accompanied by disturbances of sexual/reproductive function especially erectile dysfunction in males and inhibition of ovulation in females. The exact causes are unknown. As it is secreted in higher amounts at orgasm in both sexes it has been claimed to be a major factor responsible for 'switching off' sexual arousal but this ignores the fact that women are multiorgasmic while the evidence for this function even in males is unconvincing.[26]

Two neuropeptides, oxytocin and vasopressin (Antidiuretic Hormone, ADH) are hormones and also act as neuropeptide transmitters with distribution in parts of the brain and spinal cord. They are secreted as hormones by the hypothalamus/posterior pituitary with increases at orgasm in both males and females. Oxytocin is often claimed to be responsible for facilitating smooth muscle contractions during ejaculation and uterine contractions during orgasm but the possible role of vasopressin has been ignored. In some animals, oxytocin is involved in pair bonding but the evidence for this in humans is poor.

Males

(a) Androgens

The major steroids influencing normal sexual function in males are the androgens secreted by the testes mainly as testosterone with much smaller quantities of androstenedione and dihydrotestoterone. The adrenal cortex also manufactures and secretes androgens but this amounts only to 2 per cent of the total. Of the total testosterone (260–1000 ng/dl), 95 per cent is bound to plasma proteins; only the unbound fraction (34–194 pg/ml) is an active virilizing hormone.

The development and maintenance of the masculine musculature, bone growth, genitals, and pubic and axillary hair is androgen dependent. The mechanism of the hormonal masculinization of the brain in rodents involves the aromatization of testosterone, but in humans the role of aromatization is still uncertain.

(b) Androgens and male sexual behaviour

The involvement of androgens in adult human male sexual behaviour has been reviewed many times (see Levin[27] for references). Removal of testosterone by castration usually leads to a decrease in sexual activity and drive (libido) in the majority of subjects within

12 months. There is, however, large individual variations, and some castrates retain sexual activity and interest for years.[28] Factors such as adrenal androgens, the availability of sexual partners, and attitude to the operation influence the response. In castrates and in hypogonadal males, replacement of testosterone restores sexual interest and activity.

Females

(a) Oestrogen, progesterone, and androgens

In the female, oestrogen, progesterone, and androgens are involved in differentiating and maintaining genital and breast tissues and in influencing normal sexual function. Female genital development, unlike that of the male, does not appear to need hormonal stimulation as the development of female genital structures are the foetal prototype or default programme.[29] During puberty, ovarian oestrogen and progesterone together with androgens from the adrenal cortex induce growth and functional changes in the internal and external genitalia, breast, and nipples. Oestrogens induce growth in the fallopian tubes, uterus, vagina, and breasts and lay down subcutaneous fat largely in the breast, hip, and thigh regions to create the rounded contours of the female body that are highly attractive to the male. The fat laid down is enough to supply the energy for a pregnancy and the subsequent lactation. Females have a plasma androgen concentration some 10 times less (15–70 ng/dl) than those in the male. Androgens are produced by the ovaries (25 per cent), adrenals (25 per cent) and in peripheral tissues from adrenal-secreted androgen precursors. They are responsible for the development and maintenance of the clitoris, nipples, pubic and axillary hair, and probably the labia, periurethral glans and pelvic-striated musculature. The variation of the androgen levels in the plasma during the menstrual cycle is small and it is the free androgen level (1–21 pg/ml) not the bound that is the active principle.

(b) Role of hormones in female sexuality

While over 60 studies have been undertaken to examine whether the changing hormonal levels of the menstrual cycle influence the sexual arousal of the female,[30] neither oestrogen or progesterone have been found convincingly to play a direct role in influencing the sexual activity of the human female apart from their indirect functions in the maintenance of the structures and functions of the female genitals, especially the vagina.

The role of androgens in female sexuality is not clear-cut.[31] Some propose that, as in the male, it is the major hormonal influence on the female libido. Removal of the adrenals has been shown to reduce desire and ability to reach orgasm. Excess androgens stimulates libido but in pharmacological not physiological doses. Such doses affect the structure and sensitivity of the clitoris (an androgen-sensitive tissue), so the effects on sexuality might not be only brain mediated.

(c) Sexual behaviour during the menstrual cycle

Despite numerous studies it is still uncertain whether female sexual behaviour is influenced by the hormonal changes in the menstrual cycle. Meuwissen and Over[30] surveyed 64 published studies. A significant number showed a premenstrual peak in sexual desire and activity, others a postmenstrual peak, either at menstruation or ovulation but the latter studies often used poor methodology to determine the time of ovulation.

Male genital functions during normal sexual arousal

While the **DEOR** model characterizes the general sexual arousal of humans, a more specific detailed physiological model in males is that of excitation, erection, emission, and ejaculation with orgasm. Each of these is served by separate mechanisms. Although ejaculation and orgasm usually occur temporally together, they also have separate mechanisms.

(a) Excitation

Sexual excitation can occur through any of the five senses by psychogenic or somatogenic stimuli. In special circumstances arousal can become linked to or greatly enhanced by fetishitic association with non-sexual objects such as feminine garments of underwear, rubberware, shoes, furs, etc.

The sexual excitation is manifested in the brain by activation of numerous areas (see section on brain activation during human sexual arousal). Centres in the spinal cord for erection and ejaculation are known from animal studies and it is likely that they also exist in the human cord. They are activated by neural efferent activity from the aroused areas of the brain and initiate erection and ejaculation.

(b) Erection: the conversion of the flaccid urinary penis to the rigid sexual penis

The three longitudinal erectile chambers of the penis are arranged with a side-by-side dorsal pair of corpora cavernosa above the single ventral corpus spongiosum. The corpora cavernosa are covered by the tunica albuginea, a 2-mm thick fibrous membrane which is resistant to stretch. The corpus spongiosum surrounds the length of the penile urethra and is enlarged at its base to form the urethral bulb and distally to create the glans penis. While it becomes engorged with blood during arousal it is not involved in the rigidity of the erection but protects the urethra from closure. The unaroused penis is flaccid because the pudendal arterial blood flow into the erectile tissues is limited by the high sympathetic (adrenergic-mediated) constrictive tone in the smooth muscle of the vessels of corpus cavernosum. On sexual arousal, the sympathetic tone is reduced; the neural innervation of the arteries and cavernosal chambers is activated to release vasoactive intestinal peptide (VIP), a peptidergic vasodilator neurotransmitter that directly relaxes smooth muscle and nitric oxide (NO), the nitrergic neurotransmitter. NO activates the enzyme cyclic guanylase in the smooth muscle cells of the cavernous tissue and blood-vessel endothelium to produce cyclic guanosine monophosphate (cGMP), the second messenger that creates intracellular conditions to relax the muscle. The enzyme phosphodiesterase 5 breaks down the cGMP and inhibitors of this enzyme, which can be taken by mouth, facilitate the attainment of erection even with reduced neural inputs.

The vasodilatation of the arterial supply by VIP together with the relaxation of the vessels of the cavernosal tissue allows them to fill under arterial pressure stretching the chambers until they become stiff against their covering of unyielding tunica albuginea, and the veins (emissary) that pass obliquely through the tunica become occluded greatly reducing penile vascular drainage. The flaccid urinary penis has been converted into the erect rigid sexual penis some 7 to 8 cm longer. The rigidity is essential for successful vaginal penetration and to stimulate its walls (especially the anterior) during penile thrusting. The striated muscles of the pelvic region, namely the ischiocavernosus and bulbocavernosus, are not

normally involved in creating penile erection[32] although they can be voluntarily contracted in short bursts to aid its rigidity. The engorged corpus spongiosum is less rigid than the cavernosal chambers making the glans of the penis softer and less damaging to the female labia and vagina.

(c) Internal genitals

The genital fluids of the testes, epididymis, and accessory genital glands of the male are involved in emission. These glands are the bulbo-urethral (Cowper's gland), the prostate (approximately 30 per cent of the total volume of the ejaculate), and the paired seminal vesicles (approximately 60 per cent of the volume of the ejaculate). The fluids from all these together with that of the glands of Littré that line the penile urethra, constitute the ejaculate or semen which has a characteristic odour and rapidly forms a coagulum in contact with air. Subsequently enzymes in the semen break down the coagulum and release the trapped sperm.

(d) Emission

This phase begins with the movement of the various genital fluids into the ducts initiated by the neurally induced contraction of smooth muscles in the capsules of the testes, epididymis, and seminal vesicles. The secretions spurt into the prostatic urethra, and the sphincter of the bladder neck closes to prevent reflux into the bladder. When this happens the male experiences the sensation of 'ejaculatory inevitability' and knows that he will ejaculate within a second or two and that conscious suppression of the ejaculatory reflex is now impossible. The contractions of the smooth muscle of the glandular capsules together with the contraction of the vas deferens and peristalsis in the urethra move the semen along into the penile urethra.

(e) Ejaculation

Within a second or two later the bulbocavernosus muscle of the perineal region contracts clonically, initially at about 1 per 0.8 s, squeezing the urethra and forcing out the ejaculate. As ejaculation proceeds, the interval between each striated muscle contraction gets longer and their force weaker until they gradually die out.[15,32] Their number can vary between 5 and 60. Most of the ejaculate is expressed within the first half dozen contractions. If the striated muscles are paralysed the semen is squeezed out only by the smooth muscle peristaltic contractions which produce a dribbling ejaculate with no projectile force and little pleasurable quality.

(f) Male orgasm

Orgasm, the supreme ecstatic pleasure is experienced just before the striated contractions occur and is then associated, throbbingly, with each subsequent contraction slowly decreasing in intensity and dying away as do the contractions. It is felt as an intense pleasurable throbbing/pumping in the penis and pelvic area and can last from 5 to 60 s. Most males groan with each squirting contraction. Kinsey *et al.*[8] marshalled the evidence showing that orgasm and ejaculation were separate mechanisms. Briefly, orgasm occurs without ejaculation in preadolescent males, in some adult males orgasm does not occur until a few seconds after ejaculation, a few adult males are anatomically incapable of ejaculation but have orgasms, and males who have been prostatectomized cannot have ejaculations but some can have orgasms.

The intensity of orgasm usually varies with the duration of the sexual arousal (the longer it is maintained the greater the subsequent orgasm), the erotic excitement and novelty of the arousing stimuli, and previous ejaculation, especially the interval from the last one (initial ejaculations have usually better orgasms than subsequent ones). Males have a post ejaculation refractory time (PERT) and usually cannot have an erection or another ejaculation until some time has passed. The PERT varies with age and can be anything from minutes, when young, to hours or days when older.[15] It is not known where this inhibitory mechanism resides but animal work suggests that it is in the brain rather than the spinal cord.[33] Some men claim to be able to learn to inhibit ejaculation and yet have repeated serial orgasms.[34]

It has been stated that the larger the ejaculate volume the greater is the orgasmic pleasure,[15] the studies however were flawed because they used men who increased semen volume by abstaining from ejaculation for days. This confounds the effects of semen volume with that of ejaculatory abstinence which itself enhances subsequent sexual pleasure. In fact semen volume does not appear to be the arbiter of pleasure or the trigger for ejaculation.[35] Drugs can induce a 'dry ejaculation' but the pleasure of the orgasm appears unimpaired,[33] and in young boys the pleasure of the early dry orgasm is not noticeably changed when semen becomes added to the ejaculation around puberty.[36]

An ignored feature is whether there is a typology of male orgasms. Most orgasms arise from penile stimulation but they can also be activated by per rectum digital stimulation of the prostate gland. No laboratory study of these orgasms has ever been made but anecdotal reports say that they feel deeper, more widespread and intense and last longer than those from penile stimulation.[37]

Female genital functions during normal sexual arousal

(a) External

(i) Labia

The external female genitalia consist of the outer (majora) and inner (minora) labia containing erectile tissue that surround the vaginal introitus. Normally the outer labia meet and cover the introitus, but in some women the inner labia protrude even when they are sexually unaroused. Sexual arousal creates vasocongestion especially in the labia minora which protrude through the majora adding approximately 1 to 2 cm to the length of the vagina. The labia minora become erotically sensitive to touch and friction when engorged.

(ii) Clitoris

Although the clitoris is the homologue of the penis, its precise anatomical structure is still uncertain. The most recent description by O'Connell *et al.*[38] is of a triplanar complex of erectile tissue with a midline shaft lying in the medial sagittal plane about 2 to 4 cm long and 1 to 2 cm wide, which bifurcates internally into paired curved crura 5 to 9 cm long, and externally is capped with a glans about 20 to 30 mm long with a similar diameter. Two vaginal bulbs of erectile tissue are closely applied on either side of the urethra. The shaft's erectile tissue consists of two corpora cavernosa surrounded by a fibrous sheath (tunica albuginea) and the whole is covered by a clitoral hood formed in part by the fusing of the two labia minora. The uncertainty concerns the location and extent of the female corpus spongiosum. Some describe it as wrapped around the urethra, others state that the vaginal (vestibular) bulbs on either side of the vaginal wall are spongiosus tissue and unite ventrally to the urethral meatus to form a thin strand of erectile tissue ending

in the glans. Recent studies by van Turnhout *et al.*[39] have clarified the situation. They confirmed by dissections in fresh cadavers that the bilateral vestibular bulbs terminate into the glans clitoridis.

With sexual arousal, the blood flow to the clitoris is increased probably by a mechanism involving its vipergic (VIP) and nitrinergic (NO) innervation leading to its tumescence (swelling) but, contrary to many inaccurate descriptions, without true erection (i.e. without rigidity). The enhancement of its blood flow is paralleled by an increased sensitivity to touch and friction especially of the glans.

(iii) Periurethral glans

There is a triangular area of mucous membrane that surrounds the urethral meatus, extending from just below the glans of the clitoris to the entrance of the vagina. This area has been called the 'periurethral glans'[25] and is thought to be erotically stimulated, especially during penile thrusting, as it is known that the tissue is pushed and pulled into and out of the vagina by the penile movements.[25] The periurethral glans is actually part of the corpus spongiosum if we accept the anatomical designation of van Turnhout *et al.*[39] which suggest that it is the homologue of the male glans. The extent, mobility, density of innervation, and sensitivity of this erotic site may explain why some women find it easy to reach orgasm during penile thrusting alone.

(b) Internal

(i) Vagina

No single structure can describe the vaginal shape. In sexual quiescence it is a potential space with an H-shaped cross-section and an elongated S-shaped longitudinal section culminating in a cul-de-sac, the anterior wall of which is penetrated by the cervix. The anterior and posterior walls touch but their film of basal vaginal fluid prevents adhesion. This basal fluid is a mixture of fluids from the vagina itself (basal transudate) with uterine and cervical secretions. The squamous epithelial lining of the vagina actively transports sodium ions from the vaginal fluid back into the blood. As fluid follows this ion movement osmotically the vagina is continually producing and reabsorbing its own fluid[25,40] normally lower in sodium and higher in potassium compared to plasma.

Because sexual activity is intermittent, the basal vaginal blood flow is maintained normally at a minimal level by a high adrenergic-mediated vasoconstrictive tone and vasomotion. The latter is the mechanism by which tissue viability is maintained at a basal level by not having all the capillaries open at the same time but rather each one opens and closes randomly to supply by demand its surrounding tissue needs of oxygen and metabolic substrates. On sexual arousal, the blood supply to the vaginal walls is increased by the liberation of VIP from the vipergic neural innervation. This increases the flow through the open capillaries and recruits new ones until all are open and vasomotion disappears,[41] Within seconds this creates blood-vessel engorgement and an increased plasma transudate filters out of the capillaries and percolates between the cells of the epithelium saturating its limited reabsorptive capacity. The newly formed neurogenic vaginal transudate, with its higher sodium ion concentration, creates a lubricating film on the vaginal surface which is essential for painless penile penetration and thrusting. Poor or inadequate lubrication can lead to dyspareunia (painful coitus) and subsequent sexual dysfunction. On cessation of sexual arousal or after orgasm the blood flow returns to the basal level, the fluid is reabsorbed back into the blood following the continuous absorption of sodium ions by the vaginal epithelium and vasomotion is restored.

(ii) Coitus and the vagina

The cul-de-sac of the vagina is expanded during sexual arousal and the uterus with cervix is lifted clear of its posterior wall (vaginal tenting).[15] This vaginal tenting is an important reproductive feature that delays the transport of sperm allowing the ejaculate to decoagulate and initiate motility and pre-capacitation changes that facilitate the process of sperm capacitation (see Levin[40,41] for references). In the ventral–ventral ('missionary') coital position, penile penetration and thrusting stretches and stimulates the structures of the anterior vaginal wall which include the urethra, the 'G spot' (see below), and neural structures in Halban's fascia. All these (the anterior wall erotic complex) are thought to be capable of creating erotic excitement when so stimulated, giving rise to a significant part of the sexual pleasure normally experienced by most women during coitus.

(iii) The erotic structures of the anterior vaginal wall: urethra, 'G spot', and Halban's fascia

The urethra, approximately 4 cm long, is invested with erectile tissue which becomes engorged on sexual arousal. Ultrasound imaging during coitus has shown that the thrusting penis stretches the urethra.[42] There is an area on the anterior vaginal wall a few centimeters from the introitus, at or around the junction of the urethra with the bladder, that becomes swollen and on strong pressure stimulation can induce orgasm. This urethral area was first identified by Grafenberg[43] but the observation was overlooked until its rediscovery by Perry and Whipple[44] who named it the 'G spot' in recognition of the original discoverer. Anatomically, it probably represents the 'paraurethral' or 'periurethral' glands now referred to as the 'female prostate'. In some women these produce at orgasm a small amount of fluid secretion loosely referred to as the 'female ejaculate' (see[25,44] for references). These glands are in the space between the bladder trigone and the neck of the urethra which is filled with fibroelastic mesenchymal lamina rich in vascular lacunae and contains nerve fibres and pseudocorpuscular terminals. This area, known as Halban's fascia[45] when stimulated by pressure (penile or digital) creates intense sexual pleasurable feelings.[44–46] As any pressure stimulus on the anterior vaginal wall will in fact stimulate the G spot, Halban's fascia and the urethra at the same time it makes it difficult to apportion the generation of sexual pleasure to any specific structure.

The involvement of the spinal cord in activating the female erotic structures during sexual arousal has been poorly studied and little is known of possible mechanisms.

(c) Female orgasm

As in the male, the female orgasm creates supreme ecstatic pleasure usually accompanied by throbbing striated muscle contractions, especially of the ischiocavernous and bulbospongiosus muscles, but other pelvic-striated muscles can also be involved.[15,37] Most females groan or moan during these contractions. The induction of orgasm in women by coital stimulation alone is not as frequent as that in the male, about half not achieving orgasm unless clitoral stimulation is also used. The reason for this difference is usually ascribed either to the greater inhibitory education about sexual pleasure experienced by women or the lack of correct genital

stimulation. A major difference between males and females is that females can be multiorgasmic because they do not have a PERT after orgasm.[15,23,33] There is mounting evidence that, unlike the unitary concept of the **DEOR** model, the erotic stimulation of different genital sites (especially the anterior vaginal wall compared with the clitoris) induces different types of orgasmic response both subjectively and physiologically.[44,47,48]

Brain activation during human sexual arousal

Before the advent of brain imaging by functional magnetic resonance imaging (fMRI) or positron emission tomography (PET) to identify which areas are activated by and during sexual arousal and orgasm, only limited information was obtained by observing the effects of brain lesions on behaviour, epileptic case studies, electroencephalography (EEG) activity, and rare electrical recordings and stimulations of specific brain areas.[27] Inferences about human sexual arousal mechanisms often had to be made from animal experiments but the problem of species differences always was the spectre in the wings (see comment below).

One type of brain imaging (blood oxygen level dependent [BOLD]) relies on the concept that activation of neurones requires an increased demand for oxygen which normally entails an increase in their blood supply bringing oxygenated blood. The change in the magnetic susceptibilities of the oxygenated and deoxygenated blood is measured and is an index of the increase or decrease in flow to the neuronal area. In the PET technique a short-lived radioactive tracer (often ^{15}O) is injected intravenously and its concentration in the brain area of high blood flow localized by the radioactivity. Both techniques rely heavily on extensive computer programs to correct for a variety of essential artefact corrections one of the most important being corrections for movement. The usual protocol underlying most studies of sexual arousal is to first measure the activity of the brain area under study during a non-sexual basal state (looking at neutral videos) then switching to viewing sexually arousing videos. By subtracting the activity of the neutral from the aroused measurements the remaining activity is assumed to be due the sexual arousal per se. One difficulty is what level (threshold) of activity of a site is to be regarded as physiological. Investigators tend to choose their own criteria making direct comparisons between different studies difficult. Another difficulty is that many areas of the brain are multifunctional (e.g. the amygdala, the periaqueductal grey, the cerebellum) and are thus activated/deactivated by different stimuli (viz., pain, pleasure, fear, anger, emotional processing) the activation/deactivation may be an epiphenomenon of the arousal rather than its cause. Finally, it takes a few seconds to build up a brain scan but neuronal activation takes just hundredths of a second; the image will always lag behind the actual activation so that identifying what is the primary initiator of brain arousal will not be apparent.

Despite all the above caveats about brain imaging one outstanding conclusion is now clear from all the various studies—that there does not appear to be a single site creating arousal or orgasm, multiple site co-activation (also referred to as a neural network) is the rule in both males and females. While a number of the features of brain activation appear common to sexual arousal in the brains of both sexes, the amygdala and hypothalamus are said to be more strongly activated in men when viewing identical sexual stimuli. According to Holstege's group,[47] the first to map brain areas active during ejaculation/orgasm, the strongest activated primary region during male erection/orgasm is the ventral tegmental field region, the midbrain lateral central tegmental field, the zona incerta, the suprafascicular nucleus, the ventroposterior, midline and intralaminar nucleus, and the cerebellum. Decreased activity was seen in the amygdala and adjacent entorhinal cortex. This decrease in the amygdala is said not to occur during women's arousal.[48] Another group.[49] imaging the brain in sexually aroused males reported increased neural activity in areas that included the right frontal cortex, the inferior temporal cortex, the left anterior cingulate cortex, and the right insula. However, only a subset of these areas (anterior cingulate, insula, amygdala, hypothalamus, and secondary somatosensory cortex) were involved in creating a full erection.

Imaging studies undertaken by Komisaruk and Whipple[48] used both women who were paraplegic and some able-bodied women who had the rare ability to create orgasm by mental activity (thoughts) alone. In the former group, where arousal was induced by cervical vibratory stimulation, a large number of sites, including the hypothalamus, were activated by the stimulus used to induce orgasm but in the mental activity group, the only sites activated by thought were regions of the nucleus accumbens, the paraventricular nucleus of the hypothalamus, the hippocampus, and anterior cingulate cortex. The authors suggested that these sites may be specifically related to activate the female orgasm. Interestingly, the amygdala was not activated during these 'thought' induced orgasms. How 'normal' the 'orgasm by thought' group are is yet to be ascertained.

It is unfortunate that while there are now a number of independent studies on brain imaging during sexual arousal the resultant descriptions of the areas claimed to be the activated or deactivated are not in agreement. Part of the problem is that different stimuli, duration of stimuli, data handling, and processing protocols have been used but even these do not explain all the differences. Thus, at present, it is not possible to give a detailed and reliable account of brain activation during sexual arousal.

Summary

Normal human sexual function can be characterized simply by its biological mechanisms which are of obvious importance, not least to reproduction.[41] The mechanisms have changed little over the centuries, but their expression as behaviour, when moulded by historical time, social class, ethnic grouping, religion, and society, creates the changing complex concept of human sexuality. Indeed, it has been said that human sexuality is more about fertilizing relationships than eggs! While we have increased hugely our knowledge about many of the mechanisms involved in human sexuality, the impact of a highly successful oral therapy for erectile dysfunction being an obvious example, those of the brain and spinal cord are practically unexplored. What creates human sexual desire and sexual excitement and what causes them to fade away, where in the brain is the pleasure of orgasms created, why do men have a PERT but not women, are just a few of the fascinating questions that remain to be answered.

Further information

Goldstein, I., Meston, C.M., Davis, S.R., *et al.* (eds.) (2006). *Women's sexual function and dysfunction: study, diagnosis and treatment*. Taylor & Francis, London.

Janssen, E. (ed.) (2007). *The psychophysiology of sex*. Indiana University Press, Bloomington, IN.

Janssen, E., Prause, N., and Geer, J. (2007) The sexual response. In *Handbook of Psychophysiology* (3rd edn) (eds. J.T. Cacioppo, G. Tassinary, and G.G. Bernison), pp. 245–66, Cambridge University Press, New York.

Komisaruk, B.R., Beyer-Flores, C., and Whipple, B. (2006). *The science of orgasm*. The Johns Hopkins University Press, Baltimore, MD.

Lue, T.F., Basson, R., Rosen, R., *et al.* (eds.) (2004). *Sexual medicine- sexual dysfunctions in men and women*. Health Publications, Editions 21, Paris, France.

References

1. Morris, D. (1967). *The naked ape*. Jonathan Cape, London.
2. Heltne, P.G. and Marquardt, L.A. (eds.) (1989). *Understanding chimpanzees*. Harvard University Press, Cambridge, MA.
3. Laqueur, T. (1990). *Making sex: body and gender from the Greeks to Freud*. Harvard University Press, Cambridge, MA.
4. Foucault, M. (1980). *The history of sexuality*. Vintage, New York.
5. Tiefer, L. (1992). Historical, scientific, clinical and feminist criticism of the 'the human sexual response cycle' model. *Annual Review of Sex Research*, **2**, 1–23.
6. Gagnon, J.H. and Simon, W. (1973). *Sexual conduct: the social sources of human sexuality*. Aldine Chicago, IL.
7. Money, J. (1986). *Lovemaps: clinical concepts of sexual/erotic health and pathology, paraphilia, and gender transposition in childhood, adolescence, and maturity*. Irvington, New York.
8. Kinsey, A.C., Pomeroy, W.B., and Martin, C.E. (1948). *Sexual behaviour in the human male*. W.B. Saunders, Philadelphia, PA.
9. Kinsey, A.C., Pomeroy, W.B., Martin, C.E., *et al.* (1953). *Sexual behaviour in the human female*. W.B. Saunders, Philadelphia, PA.
10. Clement, U. (1990). Surveys of heterosexual behaviour. *Annual Review of Sex Research*, **1**, 45–74.
11. Einon, D. (1994). Are men more promiscuous than women? *Ethology and Sociobiology*, **15**, 131–43.
12. Berk, R., Abramson, P.R., and Okami, P. (1995). Sexual activities as told in surveys. In *Sexual nature, sexual culture* (eds. P.R. Abramson and S.D. Pinkerton), pp. 371–86. University of Chicago Press, Chicago, IL.
13. Michael, R.T., Gagnon, J.H., Laumann, E.O., *et al.* (1994). *Sex in America—a definitive survey*. Little, Brown, London.
14. Wellings, K., Field, J., Johnson, A.M., *et al.* (1994). *Sexual behaviour in Britain*. Penguin, Harmondsworth.
15. Masters, W.H. and Johnson, V.E. (1966). *Human sexual response*. Little, Brown, Boston, MA.
16. Robinson, P. (1976). *The modernization of sex*. Cornell University Press, Ithaca, NY.
17. Kaplan, H. (1979). *Disorders of sexual desire*. Simon and Schuster, New York.
18. Garde, K. and Lunde, I. (1980). Female sexual behaviour. A study in a random sample of 40-year-old women. *Maturitas*, **2**, 225–40.
19. Basson, R. (2000). *The female sexual response model revisited. Journal of the Society for Obstetrics and Gynecology of Canada*, **22**, 383–7.
20. Buss, D.M. (1994). *The evolution of desire: strategies of human mating*. Basic Books, New York.
21. Trivers, R.L. (1972). Parental investment and sexual selection. In *Sexual selection and the descent of man 1871–1971* (ed. B. Campbell), pp. 136–79. Aldine, Chicago, IL.
22. Grammar, K. (1989). Human courtship behaviour: biological basis and cognitive processing. In *The sociobiology of sexual and reproductive strategies* (eds. A.E. Rasa, C. Vogel, and E. Voland), pp. 147–69. Chapman & Hall, London.
23. Phillips, A. (1993). *On kissing, tickling and being bored*. Harvard University Press, Cambridge, MA.
24. Nicholson, B. (1984). Does kissing aid human bonding by semio-chemical addiction? *The British Journal* of Dermatology, **111**, 623–7.
25. Levin, R.J. (1992). The mechanisms of human female sexual arousal. *Annual Review of Sex Research*, **3**, 1–48.
26. Levin, R.J. (2007). Sexual activity, health and well being—the beneficial roles of coitus and masturbation. *Sexual and Relationship Therapy*, **22**, 135–48.
27. Levin, R.J. (1994). Human male sexuality: appetite and arousal, desire and drive. In *Appetite: neural and behavioural bases* (eds. C.R. Legg and D. Booth), pp. 127–63. Oxford University Press, Oxford.
28. Bremer, J. (1959). *Asexualisation: a follow up of 244 cases*. Macmillan, New York.
29. Wilson, J. (1978). Sexual differentiation. *Annual Review of Physiology*, **40**, 279–306.
30. Meuwissen, I. and Over, R. (1992). Sexual arousal across phases of the human menstrual cycle. *Archives of Sexual Behaviour*, **21**, 165–73.
31. Hutchinson, K.A. (1995). Androgens and sexuality. *The American Journal of Medicine*, **98**(Suppl. 1A), 111S–15S.
32. Gerstenberg, T.C., Levin, R.J., and Wagner, G. (1990). Erection and ejaculation in man. Assessment of the electromyographic activity of the bulbocavernosus and ischiocavernosus muscle. *British Journal of Urology*, **65**, 395–402.
33. Levin, R.J. (2003). *Is prolactin the biological "off switch" for human sexual arousal? Sexual and* Relationship Therapy, **18**, 239–43.
34. Robbins, M.A. and Jensen, G.D. (1978). Multiple orgasm in males. *Journal of Sex Research*, **14**, 21–6.
35. Levin, R.J. (2005). The mechanisms of human ejaculation—a critical analysis. *Sex & Relationship Therapy*, **20**, 123–31.
36. Langfeldt, T. (1990). Early childhood and juvenile sexuality, development and problems. In *Handbook of sexology*, Vol. 7 (ed. M.E. Perry), pp. 179–200. Elsevier, Amsterdam.
37. Levin, R.J. (2004). An orgasm is . . . who defines what an orgasm is? *Sexual & Relationship Therapy*, **19**, 101–7.
38. O'Connell, H.E., Hutson, J.M., Anderson, C.R., *et al.* (1998). Anatomical relation between urethra and clitoris. *The Journal of Urology*, **159**, 1892–7.
39. van Turnhout, A.A.W.M., Hage, J.J., and van Diest, P.J. (1995). The female corpus spongiosum revisited. *Acta Obstetrica Scandinavica*, **74**, 762–71.
40. Levin, R.J. (1998). Sex and the human female reproductive tract-what really happens during and after coitus. *International Journal of Impotence Research*, **10**(Suppl. 1), S14–21.
41. Levin, R.J. (2005). Sexual arousal—its physiological roles in human reproduction. *Annual Review of Sex Research*, **16**, 154–89.
42. Riley, A.J., Lees, W.R., and Riley, E.J. (1992). An ultrasound study of human coitus. In *Sex matters* (eds. W. Bezemer, P. Cohen-Kettenis, and K. Slob), pp. 29–32. Excerpta Medica, Amsterdam.
43. Grafenberg, E. (1960). The role of urethra in female orgasm. *International Journal of Sexology*, **3**, 145–8.
44. Ladas, A.K., Whipple, B., and Perry, J.D. (2005). *The G spot and other recent discoveries about human sexuality*. Henry Holt & Company Ltd., New York.
45. Minh, H.-N., Smadja, A., and De Sigalony, J.P. (1979). Le fascia de Halban: son role dans la physiologie sexuelle. *Gynaecologie*, **30**, 267–73.
46. Hoch, Z. (1986). Vaginal erotic sensitivity by sexological examination. *Acta Obstetrica et Gynaecologica Scandinavica*, **65**, 767–73.
47. Holstege, G., Geogiadis, J.R., Paans, A.M.J., *et al.* (2005). Brain activation during human male ejaculation. *Journal of Neuroscience*, **23**, 9185–93.
48. Komisaruk, B. and Whipple, B. (2005). Functional MRI of the brain during orgasm in women. *Annual Review of Sex Research*, **16**, 62–86.
49. Ferretti, A., Caulo, M., Del Gratta, C., *et al.* (2005). Dynamics of male sexual arousal: distinct components of brain activiation revealed by fMRI. *NeuroImage*, **26**, 1086–96.

4.11.2 The sexual dysfunctions

Cynthia A. Graham and John Bancroft

Introduction

Sexual relationships are central to the lives of most of us. The sexual component of those relationships can go wrong in various ways. This may be secondary to other difficulties in the relationship, mental health problems, specific sexual vulnerabilities of the individual, or the impact of disease or medication on sexual response. This chapter will describe the more common sexual problems and their prevalence. Evidence related to aetiology of sexual problems and treatment evaluation will be briefly reviewed. In the final section of the chapter, guidelines for the assessment and practical management of sexual problems will be presented.

Historical aspects and some basic concepts

Since 1970, when Masters and Johnson[1] published their groundbreaking book on the treatment of 'human sexual inadequacy', there have been two lines of development in this field, relatively detached from each other until recently: psychological methods of treatment, collectively known as 'sex therapy' and medical interventions, initially focused on erectile problems in men.

The involvement of the medical profession has been substantial, although predominantly involving urologists. Initially there were surgical procedures to implant penile splints or to improve the vascular supply to the penis, and the use of vacuum devices to induce erection mechanically. This was followed by the discovery that injection of smooth muscle relaxants, such as papaverine, phentolamine, or prostaglandin into the erectile tissues of the corpora cavernosa induced erections. Self-injections became widely prescribed. To avoid the need for penile injections, which were not popular among male patients, preparations of prostaglandin for intra-urethral administration became available. This era of medical intervention was characterized by a veritable industry of investigative procedures in attempts to identify local causes for erectile dysfunction (ED). Erectile problems were clearly differentiated into 'organic' and 'psychogenic' subtypes. There was, however, a notable lack of attention to how the brain and psychological processes interacted with these peripheral mechanisms.

Then came the 'Viagra revolution' in the early 1990s. The first oral phosphodiesterase 5 (PDE-5) inhibitor, sildenafil (Viagra®), was found to be effective in enhancing erectile response to sexual stimulation when taken about 1 h before sexual activity. This led to the next phase in the 'medicalization' of male sexual dysfunction, with a shift to the primary care physician as the principle source of treatment and a dramatic reduction in the amount of diagnostic assessment.

The progress of sex therapy has been limited since Masters and Johnson.[1] It has continued to be used, with various adaptations of the original 'sensate focus' approach, incorporating principles of psychoanalytic techniques[2] and cognitive behaviour therapy.[3] The main shortcoming has been inadequate outcome research on the efficacy of these methods.

The next phase in this recent history followed the commercial success of PDE-5 inhibitors for men, with an inevitable quest for a 'Viagra for women'. This has so far proved elusive, but has confronted the 'sexual medicine' community with the complexity of women's sexuality and the need to conceptualize it differently to the sexuality of men.

At the same time, evidence has emerged that treatment of ED with sildenafil and more recent PDE-5 inhibitors, although initially successful in the majority, was being discontinued by a substantial proportion of men.[4] In addition, the female partners of men taking these drugs do not always welcome the associated changes in the sexual relationships.[5] We are now moving into the most recent phase where the 'psychological' and 'organic' approaches, and the professional groups that have been identified with them, have started to interact. There is increasing recognition of the need to integrate psychological and medical methods of treatment,[6,7] but with the important proviso that, at least initially, treatment should focus on the couple and not the individual.

One important aspect of this evolving story is how we define a 'sexual dysfunction', with connotations of abnormal or impaired function, and how it is distinguished from a 'sexual problem' in a more general sense. This issue was epitomized by a publication in the *Journal of the American Medical Association* on the epidemiology of 'sexual dysfunction'.[8] In this widely cited paper, 43 per cent of women and 31 per cent of men were identified as having a 'sexual dysfunction', described as 'a largely uninvestigated yet significant public health problem' (p. 544). The authors commented, 'With the affected population rarely receiving medical therapy for sexual dysfunction, service delivery efforts should be augmented to target high-risk populations'(p. 544).

This dramatic example of 'medicalization', based on extremely limited information from a national survey not designed to assess sexual dysfunction, was effectively challenged by Mercer and colleagues, using data from the UK National Survey of Sexual Attitudes and Lifestyles (NATSAL).[9] This used exactly the same questions as in the Laumann *et al.* study,[8] with the important difference that it was more specific about duration of problems, asking whether particular problems had lasted 'at least 1 month', or 'at least 6 months' during the last year (Laumann *et al.* had asked if symptoms had occurred 'for several months or more' during the last year). Overall, 53.8 per cent of women and 34.8 per cent of men reported at least one sexual problem lasting at least 1 month during the previous year. In contrast, the prevalence of problems lasting 'at least 6 months in the previous year' was 15.6 per cent for women and 6.2 per cent for men. This showed that transient problems were very common, more persistent ones much less so. In both the American and the British study, such problems were related to other problems in the participants' lives, particularly involving impaired mental health (e.g. depression), relationship problems, or significant life stresses.

Relevant to the question of when a sexual problem becomes a 'dysfunction' is a theoretical approach, called the 'Dual Control Model', developed at the Kinsey Institute. This postulates that sexual response results from an interaction between excitation and inhibition, involving relatively discrete neurophysiological systems in the brain.[10] A central assumption of the model is that individuals vary in their propensity for both sexual excitation and sexual inhibition and that 'normal' levels of inhibition are adaptive, reducing sexual responsiveness in circumstances where sexual

activity is best avoided. It is predicted that high levels of inhibition may be associated with vulnerability to sexual dysfunction and low levels with an increased likelihood of engaging in high-risk sexual behaviour.[10]

This faces us with the seemingly obvious but fundamental challenge of deciding whether a loss of sexual interest or responsiveness is an understandable or even adaptive reaction to current circumstances, or is a result of 'malfunction' of the sexual response system, which can appropriately be called a 'sexual dysfunction'. This challenge is also central to the relatively new phase of integrated treatment, in which assessment identifies the key factors causing the sexual problem and how they should best be treated. A strategy for carrying out such assessment, which we have called the 'three windows approach', will be outlined below.

Clinical features of sexual problems

Sexual problems in men

The most common problems presented by men are ED and premature ejaculation (PE). Delayed or absent ejaculation is a relatively infrequent complaint. Low sexual desire may be the presenting problem, although in most cases this is combined with ED, and it is not always clear which came first.

(a) Erectile problems

Penile erection is a tangible and fundamental component of a man's experience of sexual arousal and the lack of erection in a sexual situation often has significant negative effects. Irrespective of whether or not there are peripheral explanations for impaired erections (e.g. vascular disease), the reactions of the man and his partner have a major influence on how problematic the erectile difficulty becomes. Erectile difficulties vary in severity; in some men the problem only occurs on a proportion of occasions of sexual activity. The difficulty may be in getting an erection or in maintaining it long enough for satisfactory sexual intercourse.

(b) Low sexual desire

For many men, sexual desire is linked with erectile responsiveness. Many men with low sexual desire also report a reduction in 'spontaneous' erections. However, a man can experience low sexual desire without having any erectile difficulties, although he may require more direct tactile stimulation to achieve erections.

(c) Premature ejaculation (PE)

Ejaculation results from a combination of orgasm and seminal emission, with muscular contractions as part of the orgasmic response resulting in expulsion of the seminal emission. PE is essentially a problem when the man is unable to delay orgasm and ejaculation as he would wish. Not surprisingly, this has led to considerable inconsistencies of definition in the literature. In severe cases, emission occurs before vaginal entry and the orgasmic component may be so reduced that the usual muscle spasms do not occur, resulting in semen seeping out of the urethra rather than being 'ejaculated'.

Premature ejaculation has been categorized as 'primary' (i.e. lifelong) or 'secondary'. Secondary PE is often confounded by erectile problems. If a man is taking a long time to get an erection, he may reach the stimulus intensity required for ejaculation before or soon after erection is achieved.

(d) Delayed ejaculation

Delayed or absent ejaculation occurs in men, although it is much less common than rapid ejaculation. A man might have difficulty ejaculating only during sexual activity with his partner and in some cases only during penetrative intercourse, or the problem may be evident even during masturbation. Delayed or absent ejaculation is a common side effect of selective serotonin re-uptake inhibitor (SSRI) medications, which often prevents orgasm in women as well, suggesting that the primary effect of such drugs is on the triggering of orgasm.

(e) Pain during sexual response

Pain may be associated with prolonged sexual arousal not terminated by ejaculation/orgasm. Such pain is usually experienced in the testes. Pain felt in the urethra may occur during ejaculation. Neither problem is common.

Sexual problems in women

(a) Loss of sexual arousal and/or desire

Most surveys have suggested that low sexual desire is the most common sexual problem reported by women. However, low sexual desire is a heterogeneous problem category and the relationship between sexual arousal and sexual desire in women is particularly complex. Many women do not differentiate between 'arousal' and 'desire'[11] and awareness of 'desire' is usually accompanied by some degree of central arousal, whether or not any genital response is perceived.[12] It has been argued that sexual desire in women is much more likely to be 'receptive' and triggered by a desire for intimacy with one's partner.[13] It is therefore not surprising that there is considerable overlap or comorbidity between problems related to sexual arousal and desire in women.[8]

Although traditionally seen as the counterpart to penile erection in men, vaginal response is not central to the experience of sexual arousal in women. Vaginal dryness may be a problem because of the likelihood of discomfort or pain with intercourse when the vagina is not adequately lubricated, but this symptom does not necessarily indicate lack of arousal. Conversely, a woman may experience lack of sexual arousal and yet have vaginal lubrication. The relevance of vaginal response to sexual arousal in women therefore remains unclear. An increase in vaginal blood flow has been consistently demonstrated in women reacting to sexual stimuli, whether or not they find the sexual stimulus appealing; this led Laan and Everaerd[12] to call this an 'automatic' response. There is no obvious counterpart to this in men. Tumescence of the clitoris, on the other hand, may be more directly comparable to male genital response but this is less easily assessed and less clearly perceived by women, compared with penile erection in men.

Persistent genital arousal disorder (PGAD) is a recently recognized but fairly uncommon sexual problem in women. It is characterized by genital and breast vasocongestion and sensitivity which persists for hours or days and is only temporarily relieved by orgasm; genital sensations are unaccompanied by any subjective sense of sexual desire and excitement but instead are perceived as intrusive. There is no male equivalent to this problem, probably because the post-orgasmic refractory period is more substantial in men.

Much less frequent than loss of sexual arousal or desire is extreme aversion to, and avoidance of all sexual contact with, a sexual partner. This can occur in women and in men.

(b) Problems with orgasm

Difficulty in achieving orgasm is not uncommon in women. Often this is situational in that orgasm is possible with masturbation, but not during sexual interaction with the partner. The capacity to experience orgasm varies considerably across women. Some women reach orgasm easily if sufficient arousal occurs, others may require more specific or more intense stimulation, and an estimated 10–15 per cent are unable to experience orgasm throughout their lives.[14] In identifying a problem as primarily orgasmic, one needs to first establish that appropriate sexual arousal has occurred.

(c) Problems with sexual pain and vaginismus

Pain during attempted or complete vaginal entry (dyspareunia) is a common sexual problem in women with a wide range of possible causes. Sexual pain is also frequently associated with lack of sexual desire and/or arousal. Vaginismus has traditionally been defined as recurrent or persistent involuntary spasm of the musculature of the outer third of the vagina that makes vaginal penetration difficult or impossible. This definition has recently been questioned because of a lack of empirical evidence that vaginal spasms occur in women diagnosed with vaginismus.[15] Vulvar vestibulitis syndrome (VVS) is a condition associated with pain on touching the labia or vaginal introitus. The question of whether these are primarily 'sexual' or pain disorders is currently under debate.[16]

Classification of sexual problems

DSM-IV and ICD classification

The current Diagnostic and Statistical Manual of Mental Disorders (DSM) classification[17] defines sexual dysfunction as characterized by 'disturbance in sexual desire and in the psychophysiological changes that characterize the sexual response cycle and cause marked distress and interpersonal difficulty' (p. 493). There is Hypoactive Sexual Desire Disorder and Sexual Aversion Disorder, defined in the same way for men and women. Female Sexual Arousal Disorder is defined as 'a persistent or recurrent inability to attain, or to maintain until completion of the sexual activity, an adequate lubrication-swelling response of sexual excitement' (p. 502) and, for the male version, erection is the relevant response. Orgasmic Disorder (i.e. delayed or absent orgasm) and Dyspareunia are defined in basically the same way for men and women. Vaginismus is a specifically female diagnosis and Premature Ejaculation an exclusively male disorder.

The International Statistical Classification of Diseases and Related Health Problems[18] covers sexual dysfunctions in one and a half pages, compared with nearly 30 pages in the DSM. The basic categories of dysfunction are similar to those of DSM, but there are few, if any, actual diagnostic criteria given for any of the dysfunctions. ICD-10 also does not require that personal distress or interpersonal problems are present for a diagnosis to be made. Instead, there is the statement 'sexual dysfunction covers the various ways in which an individual is unable to participate in a sexual relationship as he or she would wish' (p. 355).

There has been increasing dissatisfaction with the current classification of sexual dysfunction for women.[19,20]. Major areas of criticism include the high comorbidity between diagnoses of sexual dysfunction and the 'genital' focus of the diagnostic criteria and concomitant neglect of psychological and relationship factors. In response, alternative models of sexual response[13] and women-centred definitions of sexual problems[21] have been proposed.

Epidemiology

There have been at least 12 representative community-based surveys that have assessed the prevalence of sexual problems.[22] Prevalence rates for specific problems vary considerably and whereas several of the studies claim to be reporting prevalence of sexual 'dysfunctions',[8] it is now accepted that the detailed clinical assessment required to identify a sexual dysfunction cannot be undertaken by large-scale surveys.[23] Consequently, more recent surveys have used terms such as 'problems'[9,24] or 'difficulties'.[25] Variability in reported prevalence rates can in part be attributed to variations in how sexual problems are defined, their duration, and how and whether 'distress' about changes in sexual functioning is assessed. In studies of female sexual problems, there has been only limited overlap between what women perceive as problematic and what the researcher or clinician identifies as a problem.[26,27] The variability of prevalence rates is shown in Tables 4.11.2.1 and 4.11.2.2, which compare a number of population-based surveys involving women and men.

There has been more consistency across studies in the associations found between factors of possible aetiological relevance and sexual functioning. In women, sexual problems are more frequent in those with mental health problems and relationship difficulties.[8] In a survey of women aged 20–65 years, all in heterosexual relationships, 24.4 per cent reported marked distress about their sexual relationship and/or their own sexuality.[26] The best predictors of distress were markers of mental health and the quality of the emotional relationship with the partner. Physical aspects of sexual response in women such as arousal and orgasm were poor predictors of distress.

In men, age has a predictable negative effect on erectile function. In one study, complete erectile failure was reported by 5 per cent of men at age 40 and 25 per cent at age 70.[28] A similar age effect is found with loss of sexual desire. In another study, absence of any sexual desire was reported by 2 per cent of men aged 45–59 and 18.2 per cent aged 75+.[29] Contrary to popular belief, PE does not show a clear negative relationship with age.

The association between age and sexual problems in women is more complex. Whereas level of sexual interest typically decreases with age, older women are less likely to regard this as a problem.[26] Older women are much less likely to be in a sexual relationship than older men; the presence or absence of a partner also seems to influence women's sexuality to a greater extent than it does for men.[29] The impact of the menopause is also complex. Although the post-menopausal decline in oestrogens is relevant to vaginal lubrication, other factors such as mental health and the quality of the sexual relationship are more important determinants of sexual well-being.[30]

Aetiology

Before considering the factors that can cause sexual problems, it is worth underlining the important way that sexual function differs from most other physiological response systems. Although involving physiological mechanisms, sexual responses are most often experienced in the context of a relationship. This highlights the

Table 4.11.2.1 Prevalence of specific female problems (%) found in seven community-based surveys

Study	Low sexual interest	Impaired arousal	Impaired lubrication	Impaired orgasm	Pain	Total (one or more problems)
Richters *et al.* (2003)[25]1						
At least 1 month	54.8	–	23.9	28.6	20.3	
Mercer *et al.* (2003)[9]1						
At least 1 month	40.6	–	9.2	14.4	11.8	53.8
At least 6 months	10.2	–	2.6	3.7	3.4	15.6
Bancroft *et al.* (2003)[26]3	7.2	12.2	31.2	9.3	3.3	45
Laumann *et al.* (1999)[8]1	31.6	–	20.6	25.7	15.5	43[a]
Fugl-Meyer and Fugl-Meyer (1999)[59]1	33.0	–	12.0	22.0	6.0	47
Dunn *et al.* (1999)[24]2	–	17.0	28.0	27.0	18.0	41
Osborne *et al.* (1988)[37]2	17.0	–	17.0	16.0	8.0	33

[a] Includes additional problem categories not shown in this table.

[1] During last year.

[2] During last 3 months.

[3] During last month.

importance of keeping the interactive relationship components in mind when trying to assess and treat sexual problems. Socio-cultural factors are also crucial to understanding how sexual problems are experienced. Much of the focus in medical treatments of sexual problems has been on the individual patient, with relationship and socio-cultural aspects largely ignored. The more specific aetiological factors can now be considered using the 'three windows approach'.[26]

The first window—the current situation

Through the first window, a variety of factors in the individual's current relationship and situation may be relevant. Relationship problems, particularly resentment and insecurity within a relationship, are of particular importance. For many individuals, feeling secure and being able to 'let go' are necessary for them to enjoy sex. Other factors that may be important include: poor communication between partners about their sexual feelings and needs; misunderstandings and lack of information; unsuitable circumstances and lack of time e.g. fatigue, lack of privacy, and work pressures; concerns about pregnancy or about sexually transmitted diseases; and low self-esteem and poor body image.

Various mechanisms may mediate the effects of the above situational factors on sexual functioning. Reactive inhibition, as postulated by the Dual Control Model,[10] may well be involved in those circumstances associated with a negative emotional response. With stress and fatigue, the mechanisms are less clear, but may entail a transient reduction in the capacity for excitation (i.e. sexual arousability).

Table 4.11.2.2 Prevalence of specific male problems (%) found in five community-based surveys

Study	Low sexual interest	Erectile problem	Premature ejaculation	Delayed ejaculation	Total (one or more problems)
Richters *et al.* (2003)[25]1					
At least 1 month	24.9	9.5	23.8	6.3	–
Mercer *et al.* (2003)[9]1					
At least 1 month	17.1	5.8	11.7	5.3	34.8
At least 6 months	1.8	0.8	2.9	0.7	6.2
Laumann *et al.* (1999)[8]1	14.6	10.2	30.6	7.8	31.0[a]
Fugl-Meyer and Fugl- Meyer (1999)[59]1	16.0	5.0	9.0	–	–
Dunn *et al.* (1999)[24]2	–	26.0	14.0	–	–

[a] Includes additional problem categories not shown in this table.

[1] During last year.

[2] During last 3 months.

The second window—vulnerability of the individual

Although a wide range of factors can impact on our sexuality, it is also clear that individuals vary substantially in the extent to which they are affected by such factors, particularly in terms of an associated inhibition of sexual response. Such vulnerabilities are likely to have been evident in earlier episodes in the current relationship or in earlier relationships.

(a) Negative attitudes

Long-standing attitudes, usually stemming from childhood, that sex is inherently 'bad' or immoral are likely to interfere with an individual's ability to become involved in and enjoy a sexual relationship.

(b) Need to maintain self-control

In some individuals, a difficulty in 'letting go' sexually reflects a more general need to maintain self-control, particularly in the presence of another person.

(c) Earlier experience of sexual abuse or trauma

There is now an extensive literature on the impact of sexual abuse on subsequent sexual adjustment. Whereas the mediating mechanisms are not well understood and are likely to be complex and varied, a history of such experience should be regarded as potentially relevant to current sexual difficulties.

(d) Propensity to sexual inhibition

The Dual Control Model, discussed earlier, has led to psychometrically validated measures of propensity to sexual excitation and inhibition in men[31] and women.[32] These are new measures and research exploring their relevance to vulnerability to sexual problems has only recently started. However, using these measures, close to normal distributions of scores for both sexual excitation and inhibition proneness have been reported in men[31] and women.[32] As predicted, a clear association between low sexual excitation and/or high sexual inhibition propensity and erectile problems in men has been found, but no association with PE.[33] One study explored a possible genetic basis for such individual variability in men and found evidence of heritability for propensity to sexual inhibition but not excitation.[34] In women, a strong relationship between sexual inhibition scores and reports of sexual problems was found.[35] Particularly important was one inhibition subscale (labelled 'arousal contingency') that reflects susceptibility for sexual arousal to be easily affected by situational factors e.g. if the circumstances are not 'just right' or if the woman is distracted. Although further research is needed, these measures of sexual inhibition and excitation may prove valuable in explaining patterns of impaired sexual response and helping in the selection of appropriate treatment. As yet no other personality-related or trait measure has been shown to have clinical value in this respect.

The third window—health-related factors that alter sexual function

The variety of health-related factors are considered under three headings: mental health, physical health, and sexual side effects of medication.

(a) Mental health and sexuality

Psychiatric problems are commonly associated with sexual problems.[36,37] Reduction in sexual interest, and to some extent sexual arousability, is generally accepted as a common symptom of depressive illness.[38] In contrast, sexual interest tends to be increased in states of elevated mood such as hypomania.[39] In a study of men and women with primary loss of sexual interest, the large majority had experienced previous affective disorder, with reduced sexual desire being established during, and persisting after, one of these earlier depressive episodes.[40]

We would expect reactive inhibition of sexual interest or response in circumstances eliciting negative mood (i.e. 'reactive depression'). With more endogenous depression, however, reduction of the excitatory component of sexual response is also evident. This is shown clearly in the impairment of nocturnal penile tumescence (NPT) that typically occurs with depressive illness. NPT, while not completely understood, probably results from a 'switching off' of inhibitory tone during REM sleep. This therefore allows expression of 'excitatory tone', presumably impaired in depression with associated metabolic changes.

With anxiety, the clinical evidence is much more limited. Higher rates of sexual dysfunction in patients with anxiety disorders have been reported.[41] In the Zurich Cohort Study, a longitudinal study of men and women aged 20–35 years, loss of sexual interest was more prevalent in patients with generalized anxiety disorder, but was not associated with panic disorder, agoraphobia, or social phobia.[42] In another study, patients with panic disorder were more likely to report sexual problems, particularly sexual aversion, than social phobics; PE was the most common sexual problem in men with social phobias.[43]

Non-clinical, community-based studies have also demonstrated a relationship between mood and sexuality. In the Massachusetts Male Aging Study, an association was reported between ED and depressive symptoms, after controlling for other potential confounding variables such as age and physical health.[44] Angst[42] found a relationship between depression and loss of sexual interest, particularly marked in women. In a US survey of heterosexual women, scores on a brief measure of mental health were strongly predictive of women's distress about their sexual relationship and their own sexuality.[26]

Although most studies have focused on negative effects of mood disorders on sexual interest and response, there is evidence that a minority of individuals experience increased sexual interest during negative mood states. Angst[42] reported that among depressed men, 25.7 per cent reported decreased, and 23.3 per cent increased, sexual interest, compared with 11.1 per cent and 6.9 per cent, respectively, of their non-depressed group. In contrast, among depressed women, 35.3 per cent reported decreased sexual interest and only 8.8 per cent reported increased interest (compared with 31.6 per cent and 1.7 per cent, respectively, of their non-depressed group). This paradoxical pattern of increased sexual interest during negative mood states has also been reported in non-clinical samples of men[45] and women.[46] Although the origins of this pattern are not yet understood, it may well be problematic in various ways; for example, associated with sexual risk-taking or leading to sex being used as a mood regulator.

The impact of schizophrenia on sexuality is complex. Given the importance of dopaminergic neurotransmission to various aspects of sexuality, and the fact that most anti-psychotics are dopamine antagonists, one might expect amplification of sexual interest or response in some form, at least during the more florid type one stage of the illness. Sexual thoughts and behaviours are common in

schizophrenia and there may be a relative increase in sexual activity. However, this is typically autoerotic or 'compulsive' without any real 'object-relational' quality.[47] In view of the effects of this illness on interpersonal functioning, this is not surprising. Early studies suggested that loss of sexual interest was less likely in schizophrenia than in other types of psychiatric illness, although female schizophrenics were more likely to report a reduction in sexual interest than males.[47] 'Pre-schizophrenic personality', evident in the history of many cases, may also be associated with an interference with normal sexual development.

(b) Physical health and sexuality

The impact of poor physical health on sexuality may be relatively non-specific. For example, loss of well-being and energy associated with chronic illness is likely to cause reduced sexual interest and arousability. Psychological reactions to the illness or condition (e.g. the effects of breast cancer on a woman's body image and hence her sexual enjoyment) may also be important. In addition there are a variety of ways in which the health problem can directly affect sexual interest and/or response. In most cases we find the evidence much clearer in men than in women.

(i) Damage to the neural control of genital response

This can involve peripheral mechanisms (e.g. autonomic neuropathy and peripheral neuropathy) or disease in the spinal cord (e.g. multiple sclerosis). Injury or surgery causing nerve damage may be involved (e.g. spinal cord injury, prostatectomy, or hysterectomy). The most likely consequences are ED in men and impaired vaginal response in women. There is no clear relationship between neural damage and PE but interference with normal seminal emission can occur, with resulting loss of ejaculation or retrograde ejaculation when, due to nerve damage to the pelvic muscles, the seminal fluid passes backwards into the bladder during ejaculation. Brain abnormalities, such as epilepsy or cerebral tumour, can affect central control of sexual response, the precise effect depending on the site of the abnormality or tumour. In some cases the result is loss of sexual interest and arousability; rarely there is disinhibition of sexual behaviour.

(ii) Impairment of vascular supply of the genitalia

Genital response is dependent on increased arterial inflow as well as alteration in venous outflow. Vascular impairment can result from peripheral vascular disease, affecting the small vessels, or obstruction in a main artery. ED is also common in men who have ischaemic heart disease.

(iii) Alteration of endocrine mechanisms affecting sexual interest, arousal and response

In males any cause of lowered testosterone (T) levels is likely to produce loss of sexual interest and, to a varying extent, impairment of erectile response. Hypogonadism, if severe enough, will also be associated with loss of seminal emission (and usually orgasm).

In women, lack of oestrogen is associated with impaired vaginal lubrication. The effects of sex steroids, either oestrogens or androgens, on sexual interest and arousability, are much less predictable. The evidence is consistent with there being a proportion of women who depend on T for their normal level of sexual interest, but there are many women who can experience substantial reduction in T without obvious adverse sexual effects. The role of oestrogen in the more central aspects of women's sexuality also remains unclear.[48]

Hyperprolactinaemia, usually resulting from pituitary adenomas and hypersecretion of prolactin, may be associated with loss of sexual interest, and in men, with ED. However, this is not always the case. The precise role of prolactin in human sexuality is not understood. Its central control by dopamine and serotonin contributes to the complexity. Adverse effects of hyperprolactinaemia on sexuality are probably more likely in men than women.

(iv) Metabolic disorders

Various forms of metabolic disturbance, such as that associated with hepatic or renal disease, may be associated with adverse effects on sexual interest and response, though the mediating mechanisms are not understood.

Some diseases affect sexuality through more than one of the above mechanisms. Diabetes mellitus is a good example. ED has long been recognized as a complication of diabetes and is more common in Type 1 than Type 2 diabetes. In diabetic men, ED can result from small vessel vascular disease and also autonomic and peripheral neuropathology. Diabetes is associated with hypogonadism in men. Lowered sexual interest and impairment of genital response have also been reported in diabetic women.

(c) Side effects of medication

In considering the complexities of pharmacological effects on sexual function, it is helpful to distinguish between excitatory and inhibitory mechanisms. In the brain, sexual excitation depends to a considerable extent on dopamine (DA) and noradrenaline (NA). DA is involved in the 'incentive motivation system' and is hence relevant to sexual interest. It is the D2 receptor that is most relevant to these sexual effects. It is also involved in the hypothalamic control of genital response. NA, when acting centrally, is involved in central arousal, a key component of sexual arousal though not specific to it. In the periphery, depending on the receptor involved, NA can have inhibitory or enhancing effects on smooth muscle relaxation, a critical aspect of genital response. There are three principal NA receptors of relevance: (i) alpha-1, which is post-synaptic and mediates smooth muscle contractions, inhibiting genital response, (ii) alpha-2 which is principally pre-synaptic, where it increases re-uptake of NA and hence reduces the amount of NA at the synapse, and (iii) beta-2, a peripheral receptor which mediates a relaxing effect of NA on smooth muscle in the urogenital system and elsewhere. Serotonin (5-HT) has a central role in the inhibitory system. The two most relevant receptors are the 5-HT2 receptor, which mediates the inhibitory effects and the 5-HT1A receptor, which is pre-synaptic and, comparable to the alpha-2 receptor, reduces 5-HT transmission. We can now consider the main adverse sexual effects resulting from medication; for a review, see.[49]

(d) Anti-depressants

The clearest examples are the SSRIs which, by inhibiting the 5HT-1A receptor, increase serotonergic transmission. The most predictable effect, in both women and men, is inhibition of orgasm and ejaculation. This effect is being exploited in the treatment of PE, with the development of short-acting SSRIs which can be used as needed prior to sexual activity. Other negative effects include reduced sexual interest and arousability, though they are less predictable and not always easy to distinguish from the sexual effects of the affective disorder being treated. Tricylic anti-depressants also commonly produce sexual side effects. Two anti-depressants which

have less sexual side effects are bupropion, which is metabolized into a DA and NA re-uptake inhibitor, and nefadazone, which has a 5-HT2 antagonist effect.

(e) Anti-psychotic medication

These drugs, used for the treatment of schizophrenia and other psychotic disorders, involve a balance of DA antagonist and 5-HT agonist effects. Sexual side effects occur in around 60 per cent of men and 30–90 per cent of women using these drugs. The most common side effects are ED and ejaculatory difficulties in men and orgasmic dysfunction in women. It is not clear to what extent these effects are due to the DA antagonist or 5-HT agonist effects, or a balance of the two.

(f) Anti-hypertensive medication

Many drugs used to treat hypertension interfere with male sexual response. In the case of alpha-1 antagonists (e.g. prazosin), the peripheral effects would be expected to enhance erection, by reducing adrenergic contraction of erectile smooth muscle and indeed priapism is an occasional problem. There is a low incidence of sexual side effects, mainly failure of ejaculation, and central alpha-1 blockade might reduce central arousal. Beta-blockers (e.g. propanalol), by blocking smooth muscle relaxation and leaving the alpha-1 induced contraction unopposed, commonly cause ED. Centrally acting anti-hypertensives (e.g. guanethidine) also interfere with sexual response, impairing erection, and blocking ejaculation. Clonidine, an alpha-2 agonist, causes erectile problems in about 25 per cent of cases. In this case the principal effect is likely to be reduced central arousal. For a review of the effects of anti-hypertensive medication on male sexual response, see.[50]

The effects of drugs on women's sexuality have received far less attention. Research has mostly focused on difficulties in achieving orgasm. As orgasm only occurs after sufficient sexual arousal, these effects may reflect impairment of arousal but this has not been adequately assessed. Steroidal contraceptives, although associated with markedly reduced levels of free T, decrease sexual interest only in a minority of women.[51] This, again, has been little studied and it remains possible that the adverse effects on sexual interest only occur in those women whose sexuality is 'T dependent'.[51]

Management of sexual problems

In this section we will briefly describe the principles of sex therapy and the main forms of pharmacological treatment, followed by an outline of the process of assessment and selection of a suitable treatment plan. The growing awareness of the limitations of pharmacological treatments administered on an individual basis, discussed earlier, has led to recognition that an integration of psychological and pharmacological methods, with emphasis on the couple, may be the most appropriate treatment model in most cases.

Sex therapy

Although the approach first introduced by Masters and Johnson[1] has been adapted in various ways, the core treatment techniques remain. Originally developed for helping couples, the techniques can be modified for use with individuals and with same-sex as well as heterosexual couples.

The key elements of the therapeutic process are:

(a) clearly defined tasks are given and the couple asked to attempt them before the next therapy session

(b) those attempts, and any difficulties encountered, are examined in detail

(c) attitudes, feelings, and conflicts that make the tasks difficult to carry out are identified

(d) these are modified or resolved so that subsequent achievement of the tasks becomes possible

(e) the next tasks are set, and so on

The tasks are mostly behavioural in nature. They are chosen to facilitate the identification of relevant issues but in some cases are sufficient in themselves to produce change. The behavioural programme is in three parts. In the first part the couple are asked to avoid any direct genital touching or stimulation and to focus on non-genital contact, alternating who initiates and who does the touching. These first, non-genital steps are effective in identifying important issues in the relationship, such as lack of trust or counter-productive stereotypical attitudes (e.g. once a man is aroused, he can't be expected not to have intercourse). Once this stage can be carried out satisfactorily, and related problematic issues dealt with, the programme moves on to the second part, which allows genital touching to be combined with non-genital touching, with penile–vaginal intercourse still 'out of bounds'. In this second part more intra-personal problems, such as long-standing negative attitudes about sex, or the sequelae of earlier sexual trauma are likely to emerge. In the third part, a gradual approach to vaginal–penile contact and insertion is undertaken. Here the most relevant issues are 'performance anxiety' and fear of pain.

As the behavioural tasks reveal key issues that need to be resolved before moving on to the next stage, a variety of psychotherapeutic approaches, including cognitive behavioural techniques, can be utilized. Although the stage at which particular issues emerge does vary from case to case, there is a tendency for problems identified through the 'first window', particularly those related to lack of trust and unresolved resentment, to appear during the first stage of the programme. Intra-personal issues (e.g. as seen through the 'second window') are more likely to be recognized during the second stage.

The goals of therapy include helping the individual to accept and feel comfortable with his or her sexuality and helping the couple to establish trust and emotional security and to enhance the enjoyment and intimacy of their sexual interaction. An important point is that these goals do not include reversal of specific sexual dysfunctions. There are exceptions; for example, there are specific behavioural techniques to deal with PE and vaginismus (for details of these techniques, see Bancroft.[52] However, the overriding principle is that, assuming there is no abnormality of the basic physiological mechanisms involved in sexual response, normal sexual function (in terms of sexual desire, arousal, and genital response) will return once the above goals are achieved. In cases where impairment of physiological mechanisms does exist, the above goals of sex therapy are still helpful and integrate well with the use of pharmacological treatment.

Practical aspects

Although sex therapy varies in duration, 12 sessions over 4 to 5 months is typical. The therapist adjusts to the particular needs of the individual or couple. Treatment begins weekly with the interval between sessions extended once major issues like unexpressed

resentment or communication problems have been dealt with. The last two or three sessions are spaced out over a few months so that the couple have an opportunity to consolidate their progress and cope with any setbacks before termination. Open-ended arrangements about length of treatment are best avoided. A specified number of sessions are agreed on at the outset with the proviso that progress will be assessed and a decision made on that basis whether to continue for longer.

A more complete description of the sex therapy process is provided by Bancroft.[52]

Pharmacological treatments for men

(a) Erectile dysfunction. Phosphodiesterase 5 inhibitors

The most important development in this field was the serendipitous discovery that sildenafil (Viagra®), a PDE-5 inhibitor, enhances erectile response. Dose titration is usually required, with available tablets containing 25, 50, or 100 mg. For more severe cases of ED, 100 mg taken about 1 hour before sexual activity is required. The maximal effects last for about 4 h. The most common side effects, that are dose-related, are headache, flushing, and dyspepsia.

Two further PDE-5 inhibitors have been developed and are now in use: tadalafil (Cialis®) and vardenafil (Levitra®). They are both comparable to sildenafil in their efficacy, the main differences being their speed and duration of action. Tadalafil should be taken at least 30 min before sexual activity and has a half-life of 17.5 h; the treatment effect can persist for 24–36 h. Vardenafil is pharmacokinetically similar to sildenafil, but is more potent, with a dose range 5 mg to 20 mg. They are similar in their side effects, though the duration of these relate to the half-life of the drug. For all three drugs, the most important and dangerous drug interaction is with nitrates used for ischaemic heart disease; this is a strong contraindication for the use of PDE-5 inhibitors.

There is now an extensive literature demonstrating the efficacy of sildenafil in the treatment of ED, which has been well reviewed by Rosen and McKenna.[53] Overall, treatment is effective in about 75 per cent of cases. The effectiveness of tadalafil and vardenafil has also been demonstrated.

In spite of their efficacy, the continuing use of PDE-5 inhibitors appears to be limited. In a recent large study done in eight countries, 2912 men identified with ED were followed up; 58 per cent of them had sought medical help for the ED, but only 16 per cent of men maintained their use of PDE-5 inhibitors.[7] Various reasons were given for discontinuation, including lack of appropriate information from the physicians, fear of side effects, partner concerns, and distrust of medications.

(b) Anti-adrenergic drugs

Alpha-2 antagonists may enhance sexual arousal by their central NA effects. One example is yohimbine, which has a modest therapeutic benefit over placebo and is generally well tolerated. Phentolamine is an alpha-1 and alpha-2 antagonist used medically to treat hypertensive crises. It has been used, in combination with papaverine, to induce erection by intra-cavernosal injection, presumably mediated by its alpha-1 antagonist effect. Phentolamine administered systemically can also improve sexual arousal, presumably by blocking both the peripheral alpha-1 receptors and enhancing central arousal by blocking the alpha-2 receptors in the brain. Some evidence of beneficial effects of phentolamine has been reported, although long-term follow-up revealed high attrition.

(c) Dopamine agonists

The main effects of dopamine agonists, such as apomorphine, are to induce genital response, probably via the hypothalamic oxytocinergic system. Early studies with apomorphine in men, while showing positive effects on erection, also demonstrated substantial side effects (mainly nausea and dizziness). Sublingual administration was developed to reduce such side effects and a number of placebo-controlled studies have shown this to improve erectile function[54] but the occurrence of side effects remains a limiting factor.

(d) Melanocortin agonists

A recent development has been bremelanotide (PT-141), a metabolite of melanotan-II, and a melanocortin analogue, which probably works through the same oxytocinergic system as apomorphine. Side effects similar to those with apomorphine occur, and further phase 3 research is required before the clinical value of this compound is established.

(i) Premature ejaculation

The orgasm inhibiting effect of SSRIs (see above) has been exploited in the treatment of PE. Paroxetine, sertraline, and fluoxetine, among others, have all been shown to be effective in this respect. However, continued use is required, and effects may not be apparent during the first week. This raises the issue of side effects and the need to avoid sudden withdrawal.

More recently, PDE-5 inhibitors have been explored as a treatment for PE. However, the benefits are not clear cut and may be restricted to cases of 'secondary' PE. The most likely mechanism of action is reduction of the smooth muscle contraction involved in seminal emission.

(ii) Delayed or absent orgasm/ejaculation

At present there is no accepted pharmacological treatment for delayed or absent ejaculation or orgasm.

(iii) Low sexual desire

The most treatable but relatively infrequent cause of loss of sexual desire in men is hypogonadism. Where androgen deficiency is evident, T replacement is indicated. Currently the most favoured route of administration is transdermal, either using cream or skin patches. One advantage of the transdermal route is that the hormone is absorbed into the skin and then released more gradually into the circulation, maintaining relatively physiological levels. In comparison, intra-muscular injections of T (e.g. T enanthate) produce supraphysiological levels in the circulation, which then steadily decline. Oral routes are complicated by first pass through the liver.

If loss of sexual desire is associated with hyperprolactinaemia, treatment with dopamine agonists, such as bromocriptine, is indicated.

Pharmacological treatments for women

At present pharmacological methods for treating sexual dysfunction in women, although generating much interest and controversy, are very limited. Most attention is currently being paid to the use of T for treating low sexual desire. As considered earlier, the evidence of a role for T in women's sexuality is inconsistent and the likelihood is that it is important for some women and not others.

The most convincing evidence of the benefits of T are in women who have been ovariectomized and have therefore experienced a substantial reduction in their androgen levels. Much of this evidence, however, indicates that supraphysiological levels of T result from treatment; thus, it remains uncertain whether the benefits result from correction of T deficiency or pharmacological effects of high levels of T. For this reason, and because of uncertainty about possible long-term risks, there has been a reluctance to approve such treatments, particularly in women with intact ovaries.[48]

Attempts to treat sexual desire/arousal problems with PDE-5 inhibitors have so far proved unsuccessful, although there may be subgroups of women who benefit. Other pharmacological approaches which are being explored in women include the use of phentolamine, apomorphine, bupropion, and bremelanotide (PT-141). But further research is required before any of these can be recommended for clinical use. For more details of the above treatments, see Bancroft.[52]

Evaluation of treatments

Sex therapy

In a 1997 review, Heiman and Meston[55] concluded that there was evidence for empirically validated treatments for orgasm problems in women and ED in men. There was inadequate support for effective treatments for low sexual desire, sexual aversions, dyspareunia in women and men, and delayed orgasm in men. What is striking is that of the 90 studies involving psychological treatments cited in this review, only two were published since 1990, and 60 of them before 1980.

In the decade since this review, there have been a few controlled trials involving cognitive behaviour therapy (CBT) for women with vaginismus,[56] vulvar vestibulitus,[57] and couples presenting with problems of low sexual desire,[58] all suggesting positive effects. Further outcome studies of psychological therapies that identify predictors of successful outcomes are badly needed.

Combined psychological and medical treatment

There have now been a number of studies in men that have compared combined psychological and pharmacological treatments, with either used separately. These have been reviewed by Althof[6] and include studies combining psychotherapy with sildenafil, and counselling with intra-cavernosal injections or with vacuum devices. In each case, the combination produced significantly better results than the medical treatment alone.

Planning a treatment programme

Assessment

When couples present with sexual difficulties, one of them is usually regarded as having the problem. However, both partners should be carefully assessed whenever possible. There are three stages to assessment; first, to facilitate the decision about whether sex therapy is appropriate; second, to identify issues relevant to the sexual problem that need to be resolved, and third, to determine whether medication or other treatments are required.

Keeping in mind the distinction between a sexual problem that is adaptive or appropriate given an individual's current circumstances and one that is maladaptive (and can perhaps be considered a 'dysfunction'), we can assess each individual's case through the three conceptual 'windows' described earlier. Are there problems in the couple's current relationship or situation, which would make inhibition of sexual responsiveness in either partner understandable or adaptive? Does either individual give a history that suggests vulnerability to sexual problems? Are there are any mental or physical health issues or medication use in either partner that could be having a negative effect on sexual interest or response?

(a) The initial interview

Although not all of the details can be obtained during the initial interview, assessment of the following topics should be carried out: the nature of the problem, including an assessment of level of sexual interest and response in each partner; identification of other assessments that may be needed (e.g. physical examination, blood tests); commitment and motivation of each partner to improving the relationship; and, assessment of the mental state of each partner. If both partners are present, each should be interviewed separately following a conjoint interview. As far as possible, questions should be asked about each individual's sexual history, the nature of the current relationship, contraceptive use and reproductive history, and alcohol or recreational drug use.

At the end of the initial interview, the clinician should provide a preliminary formulation of the nature of the problem and the types of intervention that may be helpful. Whatever treatment methods are used, it is important to continue to see the couple together to monitor progress and provide counselling as needed.

If there is no evidence of causal factors of the kind viewed through the 'first' or 'second' windows (see above), then a trial of pharmacological treatment may be appropriate. In such cases a physical examination and, where relevant, laboratory investigations would normally be arranged before starting treatment. It is important to have a good 'clinical baseline' before embarking on pharmacological interventions.

If there are any indications of problems or issues, particularly of the kind that invoke inhibition of sexual response, which need to be resolved, then pharmacological interventions should not be considered as a first step. In many cases, more assessment is required before such factors can be adequately assessed. There are then two options: (i) further interview(s) to explore such issues or (ii) starting on a programme of sex therapy. The rationale for the second option is as follows. The initial two stages of sex therapy (i.e. involving non-genital and then genital touching with no attempts at vaginal intercourse) are particularly effective at identifying relevant issues underlying the problem. Furthermore, the process involved in those early stages is likely to benefit any sexual relationship, even those without obvious problems. After three or four sex therapy sessions, a re-appraisal would be made and a decision taken whether to continue with sex therapy alone, or to combine it with an appropriate pharmacological method. For example, with a couple where the man has erectile problems, it is easier to assess the indications for the use of a PDE-5 inhibitor after the couple have gone through the first two stages of the programme where there is no 'performance pressure' and no need for an erection to occur. Similarly, it is often informative to see what impact this behavioural programme has on the individual who has complained of reduced sexual desire, before attempting to deal with the problem pharmacologically or hormonally.

Ideally, the use of the pharmacological method should then proceed in combination with the continuation of the sex therapy programme. In that way a gradual transition to a satisfying sexual relationship can be achieved without renewed 'performance pressures'. If, however, progress continues to be made with sex therapy, then the addition of pharmacological interventions may not be necessary. It should be explained to the couple that, whereas pharmacological treatments often have beneficial effects on sexual response, they do not 'cure' the problem, which is likely to recur once the medication is stopped. Furthermore, there are often side effects that have to be dealt with. There is, however, some evidence that when medication is combined with sex therapy, the pharmacological component can be gradually withdrawn, or only used intermittently.

Ethical aspects of clinical management

In assessing individuals or couples with sexual problems it is clearly important to assure them that the information they provide will be treated as highly confidential. It is appropriate to have separate case files for sexual problem clinics, rather than files used in other clinics. In this way the files can be kept secure. Some overview of the information presented to the clinic by the referring clinician should be provided at the start of the assessment. Information to be passed back to the referring clinician by letter or word of mouth should be agreed with the client(s) in advance, particularly if any highly sensitive issue emerged during the assessment. When assessing a couple, and seeing each partner separately, it is important to establish whether each individual agrees to any information not known by the other being shared during the joint sessions (e.g. details of previous sexual relationships). When offering a course of treatment or further assessment, the individual or couple should be fully informed about what is involved before being asked to make their decision.

A more challenging ethical issue relates to cultural values. Much of the process of sex therapy involves the therapist encouraging and 'giving permission' for specific forms of sexual interaction, as well as more general patterns of interaction within the relationship (e.g. 'self asserting' and 'self protecting'). It is important for the therapist to keep in mind that many of these principles are based on western middle class values. When working with a couple from a different culture, or even a different social class to the therapist, there should be awareness, openness, and discussion about contrasting values of cultural or religious origin. In this way differences can be negotiated, rather than have therapist values imposed.

When to seek specialist help

The treatment programme described above can be applied by psychiatrists in practice providing that they (i) are comfortable asking detailed questions about sexual behaviour and response, (ii) have experience working psychotherapeutically with couples, and (iii) are familiar with cognitive behavioural methods of treatment. They have the advantage over many non-medically qualified sex therapists of being able to prescribe medications in combination with a behavioural approach to sex therapy.

Specialist help should be obtained if there is evidence of a physical condition which requires detailed assessment and treatment, or if, after a reasonable trial of integrated psychological and pharmacological treatment, the sexual problem remains unresolved. In some such cases, the specialist help should be from someone with expertise in sexual medicine (e.g. urologist, gynaecologist). In other cases, where the psychological aspects remain unclear or difficult to resolve, referral to an experienced sex therapist is appropriate.

Further information

American Association of Sex Counselors, Educators, and Therapists (AASECT). http://www.aasect.org/.

British Association of Sexual and Relationship Therapists (BASRT) [http://www.basrt.org.uk].

Bancroft, J. *Human sexuality and its problems* (3rd edn). Elsevier, Oxford, in press.

Kingsberg, S., Althof, S.E., and Leiblum, S. (2002). Books helpful to patients with sexual and marital problems. *Journal of Sex & Marital Therapy*, **28**, 219–28.

Leiblum, S.R. (2007). *Principles and practice of sex therapy* (4th edn). Guilford, New York.

References

1. Masters, W.H. and Johnson, V.E. (1970). *Human sexual inadequacy*. Little Brown, Boston.
2. Kaplan, H.S. (1975). *The new sex therapy*. Bailliere Tindall, London.
3. Pridal, C.G. and LoPiccolo, J. (2000). Multielement treatment of desire disorders: integration of cognitive, behavioral, and systemic therapy. In: *Principles and practice of sex therapy* (3rd edn) (eds S.R. Leiblum and R.C. Rosen), pp. 57–81. Guilford Press, New York.
4. Althof, S. (2002). When an erection alone is not enough: biopsychosocial obstacles to lovemaking. *International Journal of Impotence Research*, **14**(Suppl. 1), S99–104.
5. Potts, A., Gavey, N., Grace, V.M., *et al.* (2003). The downside of Viagra: women's experiences and concerns. *Sociology of Health and Illness*, **25**, 697–719.
6. Althof, S. (2006). Sexual therapy in the age of pharmacotherapy. *Annual Review of Sex Research*, **17**, 116–31.
7. Rosen, R.C. (2007). Erectile dysfunction: integration of medical and psychological approaches. In *Principles and practice of sex therapy* (4th edn) (ed. S.R. Leiblum), pp. 277–312. Guilford Press, New York.
8. Laumann, E.O., Paik, A., and Rosen, R.C. (1999). Sexual dysfunction in the United States: prevalence and predictors. *The Journal of the American Medical Association*, **281**, 537–44.
9. Mercer, C.H., Fenton, K.A., Johnson, A.M., *et al.* (2003). Sexual function problems and help seeking behaviour in Britain: national probability sample survey. *British Medical Journal*, **327**, 426–7.
10. Bancroft, J. (1999). Central inhibition of sexual response in the male: a theoretical perspective. *Neuroscience and Biobehavioral Reviews*, **23**, 763–84.
11. Graham, C.A., Sanders, S.A., Milhausen, R.R., *et al.* (2004). Turning on and turning off: a focus group study of the factors that affect women's sexual arousal. *Archives of Sexual Behavior*, **33**, 527–38.
12. Laan, E. and Everaerd, W. (1995). Determinants of sexual arousal: psychophysiological theory and data. *Annual Review of Sex Research*, **6**, 32–76.
13. Basson, R. (2000). The female sexual response: a different model. *Journal of Sex & Marital Therapy*, **26**, 51–64.
14. Kinsey, A.C., Pomeroy, W.B., Martin, C.F., *et al.* (1953). *Sexual behavior in the human female*. Saunders, Philadelphia.
15. Weijmar Schultz, W., Basson, R., Binik, Y., *et al.* (2005). Women's sexual pain and its management. *The Journal of Sexual Medicine*, **2**, 301–16.

16. Binik, Y.M. (2005). Should dyspareunia be retained as a sexual dysfunction in DSM-V? A painful classification decision. *Archives of Sexual Behavior*, **34**, 11–22.
17. American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th edn, text revision) (DSM-IV). Author, Washington, DC.
18. World Health Organization. (1992). *ICD-10: International statistical classification of diseases and related health problems* (10th edn). Author, Geneva.
19. Bancroft, J., Graham, C.A., and McCord, C. (2001). Conceptualizing sexual problems in women. *Journal of Sex & Marital Therapy*, **27**, 95–103.
20. Tiefer, L. (2001). A new view of women's sexual problems: why new? why now? *The Journal of Sex Research*, **38**, 89–96.
21. The Working Group for a New View of Women's Sexual Problems. (2001). A new view of women's sexual problems. In *A new view of women's sexual problems* (eds E. Kaschak and L. Tiefer), pp. 1–8, The Haworth Press, New York.
22. West, S.L., Vinikoor, L.C., and Zolhoun, D. (2004). A systematic review of the literature on female sexual dysfunction: prevalence and predictors. *Annual Review of Sex Research*, **15**, 40–172.
23. Graham, C.A. and Bancroft, J. (2005). Assessing the prevalence of female sexual dysfunction with surveys: what is feasible? In *Women's sexual function and dysfunction: study, diagnosis and treatment* (eds I. Goldstein, C. Meston, S. Davis, and A. Traish), pp. 52–60. Taylor and Francis, London.
24. Dunn, K.M, Croft, P.R., and Hackett, G.I. (1999). Association of sexual problems with social, psychological, and physical problems in men and women: a cross sectional population survey. *Journal of Epidemiology and Community Health*, **53**, 144–8.
25. Richters, J., Grulich, A.E., de Visser, R.O., *et al.* (2003). Sexual difficulties in a representative sample of adults. *Australian and New Zealand Journal of Public Health*, **27**, 164–70.
26. Bancroft, J., Loftus, J., and Long, J.S. (2003). Distress about sex: a national survey of women in heterosexual relationships. *Archives of Sexual Behavior*, **32**, 193–8.
27. King, M., Holt, V., and Nazareth, I. (2007). Women's views of their sexual difficulties: agreement and disagreement with clinical diagnoses. *Archives of Sexual Behavior*, **36**, 281–8.
28. Feldman, H.A., Goldstein, I., Hatzichristou, D.G., *et al.* (1994). Impotence and its medical and psychosocial correlates: results of the Massachusetts Male Aging Study. *The Journal of Urology*, **161**, 54–61.
29. American Association for Retired Persons (AARP). (1999). *Modern maturity sexuality study*. NFO Research, Inc., Washington, DC.
30. Hayes, R. and Dennerstein, L. (2005). The impact of aging on sexual function and sexual dysfunction in women: a review of population based studies. *The Journal of Sexual Medicine*, **2**, 317–30.
31. Janssen, E., Vorst, H., Finn, P., *et al.* (2002). The Sexual Inhibition (SIS) and Sexual Excitation (SES) Scales: i. Measuring sexual inhibition and excitation proneness in men. *The Journal of Sex Research*, **39**, 114–26.
32. Graham, C.A., Sanders, S.A., and Milhausen, R.R. (2006). The Sexual Excitation/Sexual Inhibition Inventory for Women: psychometric properties. *Archives of Sexual Behavior*, **35**, 397–409.
33. Bancroft, J., Herbenick, D., Barnes, T., *et al.* (2005). The relevance of the dual control model to male sexual dysfunction: the Kinsey Institute/BASRT collaborative project. *Sexual & Relationship Therapy*, **20**, 13–30.
34. Varjonen, M., Santilla, P., Hoglund, M., *et al.* Genetic and environmental effects on sexual excitation and sexual inhibition in men. *The Journal of Sex Research,* **44**, 359–69.
35. Sanders, S.A., Graham, C.A., and Milhausen, R.R. (2008). Predicting sexual problems in women: relevance of sexual inhibition and sexual excitation. *Archives of Sexual Behavior*, **37**(2), 241–51
36. Lindal, E. and Stefansson, J.G. (1993). The lifetime prevalence of psychosexual dysfunction among 55 to 57-year-olds in Iceland. *Social Psychiatry and Psychiatric Epidemiology*, **28**, 91–5.
37. Osborn, M., Hawton, K., and Gath, D. (1988). Sexual dysfunction among middle-aged women in the community. *British Medical Journal*, **296**, 959–62.
38. Beck, A.T. (1967). *Depression: clinical, experimental and theoretical aspects*. Staples Press, London.
39. Segraves, R.T. (1998). Psychiatric illness and sexual function. *International Journal of Impotence Research*, **10**(Suppl. 2), S131–3.
40. Schreiner-Engel, P. and Schiavi, R.C. (1986). Lifetime psychopathology in individuals with low sexual desire. *Journal of Nervous and Mental Disease*, **174**, 646–51.
41. Ware, M.R., Emmanuel, N.P., Johnson, M.R., *et al.* (1996). Self-reported sexual dysfunctions in anxiety disorder patients. *Psychopharmacology Bulletin*, **32**, 530.
42. Angst, J. (1998). Sexual problems in healthy and depressed persons. *International Clinical Psychopharmacology*, **13**(Suppl. 6), S1–4.
43. Figueira, I., Possidente, E., Marques, C., *et al.* (2001). Sexual dysfunction: a neglected complication of panic disorder and social phobia. *Archives of Sexual Behavior*, **30**, 369–77.
44. Araujo, A.B., Durante, R., Feldman, H.A., *et al.* (1998). The relationship between depressive symptoms and male erectile dysfunction: cross-sectional results from the Massachusetts Male Aging Study. *Psychosomatic Medicine*, **60**, 458–65.
45. Bancroft, J., Janssen, E., Strong, D., *et al.* (2003). The relation between mood and sexuality in heterosexual men. *Archives of Sexual Behavior*, **32**, 217–30.
46. Lykins, A., Janssen, E., and Graham, C.A. (2006). The relationship between negative mood and sexuality in heterosexual college women. *The Journal of Sex Research*, **43**, 136–43.
47. Lilleleht, E. and Leiblum, S.R. (1993). Schizophrenia and sexuality: a critical review of the literature. *Annual Review of Sex Research*, **4**, 247–76.
48. Bancroft, J. (2005). The endocrinology of sexual arousal. Starling review. *Journal of Endocrinology*, **186**, 411–27.
49. Mustanski, B. and Bancroft, J. (2006). Sexual dysfunction. In *Psychopharmacogenetics* (eds P. Gorwood and M. Hamon), pp. 479–94. Kluwer Academic Plenum, New York.
50. Bochinski, D. and Brock, G.B. (2001). Medications affecting erectile function. In *Male sexual function: a guide to clinical management* (ed. J.J. Mulcahy), pp. 91–108. Humana, Totowa.
51. Graham, C.A., Bancroft, J., Greco, T., *et al.* (2007). Does oral contraceptive-induced reduction in free testosterone adversely affect the sexuality and mood of women? *Psychoneuroendocrinology*, **32**, 246–55.
52. Bancroft, J. *Human sexuality and its problems* (3rd edn). Elsevier, Oxford, in press.
53. Rosen, R.C. and McKenna, K.E. (2002). PDE-5 inhibition and sexual response: pharmacological and clinical outcomes. *Annual Review of Sex Research*, **13**, 36–88.
54. Von Keitz, A.T., Stroberg, P., Bukofzer, S., *et al.* (2002). A European multicentre study to evaluate the tolerability of apomorphine sublingual administered in a forced dose-escalation regimen in patients with erectile dysfunction. *British Journal of Urology International*, **89**, 409–15.
55. Heiman, J.R. and Meston, C.M. (1997). Empirically validated treatment for sexual dysfunction. *Annual Review of Sex Research*, **7**, 148–94.
56. Van Lankveld, J.D.M., ter Kuile, M.M., de Groot, H.E., *et al.* (2006). Cognitive-behavioral therapy for women with lifelong vaginismus: a randomized waiting-list controlled trial of efficacy. *Journal of Consulting and Clinical Psychology*, **74**, 168–78.
57. Bergeron, S., Binik, Y.M., Khalifé, S., *et al.* (2001). A randomized comparison of group cognitive-behavioral therapy, surface electromyographic biofeedback, and vestibulectomy in the treatment of dyspareunia resulting from vulvar vestibulitis. *Pain*, **91**, 297–306.

58. Trudel, G., Marchand, A., Ravart, M., *et al.* (2001). The effect of a cognitive-behavioral group treatment program on hypoactive sexual desire in women. *Sexual & Relationship Therapy*, **16**, 145–64.
59. Fugl-Meyer, A.R. and Fugl-Meyer, K.S. (1999). Sexual disabilities, problems and satisfaction in 18–74 year old Swedes. *Scandinavian Journal of Sexology*, **2**, 79–105.

4.11.3 The paraphilias

J. Paul Fedoroff

Clinical features

Definition of the condition

The characteristic essential to all paraphilias is the presence of a persistent and/or recurrent, sexually motivated interest that causes harm. This definition has several implications.

Interest versus act

Paraphilias can exist even if they have never been acted upon. By definition, all paraphilias begin with a sexual thought and, like non-paraphilic interests, the majority of sexual fantasies are never fulfilled in reality. Sexual acts are only paraphilic if they are motivated by harmful sexual interests. For example, an individual with paedophilia (sexual interest in children) may act on this interest by masturbating while viewing non-pornographic children's television shows. However, an individual who unintentionally downloads pictures of children from the Internet while meaning to download adult pornography should not be considered paedophilic (even though he or she may still be criminally liable).

Persistent versus opportunistic

Paraphilias are characterized by their persistence. Therefore a single paraphilic thought or activity, especially if it occurs during unusual circumstances, is unlikely to be indicative of a true paraphilia. For example, a woman who while on a vacation and under the influence of alcohol exposes herself once to group of strangers in a bar, would not normally be considered to have exhibitionism (sexual arousal from exposure to strangers). Opportunistic activity that is not sexually motivated, even if it is ongoing, is also exclusionary. A pimp who coerces women to exchange sexual activities for drugs or money, would not meet criteria for sexual sadism even though he repeatedly engages in opportunistic sexual activity with otherwise unwilling participants because, for the pimp, the motivation is financial rather that sexual.

Harm versus happenstance

Many sex acts are intimate. Therefore it should come as no surprise that participants can be harmed. Unfortunately whenever sexual activities are considered, 'harm' tends to be defined somewhat solipsistically. Paraphilias are characterized by sexual interests or behaviours in which harm is more or less inevitable rather than accidental or random. For example, although sexual intercourse may expose the participants to a number of dangers including sexually transmitted diseases, possibly unintended degrees of intimacy, or subsequent events (e.g. pregnancy), by and large, consensual sexual activity between adults does not inevitably lead to disaster.

In contrast, true paraphilic interests are by definition, harmful. For example, a woman who can only become sexually aroused if she is choked to unconsciousness (asphyxophilia), exposes herself to unintended harm including cerebral anoxia and possible death. The paraphilias associated with crime (e.g. voyeurism, exhibitionism, frotteurism, criminal sadism, and paedophilia) involve non-consensual imposition of sexual activity onto others. Other non-criminal paraphilias (e.g. transvestic fetishism) become problematic when they interfere with the ability to maintain a reciprocal emotional relationship. Most men with transvestic fetishism do not seek therapy and there is no indication they should, unless the interest begins to cause harm. Paraphilic transvestites are sufficiently dependent on cross-dressing that it causes distress. Transvestic fetishism is a good example of a paraphilic condition in which the individual symptoms (wearing women's clothing while masturbating or engaging in sexual relations) are not problematic. However, when the sexual interest becomes so pervasive that it interferes with consensual sexual relations, a diagnosis is permissible. Transvestic fetishism is very responsive to treatment with selective serotonergic re-uptake inhibitors prescribed at doses low enough to avoid inhibited orgasm.

Pathologic versus unconventional

While paraphilias are characterized by deviant sexual interests, unconventional interests alone are not sufficient to meet criteria for a true paraphilic condition. This is a persistent source of confusion in two specific situations.

(a) Homosexuality

A primary sexual interest in the same sex (homosexuality) is statistically rare. However, there is nothing about a primary same-sex interest that necessarily leads to harm to anyone. Sexual interest in a woman is no more harmful for a heterosexual man than it is for a lesbian woman. Although homosexuality is statistically and socially unconventional, the absence of inevitable or likely harm excludes same-sexed sexual interest from being paraphilic.

(b) Sadomasochism

Sexual arousal from control (sadism) or from being controlled (masochism) illustrates a second way in which unconventional sexual interests may fail to meet criteria for designation as a paraphilia. While harm is a necessary criterion for paraphilias, it is not sufficient. For example, while many conventional sports involve competition and attempted domination of an opponent, the activity is not primarily sexually motivated. Therefore, although boxing involves the intentional attempt to render an opponent unconscious via infliction of a cerebral concussion (knockout), pugilism is not a paraphilia because it is not primarily sexually motivated.

The converse is also true. Many men and women engage in interactions that are sexually motivated and which involve negotiations about power and control, domination and submission. These themes have become highly organized and regulated within social groups under the general category of 'BDSM' (bondage, domination, submission, sadism, masochism). Publications describing the BDSM 'lifestyle' universally champion the credo: 'safe, sane, and

consensual'. Therefore, men and women who engage in 'BDSM' sexual activities, although they may involve statistically and/or socially unconventional activities, do not meet criteria for any paraphilia disorder. In fact, it is arguable that anyone who is sexually aroused by the idea of engaging in 'safe, sane, and consensual' activities is *less* paraphilic than those with conventional sexual interests who are willing to compromise some of these meritorious criteria in pursuit of conventional sexual interactions. For a more complete discussion of sexual violence and sexual sadism.[1]

Classification

Table 4.11.3.1 consists of a partial list of the over 100 paraphilic disorders described in the literature. Classification of the paraphilias remains controversial. This is due to two issues. The first is that many paraphilias have been assigned names based on Latin or Greek etymology. For example, retifism refers to a paraphilic interest in shoes more commonly known as a 'shoe fetish' while 'renifleurs' are individuals with sexual arousal from the smell of urine. More complete listings of paraphilic disorders has been

Table 4.11.3.1 Paraphilic sexual disorders

Paraphilia	DSM- IV TR	ICD-10	Essential feature: persistent sexual arousal towards	Comments
Acrotomophilia	302.9	F65.9	Amputees	
Apotemnophilia	302.9	F65.9	Being an amputee	
Asphyxiophilia	302.9	F65.9	Being asphyxiated	Also known as 'autoerotic asphyxia'
Biastophilia	302.9	F65.9	Non-consensual adult intercourse	Also known as paraphilic rapism or raptophilia
Exhibitionism	302.4	F65.2	Exposure to strangers	
Fetishism	302.81	F65.0	Inanimate objects	Not vibrators
Frotteurism	302.89	F65.8	Rubbing groin without consent	ICD has no specific listing
Mysophilia	302.9	F65.9	'Filth'	Typically involving 'soiled' (worn) panties
Necrophilia	302.9	F65.9	Corpses	
Paedophilia	302.2	F65.4	Children	ICD does not differentiate
Attraction				
Males				
Females				
Both				
Exclusivity				
Incest only				
Exclusive				
Non-exclusive				
Polyembolokoilamania	302.9	F65.9	Insertion of objects	Associated with Smith McGinnis syndrome. (Not clearly paraphilic)
Scoptic syndrome	302.9	F65.9	Being castrated	
Scoptophilia	302.9	F65.9	Consensual viewing	Paraphilic only if problematic
Sexual masochism	302.83	F65.5	Loss of control	ICD combines into Sadomasochism
Sexual sadism	302.84	F65.5	Non-consensual control	ICD combines into sadomasochism
Somnophilia	302.9	F65.9	sleeping sexual partner	
Telephone scatalogia	302.9	F65.9	Obscene phone calls	
Transvestic fetishism with gender dysphoria	302.3	F65.1	Wearing clothes of the opposite sex	ICD does not subclassify
Urophilia	302.9	F65.9	Urine	
Voyeurism	302.82	F65.3	Spying	
Paraphilia NOS	302.9	F65.9	Paraphilias not otherwise specified	ICD: Disorder of sexual preference unspecified
Other disorders of sexual preference		F65.8	Other paraphilic disorders	

published (c.f. Love, 1999). In addition, many of the paraphilias involve sexual interest in the characteristics of the sexual partner(s). Often some interest in assuming the same characteristics is evident and receives a unique name. The most obvious example is sado-masochism which the DSM classification divides into sadism and masochism while in the ICD classification the two complimentary conditions are combined.

The second problematic diagnostic issue in the classification of the paraphilias concerns the need to describe both unconventional sexual behaviours and problematic sexual behaviours. For example, while sexual arousal from cross-dressing is unconventional, it technically does not meet criteria for transvestic fetishism unless it causes distress. In the case of paraphilias involving criminal interests (e.g. paedophilia) issues arise if a person reports persistent sexual interest in children but no distress or wish to act on the paedophilic interest. Newer diagnostic criteria likely will address this issue.

Diagnosis and differential diagnosis

Similar to most psychiatric conditions, paraphilic disorders are diagnosed primarily on the basis of self-reported symptoms. Paraphilias differ from most other psychiatric disorders in two ways: (i) many paraphilias involve illegal interests and (ii) objective measures of the primary criteria (in this case sexual interest) are available.

Illegal sexual interests

Paraphilic interests do not necessarily lead to illegal activities, and vice versa. Therefore, 'Not all paedophiles are child molesters and not all child molesters are paedophiles'.

With few exceptions, individuals with paraphilic interests not only know they have abnormal sexual interests but also wish they could replace them with 'normal' ones. Many confuse fantasy with reality, thinking that illegal sexual interests are equivalent to having committed a sex crime. A major issue in the diagnosis of paraphilic disorders is distinguishing between legal and psychiatric concerns (see Management section below for further details).

In addition to legal issues, clinicians should also consider several other diagnostic issues:

(a) False accusations

At one time accusations by children of sexual assault by adults were considered to be almost certainly true since it was assumed that children could not know about sex. A typical assertion would be that a child could not possibly describe acts such as sexual intercourse or ejaculation unless they had been sexually assaulted. Clearly this was before the widespread availability of pornographic videos, DVD's, cable TV networks, and the Internet.

Beginning in the 1990s, adults began to report they had been sexually abused as children but had only recently recovered their memories of the assault. In part this trend seems to have been due to two factors: the decision in the United States to reset the time at which the statute of limitations required a sexual offence to be reported to the time at which the offence was recalled; and the believe that failure to recall sexual abuse was a sign that it had occurred.

(b) False confessions

While less frequent, false reports of paraphilic interests have also been described.[2] The most frequent presentation of false paraphilic symptoms takes the form of a man or woman with depression who reports obsessions involving often exceptionally troubling sex crimes. While a detailed phenomenologic examination of this phenomenon has not been published, several characteristics are typical. The individual often has a history of a mood disorder or is in circumstances in which affective disorders are more likely (e.g. post-partum). The sexual obsession typically involves children to whom the patient has access (it is rare for the patient to report spontaneous fantasies of sexual interactions with unknown children). Most importantly, the fantasies are extremely ego-dystonic. Asked if they ever masturbate to their paraphilic fantasies, they typically respond with horror and, unlike those with true paraphilias, often describe self-loathing indicative of a change in self-esteem due to depression. A danger of false confessions or admission of false paraphilic interests has also been noted in men and women with intellectual disability.

(c) Co-morbid conditions

People rarely seek treatment on their own specifically for paraphilic disorders. This is due to unfortunately widespread false beliefs that (i) there are no effective treatments for paraphilias, (ii) embarrassment about discussing paraphilic symptoms, and (iii) in the case of paedophilia, the mistaken belief that reporting a sexual interest in children necessarily requires that the patient be reported to the police (please see Management below for more comments on this problem). Since paraphilias themselves rarely motivate helpseeking, clinicians should include other conditions in the differential diagnoses both as alternative explanations for the problem and as co-morbid conditions that may be present in addition to the paraphilia. The most frequent of these are mood disorders, substance abuse problems including alcohol, marital disorder, and legal problems.

Less common are organic disorders including brain injuries.[3] Surprisingly, given the importance of sex hormones in the development and expression of sexual characteristics, endocrine disorders affecting the sex hormones are rarely implicated. This may be because testosterone in men with normal or elevated hormone levels are more closely associated with violence and aggression than with alterations in the direction of sexual interest.[1] One exception is hypogonadism associated with Klinefelter's syndrome. In some men with Klinefelter's syndrome, paraphilic problems become apparent when testosterone is prescribed to correct hypogonadism. In those men with Klinefelter's syndrome and paraphilic interests, addition of testosterone appears to unmask rather than cause previously unexpressed paraphilic conditions.

A more controversial question involves a possible association between paraphilias and intellectual disability. Some research has supported the hypothesis that, as a group, men with paedophilia have below average intelligence. However, alternative explanations are possible. For example, most men with paedophilia come to attention when they are arrested. The fact that men with intellectual disability are more likely to be arrested and are over-represented in the criminal justice system may skew the data. Those who live in supervised housing are also more likely to have private activities not only discovered by staff but also labelled by staff as deviant. In addition, men with intellectual disability frequently have impairments in social skills that can certainly contribute to problematic sexual behaviours independent of paraphilic interests. This phenomenon has been described as 'counterfeit deviance'.[4]

One meta-analytic study supporting an association between paedophilia and intellectual disability found the mean I.Q. of sex offenders to be only five I.Q. points below the mean I.Q. of non-sexual offenders.[5] While cognitive ability is important in determining level of risk and in planning treatment (see below) it would be a diagnostic mistake to confirm or refute a diagnosis of paedophilia on the basis of intelligence.

A further area of controversy involves the question of whether having one paraphilia predisposes to having other paraphilias? The answer depends on whether the assessor is a 'lumper' or a 'splitter'. John Money viewed paraphilias as 'vandalized lovemaps' that were unique. He argued that a paedophile might begin by spying on children, then surreptitiously touching them, then engaging in sexual relations. It was his view that it was more accurate to label the disorder as paedophilia (since this explains the motivation behind the varied behaviours) as opposed to making a diagnosis of voyeurism, frotteurism, and paedophilia. Clearly both approaches have strengths and weaknesses. Most important is to be aware of both diagnostic methods when evaluating incidence or prevalence reports.

Epidemiology

Prevalence of paraphilic disorders

Any discussion of the number of people with paraphilic disorders in the population must begin with a series of caveats. The most important is the fact that the majority of the information available is derived from studies of men convicted of sex crimes. This is a significant problem since not all paraphilias are associated with sex crimes. The problem is compounded by the fact that many paraphilic disorders remain undiagnosed either because assessment was never requested, clinicians did not gather sufficient information, or because the person with the paraphilic disorder was unwilling or unable to disclose the symptoms. Further degrees of confusion and subsequent dispute are added by inconsistent application of existing diagnostic criteria (for example confusing child molesters with paedophiles) or by differing opinions about whether or not to subdivide paraphilic disorders (e.g. diagnosing a person who lures children into sexual activity by exposing to them as a paedophile or as both a paedophile and an exhibitionist). To date, insufficient attention has been paid to the importance of precise definitions of what is being measured, the difference between point and period prevalence, and the potentially significant differences attributable to independent characteristics of the populations being studied. For example, a report on the incidence of 'sexual sadism' based on a point prevalence study of sexual assault of women and children in a country in which war is being waged, while important, has little to do with an analysis of the period prevalence of sexual sadism in, for example, North America. There is also an widespread but unwarranted assumption that studies based on criminal populations can be easily generalized to other populations.

Fortunately, with these above caveats in mind, a significant resource to answer epidemiologic questions is now available both in text form and, more importantly, in the form of a constantly updated Internet Web site: http://www.kinseyinstitute.org/ccies/. This Web site, currently under development and hosted on the Kinsey Institute web page, provides free access to reports by noted sexologists on sexual behaviours in 60 countries and six continents. Included in the chapters on most countries are sections dealing with both paraphilic disorders and 'unconventional sexual behaviours'. This is important since different cultures may place different significance on sexual behaviours both in public and private. The fact that each chapter is multi-authored by sexologists who live in the country being described also likely adds to the credibility of the information.

From perusal of the information available several statements about the epidemiology of paraphilic disorders can be made with reasonable confidence.

(a) Men are more likely than women to be arrested for sexually motivated crimes

Most (but not all) illegal activities are non-consensual. Those that are sexually motivated have a high likelihood of being associated with a paraphilia. Data to support this observation is derived primarily from review of the unequal ratio of men and women convicted of sex crimes. While there are increasing numbers of women (especially adolescent females) being convicted of child sexual abuse, men are still the vast majority. From a criminologic perspective men seem more willing to commit crimes of all types, especially those that involve confrontation (verbal, physical, and sexual). This is of significance since two widely employed actuarial risk assessment instruments, the Violence Risk Appraisal Guide (VRAG) and Sex Offender Risk Appraisal Guide (SORAG) rely heavily on the Hare Psychopathy Checklist (HPCL) which itself has been associated with criminal activity independent of sexual interest. This may explain why some researchers believe that a combination of high scores on the HPCL combined with deviant scores on phallometric testing are associated with high risk of re-offence (either violent and sexual) (for more information please see the Hare Web site listed below).

(b) With the exception of internet-related sex crimes, the frequency of sex crimes of all types are on the decline

This welcome finding has been reviewed extensively elsewhere.[6] It is important to note that while this trend of decreasing frequencies of sex crimes excludes countries in the midst of war or major political and social upheaval, the trend is occurring worldwide. Unfortunately the reason for this trend is unknown. One controversial explanation is a wider availability of pornographic materials, beginning with videotapes in the 1980s and now exploding with the Internet. One study noted, during a time of increased availability of pornographic materials, an 85 per cent decrease in numbers of juvenile offenders in Japan from 1803 in 1972 to 264 in 1995.[7] The increase in Internet-related crimes combined with an apparent decrease in other types of sex crimes invites the unproven speculation that one may be substituting for the other.

(c) Paraphilic disorders are not new

A final important point to be made about the epidemiology of paraphilic disorders is that they are not new. Historic texts on the topic clearly describe not only sex crimes but also the same phenomenologic characteristics seen by clinicians today. If there is a change, it is towards increasingly earlier detection of paraphilic behaviours, not only at a younger age,[8] but also before criminal acts have been committed.

Table 4.11.3.2 summarizes generally accepted prevalence estimates for the DSM-IVTR paraphilias.

Table 4.11.3.2 Frequency/prevalance of paraphilias

Paraphilia	Prevalence/frequency	Comments
Exhibitionism	No reliable data	40–60% of females report having been exposed to Evidence suggests a decrease in frequency of exhibitionism after age 40
Fetishism	No reliable data	Depending on population studied, rates range from 0.8% to 18% of men
Frotteurism	No reliable general population data	Only paraphilia claimed to decrease in frequency after age 25 (DSM-IV-TR p. 570)
Paedophilia	Frequency: 300 000 abused children per year in the USA	1988 data; does not account for repeat offences
Sexual sadism	3% to 20% of general population	Data do not distinguish between criminal and non-criminal sexual interests
Sexual masochism	5% to 10% of general population	Data does not distinguish between criminal and non-criminal sexual interests
Transvestic fetishism	1% of men	True prevalence unknown in part due to idiosyncratic diagnostic criteria
Voyeurism	No reliable data	Most studies of voyeurism involve offenders with co-morbid paraphilic interests

Aetiology

Like most psychiatric disorders, the ultimate causes of paraphilic disorders are unknown. Like most groups of psychiatric disorders, they undoubtedly have multiple contributing factors.

Four major explanatory perspectives have been identified: 'disease', 'behavioural', 'dimensional', and 'life story'.[9] Each perspective has had advocates that can be traced to the beginning of the twentieth century.

The disease perspective

The disease perspective is perhaps the most familiar to physicians since it is based upon an attempt to combine signs and symptoms into syndromes or disorders via pathophysiologic mechanisms.

In 1886 Krafft-Ebing published *Psychopathia Sexualis: A Medico-forensic Study* in which he described a large series of patients with a variety of sexual disorders, whom he had examined in his forensic psychiatry practice.[10] Using a disease perspective he advanced the hypothesis that masturbation caused a physiologic imbalance that was a major causal factor in development of sexual deviancy. Krafft-Ebing was widely criticized for his acceptance and perpetuation of the degeneration theory of masturbation (the now discarded theory that masturbation itself could cause progressive degrees of physiologic harm leading to increasingly severe psychopathology), and for his failure to clearly differentiate between extreme and mild forms of sexual deviancy.

While *Psychopathia Sexualis* is now considered a landmark text in human sexuality, its acceptance was limited because it was published at a time when other etiologic explanations for the paraphilias were becoming important. One criticism of Krafft-Ebing's disease perspective was his tendency to focus on illness rather than wellness or normality. In this context, Brautigam's criticism of Krafft-Ebing's *Psychopathia Sexualis* foreshadows current criticisms of etiologic theories of the paraphilias that rely on the disease perspective:

> This first great and distinguished complete presentation and inventory of that field has the character of a large catalogue of perversions. In a picture book-like series are offered the most monstrous cases of his time and history, particularly cases collected by Krafft-Ebing himself as a widely sought medico-legal consultant. The perversions are described in their most extreme forms. This collection of brutal necrophiliacs, anthropophages, sexual murderers,and sodomites, of cunning and cultured sadists, masochists, and transvestites has in its degree of deviation never been surpassed. By this extreme presentation Krafft-Ebing has moved sexual disorders away from general sexuality.[11]

Magnus Hirschfield was a contemporary of Krafft-Ebing who also employed a disease perspective in an attempt to identify abnormal pathophysiology in individuals with identified sexual problems. Although he is best known for his studies of castrated males and for his emphasis on the importance of hormones on sexual behaviour, Hirschfield also drew an important distinction between the simple description of observed behaviours and proof of their aetiology. His argument that describing a behaviour was not sufficient to explain its cause led him to distinguish between homosexuality, transsexualism, and fetishistic transvestism, in which the same behaviour (cross-dressing) could be shown to have different motivations and therefore presumably different causes.[12] His observation that sexual orientation and behaviour may have variable determinants was important since it emphasized the need to look for underlying causes of behaviour from a disease perspective while at the same time opening the door for collaboration with behavioural researchers.

Behavioural perspective

Although Albert Moll felt that sexual disorders could be treated with 'association therapy' (a precursor to modern behaviour modification therapy), he felt learning alone could not account for all sexual deviations:

> Were a single sexual experience, and, indeed, the first sexual experience to induce a lasting association between sex drive and the object of the first sexual experience, then we would have to find sexual perversion everywhere. Where are there to be found people who initially satisfied their sexual impulse in a normal manner?[13]

By extending consideration of sexual development from childhood to adulthood, Moll was able to answer Brautigam's criticism of Krafft-Ebing's research by studying normal as well as abnormal sexuality.

Albert Moll and Alfred Binet, used behavioural perspectives to study sexuality by directing attention to the associations between behaviours and their reinforcers. Binet re-examined the case histories of Krafft-Ebing's patients and argued that sexual deviations which had been presumed to be caused by disease or degeneration could have been caused by learned associations. He noted that many of the reported cases he studied also had 'chance associations' that could explain the development of the sexual deviation. These observations implied that, given a specific set of reinforced

experiences, any individual could become paraphilic. The proposition that all people could potentially acquire paraphilias if they were unfortunate enough to have been exposed to the necessary abnormal formative experiences was more fully developed by investigators using a dimensional perspective.

Dimensional perspective

The dimensional perspective is characterized by an avoidance of rigid categorical classifications in favour of continuous descriptors. Variations in sexual interests, including paraphilic interests, are explained as statistical extremes of normal behaviour without necessarily implying that a disease process is present. From a dimensional perspective, paedophilia would be understood as an extreme variation of the widespread tendency to equate youth with sexual attractiveness. The work of Haveloch Ellis is a good example of this approach to sex research. He is most well known for his six volume *Studies in The Psychology of Sex*.[14] A major theme in Ellis' work is the avoidance of a disease perspective as the sole explanation for sexual deviations. Instead of paraphilic disorders, he focused primarily on homosexuality and suggested that homosexuality could be 'latent'. Since degree of 'latency' is a continuous variable it was possible to consider homosexuality as a dimension rather than a disease. This dimensional perspective was later exploited by Kinsey[15] who developed his famous Kinsey scale of sexual orientation that rated men and women on a six point scale. Both Ellis and Kinsey shifted attention from the differences between homosexuals and heterosexuals to their similarities. In this way their work complemented research using disease or behavioural perspectives as well as the work of investigators using a fourth perspective, which focused primarily on the individual's 'life story'.

Life-story perspective

The life-story perspective is characterized by a search for the 'meaning' behind behaviour. This perspective appeals to the self-experienced aspect of life, so life-story perspectives tend to be very compelling and persuasive. For example, in North America, it is rare to find a man with transvestic fetishism that does not recall being dressed as a girl for Halloween. A problem with this approach is that it is often difficult to determine which 'story' is the correct one, or even if a single meaningful connection can be found. For example, there is no evidence that men with transvestic fetishism are more likely to have been dressed as girls when they were children than were other men. Sigmund Freud often used sexual motives to construct a life story that was difficult to disprove. Freud's research method consisted chiefly of interviewing patients and interpreting recurring themes in literature and art. While this approach may seem highly susceptible to bias, it has been surprisingly successful, not only for Freud but for other researchers who have adopted this method of direct interview combined with longitudinal follow-up, to understand the meanings of sexual acts and decisions. The explanations that individuals themselves develop to explain their behaviours are often as important and enlightening therapeutically as their 'true' causes.

Integration of the four perspectives

While each of the four described perspectives provide a unique basis on which to base explanatory theories about the origins of paraphilic disorders, none is mutually exclusive. In the twentieth century, some behaviourists began to acknowledge the importance of non-observable factors such as 'instinct'[16] and later, 'motivation'. Other theorists,[17] while recognizing the importance of behaviourism, emphasized that the 'motivated behaviours' of sleeping, eating, and mating could not reasonably be reduced to simple stimulus response associations. If motivated behaviours could not be fully explained from a single research perspective perhaps an integration of approaches would be more helpful. A good example of an integrated etiologic theory to explain paraphilias is the one advanced by John Money who theorized that all children are born with the potential to develop a full range of sexual interests analogous to the observation that all healthy children at the time of birth are capable of learning to speak any human language. Which language (and by analogy which sexual interests) each child eventually expresses is determined by learning during critical periods of neurologic development.

From extensive research involving patients with genetic and endocrine disorders that interfered with normal anatomic and physiologic sexual development, he theorized that sometime before the onset of puberty, a 'lovemap' was established. He defined lovemap as: 'a developmental representation or template, synchronously functional in the mind and the brain, depicting the idealized lover, the idealized love affair, and the idealized programme of sexuoerotic activity with that lover, projected in imagery and ideation, or in actual performance'.[18] He hypothesized that an individual's lovemap became 'neurologically embedded' or 'imprinted', analogous to the way an individual's native language becomes a 'native tongue'. He believed that each person has a single lovemap (analogous to having a single native language or 'mother tongue') but that each person may never be fully aware of the true nature and details of the lovemap (analogous to not learning the full vocabulary of grammar of a native language). Paraphilias represented what he termed 'vandalized' lovemaps resulting from an abnormal experience during the 'critical period' during which lovemaps are formed.

Course and prognosis

While the incidence of paraphilias (especially those that involve criminal activities) in men is at least 10 times the incidence in women,[19] the majority of both males and females do not develop paraphilias. In addition, unlike sexual orientation, for which there is some evidence of important genetic or at least prenatal influences,[20] and excluding a single retrospective report of possible familial co-morbidity,[21] there is no current evidence of any prenatal factors that cause paraphilic disorders. Similarly the hypothesis that being the victim of sexual abuse predisposes victims to become sex offenders has also been challenged.[22] There is an increasing recognition of the fact that adolescents are capable of committing sex crimes.[8] This edited book concluded that, 'approximately 20 per cent of all rapes and between 30 per cent and 50 per cent of child molestations are perpetrated by adolescent males'. The increasing number of facilities designed to assess and treat juvenile sex offenders, male and female, supports self-report descriptions of paraphilic behaviours beginning in adolescence.

Most crimes (both sexual and non-sexual) are committed by young adult males. One problem in assessing the literature on the frequency of paraphilic behaviours in this age group is the frequent failure to distinguish between current paraphilic acts

and those that are recalled. For example, it is not uncommon for a middle-aged man with exhibitionism to recall frequent exhibitionistic activities from his youth[2] but there is no reason to believe that the recollections of the sexual exploits of exhibitionists are any more reliable than the recollections of non-paraphilic men.

To summarize, it is often assumed that anyone who has a paraphilic interest has acted on it. In fact, the DSM-IV diagnostic criteria B requires that the interest has been acted upon or at least caused harm to someone. While cases have been described in which a person engaged in sexual activity without awareness (e.g. engaging in sexual acts while asleep),[23] for the most part, thought precedes action. In fact, it is arguable that a person who fantasizes about paraphilic activities is more paraphilic than one who commits a paraphilic act while fantasizing about a non-paraphilic activity.

Most paraphilic thoughts begin around the time of puberty. If the 'vandalized lovemap' explanation of paraphilias is correct, faulty development of sexual interest likely occurs prior to puberty but is not expressed until there is sufficient increase in sex drive for it to be manifested, first as fantasy and later, by action.

Prognosis is an area of considerable controversy. However, accumulated evidence indicates that the majority of convicted sex offenders do not re-offend (likely fewer that 15 per cent). In addition, evidence indicates that treatments are improving in efficacy (see below).

Evaluation of treatments

In this section 'treatment' refers to intervention with intent to reduce the frequency of unwanted paraphilic symptoms. Studies on the treatment of paraphilic disorders necessarily include participants with diagnosed paraphilias who have not only sought treatment but who have also volunteered for a treatment and have at least started the treatment under study. This represents a group that for obvious reasons may be unrepresentative of the true paraphilic population, the majority of whom likely never seek treatment.

Treatment recommendations should be individualized across paraphilic and co-morbid disorders and to account for individual variation. For example, the treatment approach for an exhibitionist with anti-social personality disorder and heroin dependency will be very different from the treatment of a woman with paedophilia or an adolescent with intellectual disability and transvestic fetishism.

In the past half-century there have been three major innovations in the treatment of paraphilic disorders that have led to significant improvement in prognosis. The first of these was the use of anti-androgen medications.

Interventions to reduce testosterone

Surgical castration has shown a high degree of success, at least in terms of reducing recidivism rates in sex offenders, with reported rates of recidivism after 30 years of follow-up as low as 2.2 per cent for convicted rapists. While this is a remarkable success rate, surgery for this purpose has been criticized on ethical grounds and because reversible interventions, in the form of anti-androgen medications have became available. For a review of the efficacy of castration in the treatment of the sex offender.[24]

Anti-androgens administered either by mouth or by intra-muscular injection were first used in the late 1950s as an alternative to surgical castration. The most common oral anti-androgens prescribed for this purpose are cyproterone acetate (Androcur) (CPA) and medroy-progesterone acetate (Provera) (MPA). Both of these medications are also available in a formulation suitable for intra-muscular injection. Reviews of treatment efficacy of both of these medications are also available.[24]

A review of eight studies involving a total of 452 sex offenders who received treatment with MPA for between 1 and 13 years, recidivism ranged from 1 per cent to 17 per cent.[25] Unfortunately, anti-androgen treatment of sex offenders has been a victim of its own success. Because the success rates were so high in open studies, and because the consequences of recidivism are so great, randomized double-blind treatment studies have been difficult to justify from an ethical perspective. Gonadotropin hormone releasing hormone (GnRH) partial analogs significantly reduce serum testosterone by down-regulating receptors in the anterior pituitary resulting in a decrease in luteinizing hormone which in turn decreases production of testosterone by leydig cells in the testes. Reviews of the apparent impressive efficacy of anti-androgens and GnRH analogs are available.[26]

Selective serotonin re-uptake inhibitors (SSRIs)

Numerous reports have confirmed the apparent efficacy of SSRI and related medications in the treatment of men with paraphilic disorders.[27] Two important prescribing issues are important:

(a) There is no evidence that the efficacy of SSRIs in the treatment of paraphilic disorders is due to suppression of sex drive. Many medications that are more effective in reducing sex drive appear to have little effect on paraphilic interest.

(b) There is no evidence that higher doses of SSRIs are more effective than low doses. In fact, if the prescribed SSRI causes inhibited orgasm, this may be counter-therapeutic since there is a tendency for individuals with paraphilic sexual interests to resort to paraphilic fantasies or acts in order to facilitate orgasm.[28,29]

Psychotherapy

The evidence in support and against various psychotherapeutic interventions has recently been reviewed.[30] Most reviews of psychotherapeutic interventions for paraphilic disorders focus on group therapy (the most common intervention in prisons) or individual psychotherapy (most common in private practices). However, other interventions are also important including marital therapy, family therapy, spousal therapy, occupational therapy, and substance abuse therapy.

Treatment of sex offenders based on psychologic interventions were carefully reviewed by a collaborative group using meta-analytic techniques.[31] A total of 43 published studies involving a total of 9454 sex offenders (5078 treated and 4376 untreated) were analysed. While noting methodologic limitations, there were two important findings: (i) sexual recidivism rates were lower in treated groups (12.3 per cent) than in comparison groups (16.8 per cent) and (ii) efficacy of psychologic treatments appear to have improved since the 1970s. The final results of one important study were not included in the Hanson *et al.* meta-analysis. This was a longitudinal study with stratified randomization into a group that received

relapse-prevention group therapy ($n = 167$) or a group that did not ($n = 225$); a third group was matched to the first two groups and consisted of sex offenders who did not volunteer to participate in the study ($n = 220$). Treatment for the first group consisted of 2 years in custody, group relapse-prevention therapy followed by a 1 year mandated programme after release to the community. While results were complicated by a dropout rate in the treatment group of 27 per cent, the third group consisting of 'non-volunteer' controls was found to have the lowest long-term rate of sexual recidivism (19.1 per cent at 12 years follow-up). In comparison, the sub-group that entered therapy but discontinued prematurely had a 35.7 per cent sexual re-offence rate compared to the group that completed the prescribed treatment which had a 21.6 per cent sexual re-offence rate at 12 years follow-up. The final conclusion of this study was that a treatment effect for the cognitive behavioural programme did not produce a significant treatment effect.[32]

The failure to demonstrate the efficacy of 'relapse-prevention' as it was constituted 20 years ago has been regarded by some as a major setback in the field. However, the fact that even in the worst outcome group the majority of sex offenders were not known to have re-offended, the fact that the treatment programme under investigation was 'manualized', and the fact that other important treatment interventions were not employed, makes any conclusion that sex offenders can not be treated unwarranted.

Management

General principles

Summaries of management strategies for individuals with paraphilic disorders and/or sex offence histories are available.[30,33] Specific recommendations for the more commonly encountered paraphilic disorders and for specific paraphilic disorders should also be consulted.[2] An initial approach to the management of paraphilic disorders is summarized in Fig. 4.11.3.1.

Individuals with paraphilic disorders typically are in extreme crisis when they present for treatment. It is important to assist the individual to not only understand that sexual urges are controllable but that sex acts are the person's responsibility. This is often a surprise to people with paraphilias who have lived with the false belief that the difference between themselves and the rest of the world is a lack of will power. It is important to assist in maintaining and establishing a healthy social support system, ideally including family members, spouses, employers, church members, and trusted friends. Group therapy is particularly helpful in assisting to confront cognitive distortions and in demonstrating that others not only do not re-offend but are capable of establishing healthy and fulfilling lives. Medications can be prescribed with the aim of decreasing sexual anxiety and impulsivity (SSRIs), moderating sex drive (anti-androgens), or eliminating sex drive (GNRH

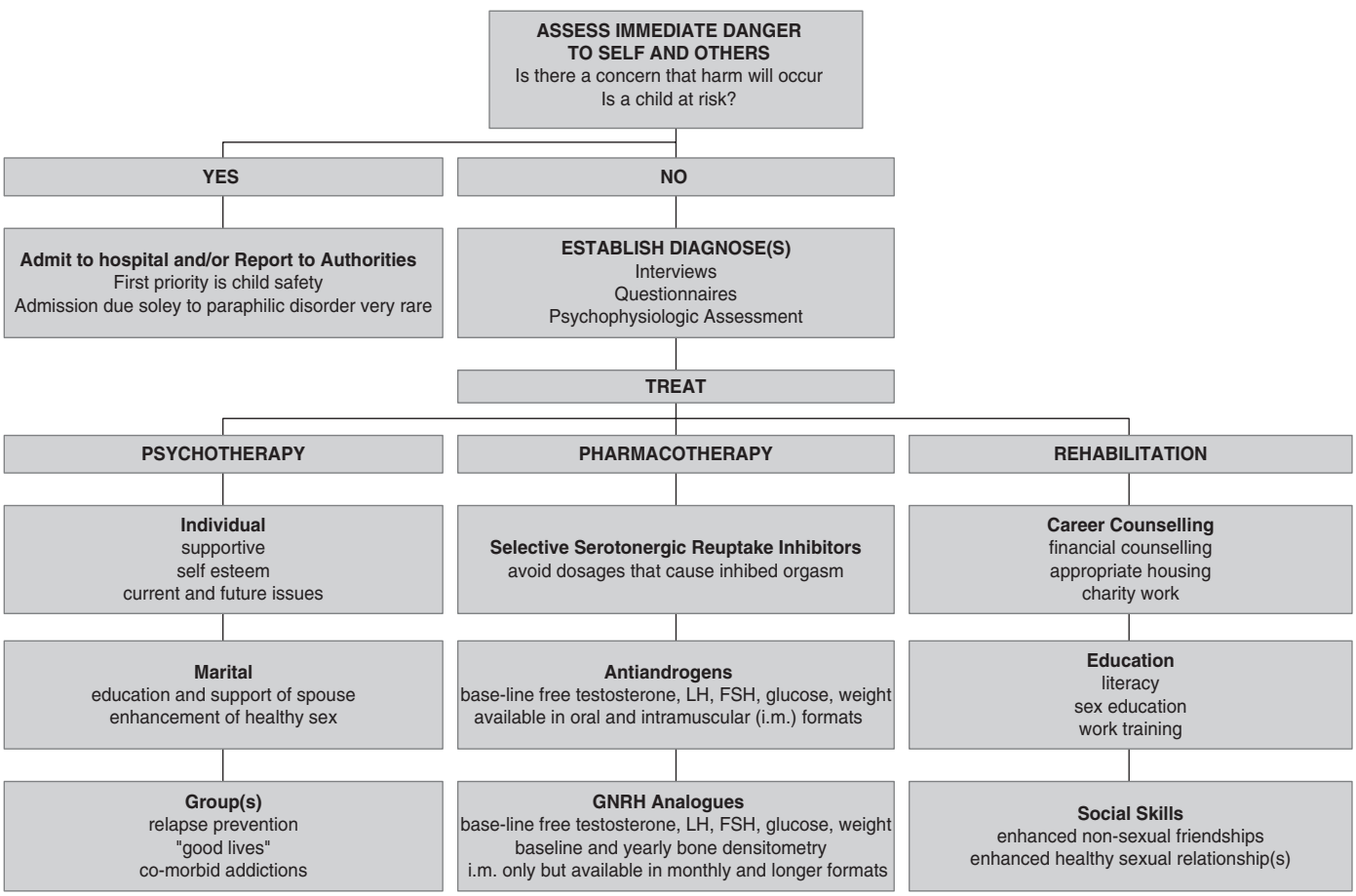


Fig. 4.11.3.1 Management of the paraphilias.

analogues). It is vital, particularly when treating men with paraphilias who have been incarcerated, to assist them in establishing pro-social lifestyles. This includes advice about appropriate work (e.g. avoiding jobs that require contact with children in the case of paedophiles), finding work (some clinics maintain lists of employers willing to hire men they know are in treatment), and ways to spend leisure time in constructive and pro-social ways. A part of the therapy often neglected is the establishment of healthy sexual activities to replace the problematic sexual interests and acts for which therapy was initiated. This frequently involves marital or couple's therapy. Some clinics establish 'spouses' groups to provide independent support, education, and treatment of the spouses of individuals with paraphilias.

In addition, the following general principles are recommended:

(a) Identify for whom you are working

Many paraphilic disorders predispose to unconventional or criminal activities. Most countries now have mandatory reporting requirements for incidents of known or potential sexual abuse of any identifiable child. It is extremely important to establish the limits of confidentiality prior to or at the first meeting. Many clinics post a description of the rules of confidentiality under which they operate. Sometimes it is difficult to be certain for whom one is working. A quick rule of thumb is to ask who is paying for the work and to whom reports will be sent.

(b) Facilitate disclosure

Men and women seeking assessment and treatment for paraphilic disorders almost always think they are the first (and worst) case of their type ever seen. They fear they will be led from the clinic room in handcuffs and be featured on the front page of the following day's newspaper. It is important to dispel these misconceptions not only because it will allow for a more in-depth assessment but also because it is important to dispel a 'me-against-authority' mentality which is common amongst individuals who see themselves as 'outsiders'.

Objective measures of sexual preference (e.g. phallometric testing) are available.[34] Their utility is enhanced if they are used as 'truth facilitators' rather than as 'lie detectors'.[35] While tests based on deception such as surreptitious measurement of time spent viewing pictures are currently widely used, their future utility is questionable since they are valid only if the deception is effective.

Perhaps the greatest truth facilitator is scrupulous adherence to disclosed rules of confidentiality.

(c) Establish what the problem is

Typically, individuals with paraphilic disorders present in crisis. However, the crisis often has only peripheral connection to the paraphilia. For example, a woman may have sexual masochism but seek treatment due to complications arising from major depression or substance abuse. This is true even if she insists that all her problems would end if she could be cured of her masochistic interests. As reviewed above, paraphilic disorders often occur together with other psychiatric, medical, and relationship problems. Clinicians should be wary of limited differential diagnoses, especially those that attribute all problems to a single paraphilia in isolation.

(d) Identify why the patient is seeking treatment now

Most paraphilias begin in childhood or at the time of puberty. However, teenagers rarely self-present for assessment of paraphilic disorders. Therefore, most adults with paraphilic disorders may be assumed to be seeking treatment for a variety of reasons besides having suddenly discovered they have (for example) paedophilia. Frequent motivations include: criminal charges, discovery of problematic behaviours by spouses or family members, and emergence of co-morbid problems such as depression or substance abuse.

For treatment to be successful it is very important to establish what the person seeking treatment thinks is the problem. Failure to do so significantly reduces the likelihood of keeping the person in therapy, establishing a therapeutic relationship, and ultimately the prognosis for a successful outcome.

(e) Avoid one-sided treatment plans

Paraphilic disorders by definition involve an interaction between the most abstract and 'high-level' cognitive processes (sexual fantasies), limbic and sub-cortical neural pathways (sexual arousal patterns), and society. Treatment plans based on single perspectives are unlikely to succeed. Most importantly, it is vital to pay attention to the worldview of the person seeking treatment. While the theory they have may be incorrect, ignoring it is likely to lead to frustration.

(f) Intervene quickly

A common mistake is to assume that because a paraphilia has been present for a long time, that it will take a long time to alleviate the symptoms. Many individuals with paraphilias think they must first review their childhood and all their previous offences before they can think about altering their current problematic behaviours. This is false. On the basis of current evidence, treatment should focus on current and future behaviours with an expectation that sexual behaviours of all types are under voluntary control. Requests to 'taper' problematic sexual behaviours should be dealt with as intentional or unintentional efforts to delay effective intervention.

(g) Be persistent

Successful treatment is not only multi-faceted but above all persistent. A useful strategy is to present treatment as a series of 'experiments' based on feedback about the effectiveness of interventions. Treatment should be presented with the expectation of success. A review of the fact that the majority of outcome studies find that the likelihood of sex offenders to recidivate is less that the likelihood of patients with major depression to have a recurrence of their life-threatening illness, is often an inspiring surprise to people entering therapy.

(h) Be inclusive with treatment

Begin with a thorough assessment and consideration of a complete differential diagnosis. If there are indications of co-morbid problems (e.g. foetal alcohol syndrome, genetic disorders, mood disorders, psychotic disorders, personality disorders, etc.) further investigations and interventions should proceed without delay. Once a diagnosis has been made, all the treatment options should be reviewed with the person seeking treatment (see above). Risks and benefits of treatment options (including the risk of declining treatment) should be thoroughly reviewed. Choice of treatment should be a collaboration between the clinician and person seeking treatment, but ultimately should be informed and voluntary. Typically treatment should include both individual and group psychotherapy. If the person has a sexual partner, couple's therapy is often of assistance. Pharmacologic treatment can include SSRIs

(often more effective at doses used for depression as opposed to the higher doses used for obsessive–compulsive disorder), anti-androgens and/or GnRH analogues. Acceptance of treatment with anti-androgens is increased by explaining that treatment effects are reversible. More information about treatment options is available.[2]

(i) Be optimistic

The frequency of sex offences is dropping. The effectiveness of treatment of sex offenders (who are not necessarily paraphilic) is improving. Sex offender re-offence rates are below 15 per cent. New as well as established treatments are available.

Beyond eliminating criminal and harmful sexual behaviours, a key to successful therapy is the establishment of non-criminal and beneficial sexual interests and behaviours. Remarkably, the literature on treatment of paraphilic disorders almost completely ignores the concept of establishing non-paraphilic sexual behaviours as a treatment goal. This is a mistake. Adult men and women rarely want to 'give up' sex. However, if they are provided with the opportunity to improve their sex lives they frequently become enthusiastic (and successful) participants.

Acknowledgements

Dr Fedoroff gratefully acknowledges Drs J. Bancroft, F. Berlin, J. Bradford, E. Coleman, P. Fagin, R. Rosen, and the late J. Money for their significant influences on his thinking about the paraphilias. However, none should be held responsible for any of the opinions expressed in this chapter.

Further information

Web sites

Robert Hare Psychopathy Web site: http://www.hare.org/

Magnus Hirschfeld Archive for Sexology: http://www2.hu-berlin.de/sexology/

Kinsey institute: http://www.indiana.edu/~kinsey/

Useful summaries

Bancroft, J. (in press). *Human sexuality and its problems*. Elsevier Health Sciences In Press.

Love, B. (1999). *Encyclopedia of unusual sex practices*. Greenwich Editions, London.

Money, J. (1986). *Venuses penuses*. Prometheus Books, Buffalo, New York.

Francoeur, R.T. and Noonan, R.J. (eds.) (2004). *The continuum complete international encyclopaedia of sexuality*. Continuum, New York.

Schwartz, B.K. (ed.) (1995–2005). *The sex offender* (Vols. 1–5). Kingston, New Jersey.

References

1. Fedoroff, J.P. (in press). Sadism, sadomasochism, sex and violence. *Canadian Journal of Psychiatry*.
2. Fedoroff, J.P. (2003). Paraphilic worlds. In *Handbook of clinical sexuality for metal health professionals* (ed. S.B. Levine). Brunner/Routeledge, New York.
3. Coleman, E. (2005). Neuroanatomical and neurotransmitter dysfunction and compulsive sexual behavior. In *Biological substrates of human sexuality* (ed. J.S. Hyde), pp. 147–69. American Psychological Association, Washington, DC.
4. Hingsburger, D., Griffiths, D., and Quinsey, V. (1991). Detecting counterfeit deviance: differentiating sexual deviance from sexual inappropriateness. *Habilitative Mental Healthcare Newsletter*, 51–4.
5. Cantor, J.M., Blanchard, R., Robichaud, L.K., *et al.* (2005). Quantitative reanalysis of aggregate data on IQ in sexual offenders. *Psychological Bulletin*, **131**, 555–68.
6. Lalumière, M., Harris, G.T., Quinsey, V., *et al.* (2005). *The causes of rape*. American Psychological Press, Washington, DC.
7. Diamond, M. and Uchiyama, A. (1999). Pornography, rape and sex crimes in Japan. *International Journal of Law and Psychiatry*, **22**, 1–22.
8. Barbaree, H. and Marshall, W.L., (eds.) (2005). *The juvenile sex offender* (2nd edn). Guilford Press, New York.
9. McHugh, P.R. and Slavney, P.R. (1983). *The perspectives of psychiatry*. Johns Hopkins University Press, Baltimore.
10. Krafft-Ebing, R.V. (1906). *Psychopathia sexualis. A medico-forensic study, with especial reference to the antipathic sexual instinct*. Physicians and Surgeons Book Company, New York.
11. Brautigam, W. (1977). Die angliche Beurtelung in Psychopathologie der Sexualitae. In (eds. H. Grese and F. Stutgart).
12. Hirshfield, M. (1910). Die Transvestiten. A. Pulvermacher, Berlin.
13. Moll, A. (1983). Untersuchungen uber die Libido Sexualis. In *Freud, biologist of the mind* (ed. F.J. Sulloway), p. 304. Basic Books, New York.
14. Ellis, H. (1936). *Studies in the psychology of sex*. New York.
15. Kinsey, A.G., Pomeroy, W.B., and Martin, C.E. (1948). *Sexual behavior in the human male*. Saunders, Philadelphia.
16. Darwin, C.R. (1859). *On the origin of species by means of natural selection, or, the preservation of favored races in the struggle for life*. John Murray, London.
17. Lashley, K.S. (1938). The experimental analysis of instinctive behavior. *Psychology Review*, **45**, 445.
18. Money, J. and Lamacz, M. (1989). *Vandalized lovemaps: paraphilic outcome of seven cases in pediatric sexology*. Prometheus Books, Amherst, NY.
19. Fedoroff, J.P., Fischell, A., and Fedoroff, B. (1999). A case series of women evaluated for paraphilic sexual disorders. *The Canadian Journal of Human Sexuality*, **8**, 127–40.
20. Blanchard, R. and Bogaert, A.F. (1996). Homosexuality in men and number of older brothers. *The American Journal of Psychiatry*, **153**, 27–31.
21. Gaffney, G.R., Lurie, S.F., and Berlin, F.S. (1984). Is there familial transmission of pedophilia? *The Journal of Nervous and Mental Disease*, **172**, 546–8.
22. Fedoroff, J.P. and Pinkus, S. (1996). The genesis of pedophilia: testing the abuse to abuser hypothesis. *Journal of Offender Rehabilitation*, **23**, 85–101.
23. Shapiro, C.M., Trajonovic, N.N., and Fedoroff, J.P. (2003). Sexomnia-a new paraphilia? *Canadian Journal of Psychiatry*, **48**, 311–7.
24. Bradford, J. (1987). Medical interventions in sexual deviance. In *Sexual deviance* (eds. D.R. Laws and W. O'Donohue), pp. 449–64. The Guilford Press, New York.
25. Grossman, L.S., Martis, B., and Fichtner, C.G. (1999). Are sex offenders treatable? A research overview. *Psychiatric Services*, **50**, 349–61.
26. Bradford, J.M.W. (2001). The neurobiology, neuropharmacology and pharmacologic treatment of the paraphilias and compulsive sexual behaviour. *Canadian Journal of Psychiatry*, **46**, 26–34.
27. Greenberg, D.M and Bradford, J.M.W. (1997). Treatment of the paraphilic disorders: a review of the role of selective serotonin reuptake inhibitors. *Sexual Abuse: A Journal of Research and Treatment*, **9**, 349–60.
28. Fedoroff, J.P. (1994). Serotonergic drug treatment of deviant sexual interests. *Annals of Sex Research*, **6**, 105–17.
29. Fedoroff, J.P., Peyser, C., Franz, M.L., *et al.* (1994). Sexual disorders in Huntington's disease. *The Journal of Neuropsychiatry and Clinical Neurosciences*, **6**, 147–53.
30. Marshall, W.L., Marshall, L.E., Serran, G.A., *et al.* (2006). *Treating sexual offenders*. Routledge, New York.

31. Hanson, R.K., Gordon, A., Harris, A.J.R., *et al.* (2002). First report of the collaborative outcome data project on the effectiveness of psychological treatment of sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, **27**, 169–95.
32. Marques, J.K., Weideranders, M., Day, *et al.* (2005). Effects of a relapse prevention program on sexual recidivism: final results from California's Sex Offender Treatment and Evaluation Progject (SOTEP). *Sexual Abuse: A Journal of Research and Treatment*, **17**, 79–107.
33. Laws, D.R. and O'Donohue, W. (eds.) (1997). *Sexual deviance: theory, assessment, and treatment*. The Guilford Press, New York.
34. Fedoroff, J.P., Kuban, M., and Bradford, J. (in press). Laboratory measurement of penile response in the assessment of sexual interests. In *Sex offenders: a multi-disciplinary approach to identification, risk assessment, treatment and legal issues* (eds. F. Saleh, J. Bradford, A. Grudzinskas, and D. Brodsky). Oxford Press, New York.
35. Grubin, D. (in press). Using the polygraph to manage risk in sex offenders. In *Assessment and treatment of sexual offenders: a handbook* (eds. A.R. Beech, L.A. Craig, and K.D. Browne). Wiley, London.

4.11.4 Gender identity disorder in adults

Richard Green

History

The behavioural phenomenon of transsexualism (now gender identity disorder) is ancient. It has been recorded for centuries and in a broad range of cultures.[1] The historic behavioural picture is comparable to that seen clinically.

In the first half of the twentieth century medical reports of sex reassignment surgery were described in Europe, primarily in Switzerland.[2] In the 1930s a wide-selling biography *Man into Woman* described a Dutch painter who underwent surgical sex reassignment.[3] Contemporary interest in transsexualism surged in 1952 when George Jorgensen, an American, travelled to Denmark and underwent hormonal and surgical treatment to become Christine Jorgensen.[4] The resultant international publicity yielded hundreds of people worldwide applying to the Danish doctors for similar treatment.[5]

By the mid-1960s there were surgeons scattered in several countries performing sex reassignment. Then in the United States at the Johns Hopkins Hospital and the University of Minnesota Hospitals and in the United Kingdom at Charing Cross Hospital, comprehensive sex reassignment programmes commenced. Extensive publicity was given to the Johns Hopkins programme as initially reported in the *New York Times* in 1966. It described the rationale for the programme and in the words of its director, 'if the mind cannot be changed to fit the body, then perhaps we should consider changing the body to fit the mind'.[6]

In 1966 the first professional text on transsexualism was written by Harry Benjamin, widely acknowledged as the 'father of transsexualism'.[7] In 1969 the first multidisciplinary text was edited by the author and John Money.[6] During the past 35 years the recognition of transsexualism, or gender identity disorder, as a treatable condition requiring psychiatric, endocrine, and surgical intervention has been accepted.

Epidemiology

The prevalence of gender identity disorder in adults is estimated from a comprehensive appraisal in the Netherlands at 1 in 10 000 males and 1 in 30 000 females.[8] At nearly all clinical centres the ratio of male-to-female patients ranges from 3:1 to 4:1, in favour of males. In some East European centres the ratio is 1:1 or reversed.[9]

Diagnosis

Diagnostic criteria of gender identity disorder in adults in DSM-IVTR[10] include a stated desire to be the other sex, a desire to live and be treated as the other sex, or the conviction that he or she has the typical feelings and reactions of the other sex. There is a preoccupation for removal of primary and secondary sexual characteristics and for procedures to alter physically the sexual characteristics to simulate the other sex. The condition is not associated with physical intersex. ICD-10 diagnostic criteria are similar but there is no mention of intersex exclusion.[11]

Origins

The search for the origins of transsexualism continues with an increasing bias towards those that are physiological. Some 20 years ago there was a false prophet in the guise of the HY antigen, on the Y chromosome believed to be influential in the development of the testes. A series of male transsexuals were found to be lacking this antigen and the tentative conclusion reached was that its absence resulted in a failure to masculinize the brain in the direction of a male identity.[12] However, the author's collaborative effort to replicate that study was not successful as all the male transsexuals studied appeared to have normal HY antigen.[13]

A more recent finding from the Netherlands implicates the brain region known as the bed nucleus of the stria terminalis. In a series of six male transsexuals studied at post-mortem over a 10-year period the size of the nucleus was comparable to that of typical females and not males.[14] A criticism of this study is that the long-term oestrogen treatment for these males may have altered the size of the nucleus. In response the researchers argue that males treated with anti-male hormone drugs or oestrogen for prostate cancer do not have an alteration in the nucleus size from typical males. However, this treatment may not be comparable to that given to transsexuals. Another criticism is that the sex difference in size of the nucleus does not manifest until early adulthood whereas the symptoms of GID often manifest earlier.

Research with male transsexuals has revealed what might be indirect markers reflecting biological distinctions. In agreement with other researchers' findings that male homosexuals have a greater likelihood of having older brothers,[15] our homosexually oriented male transsexuals also have more older brothers.[16] A theory behind this finding is that there is a progressive immunization with each pregnancy by the pregnant mother against the male foetus reflecting antigenicity of the Y chromosome. This would disrupt typical male development.

A higher ratio of aunts to uncles on the mother's side has also been found in our male transsexuals[17] a finding previously

reported by another researcher for male homosexuals.[18] A theory here is that a semilethal factor has been operant in one generation (against uncles) that in the subsequent generation influences brain development resulting in an atypical behavioural pattern (homosexual or transsexual development). The finding is explainable with genomic imprinting where a gene can be dormant in one generation depending on which parent transmitted it.[17]

We also find that both male and female transsexuals are more often non-right-handed.[19] Hand use preference begins in utero and may reflect hormonal levels or cerebral dysfunction.

For female transsexuals, a series of reports indicates a higher rate of polycystic ovarian disease.[20] Although such women secrete higher levels of androgen than typical females in adulthood, prenatal levels are unstudied. However, nearly all patients with polycystic ovarian disease are not transsexual and most female transsexuals do not have polycystic ovarian disease.

Treatment

There have been no randomized controlled trials of treatment and clinical management has evolved from decades of experience.

Prior to recognition of transsexualism as a disorder deserving medical and psychiatric attention many patients self-mutilated or committed suicide[7] Transsexual patients are helped by sympathetic assessment and intervention. However, transsexuals can be difficult patients to treat. It is a rare disorder in which patients make their own diagnosis, 'I am transsexual', and prescribe their own treatment, 'I want sex-change hormones and surgery'. Patients can be demanding and impatient for the therapist's acquiescence. They may be resentful for having to see a psychiatrist, holding the opinion that the desire for hormonal treatment and surgery should be sufficient, and psychiatric agreement should be unnecessary. Some patients will threaten self-mutilation or suicide if their demands and time schedules for demands are not met. Patients need to know that psychological stability is a key ingredient to successful negotiation of the cross-gender living trial period 'Real Life Experience' (see below) and recommendation for surgery, and that suicidal behaviour is a contraindication to going forward.

The general psychiatrist should take a full psychosexual history with emphasis on the onset and development of the gender dysphoria, attempts at treatment, and the patient's long-term goals and appreciation of the obstacles to be confronted. General psychiatric and medical status needs to be understood. A referral is to be made to a specialist centre.

During the past 30 years medical doctors and psychologists specializing in the treatment of transsexualism have worked to develop effective intervention strategies. Early on there was some optimism that extensive, prolonged psychotherapy could modify the patient's gender identity to conform to the patient's birth sex. However, in the vast majority of cases, this was not possible.

During the past 20 years one project undertaken by the Harry Benjamin International Gender Dysphoria Association, the professional body dedicated to the study and treatment of transsexualism, has been setting an extensive series of requirements for evaluation and treatment of gender identity disordered people. This set of guidelines is known as the *Standards of Care*.[21] Their principal purpose is to assure that people presenting with dissatisfaction continuing to live in the sex role to which they were born undergo comprehensive psychiatric and other medical evaluation and enter into an appropriate treatment programme. The programme includes, in addition to ongoing psychiatric or psychological monitoring, possibly endocrine therapy and, depending on the outcome of the graduated trial period of cross-gender living, possibly sex reassignment surgical procedures. The philosophy of treatment is to do reversible procedures before those that are irreversible. Thus, clothing change, name change, and cross-gender role socialization, would precede endocrine treatment with its gradual somatic changes, followed, in carefully selected cases, by surgical treatment.

The screening and evaluation of patients given the opportunity to demonstrate that they will benefit from cross-gender living, perhaps culminating in surgical treatment, is known as the 'Real Life Experience'. This requires that patients live full-time for at least a year and preferably 2 years in that role. The experience includes high doses of cross-sex hormones and full-time employment or full-time student status in the new role for at least a year. If patients can demonstrate to themselves and mental health experts that they have successfully negotiated the 'Real Life Experience' and are adjusting better in this new gender role, they can be referred for surgery.

Hormonal effects

Response to hormonal treatment is variable. This is particularly notable and potentially problematic for males. As with people born female, breast development spans a continuum. Patients may erroneously believe that more oestrogen will result in greater breast development. They neglect the fact that people born female have quite adequate female hormone production but the limiting factor is tissue response. In addition to breast development, male patients report increased hip and buttock fat, skin softening, and loss of sex drive and erection capacity. Vocal retraining is required and perhaps surgical alteration of the larynx to effect a woman's voice. Facial hair removal is required. Some patients opt for major facial contour reconfiguration, performed by a few cosmetic surgeons.

Androgen treatment to the female results in voice deepening, facial hair growth, general body hair growth, menses cessation, clitoral hypertrophy, and increased sex drive. Testosterone effects are very pronounced. The deepening voice and beard growth and perhaps scalp hair loss can metamorphose the female's appearance dramatically.

Surgery

Genital surgical treatment for the male includes penectomy, orchidectomy, and creation of a neovagina. The neovagina may be created from penile skin, perhaps augmented by other skin, or from a portion of large intestine.[22] The cosmetic result is usually very good. The extent of sexual responsivity with the neovagina is mainly anecdotal. Many patients report the subjective experience of orgasm but describe it in a different form from that experienced prior to surgery as a male. Very few physiological measures of the sexual response cycle have been reported with postoperative transsexual patients.[23]

Female transsexuals undergo bilateral mastectomy, and usually hysterectomy and ovariectomy. Genital surgery is an option taken

by perhaps half because of the limitations of phalloplasty. Two major approaches are utilized. One is creation of a micropenis from the androgen enlarged clitoris with relocation of the urethra to enable micturition while standing. Prosthetic testes can be incorporated into the labia sutured together. The microphallus will not permit vaginal penetration for intercourse but is erotically sensitive.[24] Alternatively, phalloplasty involves major surgical interventions with scarring at donor sites, particularly the arm.[25] The neophallus is not as close cosmetically to a natural penis as some patients want. It can be made more rigid with an inflatable implant and a conduit for urine may be surgically created. A procedure anastomosing a nerve from the arm to that enervating the clitoris offers promise of erotic sensation along some of the neophallus.

Sex reassignment outcome

Follow-up reports on operated transsexuals are generally quite favourable. An early review of several follow-up studies,[26] reported on 283 male-to-female transsexuals. Results were judged satisfactory for 71 per cent, uncertain for 17 per cent, and unsatisfactory for 12 per cent. For 83 female-to-male transsexuals results were judged satisfactory for 81 per cent, uncertain for 13 per cent, and unsatisfactory for 6 per cent. A more recent report considered reassignment successful in 46 of 50 male-to-female transsexuals and successful in all 61 female-to-male transsexuals.[27] In another study, of 55 male-to-female transsexuals, none regretted surgery and none had significant doubts regarding their reassignment status as women. Of 25 female-to-male transsexuals, at least 90 per cent were judged successful.[28] In a review of the English language literature over a 10-year period for operated transsexuals, 90 per cent of male-to-female transsexuals were judged to be satisfactory and 95 per cent of female-to-male transsexuals were similarly judged successful.[29] The criteria for success in some studies include objective measures of psychological and vocational status, in others the criterion is limited to an absence of regret over the reassignment process.

One study is notable because it effected some randomization of treatment conditions.[30] It was reported on 40 male-to-female transsexuals approved for surgery at Charing Cross Hospital, London. As patients qualified for surgery they were randomly assigned to two groups. Half were operated on in 3 months and the other half were kept on a waiting list for about 2 years. All patients completed a standardized assessment at acceptance for surgery and at the end of 2 years. The group that received the earlier surgery showed significant improvement in a range of psychometric measures and maintained employment. The unoperated group showed no improvement in psychological testing and deteriorated in employment.

Family management

Transsexual patients are often married and have children. When possible, the family should become part of the treatment process. Typically relations are strained with the marital partner of the transsexual and divorce is usual. Particularly when children are involved an effort should be made to deal with the feelings of betrayal or abandonment from the transsexual's spouse that contaminate the continuing relationship between transsexual parent and children. Often there is concern by the parents that the patient's transsexualism will impact adversely on the children. There is concern specifically in areas of gender identity of the children and peer group reactions to the knowledge that one of their age mate's parents is transsexual. However, in the author's research of 34 children of transsexuals who were living with or in regular contact with the transsexual parent, there were no instances of gender identity disorder in the children and no instances of peer group alienation that were especially problematic.[31] Children typically have many questions about the transsexual transformation of that parent that can be answered by the clinician, perhaps with the transsexual present.

The third sex

A recent development in the pattern of patients presenting clinically are those with a transgendered identity, popularly known as 'the third sex'. These males or females do not request 'sex change'. Rather, they want, if male, to be demasculinized and, if female, to be defeminized. Thus males may want castration and penectomy but no oestrogen treatment and no vagina, and females may want mastectomy, perhaps hysterectomy, but no androgen treatment and no phallus.

These patients pose a dilemma for clinicians. The crux of patient management for gender identity disorder is the 'Real Life Experience' (see above), including cross-sex hormonal treatment, the prelude to possible surgical alteration. Reversible procedures precede those that are irreversible in this management strategy. But with third-sex patients, no 'Real Life Experience' is possible. They do not have a trial period. Guidelines for testing the rationality and stability of their requests need to evolve from the body of clinicians currently attempting management of this unique population.

Transsexual patient subgroups

Professionals not experienced in the treatment of large numbers of male-to-female transsexual patients are often unaware that a substantial minority, perhaps a third, are not sexually oriented to male partners. Many of these patients have been married and are fathers, and many will be bisexual or remain sexually attracted to females only after surgery and live as lesbian women. A much smaller number of female-to-male transsexuals are sexually attracted to male partners and live as gay men after reassignment surgery.

In addition to the subtypes of transsexuals based on their sexual orientation, some male transsexual patients evolve through a diagnostic phase more closely fitting fetishistic transvestism. These patients have been more masculine in general lifestyle and appearance than other male transsexuals, cross-dressing has been sexually arousing, and they have usually been heterosexually oriented. However, with the passage of time gender dysphoria increases and fetishistic components of cross-dressing diminish or disappear. Many have been sexually aroused by fantasies of themselves as women.[32] They have been termed 'autogynephiles'. There is some evidence that males evolving through a fetishistic cross-dressing phase, presenting as somewhat older at gender identity clinics, have a poorer prognosis after surgery. However, it is primarily the progression through the 'Real Life Experience' that becomes the critical management guideline for patients, irrespective of their background.

Gender identity as a disorder

A growing movement among transsexual people argues for removal of gender identity disorder from the psychiatric and medical lists of disorders or diseases. These advocates argue that their sexual identity is normal male or female and that surgical correction of their anatomical anomaly is all that is required to allow them to live as normal men or women. Their condition is depicted as distinctive from the recognized traditional forms of mental disorder such as schizophrenia or major depression. Furthermore, carrying a psychiatric diagnosis is stigmatizing. One argument for inclusion of transsexualism in the APA diagnostic manual of disorders is that it follows the criteria of other entries, that is subjective distress and social disadvantage. Another considers third-party payment for treatment. It is unlikely that medical insurance carriers, private or governmental, would fund intervention for a non-medical condition.

Additionally, there is objection by some persons with gender identity disorder to being referred to as transsexual. They prefer the designation 'transwoman' for what has been known as male-to-female transsexualism, and 'transman' for female-to-male transsexualism.

Legal issues

Cultural and legal approaches to transsexualism vary widely across nations and cultures. They are beyond the scope of this chapter. However legal issues in the United States and United Kingdom can be summarized as follows.

Part of the 'Real Life Experience' of cross-gender living includes employment in the desired gender role. However, transsexuals may be the object of employment discrimination based on their transsexualism. In the United States, transsexuals were denied federal protection 20 years ago when the court held that the anti-sex discrimination statute protected men and women, but not transsexuals.[33] Finally, in 2004, another federal court extended employment protection.[34] In the United Kingdom, in 1997, a ruling on an English case before the European Court of Justice held that anti-sex discrimination law, to which all European Union members were subject, did include transsexuals. Thus discrimination in employment was illegal.[35]

Changing sex on one's birth certificate can be an important step in the life of the postoperative transsexual. In its absence, the person's legal sex may remain in the preoperative status, and pose obstacles to a full life. Marriage is a key issue. In the United Kingdom, until recently, postoperative transsexuals could not have a new birth certificate issued and change their legal sex. Thus a male-to-female transsexual could not marry a male and a female-to-male transsexual could not marry a female. This changed in a statute enacted in 2004.[36] In the United States most states permit some birth certificate change. However, that change may not be recognized in another state.

Further information

Barrett, J. (ed.) (2008). *Transsexual and other disorders of gender identity*. Radcliffe Publishing, Oxford.

Ettner, R., Monstrey, S., and Eyler, E. (eds.) (2008). *Principles of transgendered medicine and surgery*. Haworth Press, Binghampton.

References

1. Green, R. (1966). Transsexualism: mythological, historical and cross-cultural aspects. In *Appendix to The transsexual phenomenon* (ed. H. Benjamin). Julian Press, New York.
2. Abraham, F. (1931). Genital alteration in two male transvestites. *Zeitschrift Sexualwissenschaft*, **18**, 223–6.
3. Hoyer, N. (1933). *Man into woman: an authentic record of a change of sex*. E.P. Dutton, New York.
4. Hamburger, C., Sturup, G., and Dahl-Iversen, E. (1953). Transvestism: hormonal, psychiatric and surgical treatment. *The Journal of the American Medical Association*, **12**, 391–6.
5. Hamburger, C. (1953). The desire for change of sex as shown by personal letters from 465 men and women. *Acta Endocrinologica*, **14**, 361–75.
6. Green, R. and Money, J. (eds.) (1969). *Transsexualism and sex reassignment*. Johns Hopkins Press, Baltimore, MD.
7. Benjamin, H. (1966). *The transsexual phenomenon*. Julian Press, New York.
8. Kesteren, P., Gooren, L., and Megens, J. (1996). An epidemiological and demographic study of transsexuals in The Netherlands. *Archives of Sexual Behavior*, **25**, 589–600.
9. Godlewski, J. (1988). Transsexualism and anatomic sex ratio reversal in Poland. *Archives of Sexual Behavior*, **17**, 547–8.
10. American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th edn, text revision). American Psychiatric Association, Washington, DC.
11. World Health Organization. (1992). *International statistical classification of diseases and related health problems, 10th revision*. WHO, Geneva.
12. Eicher, W., Spoljar, M., Cleve, H., *et al.* (1979). H-Y antigen in transsexuality. *Lancet*, **2**, 1137–8.
13. Wachtel, S., Green, R., Simon, N., *et al.* (1986). On the expression of H-Y antigen in transsexuals. *Archives of Sexual Behavior*, **15**, 49–66.
14. Zhou, J., Hoffman, M., Gooren, L., *et al.* (1995). A sex difference in the human brain and its relation to transsexuality. *Nature*, **378**, 68–70.
15. Blanchard, R. (1997). Birth order and sibling sex ratio in homosexual versus heterosexual males and females. *Annual Review of Sex Research*, **8**, 27–67.
16. Green, R. (2000). Birth order and ratio of brothers to sisters in transsexuals. *Psychological Medicine*, **30**, 789–95.
17. Green, R. and Keverne, E.B. (2000). The disparate maternal aunt–uncle ratio in male transsexuals: an explanation invoking genomic imprinting. *Journal of Theoretical Biology*, **202**, 55–63.
18. Turner, W. (1995). Homosexuality, type 1: an Xq 28 phenomenon. *Archives of Sexual Behavior*, **24**, 109–34.
19. Green, R. and Young, R. (2001). Hand preference, sexual preference, and transsexualism. *Archives of Sexual Behavior*, **30**, 565–74.
20. Futterweit, W., Weiss, R., and Fagerstrom, R. (1986). Endocrine evaluation of 40 female-to-male transsexuals: increased frequency of polycystic ovarian disease in female transsexualism. *Archives of Sexual Behavior*, **15**, 69–78.
21. Harry Benjamin International Gender Dysphoria Association. (1981). *The standards of care for gender identity disorder*. Symposion, Düsseldorf.
22. Schrang, E. (1998). Male-to-female feminizing genital surgery. In *Current concepts in transgender identity* (ed. D. Denny), pp. 315–33. Garland, New York.
23. Green, R. (1998). Sexual functioning in post-operative transsexuals. *International Journal of Impotence Research*, **10**(Suppl. 1), 22–4.
24. Hage, J. and van Turnhout, A. (2006). Long-term outcome of metaidoiplasty in 70 female-to-male transsexuals. *Annals of Plastic Surgery*, **57**, 312–16.

25. Monstrey, S., Hoebeke, P., Dhont, M., *et al.* (2005). Radial forearm phalloplasty, review of 81 cases. *European Journal of Plastic Surgery*, **28**, 206–12.
26. Pauly, I. (1981). Outcome of sex reassignment surgery for transsexuals. *Australian and New Zealand Journal of Psychiatry*, **15**, 45–51.
27. Blanchard, R., Steiner, B., Clemensen, L., *et al.* (1989). Prediction of regrets in postoperative transsexuals. *Canadian Journal of Psychiatry*, **34**, 43–5.
28. Kuiper, B. and Cohen-Kettenis, P. (1988). Sex reassignment surgery. A study of 141 Dutch transsexuals. *Archives of Sexual Behavior*, **17**, 439–57.
29. Green, R. and Fleming, D. (eds.) (1991). Transsexual surgery follow up: status in the 1990s. In *Annual review of sex research* (eds. J. Bancroft, C. Davis, and D. Weinstein). Society for the Scientific Study of Sex, Mount Vernon, IA.
30. Mate-Cole, C., Freschi, M., and Robin, A. (1990). A controlled study of psychological and social change after surgical gender reassignment in selected male transsexuals. *The British Journal of Psychiatry*, **157**, 261–4.
31. Green, R. (1998). Transsexuals' children. *International Journal of Transgenderism*, **2**, 1–7.
32. Blanchard, R. (1989). The classification and labelling of non-homosexual gender dysphorics. *Archives of Sexual Behavior*, **18**, 315–34.
33. Green, R. (1986). Spelling relief for transsexuals: employment discrimination and the criteria of sex. *Yale Law and Policy Review*, **4**, 125–40.
34. Smith,v. City of Salem. (2004). 378 F.3d 566 (6th Cir.)
35. P v. S and Cornwall County Council. (1996). ECRI-2143 C-13/94.
36. Gender Recognition Act. (2004).

4.12

Personality disorders

Contents

4.12.1 Personality disorders: an introductory perspective

Juan J. López-Ibor Jr.

The goal of psychiatry is the study of mental illnesses. In this chapter we consider the degree to which personality disorders can be considered as mental illnesses.

Basic notions

Personality is the quality that makes each one of us both different from others and consistently recognisable throughout our lives. Hence, there are two approaches to study personality. One is transversal, consisting on description of archetypes of human beings. One of the first to take this approach was Theophrastus (372–287/5 BC) who in his book *The Characters*, portrays thirty-two such prototypes. Some of them can be are easily recognized by present-day psychiatrists, for instance those typified by poor impulse control: The offensive man (*bdeluria*), the unsociable man (*authadeia*), the show-off (*alazoneia*) and the slanderer (*kakologia*); by obsessive traits: the superstitious man (*deisidaimonia*) or by paranoid traits: the suspicious man (*apistia*). The corresponding contemporary approach consists of the isolation of psychological traits or dispositions, to describe permanent inclinations to behave in a preset way.

The longitudinal approach to the study of the personality is based on the notion that there is an initial seed that develops through the lifetime. Sir Francis Galton (1822–1911) was among the first to consider the inheritance of individual differences in humans, although for centuries breeders of dogs, horses or bulls for bullfighting, had been selecting animals for mating on to select desired characteristics whether is be hunting, running or fighting.

Twin and developmental studies have been used. For example, The 'New York Longitudinal Study'[1] on infant temperament started in the early 1950s and examined how temperament influences adjustment throughout life. Kagan *et al.*[2] followed up a cohort of babies to age 14-17 years and reported that those who were highly reactive when they were babies were more likely to be 'subdued in unfamiliar situations, to report a sour mood and anxiety over the future and to be more religious'.

There are two key features of personality, one of which is temperament and the other character. The two together constitute personality.

Temperament is the innate predisposition to behave in a particular manner. Historically the concept was part of the theory of the four humours, which had corresponding temperaments: *sanguine* (the individual is led by his own pleasure to live), *choleric* (the individual has a feeling of power and shows it), *melancholic* (the individual is dominated by doubts and ruminations) and *phlegmatic* (the individual lacks any links to life, lives without effort nor pleasure). Current research on the biological basis of personality has renewed the interest in temperament.

Character is a configuration of habits, a disposition, consisting in the actualised aspects acquired through learning and shaped by experience.

Psychiatry and abnormal behaviours

Descriptions of individuals with behavioural characteristics of a negative moral or social value exist in every culture and most societies have established institutions in which the marginalized have been confined, as recorded by Foucault.[3] The distinction between immoral behaviour and mental illness was established in France at the end of the eighteenth century, coinciding with the birth of modern psychiatry. The Marquis de Sade was expelled from the Chârenton Hospital because, in words of the director, 'he is not ill, his only madness is vice'.[3] Pinel (1745–1826) considered that, in the case of the young man who in an attack of rage threw a woman into a well, although his ability to judge was clear and intact and although he presented no delusional ideas, his behaviour was characteristic of a mental patient. Consequently, this murderer was diagnosed as suffering from *manie sans délire* and his madness was classified as **reasoning madness** (*folie raisonnante*).[4] This reasoning is similar to that of Cleckley 150 years later who proposed that the social maladaptation of psychopaths is of such high degree that should considered as the result of an underlying psychotic disturbance, being personality disorders are a *mask of sanity*.[5]

Prichard,[6] defined the concept of moral insanity from which, together with the moral degeneration described by Morel (1809–1873),[7] the modern concepts of psychopathy and personality disorders are derived.

Difficulties in the study of personality disorders

Two factors have prevented the development of scientific knowledge in this field: first the negative evaluation of the concept of moral insanity, and second the dualism inherent in psychopathology.

The stigma of personality disorder

The diagnosis of personality disorder generally implies the idea of intractability and frequently leads to a lack of proper medical care. This attitude is the expression of a negative, moralising, and, according to Tyrer *et al.*,[8] delusional attitude of the doctor towards the patient. Cusack and Malaney[9] posed the question as to whether patients with antisocial personality disorders are 'bad' or 'mad'. They attempted to establish differential criteria in order to show that if patients with an antisocial personality disorder are not 'mad', then they must be considered as 'bad' and therefore must be delivered to the judicial system, after diagnosis and treatment of secondary symptoms. In 1999, the UK Government 1999 introduced a new concept: Dangerous and Severe Personality Disorder (DSPD). This subsequently became a treatment and assessment program for individuals who satisfy three requirements: 1) have a severe disorder of personality, 2) present a significant risk of causing serious physical or psychological harm from which the victim would find it difficult or impossible to recover, and 3) the risk of offending should be functionally linked to the personality disorder.[10, 11]

Dualism in psychopathology

Dualism has been present in psychiatry since its origins as specialty. According to Griesinger:[12]

> *It is time that [mental medicine] should be cultivated as a branch of brain pathology and of [the study of] the nervous system in general, and to apply serious diagnostic methods used in all branches of medicine. . . . Besides this purely medical element, mental medicine has another essential one and which gives a special and proper character to this part of the healing art; it is the psychological study of the aberrations of the intelligence observed in mental illnesses.*

The radical separation between mental-brain illnesses and 'aberrations of intelligence' is fundamental to modern psychopathology. Schneider[13] distinguished between psychoses as pathological conditions of the brain (disease or defective structure) and variations of the psychic way of being. Abnormal personalities, personality disorders, and neurotic disorders belong to the second category.

Schneider[14] defines some **abnormal personalities** in a statistical sense, to describe those individuals whose form of feeling, experience and behaving differs to a certain degree from what is considered to be normal for most individuals in a social group. Some of these are **psychopathic personalities** who, as a result of their abnormality, suffer or make others suffer. It should be stressed that according to Schneider's statistical definition of personality and the dualism of his psychopathological system, the only possible criterion to define a clinical condition in the absence of a brain disease is the suffering, the pathos. Suffering is for Schneider the reason why some people ask for medical care, but not a sufficient criterion for to determine the presence of an illness. Schneider had to add suffering inflicted on others (social suffering) in order to be able to include certain kinds of abnormal personalities characterized by the absence of personal suffering (heartless psychopaths, sociopaths).

It seems acceptable to consider as a patient someone who suffers and asks for clinical care although the criterion for suffering is a weak one when compared with the presence of an illness of an organ. On the contrary, the criterion of induced suffering which characterises some psychopathic personalities, defined following Schneider, is not acceptable in medicine and it is surprising that this has been little criticised. The clue lies in Schneider's definition of personality which *excludes* any biological substrate.[15] The effect of viewing psychopathies as simple variations was to reduce the amount of neurobiological research into the neuroses and personality disorders because they were not considered amenable to natural scientific methods. The study of the personality was left the new psychoanalytical and psychological theories.

Nowadays it is impossible to maintain such a reductionistic perspective, and it is recognised that the morbid nature of personality disorders can be understood through the study of changes in its biological substratum. There are not two kinds of mental disorders, the psychosis which are the consequences of brain illnesses, and the variations of the psychological way of being (neurosis and personality disorders), but two inherent aspects to each disorder. It is essential to consider psychological and psychopathological aspects of psychoses, as well as the brain dysfunction of the variations of the psychic way of being.

Models of personality

The study of personality by the different schools of differential psychology provides a solid background to help to understand the disorders of personality. Unfortunately, these studies have been

conducted from different and sometimes contradictory perspectives, which are summarized in the following sections.

Categorical perspective

The categorical perspective is deep rooted in the psychiatric tradition. Categorical models consider discontinuous personality categories. This type of model is used in DSM-IV[16] and ICD-10[17] because of the need for a specific diagnostic, i.e. a categorical approach.

In modern nosology the categorisation of illness is based on the symptoms present and not on their aetiopathology, and says nothing about the nature of the disorders themselves. In the case of personality disorders, the categorisation does not affirm or deny that they are disorders or illnesses, nor does it indicate where the symptoms differ from non-morbid behaviour patterns.

This approach is supported by the notion of ideal types of Weber (1864- 1920), introduced into psychiatry by Jaspers (1883–1969) and more recently by Schwartz and Wiggins.[18] Ideal types are constructs to understand reality: An ideal type is formed by a unilateral accentuation of one or more perspectives and by the synthesis of a great deal of individual phenomena. A type describes the perfect case. Recently Doerr[19] has argued that the ideal types, when well described, become almost real types.

The experimental approach

The experimental approach looks for general laws on personality and establishes causal relations between personality variables. Wundt (1832–1920) studied the effects of modifications of stimuli on the intensity and quality of the subject's experiences introduced them. Pavlov (1849–1936) studied the conditioning of the responses to stimuli and the experimental neurosis. The behavioural approach to the personality was introduced by Watson (1878–1958) who applied objective methods to the study of human behaviour and to the relationship between stimuli and responses. Hull (1884–1952) expanded behaviourism to include learning, feelings, expectations, achievements, goals and motivations. This led to the notion that the stimulus response relationship is influenced by cognitive processes. Perception, memory, language and other functions influence the processing of information of the surrounding world and the information coming from oneself (*self*). Skinner (1904-1990) created a theory of the operating conditioning, result of a non-adaptative learning. From these perspectives, the personality is viewed as a computer which introduces, stores, transforms and produces information, including the contents of the information as well as the process in itself.[20]

The psychoanalytical approach

Freud (1856-1939)[21] proposed in the course of his life three different models of personality: The first was the speculative neuropsychological model of the *Project of a psychology for neuropsychiatrists* (1897) based on the concepts of psychic energy and psychodynamic. The second was the topographic model of the *The Interpretation of Dreams* (1900) where Freud described the conscious, preconscious and unconscious levels. The third is the structural model of *The Ego and the Id* (1923) and of *Inhibition, Symptoms and Anxiety* (1926) where Freud introduced the notions of the Id, the Ego and the Super-Ego. This last model led to a new perspective, the psychology of the Ego and the description of defence mechanisms which have a strong impact on the study of personality in clinical settings. Defence mechanisms distort reality to adapt the subject to it and to reduce anxiety. Some of them are more normal (promote adaptation to the environment), others are pathological (maladjusted or maladaptative).

The correlational approach

The correlation approach explains the individual differences based on personality traits and applying a dimensional model based on statistical correlation. Karl Pearson (1857–1936), the founder of mathematical statistics introduced the correlation coefficient (correlates of cognitive flair with variables like age, gender, weight, height and so on). Charles E. Spearman (1863–1945) applied to research on the traits of personality, a factorial analysis that groups different qualities around a series of correlational factors or dimensions.

Traits are the basic elements of a personality and individual differences are defined and classified along dimensions. The theoretical assumption is that the structure of personality is common to all individuals; it differs in the different combination of traits. Trait is a disposition to respond in a determined way to a determined situation. Traits characterize persons through brief and precise descriptions on stable ways to behave, and as behaviour is consistent, it is possible to predict behaviours. This approach has paved the way to basic dimensions of individual differences.

Dimensional models

Jung (1875-1961) made the first important contribution to the dimensional concept of the personality, based on the concept of trait or disposition.[22] A trait is a permanent inclination towards behaving in a determinant way. Traits are distributed along dimensions which make it possible to classify individuals according to their personality. The different dimensional models are based on the supposition that we all share the same personality structure, differing in the different combination of the mentioned traits. These models have benefited from the innovative statistic techniques, which allows to group different qualities of the individual character around factors of correlation or dimensions.

This dimensional approach raises several questions. How many traits define personality? Are the traits universal? Do traits relate only to manifest behaviours or are they part of feelings, values or thoughts? The problem of the number of dimensions that define personality led to the search for external validators such as biological, cultural and genetic factors. Eysenck[23] identified there are three basic types of personality: extroversion-introversion, neuroticism and psychoticism, each one including multiple levels of traits. For Eysenck and Eysenck,[24] the concept of arousal level is essential. Every individual has an optimal activation level of specific systems of the central nervous system—the better they feel, the better they will perform. This approach has been developed by many authors including Zuckermann,[25] who described sensation-seeking behaviour, Oreland *et al.*,[26] and Siever and Davis,[27] who proposed new traits and dimensions.

Cloninger[28] initially proposed three dimensions: novelty-seeking, harm avoidance, and reward dependence. Latter, he attempted to overcome the dichotomy between dimensional and categorical models by using four temperamental dimensions (novelty-seeking, harm avoidance, reward dependence, and persistence), which are life-long and stable, and three character dimensions (self-direction,

co-operation, and self-transcendency) which are variable and susceptible to environmental influences and development.[29]

The five-factor model, based on factorial studies and individual differences[30] has been widely accepted. It comprises the personality dimensions openness, conscientiousness, extraversion, agreeableness, and neuroticism, known by the acronym OCEAN. About 40 per cent of individual personality differences can be explained in terms of heredity.[31] In the five-factor model the same proportion does not apply to each factor; openness to experience appears to have the greatest hereditary input, whereas conscientiousness appears to have the least.

Mathematical tools allow recombining the data in order to find higher order factors of the Big Five. Two of them have appeared in many studies: 1) related to the Big Five trait dimensions Agreeableness, Conscientiousness, and Emotional Stability (meta-trait alpha) and 2) the dimensions Extraversion and Intellect (meta-trait beta).[32] Other have found some extra traits to be added to the Big Five, such as honesty-humility.[33]

An interesting approach is lexicographic, which is based on the examination of relations among personality-descriptive adjectives that are indigenous to various languages. They tend to reveal a structure corresponding closely to the Five-Factor Model, with some differences in the nature of the Agreeableness and Emotionality/Neuroticism factors and also in the existence of a sixth factor, Honesty-Humility.[34] This has been found in different languages such a tagalong with some differences (a Filipino extra factor resembled a Negative Valence or Infrequency dimension).[34] A study with college students yielded seven major dimensions; many of the factors were similar to recognized lexical personality factors. Big Five Conscientiousness and Neuroticism were each strongly associated with a single proverb dimension (interpreted as Restraint and Enjoys Life, respectively). Big Five Agreeableness, Extraversion, and Intellect/Imagination were all associated with several proverb dimensions. Agreeableness was most strongly associated with proverb dimensions representing Machiavellian behaviour and strong Group Ties, and both Extraversion and Intellect showed particularly notable associations with an Achievement Striving dimension. The two remaining proverb dimensions, which represented a belief that Life is Fair and an attitude of Cynicism, could not be accounted for by the Big Five.[35]

Critics of this approach have argued that a) little progress has been made in this area, b) structural models have little direct relevance for psychopathology research, c) the principal methodological tool of structural research–factor analysis–is too subjective to yield psychologically meaningful results and 4), some clinically relevant aspects, such as alexithymia or impulsivity, do not appear in the studies on the structure of personality.

Alexithymia positively correlates with Neuroticism (N) and negatively with Extraversion (E) and Openness (O), whereas no significant relations were found with Agreeableness (A) and Conscientiousness (C).[36] Impulsivity. The term impulsive madness was used in German literature and Jaspers[37] put it in relation to nostalgia and displacement. The balance between social and individual norms is related to the origin of mental disorders. Durkheim[38] introduced the concept of **anomia** when describing a particular form of suicide in individuals who perceive that their own norms and values are no longer relevant and that their relation to the community is weak or non-existent. However, the opposite may also occur. Kraus[39] coined the term **hypernomia** for premorbid personality traits of depressive patients, which consist in an exaggerated form of adaptation to social norms. This personality type is the converse of the impulsive madness that can be characterized as **hyponomic**. We have proposed the term **dysnomic** for obsessive patients[40] who show a distorted adaptation to social norms. For example, patients with obsessions and compulsions related to cleanliness usually appear to be extremely dirty because of their fear of contamination and their unrealistic compulsions, which are based more on 'magic' control than on efficient behaviour oriented towards concrete goals.

The common link in the psychopathology of obsessive-compulsive disorders and the group of impulsive disorders experienced by the **impulsivists**[41] is poor control of the impulses, in the sense that novel interior or exterior experiences are not converted into adequate behavioural patterns. Rather, obsessives abandon actions uncompleted and impulsivists behave in a disorganised manner (acting out). In both cases, 'the irrelevant substitutes the relevant'.[42]

Today, the trend is to look for a classification of personality disorder which will be dimensional, either by selecting one of the existing models by developing a common, integrative representation including the important contributions of each of the models.[43]

Models of personality and personality disorders

The ideal goal of a single structural framework to be applied to normal and abnormal personality is not easy to reach. In a study with the Eysenck Personality Questionnaire-Revised (EPQ-R) the three clusters of personality disorders of DSM found equivocal support. Exploratory principal components analysis and confirmatory factor analysis found four broad factors of personality disorder that overlapped with normal personality traits: an asthenic factor related to neuroticism; an antisocial factor associated with psychoticism; an asocial factor linked to introversion-extraversion; and an anankastic (obsessive-compulsive) factor. In spite of this, there is growing agreement about the number and type of broad personality disorder dimensions; similar dimensions may be found in clinical and non-clinical samples, suggesting that those people with personality disorders differ quantitatively rather than qualitatively from others; and there is substantial overlap between normal and abnormal personality dimensions.[44] For Livesley *et al.*,[45] personality disorders are quantitatively extreme expressions of normal personality functioning developed around four factors: Emotional Dysregulation, Dissocial Behaviour, Inhibitedness and Compulsivity. In the same year, the same group published some different results in another study:( ) they found 16 basic dispositional traits (anxiousness, affective lability, callousness, cognitive dysregulation, compulsivity, conduct problems, insecure attachment, intimacy avoidance, narcissism, oppositionality, rejection, restricted expression, social avoidance, stimulus seeking, submissiveness, and suspiciousness) and three higher-order patterns (emotional dysregulation, dissocial behaviour, and inhibitedness).[46]

Markon *et al.*[47] have delineated an integrative hierarchical account of the structure of normal and abnormal personality. This hierarchical structure integrates many Big Trait models proposed

in the literature. Similarly, O'Connor[48] reanalyzing the published studies found high level of support for both theoretically and empirically based representations of the five-factor model approach to personality disorders. The five-factor model personality dimensions of Neuroticism, Extraversion, and Agreeableness are the most apparent in the DSM-III-R conceptualizations of the personality disorders.[49,50] The five-factor model structure is present although with some variance in the current DSM-IV cluster set.[50]

Mulder and Joyce[51] have attempted to construct a simplified system for the classification of personality disorders related to normally distribute human personality characteristics. A four-factor solution of personality disorder symptoms was obtained and they labelled these factors 'the four As': Antisocial, Asocial, Asthenic and Anankastic. The factors related to the four temperament dimensions of the Tridimensional Personality Questionnaire (TPQ), but less closely to Eysenck Personality Questionnaire (EPQ) dimensions. The four factors were similar to those identified in a number of studies using a variety of assessment methods and this lends some credibility to our findings.

The masking of sanity becomes evident in some dissocial, psychopathic and even criminal behaviours that are the expression of underlying disorders. The issue is very relevant because those can be treated. One example is attention deficit hyperactivity disorder (ADHD) which has important forensic implications. For decades there has been an interest in predicting which children will become psychopaths in order to establish primary prevention interventions. Lynam[52] described the fledgling psychopath, characterized by symptoms of hyperactivity-impulsivity-attention problems and conduct problems are at the greatest risk for becoming chronic offenders. Other authors[53,54] have strongly argued against the clinical-forensic utility of tests designed to assess juvenile psychopathy.

The problems of axis II of the diagnostic and statistical manual

There are several reasons in favour of an independent axis for personality disorders. First it is important not to forget personality in the diagnostic process, especially when using symptomatic classifications (DSM-III / IV and ICD-10). Second, personality may be a predisposing factor, or something essential for the response to treatment and for prognosis.[54] Third personality traits are egosyntonic: they include traits that the subject has accepted as an integrative part of him/herself in a progressive way (when compared to axis I disorders and non-psychiatric illnesses which are something 'that occurs').

Seen from a theoretical perspective, personality disorders may be considered as the expression of the personality's functioning, which is essential for patients, be they psychiatric or not.[45]

But there are also reasons to combine Axis I and Axis II. The better a personality disorder is known and the more biological correlations are found in it, the higher the probabilities it has to be moved to Axis I: Epileptic personality (adhesivity, gliscroidy) became long ago organic disorder of the personality, cyclothymic personality became cyclothymia and depressive personality (at least in part), dysthymia in DSM-III. Schizotypal personality is schizotypal disorder in ICD-10 and several studies emphasise the strong relation of borderline personality disorder and mood (affective) disorders.[55]

Although through most of the 20th century, from K. Schneider to DSM-IV, personality disorders and mental illnesses were studied as separate fields, there has been increasing recognition of the substantial overlap of – and comorbidity between – disorders both within and across axes and interest in the joint study of normal and abnormal personality.[56] In comorbid cases, the personality disorder could be a predisposing factor, a consequence, or an attenuated form of the mental disorder, or it could be independent of the mental disorder. The fact that the association between a mental disorder and a personality disorder is not always fortuitous has been shown by the observation that effective treatment of the former can lead to the disappearance of the latter, as has been demonstrated in the treatment of obsessive-compulsive patients with pharmacotherapy and behavioural therapy.[57]

The main difference between mental and personality disorders may be that the latter are early onset variants with a very chronic course. According to the hypothesis of a spectrum of disorders, personality disorders can often be treated by the same method as those applied to the major psychiatric disorders to which they are related. Patients with anxious or avoidant personality disorder may respond to anxiolytic medication, patients with borderline personality disorder may respond to lithium and antidepressives, patients with schizotypal personality disorder may respond to antipsychotic agents, and patients with disorders characterised by poor impulse control may respond to antidepressives with a selective serotonergic action.

The diathesis-stress model

The diathesis-stress model has been used to explain the relationship between personality and mental disorders. Supposedly broad, innate temperament dimensions, sometimes correlated with somatic characteristics such as body type (Kretschmer [1888–1964]), develop and differentiate themselves through both biologically and environmental events into a hierarchical personality trait structure. At their extremes, are risk factors (diatheses) for psychopathology, especially given adverse life experiences (stress). Rosenthal[58] defined this model as an inherent constitutional predisposition which only becomes apparent under the impact of environmental stress.

The onset early in life, the variability of expression dependent on setting, the greater association with more severe disorders and the acceptance as intrinsic components of functioning by most suffering from personality disorders support the notion that they are diatheses rather than disorders.[59]

Personality disorders have been considered as belonging to the spectrum of major psychiatric disorders. However, it must be remembered that external events, such as brain damage (organic personality disorders), or the psychological impact of a catastrophic event may also lead to personality changes. Severe psychiatric disorders may have a repercussion on the personality of the patient, and other illnesses may also have this effect. For example, chronic pain (of organic nature) can be accompanied by a profound personality change (algogenic psychosyndrome). Hypochondria or dissociative symptoms and traits may become relevant only after the patient has suffered an illness, or a problem related to diagnosis or treatment, or a problem involving the patient–physician relationship.

Recent research has focused on the impact of social conditions in the neuroendocrine regulation of the individual, especially

regarding the adaptation to stressful situations. In patients with borderline personality disorder and suicidal impulsive behaviour we have found when compared to the control group, high basal concentrations of cortisol (suggesting a high level of stress) and a very blunted response to the stimulus (suggesting a reduced capacity to respond to external stimulus).[60] A clue to the interpretation of these results lies in the work of Sapolsky,[61] who has studied the adaptation to stress of baboons in the Serengeti savannah in Africa. Males of a lower rank have consistently high concentrations of the stress hormone hydrocortisone in their blood, whereas the concentration is lower in the dominant males. However, in the dominant males hydrocortisone concentrations increase rapidly at times of stress and decrease once the situation is resolved, whereas the lower-order males, who live in a permanent state of stress, are unable to initiate more adaptive resources (increase hydrocortisone secretion) when new stressful events appear. These patterns are the consequence and not the cause of the rank (if the opposite were true, the baboons who were physiologically better able to respond to stressful situations would achieve a higher rank). During periods of revolution, members of the colony hold successively different ranks and, although there is always one dominant animal, the stability of the society is lost and with it the stress-adapted physiology of the dominant males who show prolonged increased hydrocortisone concentrations like the rest of the group. When calm is established again, the normal pattern related to the hierarchical rank of the baboons is restored regardless of what their cortisol secretion pattern was prior to the revolution.

Relational disorders

This model can be applied to the relational disorders or processes. On one side every clinician has the experience of persons whose behavioural problems happen in specific environments (i.e., the family) or in relation to specific persons (i.e., the spouse) while their behaviour is totally normal in the rest of circumstances. Another observation is the case of personality disorders that once beginning to improve after a therapeutic intervention, the whole atmosphere around him or her changes. The patient himself and their relations start to consider that the personality as not a stable and incorrigible set of traits but something that can change for the better. However, it is the influence of child psychiatrists looking for childhood antecedents of mayor psychiatric disorders of adulthood[62] and family therapists that have requested a new diagnostic category to describe these situations.

It has also been claimed that in order to be able to proceed along this line, some problems have to be addressed, although the same may also be present in diagnostic categories accepted in DSM-IV and ICD-10. The main one are: 1) the little consensus on assessment means; 2) the complexity of relational assessment, which results in a lack of well-accepted, evidence-based operational definitions; 3) insufficient empirical testing of relational issues; and 4) the resistance to labeling social difficulties as disorders.[63]

These difficulties can be surmounted if new diagnostic perspectives are implemented. One is the concept of diseases as harmful dysfunctions a view which holds that disorders are harmful failures of biologically selected functions. There are evolutionarily selected functions that depend for their performance on the nature of the interaction between individuals. These relations can fail, even when both individuals are normal, because of mismatches between normal variations. Thus, there are genuine relational dysfunctions that, when harmful, are relational disorders.[64] The Structural Analysis of Social Behavior (SASB) has operationalised interpersonal theory for the research of relational aspects of psychopathology and becoming a useful diagnostic tool.[65]

The research agenda for DSM-V recognizes that the diagnosis of relational disorders is one of the most important gaps in the current DSM-IV.[66] Specific recommendations include developing assessment modules, determining the clinical utility of relational disorders, determining the role of relational disorders in the aetiology and maintenance of individual mental disorders, and considering aspects of relational disorders that might be modulated by individual mental disorders.

Personality disorders vs. personality variants

It is necessary to establish a clear distinction between personality disorders and personality variants, and to view the former as true morbid entities. ICD-10 allows us to differentiate, at least theoretically, personality disorders that appear in the chapter on mental disorders from personalities relevant for medicine in Chapter Z.[67] Personality disorders should be characterised by the presence of symptoms, and relevant personalities by their traits. Symptoms are used as diagnostic criteria in a categorical classification, while psychological traits can be classified according to dimensions.

Personality variants relevant to medicine, although not morbid, play an important role in the aetiopathogenesis of illnesses or are important for prognosis and rehabilitation. The study of the variants of personality also reminds the practitioner of the necessity to identify the uniqueness of the patient's personality.

The frontiers with normal personality are not well explained or justifiable threshold,[68] except for schizotypal disorder and borderline disorder. There are great variations between the different versions of DSM (DSM-III - DSM-III-R) leading in the case of schizotypal disorder to a reduction of prevalence from 11 per cent to 1 per cent.[69]

On the other hand, normal the population also presents maladaptative variants of traits such as: neuroticism, irritability, vulnerability, anxiety, depression, impulsivity, low consciousness, rush, negligence, hedonism, immorality, unreliable, irresponsibility, high antagonism, manipulation, disappointment, exploitation, aggressiveness, cruelty, heartless.

In everyday practice there is a need for categorical approach, essential to define the frontiers of clinical activity, for research, for management and for forensic questions and also of a dimensional approach, to identify functions and its alterations, relevant for interventions with patients. This could be done investigating in new personality models, adopting the multi-axial version of ICD-10[70] (Table 4.12.1.1.), implementing diagnostic system that include a nosographic approach and idiographic perspective[71] or adopting

Table 4.12.1.1 Multiaxial formulation of ICD-10

Axis I	Clinical disorders (DSM-IV axis I, II and III): personality disorders understood as morbid states of the personality
Axis II	Disablement
Axis III	Relevant non morbid circumstances, including variants of the personality (i.e., Type A behaviour) Normal personality traits (i.e.: five factors model)

Table 4.12.1.2 Proposed five axis system (Charney *et al.* 2002)

Axis I: Genotype	Identification of disease-/ symptom-related genes, of resiliency/protective genes and genes related to the therapeutic response and side effects of psychotropic drugs
Axis II: Neurobiological phenotype	Identification of intermediate phenotypes (neuroimaging, cognitive function, emotional regulation) related to genotype and to targeted pharmacotherapy
Axis III: Behavioural phenotype	Range and frequency expressed behaviours associated with genotype, neurobiological phenotype and environment, related to targeted therapies
Axis IV: Environmental modifiers or precipitators	Environmental factors that alter the behavioural and neurobiological phenotype
Axis V: Therapy	Therapeutic options base don the data of axes I to IV

a new multiaxial system[72] (Table 4.12.1.2). Several authors have claimed for a functional psychopathology. Van Praag[73] has recommended a two-tier diagnosis: 1) nosological diagnosis and 2) psychological dysfunctions correlated with biological variables.

The cosmetic of the personality?

Cosmetic or palliative pharmacotherapy is to use a psychotropic agent to make feel better to a person who is not ill. It usually intents to mitigate unwanted or unaccepted personality traits in order to attain a higher order of social normality and acceptability. Kamer[74] has described how a selective serotonin uptake inhibitor used for the treatment of depression and for other psychiatric alterations can remove personality traits in some people. He has considered traits previously considered as an expression of human misery or, in some cases, as the consequence of negative childhood experiences. Kamer even questions whether there could be a 'pandemic' of cosmetic psychopharmacologies which would lead to the disappearance of phenomena such as anguish which are essential for personal realisation in the arts, religion, and creativity. For example, fluoxetine can make non depressed people feel more vital, mentally more alert and become more popular, leading to an increased feeling of wellness. Shyness became a treatable illness when paroxetine was found to improve the symptoms of social phobia and atomoxetine can change the life conditions of adults with a history of ADHD. Most of these cases are probable subthreshold clinical conditions, something which may lead to a change in diagnostic habits, lowering thresholds or modifying criteria. For instance the DSM criteria of suffering or disablement are value loaded, and new values, such as wellbeing may be introduced.

Conclusions

During the last few decades there has been an impressive growth in research and knowledge on personality disorders. There is a strong growing evidence that they are "real" disorders that can be managed in ways similar to the rest of psychiatric disorders. There is also a growing consensus on the need for a new classification able to capture the nuances beyond the rigidity of present nosological systems. There is also a need for a clearer delimitation from normal personality variants something that will have important impacts, for instance in forensic settings or for reimbursement purposes.

Further information

International Society for the Study of Personality Disorders (ISSPD)
isspd@isspd.com info@isspd.com http://www.isspd.com
International Journal of the ISSPD, Guilford Publications, New York
info@guilford.com

References

1. Thomas, A. and Chess, S.1. (1984) Genesis and Evolution of Behavioral Disorders: From Infancy to Early Adult Life. *Am J Psychiatry*, **141**, 1–9.
2. Kagan, J., Snidman, N., Kahn, V., *et al.* (2007) *The preservation of two infant temperaments into adolescence*. Monographs of the Society for Research in Child Development, Serial No. 287, vol. 72, no. 2. Blackwell, Boston.
3. Foucault, M. (1961) *Folie et déraison. Histoire de la folie à l'âge classique*. Plon, Paris.
4. Pinel, P.H. (1809) *Traité médico-philosophique sur l'alienation mentale*. Brosson, Paris.
5. Cleckley, H. (1941) *The mask of sanity*. Henry Kimpton, London.
6. Prichard, J.C. (1835) *A treatise on insanity and other disorders affecting the mind*. Sherwood, Gilbert and Piper, London.
7. Morel, B. (1859) *Traité des dégénéréscences physiques, intellectuelles et morales de l'espèce humaine*. Paris.
8. Tyrer, P., Casey, P., and Ferguson, B. (1991) Personality disorder in perspective. *British Journal of Psychiatry,* **159**, 463–71.
9. Cusack, J.R. and Malaney, K.R. (1992). Patients with antisocial personality disorder. Are they bad or mad? *Postgraduate Medicine*, **91**, 341–4, 349–52, 355.
10. Maden,T. and Tyrer, P. (2003) Dangerous and severe personality disorders: a new personality concept from the United Kingdom. *Journal of Personality Disorders*, **17**, 489–96.
11. Tyrer, P. (2007) An agitation of contrary opinions. *British Journal of Psychiatry*, **190**, (suppl.49), 1–2.
12. Griesinger, W. (1872) *Gesammelte Abhandlungen. Vol. I, Psychiatrische und nervenpathologische Abhandlungen.* Reprinted by Bonset, Amsterdam, 1968.
13. Schneider, K. (1971) *Klinische Psychopathologie* (9th edn). Thieme, Stuttgart.
14. Schneider, K. (1950) *Die psychopathischen Persönlichkeiten* (9th revised edn). Deuticke, Vienna.
15. López Ibor, J.J. (1966) *Las neurosis como enfermedades del ánimo. Gredos, Madrid.*
16. American Psychiatric Association (1994*) Diagnostic and statistical manual of mental disorders* (4th ed.). American Psychiatric Association, Washington, DC.
17. World Health Organization (1992*) The ICD-10 classification of mental and behavioural disorders—clinical descriptions and diagnostic guidelines*. World Health Organization, Geneva.
18. Schwartz, M.A. and Wiggins, O.P. (1987) Diagnosis and ideal types: a contribution to psychiatric classification. *Comprehensive Psychiatry*, **28**, 277-91.
19. Personal communication
20. Pervin, L. (Ed.) (1990) *Handbook of personality: Theory and research*. The Guilford Press. New York.
21. Freud, S., Freud, A., Strachey, J. (1953–1974) *The standard edition of the complete psychological works of Sigmund Freud*. Hogarth Press Richmond UK.
22. Jung, C.J. (1966) *Two essays on analytical psychology*. Princeton University Press, Princeton, NJ.

23. Eysenck, H.J. (1970) *The structure of human personality*. Methuen, London.
24. Eysenck, H.J. and Eysenck, M.W. (1985*) Personality and individual differences: A natural science approach*. Plenum Press, New York.
25. Zuckermann, M. (1979) *Sensation seeking: Beyond the optimal level of arousal*. Erlbaum, Hillsdale, NJ.
26. Oreland, L., Wiberg, A., Åsberg, M., *et al.* (1981) Platelet MAO activity general clinical description of personality disorders and monoamine metabolism in CSF in depressed and suicidal patients and in healthy controls. *Psychiatric Research*, **4**, 21–9.
27. Siever, L.J. and Davis, K.L. (1991) A psychobiological perspective on the personality disorders. *American Journal of Psychiatry*, **148**, 1647–58.
28. Cloninger, C.R. (1987) A systematic method for clinical description and classification of personality variants. *Archives of General Psychiatry*, **44**, 573–88.
29. Cloninger, C.R. (1996) A psychobiological model of temperament and character. Fundamental findings for use in clinical practice. In *New research in psychiatry* (ed. H. Häfner and E.M. Wolpert), pp. 95–112. *Hogrefe and Huber, Göttingen.*
30. McCrae, R.R. and John, O.P. (1992) An introduction to the five-factor model and its applications. *Journal of Personality*, **60**, 175–213.
31. Loehlin, J.C. (1992) *Genes and environment in personality development*. Sage, Newbury Park, CA.
32. Digman, J.M. (1997) Higher-order factors of the Big Five. *Journal Personality and Social Psychology*, **73**, 1246–56.
33. Ashton, M.C. and Lee, K. (2005) Honesty-humility, the big five, and the five-factor model. *Journal of Personality*, **73**, 1321–53.
34. Church, A.T., Reyes, J.A., Katigbak, M.S., *et al.* (1997) Filipino personality structure and the big five model: a lexical approach. *Journal of Personality*, **65**, 477–528.
35. Haas, H.A. (2002) Extending the search for folk personality constructs: the dimensionality of the personality-relevant proverb domain. *Journal of Personality and Social Psychology*, **82**, 594–609.
36. Luminet, O., Bagby, R.M., Wagner, H., *et al.* (1999) Relation between alexithymia and the five-factor model of personality: a facet-level analysis. *Journal of Personality Assessment*, **73**, 345–58.
37. Jaspers, K. (1909) *Heimweh und Verbrechen*. F.C.M. Vogel, Leipzig.
38. Durkheim, E. (1951) *Suicide*. Free Press, Glencoe, IL.
39. Kraus, A. (1977) *Soziales Verhalten und Psychosen manisch-depressiver*. Enke, Stuttgart.
40. López-Ibor, J.J. (1991) Obsessive-compulsive disorder and other disorders. *European Neuropsychopharmacology*, **1**, 275–9.
41. Lacey, J.H. and Evans, C.D.H. (1986). The impulsivist: a multi-impulsive personality disorder. *British Journal of Addiction*, **81**, 641–9.
42. Janet, P. (1903) *Les obsessions et la psychasthenie*. Alcan, Paris.
43. Widiger, TA. (2007) Dimensional models of personality disorder. *World Psychiatry,* **6**, 15–9.
44. Deary, I.J., Peter, A., Austin, E., *et al.* (1998) Personality traits and personality disorders. *British Journal of Psychology*, **89**, 647–61.
45. Livesley, W.J., Jang, K.L., Vernon, P.A. (1998) Phenotypic and genetic structure of traits delineating personality disorder. *Archives of General Psychiatry*, **55**, 941–8.
46. Livesley, W.J. (1998) Suggestions for a framework for an empirically based classification of personality disorder. *Canadian Journal of Psychiatry*, **43**, 137–147.
47. Markon, K.E., Krueger, R.F., Watson, D. (2005) Delineating the structure of normal and abnormal personality: an integrative hierarchical approach. *Journal of Personality and Social Psychology*, **88**, 139–57.
48. O'Connor, B.P. (2002) The search for dimensional structure differences between normality and abnormality.. *Journal of Personality and Social Psychology*, **83**, 962–82.
49. Trull, T.J. (1992) DSM-III-R personality disorders and the five-factor model of personality: An empirical comparison. *Journal of Abnormal Psychology*, **101**, 553–60.
50. Blais, M.A. (1997) Clinician ratings of the five-factor model of personality and the DSM-IV personality disorders. *Journal of Nervous and Mental Diseases*, **185**, 388–93.
51. Parker, G., Hadzi-Pavloic, D., Wilhelm, K. (2000) Modeling and measuring the personality disorders. *Journal of Personality Disorders*, **14**, 189–98.
52. Mulder, R.T., Joyce, P.R. (1997) Temperament and the structure of personality disorder symptoms. *Psychologial Medicine*, **27**, 99–106.
53. Lynam, D.R. (1996) Early identification of chronic offenders: who is the fledgling psychopath? *Psychological Bulletin*, **120**, 209–34.
54. Seagrave, D., Grisso, T. (2002) Adolescent development and the measurement of juvenile psychopathy. *Law and human behavior*, **26**, 219–39.
55. Hart, S.D., Watt, K.A., Vincent, G.M. (2002) Commentary on Seagrave and Grisso: impressions of the state of the art. *Law and human behavior*, **26**, 241–5.
56. Widiger, T.A., Frances, A. (1985) Axis II personality disorders: diagnostic and treatment issues. *Hospital and Community Psychiatry*, **36**, 619–27.
57. Akiskal, H.S., Akiskal, K.K., Perugi, G., *et al.* (2006) Bipolar II and anxious reactive 'comorbidity': Toward better phenotypic characterization suitable for genotyping. *Journal of Affective Disorders*, **96**, 239–47.
58. Clark, L.A. (2005) Temperament as a unifying basis for personality and psychopathology. *Journal of Abnormal Psychology*, **114**, 505–21.
59. Ricciardi, J.N., Baer, L., Jenike, M.A., *et al.* (1992) Changes in DSM-III-R axis II diagnoses following treatment of obsessive-compulsive disorder. *American Journal of Psychiatry*, **149**, 829–31.
60. Rosenthal, D. (1970) *Genetic theory and abnormal behavior*. New York: McGraw Hill.
61. Tyrer, P. (2007) Personality diatheses: a superior explanation than disorder. *Psychological Medicine*, **12**, 1–5.
62. López-Ibor, J.J., Jr, Lana, F., and Sáiz, J. (1991) *Serotonin, impulsiveness and aggression in humans*. In Serotonin related psychiatric syndromes: clinical and therapeutic links (ed. G.B. Cassano and H.S. Akiskal), pp. 35–9. Royal Society of Medicine, London.
63. Sapolsky, R.M. (1990) Adrenocortical function, social rank, and personality among wild baboons. *Biological Psychiatry*, **15**, 862–78.
64. Manzano, J., Zabala, I., Borella, E., *et al.* (1992) Continuity and discontinuity of psychopathology: a study of patients examined as children and as adults. I. Antecedents of adult schizophrenic disorders. *Schweierz Archiv für Neurologie und Psychiatrie,* **143**, 5–25.
65. Lebow, J. and Gordon, K.C. (2006) You cannot choose what is not on the menu—obstacles to and reasons for the inclusion of relational processes in the DSM-V: comment on the special section. *Journal of Family Psychology*, **20**, 432–7.
66. Wakefield, J.C. (2006) Are there relational disorders? A harmful dysfunction perspective: comment on the special section. *Journal of Family Psychology*, **20**, 423–7.
67. Erickson, T.M. and Pincus, A.L. (2005) Using Structural Analysis of Social Behavior (SASB) measures of self- and social perception to give interpersonal meaning to symptoms: anxiety as an exemplar. *Assessment*, **12**, 243–54.
68. Beach, S.R., Wamboldt, M.Z., Kaslow, N.J., *et al.* (Eds.) (2006) *Relational Processes and DSM-V. Neuroscience, Assessment, Prevention, and Treatment*. American Psychiatric Publishing Inc., Washington.
69. World Health Organization (1992) *The ICD-10 International Classification of Mental and Behavioral Disorders. Clinical descriptions and diagnostic guidelines*. Geneva, WHO.
70. Widiger, T.A., Costa, P.T. Jr. (1994) Personality and personality disorders. *Journal of Abnormal Psychology*, **103**, 78–91.

71. Blashfield, R., Blum, N., Pfohl, B. (1992) The effects of changing axis II diagnostic criteria. *Comprehensive Psychiatry*, **33**, 245–5.
72. Janca, A., Kastrup, M., Katschnig, H., *et al.* (1996) The ICD-10 Multiaxial System for Use in Adult Psychiatry: Structure and Applications. *The Journal of Nervous and Mental Diseases*, **184**, 191–192.
73. Mezzich, J.E. (2007) Psychiatry for the Person: articulating medicine's science and humanism. *World Psychiatry*, **6**, 1–3.
74. Charney, D.S., Barlow, D.H., Botteron, K., *et al.* (2002) *Neuroscience research agenda to guide development of a pathophysiologically based classification system. In: A Research Agenda for DSM-V. Kupfer, D., First, M., Regier, D.A. (Eds.), pp:31–83*, American Psychiatric Association, Washington.
75. Van Praag, H.M. (1990). Two-tier diagnosing in psychiatry. *Psychiatry Research*, **34**, 1–11.
76. Kamer, P.D. (1993) *Listening to Prozac*. Viking, New York.

4.12.2 Diagnosis and classification of personality disorders

James Reich and Giovanni de Girolamo

Definitions of personality disorders

There has been considerable interest in the study of personality and personality disorder (PD) since early times and in many different cultures. However, as noted by Tyrer *et al.*[1] 'The categorization of personality disorder did not receive any firm support until the time of Schneider'. Schneider[2] regarded abnormal personalities as 'constitutional variants that are highly influenced by personal experiences' and identified 10 specific types or classes of 'psychopathic personality'. The classification system proposed by Schneider has deeply influenced subsequent classification systems[1]: of the 10 types of PD identified by Schneider, eight are closely related to similar types of PD as classified in DSM-III.[3] Many of these categories are also represented in DSM-IV[4] and ICD-10.[5]

Personality is defined in the second edition of the WHO *Lexicon of Psychiatric and Mental Health Terms*[6] as 'The ingrained patterns of thought, feeling, and behaviour characterising an individual's unique lifestyle and mode of adaptation, and resulting from constitutional factors, development, and social experience'. Personality disorders, according to the ICD-10 diagnostic guidelines[5]:

> . . . comprise deeply ingrained and enduring behaviour patterns, manifesting themselves as inflexible responses to a broad range of personal and social situations. They represent either extreme or significant deviations from the way the average individual in a given culture perceives, thinks, feels, and particularly, relates to others. Such behaviour patterns tend to be stable and to encompass multiple domains of behaviour and psychological functioning. They are frequently, but not always, associated with various degrees of subjective distress and problems in social functioning and performance.

For example, a dependent PD in a favourable environment might not cause dysfunction, but nevertheless might be considered a disorder since it is clinically identical to the same disorder that usually causes dysfunction.

DSM-IV[4] defines a PD as 'an enduring pattern of inner experience and behaviour that deviates markedly from the expectations of the individual's culture'. The pattern is manifested in two or more of the following areas: cognition, affectivity, interpersonal functioning, and impulse control. The pattern is inflexible and pervasive across a broad range of situations, has an early onset, is stable and leads to significant distress or impairment.

Personality traits, according to DSM-IV,[4] 'are enduring patterns of perceiving, relating to and thinking about the environment and oneself that are exhibited in a wide range of social and personal contexts. Only when personality traits are inflexible and maladaptive and cause significant functional impairment or subjective distress do they constitute PDs.'

ICD and DSM classifications of personality disorders

Table 4.12.2.1 lists the specific PDs as classified in ICD-9,[7] ICD-10, DSM-IIIR,[8] and DSM-IV.

In the ICD-10 classification, which does not have a multiaxial system for the separate recording of the personality status, PD can be diagnosed together with any other mental disorder, if present. Although a multiaxial system for ICD-10 is being developed, this will not include a separate axis for PDs, as in DSM-IV.

Despite the importance given to behavioural manifestations for the classification and assessment of PDs, personality traits and attitudes are also considered when a diagnosis is made. The ICD-10 diagnostic guidelines subdivide PDs 'according to clusters of traits that correspond to the most frequent or conspicuous behavioural manifestations'. As stressed by Widiger and Frances,[9] the reliance on behavioural indicators can improve inter-rater reliability, which reduces the amount of inferential judgement required for the diagnosis, but it does not ensure that the same diagnosis will be made at different times. Moreover, the diagnosis of a PD cannot be based on a single behaviour, as any given behaviour may have multiple causes (e.g. situational and role factors).

There have been four studies that have explored the diagnostic categories for PDs contained in ICD-10 and compared them with the DSM classification. The first[10] was carried out among 177 American clinicians who found some degree of overlap between the different categories. When the authors compared the diagnostic categories in ICD-10 with those in DSM-IIIR, they found that only anankastic (ICD) and obsessive–compulsive (DSM) PDs showed a high level of correspondence. The second study[11] looked at 52 outpatients and compared DSM-IIIR to ICD-10. It found fair concordance for the diagnosis of 'any PD', but poor agreement for individual PDs; the ICD-10 tended to overdiagnose PDs relative to DSM-IIIR. The third report[12] compared ICD-10 and DSM-IV in 58 patients with panic disorder. There was good agreement for the presence of 'any PD', and a reasonable agreement between individual diagnoses (κ ranged from 0.51 to 0.83.), with a tendency for ICD-10 to overdiagnose PDs relative to DSM-IV. In the fourth study,[13] ICD-10 criteria were found to have satisfactory inter-rater reliability in a sample of homeless adults.

In the American taxonomic system, a multiaxial classification was first introduced in DSM-III. With the development of DSM-IIIR, more than 100 changes in the classification of PDs were introduced compared with DSM-III.[14,15] While the multiaxial and categorical style of classification was maintained, the diagnostic criteria were revised to form a list of symptoms for each PD, of

Table 4.12.2.1 Comparison of different classification systems of personality disorders: ICD-9, ICD-10, DSM-IIIR, and DSM-IV

ICD-9	ICD-10	DSM-III-R	DSM-IV
Paranoid personality disorder	Paranoid personality disorder	Paranoid personality disorder	Paranoid personality disorder
Schizoid personality disorder	Schizoid personality disorder	Schizoid personality disorder	Schizoid personality disorder
Personality disorder with predominantly sociopathic or asocial manifestations	Dissocial personality disorder	Antisocial personality disorder	Antisocial personality disorder
	Emotionally unstable personality disorder:		
Explosive personality disorder	Impulsive type	NA	NA
NA	Borderline type	Borderline personality disorder	Borderline personality disorder
Histrionic personality disorder	Histrionic personality disorder	Histrionic personality disorder	Histrionic personality disorder
Anankastic personality disorder	Anankastic personality disorder	Obsessive-compulsive personality disorder	Obsessive-compulsive personality disorder
NA	Anxious [avoidant] personality disorder	Avoidant personality disorder	Avoidant personality disorder
NA	Dependent personality disorder	Dependent personality disorder	Dependent personality disorder
Affective personality disorder	Other specific personality disorders	Passive-aggressive personality disorder	NA
Asthenic personality disorder		Schizotypal personality disorder	Schizotypal personality disorder
		Narcissistic personality disorder	Narcissistic personality disorder
		Self-defeating personality disorder	NA
		Sadistic personality disorder	NA
			Personality disorder not otherwise specified

which only a certain number were required for a diagnosis to be reached. In DSM-IIIR, each category of PD comprised 7 to 10 criteria, with the presence of four to six criteria required for diagnosis. DSM-IIIR contained 11 PDs (see Table 4.12.2.1), plus two new disorders (self-defeating PD and sadistic PD) that were not included in DSM-III but were considered as diagnostic categories needing further study. As in DSM-III, the 11 PDs were divided into three clusters:

- cluster A (the 'odd' or 'eccentric' cluster), which included paranoid, schizoid, and schizotypal PD;
- cluster B (the 'dramatic' or 'erratic' cluster), which included histrionic, narcissistic, antisocial, and borderline PDs; and
- cluster C (the 'anxious' cluster), which included avoidant, dependent, obsessive–compulsive, and passive–aggressive PDs.

One study in the United States examined changes in personality diagnoses using DSM-III versus DSM-IIIR.[16] For some categories there was a considerable difference in the frequency of diagnosis: for example, there was an 800 per cent increase in the rate of schizoid PD and a 350 per cent increase in the rate of narcissistic PD diagnosed by the clinicians when DSM-IIIR criteria were applied.

DSM-IV was designed to be a conservative evolution from DSM-IIIR; however, some differences in diagnoses between DSM-IIIR and DSM-IV can be expected.[17] In general, the different DSMs should not be considered interchangeable unless there is specific data supporting agreement of a diagnosis across systems. As shown in Table 4.12.2.1, DSM-IV includes 11 PDs as in the DSM-IIIR classification; slight changes were introduced in the diagnostic criteria, and a new category 'PD not otherwise specified' added. Passive–aggressive, self-defeating, and sadistic PDs (provisionally included in DSM-IIIR) were dropped. The overall effect of these changes will be to increase the concordance between the DSM-IV and the ICD-10 classification systems compared with that between DSM-IIIR and ICD-10. DSM-IV also includes the three clusters present in DSM-IIIR.

Similarities differences between ICD-10 and DSM-IV

Table 4.12.2.1 shows that for seven categories of PD (paranoid, schizoid, dissocial/antisocial, histrionic, anankastic/obsessive–compulsive, anxious/avoidant, and dependent), there is a specific correspondence between ICD-10 and DSM-IV. For three categories, there are differences in nomenclature between the two systems; in particular ICD-10 uses the term 'anankastic' instead of 'obsessive–compulsive', to avoid the erroneous implication of an inevitable link between this type of personality and obsessive–compulsive disorder. ICD-10 also uses the term 'dissocial' instead of 'antisocial', to prevent any possible connotation of stigmatization, and the term 'anxious' instead of 'avoidant'. Moreover, while DSM-IV classifies borderline PD as a specific category, ICD-10 includes it as a subcategory of emotionally unstable PD. Narcissistic and passive–aggressive PDs (present in DSM-IV) are included in ICD-10 under the category of 'other specific PDs'. Finally, while DSM-IV includes schizotypal PD as a PD, ICD-10 classifies it in the overall group of 'Schizophrenia, schizotypal and delusional disorders', to highlight the contiguity between this disorder and the schizophrenia group disorders, as shown by genetic and clinical studies. DSM-IV has the category 'Personality disorders not otherwise

specified', while ICD-10 has the category 'Other specific personality disorders'.

Changes in the conceptualization of DSM personality disorders since the last edition of this chapter

Empirical research has advanced in the years following the original chapter in an earlier volume. These changes have impacted our understanding and use of DSM measurement instruments. These changes are that the personality disorders as described by DSM are not as enduring as we once thought. The instruments to measure the DSM PDs have modest agreement at best on the categorical level. Finally these instruments do not seem to adequately fit most of the disorders diagnosed which are diagnosed in the remainder category, 'Personality Disorder NOS'.

1. Research indicating lack of enduring quality of personality disorders.

There has now been considerable research indicating that some aspects of personality are state like. This was a line of research pursued by Reich[18,19] and later confirmed by others.[20] This means that some personality traits may disappear relatively rapidly—the state component. Experienced researchers have also found that even when personality disorders are selected for long-term study by careful methods, significant percentages of these will not be found on retest within a 6 month to several year periods.[21,22]

2. Modest agreement of personality DSM personality disorder instruments.

Research has been fairly consistent from the beginning of DSM instruments that while they may measure aspects of clinical relevance even well-designed instruments of the same design (self-report or semi-structured interview) are compared on categorical diagnoses, the agreements are usually modest at best.[23] (The results improve somewhat when using dimensional measures.) In addition when personality disorders are measured in a clinical population very few fit into the established categories and most wind up in the 'remainder bin' of personality disorder NOS.[24]

3. The problem of chasing changing criteria.

Each time the wording of a questionnaire is changed it can make significant difference in the outcome of the questionnaire or survey even though the changes may seem minor. This means that for every change in a version of the DSM, the DSM personality test developers would theoretically have to do extensive reliability and validity testing all over again. Unfortunately this is a time-consuming and expensive process which is beyond the resources of most DSM test developers. What happens instead is that either they do not update their instrument or update only the questions to conform to the new DSM criteria without doing new validity testing. Although these updated tests may have 'face validity' they (understandably) do not have extensive reliability and validity testing.

4. Conclusions from intervening research on DSM personality instruments.

At the end of the day we are left with the DSM personality measurement instruments being limited by the conceptualizations behind them which may be incorrect. These errors may either by in the enduring nature of the personality disorders and the nature of the proper categories or the nature of the specific criteria for a category. These measures then become good dimensional measures of various aspects of personality pathology and a rough clinical guide for 'disorder'. However, no one should expect that what one of these instruments is measuring is really the same as another. They do, however, conform to the DSM nomenclature.

Categorical versus dimensional styles of classification

In general, researchers involved in the assessment of personality traits tend to use dimensional measures based on normal populations. In contrast, those concerned with personality types and, even more, clinicians concerned with PDs, tend to employ categorical concepts and assessment measures based on these concepts.[15]

Each of the two approaches has specific advantages and disadvantages. The drawbacks of the categorical approach are represented by the difficulty of classifying patients who are at the boundary of different categories or who do not meet the diagnostic criteria for any specific PD, but who still have significant pathology. Other points that should be addressed include the wide variation of symptomatology found within each given category, the need for heterogeneous categories, such as 'mixed' and 'atypical', the need to simplify necessarily complex conditions, the need to define valid cut-off points, and the use of a nominal rather than an ordinal scale.

Those in favour of a dimensional approach argue that PDs differ from normal variations in personality only in terms of degree, and to some extent this is supported by empirical data.[25,26] However, there is evidence that some dimensional models do not account for all of the abnormal personality traits (see discussion on the 'Big five' models below.)

Moreover, it is still unclear whether normal and abnormal personality traits are the same or whether they are qualitatively different. Two main findings seem to support the latter hypothesis: first, normal personality traits are at least moderately heritable, while abnormal personality traits appear less heritable; second, the prevalence rates of PDs found in surveys are much greater than would be expected from the prevalence rates of normal personality traits in the general population. Some researchers[27] have even suggested that an extreme form of a normal personality trait is not necessarily pathological. Although some authors argue that there is no difference between normal and abnormal personality traits,[25] and this is still an area of open inquiry, it does appear more and more likely that there are some abnormal personality traits that differ from normal ones. It is possible that different models may be appropriate in different situations.

1. 'Big Five' personality measures and DSM-IV personality disorders

Since the first edition of this chapter there has been a fair amount of research using the 'big five' personality factors. There is evidence that the big five factors can account for some, but not all of the pathological personality pathology described by the DSM system.[28] However the amount of information also depends on the specific instrument with some with more facets or detail being more able account for aspects of pathological personality. These would be instruments such as the NEO-PI[29] and the five factor model of Cloninger.[30] Also of possible greater utility are dimensional instruments which were designed from the start to measure

abnormal as well as normal personality. These include the Schedule for Non-adaptive and Adaptive Personality (SNAP)[31] and Dimensional Assessment of Personality Pathology (DAPP).[32] Also worth noting is a method by Widiger which combines a five factor self-report with a five factor interview in order to obtain more comprehensive personality information.[33]

2. Other non-DSM personality measures from which DSM PD diagnoses can be derived.

There is one measure of DSM-IV personality which is derived from the interpersonal circumplex model. This is the Wisconsin Personality Disorders Inventory (currently in version IV).[34] Although from a fairly different theoretical perspective from the DSM-IV it does have translations to make some DSM-IV diagnoses.[35]

Another unique approach is the Shedler–Westen Assessment Procedure which does not have a patient interview but rather has a clinician who knows the patient do a Q sort procedure. The process is based on distinguishing personality disorders based on prototypes. This procedure can be translated into DSM personality diagnoses. It tends to create fewer diagnoses because the number of descriptors that can be used is limited by the Q sort procedure.[36]

The Millon Clinical MultiAxial Inventory (now in version three) is a very commonly used personality measure. It is now based on evolutionary theory according to its manual. One of the best validated instruments in data-based terms. Unfortunately this data is mostly published in its own manual and not the journal literature. As the population that the instrument was validated on included some fairly severe disorders it is not clear how well the findings would translate into a less severely ill population.[37]

Assessment methods for personality disorders

Table 4.12.2.2 shows the main methods currently available for assessing all PDs. Additional instruments for assessing specified PDs have also been developed—some of the methods listed are

Table 4.12.2.2 Some commonly used assessment methods for all DSM personality disorders

Name of the instrument	Author(s) [a]	Method of assessment	Number of Questions	Time required (minutes)
Diagnostic Interview for Personality Disorders (DIPD)	Zanarini	Semistructured interview with patient using DSM-IV criteria	398	60–120
Dimensional Assessment of Personality Disorders (DAPP-BQ)	Livesley & Jackson	Dimensional based of normal and abnormal personality. Has components of DSM-IV PDs	290	45
International Personality Disorder Examination (IPDE)	Loranger *et al.*	Semistructured interview with patient using ICD-10 and DSM-IV criteria	537	150
Millon clinical Multiaxial Inventory (MCMI)	Millon	Self-report by patient using DSM-IV criteria	175	20–30
NEO Personality Inventory-Revised (NEO PI-R)	Costa & McRae	Comprehensive measurement of normal and abnormal personality traits.	240	30–40
Personality Assessment Schedule (PAS)	Tyrer *et al.*	Semistructured interview with informant (s) can derive ICD-10 and DSM-IV diagnoses.	24	60
Personality Diagnostic Questionnaire –Revised (PDQ-4)	Hyler *et al.*	Self-report by patient or informant(s) using DSM-IV criteria	99	30
Personality Disorders Interview-IV	Widiger *et al.*	Semistructured interview with patient using DSM-IV criteria	325	60–120
Schedule for Nonadaptive and Adaptive Personality (SNAP)	Clark	Self report by patient using DSM-IV and dimensional criteria	375	30–60
Shedler-Westen Assessment Procedure	Westen & Shedler	Clinician rated Q sort procedure. Gives abnormal traits, symptoms and defences and can generate DSM-IV diagnoses	200	No interview, based of clinician knowledge of client.
Structured Clinical Interview for DSM-IV Personality Disorders (SCID-II)	First & Gibbon	Semistructured interview with patient using DSM-IV criteria	303	60–90
Structured Interview for DSM Personality Disorders-IV (SIPD)	Pfohl *et al.*	Semistructured interview with patient or informant(s) using DSM-III criteria	337	90
Temperament and Character Inventory (TPQ)	Cloninger *et al.*	Self-report by patient. Temperament and Character dimensions.	240	30–40
Wisconsin Personality Disorders Inventory (WISPI)	Klein *et al.*	Self-report by patient using DSM-III-R criteria. Correlated with interpersonal object relation theory.	214	30

[a] For specific references to each instrument please see references.[30, 31, and 33]

new, while others have been revised two or three times.[38,39] The following points related to these methods need to be mentioned.

1 The interview measures have generally shown a satisfactory inter-rater reliability, while test–retest reliability has not been well established. However, three methods do show some evidence of good test–retest reliability—the Personality Assessment Schedule (**PAS**),[40] the International Personality Disorder Examination (**IPDE**),[41] and the Structured Clinical Interview for Personality Disorders (**SCID-II**).[42] Many of the methods have been standardized on psychiatric inpatient or outpatient populations; their applicability in epidemiological community studies is largely unknown.[43,44]

2 The various measures tend not to agree with each other on specific diagnoses.[38,39] A measurement on one standardized instrument is not necessarily the same as the measurement on another.

3 Some authors, mostly developers of interview instruments, have in the past stated that self-report measures are not as valid for the measurement and study of personality as interview measures. These arguments tend to be based on the finding that self-report instruments do not agree well with interview instruments, and that a PD diagnosis cannot be made without a clinical interview. The first argument does not hold water, since none of the interview instruments agree well with each other either. Whether self-report instruments can reliably diagnose a PD (as opposed to personality traits) is an open question. However, dimensional interview instruments have high test–retest reliability (some for as long as 30 years) and have been, and will continue to be, a valuable component of personality research, especially where the focus is on dimensions.[45] Self-report instruments measuring DSM disorders tend to disagree with each other in a similar way to the interview instruments. In conclusion, the instrument chosen for any given clinical or research endeavour should reflect the ability of the individual instrument to meet the specific needs of the project. Most researchers now believe that using a self-report, which may be more sensitive but less specific is a good prelude to a semi-structured interview measure.

4 Standardized testing of personality and clinician impression do not tend be in good agreement. This is due both to the tendency of instruments to report more disorders and for clinicians to use idiosyncratic criteria in their diagnoses.[46,47] Standardized measures are a must, however, if the data is to be used for research or public policy purposes.

5 Most personality measurement instruments are affected by the comorbid presence of a non-personality emotional disorder (Axis I disorder in the DSM system). Some structured interviews try to correct for this by asking patients about times when they were not suffering from an Axis I disorder; however, when the Axis I disorder is chronic, this may be difficult to achieve.[41,42] This problem also affects the ability of questionnaires to differentiate between current Axis I disorders and PDs, as this self-judgement of patients who are suffering from a psychiatric condition is frequently impaired.[41] The PDE, PAS, and SCID-II may be less affected by this problem.

6 There is disagreement among experts about the use of informants. Many authors have argued that, besides the patient, a key informant should also be interviewed, given the likelihood that many patients will not reply reliably to questions about their personality and the possibility that informant ratings will differ substantially from patient ones.[41,48] However, even if an informant is interviewed, it is often unclear which source to use to score the test if the interviewee and informant disagree. This tends to reduce their value. More recent research indicates that informants may be of value if their information is in certain areas that are easily observable to then and then are incorporated into the evaluation process in a standardized way.[49]

7 Discriminant validity refers to the ability of a diagnostic system and measurement system to diagnose non-overlapping disorders. Discriminant validity is not high with the ICD-10 and DSM-IV PDs.[50] This means that it is the rule, rather the exception, that multiple personality diagnoses will be made: some studies have provided evidence of this assumption. For instance, in four studies that examined personality comorbidity in a total of 568 patients, the average percentage of multiple diagnoses was 85 per cent.[51] To what extent this reflects the real coexistence of different disorders—with distinct patterns of symptoms, correlates, and course—or if it is simply the effect of an insufficient discriminant validity of current diagnostic systems has still to be clarified.

Further information

Reich, J. (2005). State and trait in personality disorders. In *Personality disorders: current research and treatments* (ed. J. Reich), pp. 3–20. Taylor & Francis, New York.

Reich, J. (2007). State and trait in personality disorders. *Annals of Clinical Psychiatry*, **19**, 37–44.

Widiger, T.A., Costa, P.T., and Samuel, D.B. (2006). Assessment of maladaptive personality traits. In *Differentiating normal and abnormal personality* (2nd edn) (eds. S. Strack and M. Lorr), pp. 311–55. Springer, New York.

References

1. Tyrer, P., Casey, P., and Ferguson, B. (1991). Personality disorder in perspective. *British Journal of Psychiatry*, **159**, 463–71.
2. Schneider, K. (1923). *Die psychopathischen personlichkeiten*. Springer, Berlin.
3. American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd edn). American Psychiatric Press, Washington, DC.
4. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
5. World Health Organization. (1992). *International statistical classification of diseases and related health problems, 10th revision*. WHO, Geneva.
6. World Health Organization. (1994). *Lexicon of psychiatric and mental health terms* (2nd edn). WHO, Geneva.
7. World Health Organization. (1977). *International classification of diseases, injuries and causes of death, ninth revision*. WHO, Geneva.
8. American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd edn, revised). American Psychiatric Press, Washington, DC.
9. Widiger, T.A. and Frances, A. (1985). Axis II personality disorders: diagnostic and treatment issues. *Hospital and Community Psychiatry*, **36**, 619–27.

10. Blashfield, R.K. (1991). An American view of the ICD-10 personality disorders. *Acta Psychiatrica Scandinavica*, **82**, 250–6.
11. Sara, G., Raven, P., and Mann, A. (1996). A comparison of DSM-III-R and ICD-10 personality disorder criteria in an outpatient population. *Psychological Medicine*, **26**, 151–60.
12. Starcevic, V., Bogojevic, G., and Kelin, K. (1997). Diagnostic agreement between the DSM-IV and ICD-10-DCR personality disorders. *Psychopathology*, **30**, 328–34.
13. Merson, S., Tyrer, P., Duke, P., *et al.* (1994). Interrater reliability of ICD-10 guidelines for the diagnosis of personality disorders. *Journal of Personality Disorders*, **8**, 89–95.
14. Gorton, G. and Akhtar, S. (1990). The literature on personality disorders, 1985–1988: trends, issues and controversies. *Hospital and Community Psychiatry*, **41**, 39–51.
15. Widiger, T.A., Frances, A., Spitzer, R.L., *et al.* (1988). DSM-III-R Personality disorders: an overview. *American Journal of Psychiatry*, **145**, 786–95.
16. Morey, L.C. (1988). Personality disorders in DSM-III and DSM-III-R: convergence, coverage and internal consistency. *American Journal of Psychiatry*, **145**, 573–7.
17. Coolidge, F.L. and Segal, D.L. (1998). Evolution of personality disorder diagnosis in the diagnostic and statistical manual of mental disorders. *Clinical Psychology Review*, **19**, 583–99.
18. Reich, J. (2005). State and trait in personality disorders. In *Personality disorders: current research and treatments* (ed. J. Reich), pp. 3–20. Taylor & Francis, New York.
19. Reich, J. (2007). State and trait in personality disorders. *Annals of Clinical Psychiatry*, **19**, 37–44.
20. Clark, L.A. (2005). Stability and change in personality pathology: revelations of three longitudinal studies. *Journal of Personality Disorders*, **19**, 524–32.
21. Zanarini, M.C., Frankenburg, F.R., Hennen, J., *et al.* (2006). Prediction of the 10-year course of borderline personality disorder. *American Journal of Psychiatry*, **163**, 827–32.
22. Skodol, A.E., Gunderson, J.G., Shea, M.T., *et al.* (2005). The collaborative longitudinal personality disorders study (CLPS). *Journal of Personality Disorders*, **19**, 487–504.
23. Widiger, T.A. and Coker, L.A. (2002). Assessing personality disorders. In *Clinical personality assessment. Practical approaches* (2nd edn) (ed. J.N. Butcher), pp. 407–34. Oxford University Press, New York.
24. Verheuk, R. and Widiger, T.A. (2004). A meta-analysis of the prevalence and usage of the personality disorders not otherwise specified (PDNOS) diagnosis. *Journal of Personality Disorders*, **18**, 309–19.
25. Livesley, W.J., Schroeder, M.L., Jackson, D.N., and Jang, K.L. (1994). Categorical distinctions in the study of personality disorders: implications for classification. *Journal of Abnormal Psychology*, **103**, 6–17.
26. Schroeder, M.L. and Livesley, W.J. (1991). An evaluation of DSM-III-R personality disorders. *Acta Psychiatrica Scandinavica*, **84**, 512–59.
27. Birtchnell, J. (1988). Defining dependence. *British Journal of Medical Psychology*, **61**, 111–23.
28. Clark, L.A. and Harrison, J.A. (2001). Assessment instruments. In *Handbook of personality disorders: theory, research and treatment* (ed. W.J. Livesley), pp. 277–306. Guilford Press, New York.
29. Costa, P.T. and McCrae, R.R. (1992). *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI) professional manual*. Psychological Assessment Resources, Odessa, FL.
30. Cloninger, C.R., Przybeck, T.R., Svrakic, D.M., *et al.* (1994). *The temperament and character inventory (TCI): a guide to its development and use*. Center for Psychobiology of Personality, Washington University, St. Louis, MO.
31. Clark, L.A. (1993). *Schedule for adaptive and nonadaptive personality*. University of Minnesota Press, Minneapolis, MN.
32. Livesley, W.J. and Jackson, D.N. (2006). *Manual for the dimensional assessment of personality pathology*. Sigma, Port Huron, MI.
33. Widiger, T.A., Costa, P.T., and Samuel, D.B. (2006). Assessment of maladaptive personality traits. In *Differentiating normal and abnormal personality* (2nd edn) (eds. S. Strack and M. Lorr), pp. 311–55. Springer, New York.
34. Klein, M.H. and Benjamin, L.S. (1996). *The Wisconsin Personality Inventory-IV*. Wisconsin Psychiatric Institute, Madison, WI.
35. Smith, T.L., Klein, M.H., and Benjamin, L.S. (2003). Validation of the Wisconsin Personality Inventory-IV with the SCID-II. *Journal of Personality Disorders*, **17**, 173–87.
36. Westen, D., Shedler, J., and Bradley, R. (2006). A prototype approach to personality disorder diagnosis, *American Journal of Psychiatry*, **163**, 846–56.
37. Millon, T., Millon, C., Davs R., *et al.* (2006). *MCMI-III manual* (3rd edn). Dicandrien Inc., Minneapolis, MN.
38. de Girolamo, G. and Reich, J.H. (1993). *Personality disorders*. WHO, Geneva.
39. Reich, J.H. (1989). Update on instruments to measure DSM-III and DSM-III-R personality disorders. *Journal of Nervous and Mental Disease*, **177**, 366–70.
40. Tyrer, P., Casey, P., and Gall, J. (1983). Relationship between neurosis and personality disorder. *British Journal of Psychiatry*, **142**, 404–8.
41. Loranger, A.W., Sartorius, N., Andreoli, A., *et al.* (1994). The international personality disorder examination: the World Health Organization/alcohol, drug abuse, and mental health administration international pilot study of personality disorders. *Archives of General Psychiatry*, **51**, 215–24.
42. Ouimette, P.C. and Klein, D.N. (1995). Test-retest stability, mood-state dependence and informant-subject concordance in the SCID-Axis II questionnaire in a non clinical sample. *Journal of Personality Disorders*, **9**, 105–11.
43. Zimmerman, M. (1994). Diagnosing personality disorders. A review of issues and research methods. *Archives of General Psychiatry*, **51**, 225–45.
44. Jackson, H.J., Gazis, J., Rudd, R.P., *et al.* (1991). Concordance between two personality disorder instruments with psychiatric inpatients. *Comprehensive Psychiatry*, **32**, 252–60.
45. Clark, L.A., Livesley, W.J., and Morey, L. (1997). Special feature: personality disorder assessment: the challenge of construct validity. *Journal of Personality Disorders*, **11**, 205–31.
46. Blashfield, R.K. and Herkov, M.J. (1996). Investigating clinician adherence to diagnosis by criteria: a replication of Morey and Ochoa (1989). *Journal of Personality Disorder*, **10**, 219–28.
47. Perry, J.C. (1992). Problems and considerations in the valid assessment of personality disorders. *American Journal of Psychiatry*, **149**, 1645–53.
48. Dodwell, D. (1988). Comparison of self-ratings within informant-ratings of pre-morbid personality on two personality rating scales. *Psychological Medicine*, **18**, 495–502.
49. Ready, R.E., Watson, D., and Clark, L.A. (2002). Psychiatric patient- and informant-reported personality. *Assessment*, **9**, 361–72.
50. Bornstein, R.F. (1998). Reconceptualizing personality disorder diagnosis in the DSM-IV: the discriminant validity challenge. *Clinical Psychology: Science and Practice*, **5**, 333–43.
51. Widiger, T.A. and Rogers, J.H. (1989). Prevalence and comorbidity of personality disorders. *Psychiatric Annals*, **19**, 132–6.

4.12.3 Specific types of personality disorder

José Luis Carrasco and Dusica Lecic-Tosevski

Cluster A personality disorders

Paranoid personality disorder

Pervasive suspiciousness, mistrust, hypersensitivity to criticism, and hostility are the essential features of paranoid personality disorder. These individuals live a restricted emotional and interpersonal life because they fear the malevolent intent of others. As a rule, paranoid people are ready to counter-attack, provoking repeated confrontations. In this way, they induce hostility and resentment in others.

The term paranoia may lead to some confusion if it is not properly delimited. Paranoid had been used as an adjective to label various delusional representations or syndromes. Kraepelin[1] differentiated paranoia as a distinct condition characterized by chronic and highly systematized delusional ideas (see Chapter 4.4). Schneider[2] described people with this paranoid personality as fanatic psychopaths, stressing their intensity, and rigidity in confrontation with others. He denied any relationship with paranoia. Freud[3] and other psychoanalysts construed the paranoid character as a pattern of mistrust and feeling of being attacked, based on distortions and externalization of the person's inner world.

Paranoid personality disorder was included in DSM-III with criteria of suspiciousness, mistrust, hypersensitivity, and restricted affectivity. This last criterion does not appear in DSM-IV and ICD-10, since restricted affectivity is neither necessary nor specific for paranoid personalities. Instead, emphasis is placed on mistrust and sensitivity to setbacks. The DSM-IV criteria for paranoid personality disorder are shown in Table 4.12.3.1.

Table 4.12.3.1 DSM-IV diagnostic criteria for paranoid personality disorder

A. A pervasive distrust and suspiciousness of others such that their motives are interpreted as malevolent, beginning by early adulthood and present in a variety of contexts, as indicated by four (or more) of the following
1 Suspects, without sufficient basis, that others are exploiting, harming, or deceiving him or her
2 Is preoccupied with unjustified doubts about the loyalty or trustworthiness of friends or associates
3 Is reluctant to confide in others because of unwarranted fear that the information will be used maliciously against him or her
4 Reads hidden demeaning or threatening meanings into benign remarks or events
5 Persistently bears grudges, i.e. is unforgiving of insults, injuries, or slights
6 Perceives attacks on his or her character or reputation that are not apparent to others and is quick to react angrily or to counterattack
7 Has recurrent suspicions, without justification, regarding fidelity of spouse or sexual partner
B. Does not occur exclusively during the course of Schizophrenia, a Mood Disorder With Psychotic Features, or another Psychotic Disorder and is not due to the direct physiological effects of a general medical condition.

Note If criteria are met prior to the onset of Schizophrenia, add 'Premorbid', e.g. 'Paranoid Personality Disorder (Premorbid)'

(a) Epidemiology

The prevalence of paranoid personality disorder is estimated at about 0.5 to 1 per cent in the general population and at 10 to 20 per cent among psychiatric patients. The disorder is more commonly diagnosed in males.

(b) Aetiology

This personality disorder has a familial relationship with delusional disorders and with schizophrenia,[4] and has been included in the so-called schizophrenic spectrum.[5] Deficits in cortical dopamine activity may be associated with a poor conceptual organization that could in turn be responsible for suspiciousness and distorted interpretations.[6]

Mistrust and lack of confidence may reflect deficits arising in early developmental stages and resulting in a lack of basic self-confidence.[7] Lack of protective care and affective support in childhood could perhaps facilitate the development of paranoid features.

(c) Clinical picture

Paranoid individuals do not often ask for help from psychiatrists. They have no wish to be cured; instead, they believe that they have to be protected from other people's hatred and attacks. Subjects with this personality disorder suspect that others are acting to harm, exploit, or deceive them. These suspections are based not on objective evidence, but on internal conviction and an attempt to find a rational explanation for the supposed wrongs.

Paranoids are reluctant to confide in others; they tend to feel that others are plotting against them, and that the enemy may be found in unexpected places. They do not readily tell others about their suspicions. The disorder may be manifested by irritability, unusual defensive or self-protective behaviours (e.g. locking doors and closing windows and curtains to avoid being spied on, and hiding papers or documents), or emotional detachment.

Paranoid people lack confidence in others. They doubt the loyalty or trustworthiness of friends and partners, and check their behaviour repeatedly for evidence of malevolent intentions. They assume that others are not trustworthy, to the extent that they cannot believe it when friends demonstrate their loyalty. They withhold personal or significant information from friends, fearing that it will be used maliciously against them. They do not form close friendships and are often isolated. When in trouble, paranoids do not expect help from friends or others close to them; instead, they expect to be attacked or ignored.

Many of the suspicious and distrustful attitudes of paranoids are perpetuated by their intense interpersonal sensitivity. They react intensely to any comment or event that may relate to them. Hidden meanings that are demeaning and threatening may be read into benign events or the remarks of others. Unintended errors by colleagues or public servants are taken as deliberate attempts to harm or deceive them. Humorous remarks or jokes may be interpreted as attacks on their character. Paranoids are easily hurt, and their pride is easily damaged by minor critical comments or questioning. They are excessively preoccupied with attacks on their reputation or character, and minor slights may arouse major hostility and a

counter-attack. They bear grudges and harbour hostile feelings for a long time, and are unwilling to forgive the insults, injuries, or slights that they think they have received.[8]

Pathological jealousy is a common presentation of paranoid individuals. They have unreasonable doubts about the loyalty and faithfulness of their partners, based on little or no evidence. They may try to gather trivial and circumstantial facts to justify their beliefs. To avoid betrayal they attempt to gain complete control of intimate relationships, continuously questioning, and challenging partners about their whereabouts and intentions.

The interpersonal world of paranoids is a consequence of their suspiciousness and distrust. They have difficulty in relating to others, especially with close relationships. Hostility is always present and can be manifested as excessive argumentativeness, recurrent complaint and confrontation, or hostile aloofness.[8] Although they may appear rational, unemotional, and cold, the affect of paranoids is labile and oversensitive and they may be hostile, stubborn, and sarcastic. This mixture of secretive, cold, hostile, and sarcastic behaviours often elicits a hostile response in others, which confirms the paranoid person's beliefs.

Paranoids blame others for their shortcomings. They are querulous and quick to counter-attack, so that they may become involved in frequent litigation. Since they do not confide in others, paranoids need self-confidence and a sense of autonomy and independence. They need to control people who might be harmful. While they do not accept criticism, they are highly critical.

One group of paranoids are close to Schneider's 'fanatics'.[2] They have hidden grandiose fantasies of power and negative views of other people, especially those belonging to another group who come to be considered as natural enemies. They simplify issues and avoid any ambiguous perspective. Some form cults or other tightly knit groups with people who share their paranoid belief systems.

(d) Course

Paranoid features may be present in childhood and early adolescence in the form of hypersensitivity, social anxiety, poor peer relationships, and eccentricity. These features sometimes elicit teasing from other children, which in turn may aggravate the paranoid attitudes.

In situations of stress, individuals with paranoid personality disorder may respond with brief psychotic episodes. During these episodes, they may have frank delusional ideas or distorted perceptions. Some paranoid personality disorders are the premorbid state for a delusional disorder or even schizophrenia.

Individuals with this personality disorder may be at increased risk for agoraphobia, obsessive–compulsive disorder, and substance abuse or dependence. This personality disorder is often co-diagnosed with schizoid, schizotypal, narcissistic, and avoidant personality disorders.

(e) Differential diagnosis

Paranoid personality disorder should be distinguished from suspicious attitudes towards examination among immigrants, ethnic groups, or political groups. Members of these groups may display defensive and mistrustful behaviours owing to lack of familiarity with the language or the rules of a society, or in response to perceived neglect or rejection. Their behaviour may elicit further rejection from the majority, thus reinforcing the defensive behaviours.

Paranoid personality disorder is distinguished from delusional disorder, paranoid schizophrenia, and depression with psychotic symptoms, all of which are characterized by periods of persistent psychotic symptoms. Paranoid personality disorder present before the occurrence of these syndromes should be diagnosed as 'premorbid'.

People with schizotypal personality disorder are suspicious, have paranoid ideas, and keep their distance from others. However, they also experience perceptual distortions and magical thinking, and are usually odd and eccentric. Schizoid personality disorder is characterized by aloofness, coldness, and eccentricity, but these individuals usually lack prominent suspiciousness or paranoid ideation. Individuals with **avoidant personality disorder** are hypersensitive and do not confide in others. However, their lack of confidence is based on fear of being embarrassed or found inadequate rather than fear of other people's malicious intentions. Some antisocial behaviour by paranoid individuals originates in a wish for revenge or counter-attack, rather a desire for personal gain as in antisocial personality disorder. Paranoid features are often present in narcissistic individuals who fear that their imperfections could be revealed. The differential diagnosis should be based on the predominance of persistent need of praise versus persistent suspiciousness and distrust.

(f) Treatment

Antidepressant and anxiolytic treatment may be useful for anxiety and depression resulting from a paranoid response to stressful situations. Low-dose antipsychotics may be indicated during brief psychotic episodes or when ideas of reference are present.

Psychological treatment is difficult owing to the lack of insight. The approach is to attempt to gain the patient's confidence, avoiding early confrontation of distorted ideas, followed by a slow gentle attempt at cognitive restructuring.

Schizoid personality disorder

Schizoid personality disorder is characterized by a persistent pattern of social withdrawal. Schizoid individuals show discomfort in social interactions and are introverted. They are seen by others as eccentric, isolated, or lonely. DSM-IV diagnostic criteria are shown in Table 4.12.3.2.

This type of personality became recognized in the first two decades of the twentieth century. August Block's description of the shut-in personality and Eugen Bleuler's description of autism distinguished between shy and lonely persons and those who engage in relationships only in fantasy. Psychoanalysts included this term in their writings and developed an approach based on deficient object relations and individuation.[9] Some schizoid personalities have probably been sweet children who were very easy to care for, although giving less joy to their parents and eliciting less stimulation and fewer expressions of emotion than more expressive children.[7]

(a) Epidemiology

The epidemiology of schizoid personality disorder is not clearly established. Recent studies give a median prevalence of 0.5 to 1 per cent (see Chapter 4.12.5).

(b) Aetiology

A familial association may exist between schizotypal personality disorder and schizophrenia.

Table 4.12.3.2 DSM-IV diagnostic criteria for schizoid personality disorder

A.	A pervasive pattern of detachment from social relationships and a restricted range of expression of emotions in interpersonal settings, beginning by early adulthood and present in a variety of contexts, as indicated by four (or more) of the following
	1 Neither desires nor enjoys close relationships, including being part of a family
	2 Almost always chooses solitary activities
	3 Has little, if any, interest in having sexual experiences with another person
	4 Takes pleasure in few, if any, activities
	5 Lacks close friends or confidants other than first-degree relatives
	6 Appears indifferent to the praise or criticism of others
	7 Shows emotional coldness, detachment, or flattened affectivity
B.	Does not occur exclusively during the course of Schizophrenia, a Mood Disorder With Psychotic Features, another Psychotic Disorder, or a Pervasive Developmental Disorder and is not due to the direct physiological effects of a general medical condition

Note If criteria are met prior to the onset of Schizophrenia, add 'Premorbid.' e.g. 'Schizoid Personality Disorder (Premorbid)'.

(c) Clinical picture

People with schizoid personality disorder appear cold, reserved, distant, and unsociable. They lack involvement in everyday events and in the concerns of others. They rarely tolerate eye contact, usually give short answers, and appear uneasy when asked about emotions or feelings. However, they may invest much energy in abstract ideas such as those of mathematics or philosophy.

There is a characteristic lack of emotional expression and low energy. Speech is typically slow and monotonous, and seems to lack associated emotion. Affect is excessively serious or constrained, although some inner fear may be detected by an experienced clinician. If they try to be humorous, they usually give a child-like impression. Psychomotor activity tends to be lethargic, lacking gesture, and rhythmic movement. They may seem absorbed in insignificant matters, keeping quiet and not annoying anybody, as if in their own world. They do not express joy, anger, sadness, or other emotions. Interpersonal communication tends to be formal and impersonal, although not irrational. Threats, real or imagined, are dealt with by fantasized omnipotence or resignation. Aggressive acts are infrequent.

People with schizoid personality disorder characteristically seem to lack interest in the lives and concerns of others. When in a group, they stay unnoticed and detached, seeming indifferent to criticism or praise or to the reactions of others. Schizoids are attracted to solitary hobbies, and may be successful in lonely jobs that others find difficult to tolerate. Many prefer working at night. Usually, they do not seem to suffer because of this detachment and they have no desire for closeness or intimacy. They seldom have close friends or relationships, except with immediate relatives. Their sexual lives may be poor or exist only in fantasy, and some postpone mature sexuality indefinitely. They do not usually marry, although some, especially schizoid women, may passively agree to marriage. However, schizoid individuals may make emotional attachments with animals or inanimate objects.

Schizoid personalities lack insight, and generally have a poorly developed sense of identity and a poor capacity for evaluating interpersonal events. They may appear to be self-absorbed and engage in excessive daydreaming. However, some schizoid individuals have original and creative ideas.

(d) Differential diagnosis

Schizoids have better occupational functioning than patients with **schizophrenia** or **schizotypal personality disorder**, and, although isolated, can have successful careers. Schizophrenic patients exhibit delusional thinking or hallucinations and psychotic episodes. Schizotypal individuals show greater eccentricity and oddness than schizoids, and also have perceptual and thought disturbances including magical thinking.

People with **paranoid personality disorder** may also show social detachment and lack close relationships. However, they show more social engagement than schizoids and may have a history of aggressive behaviour.

Emotional constraint is also present in **obsessive–compulsive personality disorder**, but obsessional patients are more involved in everyday life and concerns, and may be worried by criticism. People with **avoidant personality disorder** are also detached and aloof. However, although they actively avoid interpersonal contact because of fear of rejection or being found inadequate, they have an intense desire for close relationships.

(e) Course

Schizoid personality disorder is usually apparent in early childhood. As with all personality disorders, it is usually long-lasting; however, it is not necessarily lifelong although there is seldom any rapid or profound change. If their deficits are moderate and social circumstances are favourable, some schizoids achieve social and vocational adaptation.

Although this personality disorder is sometimes a precursor of schizophrenia, the number of schizoid patients who go on to develop schizophrenia is unknown.

(f) Treatment

Because they lack insight and have little motivation for change, schizoids seldom seek treatment. Motivation for change may depend on life circumstances and external pressures.

Low-dose antipsychotic medication is useful in some situations. Antidepressants and psychostimulants have also been used with some positive effects.

The psychotherapy of patients with schizoid personality disorder must be based on gaining a therapeutic alliance. Unlike paranoid patients, they may become involved in therapy and reveal fantasies, imaginary friends, and fears of unbearable dependency. Ambivalence may appear because of fear of dependence on the therapist, who must keep the necessary distance to allow a tolerable relationship for the patient.

Social skills training is sometimes useful in improving their awareness of social cues.

Schizotypal personality disorder

Schizotypia is a controversial term in psychiatry. The term was used by Kretschmer[10] to denominate the phenotypic characters that antedated the development of schizophrenia. Nevertheless, the term schizotypal personality disorder was not included in psychiatric classifications until the publication of DSM-IIIR in 1987.[11] Before

Table 4.12.3.3 DV-IV diagnostic criteria for schizotypal personality disorder

A. A pervasive pattern of social and interpersonal deficits marked by acute discomfort with, and reduced capacity for, close relationships as well as by cognitive or perceptual distortions and eccentricities of behaviour beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following
   1 Ideas of reference (excluding delusions of reference)
   2 Odd beliefs or magical thinking that influences behaviour and is inconsistent with subcultural norms (e.g. superstitiousness, belief in clairvoyance, telepathy, or 'sixth sense'; in children and adolescents, bizarre fantasies or preoccupations)
   3 Unusual perceptual experiences, including bodily illusions
   4 Odd thinking and speech (e.g. vague, circumstantial, metaphorical. overelaborate, or stereotyped)
   5 Suspiciousness or paranoid ideation
   6 Inappropriate or constricted affect
   7 Behaviour or appearance that is odd, eccentric, or peculiar
   8 Lack of close friends or confidants other than first-degree relatives
   9 Excessive social anxiety that does not diminish with familiarity and tends to be associated with paranoid fears rather than negative judgements about self

B. Does not occur exclusively during the course of Schizophrenia, a Mood Disorder With Psychotic Features, another Psychotic Disorder, or a Pervasive Developmental Disorder.

Note If criteria are met prior to the onset of Schizophrenia, add 'Premorbid', e.g. 'Schizotypal Personality Disorder (Premorbid)'.

that date, schizotypal individuals were allocated either with schizoids or with schizophrenics, and were usually labelled as latent schizophrenics or pseudoneurotic schizophrenics. However, the validity of this nosological entity is still controversial and, despite its acceptance in DSM-IV, ICD-10 does not recognize it as a separate personality disorder. Instead, ICD-10 includes the schizotypal syndrome among the psychotic disorders and not as a personality disorder, based on the biological affinities of schizotypal individuals with other schizophrenic patients. DSM-IV diagnostic criteria are shown in Table 4.12.3.3.

(a) Epidemiology

Schizotypal personality disorder is present in 0.5 to 3 per cent of the general population, with no demonstrated differences between sexes. It is more commonly diagnosed in relatives of schizophrenic patients, and the incidence is much higher in monozygotic than in dizygotic twins (33 per cent versus 4 per cent).[4]

(b) Clinical picture

The essential feature of schizotypal individuals is a pattern of peculiarity and oddness in interpersonal relationships with resulting social detachment and lack of close relationships. Because of their distorted reality processing schizotypal individuals feel intensely uncomfortable in the presence of others. Conversely, others feel uneasy in the presence of schizotypals because of their unusual ways of thinking and expressing emotions.

Like schizoids, schizotypals have a decreased desire for intimate contacts, although they may sometimes express unhappiness about their lack of relationships. As a consequence they do not have close friends or confidants other than relatives. They experience intense anxiety in social situations with unfamiliar people. They can interact if necessary, but they prefer to keep aloof because they feel different and are not interested in the concerns of others. Their anxiety in these situations is not based on feelings of inadequacy or fear of humiliation. Rather, it is due to suspicion of the motivation of others, and therefore it is not alleviated as time passes and the situation becomes more familiar. Thus schizotypals feel progressively worse and more reluctant to confide in other people.

Individuals with schizotypal personality disorder often have ideas of reference that is interpretations of casual events as having specific and unusual meanings related to themselves. However, these ideas do not achieve the pathological conviction of delusions. Similarly, these individuals may be preoccupied with superstitions or paranormal phenomena. They may feel that they may read other people's thoughts or influence their behaviour by the power of thought. Their magical thinking is often manifested by ritualized behaviours aimed at avoiding harmful events.

Perceptual disturbances are frequent in schizotypal personality disorder. An experience of a sixth sense is typical, with the 'ability' to notice someone's presence. Distorted perceptions are present in the form of sounds perceived as calling voices or shadows transformed into figures and faces.

Thought processing and speech are characteristically affected. Speech may be constructed in an unusual and idiosyncratic way-generally loose, digressive, or vague, but without actual derailment or incoherence. Responses may be either excessively concrete or far too abstract, and words may be used in unusual ways.

The interpersonal relationships of schizotypal individuals are primarily affected by paranoid and suspicious ideation. They may believe that colleagues at work want to damage their reputation. In addition to the social anxiety of these individuals, this leads to a stiff and constricted contact and affect. They are considered odd and eccentric by others: they have peculiar mannerisms, dress in an unusual and unkempt manner, adopt extravagant postures and clothing combinations, do not obey normal social conventions, and generally avoid eye contact.

(b) Course

Schizotypal features may be present in childhood and adolescence in the form of solitariness, academic underachievement, hypersensitivity, and bizarre fantasies. Schizotypals do not seek treatment because of their personality disorder, but rather because of the presentation of associated depression, dysphoria, and anxiety. In response to stressful situations, these patients may experience transient psychotic episodes lasting from minutes to hours. In some cases, clinical symptoms and duration reach the degree of brief psychotic disorder, schizophreniform disorder, or schizophrenia, with the schizotypal personality disorder as the premorbid state. The prevalence of major depressive episodes is notoriously high, as is co-diagnosis with paranoid, schizoid, avoidant, and borderline personality disorders.

(c) Differential diagnosis

Delusional disorder, **schizophrenia**, and **mood disorder with psychotic symptoms** have to be excluded based on the greater intensity and persistence of psychotic symptoms.

In childhood, it can be difficult to distinguish schizotypal personality disorder from other forms of disorders characterized by odd behaviour, isolation, eccentricity, and peculiarities of

language. These include **autistic disorder**, **Asperger's disorder**, and some **language disorders**. The differentiation with communication disorders is based on the prominence of language symptoms in these children and the compensatory efforts to communicate by gesture and other means. Autism and Asperger's disorder present an even more intense social isolation and indifference, stereotyped behaviours and interests.

Paranoid and **schizoid personality disorders** lack the perceptual and speech impairment of schizotypal personality disorder, as well as the marked eccentricity and oddness. **Avoidant personality disorder**, while including social anxiety and isolation, differs from schizotypal personality disorder in that avoidants have an intense desire for closeness, which is constrained by fear of rejection. Schizotypals do not have a desire for relationships. **Borderline personality disorder** has a high rate of co-occurrence with schizotypal personality disorder and frequently the two disorders cannot be differentiated. Brief psychotic episodes in people with borderline personality disorder are more dissociative-like and generally follow affective shifts in response to stress or frustration. Social isolation in borderline personality patients is generally due to repeated interpersonal failures rather than a persistent lack of desire for relationships and intimacy.

Finally, schizotypal personality disorder must be diagnosed in the cultural context of the patient. It should be noted that some perceptual peculiarities and magical beliefs may be due to culturally determined characteristics. For example, mind reading, voodoo, shamanism, evil eye, and so on should not be considered as personality disorders in some cultural areas.

(d) Treatment

Low-dose antipsychotic medication may be useful for ideas of reference, perceptual disturbances, and other psychotic-like symptoms. Antidepressants are effective when depressive states are associated.

The psychological management of schizotypals should include a prolonged period of gaining the confidence of the patient. However, a particularly careful approach must be adopted owing to the peculiar thought processing of these patients.

Cluster B personality disorders

Antisocial personality disorder

Antisocial personality disorder is characterized by a pattern of disregard for the safety and the rights of others, without feeling remorse. Individuals with this disorder are unreliable, manipulative, incapable of lasting relationships, and unable to conform to social norms. The disorder starts early (before the age of 15), is pervasive, and manifests in variety of contexts. Although social deviance is one of the core features of antisocial personality disorder, it is not synonymous with criminality. Antisocial personality disorder uncomplicated by other disorders is not often met in clinical settings, except forensic psychiatry. However, owing to its impact on family and social environment, it has major public health significance and has been extensively studied in academic psychiatry, psychoanalysis, law, sociology, theology, and literature.

The description of antisocial personality in the last 1970s was mainly based on criminal behaviour[12] and the disorder was conceptualized as synonymous of criminality. Later classifications modified this approach and focused on the personality traits and emotional patterns described in classic descriptions included the classic personality traits leading to the current DSM-IV and ICD-10 classification criteria for antisocial personality disorder. (Tables 4.12.3.4 and 4.12.3.5).

Table 4.12.3.4 SM-IV diagnostic criteria for antisocial personality disorder

A. There is a pervasive pattern of disregard for and violation of the rights of others occurring since age 15 years, as indicated by three (or more) of the following 1 Failure to conform to social norms with respect to lawful behaviours as indicated by repeatedly performing acts that are ground for arrest 2 Deceitfulness, as indicated by repeated lying, use of aliases, or conning others for personal profit or pleasure 3 Impulsivity or failure to plan ahead 4 Irritability and aggressiveness, as indicated by repeated physical fights or assaults 5 Reckless disregard for safety of self or others 6 Consistent irresponsibility, as indicated by repeated failure to sustain consistent work behaviour or honour financial obligations 7 Lack of remorse, as indicated by being indifferent to or rationalizing having hurt, mistreated, or stolen from another
B. The individual is at least age 18 years
C. There is evidence of conduct disorder with onset before 15 years
D. The occurrence of antisocial behaviour is not exclusively during the course of Schizophrenia or a Manic Episode.

(a) Epidemiology

A prevalence rate of about 3 per cent is consistently found in the general population, and it is more frequent in males than females, with sex ratios ranging from 2:1 to 7:1. It is more common among younger adults, people living in urban areas and lower socio-economic groups.[13]

Table 4.12.3.5 ICD-10 diagnostic criteria for disocial personality disorder

Personality disorder, usually coming to attention because of a gross disparity between behaviour and the prevailing social norms, and characterized by (a) callous unconcern for the feelings of others (b) gross and persistent attitude of irresponsibility and disregard for social norms, rules and obligations (c) incapacity to maintain enduring relationships, though having no difficulty in establishing them (d) very low tolerance to frustration and a low threshold for discharge of aggression, including violence (e) incapacity to experience guilt and to profit from experience, particularly punishment (f) marked proneness to blame others, or to offer plausible rationalizations, for the behaviour that has brought the patient into conflict with society
There may also be persistent irritability as an associated feature. Conduct disorder during childhood and adolescence, though not invariably present, may further support the diagnosis.
Includes: amoral, antisocial, psychopathic, and sociopathic personality (disorder)
Excludes: conduct disorders, emotionally unstable personality disorder

(b) Aetiology

The aetiology of antisocial personality disorder is complex and multifactorial, involving biological, early developmental, and social determinants.

Twin, adoption, and family studies have demonstrated that genetic factors strongly contribute to the development of antisocial personality.[14] Antisocial personality in males is often associated with hysteria in women of the same family which suggests that the two conditions might be alternative expressions of the same genetic endowment, belonging to 'spectrum conditions'.[15] Longitudinal studies of hyperactive children have reported high rates of later (adult) antisocial behaviour, and have suggested a 'developmental' relationship between antisocial behaviour and childhood hyperactivity.

Aggression in antisocial personality disorder is associated with indexes of reduced brain serotonin activity such as low levels of the serotonin metabolite 5-hydroxyindole-acetic acid in the cerebrospinal fluid and[16,17] low platelet monoamine oxidase activity. Reports on minimal brain dysfunctions resulting on frontal-lobe deficiencies and lack of inhibition have also been described.

Parental deprivation, inconsistent maternal care, family violence, and severe childhood physical) abuse have been reported as strong predictors for development of antisocial personality disorders.[12,18]

Social disintegration and chronic criminality can cause episodic antisocial behaviour, reflecting a normal adaptation to an abnormal social environment.[19] However, the multifactorial origin of the antisocial personality disorder and its early onset and manifestations indicate that it cannot be attributed to cultural conflicts and social determinants.

(c) Clinical features and diagnosis

Patients with antisocial personality disorder often appear quite normal, charming, and understanding. However, their history reveals disturbed functioning in the domains of behaviour and self-concept, love and sexuality, interpersonal relations, and cognitive style.[20]

Reckless behaviour unaffected by punishment is typical of antisocial individuals, who are also exploitative, manipulative, demanding, and lacking in a sense of responsibility. An easy-going hedonistic attitude may be interrupted by rage, cruelty, and violence. The absence of internalized moral values is manifested by lying, truancy, running away from home, thefts, fights, substance abuse, and illegal activities may be typical experiences, beginning in early childhood.

An impaired control of impulses and a reduced ability to anticipate the negative consequences of behaviour is typical associated to a marked intolerance to anxiety. Antisocial individuals are egocentric, and unable to feel genuine guilt and remorse. They exhibit intense and persistent anger usually expressed as hostility towards others and they have an incapacity for reflective mourning or sadness. Frequent suicide threats and attempts are also common, as is somatic preoccupation.

Interpersonal relationships of antisocial subjects are characterized by manipulation, exploitation, instability and incapacity for love, and comprehension. Sexual perversions, abuse, and paedophilia are frequent. They display deficient parenting and social dysfunction, and resistance to authorities is pronounced.

The cognitive style of antisocial subjects is characterized by glibness, superficiality of knowledge, and paranoid view of reality.

(d) Comorbidity and differential diagnosis

Antisocial personality disorder is frequently comorbid with **depression**, which usually has atypical features. **Bipolar disorder** (manic phase) and **mental retardation** (learning difficulties) should be excluded. Substance abuse may be comorbid from childhood, and antisocial behaviour may be secondary to premorbid alcoholism type 2. **Atypical schizophrenic disorder** (pseudopsychopathic schizophrenia), **temporal-lobe epilepsy**, or a **limbic-lobe syndrome** should also be excluded.

The presentation of antisocial and criminal behaviour in borderline personality disorder is frequent. However, borderline behaviours are marked by intense affective instability and reactivity and may show some remorse or guilt. Unlike antisocial patients, borderline personality disorders do not lack the capacity for intimacy and emotional investment of others and do not show sadistic behaviours. Self-aggression and suicide attempts are much more prevalent among BPD than in antisocial personality.

Aggressive and defiant behaviours are often present in histrionic personality disorder. Although some aetiologic relationship among both disorders might be possible, as described above, histrionic patients are more impulsive and emotionally driven than antisocial and display intense emotions related with attachments and losses.

(e) Course and prognosis

Antisocial behaviour is most pronounced in early adult years, and gradually decreases with age. Professional motivation and establishing a stable couple or partnership may have beneficial effects. Maturation of the personality might also take with depression or hypochondriasis emerging when rage and aggression are abandoned. Substance abuse and promiscuity are risky behaviours for developing HIV infection.

(f) Treatment

Medication is used to deal with incapacitating symptoms, such as anxiety, rage, depression, and somatic complaints. Selective serotonin reuptake inhibitors, lithium, carbamazepine, clonazepam, and other anticonvulsants have been used to control aggressive behaviour but the effects are much less pronounced than in borderline personality disorder or intermittent explosive disorder. Psychostimulants such as methylphenidate may be useful if there is evidence of attention-deficit hyperactivity disorder. Benzodiazepines are contraindicated since they might cause behavioural disinhibition.

Efficacy of psychotherapy is very little in antisocial patients. Fear of intimacy causes difficulties in establishing a therapeutic alliance, which should be oriented to find alternative defence mechanisms to acting-out and to self-defeating behaviours. Therapeutic communities based on the principles outlined by Maxwell Jones[21] with a general social adjustment as a main task, might give positive results.

Borderline personality disorder

Borderline personality disorder (BPD) is the denomination of a syndromal picture characterized by intense affective instability and impulsivity together with an unstable sense of self-identity. It is often manifested by impulsive self-aggression and suicide attempts, substance abuse, chronic feelings of emptiness, and persistent pattern of severely unstable interpersonal relationships.

The term borderline was first used by Stern[22] in 1938 to denominate a group of syndromes placed in the border between neuroses and psychoses and included also the current label of schizotypal personality disorder and a group of disorders currently classified as psychotic disorders. Only some decades later the term borderline began to be understood as a disorder of character[23] and introduced in DSM-III as a personality disorder, after being separated from schizotypal personality disorder.

Borderline personality disorder derives but is not fully equivalent to the concept of borderline personality organization developed by Kernberg.[24] BPO is a stable permanent state based on three criteria: diffuse identity, primitive defence mechanisms (splitting, denial, and projective identification), and intact reality testing. This personality organization can be found not only in BPD but also in other severe personality disorders and Axis I conditions.

Borderline personality disorder itself can be found in association with so many Axis I and Axis II disorders that its validity as an independent diagnostic category is still weak compared with other personality disorders. Some authors have suggested that borderline personality disorder reflects rather a state of severely impaired personality function than a discrete diagnostic entity.[25] Others have suggested that BPD is an atypical variant of affective disorder and should be included in the affective disorder spectrum.[26] In the ICD-10, this disorder is named as 'emotionally unstable personality disorder', with two subtypes: impulsive and borderline. Borderline subtype is specifically linked to the presence of self-identity weakness and diffusion.

(a) Epidemiology

The number of people suffering from borderline personality disorder ranges from 1.5 to 5 per cent of general population with wide differences between studies because of lack of reliable measures. The prevalence is greater in clinical samples of patients at the outpatient clinics, ranging form 10 to 15 per cent. The disorder is more common in women than in men and is commonly initiated between 18 and 35 years old.[27]

(b) Aetiology

Several factors have been associated with a higher prevalence of borderline personality disorder, including genetic, biological, and developmental findings.[28] Family studies indicate that parents of patients with BPD have a greater incidence of mood disorders but not of schizophrenia. Additionally, there is also high family incidence of antisocial personality disorder and alcoholism.

Among the biochemical findings, those indicating a brain serotonin deficiency are the more consistent. Reduced levels of 5-hydroxyindoleacetic acid in cerebrospinal fluid and blunted prolactin response to serotonin agonists have been demonstrated in association with impulsive aggression, which is a core feature of BPD.[29] Hypothalamic-pituitary—adrenal axis dysfunctions, suggesting increased feedback inhibition, as well as increased sensitivity of some areas of the amigdala, have been reported in samples of BPD patients. Current available data suggest that BPD might be associated with abnormal emotional reactivity in the limbic areas and insufficient regulatory function at the cingulated and prefrontal areas of the brain.[30]

The role of childhood trauma in the development of borderline personality disorder could be crucial. Higher incidence of childhood traumatic experiences, both for sexual/physical abuse or for neglect, has been demonstrated in these patients.[31] Other proposed developmental factors include deficiencies in self and identity development linked to attachment failures with parental figures in the early developmental phases.[32,33]

The onset of BPD needs the interaction of predisposing factors, both biological and developmental, and environmental precipitants. BPD patients seem to be extremely sensitive to frustrations in the intimate relationships, which are commonly detected at the onset of the disorder.

(c) Clinical features and diagnosis

Impulsivity and affective instability, self-aggression, identity disturbance, and unstable/intense interpersonal relationships are the most characteristic manifestations of borderline personality disorder.

Identity weakness and diffusion explain several aspects of borderline personality disorder (Table 4.12.3.6). It is clinically manifested by contradictory character traits and sense of discontinuity of the self and feelings of emptiness.[34] Probably related with this is also the intolerance to be alone and the desperate efforts to avoid abandonment by significant others. The chronic feeling of emptiness is recurrently intensified and unbearable leading to drug abuse and self-defeating behaviours.

The affect of borderline patients is chronically dysphoric and irritated. Their unstable mood is a mixture of depressed affect, anger, loneliness, and emptiness. Impulsive–aggressive behaviour is a core feature of borderline personality disorder and is related with this abnormal affective state.

Cognitive style of borderline patients are easily suggestible and frequently change their decisions. Things and people are seen in black-and-white terms. Transient and brief psychotic episodes are frequent in BPD patients associated with unstructured stressful

Table 4.12.3.6 DSM-IV diagnostic criteria for borderline personality disorder

A pervasive pattern of instability of interpersonal relationships, self-image, and affects and marked impulsivity beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following

1. Frantic efforts to avoid real or imagined abandonment. **Note:** Do not include suicidal or self-mutilating behaviour covered in Criterion 5
2. A pattern of unstable and intense interpersonal relationships characterized by alternating between extremes of idealization and devaluation
3. Identity disturbance: markedly and persistently unstable self-image or sense of self
4. Impulsivity in at least two areas that are potentially self-damaging (e.g spending, sex, substance abuse, reckless driving, binge eating). **Note:** Do not include suicidal or self-mutilating behaviour covered in Criterion 5
5. Recurrent suicidal behaviour, gestures, or threats, or self-mutilating behaviour
6. Affective instability due to a marked reactivity of mood (e.g. intense episodic dysphoria, irritability, or anxiety usually lasting a few hours and only rarely more than a few days)
7. Chronic feelings of emptiness
8. Inappropriate intense anger or difficulty controlling anger (e.g. frequent displays of temper, constant anger, recurrent physical fights)
9. Transient stress-related paranoid ideation or severe dissociative symptoms

situations. Psychotic symptoms may have a typical dissociative-like nature or present as transient self-referential ideation. Rejection sensitivity and suspiciousness usually colours the interpretations of behaviours of others.

Borderline patients are both intensely dependent and hostile towards significant others. Interpersonal relationships are unstable, intense, demanding, clinging, and characterized by alternation between extremes of idealization and devaluation, deriving from the defence mechanism of splitting. Infatuations are followed by devaluation of love objects. There is a tendency towards promiscuity and perversions.

(d) Comorbidity and differential diagnosis

Borderline personality disorder is frequently comorbid with **affective disorders** (major depression, dysthymia, and 'double depression'), **anxiety disorders**, **somatization disorder**, **post-traumatic stress disorder**, and **alcohol abuse**.

Differential diagnosis has to be made with **type II bipolar disorder**. Bipolar patients more often present emotional lability from sadness/apathy to euphoria while BPD patients are characterized by intense and reactive affective instability and shift rather from sadness to tolerable dysphoria. Intermittent explosive disorder also shows impulsive and aggressive behaviours but lacks identity disturbances and affective instability typical of BPD. Mild mental retardation might have intense irritability, lability, and impulsive/aggressive behaviour but lacks chronic feelings of emptiness and self-identity diffusion. Transient psychotic episodes and stress-related referential ideas of BPD should be differentiated from pervasive psychotic-like experiences of schizotypal personality disorder.

Borderline personality disorder has been shown to be associated with most personality disorders, especially with those from the dramatic cluster. The high prevalence of comorbid personality disorders may result from overlapping of diagnostic criteria or reflect the confirmation that there is the underlying borderline personality organization of all severe personality disorders. However, some features like chronic feelings of emptiness, self-mutilation, short-lived psychotic episodes, intense and episodic drug abuse, and intense ambivalent dependency in close relationships suggest a primary diagnosis of BPD.

(e) Course and outcome

Borderline patients often experience profound dysfunction in many important aspects of life including education, jobs, partner relationships, and marriage. Alcohol and psychosexual problems are also frequent. Repeated suicide attempts and premature death from suicide are frequent complications of borderline personality disorder; therefore suicidal gestures and intentions should be always taken seriously. It has been reported that 8 to 10 per cent per cent of all persons with borderline personality disorder commit suicide.[35]

The long-term outcome of borderline patients has not been studied, but the diagnosis is rarely made in patients aged over 40. It is speculated that neural structures and defence mechanisms mature with age and that these changes, together with social learning, reduce symptomatology.

(f) Treatment

Pharmacotherapy is targeted to symptoms such as affective changes (depression, anxiety, rage, dysphoria), cognitive disturbances (brief psychotic episodes or interpretative distortions), and impulsive behavioural dyscontrol.[36] New antidepressants, including SSRIs and venlafaxine, have shown positive effects in treating a broad spectrum of acute symptoms, including depression, hostility, irritability, anxiety, obsessive–compulsive symptoms, suicidal attempts, and impulsivity. Antipsychotics and anticonvulsants may help some patients, even in the absence of EEG or organic changes. There are still no clinical predictors for efficacy, therefore, a pragmatic approach is indicated with patients being treated with two or three drugs in a sequential trial.[36] Suicidal and abusive use of drugs prescribed and non-compliance may be serious problems for treatment of BPD.

Various psychotherapeutical modalities are used, including psychodynamic psychotherapy, supportive psychotherapy, and dialectical–behavioural therapy.

A more structured form of psychoanalytic approach, involving expressive psychotherapy and an active role of psychotherapist have been proposed specifically for BPD. The aims are confrontation of maladaptive defences and interpretation of transference, focusing on the 'here and now', without attempting the achievement of a full genetic reconstruction.[37]

Short-term psychotherapy is useful for managing crises or introducing long-term forms of therapy. Supportive psychotherapy is suggested for more fragile borderline patients, who are prone to serious regression in treatment. In practice, supportive therapy, with a psychoeducational component, has been the most frequently used form of treatment for borderline personality disorder. It is also possible to combine elements of intensive dynamic therapy with supportive therapy, depending on the ego strength of the patients.[38]

Dialectical–behavioural therapy[39] is based on cognitive techniques associated with reality confrontation. The major aim is emotional self-regulation and behavioural self-control and is particularly indicated for control of suicidal and impulsive behaviours and treatment of emotional reactivity.

Histrionic personality disorder

(a) Definition

Histrionic personality disorder derives from the concept of hysterical personality, supported by descriptive literature and clinical tradition but not so much by valid empirical research. It is characterized by excessive emotionality and attention seeking, and by dramatic, colourful, and extroverted behaviour. Egocentric, dependent, and demanding interpersonal relationships are typical of this disorder, which begins in early adulthood and is present in a variety of contexts. The DSM-IV and ICD-10 diagnostic criteria for this disorder are shown in tables 4.12.3.7 and 4.12.3.8.

It was included in scientific medicine by Kraepelin,[40] who described multiple symptoms, including capricious and inconsistent behaviour, histrionic exaggeration, and a life of illness, which captured the core pattern of the illness. Freud[41] recognized the relationship between hysterical neurosis and what he called the 'erotic personality, whose major goal in life is the desire to love or above all to be loved'. Psychoanalytic theorists often distinguish hysterical ('healthier') and histrionic ('sicker') personalities, where the latter has borderline organization and is an exaggeration of the former. Differences between the two concepts are shown in Table 4.12.3.9.

Table 4.12.3.7 DSM-IV diagnostic criteria for histrionic personality disorder

A pervasive pattern of excessive emotionality and attention seeking, beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following

1 Is uncomfortable in situations in which he or she is not the centre of attention
2 Interaction with others is often characterized by inappropriate sexually seductive or provocative behaviour
3 Displays rapidly shifting and shallow expression of emotions
4 Consistently uses physical appearance to draw attention to self
5 Has a style of speech that is excessively impressionistic and lacking in detail
6 Shows self-dramatization, theatricality, and exaggerated expression of emotion
7 Is suggestible, i.e. easily influenced by others or circumstances
8 Considers relationships to be more intimate than they actually are

(b) Epidemiology

The prevalence is found to be 2 to 3 per cent of the general population and 10 to 15 per cent in clinical settings.[42] No significant difference has been found in terms of race and education and is more frequently diagnosed in women than in men. However, some cultural biases associated with sex role stereotypes and emotional expressiveness could lead the lower diagnostic rates in men.[43]

(c) Aetiology

Some studies suggest that histrionic personality runs in families, but evidence for a biological or learning transmission is not yet consistent.[44] Traits such as extraversion, emotional expression, and reward dependence have a strong genetic origin and might be constitutional. Biological findings associated with impulsivity, such as serotonin deficiency, can be found in histrionic patients with marked emotional instability and impulsive behaviours. It has been proposed that histrionic personality in women is genotypically linked to antisocial personality in men.[45]

From a development perspective, histrionic personality is considered to be a result of abnormally intense attachment with parental figures. The erotization component of this attachment in the oedipical phase was classically emphasized in the psychoanalytic research, although recent approaches suggest that there oral/dependent factors, derived from anomalies in earlier phases are of greater importance in the development of the disorder.[46]

Table 4.12.3.8 ICD-10 diagnostic criteria for histrionic personality disorder

Personality disorder characterized by at least three of the following:

(a) self-dramatization, theatricality, exaggerated expression of emotions
(b) suggestibility, easily influenced by others or by circumstances
(c) shallow and labile affectivity
(d) continually seeking for excitement, appreciation by others, and activities in which the patient is the centre of attention
(e) Inappropriate seductiveness in appearance or behaviour
(f) Overconcern with physical attractiveness

Associated features may include egocentricity, self-indulgence, continuous longing for appreciation, feelings that are easily hurt, and persistent manipulative behaviour to achieve own needs

Includes: hysterical and psycho-infantile personality (disorder)

Table 4.12.3.9 Types of histrionism

Hysterical personality	Histrionic personalitya
Neurotic personality organization	Borderline personality organization
Integrated identity	Diffuse identity
Predominance of repression	Predominance of splitting
Intact reality testing	Intact reality testing (proneness to distortion)
Integrated superego	Marked superego defects
Strongly bonded families	Disturbed, often broken families
Steady educational and vocational careers	Erratic careers
Capable of maintaining long-term friendships	Chaotic interpersonal relationships
Suggestible in triangular relationships	Diffuse suggestibility
Inauthenticity	Multiple identifications
Changing moods	Frequent dysphoria
Sexual inhibition	Promiscuity, perverse tendencies
Competitiveness with the same sex	Less differentiated behaviour toward sexes
Genital traits	Oral/pregenital traits

[a] Includes hysteroid, hysteriform, oral hysteric, and sick hysteric personality disorders. After Akhtar.[31]

(d) Clinical features and diagnosis

Emotionality, dramatization, exhibitionism, egocentricity, and sexual provocativeness are typical of histrionic personalities. However, behavioural expression is not always as manifested and other emotional and cognitive aspects may help for diagnosis.

Histrionic individuals are inappropriately seductive and aggressively demanding. They are self-centred, crave for novelty and excitement, and are prone to temper tantrums. Histrionic subjects are hyperemotional and impulsive, but their emotional enthusiasm is superficial and transient and their mood is labile. They describe their emotions in an inappropriate and exaggerated way in an attempt to obtain attention. Histrionic individuals are suggestible, demanding, accusative, and guilt inducing. In intense stressful situations they can show a dissociative-type of indifference and infantilism.

Histrionic personalities are inclined to sexualize all non-sexual relations, often indiscriminately, not only with a chosen partner but also with a wide variety of persons in various social, occupational, and professional relationships. Pseudosexuality is often accompanied by frigidity. A romantic outlook or a superficially adoring attitude often disguises needs for dependency and emotional attachment to a significant protective figure. Sicker individuals may be promiscuous, and may engage in multiple perverse activities.

Cognitive style is global, impressionistic, and diffuse, and lacks sharpness of detail. Non-verbal communication is better than verbal, speech is inhibited, and education is often superficial. Speech may show malapropisms or slips of the tongue.

The basic belief of histrionic personalities is that others should be impressed, and their basic strategy is dramatization. They blossom when they are the centre of attention and are highly disappointed when they are not, and draw attention to themselves by acting and speaking in a charming flirtatious way. Histrionic individuals quickly respond to others in an intimate way, often treating superficial acquaintances as if they were friends.

(e) Comorbidity and differential diagnosis

There is current evidence that **somatization disorder** (Briquet's syndrome) and **conversion disorder** can occur in conjunction with histrionic personality disorder, as well as **dissociative disorder** and **brief reactive psychosis**.[21] Differential diagnosis should not be difficult, because histrionic personality disorder is a lifelong disturbance with a chronic course, unlike Axis I disorders which are episodic. **Hypomanic and manic states** may be accompanied by seductive behaviour and exaggerated expression of emotions, but can be distinguished from histrionic personality by their episodic nature and the presence of other characteristic symptoms.

A great deal of overlap has been found between histrionic personality disorder and other **Axis II disorders**, defined by DSM-IIIR criteria; of these, the borderline, narcissistic, antisocial, and dependent personality disorders are the most frequent.

Borderline patients have more chaotic interpersonal relationships, make frequent suicide attempts, and are prone to regressive episodes of a psychotic nature. Histrionic individuals share sexual promiscuity, corruptibility, shallow emotions, and a self-centred attitude with antisocial personalities.[47] However, they do not show sustained, calculated, and ruthless disregard of social norms. Narcissists may also seek attention, but they want to be admired for their superiority while histrionic persons are clinging and dependent. Unlike narcissists, histrionic individuals have empathy for other persons. However, the features of the two disorders can be combined.

(f) Course and prognosis

Depressive symptoms, suicide attempts, and frequent use of medical services are common. Histrionic personality may gradually improve with age, as if a maturation of histrionic infantilism occurs over the years.

(g) Treatment

Depressive and anxious symptoms are frequent in histrionic personality disorder and can be alleviated with the use of antidepressants and anxiolytic medications. However, extreme care should be taken for treatment due to the vulnerability of these patients for medication abuse and non-compliance.

Supportive therapy is indicated for acutely distressed histrionic patients, as well as for those at the sicker end of the continuum. Psychoanalytic techniques in histrionic patients[47] are oriented to clarification of the patient's covert inner feelings. Patients are often demanding, want to take a special place in the therapist's life, and act out during therapy sessions, threatening to abandon treatment or undertake dangerous actions. Clear limits should be set and demanding dependent behaviour should not be rewarded.

Narcissistic personality disorder

Narcissistic personality disorder is characterized by an exaggerated sense of self-importance with a lack of sustained positive regard for others. Grandiosity (in fantasy or behaviour) and constant craving for admiration and external gratification are additional features of this disorder. They are present in a variety of contexts and begin by early adulthood.

(a) Historical perspective

The term narcissism originates from the Greek myth of Narcissus who was infatuated with his own reflection in the mirror-lake. Its contemporary usage has many meanings and implications, from its colloquial usage denoting self-centred persons, often with pejorative connotations, to a pathological clinical syndrome. Despite the popularity of the construct, there is still considerable disagreement on the aetiology and phenomenology of narcissistic personality disorder. There is little empirical evidence regarding its description, clinical utility, and validity.

A narcissistic personality has a pathological grandiose self, which hides a diffuse and aimless inner identity. Kernberg argues that self-hatred, rather than self-love, lies at the root of pathological narcissism, and distinguishes between narcissism in the broad sense and the specific pathological structures of the narcissistic personality. According to Kernberg, narcissistic patients function on a borderline level. Malignant narcissism,[37] which develops when primitive aggression infiltrates the pathological grandiose self, lies at the extreme end of a continuum. It is a combination of narcissistic personality disorder, antisocial behaviour, egosyntonic aggression or sadism directed against others, and a strong paranoid orientation (Table 4.12.3.10).

Narcissistic personality disorder was officially accepted in DSM-III. Somewhat refined criteria were adopted in DSM-IV (Table 4.12.3.11), because some studies showed a substantial lack of diagnostic reliability when the DSM-III criteria were used. Narcissistic personality disorder is not included in ICD-10, being mentioned only in the category 'Other specific personality disorders'.

Table 4.12.3.10 Types of narcissism

Normal	Pathological	Malignant
Infantile	Grandiose self-image	Grandiose self-image
Regression or fixation to infantile narcissistic goals in personality disorders (personality traits)	Low self-esteem Primitive defences Superego defects Borderline organization	Aggression Paranoid traits Antisocial behaviour Explosive traits Borderline organization
Adult		
Healthy self-esteem regulated by normal self-structure		
Integrated object representations		
Capacity for deep object relations		
Integrated superego		

After Kernberg.[32,37]

Table 4.12.3.11 DSM-IV diagnostic criteria for narcissistic personality disorder

A pervasive pattern of grandiosity (in fantasy or behaviour), need for admiration, and lack of empathy, beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following

1 Has a grandiose sense of self-importance (e.g. exaggerates achievements and talents, expects to be recognized as superior without commensurate achievements)
2 Is preoccupied with fantasies of unlimited success, power, brilliance, beauty, or ideal love
3 Believes that he or she is 'special' and unique and can only be understood by, or should associate with, other special or high-status people (or institutions)
4 Requires excessive admiration
5 Has a sense of entitlement, i.e. unreasonable expectations of especially favourable treatment or automatic compliance with his or her expectations
6 Is interpersonally exploitative, i.e. takes advantage of others to achieve his or her own ends
7 Lacks empathy, i.e. is unwilling to recognize or identify with the feelings and needs of others
8 Is often envious of others or believes that others are envious of him or her
9 Shows arrogant haughty behaviours or attitudes

(b) Epidemiology

The prevalence of narcissistic personality disorder in the community has been found to be around 0.5 per cent.[48] Its prevalence in clinical populations is estimated to range from 1 to 3 per cent, and is more frequently diagnosed among males.

(c) Aetiology

Although no demonstrating evidence is yet available, some aspects of narcissism might be related with inappropriate seeking for excitement and reward and associated to monoamine function abnormalities at the mesolimbic reward systems.

Severe frustration with early objects is considered important in the defensive genesis of narcissistic personality disorder. Behind the compensatory grandiose self, a hungry and inferior real self-resides, as the core problem of narcissistic personality disorder. Often, high parental expectations and harsh criticism of the child is present in the family.

(d) Clinical characteristics and diagnosis

Narcissistic patients only seek for help when depressed or involved in interpersonal problems.

An often engaging and attractive appearance masks intense preoccupation with self-regard and an unusual absence of concern for others. Narcissistic individuals may be energetic, capable of consistent work, and socially successful, but this is done in order to obtain admiration. Their successes provide no inner satisfaction and always end with frustration and a feeling of emptiness. Narcissistic grandiosity is often masked by opposing tendencies (false modesty, social aloofness, and a pretended contempt for status). Pathological lying is frequent.

Narcissistic individuals feel bored when the external glitter wears off and there are no new sources to feed their self-esteem, which is extremely fragile. Lacking emotional depth, they do not have genuine feelings of sadness or longing. Anger and resentment laden with vengeful wishes are frequent as a reaction to injured self-esteem. Chronic intense envy is present, as are defences against such envy, particularly devaluation, omnipotent control, and narcissistic withdrawal. They have frequent mood swings, and hypomanic exaltation is often part of the clinical picture.

Narcissistic persons are unable to fall in love, and only have fantasies of ideal love. Sexuality is trivialized, and intercourse is a purely physical pleasure. Promiscuity, perverse fantasies, devaluation of objects, and boredom in relationships are frequent.

Interpersonal relationships are frequently manipulative and exploitative. They idealize people whom they expect to feed their narcissism, but depreciate and treat with contempt others (often former idols) from whom they do not expect to receive anything. They lack empathy and concern for others, who are welcome only as an applauding crowd and as mirrors of success.

Typical defence mechanisms (omnipotence, omniscience, intellectualization, rationalization, idealization, and devaluation) are derived from splitting.

(e) Comorbidity and differential diagnosis

Narcissistic personality disorder is often comorbid with major depression, dysthymic disorder, substance abuse, and anorexia nervosa. Patients meeting criteria for narcissistic disorder have a high overlap with **histrionic, borderline, and antisocial personality disorders**, and also with **schizotypal, paranoid, and passive–aggressive personality disorders**.

Narcissistic personality disorder may display some features of bipolar disorder (manic and hypomanic episodes). However, the mood swings are of limited duration and change rapidly, while insight is maintained and the general integrity of the personality is preserved.

Narcissistic personality disorder is strikingly similar to borderline personality disorder. Phenomenologically, grandiosity is the best discriminator between the two disorders.[49] In narcissistic personality disorder, there is also better impulse control, greater social adjustment and anxiety tolerance, less frequent suicide attempts, and less danger of regressive fragmentation and psychotic episodes.

Narcissistic individuals, especially those manifesting malignant narcissism, may demonstrate antisocial behaviour. However, antisocial individuals are more impulsive and less capable of concentrating on work and career, and they are devoid of guilt feelings. Similarities with histrionic and obsessive–compulsive personalities are superficial, since people with these disorders have a capacity for empathy and a concern and love for others.

(f) Course and prognosis

Patients often become depressed or defensively hypomanic during middle age, when their internal life gradually deteriorates owing to a vicious circle of frustrations and disappointments and diminishing narcissistic supplies. Hypochondriasis and anxiety disorders are frequent complications.

(g) Treatment

Anxiolytic agents and antidepressants may be helpful for alleviating target episodes of mood and anxiety symptoms.

(i) Psychotherapy

Individual psychotherapy is aimed to the analysis of idealizing transference and interpretation of self-grandiosity. However, during

the first stages only supportive therapy is recommended with interpretations delayed until confident and integrated attachment with therapist is achieved.

The treatment of narcissistic individuals inevitably arouses serious countertransference problems, because of the emotional detachment, demanding behaviour, and devaluative actions of narcissistic patients. The therapist should have worked through his or her own narcissism and retain an empathic and non-judgemental attitude.

Cluster C personality disorders

Avoidant personality disorder

Avoidant personality disorder was first introduced into psychiatric classification in DSM-III.[11] Before this, such patients were included among the schizoid or dependent patients. The emphasis on avoidant behaviour as a consequence of an intense sensitivity to rejection led to the differentiation of this new personality type.

The characteristic behaviour of the avoidant personality is active isolation from the social environment. Unlike schizoids, who are characterized by passive social isolation, avoidant subjects turn inward to protect themselves because they are extremely sensitive to rejection.[50] They desire interpersonal relationships but they avoid any chance of disapproval. Thus, only relationships that are likely to lead to complete non-critical acceptance are established. The extreme sensitivity to criticism is based on intense feelings of inferiority, poor self-concept and low self-esteem. This disorder is termed anxious personality disorder in ICD-10, since anxiety is considered to be the basic affective feature. The DSM-IV criteria for avoidant personality disorder are shown in Table 4.12.3.12.

(a) Epidemiology

The prevalence of avoidant personality disorder is estimated to be less than 1 per cent in the general population, but almost 10 per cent in clinical populations. No differences between sexes are found.

(b) Aetiology

Some familial transmission is possible, perhaps involving learning and identification, but genetic transmission may also be involved.[51] The biological mechanisms involved in anxiety disorders and social phobia may have a role in the development of this personality disorder. It has been suggested that hypersensitivity of brain areas involved in the separation-anxiety response and overactivity of serotonin limbic neuronal circuits may underlie the avoidant temperament trait.[51]

Psychosocial factors mediate the extent to which biological vulnerability is expressed. Children who are belittled, criticized, and rejected by parents have decreased self-esteem, resulting in social avoidance. These problems are reinforced and perpetuated at school and may generate the expectation of rejection from everyone.[50]

Table 4.12.3.12 DSM-IV diagnostic criteria for avoidant personality disorder

A pervasive pattern of social inhibition, feelings of inadequacy, and to negative evaluation, beginning by early adulthood and present in a variety of contexts, as indicated by four (or more) of the following

1 Avoids occupational activities that involve significant interpersonal contact, because of fears of criticism, disapproval, or rejection
2 Is unwilling to get involved with people unless certain of being liked
3 Shows restraint within intimate relationships because of the fear of being shamed or ridiculed
4 Is preoccupied with being criticized or rejected in social situations
5 Is inhibited in new interpersonal situations because of feelings of inadequacy
6 Views self as socially inept, personally unappealing, or inferior to others
7 Is unusually reluctant to take personal risks or to engage in any new activities because they may prove embarrassing

(c) Clinical picture

Avoidant people are characterized by extreme shyness. They appear distant from others and do not express wishes, demands, or opinions. However, this behaviour contrasts with an extreme internal need for warmth and closeness. This contradiction is explained by an exaggerated sensitivity to rejection by others. People with this personality disorder are easily hurt and humiliated by comments from others, which they misinterpret as degrading and disapproving. They tend to be shy, quiet, and inhibited. They say nothing rather than risk being wrong, and they react strongly to any possible indications of mockery or criticism. They usually appear anxious, self-doubting, and insecure when speaking, often use self-defeating expressions, and try to please others. Their tense and fearful demeanour may elicit ridicule from others, which confirms their insecurity. They are concerned with reacting to scrutiny by blushing or crying, which is a cause of further interpersonal avoidance. These patients often choose occupations where no social interaction is needed, and strongly avoid talking in public. The avoidant person lacks intimate relationships with friends or sexual partners unless they anticipate non-critical acceptance.

Patients with avoidant personality disorder perceive themselves as inept and inadequate, and assume that they are unattractive. They tend to see others as negative and potentially harmful. They are inattentive and repeatedly distracted by intrusive thoughts, but they attend intensely to signals of rejection. These people tend to make negative evaluations of situations and exaggerate risk. They have a low tolerance for dysphoric affects, which they avoid by escaping. Escape from reality through fantasy is their usual way of satisfying their needs and relieving frustration.

At interview they may be quite open if they feel accepted. This happens when good rapport is made, which is often easier in clinical than in social situations. However, social limitation outside the office may be intense. Avoidant patients usually feel ashamed about many aspects of their lives and are excessively self-critical, although most of the concerns expressed seem to be trivial.

(d) Course

Avoidant personality disorder may follow childhood fear of strangers and shyness and isolation during school years. However, most shyness in childhood gradually dissipates in adolescence. When it evolves into avoidant personality disorder, the shyness may worsen in adolescence when social and interpersonal relationships become more complex and demanding. The disorder tends to remit or to become less evident in older people.

Avoidant personality disorder is often associated with depressive episodes, dysthymia, and anxiety disorders, particularly social phobia.

(e) Differential diagnosis

It is often difficult to differentiate avoidant personality disorder and social phobia of the generalized type. Impairment and distress due to the phobic situations is more intense in social phobia, which may have started in middle adulthood rather than adolescence. It is not clear whether the disorders are alternative manifestations of the same condition, or are separate disorders.

Hypersensitivity to rejection and criticism, low self-esteem, and feelings of inadequacy are also features of dependent personality disorder. While the avoidant patient avoids contact, the dependent patient focuses on being cared for. However, the disorders often co-occur and must be diagnosed together.

Schizoid and schizotypal personality disorder are also characterized by social isolation. However, avoidants want to have relationships and suffer for their isolation, while schizoids and schizotypals accept isolation.

People with paranoid personality disorders lack confidence in others. However, avoidants do not confide in others because they fear being found inadequate, whereas paranoids fear malicious intent.

(f) Treatment

Anxiety and hypersensitivity to rejection may improve with anxiolytic medication, β-blockers, monoamine oxidase inhibitors, and antidepressant medication. Medication should be combined with psychological treatment based on reinforcing assertiveness and self-esteem, and restructuring cognitive distortions concerning the self and others. Conscious and unconscious dependency needs should be addressed.

Dependent personality disorder (JLC)

Individuals with this disorder show a persistent and global pattern of behaviours directed at avoidance of the loss of intimate others. To attain this goal, they relinquish their own needs, opinions, expression of feelings, and even their self-identity. In exchange, they get others to take over responsibility for major areas of their lives and to protect them. Their self-concept is characterized by weakness and helplessness, while others are perceived as powerful and protective.

These people were formerly included in different classificatory categories. They belong to the abulic type of Kraepelin and of Schneider[2] and were considered as immature personalities.

(a) Aetiology

Classic hypotheses attributing dependent personality to fixation in oral phase of psychosexual development have given way to others indicating that the cause is rather deprivation than overgratification in the oral phase.

Dependent personality was recognized first in DSM-III.[11] The description has changed little in DSM-IV (Table 4.12.3.13) and ICD-10.[52] The crucial feature in both systems is the urgent need of patients to be cared for by others, with dependence, attachment, and fear of abandonment. Lack of self-confidence was required in DSM-III, but was eliminated from recent classifications because it is not specific to this disorder.

(b) Clinical picture

Dependent patients are passive. They rarely express needs or feelings, especially those that are sexual or aggressive. They tend to avoid responsibilities or decisions in major areas of their lives, such as work and financial or interpersonal relationships. Instead they get others, particularly family or partner, to decide for them or to provide continuous guidance. They depend on others (often one other person, usually the partner or a parent) to decide where they should go, what they should do, and even which clothes they should wear. They manifest self-doubt, pessimism, and a need for affection. They lack aggressiveness and appear helpless. The dependent patient avoids jobs that demand taking responsibility and managing others, and becomes anxious when forced into such situations. These patients seek intensely for companionship and do not tolerate being alone. They may function at an adequate level if in a close and protective relationship, but when left alone they are unable to survive. They believe that they are incapable of functioning independently and require constant assistance. They do not initiate projects, but wait for others who, they believe, will do them better. However, dependent individuals can perform such tasks for other people whom they want to please and to whom they want to attach themselves.

They accept unpleasant tasks, are self-sacrificing, and tolerate verbal, physical, or sexual abuse. Abusive relationships may be accepted as long as the attachment is preserved and is not excessively distorted.

An excessive and unrealistic fear of abandonment is constant in dependent individuals. When an intimate relationship is terminated by separation or death, dependent individuals urgently seek another person who will provide the care and support they seek. Thus they become rapidly and indiscriminately attached to other persons when left alone.

These people are pessimistic, self-doubting, and have low self-esteem. They belittle their capacities and successes and present themselves as inept. They take criticism as a proof of their ineptness and confirmation of their lack of self-confidence.

Table 4.12.3.13 DSM-IV diagnostic criteria for dependent personality disorder

A pervasive and excessive need to be taken care of that leads to submissive and clinging behaviour and fears of separation, beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following

1. Has difficulty making everyday decisions without an excessive amount of advice and reassurance from others
2. Needs others to assume responsibility for most major areas of his or her life
3. Has difficulty expressing disagreement with others because of fear of loss of support or approval. **Note** Do not include realistic fears of retribution
4. Has difficulty initiating projects or doing things on his or her own (because of a lack of self-confidence in judgement or abilities rather than a lack of motivation or energy)
5. Goes to excessive lengths to obtain nurturance and support from others, to the point of volunteering to do things that are unpleasant
6. Feels uncomfortable or helpless when alone because of exaggerated fears of being unable to care for himself or herself
7. Urgently seeks another relationship as a source of care and support when a close relationship ends
8. Is unrealistically preoccupied with fears of being left to take care of himself or herself

(c) Epidemiology

Recent studies have found a median prevalence of 0.7 per cent (see Chapter 4.12.5). Although dependent personality disorder is diagnosed more frequently in women, structured interviews have not shown significant differences between the sexes. Cultural factors may affect the reported prevalence, as passivity, politeness, and submission are normal in some societies.

(d) Course

Dependent personality features present in adolescence may evolve positively in adulthood or lead to a personality disorder. Dependent individuals are at increased risk of depressive, anxiety, and adjustment disorders, particularly in relation to loss of close relationships. Dependent personality disorder may follow separation anxiety in childhood, or chronic physical illnesses in childhood requiring long periods of care and attention.

(e) Differential diagnosis

Dependent personality disorder has some similarities with histrionic personality disorder. Histrionic patients, like dependent patients, adjust their conduct to please other people. Their lives are centred in these others. However, people with histrionic personality disorder obtain attention and care by seductive or manipulative behaviours, whereas people with dependent personality disorder wait passively for others to care for them.[7]

Like people with avoidant personality disorder, dependent individuals may feel devastated by minor criticism or lack of attention from others. However, they lack the sense of embarrassment and social shyness of the avoidant, and fear loneliness or abandonment.

People with dependent and borderline personality disorders share an excessive fear of abandonment. However, borderline individuals react to separation with feelings of emptiness and rage, and are demanding, in contrast with the submissive and appeasing attitude of dependent individuals, which is directed towards finding another person to provide support.

Dependent personality disorder must be differentiated from normal dependent behaviours in specific life situations; for example, elderly people with chronic or debilitating disease may become dependent.

(f) Treatment

Pharmacological treatment is indicated only when depressive or anxiety symptoms are present, especially when associated with separation or loss.

In psychotherapy, the therapist must avoid the development of excessively dependent attachments. Self-confidence and self-esteem should be enhanced and the patient helped to enjoy the feeling of personal autonomy and independence. Cognitive restructuring and social skills training are often useful in bringing these changes about.

Obsessive–compulsive (anankastic) personality disorder

While DSM-IV labels this personality disorder as obsessive–compulsive personality disorder (Table 4.12.3.14), ICD-10 prefers the term anankastic, previously used in European psychiatry to refer to fearful, insecure, and compulsive individuals. The cardinal feature of this disorder is an exaggerated and pervasive attempt to control. Anankastic patients need to control those who are close to them, to control every uncertainty, and to control their own thoughts and emotions. The anankastic lacks an internal sense of security and tries to make the external world totally predictable. The anankastic is afraid of his own internal aggressive drives and avoids free emotional expression. Others perceive this kind of personality as characterized by inflexibility and stubborn inefficiency.

Table 4.12.3.14 DSM-IV diagnostic criteria for obsessive–compulsive personality disorder

A pervasive pattern of preoccupation with orderliness, perfectionism, and mental and interpersonal control, at the expense of flexibility, openness, and effciency, beginning by early adulthood and present in a variety of contexts, as indicated by four (or more) of the following

1. Is preoccupied with details, rules, lists, order, organization, or schedules to the extent that the major point of the activity is lost
2. Shows perfectionism that interferes with task completion (e.g. is unable to complete a project because his or her own overly strict standards are not met)
3. Is excessively devoted to work and productivity to the exclusion of leisure activities and friendships (not accounted for by obvious economic necessity)
4. Is overconscientious, scrupulous, and inflexible about matters of morality, ethics, or values (not accounted for by cultural or religious identification)
5. Is unable to discard worn-out or worthless objects even when they have no sentimental value
6. Is reluctant to delegate tasks or to work with others unless they submit to exactly his or her way of doing things
7. Adopts a miserly spending style toward both self and others; money is viewed as something to be hoarded for future catastrophes
8. Shows rigidity and stubbornness

(a) Epidemiology

The prevalence of obsessive–compulsive personality disorders is about 1 per cent in community samples and up to 10 per cent in psychiatric patients, especially those with depressive and anxiety disorders. It is most frequent among males. Some obsessive–compulsive traits are sanctioned in some cultures, and a personality disorder should not be diagnosed unless the traits are markedly beyond the average for the culture.

(b) Aetiology

Biological factors and learning seem to be involved in the aetiology of obsessive–compulsive personality disorder. The personality may be partly inherited.[53] Early psychodynamic theories linked obsessive personality to the anal phase of psychosexual development between the ages of 2 and 4, when libidinal drives come into conflict with parental attempts to socialize the child, especially in sphincter control and toilet training. Later psychoanalytic theory[54] emphasized earlier manifestations of the child's autonomy versus parental wishes. The expression of drives and emotions, including anger, is shaped by parental responses and may evoke shame and criticism.

This dynamic sequence is reinforced in societies which are strongly influenced by the Protestant work ethic, in families where individual emotions are subordinated to the group, and in societies in which open expression of emotions is discouraged.

(c) Clinical picture

The behaviour of an obsessive–compulsive personality has been consistently described as one of orderliness. The patient is preoccupied with details, and pays attention to rules, procedures, schedules, and punctuality. Patients with obsessional personalities often produce their own detailed lists of symptoms and are annoyed if any item is neglected or misinterpreted. They repeat actions and check for mistakes, despite the inconvenience and annoyance that result from this behaviour. As a consequence, their conduct is frequently inefficient. For example, the combination of unproductive perfectionism and rigidity may lead to difficulty in finishing a written report on time because of excessive correction and rewriting. Since this striving for perfection and order is time consuming, other areas of their lives often appear disorganized. One room or one desk drawer may fall into disarray, or parts of their social or family lives may be disorganized.

People with obsessive–compulsive personality focus on work and productivity. It is difficult for them to take vacations or even to have free time. They do not enjoy leisure activity, which they may consider a waste of time. Often, they need to take work home to alleviate their anxiety. Hobbies and leisure pursuits become formally organized activities. They insist on perfect performance of sports or games and transform them into a serious task requiring careful organization and hard work. Leisure activities may be an unpleasant experience for the others involved, owing to the insistence on rules and standards.

Stubbornness is another characteristic of these people. They need things to be done in their way, and realistic arguments do not usually make them change their insistence. They need others to submit to their way of doing things, and often believe that no one can do the tasks as perfectly as they can. They give detailed instructions, insisting that their way is the only way of doing things, and are irritated if others suggest alternatives. Therefore, they generally insist in doing everything themselves and are unable to delegate, which increases their inefficiency at work. Paradoxically, their stubbornness is associated with doubt. Indecisiveness is a constant characteristic unless they have structured guidelines. They fear making mistakes or misjudgements, and delay repeatedly until they have enough data to take what they consider the only right decision. When rules do not dictate the correct answer to a problem or when procedures for tasks are not laid down, decision-making or task initiation may become a lengthy and painful process.

People with this personality disorder are characterized by excessive conscientiousness and scruples. They are inflexible about matters of morality, ethics, or values. Moral principles and standards of performance have to be followed rigidly, and respect for authority and rules is absolute. Failure to do these things leads to irritation, anger, and self-criticism.

These people are stingy and mean, and often live with standards far below their actual socio-economic status. They dislike spending, believing that money should be saved in case of future difficulties. They have great difficulty in discarding worn-out or worthless objects, believing that they might be useful some day. They may hoard objects such as newspapers or broken appliances, even when they have no sentimental value.

These people are humourless and lack spontaneity of emotional expression. Usually they do not express anger directly. However, they are often angry in situations in which they are unable to control the behaviour of themselves or others. Anger is generally manifested by indirect aggressive acts (such as leaving a small tip or not providing minor help when expected). Their management of anger is closely related to their attitude of dominance–submission towards authority figures. They may be excessively submissive to a person in authority whom they respect, but obstructive with an authority figure whom they do not respect.

The affect of the obsessive person is controlled and stilted. It is not flat or blunted, but constricted. They do not laugh or cry, and feel uncomfortable with people who express their feelings. Their mood is usually serious but may appear anxious or depressed. In a clinical interview they may sit in a stiff unnatural posture, and seldom make spontaneous comments about their emotions. They usually relate their history in a pedantic and circumstantial manner. If interrupted by a question from the doctor, they have to finish their monologue before answering. When asked about feelings, they answer with lists of facts and circumstances. They can label emotions and feelings, but are unable to display them.

In summary, obsessive personalities love order, neatness, and sameness, and hates novelty, spontaneity, and change. They need control, security, and certainty, and avoid creativity, art, and excitement. They mitigate anxiety by following strict rules and repress emotional expression by avoiding spontaneity. They fear their inner fragile and aggressive emotional world.

(d) Course

Like other personality disorders, obsessive–compulsive personality disorder is present in early adulthood and tends to be persistent and constant. However, some adolescents with marked obsessive traits become warm, loving, and tender adults. On the other hand, intense obsessional traits in adolescence are occasionally a premorbid stage of schizophrenia ('pseudoneurotic schizophrenia'). The developmental relationship between obsessive–compulsive personality disorder and obsessive–compulsive disorder is controversial. In the past, it was suggested that most obsessive–compulsive personality disorder evolved to a full obsessive–compulsive disorder, indicating that the two syndromes were expressions of the same basic disorder. More recent investigations[55] indicate that most obsessive–compulsive disorder patients do not have a comorbid obsessive–compulsive personality disorder. A variety of psychiatric disorders may present in a patient with obsessive personality, but depressive and anxiety disorders are the most common, followed by phobic, somatoform, and obsessive–compulsive symptoms. Hypochondriacal syndromes are commonly found in obsessive individuals when they lose control of situations.

Persons with this personality disorder may do well in jobs that demand working with detail, order, and structured procedures, and may adjust to interpersonal relationships with submissive spouses. However, they are particularly vulnerable to unexpected changes in their occupational and social environment. Late-onset depression is a common occurrence in obsessive–compulsive personalities.

(e) Differential diagnosis

The main difficulty in diagnosing obsessive–compulsive personality disorder is to differentiate it from obsessive–compulsive disorder. The latter diagnosis is made when occupational and personal functioning is severely impaired as a consequence of doubt, indecisiveness, hoarding, or any other obsessive behaviour. In many, but not all, cases of obsessive personality, the traits and behaviours

are egosyntonic and no resistance is present, in contrast with obsessive–compulsive disorder.

The perfectionism of obsessive personalities may be present in narcissistic personality disorder. However, narcissistic individuals tend to believe that they have achieved perfection, while obsessive individuals tend to be highly critical of their own achievements.

Social detachment and the lack of empathy and warmth may suggest schizoid personality disorder. However, obsessive individuals constrain their emotional expression to keep control of a situation, while schizoids lack the fundamental capacity for affective display or intimacy.

Not all individuals with obsessive traits have obsessive–compulsive personality disorder. Obsessive traits can be adaptive in some situations; it is only when they are maladaptive, inflexible, and persistently cause functional impairment that a personality disorder be diagnosed.

(f) Treatment

Pharmacological treatment may be tried in patients with anxiety and distress due to intense doubts, indecisiveness, and scruples. Benzodiazepines may alleviate tension in these cases. Antidepressants with a serotonergic profile sometimes improve mood and global functioning.

Psychological cognitive treatment, focusing on perfectionism, rigidity, scrupulousness, and intolerance of failure, is the main therapeutic approach. Repressed aggression, guilt, and dependency needs should be addressed using a psychodynamic approach.

Other personality disorders (not included in DSM-IV)

Passive–aggressive (negativistic) personality disorder

(a) Definition

Resistance to demands for adequate social and occupational performance and negativistic attitude are considered to be central features of passive–aggressive personality disorder. A pervasive pattern of argumentativeness, oppositional behaviour, and defeatist attitudes are typical, and are thought to be a covert manifestation of underlying aggression, which is expressed passively and indirectly. Passive–aggressive personalities have interpersonal and cognitive dysfunction and severe impairment in terms of self-image, global mental health, and ability to function at work and in intimate relationships.[56]

Passive–aggressive personality disorder was officially included in DSM-I as the passive–aggressive personality 'trait disturbance' depicted as an immature reaction to military stress by helpless, passive, and obstructive resistant behaviour. However, passive–aggressive disorder was not included in DSM-IV because of the many unsolved problems related to its concept in previous classifications. Instead, it is placed in Appendix B of DSM-IV where it is alternatively called negativistic personality disorder. Research criteria are proposed which are expected to be empirically evaluated and to determine the validity and reliability of this diagnosis (Table 4.12.3.15).

There has been much debate as to whether passive aggression constitutes a personality disorder, a defence mechanism, or a specific maladaptive personality trait (coping style).[57] Surprisingly, empirical literature on the subject is scarce, although passive–aggressive behaviour has been widely recognized by clinicians. An overlap with other personalities has been shown, and it has never been included as a separate category in the *International Classification of Diseases*. The passive–aggressive dimension, as assessed by self-reports, is always high in depressed patients and is state-dependent.[58] Perhaps it would be best to conceptualize passive aggression as a continuum: a passive–aggressive defence mechanism may be normal in some situations, it could be a trait of many personality disorders, and when pronounced and long-lasting it should be designated as passive–aggressive personality disorder.

Table 4.12.3.15 DSM-IV research criteria for passive–aggressive personality disorder

A.	A pervasive pattern of negativistic attitude and passive resistance to demands for adequate performance, beginning by early adulthood and present in variety of contexts, as indicated by four (or more) of the following
	1 Passively resists fulfilling routine social and occupational tasks
	2 Complains of being misunderstood and unappreciated by others
	3 Is sullen and argumentative
	4 Unreasonably criticizes and scorns authority
	5 Expresses envy and resentment toward those apparently more fortunate
	6 Voices exaggerated and persistent complaints of personal misfortune
	7 Alternates between hostile defiance and contrition
B.	Does not occur exclusively during major depressive episodes and is not better accounted for by dysthymic disorder

(b) Epidemiology

The population prevalence ranges from 0.9 to 3 per cent, but in those cases in which a secondary co-occurring diagnosis was assigned, the secondary frequency of passive–aggressive personality disorder was about 10 per cent.[59] Some studies found a higher prevalence in women and others in men.

(c) Aetiology

The cause of the disorder is multidimensional, comprising biological, psychoanalytical, behavioural, interpersonal, and social learning perspectives.

Ambivalence is considered to be a core conflict of passive–aggressive personalities, which originates from fixation to the biting or sucking stages of the oral phase of psychosexual development. Some authors consider masochism to be another precursor of the passive–aggressive personality.

According to the behavioural model, passive–aggressive behaviour is the expression of anger in maladaptive verbal and non-verbal ways that do not lead to rewarding problem-solving Failure to learn appropriate assertive behaviour would be the main aetiological factor.[60]

The primary social factor influencing the development of passive–aggressive patterns would be contradictory parental attitudes in childhood which, being conflicting and incompatible, prevent the child from expressing his feelings directly and thus urge him to develop passive resistance.

(d) Clinical characteristics and diagnosis

Passive–aggressive personalities seek novel and stimulating situations in impulsive ways, while remaining unpredictable.[61] Procrastination and inefficiency are behaviours used to avoid responsibility, which they show by stubborn resistance to the fulfilment of expectations and claiming forgetfulness.

Passive–aggressive individuals easily become irritable and, gloomy and they are resentful and discontent with life. Accumulated anger may be expressed by verbal acting-out, after which passive–aggressive individuals feel guilty and gloomy. Since they have difficulties in expressing emotions directly, they are prone to diffuse somatic complaints, hypochondriasis, and psychosomatic disorders.

Interpersonal ambivalence is a core feature of passive–aggressive personality disorder. Negativism is particularly expressed towards authorities, with whom they are never satisfied and whom they criticize constantly. The argumentative self-detrimental behaviour of passive–aggressive personalities is often experienced by others as punitive and manipulative. Negative verbal comments, which are often caustic, and irritable and moody patterns of communication are typical.

Passive–aggressive individuals are cynical, sceptical, hypercritical, and mistrustful. Disillusioned with life, discouraged, discontented with themselves and with others, they are also pessimistic about the future. They persistently complain and blame others for their own bad luck, feeling themselves to be misunderstood martyrs and victims of destiny.

(e) Comorbidity and differential diagnosis

Passive–aggressive personality disorder is frequently comorbid with major depression, dysthymic disorder, anxiety, panic disorder, and hypochondriasis. Patients with depressive disorders are more aware of their feelings of inferiority and more likely to feel guilty, and their depressed mood is continuous rather than erratically hostile and moody, as in passive–aggressive personality.

Comorbidity with many personality disorders (histrionic, borderline, obsessive–compulsive, dependent, narcissistic) is also frequently observed. People with these personality disorders may use passive aggression as a defence mechanism. Suicide attempts are not as frequent as in histrionic and borderline personality disorders, and features of passive–aggressive personality are less dramatic, affective, openly aggressive, and severe.

(e) Course and prognosis

There are insufficient data on the course and prognosis of passive–aggressive personality disorder. When passive–aggressive people are unable to control their anger, they may experience anxiety, panic states, depressive episodes, chronic depression, and psychosomatic disorders. They are prone to alcohol abuse, and their careers are erratic and stunted despite their abilities (frequent changes of jobs are common). Suicide attempts may complicate this disorder.

(f) Treatment

(i) Pharmacotherapy

Target symptoms, such as depression, anxiety, and somatic complaints, should be treated. Benzodiazepines may be warranted and helpful in the initial period of psychotherapy. The side effects of medication are often the reason for complaints about their psychiatrist's inefficiency. Non-compliance is frequent, reflecting resistance to the therapist. Abuse of medicaments should be controlled and considered seriously.

(ii) Psychotherapy

Psychotherapy is the treatment of choice. Various modes are used, including supportive, psychodynamic, behavioural assertiveness training, and the paradoxical approach in group or individual settings. The goal of treatment should be to help the patient escape from the vicious circle of self-defeating behaviour and develop a mature way of expressing anger and other feelings.

Self-defeating (masochistic) personality disorder

Individuals with masochistic personality disorder persistently seek humiliation and failure, and submit to the will of others.

The term masochism was introduced to psychiatry by Krafft-Ebing in 1882. It is derived from a character in a novel by the German author Leopold von Sacher-Masoch, a man who endured torture, scorn, and humiliation from a woman. Later, Freud conceptualized masochism as the result of aggressive instinct directed towards the self instead of an external object. Reich[62] described the masochistic character as a person who suffered deep frustrations in early developmental stages and needed to express this frustration through suffering inflicted by the love 'objects'. Thus defiance is always present in the masochistic search for love. According to Horney,[63] helplessness and victimization may be a masochistic way of expressing hostility by making others feel shameful. Masochistic suffering may also be used to avoid reproach and responsibility and as a way of restoring a sign of personal value.

(a) Clinical picture

People with self-defeating personality disorder avoid pleasurable experiences and undermine their achievements. They neglect their appearance and live below their means. They accept and endure humiliation, expecting that others will sympathize with them. In this way, they fulfil their expectation that submission will bring love and care. They prefer to relate to people who abuse them and consider those who consistently treat them well to be boring.

These people fail to accomplish tasks of which they are capable and adopt an inferior role. When making appropriate demands, they feel that they are taking advantage of others and adopt an apologetic manner. For them, success is inversely related to inner security. Successful relationships do not make them feel confident, but increase their fears. They tend to believe that all experiences involve future frustration and pain. They may respond to positive personal events with depression and behaviour that negates their accomplishments. They look for people who will respond to their behaviour with disdain, rejection, or cruelty.

People with self-defeating personality disorder do not defend themselves against expressions of disgust and resentment directed towards them and rarely accuse or reproach others. They do not feel confident and are not assertive. They fear that optimism may lead to greater problems.

They are not worried by these attitudes. Rather, masochists believe that by exaggerating their weaknesses and inefficiency they will protect themselves from aggression by others. They feel protected when someone needs something from them, and many non-assertive masochists engage in self-sacrifice for their own protective feelings rather than for the welfare of others. What seems to be a comprehensive and self-sacrificing attitude reflects a lack of confidence and empathy.

The mood of these individuals is usually dysphoric, fluctuating between anxiety and sadness.

(b) Differential diagnosis

Since masochistic personality disorder is not included in DSM-IV, little research has been done on comorbidity or on the validity or specificity of this diagnosis.

Self-defeating attitudes, low self-esteem, and depression may be found in individuals with dependent, borderline, and depressive personality disorders. Some of the characteristics are also found in avoidant personality disorder.

Masochists are prone to mood disorder and dysthymia. Anxiety disorders are frequent. Their fears of abandonment are a persistent source of anxiety. Hypochondriasis and somatization are also common, sometimes as a way of obtaining attention.

(c) Treatment

Antidepressants and anxiolytics may be useful to alleviate a dysphoric state. Psychological treatment should take into account that masochistic patients sometimes induce an aggressive countertransference as a response to their own wish to be hurt. The therapist should gradually clarify the behaviours, which provoke hostile responses and seek to reward adaptive interpersonal behaviours. Training in assertiveness and social skills is sometimes helpful.

Sadistic personality disorder

Sadistic personality disorder is a controversial category, which is not included in either DSM-IV or ICD-10. Some authors, especially those working with perpetrators of abuse, support its inclusion in the diagnostic nomenclature, believing that it is a valid clinical entity, which deserves special treatment. Sadism as a term describing desire to inflict pain upon the sexual object was originally used by Krafft-Ebing.[64] Kernberg[32] connects two dispositions (sadistic and masochistic) into a sadomasochistic character, which includes 'help-rejecting complainers' and often has underlying borderline personality organization.

(a) Aetiology

Subjects with sadistic personality disorder often had a history of significant childhood loss and physical, emotional, or sexual abuse during childhood.[65] Despite their significant psychopathology, sadistic individuals are surprisingly highly functioning, with steady employment and intense long-lasting relationships.

(b) Clinical picture

Sadistic individuals demonstrate a long-standing maladaptive pattern of cruel, demeaning, and aggressive behaviour towards others in order to cause suffering and pain and to establish dominance and control. Sadistic persons are fascinated by violence, weapons, injuries, and torture, and are most frequently encountered in forensic settings among child and partner abusers.

(c) Differential diagnosis

The major distinction for the diagnosis of sadistic personality disorder is from antisocial personality disorder. Sadistic persons may simply represent an aggressive (antagonistic) subtype of psychopathy. Intimidation and sadistic control of others, as well as a fascination with weapons, martial arts, and torture, may be manifested by both antisocial and sadistic individuals. Moreover, both disorders may display 'malignant narcissism',[37] with an admixture of narcissistic, antisocial, sadistic, and paranoid features.

(d) Epidemiology

Existing data[66] suggest that sadistic personality disorder is relatively uncommon, although it may have a higher prevalence in specific forensic populations. Several studies found a high overlap with narcissistic, paranoid, and antisocial personality disorders, which raised the possibility that sadistic personality disorder may not be a distinct entity.

(e) Course

No data are available on the course of this disorder.

(f) Treatment

Sadistic individuals seldom seek treatment, and are usually encountered in forensic settings. No treatment has been successful for this disorder. Since the aetiology is probably multifactorial, there should be multiple approaches to treatment. The primary aim of treatment is control of cruelty and malignant aggression. As with antisocial personalities, selective serotonin reuptake inhibitors and lithium may be beneficial in regulating the serotonergic function that probably underlies the aggressive action, and carbamazepine and clonazepam may act to regulate ictal aggressive outbursts.

Psychotherapy is usually difficult because these individuals lack any desire to change and because there may be serious countertransference problems for the therapists. Small groups may be helpful because there is a dissolution of transference and peer confrontation is accepted more easily.

Depressive personality disorder

This personality disorder is not included in ICD-10 or DSM-IV, although it is considered as a subject for further study in the latter. However, the concept of depressive personality was well recognized in previous decades (e.g. by Kraepelin and Schneider). Depressive personality was seen as a pattern of brooding, pessimism, and low self-confidence,[2] and as a tendency to physical lassitude and suffering. Phenomenological accounts[67] emphasize dependency, orderliness, and adherence to social conventions. The psychoanalytic concept of masochism[47] overlaps to some extent with the classical depressive personality. More recently, Akiskal[68] has discussed the depressive personality as part of a spectrum of affective disorders, reviving Kretschmer's ideas[10] on the role of temperament as the base from where psychiatric disorders develop.

(a) Aetiology

Psychological and biological factors have been suggested as causes of depressive personality disorder. Early losses, inadequate attention from parents, and a punitive superego have been postulated by psychoanalytic authors. Others have suggested that a depressive temperament is genetically related to affective disorder.

(b) Clinical picture

People with depressive personality disorder are submissive, quiet, introverted, and unassertive. Their cognitive style is marked by pessimism, dejection, and self-reproach. They appear gloomy, joyless, cheerless, and unhappy. They are serious and lack a sense of humour. They do not believe that they deserve to be happy. They have a negative view of the past and present and do not expect things to improve. They anticipate failure and dwell on their negative perspective of life. They have a low tolerance to shortcomings and failures, which are seen as confirming their own pessimistic assumptions. They are prone to guilt and judge themselves severely.

Their self-esteem is low and they feel inadequate. They focus on the failings of others and are critical of themselves.

(c) Epidemiology

No data are available on the frequency of this disorder in the population.

(d) Differential diagnosis

Some clinicians doubt whether a distinction can be made between depressive personality disorder and dysthymic disorder. The diagnosis of depressive personality disorder emphasizes cognitive and behavioural aspects rather than the depressed mood. Also, dysthymia, although chronic, has a fluctuating course.

Dysthymic patents generally experience their symptoms as egodystonic, while people with depressive personality disorder think that they have a realistic view of their situation. People with depressive personality disorder may meet the criteria for dependent personality disorder and self-defeating personality disorder, and it is difficult to distinguish between these disorders.

(e) Course

Patients with depressive personality disorder have depressive episodes and dysthymia more frequently than other individuals, and may have difficulties in adapting to stressful or uncertain situations.

(f) Treatment

Antidepressants may be useful when the person is at higher risk for developing a depressive episode. Psychological treatment may be helpful, since these people have a good capacity for introspection and reality testing. Cognitive approaches can help patients to understand their negative views and derogatory attitudes. These patients usually establish a good therapeutic relationship with the clinician.

Personality changes

Enduring personality changes after traumatic experiences

ICD-10 has two categories for personality changes: those occurring after catastrophic situations, and those starting after psychiatric illness. Either diagnosis should be made only when there is evidence of a definite personality change, including cognition, behaviour, and interpersonal relationships. The changes must not be a manifestation of a current mental disorder or the residue of a previous mental disorder. It must not be an exacerbation of a pre-existing personality disorder.

The **aetiology** of the personality change presumably relates to the extreme existential experience of the catastrophic situation or the psychiatric illness.

Examples of catastrophic experiences include life in concentration camps, experiences of disaster, and prolonged exposure to other life-threatening situations. Personality changes following short-term exposure to life-threatening situations, such as a car accident, should not be included in this group, since such changes probably depend on a previous psychological vulnerability. In ICD-10, the symptoms of personality change after catastrophic experiences include hostility and mistrust of the environment, social withdrawal, feelings of emptiness or hopelessness, estrangement, and alertness or feeling on the edge.

When the personality change follows a mental disorder, the cause is related to the stressful experience and the perceived damage to the patient's self-image arising from the disease. Other people's attitudes towards the illness, the subjective emotional experience, and previous psychological adjustment are also important. Features of the personality change after mental disorders include feelings of being stigmatized and consequent withdrawal, passivity, and loss of interests, dependence, excessive demandingness and complaining, and dysphoric–labile mood.

Personality change due to a general medical condition (JLC)

This category, which is included in both ICD-10 and DSM-IV, describes syndromes affecting global features of behaviour, cognition, and emotions, and secondary to the physiological effects of general medical diseases.

The essential feature is a change in personality after a general medical disease. In childhood before a stable pattern of personality has been established, a marked deviation from normal development suggests the diagnosis.

A central feature is loss of control over the expression of emotions and impulses. Affect is commonly labile and shallow, although persistent mild euphoria or apathy may be present, especially when the frontal lobes are affected. The elevated mood of these patients is hollow and silly, unlike that of hypomania. Patients may appear childish, expansive, and disinhibited, but they may admit to not feeling happy. Others are indifferent and apathetic.

Exaggerated expressions of rage and aggression are usually present, often out of proportion to any stressor. Loss of impulse control is also shown in social and sexual disinhibition, inappropriate jokes, and overeating.

(a) Aetiology

In most cases this disorder is associated with structural damage to the brain. Head trauma, cerebral neoplasms, vascular accidents, multiple sclerosis, Huntington's disease, and complex partial epilepsy may all cause personality change, especially when affecting frontal and temporal lobes. Systemic diseases involving the central nervous system, including endocrine disorders, AIDS, lupus erythematosus and chronic metal poisoning, may have the same effect.

Patients with this disorder generally have a clear sensorium but may be inattentive and have some mild cognitive dysfunction. They do not show intellectual deterioration.

(b) Differential diagnosis

In dementia, personality change is accompanied by intellectual deterioration. However, personality change may predate the dementia. Distinction from schizophrenia and other personality disorders is based on the clinical history and the presence of a general medical disease capable of causing personality change.

(c) Treatment

If possible, treatment is directed to the causative condition. Pharmacological treatment of specific symptoms may be useful when depression or inappropriate anger is present.

Further information

Millon, T., and Davis, R. (2000). *Personality disorders in modern life*. John Wiley and sons, Inc., New York.

Schneider, K. (1950). *Psychopathic personalities* (9th edn.). Casell, London. (Original work published in 1923).

Menninger, K. (1940). Character disorders. In *The psychodynamics of abnormal behaviour* (ed. J.F. Brown), pp. 384–403. McGraw-Hill, New York.

Livesley, W.J., Jackson, D.N., and Schroeder, M.L. (1992). Factorial structure of traits delineating personality disorders in clinical and general population samples. *Journal of Abnormal Psychology*, **101**, 432–40.

References

1. Kraepelin, E. (1921). *Manic-depressive insanity and paranoia*. Lingstone, Edinburgh.
2. Schneider, K. (1950). *Psychopathic personalities* (9th edn.). Cassell, London.
3. Freud, S. (1911). Psychoanalytic notes upon an autobiographical account of a case of paranoia. In *Collected papers*, Vol. 3. Hogarth Press, London, 1925.
4. Perry, J.C., and Vaillant, G.E. (1989). Personality disorders. In *Textbook of psychiatry* (eds. H.I. Kaplan and B.J. Sadock), pp. 1352–87. Williams and Wilkins, Baltimore, MD.
5. Kendler, K.S., Masterson, C.C., Ungaro, R., *et al.* (1984). A family history study of schizophrenia related personality disorders. *The American Journal of Psychiatry*, **141**, 424–7.
6. Seiver, L.J. and Davis, K.L. (1991). A psychobiological perspective on the personality disorders. *The American Journal of Psychiatry*, **148**, 1647–58.
7. Millon, T. (1997). *Disorders of personality: DSM-IV and beyond* (2nd edn.). Wiley, New York.
8. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn.). American Psychiatric Association, Washington, DC.
9. Sullivan, H.S. (1953). *The interpersonal theory of psychiatry*. Norton, New York.
10. Kretschmer, E. (1936). *Physique and character*. Kegan Paul, Trench, Trubner, London.
11. American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd edn. revised). American Psychiatric Association, Washington, DC.
12. Robins, L.N. (1966). *Deviant children grown up: a sociological and psychiatric study of sociopathic personality*. Williams and Wilkins, Baltimore, MD.
13. de Girolamo, G. and Reich, J.H. (1993). *Personality disorders*. WHO, Geneva.
14. Crowe, R.R. (1974). An adoption study of antisocial personality. *Archives of General Psychiatry*, **31**, 785–91.
15. Cadoret, R.J. (1978). Psychopathology in adopted-away offspring of biologic parents with antisocial behaviour. *Archives of General Psychiatry*, **35**, 176–84.
16. Coccaro, E.F., Siever, L.J., Klar, H.M., *et al.* (1989). Serotonergic studies in patients with affective and personality disorders. *Archives of General Psychiatry*, **46**, 587–99.
17. Dee Higley, J., Mehlman, P.T., Taub, D.M., *et al.* (1992). Cerebrospinal fluid monoamine and adrenal correlates of aggression in free-ranging rhesus monkeys. *Archives of General Psychiatry*, **49**, 436–41.
18. Glueck, B. and Glueck, E. (1956). *Physique and delinquency*. Harper, New York.
19. Gunderson, J.G. (1984). *Borderline personality disorder*. American Psychiatric Press, Washington, DC.
20. Akhtar, S. (1992). *Broken structure*. Aronson, Northvale, NJ.
21. Jones, M. (1952). *Social psychiatry: a study of therapeutic communities*. Tavistock Publications, London.
22. Stern, A. (1938). Psychoanalytical investigation and therapy in borderline group of neuroses. *Psychoanalytic Quarterly*, **7**, 467–89.
23. Gunderson, J.G. and Kolb, J.E. (1978). Discriminating features of borderline patients. *The American Journal of Psychiatry*, **135**, 792–6.
24. Kernberg, O. (1989). The narcissistic personality disorder and the differential diagnosis of antisocial behaviour. *The Psychiatric Clinics of North America*, **12**, 553–70.
25. Quality Assurance Project. (1991). Treatment outlines for borderlines, narcissistic and histrionic personality disorders. *The Australian and New Zealand Journal of Psychiatry*, **25**, 392–403.
26. Akiskal, H.S. (1981). Subaffective disorders: dysthymic, cyclothymic and bipolar II disorders in the 'borderline realm'. *The Psychiatric Clinics of North America*, **4**, 25–46.
27. Widiger, T.A. and Frances, A. (1989). Epidemiology, diagnosis, and comorbidity of borderline personality disorder. In *American PsychiatricPress review of psychiatry*, Vol. 8 (eds. A. Tasman, R.E. Hales, and A.J. Frances), pp. 8–24. American Psychiatric Press, Washington, DC.
28. Stone, M.H. (1980). *The borderline syndromes. Constitution, personality and adaptation*. McGraw-Hill, New York.
29. Coccaro, E.F., Siever, L.J., Klar, H.M., *et al.* (1989). Sertonergic studies in patients with affective and personality disorders: correlates with suicidal and impulsive aggressive behaviour. *Archives of General Psychiatry*, **46**, 587–99.
30. New, A.S., Hazlett, E.A., Buchsbaum, M.S., *et al.* (2007). Amygdala-prefrontal disconnection in borderline personality disorder. *Neuropsychopharmacology*, **32** (7), 1629–40.
31. Zanarini, M.C., Williams, A.A., Lewis, R.E., *et al.* (1997). Reported pathological childhood experiences associated with the development of borderline personality disorder. *The American Journal of Psychiatry*, **154**, 1101–6.
32. Kernberg, O.F. (1975). *Borderline conditions and pathological narcissism*. Aronson, New York.
33. Gunderson, J.G. (1996). The borderline patient's intolerance of aloneness: insecure attachment and therapist availability. *The American Journal of Psychiatry*, **153**, 752–8.
34. Akhtar, S. (1984). The syndrome of identity diffusion. *The American Journal of Psychiatry*, 141, 1381–5.
35. Gunderson, J.G. and Phillips, K.A. (1995). Personality disorders. In *Comprehensive textbook of psychiatry* (5th edn.) (eds. H.I. Kaplan and B.J. Sadock), pp. 1438–41. Williams and Wilkins, Baltimore, MD.
36. Stein, G. (1992). Drug treatment of personality disorders. *The British Journal of Psychiatry*, **161**, 167–84.
37. Kernberg, O.F. (1984). *Severe personality disorder-psychotherapeutic strategies*. Yale University Press, New Haven, CT.
38. Paris, J. (1994). *Borderline personality disorder-a multidimensional approach*. American Psychiatric Press, Washington, DC.
39. Linehan, M.M., (1987). Dialectical behavior therapy: a cognitive behavioral approach to parasuicide. *Journal of Personality Disorders*, **1**, 328–33.
40. Kraepelin, E. (1913). Hysterical insanity. In *Lectures on clinical psychiatry* (trans. T. Johnstone), pp. 249–58. William Wood, New York.
41. Freud, S. (1931). Libidinal types. In *Standard edition of the complete psychological works of Sigmund Freud*, Vol. 21 (ed. J. Strachey), p. 266. Hogarth Press, London, 1961.
42. Zimmerman, M. and Coryell, W.H. (1990). Diagnosing personality disorders in the community. *Archives of General Psychiatry*, **47**, 527–31.
43. Marmor, J. (1953). Orality in the hysterical personality. *The Journal of the American Psychoanalytic Association*, **1**, 656–71.

44. Torgersen, S. (1980). The oral, obsessive and hysterical personality syndromes. A study of hereditary and environmental factors by means of the twin method. *Archives of General Psychiatry*, **37**, 1272–7.
45. Cadoret, R.J. (1978). Psychopathology in adopted-away offspring of biologic parents with antisocial behaviour. *Archives of General Psychiatry*, **35**, 176–84.
46. Zetzel, E. (1968). The so-called good hysteric. *The International Journal of Psychoanalysis*, **49**, 256–60.
47. Kernberg, O. (1988). Hysterical and histrionic personality disorders. In *The personality disorders and neuroses* (eds. A. Cooper, A. Frances, and M. Sacks), pp. 231–41. J.B. Lippincott, Philadelphia, PA.
48. Zimmerman, M. and Coryell, W.H. (1990). Diagnosing personality disorders in the community. *Archives of General Psychiatry*, **47**, 527–31.
49. Ronningstam, E. and Gunderson, J. (1991). Differentiating borderline personality disorder from narcissistic personality disorder. *Journal of Personality Disorders*, **5**, 225–32.
50. Millon, T. (1981). *Disorders of personality: DSM-III Axis II*. Wiley, New York.
51. Cloninger, C.R. (1986). A unified biosocial theory and its role in the development of anxiety states. *Psychiatric Development*, **3**, 167–226.
52. World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders-clinical descriptions and diagnostic guidelines*. World Health Organization, Geneva.
53. Clifford, C.A., Murray, R.M., and Fulker, D.W. (1984). Genetic and environmental influences on obsessional traits and symptoms. *Psychological Medicine*, **14**, 791–800.
54. Erikson, E. (1959). Growth and crises of the healthy personality. In *Psychological issues* (ed. G.S. Klein). International Universities Press, New York.
55. Baer, L. and Jenike, M.A. (1992). Personality disorders in obsessive compulsive disorder. *The Psychiatric Clinics of North America*, **15**, 803–12.
56. Drake, R.E. and Vaillant, G.E. (1985). A validity study of axis II of DSM-III. *The American Journal of Psychiatry*, **142**, 553–8.
57. Perry, J.C. and Flannery, R.B. (1982). Passive-aggressive personality disorder: treatment implications of a clinical typology. *The Journal of Nervous and Mental Disease*, **170**, 164–73.
58. Lecic-Tosevski, D. and Divac-Jovanovic, M. (1996). Effects of dysthymia on personality assessment. *European Personality*, **11**, 244–8.
59. Millon, T. (1993). Negativistic (passive-aggressive) personality disorder. *Journal of Personality Disorders*, **7**, 78–85.
60. Millon, T. (1981). *Disorders of personality. DSM-III: Axis II*. Wiley, New York.
61. Cloninger, C.R. (1987). A systematic method for clinical description and classification of personality variants. *Archives of General Psychiatry*, **44**, 573–88.
62. Reich, W. (1933). *Character analysis* (3rd edn.) (trans. V.R. Carfagno). Farrar, Straus and Giroux, New York.
63. Horney, K. (1945). *Our inner confl icts*. Norton, New York.
64. Krafft-Ebing, R. (1989). *Psychopathia sexualis* (10th edn). Enke, Stuttgart.
65. Gay, M. and Fiester, S. (1991). Sadistic personality disorder. In *Psychiatry* (ed. R. Michaels). J.B. Lippincott, Philadelphia, PA.
66. Fiester, S.J. and Gay, M. (1995). Sadistic personality disorder. In *The DSM-IV personality disorders* (ed. W.J. Livesley), pp. 329–40. Guilford Press, New York.
67. Tellenbach, H. (1980). *Melancholia*. Duquesne University Press, Pittsburgh, PA.
68. Akiskal, H.S. (1989). Validating affective personality types. In *The validity of psychiatric diagnosis* (ed. L. Robins), pp. 217–27. Raven Press, New York.

4.12.4 Epidemiology of personality disorders

Francesca Guzzetta and Giovanni de Girolamo

Introduction

tbegun to be scientifically investigated. This development has taken place because a number of standardized instruments to assess personality and PD in an empirical fashion have been developed, in parallel with the refinement of a valid and reliable diagnostic system based on a categorical approach.

The need for the epidemiological investigation of PDs seems justified for several reasons.

1 As seen in recent epidemiological surveys, PDs are frequent and have been found in different countries and sociocultural settings.

2 PDs can seriously impair the life of the affected individual and can be highly disruptive to societies, communities, and families.

3 Personality status is often a major predictive variable in determining the outcome of Axis I mental disorders and the response to treatment.

In this chapter, we review the epidemiological literature on PDs up to October 2007, focusing on studies carried out since the development of the DSM-III. First, community prevalence studies of PDs are reviewed. We then look at the prevalence of individual PDs in the community. Finally, we consider the prevalence of PDs in clinical populations, and in special settings (e.g. primary care, prisons, etc.).

Community epidemiological studies of unspecified personality disorders

Until the development of the DSM-III diagnostic criteria for PDs and the subsequent availability of standardized assessment instruments, epidemiological studies aimed at assessing the prevalence rate of PDs were hampered by severe methodological limitations, including differences in sampling methods and in diagnostic criteria, the known unreliability of PD diagnoses based on clinical judgement, and the lack of standardized assessment methods. Since 1980, twelve main studies with at least 200 subjects sampled have ascertained the prevalence rate of PDs in different community samples using assessment instruments specific for PD; they are shown in Table 4.12.4.1.

In these studies, the sample sizes ranged between 200 and 2053 subjects, with an average sample of 565.4; all surveyed individuals were evaluated by means of a specific PD assessment instrument, mainly a structured interview. While most studies were carried out in one stage, Lenzenweger *et al.*[6] first screened a large sample of university students with a self-administered Axis II inventory, and then interviewed a subgroup of 258 subjects using the International Personality Disorder Examination. The median prevalence rate of any PDs in these eight studies is 12.5 per cent.

Two large community studies[13,14] carried out in the USA were not included in Table 4.12.4.1 since PD prevalence rates were based

Table 4.12.4.1 Prevalence rates of personality disorders in epidemiological surveys

Reference	Country	Sample size	Sample features	Diagnostic Criteria	Assessment method	PD prevalence rate (%)
Black *et al.*[1]	USA	247	Relatives of obsessive-compulsive and normal control probands	DSM-III-R	SIPD	**22.3**[a]
Casey & Tyrer[2]	UK	200	Urban and rural residents	ICD-9	PAS	**13.0**
Coid *et al.*[3]	UK	626	Urban and rural residents aged 16–74 and selected in a 2 phase survey (weighted data)	DSM-IV	SCID-II	**4.4**
Crawford *et al.*[4]	USA	644	Individuals re-interviewed from previous surveys, mean age 33.1 years (range 27.7–40.1)	DSM-IV	SCID-II	**15.7**
Klein *et al.*[5]	USA	229	Relatives of normal controls	DSM-III-R	PDE	**14.8**
Lenzenweger *et al.*[6]	USA	258	University students age 18–19 years (two-stage procedure)	DSM-III-R	IPDE	**3.9**[b]
Maier *et al.*[7]	Germany	452	Normal controls, their partners, and relatives	DSM-III-R	SCID-II	**10.0**
Moldin *et al.*[8]	USA	302	Normal controls, parents and their children	DSM-III-R	PDE	**7.3**
Reich *et al.*[9]	USA	235	Urban residents	DSM-III	PDQ	**11.1**
Samuels *et al.*[10]	USA	742	Individuals re-interviewed from previous survey, aged 34–94 years (weighted data)	DSM-IV	IPDE	**9.0**
Torgersen *et al.*[11]	Norway	2,053	Individual from National Register (weighted data)	DSM-III-R	SIPD	**13.4**
Zimmerman & Coryell[12]	USA	797	Relatives of patients and normal controls	DSM-III	SIDP	**14.3**[a]

PAS, Personality Assessment Schedule, IPDE, International Personality Disorder Examination; PDE, Personality Disorder Examination; PDQ, Personality Diagnostic Questionnaire; SCID—II, Structured Clinical Interview for DSM-IV Axis II disorders; SIDP, Structured Interview for DSM-III-R Personality.

[a] Prevalence includes those with 'mixed' and 'not otherwise specified' disorder.

[b] Prevalence was 6.7% 'definite', 11% 'possible', including 'not otherwise specified disorder'.

on screening questions[13] and on a newly developed fully diagnostic structured interview carried out by lay interviewers rather than clinicians, which lacked any accompanying validity data.[14]

In the surveys considered here, the rate of PDs decreases in older age groups; although the sex ratio is different for specific types of PD (e.g. more schizoid, narcissistic, and antisocial PDs among males, more dependent, avoidant, and histrionic PDs among females), the overall rates of PD are about equal for both sexes. Finally, prevalence rates are generally higher in urban populations and lower socio-economic groups.

Community epidemiological studies of specified personality disorders

Table 4.12.4.2 lists the median prevalence rates for specified PDs based on studies that surveyed different types of randomly selected community samples. We will comment on some of the findings. The first column shows the number of studies on which the median prevalence rate is based.

Paranoid personality disorder

The median prevalence rate of paranoid PD is 1.6 per cent. In the study by Baron,[15] paranoid PD was remarkably more common among relatives of schizophrenic probands (7.3 per cent) than among relatives of control probands (2.7 per cent).

Schizoid personality disorder

There have been 13 studies evaluating the prevalence of schizoid PD in the community, with a median prevalence rate of 0.8 per cent. Baron[15] reported a rate of 1.6 per cent of schizoid PD among relatives of schizophrenic probands, but no cases among relatives of control probands.

Table 4.12.4.2 Median prevalence rates of specified personality disorders in epidemiological surveys

PD Category	Number of studies (*N*)	Median prevalence rate (%)
Paranoid	13	1.6
Schizoid	13	0.8
Schizotypal	13	0.7
Antisocial (dissocial)	24	1.5
Borderline	15	1.6
Histrionic	12	1.8
Narcissistic	10	0.2
Obsessive-compulsive	13	2.0
Avoidant (anxious)	13	1.3
Dependent	12	0.9
Passive-aggressive	8	1.7

Schizotypal personality disorder

The median prevalence rate of schizotypal PD is 0.7 per cent. However, in the study by Baron[15] schizotypal PD was remarkably more common among relatives of schizophrenic probands (14.6 per cent) than among relatives of control probands (2.1 per cent). In a similar fashion, Asarnow *et al.*[16] reported significantly higher rates of schizotypal personality disorders in relatives of probands with childhood onset of schizophrenia than in relatives of community controls (4.2 per cent vs. 0 per cent). These results provide additional support for the specific relationship between schizophrenia and schizotypal PD.

Antisocial (dissocial) personality disorder

Antisocial is the most studied PD. Its prevalence has been assessed in 24 epidemiological surveys, with a mean sample size of 2 943 subjects; nine studies used the Diagnostic Interview Schedule as assessment instrument. Antisocial PD seems to have a prevalence of around 1.5 per cent in the general population and to be substantially more frequent among males than females, with sex ratios ranging from 2:1 to 7:1. It is also more common among younger adults, those living in urban areas, and the lower socio-economic classes. People with a diagnosis of antisocial PD are also high users of medical services.

Borderline personality disorder

Borderline PD has been investigated in 15 studies. Swartz *et al.* [17] carried out a study among 1 541 community subjects (between 19 and 55 years of age) at the North Carolina site of the Epidemiologic Catchment Area (ECA) Program, using a diagnostic algorithm derived from the Diagnostic Interview Schedule (**DIS**). They found a rate of 1.8 per cent for borderline PD; the disorder was significantly more common among females, the widowed, and the unmarried. There was a trend towards an increase in the diagnosis in younger, non-white, urban, and poorer respondents. The highest rates were found in the 19 to 34 age range, with the rates declining with age. All borderline respondents had also a DIS DSM-III Axis I lifetime diagnosis.

Although some believe there is a preponderance of females with borderline PD, they do not always take into account that there is also a preponderance of females in the populations studied. There were two studies that did not find a higher female prevalence.[11,18]

Histrionic personality disorder

Histrionic PD has a median prevalence rate of 1.8 per cent, ranging from 0 per cent[3] to 3.2 per cent.[1] A study by Nestadt *et al.*[19] carried out at the Baltimore (Maryland) site of the Epidemiologic Catchment Area Program, ascertained the prevalence of histrionic PD in the community. The authors found a prevalence of 2.1 per cent in the general population, with virtually identical rates in men and women. No significant differences were found in terms of race and education, but the prevalence was significantly higher among separated and divorced persons. It should be noted that the study derived the diagnoses from instruments not originally intended to diagnose PDs; it might be possible that, in some cases, this study has identified personality traits rather than 'true' PDs.

Narcissistic personality disorder

No cases of narcissistic PD were found in five studies. However, Reich *et al.*[9] and Lenzenweger *et al.*[6] found rates of 0.4 per cent and 1.2 per cent respectively, even higher rates were found by Klein *et al.*[5] and Crawford *et al.*[4] who reported prevalence rates of 3.9 and 2.2 per cent respectively.

Obsessive compulsive (anankastic) personality disorder

The median prevalence rate of obsessive compulsive PD, obtained from 13 studies, was found to be 2 per cent, the highest of all PDs. The rate of compulsive PD was especially high in a study in which the Personality Diagnostic Questionnaire was used (6.4 per cent).[6] However, lower rates were reported with structured interviews. A community study, carried out at the Epidemiologic Catchment Area Program Baltimore site, found a prevalence of 1.7 per cent.[20] Males had a rate about five times higher than females. The disorder was also more frequent among white, highly educated, married, and employed subjects, and it was associated with anxiety disorders. However, the study derived the diagnosis from an interview originally not intended to diagnose PDs; this may mean that adaptive obsessive compulsive traits, rather than a 'true' PD, were identified. In the study by Black *et al.* [1] rates of obsessive compulsive PD were higher among relatives of probands with obsessive compulsive disorder compared to relatives of comparison probands (10.8 per cent vs. 7.9 per cent, respectively), however this difference did not reach statistical significance.

Avoidant (anxious) personality disorder

A total of 13 studies have investigated the prevalence of avoidant PD in community samples, with a median prevalence rate of 1.3 per cent. In the study by Asarnow *et al.*[16] avoidant PD occurred more frequently in relatives of schizophrenia probands compared to comparison control probands, also when controlling for schizoid or paranoid PD, and the authors suggest that avoidant PD might be a separate schizophrenia spectrum disorder, and not merely a sub-clinical form of schizoid or paranoid PD.

Dependent personality disorder

In 12 studies in which the frequency of dependent PD was assessed, the median prevalence rate was 0.9 per cent.

Passive-aggressive personality disorder

The median prevalence rate of passive-aggressive PD, obtained from 8 studies, was found to be quite high (1.8 per cent); interestingly, this type of PD has not been included either in DSM-IV or in ICD-10.

Epidemiological studies of personality disorders carried out in psychiatric settings

Table 4.12.4.3 lists the median prevalence rates for any PDs found in 61 studies carried out in inpatient and outpatient psychiatric samples and published between 1981 and October 2007. Only prospective studies that surveyed clinical samples (either inpatients or outpatients) of more than 100 subjects have been considered for

Table 4.12.4.3 Median prevalence rates of PDs among psychiatric patients in prospective studies including more than 100 subjects

Diagnostic category	Number of studies (*N*)	Median sample (*N*)	Median prevalence rates (%)
Alcohol and substance abuse	15	250	57.0
Affective disorder	19	200	49.2
Anxiety disorders	7	200	40.4
Any Axis disorder	20	131	51.0

this analysis. The second column shows the number of studies on which the median prevalence rate is based.

In these studies, subjects have been directly evaluated for the purpose of obtaining PD rates, by means of a standardized assessment instrument specific for PDs. Several other studies, which have evaluated only the prevalence of specified PDs in clinical samples, are not shown here.

In general, the prevalence of PDs among psychiatric outpatients and inpatients is quite high, with a substantial number of studies ($n = 29$) showing a PD prevalence rate equal or higher than 50 per cent of the sample. However, it is difficult to draw more definite conclusions from these studies, because of substantial differences in sampling, diagnostic criteria, timing of the assessment, assessment methods, availability of mental health services, prevalence of Axis disorders, and sociocultural factors.

There are, however, some consistencies across studies that deserve consideration. The most prevalent PD seems to be borderline, both in inpatient and in outpatient settings. The next most common PDs is histrionic, whereas schizoid PD is infrequently diagnosed. Borderline and histrionic PDs are also characterized by the lowest social functioning. They are especially common in inpatient settings, as their symptomatology often results in the patient being admitted to hospital due to their suicidal behaviour, substance abuse, and cognitive-perceptual abnormalities. In outpatient settings, dependent, and passive-aggressive PD are also common.

Especially in inpatient settings, many people who meet the criteria for one PD also meet the criteria for other PDs [21–23]. The highest comorbidity rate appears to occur with borderline PD, with the frequent coexistence of borderline and histrionic PDs, antisocial, schizotypal, and dependent PDs.

With regard to comorbidity between PDs and Axis I disorders, the most common and best-studied patterns are between substance abuse and PDs, affective disorders and PDs, and anxiety disorders and PDs (particularly borderline, antisocial, avoidant, and dependent PDs). Other clinically significant associations have been found between PDs and eating disorders: obsessive-compulsive, avoidant, and dependent PDs are most commonly associated with anorexia nervosa whereas borderline, avoidant, dependent, and paranoid PDs are the most common among individuals with bulimia.[24] High rates of PD (especially borderline and antisocial PDs) have also been detected in patients with selected medical conditions, such as HIV-positive patients.[25,26]

Some studies have assessed the treated prevalence of PD using administrative data (e.g. discharge figures, psychiatric case register data, etc.). In the United States, using data from the 1993 National Hospital Discharge Survey, Olfson and Mechanic[27] found that almost 12 per cent of patients discharged from public general hospitals had a diagnosis of PD, compared with 11 per cent of patients from non-profit hospitals and 5 per cent of patients from proprietary general hospitals.

Some investigations, which compared the hospital admission rates for PD over time, allow us to assess the impact of diagnostic changes. In Denmark, sex- and age-standardized rates of first-admitted borderline patients significantly increased during the 16-year interval between 1970 and 1985, and this might be explained in terms of a change in diagnostic habits.[28] In the United States, comparing the diagnoses given to inpatients in a large university-affiliated mental hospital in the last 5 years of the DSM-II era ($n = 5143$) with those given in the first 5 years of the DSM-III era ($n = 5771$), a marked increase (from 19 per cent to 49 per cent) was found in the diagnosis of PD, together with a decrease in the diagnosis of schizophrenia and a corresponding increase in the diagnosis of affective disorders.[29]

The epidemiological findings in treated samples are especially important if we bear in mind that the presence of a PD among those suffering from other mental disorders can be a major predictor of the natural history and treatment outcome. Therefore, an important clinical implication of these findings is that patients in treatment because of severe Axis I disorders must be carefully assessed with an assessment instrument specific for PDs, because of the high likelihood of diagnosing a PD and the subsequent need to adjust their treatment accordingly.

Epidemiological studies of personality disorders carried out in other settings

A few epidemiological studies on PDs have been carried out among patients attending primary healthcare settings; in these studies between 5 and 8 per cent of patients have been identified as having a primary diagnosis of PD.[21] When the assessment is made independent of the primary diagnosis, however, the average prevalence rate can rise several-fold because of state effects. In a consecutive sample of primary care *attenders*, a PD was diagnosed in 24 per cent of the sample (N=303) and was associated with the presence of common mental disorders and unplanned surgery attendance, indicating that PDs may represent a significant source of burden in primary care.[30]

In other institutional settings, such as prisons, several studies have found very high rates of PDs. In the United Kingdom, two large-scale studies have been completed; in the first, carried out among 750 prisoners representing a 9 per cent cross-sectional sample of the entire male unconvicted population, a PD was diagnosed in 11 per cent of the sample.[31] In the second study, a representative sample of the entire prison population of England and Wales was evaluated; a sub-sample was assessed with the SCID-II administered by a clinician.[32] The prevalence rates for any PD were 78 per cent for male remand prisoners, 64 per cent for male sentenced prisoners, and 50 per cent for female prisoners. In a large meta-analysis by Fazel & Danesh [33] of 28 studies,

including a total of 13 844 prisoners, antisocial PD was diagnosed in 47 per cent of subjects. High rates of borderline and antisocial PDs have also been found in a sample ($n = 805$) of women felons entering prison in a North American State.[34]

Conclusions

Up to 30 years ago, the epidemiology of PDs had not received the same amount of attention as that of many other psychiatric disorders. Since then the situation has changed, and we now have data on the prevalence of PD in the community and in psychiatric facilities. Community data come primarily from 12 studies, with a total sample of 6 785 subjects from four countries (Germany, Norway, the United Kingdom, the United States). There are excellent national and cross-national epidemiological data on antisocial personality disorder based on the same diagnostic methods. There are almost no data on other PDs from countries other than the United States, the United Kingdom, Germany and Norway.

One important methodological problem is that some PDs have a very low prevalence rate. Consequently, epidemiological surveys carried out among the general population may require very large samples in order to identify a sufficient number of cases to study demographic correlates and the association of PD with other psychiatric disorders. Future studies should try to address this problem and provide us with more definite epidemiological data. These data will also be invaluable in showing the validity of current classifications and in better delineating the boundaries between different PDs.

Further information

The complete bibliography of studies included in tables 4.12.4.2–4.12.4.3 can be asked from the authors at the e-mail address: f.guzzetta@gmail.com.

Torgersen S. (2005). Epidemiology. In *The American Psychiatric Publishing Textbook of Personality Disorders* (eds. Oldham J.M., Skodol, A.E., Bender, D.S.), pp. 129–41. The American Psychiatric Publishing.

References

1. Black, D.W., Noyes, R., Jr., Pfohl, B., *et al.* (1993). Personality disorder in obsessive-compulsive volunteers, well comparison subjects, and their first-degree relatives. *American Journal of Psychiatry,* **150**(8) 1226–32.
2. Casey, P.R., Tyrer, P.J. (1986). Personality, functioning and symptomatology. *Journal of Psychiatric Research,* **20**(4), 363–74.
3. Coid,J. Yang, M., Tyrer, P. *et al.* (2006). Prevalence and correlates of personality disorder in Great Britain. *British Journal of Psychiatry*, **188**, 423–31.
4. Crawford, T.N., Cohen, P., Johnson, J.G., *et al.* (2005). Self-reported personality disorder in the children in the community sample: convergent and prospective validity in late adolescence and adulthood. *Journal of Personal Disorder,* **19**(1), 30–52.
5. Klein, D.N., Riso, L.P., Donaldson, S.K., *et al.* (1995). Family study of early-onset dysthymia. Mood and personality disorders in relatives of outpatients with dysthymia and episodic major depression and normal controls. *Archives of General Psychiatry*, **52**(6), 487–96.
6. Lenzenweger, M.F., Loranger, A.W., Korfine, L., *et al.* (1997). Detecting personality disorders in a non clinical population. Application of a 2-stage procedure for case identifi cation. *Archive of General Psychiatry*, **54**(4), 345–51.
7. Maier, W. (1992). Prevalence of personality disorders (DSM-III-R) in the community. *Journal of Personal Disorder*, **6**, 186–96.
8. Moldin, S.O., Rice, J.P., Erlenmeyer-Kimling, L. *et al.* (1994). Latent structure of DSM-III-R Axis II psychopathology in a normal sample. *Journal of Abnormal Psychology*, **103**(2), 259–66.
9. Reich, J., Yates, W., Nduaguba, M. (1989). Prevalence of DSM-III personality disorders in the community. *Social Psychiatry Psychiatric Epidemiology*, **24**(1), 12–6.
10. Samuels, J., Eaton, W.W., Bienvenu, O.J., III, *et al.* (2002). Prevalence and correlates of personality disorders in a community sample. *British Journal of Psychiatry*, **180**, 536–42.
11. Torgersen, S., Kringlen, E., Cramer, V. (2001). The prevalence of personality disorders in a community sample. *Archive of General Psychiatry*, **58**(6), 590–6.
12. Zimmerman, M. and Coryell, W. (1989). DSM-III personality disorder diagnoses in a nonpatient sample. Demographic correlates and comorbidity. *Archive of General Psychiatry*, **46**(8), 682–9.
13. Lenzenweger, M.F., Lane, M.C., Loranger, A.W., *et al.* (2007). DSM-IV personality disorders in the National Comorbidity Survey Replication. *Biological Psychiatry,* **62**(6), 553–64.
14. Grant, B.F., Hasin, D.S., Stinson, F.S., *et al.* (2004). Prevalence, correlates, and disability of personality disorders in the United States: results from the national epidemiologic survey on alcohol and related conditions. *Journal of Clinical Psychiatry*, **65**(7), 948–58.
15. Baron, M., Gruen, R., Rainer, J.D., *et al.* (1985). A family study of schizophrenic and normal control probands: implications for the spectrum concept of schizophrenia. *American Journal of Psychiatry*, **142**(4), 447–55.
16. Asarnow, R.F., Nuechterlein, K.H., Fogelson, D., *et al.* (2001). Schizophrenia and schizophrenia-spectrum personality disorders in the fi rst-degree relatives of children with schizophrenia: the UCLA family study. *Archives of General Psychiatry*, **58**(6), 581–8.
17. Swartz, M., Blazer, D., George, L., *et al.* (1990). Estimating the prevalence of borderline personality disorder in the community. *Journal of Personal Disorder,* (4), 257–72.
18. Zimmerman, M. and Coryell, W.H. (1990). Diagnosing personality disorders in the community. A comparison of self-report and interview measures. *Archives of General Psychiatry*, **47**(6), 527–31.
19. Nestadt, G., Romanoski, A.J., Chahal, R., *et al.* (1990). An epidemiological study of histrionic personality disorder. *Psychological Medicine*, **20**(2), 413–22.
20. Nestadt, G., Romanoski, A.J., Brown, C.H., *et al.* (1991). DSM-III compulsive personality disorder: an epidemiological survey. *Psychological Medicine*, **21**(2), 461–71.
21. de Girolamo, G. and Tyrer, P. (1993). *Personality disorders*. Geneva: WHO.
22. Dolan, B., Evans, C., Norton, K. (1995). Multiple axis-II diagnoses of personality disorder. *British Journal of Psychiatry*, **166**(1), 107–12.
23. Zimmerman, M., Rothschild, L. and Chelminski, I. (2005). The prevalence of DSM-IV personality disorders in psychiatric outpatients. *American Journal of Psychiatry*, **162**(10),1911–8.
24. Cassin, S.E. and von Ranson, K.M. (2005). Personality and eating disorders: a decade in review. *Clinical Psychology Review*, **25**(7), 895–916.
25. Erbelding, E.J., Hutton, H.E., Zenilman, J.M., *et al.* (2004). The prevalence of psychiatric disorders in sexually transmitted disease clinic patients and their association with sexually transmitted disease risk. *Sexually Transmitted Diseases*, **31**(1), 8–12.
26. Golding, M. and Pekins, D.O. (1996). Personality disorder in HIV infection. International Review of Psychiatry, (8), 253–8.
27. Olfson, M. and Mechanic, D. (1996). Mental disorders in public, private nonprofit, and proprietary general hospitals. *American Journal of Psychiatry*, **153**(12), 1613–9.
28. Mors O. (1988). Increasing incidence of borderline states in Denmark from 1970–1985. *Acta Psychiatrica Scandinavica*, **77**(5), 575–83.
29. Loranger, A.W. (1990). The impact of DSM-III on diagnostic practice in a university hospital. A comparison of DSM-II and DSM-III in 10,914 patients. *Archives of General Psychiatry*, **47**(7), 672–5.

30. Moran, P., Jenkins, R., Tylee, A., *et al.* (2000). The prevalence of personality disorder among UK primary care attenders. *Acta Psychiatrica Scandinavica*, **102**(1), 52–77.
31. Brooke, D., Taylor, C., Gunn, J., *et al.* (1996). Point prevalence of mental disorder in unconvicted male prisoners in England and Wales. *British Medical Journal*, **313**(7071), 1524–7.
32. Singleton, N., Meltzer, H., Gatward, R., *et al.* (1998). *Psychiatric morbidity among prisoners in England and Wales*. London: Office of National Statistics.
33. Fazel, S. and Danesh, J. (2002). Serious mental disorder in 23 000 prisoners: a systematic review of 62 surveys. *Lancet*, **359**(9306), 545–50.
34. Jordan, B.K., Schlenger, W.E., Fairbank, J.A., *et al.* (1996). Prevalence of psychiatric disorders among incarcerated women. II. Convicted felons entering prison. *Archives of General Psychiatry*, **53**(6), 513–9.

4.12.5 Neuropsychological templates for abnormal personalities: from genes to biodevelopmental pathways

Adolf Tobeña

The scaffolding of personality

Research on human personality has converged upon a 'consensual pathway' indicating that a small number of dimensions can provide the framework for describing the rich variety of human temperaments. These high-level temperamental traits are factorially derived from psychometric measures of individual variation in behaviour, feeling, and thinking,[1,2] and it is assumed that they may reflect the operation of brain systems that are probably multifaceted and multipurpose.[3–5] This global outline of the structure of personality depends on the notion that genetic and developmental dispositions combine with critical nurturing and social conditioning events to form the tapestry of human uniqueness within temperamental clusters. In other words, personality types are expressed through relatively clear-cut and stable phenotypic traits that are accessible to objective measurement at behavioural, emotional, and cognitive levels. These depend, in turn, upon the specific and early organization of particular neurocognitive and neuroendocrine templates.

A handful of 'superfactors' (broad traits or dimensions) apparently capture the essential components of the mosaic of terms and traits used to describe normal personality. These dimensions are *neuroticism*, *extraversion*, *agreeableness/friendliness*, *conscientiousness*, and *intellectual openness*. Neuroticism and extraversion always appear as main stars in these factorial *solutions* whereas the remaining three superfactors—conscientiousness (reliability/persistence), friendliness (as opposed to aggressiveness/hostility) and intellectual curiosity (openness/creativity)—have less regularity on such high-order taxonomies. A five-dimensional structure is advocated by many researchers though dissent is still strong regarding the nature and scope of these superfactors that would define the 'core' of human temperament.[6]

Biological rooting of personality types

Searching for biological substrates of personality dimensions would reinforce their validity as useful constructs but this endeavour was largely neglected by psychometricians devoted to purely descriptive studies and by clinical researchers as well. Some pioneers, like Hans Eysenck at the Institute of Psychiatry, London, tried to root behavioural trait variations within neurobiological concepts[1] following a venerable tradition, which can be traced back as far as Pavlov. These early proposals were rather rudimentary though they served as drivers of subsequent models which focused more tightly on certain brain subsystems as possible sites for the factors underlying normal and abnormal temperaments.[3–5,7–9] Jeffrey Gray, Robert Cloninger, and Larry Siever's ideas were among the more fruitful in an area which has grown steadily and is now an lively field of personality research.[9–11] Progress in basic neuroscience has made it possible to relate a variety of biological measures to paper and pencil or neurocognitive tests distinguishing normal and anomalous temperaments. Biological screening has also increasingly been applied to patients with personality disorders, using the clinical clusters as defined by Diagnostic Systems. Besides these attempts to build psychobiological profiles of normal and abnormal temperaments, converging evidence is used to advocate that categorical and dimensional models for diagnosing personality disorders should be integrated.[12]

To give a broad overview of an area that may be crucial to illuminate the genesis of personality disorders, I shall discuss the studies that, during the last decade, have tried to find genetic traces for personality traits that are both behaviourally consistent and biologically well rooted. Previous work using classical (familial or twin) methods had found substantial heritability estimates for several personality traits.[13] It was thus unsurprising that genetic tracking methods impulsed research aimed at showing that temperamental traits contribute to personality scaffolding via neuroendocrine targets specified by particular genes. I'll be discussing the outcome of some of these efforts and I'll explore afterwards how other basic temperamental traits, rooted within biodevelopmental processes, do mediate enduring neurocognitive organization resulting in long-lasting behavioural styles. Finally I'll outline new avenues for the neuropsychology of personality. My approach is deliberately selective, discussing relevant evidence rather than performing a systematic assessment of the field. For reasons of convenience and possible clinical relevance, I have selected some of the traits heralding sound biological foundations, although they are not necessarily prominent in the state-of-the-art dimensional '*solutions*' for normal and abnormal temperaments.

The genetic saga for novelty-seeking

In 1996 two independent teams reported[14,15] that a particular chromosomal 'locus' was associated with a well-established trait of human temperament—the hunger for novelty and excitement that lies behind sensation-seeking, risk taking, and impulsive behaviours.[5,10] A polymorphism in the sequence of the gene expressing the D4 dopamine receptor (D4DR), located on the short arm of chromosome 11, explained 10 per cent of the genetic variance due to this trait. Individuals with the longer repeat allele at exon III of the *D4DR* gene scored higher in novelty-seeking

behaviour (explorers, risk-seekers), whereas those with the shorter allele had lower scores (prudent, cautious). The first of these studies[14] investigated a heterogeneous sample of young Israelis, and showed the association to be independent of ethnicity (Ashkenazim versus Sephardim), sex, or age. The second study, carried out in the United States,[15] used a random sample of people who had initially been recruited in a search for chromosomal regions possibly associated with sexual orientation; this sample mainly comprised white men, although some ethnic minorities were also included. The personality questionnaires were different but very popular in personality research: the Israelis were evaluated using Cloninger's Tridimensional Personality Questionnaire (**TPQ**),[16] which gives direct scores of novelty-seeking, whereas the American study used the Revised NEO Personality Inventory[17] which measures the five superfactors mentioned above, from which scores for novelty-seeking were derived. The results of the two studies were highly concordant.

Despite the modest explanatory power of this reported association, the link between temperamental variability for one trait and a chromosomal polymorphism was the first hint for a direct relationship between a putative 'genetic marker' and a core dimension of normal personality. In this case, the potential genetic marker appeared promising because of the amount of basic and clinical research linking dopaminergic function with the regulation of stimulus-seeking and sensitivity to incentives. In theory, if similar degrees of explained variance were assignable to other sound gene markers associated to *approach/exploring* phenotypes, a substantial part of the heritability of the trait could be explained. Subsequent studies[18,19] failed to replicate these early findings in a consistent way and the optimism receded. The heterogeneity of the samples and the subtleties of the genetics of complex traits were blamed for the disparate results, though the research saga was quite productive: the links between dopamine receptor polymorphisms and novelty-seeking have been intensively searched and the race to find other markers for the same trait was impressive.

Metanalyses suggest that there are subtle connections between dopamine receptor gene variants and *approach/exploring* propensities as measured by personality questionnaires, though the strength of the contribution of every variant is small and hard to establish.[19,20] Moreover, parallel research has established suggestive connections between gene variants regulating other molecular targets (i.e. tryptophan hydroxylase, dopamine transporter, dopamine-beta-hydroxylase, serotonin transporter, MAOA, COMT) that modulate risk-taking behaviours and impulsivity. A handful of genes, thus contribute to differential vulnerabilities for addictive behaviours, a congruent result at the extreme of stimulus-seeking tendencies. Although further and more refined research is required, these data seem to confirm pioneer work, mostly with twins, which had consistently established that novelty-seeking behaviour was moderately heritable (40 to 50 per cent).

Genetics of fearfulness/neuroticism

The aforementioned American team that reported the first associations between novelty-seeking and variants of D4DR gene informed that there was an association between the neuroticism trait and a chromosomal region linked to serotonin neurotransmission involved in modulating anxiety-related traits.[21] The 5-hydroxytryptamine transporter protein (5-HTT) that promotes the reuptake of serotonin into cell membranes is encoded by a gene (*SLC6A4*) located in the q11–q12 segment of chromosome 17. The region governing the transcriptional control of the protein shows a polymorphism that influences its expression and functioning. Individuals carrying the short variant of the polymorphism show a reduced efficiency of serotonin reuptake compared with those possessing the longer variant. The study measured these parameters in the lymphoblasts of two independent samples totalling more than 500 volunteers. Using two different personality questionnaires (NEO and Cattell's 16PF Personality Inventory) and estimated scores on various dimensions of Cloninger's TPQ, the evidence showed that subjects who carried the short variant in the 5-HTT gene polymorphism had higher neuroticism (NEO), anxiety (16PF), and harm-avoidance (TPQ) scores. The results were equally consistent across and within pedigrees. Across the three personality measures, the 5-HTTLPR contributed a modest 3 to 4 per cent of the total variance and 7 to 9 per cent of the genetic variance in anxiety-related traits. It was suggested that, if other genes contributed similar dosage effects to anxiety traits, approximately 10 to 15 genes might be involved in the heritability of neuroticism.

The implication of the serotonin transporter in the potential genetic predisposition towards emotionality traits agrees with many other results. Serotonergic neurotransmission is involved in multiple brain functions with little or no relationship with fear/anxiety regulation, but there is a large body of evidence linking it with the modulation of adaptive responses to serious conflict and emotionally demanding situations.[3] Moreover, many drugs currently used to treat anxious/depressive dysphorias and personality disorders depend on mending serotonergic function. Finally, several studies have shown that variants of the 5-HTTLPR predict differential response on anxious phenotypes: fear-driven amygdala activation[22] and response to pharmacological challenges.[23] Therefore, although the exact role of serotonergic systems in the modulation of emotionality is not fully understood, it is improbable that the neurohormonal adaptations that participate in individual responses to serious emotional conflicts would not include serotonin modulation either through the cell transporter or through the extended family of serotonin receptors and their intracellular targets. Defensive adaptations require however the participation of other central neuromodulators: the CRH-ACTH regulatory cascade, γ-aminobutyric acid, neuropeptide Y, and substance P,[3] are major contenders in this respect and they can be expected to contribute to the genetic mediation of neuroticism.

Polymorphisms in anxious humans versus QTL-genes for fearful rodents

Dozens of studies investigated whether the particular polymorphism in the 5-HTT gene contributes to the tendency for individuals who score higher on neuroticism, in personality tests, to be at higher risk for 'internalizing disorders' (anxiety/depression) and personality disorders (anxiety/affective clusters). The global outcome of that research has been unreliable: though stringent metanalyses confirmed the original association with a modest relevance at explaining the trait variance,[23] subsequent results in large samples of siblings and singletons have been negative.[24] There are other lines of evidence, however, mainly from animal research, that support the claim of a possible genetic basis for fearfulness/emotionality. This evidence is derived from studies of the

psychogenetics of emotional susceptibility searching for chromosome loci. In many biological and behavioural tests, comparisons of several reactive and non-reactive strains of mice and rats obtained through artificial selection (forced interbreedings) have narrowed the search for genetic loci thanks to increasingly powerful methods of chromosomal mapping. In a pioneer work with progeny obtained by crossing two strains of mice selected for activity and defecation in an open-field test, three loci (QTLs) which explained most of the genetic variance in emotionality were found on chromosomes 1, 12, and 15 of the murine genome.[25] These data were confirmed and extended by measuring fear responses towards particular cues: the same segment of chromosome 1 was identified as a relevant 'locus' for emotional susceptibility besides other murine chromosomal zones.[9] The importance of the loci at chromosome 1 has been established in studies using heterogeneous and inbred stocks, combining techniques which have permitted to focus the suspicious segment to less than 1 cM and leading to the identification of the first gene linked to murine emotionality: Rgs2, a regulator of G-protein functioning which is highly expressed in the brain.[26] The complexities are nevertheless tremendous because even a QTL like that contains several genes each contributing a very modest part of variation on the phenotype of interest.[9,26]

Several research programmes were concurrently started to determine whether there is also concordance in the chromosomal marking of emotionality in strains of rats differing in fearfulness. The hypoemotional Roman high-avoider (**RHA**) versus the hyperemotional Roman low-avoider (**RLA**) rat lines represented a particularly interesting assay because of the very large body of evidence showing their usefulness as animal models of 'temperamental' styles.[27] Several plausible QTLs were detected but only those located on chromosome 5 and 15 predicted a wide array of anxious/fearful behaviours including spontaneous and learned fears.[28,29] The QTL located at the middle of rat chromosome 5 looks particularly promising because this region is partially syntenic with the human 1p segment where QTLs for neuroticism and liabilities for anxious/depressive dysphorias have been detected.[9,30] The search for plausible genes is, then, much more focused now though they will explain, in all probability, very modest portions of the phenotypic trait variance.[9]

Biodevelopmental mechanisms for affiliative traits

Affiliation (friendliness, sociability, gregariousness, empathy) is another core personality trait that can be measured consistently. Poor or distorted affiliative behaviour is the most predictive marker for a reliable diagnosis of personality disorder.[7,31] Deficits or alterations in affiliative tendencies may show a variety of clinical manifestations: extreme aloofness and detachment, manipulative, non-empathic or exploitative attitudes, and even definite asocial or antisocial tendencies. These behavioural styles appear, in different degrees and combinations, in several clinical categories of abnormal temperament. They could reflect alterations in the functioning of neuroendocrine systems specialized in mediating affective attachments, possibly including subsystems for social reward and social distress.[31] In this respect, an impressive amount of evidence has been gathered (mostly in animals but in humans as well) showing that prosocial behaviours, such as maternal nurturing, friendliness/gregariousness, playful/sexual behaviours, and even cooperativeness in economic interactions, share some neurochemical controls.[31–33]

Central oxytocin and opioid systems are among the more relevant of these modulatory molecules, because several types of attachments (mother–infant bonds, young peer bonds, pair-mating bonds, in-group tendencies) are dramatically altered when the functions of oxytocinergic or opioid systems in the brain are modified. Other centrally acting neuropeptides, such as prolactin and vasopressin, also contribute to particular types of species-specific social bonding. In addition, both the central regulatory monoaminergic systems and the corticotrophin–adrenal cascades that mediate stress adaptations help to organize responses to everyday social challenges.[31] The application of these ideas to personality is still in its infancy and requires the development of consistent scales to be related with sound biological markers.

Neuropeptides, social bonding, and early rearing practices

Theory and research in the psychobiology of social attachments[8,31] has linked the impact of early rearing practices (secure/nurturing mothering versus peer rearing or isolation) with the future organization and functioning of several neuroendocrinological systems. This research has mainly explored the function of the central modulatory monoamines norepinephrine, dopamine, and serotonin, and the hypophyseal–adrenal axis responses to social challenges. The evidence has shown that socially deprived monkeys differ physiologically, behaviourally, and cognitively from mother-reared infants in almost every aspect of what it means to be a social monkey; if the privation extends over the first 6 to 9 months of post-natal life, most effects persist into the adulthood. According to Kraemer[34]: 'The way in which socially deprived individuals orient to and respond physiologically to social stressors is altered and the kind of behavioural differences that seem to be the most important are those usually assigned to the domain of "temperament"'.

With the addition of the central neuropeptide systems that specifically modulate affiliative tendencies, the study of some crucial experiences during early infancy (and probably adolescence) will provide clues to the clarification of the role that developmental processes play in shaping attachment styles. Neural organization depends, to a great extent, on critical environmental inputs to produce enduring behavioural and cognitive adaptation in all domains. Therefore ontogenic factors must be particularly important in modelling social behaviour and in sustaining profiles of affiliative versus non-affiliative temperaments, in the same way as has already been demonstrated for other traits. For instance, in reactive/fearful monkeys, maternal and even grandparental rearing practices can significantly modify future neuroendocrine and behavioural adaptations[35] with parallel data obtained in rats.[36]

Affiliative genes?

These findings do not exclude the participation of genetic dispositions in attachment styles. Some authors have suggested that the operation of 'communicative' or 'affiliative' genes could prime individual tendencies through different sorts of emotional affects.[37] There is a paucity of data on the putative genetic basis of particular attachment styles. When a molecular approach has been used,

in rodents, to establish a genetic link between variants of the vasopressin V1a receptor gene with a conspicuous social behaviour such as monogamous pair bonding, the results have been spectacular. After substituting and inserting the V1a receptor gene characteristic of a monogamous species (the prairie voles), both promiscuous mice and meadow voles adopted the pattern of partner preferences and parental behaviours distinctive of the monogamous species. It has been shown, in fact, that socio-behavioural trait differences depend on polymorphisms on the regulatory region of the V1a gene.[38] In rodents and other mammals the neurochemical circuits in the brain regulating attachment/affiliative behaviours are increasingly detailed. In humans the task is just starting and will deeply influence personality research.

Biodevelopmental processes for aggressiveness

Aggressiveness is another temperamental trait that has a very well-founded biodevelopmental basis.[39] Dimensional descriptions of the structure of personality do not always include aggressiveness as a high-level factor, but it is embedded in other traits such as impulsiveness, poor control, or explosive/desinhibited behaviour. However, aggressiveness is a major behavioural trait that has to be taken into account in the routine management of mental disorders, and it is also a prevalent characteristic in personality disorders (generally as an excess, but sometimes because of its absence). In the past, it was extremely difficult to prove that individual variations in aggressive behaviour might be correlated with neurohormonal characteristics. However, this was because of inadequate methods of quantifying biological and behavioural variations.[40]

In humans, the link between lifelong aggressive profiles and familial/subcultural problems is strong indicating that lower socio-economic status, increased rates of abuse, coercive family interactions, and neglect or other adversities contribute to violent behaviour from infancy to adolescence and into adulthood.[39,41] But this cannot obscure the effects of enduring biological dispositions that could, in part, result from the influence of socio-environmental insults to the developing brain. Intensive research in behavioural neuroendocrinology and molecular neuroscience using animal models has shown that aggressive temperaments are associated with both genetic dispositions and critical developmental processes, which affect the function of neuroregulatory controls that either promote or inhibit aggression.

MAO-A functioning and aggression-proneness

It was not surprising, therefore, that an association between a specific gene and an aggressive disorder in humans was early reported[42]: the males of a Dutch family that carried a mutation of the gene for monoaminoxidase A (MAO-A gene situated at chromosome Xp11.23) had a record of severe aggressive incidents in different generations. Subsequent investigations in mice showed that ablating the gene for MAO-A resulted in the 'knockout' mutants being much more aggressive[43] MAO plays an important part in the breakdown of serotonin and other central monoamines, and previous research in humans had found consistent relationships between impulsive or 'poorly controlled' behavioural styles and low-MAO activity, a feature also seen in sensation-seekers. A common polymorphism in the promoter region of the MAO-A gene has been shown to interact with early abuse/neglect in longitudinal studies of large samples of children[44]: only abused kids carrying the genetic variant giving low-MAO-A activity are later at risk of antisocial tendencies or violent/criminal behaviour, during adolescence or young adulthood. Further studies with normative samples differentiated by this polymorphism have permitted to map, through structural and functional neuroimaging, cortico-limbic singularities associated with emotional regulation in the low-MAO-A carriers.[45]

Serotonin/vasopressin ratios

Many other genetically altered animals have been produced that are highly aggressive. Several of them typically show anomalies in serotonin function. Knockout mice that do not express the serotonin 5-HT1B receptors are much quicker to attack intruders[46] an action that can be blocked with targeted serotonergic drugs. All this adds to the evidence obtained from mentally ill patients presenting aggressive outbursts and from chronic offenders in prisons, which shows an inverse relationship between serotonergic function and violent attacks directed towards others or themselves. Taken together, the data appear to suggest that preserved brain serotonin function helps to attenuate aggressive impulses, probably by blocking other neuroregulatory systems that promote aggression such as sex steroids, insulin, vasopressin, and others. Vasopressin is involved in the establishment of pair bonds and territorial tendencies in mammals, and promotes aggressive behaviour when social/dominance challenges are perceived. There is evidence from adolescent and juvenile hamsters showing that proserotonergic interventions attenuate vasopressin-induced attacks. Patients with personality disorders[47] showed that cerebrospinal fluid levels of vasopressin were positively correlated with a life history of aggression, and with attacks against persons in particular. This was a more poweful relationship than the negative one usually obtained from measurements of serotonergic function and aggression. Other neuromodulators help serotonin to attenuate aggressive outputs: in animals, highly specific genetic techniques of knocking down or targeted regional brain expression of steroid and oxytocin receptors, have been used to identify ensambles of neuromodulators devoted to control aggressiveness, to the point of postulating a landscape of 50 genes to get a sound description to this function.[48]

A prospective landscape including other traits

Temperamental styles in animals show differential clustering of behavioural traits, which correspond to specific (and very complex) neurohormonal profiles. Sometimes, however, a specific genetic modification is sufficient to promote a fully differentiated temperament, as in the case of mice deficient in α-calcium–calmodulin-dependent kinase II (**α-CaMKII**) which show decreased fear responses and increased defensive aggression associated with low-serotonin levels[49] (Table 4.12.5.1).

It is possible that this may also hold for exceptional temperamental combinations in humans. However, we have already established that, when trying to explain the genetic contribution to basic ('universal') personality traits, a multigenic/interactional approach is compulsory to explain just part of the measured variance in each trait (Fig. 4.12.5.1). Nevertheless, this contribution can be very

Table 4.12.5.1 Summary of behavioural phenotypes in α-CaMKII mutant mice

Behavioural phenotype	Heterozygote	Homozygote
Fear-related responses	Decrease	Decrease
Offensive aggression	Normal	Decrease
Defensive aggression	Increase	Decrease
Pain sensitivity	Normal	Increase
Startle response	Normal	Increase
Vigilance	Normal	Increase
Mating	Decrease	Decrease
Maze learning	Normal	Decrease

Reproduced from C. Chen *et al.* Abnormal fear response and aggressive behaviour in mice deficient for alpha–calcium–calmodulin–leinase II. *Science*, **266**, 291–4, copyright 1994, with permission from the American Association for the Advancement of Science.

important, as most of research involving twins has yielded estimates of just over 40 per cent for the genetic input to typical personality traits, with very modest estimations assignable to the so-called shared (familial/cultural) environment. Non-shared environmental influences (from the womb onwards) and genetic–environmental interactions make a well-known contribution to each individual temperament. These complex influences may act first by modelling the development of basic neuroendocrine regulatory systems that cope with natural and social challenges, and second by shaping the neurocognitive architecture that results in an autonomous and particular lifestyle. An interactional approach along these lines is now laying the foundation for a better description of the different factors that contribute to building the typical profiles of normal or abnormal personalities.[50,51]

Such a general scheme must include other traits, in addition to the ones considered so far. For instance, the detection of a substantial genetic contribution to the baseline level of happiness,[52] which all individuals show throughout life independently of the events or episodes that they encounter, must be important for

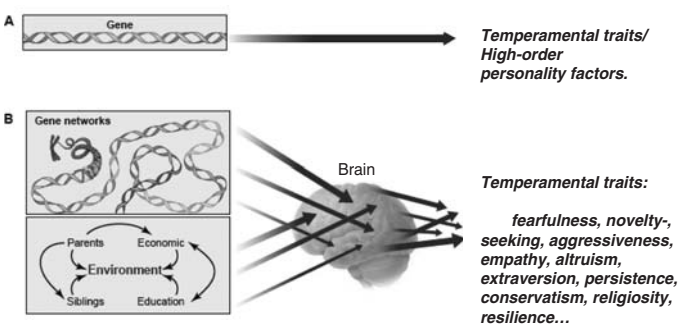


Fig. 4.12.5.1 Two views of the relationships between genes and personality. **A**. Early studies looked for linear relationships between gene markers and temperamental traits or high-order personality factors. **B**. Reality is likely to be far more complex with gene networks interacting with environmental inputs impacting on brain development and leaving enduring neural dispositions that, in turn, influence behavioural, cognitive, and affective styles (temperamental traits). (Reproduced from D. Harner. Rethinking behavior genetics, *Science*, **298**, 71–2, copyright 2004, with permission from the American Association for the Advancement of Science.).

personality diagnoses, because such a 'calibration point' has a direct relationship with the affective tone of optimism or pessimism. Moreover, other traits and measures in the domain of cognitive performance (e.g. attentional spans, perceptual/appraisal reliability, thinking styles, and memory biases) or of character (e.g. religiosity/transcendence, conservatism, altruism, self-directness/self-esteem, and the drive to achieve/enthusiasm) should be incorporated into the whole description of personality structure. This is essential if the aim is to produce a general framework powerful enough to contain the complex categories that clinicians have tried to construct on the basis of systematic observation for more than a century. Cloninger[8,11] and others have advanced proposals along these lines, which have functioned as useful steps to guide empirical work. It must be added that structural and functional neuroimaging techniques have been used to map relations between personality dimensions and brain's regional organization and patterns of activation/deactivation. Although these studies are very preliminary and have been done with convenient samples, they have offered hints of neurocognitive architectures, which tend to confirm, on a broad basis, the neurochemical mechanisms behind the main temperamental traits.[53–55]

Sophisticating the diagnoses of personality disorders

The evidence discussed so far is starting to have a major impact on the reconceptualizing of the approach that has been applied in psychiatry to the detection and categorization of the elusive profiles of normal and abnormal personalities. It is clear that anchoring some temperamental traits to a sound genetic or biodevelopmental base, and accruing a plausible neurocognitive architecture for it, will not provide a complete solution to the problems encountered in the aetiology and diagnosis of personality disorders. Many additional steps will have to be worked out. But in view of the increasing links that are being established between core personality traits and some genetic/developmental mechanisms, the task of building a solid neuropsychological framework, which would allow improved identification and differentiation of abnormal personalities no longer seems hopeless. In fact, such research is opening up many new avenues for understanding the effects of different factors (innate dispositions, neurodevelopmental organization, neurocognitive architectures, critical social transitions, and repeated stress episodes) on an individual's vulnerability to developing a personality disorder.

Hence, the use of behaviourally well defined and biologically well-established personality measures must be the starting point for achieving the fine-tuned diagnoses increasingly required in modern medicine. Complexity in measurement will increase, but there is no other way of obtaining data that are sufficiently valid to allow understanding of the classical personality disorders. It is hoped that advances in the neuropsychological detection of the more salient 'clinical' profiles within each of the personality subspaces or 'clusters' (at the level of either traits or dimensions), together with a refinement of the neurocognitive and neurohormonal data, will produce much better solutions. Some of the old-established and consistent categories of personality disorders will be confirmed, but it is possible that unexpected 'types' of abnormal personalities, with clinical relevance, will emerge. In this new

framework it may be easier perhaps to detect at an early stage those 'exceptional' and 'charismatic', although anomalous, personalities who often impose great social costs and dramatic consequences not only for themselves but also for the group or the society in which they live.[56]

Further information

Archer, J. (2006). Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neuroscience and Biobehavioral Reviews*, **30**, 319–45.

Caspi, A., Roberts, B.W., and Shiner, R.L. (2005). Personality development: stability and change. *Annual Review of Psychology*, **58**, 453–84.

Gosling, S.D. (2001). From mice to men: what can we learn about personality from animal research. *Psychological Bulletin*, **127**, 45–86.

Hariri, A.R. and Holmes, A. (2006). Genetics of emotional regulation: the role of serotonin transporter in neural function. *Trends in Cognitive Sciences*, **10**, 182–91.

Kreek, M.J., Nielsen, D.A., Butelman, E.R., *et al.* (2005). Genetic influences on impulsivity, risk taking, stress responsivity vulnerability to drug abuse and addiction. *Nature Neuroscience*, **8**, 1450–7.

Markon, K.E., Krueger, R.F., Bouchard, T.J. Jr, *et al.* (2002). Normal and abnormal personality traits: evidence for genetic and environmental relationships in the Minnesota study of twins reared apart. *Journal of Personality*, **70**, 660–93.

McGue, M. and Bouchard, T.J. (1998). Genetic and environmental influences on human behavioral differences. *Annual Review of Neuroscience*, **21**, 1–24.

McNaughton, N. and Corr, P..J. (2004). A two-dimensional neuropsychology of defense: fear/anxiety and defensive distance. *Neuroscience and Biobehavioural Reviews*, **28**, 285–305.

Paris, J. (2005). Neurobiological dimensional models of personality: a review of the models of Cloninger, Depue and Siever. *Journal of Personality Research*, **19**, 156–70.

Tobeña, A. (2004). *Deadly martyrs: the biology of lethal altruism*. Bromera Publish.-Valencia University Press, Valencia, Spain.

Young, L.J. and Wang, Z. (2005). The neurobiology of pair bonding. *Nature Neuroscience*, **7**, 1048–54.

References

1. Eysenck, H.J. (1967). *The biological basis of personality*. Charles Thomas, London.
2. McRae, R.R. and Costa, P.T. (1997). Personality trait structure as a human universal. *The American Psychologist*, **52**, 509–16.
3. Gray, J.A. and McNaugthon, N. (2000). *Anxiety: an enquiry into the functions of the septohippocampal system* (2nd edn). Oxford University Press, New York.
4. Cloninger, C.R. (1987). A systematic method for clinical description and classification of personality variants. *Archives of General Psychiatry*, **44**, 573–88.
5. Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking*. Cambridge University Press, New York.
6. Sheller, J. and Westen, D. (2004). Dimensions of personality pathology: an alternative to the five-factor model. *The American Journal of Psychiatry*, **161**, 1743–54.
7. Siever, L.J. and Davis, K.L. (1991). A psychobiological perspective on the personality disorders. *The American Journal of Psychiatry*, **148**, 1647–58.
8. Cloninger, C.R., Svrakic, D.M., and Przybeck, T.R. (1993). A psychobiological model of temperament and character. *Archives of General Psychiatry*, **50**, 975–90.
9. Flint, J. (2004). The genetic basis of neuroticism. *Neuroscience and Biobehavioural Reviews*, **28**, 307–16.
10. Corr, P.J. (2004). Reinforcement sensitivity theory and personality. *Neuroscience and Biobehavioural Reviews*, **28**, 317–32.
11. Sravic, D.M., Draganic, S., Hill, K., *et al.* (2002). Temperament, character and personality disorders: ethiogic, diagnostic, treatment issues. *Acta Psychiatrica Scandinavica*, **106**, 189–95.
12. Widiger, T.A. and Frances, A.J. (2001). Towards a dimensional model for personality disorders. In *Personality disorders and the five-factor model of personality* (2nd edn) (eds. P.T. Costa and T.A. Widiger). American Psychological Association Press, Washington, DC.
13. Markon, K.E., Krueger, R.F., Bouchard, T. J. Jr, *et al.* (2002). Normal and abnormal personality traits: evidence for genetic and environmental relationships in the Minnesota study of twins reared apart. *Journal of Personality*, **70**, 660–93
14. Ebstein, E.P., Novick, O., Umansky, R., *et al.* (1996). Dopamine D4 receptor (D4DR) exon III polymorphism associated with the human personality trait of novelty seeking. *Nature Genetics*, **12**, 78–80.
15. Benjamin, J., Li, L., Greenberg, B.D., *et al.* (1996). Population and familial association between the D4 dopamine receptor gene and measures of novelty seeking. *Nature Genetics*, **12**, 81–4.
16. Cloninger, C.R., Przybeck, T.R., and Svrakic, D.M. (1991). The tridimensional personality questionnaire: US normative data. *Psychological Reports*, **69**, 1047–57.
17. Costa, P.T. and McCrae, R.R. (1992). *Revised NEO personality inventory and NEO five inventory*. Psychological Assessment Resources, Odessa, FL.
18. Lusher, J.M., Chandler, C., and Ball, D. (2001). Dopamine D4 receptor gene (DRD4) is associated with novelty seeking (NS) and substance abuse: the saga continues. *Molecular Psychiatry*, **6**, 497–9.
19. Shinka, J.A., Letsch, E.H., and Crafword, F.C. (2002). DRD4 and novelty seeking: results of a metanalyses. *American Journal of Medical Genetics*, **44**, 643–8.
20. Munafo, M.R., Clark, T.G., Moore, L.R., *et al.* (2003). Genetic polymorphisms and personality in healthy adults: a systematic review and metanalyses. *Molecular Psychiatry*, **8**, 471–84.
21. Lesch, K.P., Bengel, D., Heils, A., *et al.* (1996). Association of anxiety related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science*, **274**, 1527–31.
22. Hariri, A.R., Mattay, V.S., Tessitore, A., *et al.* (2002). Serotonin transporter genetic variation and the response of the human amygdala. *Science*, **297**, 400–03.
23. Smith, G.S., Lothich, F.E., Molhutra, A.K., *et al.* (2004). Effects of serotonin transporter promoter polymorphisms on serotonin function. *Neuropsychopharmacology*, **29**, 2226–34.
24. Willis-Owen, S.A.G., Turri, M.G., Munafò, M.R., *et al.* (2005). The serotonin transporter length polymorphism, neuroticism and depression: a comprehensive assessment of association. *Biological Psychiatry*, **58**, 451–6.
25. Flint, J., Carley, R., DeFries, J.C., *et al.* (1995). A simple genetic basis for a complex psychological trait in laboratory mice. *Science*, **268**, 1432–5.
26. Yakin, B., Willis-Owen, S.A.G., Fullerton, J., *et al.* (2004). Genetic dissection of a behavioral quantitative trait locus shows the Rgs2 modulates anxiety in mice. *Nature Genetics*, **36**, 1197–202.
27. Driscoll, P., Escorihuela, R.M., Fernández-Teruel, A., *et al.* (1998). Genetic selection and differential stress responses: the Roman lines/strains of rats. *Annals of the New York Academy of Sciences*, **851**, 521–30.
28. Fernandez-Teruel, A., Escorihuela, R.M., Gray, J.A., *et al.* (2002). A quantitative trait locus influencing anxiety in the laboratory rat. *Genome Research*, **12**, 618–26.
29. Aguilar, R., Gil, L., Flint, J., *et al.* (2002). Learned fear, emotional reactivity and fear of heights: a factor analytic map from a large F2 intercross of Roman rat strains. *Brain Research Bulletin*, **57**, 17–26.
30. Nash, M.W., Huezo Diaz, P., Williamson, R.J., *et al.* (2004). Genome-wide linkage analyses of a composite index of neuroticism and mood related scales in extreme selected sibships. *Human Molecular Genetics*, **13**, 2173–82.

31. Panksepp, J. (1998). *Affective neuroscience*. Oxford University Press, New York.
32. Insel, T.R., Young, L., and Wang, Z. (1997). Molecular aspects of monogamy. *Annals of the New York Academy of Sciences*, **807**, 302–16.
33. Kosfeld, M., Heinrichs, M., Zak, P.J., *et al.* (2005). Oxytocine increases trust in humans. *Nature*, **435**, 673–6.
34. Kraemer, G.W. (1997). Psychobiology of early social attachment in rhesus monkeys. In *The integrative neurobiology of affiliation* (eds. C.S. Carter, I.I. Lederhendler, and B. Kirpatrick). *Annals of the New York Academy of Sciences*, **807**, 401–18.
35. Suomi, S.J. (1991). Uptight and laid-back monkeys: individual differences in the response to social challenges. In *Plasticity and development* (eds. S.E. Brauth, W.S. Hall, and R.J. Dooling), pp. 27–56. MIT Press, Cambridge, MA.
36. Liu, D., Diorio, J., Tannenbaum, B., *et al.* (1997). Maternal care, hippocampal glucocorticoid receptors and hypothalamic-pituitary-adrenal responses to stress. *Science*, **277**, 1659–62.
37. Buck, R. and Ginsburgh, B. (1997). Communicative genes and the evolution of empathy: selfish and social emotions as voices of selfish and social genes. *Annals of the New York Academy of Sciences*, **807**, 481–3.
38. Hammock, E.A.D. and Young, L.J. (2005). Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science*, **308**, 1630–4.
39. Tremblay, R.E., Hartup, W.W., and Archer, J. (eds.) (2005). *Developmental origins of aggression*, Guilford Books, New York.
40. Adams, D.B. (2006). Brain mechanisms of aggressive behavior: an updated review. *Neuroscience and Biobehavioral Reviews*, **30**, 304–18.
41. Ferris, C.F. and Grisso, T. (eds.) (1996). Understanding aggressive behavior in children. *Annals of the New York Academy of Sciences*, **794**, 98–103.
42. Brunner, H.G., Nelen, M., Breakefiled, X.O., *et al.* (1994). Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. *Science*, **262**, 578–80.
43. Cases, O., Seif, I., Gromsby, J., *et al.* (1995). Aggressive behavior and altered amounts of brain serotonin and noradrenaline in mice lacking MAO-A. *Science*, **268**, 1763–6.
44. Caspi, A., McClay, J., Moffit, T., *et al.* (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, **297**, 851–4.
45. Meyer-Lindenberg, A., Buckholtz, J.W., Kolachana, B., *et al.* (2006). Neural mechanisms of genetic risk for impulsivity and violence in humans. *PNAS*, **103**, 6269–74.
46. Saudou, F., Amarada, G., Dierich, A., *et al.* (1994). Enhanced aggressive behavior in mice lacking 5-HT1b receptor. *Science*, **265**, 1875–8.
47. Coccaro, E.F., Kavoussi, R.J., Hauger, R.L., *et al.* (1998). Cerebrospinal fluid vasopressin levels: correlates with aggression and serotonin function in personality-disordered patients. *Archives of General Psychiatry*, **55**, 708–14.
48. Ogawa, S., Choleris, E., and Pfaff, D. (2004). Genetic influences on aggressive behaviors and arousability in animals. *Annals of the New York Academy of Sciences*, **1036**, 257–66.
49. Chen, C., Rainnie, D.G., Greene, R.W., *et al.* (1994). Abnormal fear response and aggressive behavior in mice deficient for alpha-calcium-calmodulin-kinase II. *Science*, **266**, 291–4.
50. Saudino, K.J. (2005). Behavioral genetics and child temperament. *Journal of Developmental and Behavioral Pediatrics*, **26**, 214–23.
51. Blackburn, R., Logan, C., Renwick, S.D., *et al.* (2005). Higher-order dimensions of personality disorder: hierarchical structure and relationships with the five factor model, the interpersonal cycle and psychopathy. *Journal of Personality Disorders*, **19**, 597–623.
52. Lykken, D.T. and Tellegen, A. (1996). Happiness is a stochastic phenomenon. *Psychological Science*, **7**, 186–9.
53. Pujol, J., Lopez, A., Deus, J., *et al.* (2002). Anatomical variability in the anterior cingulate gyrus and basic dimensions of personality, *Neuroimage*, **15**, 847–55.
54. Kumari, V., Ffytche, D.H., Williams, S.C.R., *et al.* (2004). Personality predicts brain responses to cognitive demands. *The Journal of Neuroscience*, **24**, 10636–41.
55. O'Gorman, R.L., Kumaro, V., Williams, S.C.R., *et al.* (2006). Personality factors correlate with regional cerebral perfusion. *Neuroimage*, **31**, 489–95.
56. Henry, D., Geary, D., and Tyrer, P. (1993). Adolf Hitler: a re-assessment of his personality status. *Irish Journal of Psychological Medicine*, **10**, 148–51.

4.12.6 Psychotherapy for personality disorder

Anthony W. Bateman and Peter Fonagy

Introduction

Psychotherapy has historically been the mainstay of treatment for personality disorder (PD). It remains so. Psychoanalysis was probably the earliest formal treatment for PD, which led to the first clinical descriptions of borderline personality disorder. A parallel but linked development was the application of psychoanalytic ideas in therapeutic communities which have been in existence for over 60 years and remain a treatment context and method for patients with PD. It was only in the 1960s that modified psychotherapeutic treatments were developed. Initially these were based on psychodynamic understanding of PD, but gradually other theoretically and practically driven models have developed, leading to the current situation in which there are behavioural, cognitive, dynamic, and supportive treatments offered in a range of contexts. Some of these methods have more empirical support than others. These methods will be described in this chapter.

Psychological therapies for personality disorders take place against the background of the natural course and outcome of the disorder. Until recently, the natural history of personality disorder had not been systematically studied. Several major cohort follow-along studies have yielded surprising data concerning the rate of symptomatic remissions in a disorder that was assumed to have a lifelong course.[1] For example, over a 10-year follow-along period, 88 per cent of those initially diagnosed with borderline personality disorder appeared to remit in the sense of no longer meeting DIB-R or DSM-III criteria for BPD for 2 years.[2] The symptoms that remit most readily, irrespective of treatment, appear to be the acute ones, such as parasuicide and self-injury, which are the most likely to trigger psychotherapeutic intervention. Temperamental symptoms, such as angry feelings and acts, distrust and suspicion, abandonment concerns, and emotional instability, appear to resolve far more slowly. In the Collaborative Longitudinal Personality Disorder Study (CLPS),[3] when remission was defined as 12 months at two or fewer criteria for PDs, over half of BPD and 85 per cent of major depressive disorder (MDD) patients were reported to remit over a 4-year period. Psychosocial functioning recovered far more slowly than acute symptoms.[1]

There is a considerable body of literature on psychotherapeutic interventions for personality disorders, but significant evidence

for effective treatment remains sparse. Much of the literature is dominated by expert opinion, which is not invariably the most helpful guide. In this chapter, we focus on psychological treatments where at least some evidence for treatment effectiveness exists. The evidence is strongest for borderline personality disorder (BPD). Treatment of some other personality disorders, for example schizoid, narcissistic, obsessive–compulsive, dependent, is evidenced mainly by clinical case reports in which theory is combined with clinical description and where, if outcome is measured at all, it is measured for the purpose of illustration and has little probative value.

Assessment of treatment

Any study that seeks to demonstrate the effectiveness of a treatment for PD must fulfil the following requirements:

(a) *Carefully define the target population*. This can be problematic, because the definition of personality disorder (PD) remains controversial, and there is little evidence that the categories of personality disorder have any predictive value in determining response to treatment. Comorbidity must also be considered. Lifetime comorbidity should not be an exclusion criterion for studies, but there is an argument for excluding individuals with current comorbidity. However, in clinical practice, such exclusion is almost impossible.

(b) *Adequately define the treatment and assessment of its specificity*. Personality disorder is a multifaceted condition that is susceptible to a variety of influences, and it justifies the use of complex interventions. These require complex evaluations, which increase the difficulty of interpreting results.

(c) *Establish that treatment is superior to no treatment* since personality disorders show gradual improvement over time.

(d) *Take account of Axis I disorders*. This can be done by excluding patients with Axis I disorders, but PDs are almost always associated with significant Axis I psychopathology. An alternative, which no trial has, to date, attempted, would be to assign patients to treatment group on the basis of matched Axis I disorders. In any case, it must be demonstrated that treatment impacts on personality rather than merely causing a change in mood or psychiatric symptoms.

(e) *Include adequate follow-up*, as some trials report reduced treatment effects during follow-up.

(f) *Address cost-effectiveness* relative to alternative interventions.

(g) *Study treatment effects in standard practice* (pragmatic trials) as well as under strict experimental conditions.

Research trials investigating the effectiveness of treatments for personality disorder have so far singularly failed to meet most of these requirements.

Adverse effects of psychotherapy for personality disorder

It is possible that some psychosocial treatments for personality disorder may have impeded the patient's capacity to recover following the natural course of the disorder and/or prevented them from taking advantage of changes in social circumstances. In Stone's[4] classic follow-up of patients treated nearly 40 years ago, a 66 per cent recovery rate was only achieved in 20 years, about four times longer than reported in the more recent studies. It seems unlikely that the nature of the disorder has changed or that treatments have become markedly more effective rather it is possible that treatments with these adverse effects are being offered less frequently now than in the past, perhaps because of changed patterns of health care, particularly in the United States.

Meta-analyses of psychotherapy and psychosocial treatments

It remains unclear whether the literature is robust enough to withstand the methodology of meta-analysis. The lack of good quality studies, especially randomized trials, the small number of patients in the trials, the heterogeneity of the personality disorders studied, and the variability of outcome measures across studies means that conclusions must remain tentative. One meta-analysis[5] included 15 studies that reported data on pre- to post-treatment effects and/or recovery at follow-up, including three randomized, controlled trials, three randomized comparisons of active treatments, and nine uncontrolled observational studies. They included psychodynamic/interpersonal, cognitive behavioural, mixed, and supportive therapies. All studies reported improvement in personality disorders with treatment. The mean pre–post effect sizes within treatments were 1.11 for self-report measures and 1.29 for observational measures. Among the three randomized, controlled treatment trials, active psychotherapy was more effective than no treatment according to self-report (ES = 0.75), though none of the controls employed an active therapy. Only four studies reported the percentage of cases no longer meeting the criteria for personality disorder. At follow-up (at a mean of 67 weeks), 52 per cent met this criterion. Treatment length was associated with the likelihood of recovery.

A subsequent meta-analysis[6] included psychodynamic therapy and cognitive behaviour therapy (CBT) in the treatment of personality disorders. There were 22 studies of psychodynamic therapy and CBT published between 1974 and 2001 that (1) used standardized methods of diagnosis, (2) applied reliable and valid instruments for the assessment of outcome, and (3) reported data that allowed calculation of within-group effect sizes or included assessment of recovery rates. Because only 11 of the 22 studies were RCTs, ESs were calculated on the basis of pre- to post-therapy change. Fourteen studies included psychodynamic therapy, and 11 studies included CBT. The psychodynamic studies had a mean follow-up of 1.5 years, compared to only 13 weeks for CBT. Psychodynamic therapy yielded an overall effect size of 1.46 ($k = 15$ contrasts), with effect sizes of 1.08 for self-report measures and 1.79 for observer-rated measures. For CBT ($k = 10$ contrasts), the corresponding values were 1.00, 1.20, and 0.87. For more specific measures of personality disorder pathology, a large overall effect size (1.56) for psychodynamic therapy suggested long-term rather than short-term change in personality disorders. For BPD, the ES for psychodynamic therapy was 1.31 ($N = 8$), and for CBT 0.95 ($N = 4$). Treatment length showed a positive but nonsignificant correlation with outcome in psychodynamic studies ($r = 0.41$); there were too few CBT trials for an equivalent analysis. In the 5 years since 2001, there have been nearly as many new randomized trials of treatments for personality disorder as were included in this meta-analysis, so a further systematic review including only randomized trials is due.

Psychotherapy for borderline personality disorder (BPD)

Several psychosocial treatments for BPD have emerged over the past decade, with one US guideline recommending psychotherapy as the primary treatment for this condition.[7] It is impossible to recommend one specific therapy, because information from research remains inadequate. It has become clear not only that several treatments may be of use, but also that any one treatment by itself, is helpful in only about a half of all cases.[8] Also, there is general consensus that some of the nonspecific elements of psychotherapy may be as important in determining the success of a treatment as the specific techniques. We reviewed treatments shown to be moderately effective and concluded that they share certain common features.[9] They tend to (a) be well-structured, (b) devote considerable effort to enhancing collaboration, (c) have a clear focus, (d) be theoretically highly coherent to both therapist and patient, (e) be relatively long-term, (f) encourage a powerful attachment relationship between therapist and patient, and (g) be well integrated with other services available to the patient. While some of these features may seem to pertain more to a successful research study rather than a successful therapy, the manner in which clinical treatment protocols are constructed and delivered is probably as important in the success of treatment as the intervention itself.

(a) Dynamic psychotherapy

Dynamic psychotherapy has long been recommended for BPD and has now been modified to target the characteristic features of the disorder. Almost all of the first studies examined inpatient treatment using prospective one group pre- to post-test designs.[10] These studies failed to rule out other plausible causes of change, such as passage of time or subsequent outpatient treatment. Stone's[11] report of up to 20 years follow-up of 550 inpatients, most of whom had received some sort of psychosocial intervention, indicated that 66 per cent of patients were functioning well. However, a naturalistic 5-year follow-up of individuals receiving inpatient treatment at the Cassel Hospital in London indicates the need for caution in ascribing benefits to inpatient treatment.[12] Whilst longer term follow-ups are to be applauded, they are hard to interpret, because other therapies are often given subsequent to the original treatment.

Several nonrandomized trials of dynamic psychotherapy have been undertaken by Stevenson, Meares, and colleagues. In an open trial for 48 patients receiving twice-weekly interpersonal-psychodynamic outpatient therapy for 12 months,[13] 30 per cent of the treatment group no longer met DSM-III-R criteria for BPD, while the waiting-list group changed little. Cost–benefit analysis found significant reduction in costs, largely attributable to reduced inpatient stays. A replication study[14] also found significant reduction in symptom severity with the same treatment.

A randomized study of 38 patients with BPD compared an 18-month programme of partial hospitalization using mentalization-based treatment (MBT) with standard psychiatric care.[15] Mentalization entails making sense of the actions of oneself and others on the basis of intentional mental states such as desires, feelings, and beliefs. Outcome measures included frequency of suicide attempts, acts of self-harm, number and duration of inpatient admissions, use of psychotropic medication, and self-report measures including depression, interpersonal function, and social adjustment. After 6 months, patients given MBT showed a statistically significant decrease on all measures in contrast to the control group which showed limited change or deterioration over the same period. This was sustained at the end of the 18 months of treatment and showed further improvement on follow-up after another 18 months. Long-term follow-up 5 years after the initial treatment suggested that the differences between the groups continued, but general social function remained impaired in both groups.[16] The treatment was cost-effective and has been manualized,[17] but as yet the active components remain unclear, and it has not been shown that positive outcomes are correlated with an improvement in mentalizing. (An outpatient version of MBT is currently being evaluated for borderline and antisocial PD in a further randomized controlled trial.) Although promising, this treatment needs further validation by research carried out independently of the originators. Favourable data has recently become available on the effectiveness of a similar programme established in the Netherlands.[18]

Transference-focused psychotherapy (TFP) has also shown good results. In a cohort study,[19] 23 female borderline patients were treated for 12 months. Compared with the year before treatment, the number of patients who made suicide attempts decreased significantly, as did the medical risk and severity of self-injurious behaviour. Also, compared with the previous year, there were significantly fewer hospitalizations and fewer days of psychiatric hospitalization. However, one in five patients dropped out of treatment. A subsequent trial compared TFP, DBT, and supportive therapy.[20] Ninety patients (all but nine of whom were women) were randomized. At the end of 1 year of treatment, the groups did not differ on global assessment of functioning, social adjustment, scores of depression and anxiety, and measures of self-harm. TFP patients improved significantly more than those receiving DBT or supportive therapy on irritability and verbal and direct assault. Patients who received TFP improved most in reflective function, an operationalization of the mentalization construct,[21] but it is not known whether improved reflective function relates to treatment gains at follow-up.

Schema-focused therapy (SFT) has been compared with TFP.[22] Treatment was given by therapists with approved training in the treatment methods administered treatment to 88 patients with borderline personality disorder. In an 'intent to treat' analysis, patients who received TFP showed significantly less improvement than those who received schema-focused CBT over 3 years, and TFP was more expensive. Both groups showed improvement, but changes in the combined measure of outcome in the schema-focused therapy group were greater and more prolonged than in the TFP group. There are several reasons for caution regarding these conclusions. (i) Differences in outcome between the groups can be accounted for almost entirely by the larger dropout early in treatment of patients treated with TFP and disappear when 'completers' are compared. It would be valuable to know why more patients dropped out from TFP at an early stage than from SFT. (ii) Follow-up is required to determine whether treatment gains and group differences are maintained. (iii) In the duration of the treatment period, around 40 per cent of patients could be expected to improve without the treatment.[23] This study also raises the question of how successfully a treatment (TFP) from the US can be transported to a European context.

(b) Group psychotherapy

Noncontrolled studies of day hospital stabilization followed by outpatient dynamic group therapy indicate its utility for BPD.[24] Marziali and Monroe-Blum used group therapy focused on relationship management and without the milieu and social components. A randomized controlled trial found equivalent results between group and individual therapy, suggesting that group therapy is more cost-effective.[25] Further studies are needed to confirm their findings, especially since dropout rates were high.

(c) Cognitive analytical therapy (CAT)

CAT has been manualized for treatment of BPD, and many are enthusiastic about its effectiveness. There are some indications that it may be effective. In a series of 27 patients with borderline personality disorder treated with 24 sessions of CAT, half no longer met diagnostic criteria for personality disorder at 6-month follow-up.[26] More definitive statements about efficacy await results of a randomized trial in progress. Ryle (personal communication) has indicated that patients treated with CAT showed significant improvement on a range of clinical measures but reported no difference between people receiving CAT and those undergoing other psychological treatments. Thus, the effects may be nonspecific. A second randomized trial is in progress comparing CAT with 'best available standard care' for adolescent patients with borderline personality disorder.[27]

(d) Cognitive therapy

Cognitive behavioural formulations of BPD are diverse. In a model derived from 'standard' CBT and modified for personality disorders, Beck and associates[28] define personality in terms of patterns of social, motivational, and cognitive-affective processes, thereby moving away from a primary emphasis on cognitions. However, personality is considered to be determined by 'idiosyncratic structures', known as schemas, whose cognitive content gives meaning to the person, and these schemas are the cornerstone of cognitive formulations of BPD. Young[29] has developed a treatment programme for BPD based on early maladaptive schemas (EMS). These are stable, enduring patterns of thinking and perception that begin early in life and are continually elaborated. EMS are unconditional beliefs linked together to form the core of an individual's self-image. Challenge to these beliefs threatens the core identity which is defended with alacrity, guile, and desperation since activation of the schemas may evoke aversive emotions. The EMS give rise to 'schema coping behaviour', which offers the best adaptation to living that the borderline has found. Schema-focused therapy (SFT) is only just being evaluated, but its adherence to the general requirements of an effective treatment enumerated above suggests that it should be reasonably successful. The recent report comparing SFT with TFP[22] is described above, but SFT has yet to be shown to be more effective than treatment as usual. It is possible that TFP simply induced more negative effects in patients than SFT.

A small ($N = 34$), randomized controlled trial assessed brief cognitive therapy, linked to a manual and incorporating elements of dialectical behaviour therapy, in the treatment of recurrent self-harm in people with cluster B personality difficulties and disorders.[30]

Manual-assisted cognitive treatment (MACT) is a complex six-session treatment based on the theory that deliberate self-harm and suicide attempts stem from distorted cognitive schemas and coping skills deficits.[31] It incorporates elements of bibliotherapy, CBT, and DBT, as well as psychoeducation in relation to self-harm and suicide attempts, and a functional analysis of specific episodes. The treatment also involves strategies to regulate emotion, such as distraction, crisis planning, and problem-solving strategies. Cognitive restructuring strategies and management of negative thinking are incorporated in the second phase of the programme, which includes components for the management of substance abuse and relapse prevention. Its brevity makes MACT a potentially valuable intervention from a public health standpoint. In a clinical trial, 34 self-harm repeaters with a parasuicide attempt in the preceding 12 months were randomly allocated to MACT or treatment as usual (TAU). The rate of suicide acts was lower with MACT, and self-rated depressive symptoms also improved. The mean treatment time was 2.7 sessions, and the average cost of care was 46 per cent less with MACT. A subsequent larger study ($N = 480$) did not find evidence that time to repeat parasuicide was extended following MACT, although there was a decrease in the cost of care.[32] A randomized controlled trial with 104 people with borderline personality disorder with a longer period of treatment (up to 30 sessions) found significant benefit with regard to suicidal behaviour but a nonsignificant increase in emergency presentations in the cognitive behaviour therapy group.[33] The CBT arm used less resources, although no significant cost-effectiveness advantage was demonstrated. Those who received CBT showed less evidence of dysfunctional beliefs, lower state anxiety scores, and less positive symptom distress.

Systems training for emotional predictability and problem solving (STEPPS)[34] is a group treatment offered as an adjunct to other treatments rather than as a sole intervention. It is a 20-week manualized programme of psychoeducation and behavioural management focusing on maladaptive schemas and including both professional and family carers. Subjects are encouraged to continue their usual care, including individual psychotherapy, medication, and case management, and are required to designate a mental health professional who would provide ongoing care and could be reached in a crisis. Data from 52 patients suggests some reduction in impulsive and suicidal behaviour and some improvement on measures of depression, but no follow-up data is yet available. An RCT is currently ongoing. In Holland, a retrospective assessment of the experience of 85 patients enrolled in STEPPS groups[35] reported significant improvement on all subscales of the SCL-90, particularly those assessing anxiety, depression, and interpersonal sensitivity. Patients and therapists reported moderate to high levels of acceptance for the treatment in both studies. As most of the effective programmes for BPD are long-term and expensive, a short-term efficacious treatment will be of great value. However, even if found to be effective in a clinical trial, its effectiveness will depend on the nature of the treatment as usual.

(e) Dialectical behaviour therapy (DBT)

DBT is a special adaptation of CBT, originally used for the treatment of a group of repeatedly parasuicidal female patients with borderline personality disorder. DBT is a manualized therapy[36] which includes techniques at the level of behaviour (functional analysis), cognitions (e.g. skills training), and support (empathy, teaching management of trauma), with a judicious mix of ideas derived from Zen Buddhism. The initial aim of DBT is to control self-harm, but its main aim is to promote change in the emotional

dysregulation that is judged to be at the core of the disorder. Thus, the goal of DBT goes far beyond self-harm reduction. The first trial[37] involved 44 females with borderline personality disorder who had made at least two suicide attempts in the previous 5 years, with one in the preceding 8 weeks. Half were assigned to DBT and half to the control condition. Assessments were made during and at the end of therapy and at a 1-year follow-up. Control patients were significantly more likely to attempt suicide spent significantly more time as inpatients over the year of treatment (mean 38.8 and 8.5 days, respectively), and were significantly more likely to drop out of the therapies to which they were assigned (attrition 50 per cent versus 16.7 per cent, respectively).

DBT was less superior at the 1-year follow-up. Follow-up was naturalistic, because the morbidity of the group was thought to preclude termination of therapy at the end of the experimental period. At 6-month follow-up, DBT patients continued to show less parasuicidal behaviour than controls, though there were no between-group differences in days in hospital. At 1 year there were no between-group differences in suicidal behaviour, but DBT patients had had fewer days in hospital. Treatment with DBT for 1 year compared with treatment as usual led to a reduction in the number and severity of suicide attempts and decreased the frequency and length of inpatient admission. However, there were no between-group differences on measures of depression, hopelessness or reasons for living.

The widespread adoption of DBT for BPD and other PDs is a tribute to both the effectiveness of the treatment and its acceptability to patients and families. Several studies have replicated the original Linehan study. Turner observed a decrease in parasuicidal acts and deliberate self-harm at 6 and 12 months in 12 patients treated with DBT compared to 12 treated with client-centred therapy.[38] Koons and colleagues[39] compared DBT with outpatient treatment as usual in 28 participants and found decreases in frequency of parasuicidal acts and self-injury at 3 and 6 months. Bohus and colleagues[40] explored an inpatient adaptation of DBT and found significant improvement in deliberate self-harm but a higher dropout rate. Other studies, however, have shown DBT to be no better than other active treatments such as the 12-step programme for opioid dependence[41] or treatment as usual in a UK context.[42]

Because of severity, symptomatology, and high rates of co-occurring disorders, BPD affects family members, and interventions addressing the needs of family members have been developed using a DBT frame.[43] In a study of a 12-week community-based BPD family education programme, Family Connections (FC),[44] family members showed significant improvements on burden, grief and empowerment, and a reduction in depression. However, the long-term effect of family interventions on the patient's well-being remains to be demonstrated.

It is uncertain which are the active elements of DBT-individual psychotherapy whether skills training, phone consultation, or the consultation team. Two studies examining the process of change in DBT[41,45] had inconclusive results. Nevertheless, adding a DBT skills training group to outpatient individual (non-DBT) psychotherapy does not seem to enhance treatment outcomes. Given that DBT is described as primarily a skills training approach, this finding might indicate that the central skills training component of DBT may not be of primary importance. However, individual DBT focuses on the strengthening of skills learned in the skills groups, and trying to combine a skills training group with an individual therapy that ignores or pays minimal attention to skill strengthening, may invalidate what the patient has learned about utilizing learned skills in an attempt to cope with everyday functioning. Disagreement remains regarding the policy of not admitting patients to hospital, except for a minimum period, since some studies show that the time and structure of an inpatient setting can be used to apparently good effect.[46]

(f) Therapeutic community treatments

A therapeutic community (TC) may be defined as an intensive form of treatment in which the environmental setting provides the core means through which behaviour can be challenged and modified, essentially through group interaction and interpersonal understanding. Therapeutic communities are described in Chapter 6.3.9. Several studies have been completed at the Cassel Hospital, a tertiary referral centre with a psychosocial residential treatment programme which includes daily unit meetings, community meetings, structured activities, co-responsibility planning for the running of the therapeutic community and other structured activities, and formal psychoanalytic psychotherapy (individual and small group). Patients in two different specialist psychosocial programmes (step-down and long-term inpatient) and in general psychiatric treatment as usual (TAU) were assessed at 12 and 24 months. By 24 months, patients in the step-down condition showed significant improvements on all measures.[47] Patients in the long-term residential condition showed significant improvements in symptom severity, social adaptation, and global functioning but no changes in self-harm, attempted suicide, and readmission rates. Over the same period, patients in the TAU group showed no improvement on any variable except self-harm and hospital readmissions. All three groups were followed for 72 months after intake. The specialist step-down condition led to significantly greater change than either the solely inpatient model or TAU in most key dimensions of functioning, even 5 years after the 12 months hospitalization.[48]

While the study appears to show that a step-down programme leads to significant improvement, this conclusion should be qualified by the study's design limitations, including the lack of random assignment to the three conditions and the naturalistic geographical allocation. Overall, these findings are consistent with the general view that extended hospital admission, even to a psychotherapeutically oriented unit, may engender pathological dependency and regression. Against this, a prospective study of 216 patients with severe personality disorder treated at the Menninger Clinic in two psychoanalytically orientated inpatient units[49] found positive change at discharge and 1 year follow-up, with no evidence of deleterious effects due to regression and dependency. As there are now many treatments for personality disorder, therapeutic communities must be evaluated by comparison studies using acceptable experimental designs if they are to be considered a serious treatment approach.

(g) Nidotherapy

Nidotherapy is the name of a new form of systematic management for chronic and persisting mental disorders that concentrates on making environmental changes to bring about a better fit between person and environment.[50] All types of environmental change—physical, social, and personal—are considered relevant.

Nidotherapy is a collaborative treatment 'involving the systematic assessment and modification of the environment to minimize the impact of any form of the disorder on individual or on society'.[50] The word is derived from the Latin nidus or nest, an environment adapted to the object that is occupying it. The therapeutic aim is not to change the person but rather to change the environment so it better fits the person. The therapist should repeatedly question whether interventions given in the course of management incorporate analytic, cognitive, or behavioural psychotherapy—if they do, the intervention is no longer nidotherapy. Nidotherapy involves seeing the world from the patient's standpoint, joint planning of agreed environmental targets, a concentration on finding ways to improve social function, sharing responsibility for the programme so that the patient is the final owner, and the use of independent arbitrators to resolve disagreements about the change that is needed. The programme has several phases: identification of the boundaries of the nidotherapy, full environmental analysis, implementation of agreed environmental change, monitoring progress, and resetting targets. It addresses the patient's physical, social, and personal environments.

This therapy is at its early stages, with little other than anecdotal evidence.[51] Nevertheless, it represents a healthy contrast to the behavioural, cognitive, and analytic approaches which dominate most psychosocial interventions. There is no doubt that a person's environment is an important determinant of their experience and behaviour. This is accepted regardless of theoretical perspective, yet none of these perspectives offers a model of how environments should be modified to be more in line with the capacities of the person who exists within them. Modifying situational constraints may indeed be an independent and effective adjunct to other treatment protocols for personality disorder.

(h) Summary

The findings from the above studies suggest that some of the problems encountered by people with BPD may be amenable to talking/behavioural treatments. Several studies show that the effort by the recipient of care in sticking with the care package is rewarded by a decline in anxiety, depression, self-harm, hospital admission, and use of prescribed medication. A review using Cochrane methodology[52] concluded that the studies are too few and too small to justify full confidence in their results so that current therapies remain experimental. The reviewers suggest that people with BPD, if offered entry into a randomized study of therapies, may wish to consider that outcomes of both experimental and control groups will probably be better than those of standard care outside the trial.

Psychotherapy for avoidant personality disorder

Avoidant PD is the most prevalent personality disorder in the general population.[53] However, avoidant personalities tend to suffer quietly and reject emotional involvement with others. Nevertheless, their distress and functional impairment are probably comparable to those of people with BPD.

Studies of the efficacy of treatment for avoidant PD have been hampered by the close connection of the condition to social phobia. About half of avoidant patients also have a social phobia. There is controversy over the distinction between generalized social phobia and avoidant personality disorder, many authors regarding the latter as an exaggerated type of social phobia.[54] Many studies of social phobia include individuals with avoidant personality disorder. Few trials explicitly focus on this disorder. Most APD patients are treated using behavioural methods including exposure, social skills training, and systematic desensitization. Early studies suggested treatment gains that were maintained for up to 1 year after finishing active treatment.[55]

For phobic anxiety, there is some agreement that the cognitive behavioural strategies involved in traditional treatment tend sometimes to fail with the more severely disturbed patients because of resistance evoked by maladaptive personality traits.[56] Most studies report a pattern of outcome similar to that for generalized social phobia, with similar rate of gains to those without APD but lower end-point functioning,[57] though some report a slower rate of change.[58] Massion *et al.*[59] report a 5-year prospective study examining the impact of personality disorders in 514 patients in the Harvard/Brown anxiety research programme. The presence of a personality disorder reduced the probability of remission in social phobia by 39 per cent. Much of this reduction was explained by the presence of avoidant personality disorder. Alden[60] found that while there were significant improvements on a range of measures, only 9 per cent of patients treated for APD rated themselves as completely improved.

Therapeutic strategies for severe cases of APD should be tailored to the core pathology of the disorder—(1) a despised and unworthy sense of self connected to a punitive and blaming internal other; (2) an avoidant interpersonal style; (3) affect phobia; and (4) pseudomentalization (intellectualization and pretend relations). In achieving somewhat better functioning for avoidant patients, short-term dynamic psychotherapy and cognitive therapy have been shown to have good results.[61] For the more severely disturbed avoidant patients, day treatment programmes are probably more effective[62] than outpatient therapy, where they tend to ossify.[63] An RCT of cognitive behaviour therapy versus psychodynamic therapy for AVPD[64] found greater benefit from cognitive behaviour therapy. Interpersonal therapy and supportive-expressive therapy have also been tried. Patients are often reported to have made substantial gains in all treatment programmes,[65] but many patients who complete treatment still demonstrate a low level of social function.

Psychotherapy for antisocial personality disorder (ASPD)

There are few evaluated therapies for ASPD. The disorder is a common comorbid diagnosis for individuals with substance abuse problems,[66] and a small number of trials have contrasted those with ASPD alone with those with ASPD and depression in the treatment of substance misuse. Outcomes with antisocial personality disorder alone are often poorer than when ASPD is associated with depression. Three studies[67–69] have reported that ASPD alone was an obstacle in the treatment of substance misuse.

Several studies (though few controlled trials) have examined the impact of treatment on individuals within the penal system, at least some of whom, it is likely meet criteria for antisocial personality disorder. Most studies focus on specific categories of offending behaviour (e.g. sexual offences), or on prisoners with problem behaviours, particularly violence.[70] Some studies have focused on individuals who manifest callous-unemotional traits.[71]

There are a few observational studies of individuals detained in high security settings using group CBT,[72] individual and group CBT in the context of a therapeutic milieu,[73] or psychodynamic therapy.[74] While improvements in functioning are noted, there are severe methodological problems such as the absence of control groups and highly reactive measurement of outcome.

Studies reviewed by Warren and colleagues[70] suggest the usefulness of cognitive behavioural or (less frequently) systemic group-based interventions of various kinds for mentally disordered offenders. These include: seven studies of problem-solving skills training, five studies of anger/aggression management, and three studies aimed at social skills problems. Three studies of problem-solving training reported no statistically significant positive effects, while a further three reported some positive results with measures of outcome closely linked to the intervention. Anger management studies were predominantly negative with no statistically significant positive effects reported by the majority of studies and positive results reported only with quite reactive measures or within quasi-experimental studies. Social skills interventions also enjoyed mixed success with either no statistically significant positive effects being reported or positive effects observed only with reactive measures or in uncontrolled studies. While it is not possible to claim that psychotherapy is effective in addressing problems associated with ASPD, there is sufficient preliminary data to justify further inquiry and to counter the therapeutic nihilism that usually pervades this field.

Psychotherapy for other personality disorders

Individuals with **paranoid PD** are highly suspicious of other people, including doctors, and are often unable to acknowledge their own negative feelings towards others. This fact alone means that treatment is problematic, but given that other characteristics include concern that other people have hidden motives, expectation of being exploited by others, inability to collaborate, social isolation, emotional detachment, and overt hostility, treatment becomes nearly impossible. This constellation of characteristics means research on this group of patients is very difficult to conduct, since researchers are inevitably seen as having hidden motives. Whatever treatment method is employed, the first step is the development of a moderately trusting relationship. There is no research specifically investigating the outcome of psychotherapeutic treatment for paranoid personality disorder. However, a number of studies have included patients with paranoid PD in assessment of treatment. In general, the presence of paranoid PD diminishes the effectiveness of psychotherapeutic treatment for other co-occurring personality disorders.

Most of the interest in **schizotypal** personality disorder (STPD) has been in its relationship to schizophrenia, and hence there is little data on outcome of treatment using psychosocial methods. The Chestnut Lodge follow-up study yielded some data suggesting that patients who had schizophrenia plus STPD and had been treated within the psychotherapeutic milieu did slightly better than patients without STPD.[75] This unexpected finding has not been explained.

Winston *et al.*[76] treated 32 patients with a range of personality disorders, predominantly in DSM-III-R cluster C, using two forms of brief psychodynamic therapy (short-term dynamic psychotherapy or brief adaptational therapy) with 40 weekly sessions and waiting-list control. Because of ethical constraints, waiting-list subjects began treatment after 15 weeks. Although both therapies used psychodynamic techniques, brief adaptational therapy employed more cognitive strategies. Contrasted to waiting-list controls, treated patients showed moderate improvements in symptoms and target complaints at post-therapy. Though there were few between-treatment differences, there was more variance in outcome for the patients treated with short-term dynamic therapy, suggesting that for some patients the technique was unhelpful. Gains appeared to be maintained at 18-month follow-up. A larger trial of these therapies with mainly cluster C disorder[77] led to similar findings.

Conclusions

Effectiveness

It is difficult to reach definitive conclusions about the effectiveness of psychotherapy in the treatment of personality disorders, chiefly because RCTs of psychological treatments for PDs are relatively scarce. There is evidence that personality disorders impact negatively on treatments for Axis I presentations. However, the strength and specificity of this evidence varies both across Axis I diagnoses and also across personality disorders.

Historically, studies of psychotherapy for personality disorder have focused on cluster B PDs. Borderline personality disorder is the best studied of the PD diagnoses both from a biological and a psychotherapeutic perspective. The recent growth in the evidence base indicates the feasibility of coherent research programmes with a patient group usually seen as presenting substantial problems of engagement. Evidence from randomized trials suggests that structured treatments employing DBT or a psychodynamic approach are more effective than routine care. Data gathered so far support the use of both behavioural and psychodynamic interventions, while evidence for more short-term cognitive interventions is somewhat equivocal. Since most studies present data from open trials or case-series, only two approaches can be considered 'evidence-based'. The available randomized trials of DBT and psychodynamic treatment are methodologically sound but limited in the latter case by relatively small sample sizes; further, both sets of studies were conducted in the context of clear leadership from the developers. Because, for the most part, contrast was to routine care, it is difficult to ascertain whether outcomes are due to the structured nature of the programmes or their therapeutic orientation.

In relation to antisocial personality disorder, research into treatments has been centred on patients seeking help for substance abuse or individuals seen within the penal system for problems associated with ASPD. There is no clear evidence of treatment efficacy for any one form of intervention. Conclusions in relation to ASPD are limited not only by the small number of trials but also by the methodological limitations of these studies, particularly reactive measures of outcome and complex multifaceted interventions (such as therapeutic communities) where effective transportable components of the treatment package are impossible to disaggregate.

Avoidant personality disorder has received only a modest amount of research attention. Individuals presenting with APD would normally be offered social skills training or cognitive (exposure-based) techniques which are likely to be successful, but the generalization of improvements to other social contexts has not yet been well demonstrated.

Implications for clinical practice

With the exception of cluster B disorders, many individuals with PDs present to services for help with Axis I disorders. Clinicians should be aware that while treatments normally recommended for these problems can still work, they may do so to a lesser extent than would be expected if comorbid PD were not part of the clinical problem. However, they should not assume that this will definitely be the case. The situation is somewhat different in the case of the so-called dramatic PDs. Many of the features of the behaviour of individuals with borderline or antisocial PD create such distress for the people around them that clinicians tend to privilege these features over symptomatic states related to comorbid Axis I disorders when setting goals for treatment.

There is little doubt that psychological therapies have a place in the management of personality disorders, but the methodological shortcomings of the studies so far conducted limit our ability to make practical recommendations. Given the nature of the problem (supposedly enduring problems of character) the fact that few trials conduct follow-up over a longer period or include a contrast or control group over follow-up is a grave limitation of the database. Only two studies (Chiesa's 72 months follow-up and the MBT 5-year follow-up) have demonstrated that post-treatment differences are maintained for extended periods. Clinical experience, somewhat contradicted by recent follow-along data, suggests that many individuals may show periods of complete remission followed by episodes of great severity, and therefore it is crucial to investigate the impact of treatment on this often chronic and cyclic course. DBT has good evidence for its immediate impact on behavioural problems of impulsivity and suicidality, but the evidence for long-term benefit without further treatment is less clear. Psychodynamic approaches aim to change the way the person thinks, and there is some evidence that this approach has some impact on mood states and interpersonal functioning. However, this emphasis on cognitive change requires longer term therapy, with some indication of slower rates of change but perhaps also more sustained gains over follow-up. While shorter interventions are likely to have lower immediate costs, longer ones would be justified if they were more cost-effective in terms of service utilization and other cost-offsets. However, evidence for these kinds of effects is scarce.

Debate about particular forms of therapy may be less relevant than attention to nonspecific factors, especially in the treatment of individuals with cluster B PDs. We have seen that successful approaches usually emphasize the importance of structure, a coherent theoretical base, a higher intensity of treatment than for many Axis I disorders, and especially, attention to the unique set of problems presented by the individual, with a clear psychological formulation guiding treatment. Decisions about which treatment to opt for might be best made with regard to pragmatic considerations rather than argument about which particular course is 'best'. It may even be better to offer patients intensive treatment initially and to follow with short bursts of treatment over a long period of time. The competence and training of senior clinicians who can offer supervision will be especially important, as will the skill mix of staff and the resources (including staffing level) available to the service. These 'nonspecific' issues may be especially pertinent when considering the performance of evidence-based treatments in routine practice. Since systemic factors may be as relevant to success as the type of treatment offered, pragmatic trials would be useful to give an indication of the conditions required to implement evidence-based therapies in routine services.

Further information

Fonagy, P. (ed.) (2007) Special issue on treatment of personality disorder. *Journal of Mental Health*, **16**(1).

Gunderson, J.G. (2001) *Borderline personality disorder: a clinical guide*. American Psychiatric Publishing, Washington, DC.

Lenzenweger, M.F. and Clarkin, J.F. (2004). *Major theories of personality disorder* (2nd edn). Guilford, New York.

Roth, A. and Fonagy, P. (2005). Personality disorders. In *What works for whom? A critical review of psychotherapy research* (2nd edn) (eds. A. Roth and P. Fonagy), pp. 297–319. Guilford, New York.

References

1. Skodol, A.E., Gunderson, J., Shea, M.T., *et al.* (2006). The Collaborative Longitudinal Personality Disorders Study (CLPS): overview and implications. *Journal of Personality Disorders*, **19**(5), 487–504.
2. Zanarini, M.C., Frankenburg, F.R., Hennen, J., *et al.* (2006). The McLean Study of Adult Development (MSAD): overview and implications of the first six years of prospective follow-up. *Journal of Personality Disorders*, **19**(5), 505–23.
3. Grilo, C.M., Sanislow, C.A., Gunderson, J.G., *et al.* (2004). Two-year stability and change of schizotypal, borderline, avoidant, and obsessive-compulsive personality disorders. *Journal of Consulting and Clinical Psychology*, **72**(5), 767–75.
4. Stone, M.H. (1990). *The fate of borderline patients: successful outcome and psychiatric practice*. Guilford Press, New York.
5. Perry, J.C., Banon, E., and Ianni, F. (1999). Effectiveness of psychotherapy for personality disorders. *The American Journal of Psychiatry*, **156**(9), 1312–21.
6. Leichsenring, F. and Leibing, E. (2003). The effectiveness of psychodynamic therapy and cognitive behavior therapy in the treatment of personality disorders: a meta-analysis. *The American Journal of Psychiatry*, **160**(7), 1223–32.
7. Oldham, J.M., Phillips, K.A., and Gabbard, G.O. (2001). Practice guideline for the treatment of patients with borderline personality disorder. *The American Journal of Psychiatry*, **158**(Suppl. 10), 1–52.
8. Roth, A. and Fonagy, P. (2005). *What works for whom? A critical review of psychotherapy research* (2nd edn). Guilford Press, New York.
9. Bateman, A.W. and Fonagy, P. (2000). Effectiveness of psychotherapeutic treatment of personality disorder. *The British Journal of Psychiatry*, **177**, 138–43.
10. McGlashan, T. (1986). The chestnut lodge follow-up study III: long-term outcome of borderline personalities. *Archives of General Psychiatry*, **43**, 20–30.
11. Stone, M. (1993). Long-term outcome in personality disorders. *The British Journal of Psychiatry*, **162**, 299–313.
12. Rosser, R., Birch, S., Bond, H., *et al.* (1987). Five year follow-up of patients treated with in-patient psychotherapy at the Cassel Hospital for nervous diseases. *Journal of the Royal Society of Medicine*, **80**, 549–55.
13. Stevenson, J., Meares, R., and D'Angelo, R. (2005). Five-year outcome of outpatient psychotherapy with borderline patients. *Psychological Medicine*, **35**(1), 79–87.
14. Korner, A., Gerull, F., Meares, R., *et al.* (2006). Borderline personality disorder treated with the conversational model: a replication study. *Comprehensive Psychiatry*, **47**, 406–11.
15. Bateman, A.W. and Fonagy, P. (2001). Treatment of borderline personality disorder with psychoanalytically oriented partial hospitalization: an 18-month follow-up. *The American Journal of Psychiatry*, **158**(1), 36–42.

16. Bateman, A.W. Fonagy P. 8-year follow-up of patients treated for borderline personality disorder—mentalization based treatment versus treatment as usual. *The American Journal of Psychiatry*. submitted.
17. Bateman, A.W. and Fonagy, P. (2004). *Psychotherapy for borderline personality disorder: mentalization based treatment*. Oxford University Press, Oxford.
18. Bales, D., Andrea, H., Smits, M., *et al.* (2007). Mentalization based treatment in the Netherlands: first results. *Biannual meeting of the international society personality disorder*. Den haag, Netherlands.
19. Clarkin, J.F., Foelsch, P.A., Levy, K.N., *et al.* (2001). The development of a psychodynamic treatment for patients with borderline personality disorder: a preliminary study of behavioral change. *Journal of Personality Disorders*, **15**, 487–95.
20. Clarkin, J., Levy, K.N., Lenzenweger, M.F., *et al.* (2007). Evaluating three treatments for borderline personality disorder: a multiwave study. *The American Journal of Psychiatry*, **164**, 922–8.
21. Levy, K.N., Meehan, K.B., Kelly, K.M., *et al.* (2006). Change in attachment patterns and reflective function in a randomized control trial of transference-focused psychotherapy for borderline personality disorder. *Journal of Consulting and Clinical Psychology*, **74**(6), 1027–40.
22. Giesen-Bloo, J., van Dyck, R., Spinhoven, P., *et al.* (2006). Outpatient psychotherapy for borderline personality disorder: randomized trial of schema-focused therapy vs transference-focused psychotherapy. *Archives of General Psychiatry*, **63**(6), 649–58.
23. Zanarini, M.C., Frankenburg, F.R., Hennen, J., *et al.* (2003). The longitudinal course of borderline psychopathology: 6-year prospective follow-up of the phenomenology of borderline personality disorder. *The American Journal of Psychiatry*, **160**(2), 274–83.
24. Wilberg, T., Friis, S., Karterud, S., *et al.* (1998). Outpatient group psychotherapy: a valuable continuation treatment for patients with borderline personality disorder treated in a day hospital? A 3-year follow-up study. *Nordic Journal of Psychiatry*, **52**, 213–22.
25. Marziali, E. and Monroe-Blum, H. (1995). An interpersonal approach to group psychotherapy with borderline personality disorder. *Journal of Personality Disorders*, **9**, 179–89.
26. Ryle, A. and Golynkina, K. (2000). Effectiveness of time-limited cognitive analytic therapy of borderline personality disorder: factors associated with outcome. *British Journal of Medical Psychology*, **73**(Pt 2), 197–210.
27. Ryle, A. (2004). The contribution of cognitive analytic therapy to the treatment of borderline personality disorder. *Journal of Personality Disorders*, **18**(1), 3–35.
28. Alford, B. and Beck, A. (1997). *The integrative power of cognitive therapy*. Guilford, New York.
29. Young, J.E. (1990). *Cognitive therapy for personality disorders: a schema-focused approach*. Professional Resource Exchange, Sarasota, Florida.
30. Evans, K., Tyrer, P., Catalan, J., *et al.* (1999). Manual-assisted cognitive-behaviour therapy (MACT): a randomized controlled trial of a brief intervention with bibliotherapy in the treatment of recurrent deliberate self-harm. *Psychological Medicine*, **29**(1), 19–25.
31. Schmidt, U. and Davidson, K. (2004). *Life after self-harm*. Routledge, London.
32. Byford, S., Knapp, M., Greenshields, J., *et al.* (2003). Cost-effectiveness of brief cognitive behaviour therapy versus treatment as usual in recurrent deliberate self-harm: a decision-making approach. *Psychological Medicine*, **33**(6), 977–86.
33. Davidson, K., Tyrer, P., Gumley, A., *et al.* (2006). A randomized controlled trial of cognitive behavior therapy for borderline personality disorder: rationale for trial, method, and description of sample. *Journal of Personality Disorders*, **20**(5), 431–49.
34. Blum, N., Pfohl, B., John, D.S., *et al.* (2002). STEPPS: a cognitive-behavioral systems-based group treatment for outpatients with borderline personality disorder—a preliminary report. *Comprehensive Psychiatry*, **43**(4), 301–10.
35. Freije, H., Dietz, B., and Appelo, M. (2002). Behandling van de borderline persoonlijk heidsstoornis met de VERS: de Vaardigheidstraining emotionele regulatiestoornis. *Directive Therapies*, **4**, 367–78.
36. Linehan, M.M. (1993). The skills training manual for treating borderline personality disorder. Guilford Press, New York.
37. Linehan, M.M., Heard, H.L., and Armstrong, H.E. (1993). Naturalistic follow-up of a behavioral treatment for chronically parasuicidal borderline patients. *Archives of General Psychiatry*, **50**, 971–4.
38. Turner, R.M. (2000). Naturalistic evaluation of DBT oriented treatment for BPD. *Cognitive Behaviour Practice*, **7**, 413–9.
39. Koons, C.R., Robins, C.J., Tweed, J.L., *et al.* (2001). Efficacy of DBT in women veterans with BPD. *Behavior Therapy*, **32**, 371–90.
40. Bohus, M., Haaf, B., Simms, T., *et al.* (2004). Effectiveness of inpatient dialectical behavioral therapy for borderline personality disorder: a controlled trial. *Behaviour Research and Therapy*, **42**(5), 487–99.
41. Linehan, M.M., Dimeff, L.A., Reynolds, S.K., *et al.* (2002). Dialectical behavior therapy versus comprehensive validation therapy plus 12-step for the treatment of opioid dependent women meeting criteria for borderline personality disorder. *Drug and Alcohol Dependence*, **67**(1), 13–26.
42. Feigenbaum, J.D., Fonagy, P., Pilling, S., *et al.* A pilot randomized control trial of the effectiveness of Dialectical Behavioural Therapy (DBT) for cluster B personality disorder. Submitted.
43. Hoffman, P.D. and Fruzzetti, A.E. (2007). Advances in interventions for families with a relative with a personality disorder diagnosis. *Current Psychiatry Reports*, **9**(1), 68–73.
44. Hoffman, P.D., Fruzzetti, A.E., and Buteau, E. (2007). Understanding and engaging families: an education, skills and support program for relatives impacted by borderline personality disorder. *Journal of Mental Health*, **16**(1), 69–82.
45. Shearin, E. and Linehan, M.M. (1992). Patient-therapist ratings and relationship to progress in dialectical behaviour therapy for borderline personality disorder. *Behavior Therapy*, **23**, 730–41.
46. Bohus, M., Haaf, B., Simms, T., *et al.* (2002). Effectiveness of inpatient dialectical behavioural therapy for borderline personality disorder: a controlled trial. *5th ISSPD European Congress on Personality Disorders—Abstracts*. Munich, Germany: 18.
47. Chiesa, M., Fonagy, P., Holmes, J., *et al.* (2004). Residential versus community treatment of personality disorders: a comparative study of three treatment programs. *The American Journal of Psychiatry*, **161**(8), 1463–70.
48. Chiesa, M., Fonagy, P., and Holmes, J. (2006). Six-year follow-up of three treatment programs to personality disorder. *Journal of Personality Disorders*, **20**(5), 493–509.
49. Gabbard, G.O., Coyne, L., Allen, J.G., *et al.* (2000). Evaluation of intensive inpatient treatment of patients with severe personality disorders. *Psychiatric Services*, **51**(7), 893–8.
50. Tyrer, P., Sensky, T., and Mitchard, S. (2003). Principles of nidotherapy in the treatment of persistent mental and personality disorders. *Psychotherapy and Psychosomatics*, **72**(6), 350–6.
51. Tyrer, P. and Kramo, K. (2007). Nidotherapy in practice. *Journal of Mental Health*, **16**(1), 117–29.
52. Binks, C.A., Fenton, M., McCarthy, L., *et al.* (2006). Pharmacological interventions for people with borderline personality disorder. *Cochrane Database of Systematic Review*, 2006(1), CD005653.
53. Samuels, J., Eaton, W.W., Bienvenu, O.J. III, *et al.* (2002). Prevalence and correlates of personality disorders in a community sample. *The British Journal of Psychiatry*, **180**, 536–42.
54. Rettew, D.C. (2000). Avoidant personality disorder, generalized social phobia, and shyness: putting the personality back into personality disorders. *Harvard Review of Psychiatry*, **8**(6), 283–97.
55. Renneberg, B., Goldstein, A.M., Phillips, D., *et al.* (1990). Intensive behavioral group treatment of avoidant personality disorder. *Behavior Therapy*, **21**, 363–77.

56. Alden, L.E., Taylor, C.T., Laposa, J.M., *et al.* (2006). Impact of social developmental experiences on cognitive-behavioral therapy for generalized social phobia. *Journal of Cognitive Psychotherapy: An International Quarterly*, **20**, 7–16.
57. Scholing, A. and Emmelkamp, P.M. (1999). Prediction of treatment outcome in social phobia: a cross-validation. *Behaviour Research and Therapy*, **37**(7), 659–70.
58. Oosterbaan, D.B., van Balkom, A.J., Spinhoven, P., *et al.* (2002). The influence on treatment gain of comorbid avoidant personality disorder in patients with social phobia. *The Journal of Nervous and Mental Disease*, **190**(1), 41–3.
59. Massion, A.O., Dyck, I.R., Shea, M.T., *et al.* (2002). Personality disorders and time to remission in generalized anxiety disorder, social phobia, and panic disorder. *Archives of General Psychiatry*, **59**(5), 434–40.
60. Alden, L.E. (1989). Short term structured treatment for avoidant personality disorder. *Journal of Consulting and Clinical Psychology*, **56**, 756–64.
61. Svartberg, M., Stiles, T.C., and Seltzer, M.H. (2004). Randomized, controlled trial of the effectiveness of short-term dynamic psychotherapy and cognitive therapy for cluster C personality disorders. *The American Journal of Psychiatry*, **161**, 810–7.
62. Karterud, S., Pedersen, G., Bjordal, E., *et al.* (2003). Day treatment of patients with personality disorders: experiences from a Norwegian treatment research network. *Journal of Personality Disorders*, **17**, 243–62.
63. Wilberg, T., Karterud, S., Pedersen, G., *et al.* (2003). Outpatient group psychotherapy following day treatment of patients with personality disorders. *Journal of Personality Disorders*, **17**, 510–21.
64. Emmelkamp, P., Benner, A., Kuipers, A., *et al.* (2006). Comparison of brief dynamic and cognitive-behavioural therapies in avoidant personality disorder. *The British Journal of Psychiatry*, **189**, 60–4.
65. Barber, J.P., Morse, J.Q., Krakauer, I.D., *et al.* (1997). Change in obsessive-compulsive and avoidant personality disorders following time-limited supportive-expressive therapy. *Psychotherapy*, **34**(2), 133–43.
66. Brooner, R.K., King, V.L., Kidorf, M., *et al.* (1997). Psychiatric and substance use comorbidity among treatment seeking opiate users. *Archives of General Psychiatry*, **54**, 71–80.
67. Woody, G.E., McLellan, T., Luborsky, L., *et al.* (1985). Sociopathy and psychotherapy outcome. *Archives of General Psychiatry*, **179**, 188–93.
68. Alterman, A.I., Rutherford, M.J., Cacciola, J.S., *et al.* (1996). Response to methadone maintenance and counseling in antisocial patients with and without major depression. *The Journal of Nervous and Mental Disease*, **184**(11), 695–702.
69. King, V.L., Kidorf, M.S., Stoller, K.B., *et al.* (2001). Influence of antisocial personality subtypes on drug abuse treatment response. *The Journal of Nervous and Mental Disease*, **189**(9), 593–601.
70. Warren, F., Preedy-Fayers, K., McGauley, G., *et al.* (2003). Review of treatments for severe personality disorder home office online report 30/03 (http://www.homeoffice.gov.uk/rds/pdfs2/rdsolr3003.pdf).
71. Hare, R.D., Hart, S.D., and Harpur, T.J. (1991). Psychopathy and the DSM-IV criteria for antisocial personality disorder. *Journal of Abnormal Psychology*, **100**, 391–8.
72. Quayle, M. and Moore, E. (1998). Evaluating the impact of structured groupwork with men in a high security hospital. *Criminal Behaviour and Mental Health*, **8**, 77–92.
73. Hughes, G., Hogue, T., Hollin, C., *et al.* (1997). First-stage evaluation of a treatment programme for personality disordered offenders. *Journal of Forensic Psychiatry*, **8**, 515–27.
74. Reiss, D., Grubin, D., and Meux, C. (1996). Young psychopaths in special hospital: treatment and outcome. *The British Journal of Psychiatry*, **168**, 99–104.
75. McGlashan, T.H. (1986). Schizotypal personality disorder. Chestnut Lodge follow-up study: VI. Long-term follow-up perspectives. *Archives of General Psychiatry*, **43**(4), 329–34.
76. Winston, A., Pollack, J., McCullough, L., *et al.* (1991). Brief psychotherapy of personality disorders. *Journal of Nervous and Mental Disease*, **179**, 188–93.
77. Winston, A., Laikin, M., Pollack, J., *et al.* (1994). Short-term psychotherapy of personality disorders. *The American Journal of Psychiatry*, **151**, 190–4.

4.12.7 Management of personality disorder

Giles Newton-Howes and Kate Davidson

Introduction

The treatment of personality disorders is a complex but rapidly evolving subject. It is to some extent described elsewhere in Chapters 3.3, 4.12.6, 4.13.1, 5.2.9, 6.3.5, 6.3.9 8.5.6, 9.2.4, 11.3.2, 11.16, and 11.17, and so this chapter excludes a full discussion of psychodynamic interventions, therapeutic communities, interventions for older people and the management of both adolescent and adult offenders.`

Methodological difficulties in evaluating the efficacy of treatment

The requirements for establishing whether a treatment is effective for personality disorders are much more exacting than those for mental state disorders. These can usefully be described under four headings:

- duration of treatment
- comorbidity
- adherence to treatment
- outcome measures.

(a) Duration of treatment

For most mental state disorders it is relatively easy to choose the period over which efficacy has to be demonstrated. In conditions that develop suddenly (e.g. panic), treatment trials could be for a very short time indeed. For others, particularly when maintenance treatment is being evaluated, at least 6 months may be necessary to establish continued efficacy. In the case of personality disorders, it has been thought that efficacy of treatment could not be judged adequately without at least a 2 to 3 year treatment phase. Personality disorder was regarded as being unlikely to change in the short-term. However, these ideas are changing in the light of evidence from longer-term follow-up studies of patients with personality disorder that show that these conditions do change over time in a consistent and predictable manner with substantial numbers of patients achieving full remission in the longer term.[1] If these longer-term follow-up studies are replicated, it would suggest that therapy should aim to accelerate the process of recovery. Determining what constitutes an adequate amount of therapy and over what length of time is an empirical question. More recent studies have offered treatment over 1 year with a 1 year follow-up

to examine maintenance of effect[2,3] but other studies have chosen lengthier treatment phases of up to 3 years[4] with some reporting continued therapy in the follow-up phase which does not allow the effect of maintenance of the original effect to be judged.[5] More recent studies, examining the effect of psychological treatments, have included a 1 year follow-up. The purpose of this follow-up period is to determine if treatment effects are maintained following the termination of treatment. Such a requirement is not a purist position; if a treatment for personality disorder appears to be effective over a shorter period, this may be due to change in a concurrent (comorbid) condition. In addition, if a treatment is to be judged efficacious in personality disorder its effects should be lasting beyond the active treatment component.

(b) Comorbidity

Comorbidity has been defined as 'the presence of any distinct clinical entity that has existed or that may occur during the clinical course of a patient who has the index disease under study'.[5] The key word here is 'distinct'. True comorbidity implies the presence of two completely separate disorders in the same person which are not causally related to each other in any way. Co-occurrence ranges from true comorbidity to the presence of the same disorder in two or more different forms.[6]

Comorbidity is the norm for most personality disorders both with other personality disorders or with mental state disorders. Borderline personality disorder is a major offender in this regard. Only about 1 in 20 of such disorders constitutes the pure condition,[7] and multiple comorbidity with four or more disorders is common.

In deciding on the efficacy of any treatment for personality disorder it is impossible to be certain whether observed improvement is in the personality disorder or in a comorbid condition. This problem is made worse because personality assessment is allegedly confounded, or 'contaminated', by the effect of a concurrent mental state disorder. Thus personality status apparently changes during the presence of a mental state disorder such as depression, only to return to the baseline normal subsequently.[8,9] Apparent improvement in a personality disorder following a treatment may be entirely due to improvement in a concurrent mental state disorder. However, this conclusion does not mean that *personality function*, as opposed to *personality disorder*, does not change. The underlying personality may remain stable, but if the setting and circumstances change, and this includes mental state changes, there can be marked changes in adjustment and so the manifestation of disorder will also change.[10]

In view of these problems, a treatment for a personality disorder should ideally be tested in those patients who have that personality disorder only. As these patients are uncommon and atypical, it is difficult to interpret the data from clinical trials.

(c) Adherence to treatment

People with personality disorders do not usually form good relationships with therapists. Although this is in keeping with their problems with relationships elsewhere, it can be a major problem in any form of therapy. The problem is particularly marked with psychotherapy, in which long-held views are challenged by the therapist. The consequence is that many patients dropout of care, and sometimes no amount of therapeutic skill can maintain them in care.[11] The failure to maintain prescribed treatment, in whatever form, is a constant handicap in accumulating an evidence base for interventions in personality disorder. Even within personality disorder there are differences between sub-groups. Most therapeutic trials have been inpatients with borderline personality disorder while those with schizoid, paranoid, histrionic, narcissistic, and antisocial personality disorders appear much less frequently. This is probably related to treatment attitudes. Borderline, anxious, and avoidant personality disorders contain a much higher proportion of treatment seeking (Type S) personality disorders as opposed to treatment resisting (Type R) ones, which are most prominent in those with antisocial, schizoid, and paranoid personalities.[12]

Any study of personality disorder is likely to have a large proportion of dropouts and this complicates the interpretation of the effects of treatment. The exception is when patients are treated in restricted settings such as prisons and other closed facilities,[13] but as these circumstances are abnormal it is difficult to generalize from them.

(d) Outcome measures

The choice of outcome measures is a problem in the assessment of all psychiatric disorders, but difficulty is particularly great in studies of personality disorders. These disorders affect both the individual and society, and a range of outcomes can be measured to cover these possibilities. Forensic psychiatrists and the general public usually consider that the outcome of mentally disordered offenders is best measured by the frequency of reoffending. This is an easily measured and reliable statistic, but it does not necessarily record symptomatic or personality change, and may be distorted by a range of other factors (e.g. patients who spend a long time in hospital or prison are not likely to reoffend). Changes in symptoms also have limited use since they may be a consequence of changes in mental state disorders quite independent of personality. Repeated measures of personality status also have disadvantages since, as noted earlier, they may be affected by changes with concurrent mental state disorder. Personality also changes with ageing irrespective of treatment.[14]

Because of these difficulties, global outcome measures are often used to determine the degree of improvement in personality disorders in long-term follow-up studies, although a battery of measurements is normally used in short-term treatment studies. Unfortunately, there is no standardized set of measures of global outcome. It is reasonable to take into account symptomatic change, social functioning, quality of life, incidents of societal conflict (e.g. police contacts), and reports from informants. Even these may not correctly reflect change in personality status. Thus a person whose personality disorder does not change in any basic way may find an environmental niche in which the personality disturbance does not manifest itself as conflict. Such a person would show improvement on all the items listed above, but the improvement would be a consequence of environmental change, not of personality alteration.

(e) Minimum requirements for establishment of efficacy

The evidence base for effectiveness of treatment in personality disorders is also exacting:

1 The treatment should be effective when used in the pure form of the personality disorder (in an explanatory trial) and subsequently in other forms of the disorder in which comorbidity is more common (pragmatic trial).

2 Efficacy cannot be established satisfactorily unless the treatment is tested in a randomized controlled trial.

3 A suitable control treatment or management needs to be tested against the experimental treatment.

4 Efficacy should only be determined after a period of at least 1 year because of the long duration of personality disorder. More recent randomized controlled trials in borderline personality disorder examining psychological therapies have met these more demanding criteria.[15,16]

Dynamic psychotherapy

This is discussed in chapter 4.12.6.

Cognitive therapies

(a) Cognitive analytical therapy

Cognitive analytical therapy combines cognitive and analytical approaches and has been applied to the treatment of personality disorders, particularly the borderline group.[17] The clinical manifestations of this condition are postulated to be a set of partially dissociated 'self-states' which account for the clinical features of this disorder. Such patients typically describe rapid switching from one state of mind to another, experiencing intense uncontrollable emotions or alternatively feeling muddled, or emotionally cut-off. Such 'dissociative states' (different from the conditions of similar name formerly linked to hysteria) are said to be activated by severe external threats and to be maintained by repetitions of threat and reinforcement by memories or situations which are similar to the original source of threat.

Cognitive analytical therapy is concerned with the identification of these different self-states and helping patients to identify 'reciprocal role procedures', or patterns of relationships which are learned in early childhood and are relatively resistant to change.[18] The patient is taught to observe and try to change damaging patterns of thinking and behaviour which relate to these self-states and to become more self-aware. The therapist's role is to gather information about the patient's experience of relationships and the different states he or she experiences, including interpretation where necessary. Although the standard measure of evidence of effectiveness, the randomized controlled trial, has still not yet been reported for this treatment it has gathered an impressive group of adherents and has become widely used and now has a good theoretical and pragmatic base.[18]

(b) Cognitive behaviour therapy

In its original form, cognitive therapy for depression was used to help the patient to identify and modify dysfunctional thoughts and beliefs through the use of specific cognitive techniques such as Socratic questioning. The focus of therapy was here and now the aim was to return the patient to his or her usual functioning by relieving current symptoms. The cognitive model of personality disorder does emphasize cognitive, emotional, and behavioural factors but the origins of personality problems are regarded as originating in the temperament of the child, childhood development, and experiences. Early infant attachment patterns, the child's internal working model of relationships, self-identity, self-worth, and the emotional availability of the infant's caregivers are central to how the child develops and these shape the adult self-identity, interpersonal relationships in adulthood, and behavioural and emotional coping responses.[19]

One of the first tasks of cognitive therapy in personality disorders is to gain an historical account of the patient's childhood development and background from which the therapist can derive a cognitive formulation linking past difficulties and presenting problems. Through the formulation, and understanding of the patient's view of self and others, unique core beliefs are identified that are linked to affect and to overdeveloped behavioural patterns that prevent the individual from functioning in an adaptive manner, particularly in interpersonal contexts. Therapy focuses on beliefs that concern core concepts about the self and others that have developed from childhood onwards and associated behaviours that have developed as coping strategies. The content and meaning of the beliefs have had an impact on past and present relationships and are likely to impact the therapeutic relationship. These beliefs, formed through negative, possibly abusive and neglectful experiences with others, are likely to have resulted in low self-esteem, hypersensitivity to criticism, and poor relationships with peers, caregivers and others in adolescence. Once a clear understanding of the content of patient's core beliefs and associated overdeveloped or compensatory behavioural patterns has been established, patients are encouraged to test out their beliefs and assumptions about others by learning new, more adaptive strategies for relating to others and to themselves. In borderline personality disorder, typically patients hold beliefs such as 'I am a bad and inadequate person' and 'others will abandon or reject me'. Having formed these beliefs through experiences in childhood, borderline patients, for example, may have learnt to avoid close relationships, are highly sensitized to signs of disapproval in others and have developed a punitive, self-critical style of thinking and behaviour, including self-harm. The emphasis in cognitive therapy is in developing new ways of thinking about self and others and in testing out new ways of behaving that are less self-defeating and more likely to improve the patient's interpersonal skills.[19] In comparison with the treatment of Axis I disorders, cognitive therapy with personality-disordered individuals takes more sessions and spans a longer time because the underlying problems are more pervasive and ingrained. There are other important elements of cognitive and related therapies, of which schema therapy and dialectical behaviour therapy are the most prominent. One of these differences is on the emphasis and attention paid to the therapeutic relationship. In cognitive therapy for personality disorder, more emphasis is placed on establishing and maintaining a therapeutic alliance, as interpersonal difficulties which occur in the patient's life outside therapy are also likely to arise within therapy. This is based on the hypothesis that the patient's core beliefs are consistent across a wide range of settings and therefore are also likely to be manifest in therapeutic relationships. Patients are therefore likely to be highly sensitive to signs of criticism and approval in their therapists. The models of treatment for personality disorder proposed by Beck and Freeman,[20] Davidson,[19] Young,[21] and Linehan[22] have in common an attempt to integrate biological and psychosocial factors. All models of treatment recognize the importance of building a secure therapeutic relationship and transference and countertransference issues in therapy are increasingly recognized as important mediators of the therapeutic process. These therapies utilize cognitive techniques to repair breakdown in communication that can occur during therapy.

Cognitive analytical therapy[18] also gives special attention to this aspect of therapy.

One of the goals of cognitive therapy with personality-disordered patients is to take advantage of these interpersonal difficulties in treatment by identifying and modifying the beliefs underlying them and, by extension, other relationships. Although people with personality disorders can recognize difficulties, they experience the problems as egosyntonic[23] (i.e. accepted as normal because they are an intrinsic part of usual functioning). As a result, alternative and potentially more adaptive beliefs about the self and others need to be identified and evaluated to see if they are indeed more adaptive and embraced as a consequence. These alternative more adaptive beliefs require to be systematically reviewed and reinforced, and new behaviours and ways of relating to others need to be practiced repeatedly if changes are to be consolidated. To achieve these changes, the therapist usually has to adopt a more directive approach than in cognitive therapy for depression and other Axis I disorders, and throughout will be more concerned with identifying and overcoming cognitive, emotional, and behavioural avoidance which maintains core beliefs.

(c) Other related psychological therapies

(i) STEPPS

Systems Training for Emotional Predictability and Problem Solving (STEPPS) is affiliated to the other cognitive psychotherapies. It was developed by Nancee Blum in Iowa and has been extended across several states within the United States and to the Netherlands. It has some of the elements of standard cognitive behaviour therapy and dialectical behaviour therapy and is a manualized programme involving 20 2 h weekly group meetings; with specific goals (or lessons) identified for each session.[24] A randomized controlled trial has just been completed and this shows significant gains in some areas compared with treatment as usual (Black, 2007, APA meeting, San Diego, USA).

(ii) Schema-focused therapy

Schema-focused therapy[21] is now becoming increasingly used in the treatment of borderline and antisocial personality disorders. It is a compendium of cognitive behaviour therapy, object relations theory, and gestalt therapy, and also involving what Young calls 'limited reparenting'. It is given in a relatively intensive form—two to three sessions a week for 1–2 years—but has been shown to be both more effective and cost-effective than transference-focused psychotherapy in a trial of treatments for borderline personality disorder.[28]

(iii) Dialectical behaviour therapy

The era of evidence-based therapy in personality disorder began with a formal trial of dialectical behaviour therapy, a form of cognitive behaviour therapy linked to skills training and detached acceptance (or mindfulness), was compared with treatment as usual in a group of repeatedly self-harming female patients with borderline personality disorder.[28,69] The hypothesis that dialectical behaviour therapy was effective in reducing self-harm was supported. Now several other randomized trials have taken place that show that DBT is particularly effective in reducing self-harm[4,29,30] though in another study, DBT improved hopelessness, depression, anger, and suicidal ideation but showed no difference in suicide attempts.[31]

This treatment has also been used systematically in the treatment of borderline personality disorder and those with comorbid substance abuse.[30] According to Linehan,[22] borderline personality disorder is primarily a dysfunction of emotional regulation which is assumed to have resulted from biological irregularities combined with certain dysfunctional environments. Others in contact with the patient are postulated as reinforcing this dysfunction by discounting or, in Linehan's preferred term, 'invalidating' their emotional experiences. Borderline patients are emotionally vulnerable and have difficulty in regulating patterns of responses associated with emotional states. The maladaptive behaviours which form part of the borderline syndrome can be viewed as either the product of emotional dysregulation or as attempts by the individual at regulating intense emotional states by maladaptive problem-solving strategies. Dialectical behaviour therapy, as its name suggests, contains within it the notion of opposites; common themes that emerge in therapy with borderline patients, such as acceptance of things as they are (so that there is no need for suicidal action), and change (from former maladaptive types of response) may appear incompatible but are synthesized in the therapy.

The essentials of dialectical behaviour therapy[22] are manualized weekly individual psychotherapy, group psychoeducational behavioural skills training, and telephone consultation when considered necessary. Therefore the content comprises a variety of problem-solving techniques including teaching the patient skills to help regulate emotions and tolerate distress, methods for validating the patient's perceptions, and behavioural and psychological versions of meditation skills. Therapists are also trained in case management. 'Core mindfulness skills' are also part of the treatment and involve teaching the patient to observe, describe, and take part in events and responses to events without dissociating from what is happening. The treatment encourages patients to take a non-judgemental approach to events and interactions and to do what is effective in situations rather than what they may feel is the 'right' thing to do.

Summary of cognitive therapies

The collective view of the effectiveness of cognitive and dialectical behaviour therapies is that they are undoubtedly effective Leichsenring and Leibing[32] examined 25 studies (but very few randomized controlled trials) and found cognitive behaviour therapy to be effective (effect size 1.0) but with psychodynamic therapies the effect size was larger (1.46), but the authors emphasized these results were preliminary and need updating. More randomized controlled trials have been carried out into treatments for personality disorder in the last 4 years than in the previous 50 years and further reviews are needed. To date we only have good evidence of effectiveness for borderline personality disorder, but cognitive behaviour therapy has also been shown to be effective in a randomized trial of avoidant personality disorder.[33]

Nidotherapy

Nidotherapy is 'the collaborative systematic assessment and modification of the environment to minimize the impact of any form of mental disorder on the individual or on society'.[25] It was developed specifically for the care of patients with multiple personality and major psychiatric pathology and is particularly focused on Type R (treatment resisting personality disorders). It makes no attempt to treat the patient directly but instead attempts to change the environment so there is a better fit between individual and

setting. It is normally carried out individually by a nidotherapist working independently from other clinical services.[26] It has been tested in a randomized trial in which the main gain was in cost savings as patients spent less time in hospital[27] and has also been used for antisocial personality disorder for the reduction of aggression and improvement in social functioning (http://www.controlled-trials.com (ISRCTN96256106)).

Therapeutic community treatments

There has been some confusion regarding the term 'therapeutic community'. It can apply to any form of environment created for a specific treatment but was really introduced and defined in its modern context by Maxwell Jones[34] who created a structure that ran completely counter to that of the traditional (authoritarian) mental hospital. This can be defined as a socially cohesive structure depending on intensive group treatments carried out by its residents. This approach, the democratic therapeutic community, differs from other forms of more coercive communities linked to criminal justice in the United States, often used for substance abusers, which are called concept therapeutic communities.[35,36] Therapeutic communities have very strong advocates and the complexity of the intervention makes it difficult to create a suitable comparison group to act as a control for a randomized trial; to date no such trial has been carried out.

Drug treatment

Drugs are often used for treating personality disorders although it is important to note that none are licensed for the treatment of these conditions. As with other forms of treatment, borderline personality disorder constitutes the largest group in which drug treatment is being used and is therefore worth examining separately. The evidence base for drug treatment is growing and there are now sufficient randomized controlled trials to evaluate and for other studies to be ignored in this review.

Borderline personality disorder

Borderline personality disorder is one of the most heterogeneous of all groups within the personality classification and includes extensive comorbidity with other personality disorders as well as with mental state disorders, particularly mood and stress-related disorders. This hinders the interpretation of Table 4.12.7.1, noting the problems of evaluation referred to at the beginning of this

Table 4.12.7.1 Summary of randomized controlled trials of drugs in the treatment of personality disorder (borderline except where indicated)

Drug group	Individual drug	Size of trial—*n* (ref)	Main outcome measures	Results
Tricyclic antidepressants@	Amitriptyline	61[37]	Depression, aggression, global improvement	Amitriptyline no better than placebo
Selective serotonin reuptake inhibitors (SSRIs)	Fluoxetine	40[38]	Aggression depression,	Fluoxetine superior to placebo
	Fluvoxamine (cross-over study)	38[39]	Mood shifts, aggression, impulsivity	Positive effects of fluvoxamine on mood only
Antipsychotic drugs (both first and second generation)	Haloperidol	61[37] and 108[40]	Depression, hostility, impulsiveness	Haloperidol effective in reducing depression and hostility in first study but not in second
	Olanzapine (wide-dosage range)	40[41]	Aggression, depression, anxiety	Significant improvement in an unrecognized scale (Clinical Global Improvement-BPD) only
	Olanzapine (mean 5.3 mg)	28[42]	'Random effects regression modelling of panel data'	Alleged improvement on composite measure (unsatisfactory)
	Aripiprazole 15 mg/day*	52[43]	Hostility, anger, depression, self-harm	Significant improvement in all measures and reduction in self-harm
Monoamine oxidase inhibitors (MAOIs)	Phenelzine (60 mg/day)	108[40]	Depression, hostility, impulsiveness, anxiety	Phenelzine superior to haloperidol and placebo for anger and hostility
Mood stabilizers and anti-convulsants	Carbamazepine (7 mg/day)†	20[44]	Depression, hostility	
	Sodium valproate (850 mg/day)*	30[45]	Depression, aggression, hostility	
	Sodium valproate*	91[46]	Aggression	
	Lamotrigine*	27[47]	Self-rated anger	
	Topiramate*	56[48]	Anger	

*Patients recruited by advertisement.

†Patients recruited as inpatients.

@A study involving mainly cluster C personality disorders compared the tricyclic antidepressant, dothiepin, with cognitive behaviour therapy and self-help over a 2-year period and showed significantly better response to dothiepin[49] (but study not included as after 2 years there were many deviations from the original protocol.

chapter. The studies are complex, yet the numbers are generally small, the period of treatment variable (8 weeks to 6 months), the dosage of drugs is usually flexible, and the outcomes manifold (and so variable that it is almost impossible to combine the data systematically). Positive publication bias and recruitment of volunteers by advertisement diminish the relationship between this patients group and those in clinical practice.

(i) Antidepressants

Tricyclic antidepressants do not appear to be effective in borderline personality disorder and, interestingly, do not help depressive symptoms preferentially, suggesting that there are subtle differences between the despair and emptiness of the borderline personality disorder and the anhedonia of depressive illnesses. Selective serotonin reuptake inhibitors (SSRIs), have been used in many trials (including some not cited here in which comparisons of inadequately small numbers have been made with psychological treatments) and provide some evidence of a positive effect on aggression and impulsiveness in antisocial personality disorder.[38] Monoamine oxidase inhibitors show reduction of anger and hostility in one trial but conventional risk management uggests that this treatment should only be used in exceptional circumstances.

(ii) Antipsychotic drugs

Although antipsychotic drugs have been tested in the treatment of borderline personality disorder more frequently than any other drug treatment, the results are equivocal. Haloperidol may be effective[37] in the short-term, however continuation studies do not show sustained improvement.[50] Olanzapine has been tested in three trials and, apart from showing consistent weight gain in all of these, has not shown any real evidence of benefit for any of the core symptoms of borderline personality disorder. The only trial showing substantial benefit was carried out with aripiprazole in symptomatic volunteers[43]; as aripiprazole has a complex pharmacological profile with 5-HT and dopamine agonist and antagonist actions this could represent a novel intervention and is worthy of replication.

(iii) Mood stabilizers

Borderline personality disorder is characterized by rapid swings in mood and emotional lability is a core feature. It is therefore not surprising that mood stabilizers have been used in its treatment. Again the results are equivocal and this subject urgently needs the benefit of a large independent trial. The most impressive result has been shown with the anticonvulsant drug, topiramate, which reduces aggression and hostility[48,51] but this finding needs replication. Lithium may also be effective in aggression but a satisfactory level of evidence is lacking.

Critique of drug treatment for borderline personality disorder

The evidence for the value of drug treatment has been influenced greatly by a guideline issued by the American Psychiatric Association in 2001.[49] This was a bold attempt to give clear recommendations to clinicians desperate to find a way through the fog of uncertainty with the abyss of suicide yawning on one side and iatrogenic poly-poly-pharmacy (the patients with this disorder are multiple consumers of prescribed drugs) on the other. There were four recommendations:

1 psychotherapy—both psychoanalytic/psychodynamic therapy, and dialectical behaviour therapy 'have been shown to have efficacy'
2 pharmacotherapy for 'affective dysregulation symptoms' should'be treated initially with a selective serotonin reuptake inhibitor or related antidepressant such as venlafaxine'
3 treatment of 'impulsive-behavioural dyscontrol symptoms' also suggests 'SSRIs are the initial treatment of choice'
4 'low-dose neuroleptics are the treatment of choice for "cognitive-perceptual" symptoms'.

These were criticized heavily at the time for going far beyond the evidence[52–54] and for some treatments, notably cognitive behaviour therapy, neglecting available evidence,[53] and further data accrued over the ensuing years has reinforced these concerns.[15,28,55] The recommendations concerning drug treatments are particularly suspect. There is no evidence worthy of the name that justifies the sub-grouping of borderline personality disorder into 'affective dyscontrol', 'impulsive-behavioural dyscontrol', and 'cognitive-perceptual' categories and these appear to be entities created to justify the use of particular drugs rather than provide a valid subdivision of a complex disorder. The creation of these essentially pseudo-diagnoses allows general conclusions to be made that all the drug treatments are efficacious and that 'taken together, the results of these studies suggest that the choice of medication can be guided as much by tolerability and safety as by symptom presentation'.[56] On the evidence analysed dispassionately this conclusion applies equally well to placebo, whose tolerability and safety are unparalleled.

Flamboyant group (cluster B, not borderline)

The evidence for drug use in other cluster B disorders is very limited. There may be a limited role for the use of mood stabilizers in reducing anger in dissocial personality. Lithium was found to be effective[57] and this action has been found in other settings,[58] but not reproduced in dissocial personality disorder. Carbamazepine has also been found to reduce impulsivity[59] but this result requires replication.

Odd eccentric group (cluster A)

No placebo-controlled, explanatory trials in this group have been conducted and no pragmatic trials provide evidence for drugs that remain in use. A mistrust of treatment given by others limits research and is probably an intrinsic part of the condition.

Anxious fearful group (cluster C)

The diagnostic overlap between cluster C personality disorders and neurotic disorders makes drawing conclusion about drug treatment for this group difficult. No clear explanatory trails have been conducted and the evidence of the effectiveness of antidepressants is confounded by the presence of neurotic disorder.[60,61]

Management

The management of borderline personality disorder can now be directed by a combination of the evidence and clinical judgement. It is clear from the randomized trials of both pharmacological and psychological treatment that an organized plan of care leads to improvement. This includes a consistent approach, a constancy of

personnel and adequate access to inpatient care during crisis situations. The cognitively based therapies are effective and pragmatic, providing a higher degree of input to ensure effectiveness. The intermittent addition of low-dose haloperidol or aripiprazole may assist with worsening impulsivity. A trial of topiramate for anger or SSRIs for impulsivity may also be of value. These need to be trialed for a specific period, 4 to 8 weeks, with a clear plan to review or stop if effectiveness is not clear or deteriorates over time. The risks of prescribing need to be carefully weighed against the potential benefits. Polypharmacy increases risk with no evidence of benefit. It is ethically important to ensure all patients have the capacity to consent to treatment, particularly for unlicensed drug use. Because the treatment of personality disorder is among the most complex of complex interventions[62] it is likely that several treatments, given in combination in a systematic way, are best suited to many of those who have more severe personality disorders.

The management of other personality disorders remains largely guided by clinical experience. The general principles of service organization remain the same, however decisions regarding psychotherapy and pharmacotherapy need to be tailored to individuals with an expectation of building a body of patient-based evidence, as other evidence is weak. Regular review of a management plan minimizes the likelihood of harm from any one course of action or the neglect of the patient's presentation.

Some management options can now be considered to have a negative risk: benefit ratio. These include behaviour therapy alone, tricyclic antidepressant therapy, and monoamine oxidase therapy.

Organization of services for personality disorder

It is likely that the organization of care for those with personality disorder has a much greater part to play than any specific single treatment in the clinical success of management but it is also fair to add that this conclusion is not based on excellent evidence. What is, however, clear from the randomized trials of both pharmacological and psychological treatment is that an organized plan of care leads to great improvement in those with personality disorder irrespective of the exact nature of the intervention.

Services for the management of personality disorder can include the 'sole practitioner' model (a single therapist seeing the patient), the 'divided function' model (in which part of the patient's care is taken over by specially trained staff whilst others look after other parts), and the 'specialist team' model (in which a specific personality disorder service takes on all aspects of care). The general conclusion is that the 'divided function or specialist team model is probably best for reducing risk and improving outcomes'.[63] The general principles of management apply to all personality disorders although the consequences of ignoring them will be greater for the flamboyant cluster of personality disorders than for others.

(a) Consistency

One of the reasons why those with personality disorders create so many problems in therapy is that they are highly sensitive to perceived criticism and are therefore able to detect any inconsistency in their treatment. Sometimes this is a way of deflecting attention away from fundamental problems associated with the personality disorder, but they are nonetheless effective in creating a screen that prevents other issues from being addressed. Clearly the fewer the people involved in care, the less are the chances of creating inconsistency. Keeping the number of main workers to a small number, and maintaining good communication, is a sound goal.

(b) Constancy

It is helpful to avoid changes in staff as far as possible. This is of particular relevance in the treatment of borderline personality disorder in which changes in therapists often re-enact the problems of loss and despair that the patient experiences so commonly in relationships.

(c) Adequate inpatient support

Staff involved in hospital care often believe that people with personality disorders should be kept out of hospital. This belief is based on the observation that such people exploit the opportunities offered by admission and create circumstances whereby it is difficult to discharge them. Much of this is opinion and not founded on evidence. Patients with comorbid mental state and personality disorders actually have better outcomes if they have a hospital-oriented programme of care than treatment in the community, whereas the opposite is true in the absence of personality disorder.[64] However, this work was carried out before specific services for personality disorder were set-up in England, the first country in the world to create such services.[65] The initial pilot service has just been evaluated and the results are encouraging, but the problem remains that those who wish to seek treatment from these services comprise only a small proportion of the total who suffer.[72]

Problems with comorbidity

Perhaps the most important error in management is failure to recognize personality disorder when other psychiatric disorders are more prominent and appear to be the only presenting problem. This is becoming recognized in the development of treatment protocols. This problem is encountered widely in the mental health services among people presenting to emergency psychiatric clinics,[73] in services for the homeless mentally ill,[74] and among heavy users of psychiatric services[75] and those with multiple admissions.[68] In all these settings, personality disorder is often not recognized early enough. This is to some extent understandable as the assessment of personality disorder is difficult in, for example, a busy emergency clinic. Nevertheless, failure to achieve a predicted response is often due to an earlier failure to detect the presence of personality disorder.

Conclusions

Some years ago that most sceptical of academic psychiatrists, Michael Shepherd, in referring to the contents of a book entitled *Recent Advances in Psychiatry* commented that the content was more accurately defined as 'recent activities', as 'advances' was too generous a word. Whilst not going quite as far as this in regard to advances in the treatment of personality disorder it is fair to add that the promise of effective therapies across the spectrum of personality dysfunction remains a long way off and we must be very careful not to oversell the evidence. The complexity of personality disorders often requires complex intervention, however, until we are confident that single treatments are effective the arguments for evaluating them in combination have to be very strong on theoretical grounds to justify the cost.

We are still at the stage in which explanatory trials (ones demonstrating efficacy under optimal conditions) are at least as necessary as pragmatic ones (demonstrating benefit in conditions of ordinary practice). These need to be carried out with adequate numbers of patients (at least 50 in each treatment arm rather than an artificially derived sample size to justify a small number) and with good independent assessments carried out by research workers who are masked as much as possible from disclosure of treatment. These requirements are exacting but can be achieved.[67] We also need better pragmatic trials of patients seen in ordinary mental health practice whose treatment and characteristics are both representative and from whom it is possible to generalize findings with confidence. Currently there are very few studies that satisfy this requirement; one recent study combining an educational intervention with problem-solving is an exception.[68]

Despite the caution of these statements these are exciting times in the management of personality disorder. We are no longer listening to the once powerful lobby that claimed that patients with these conditions should not be treated by psychiatric services, or to the pessimists that still regard these conditions as untreatable. We are in the equivalent position as those in the early 1950s who suddenly became aware of the possibility that powerful treatments for severe mental illness were ready and waiting to be used. They do indeed appear to be within reach, but we must use them wisely.

Acknowledgement

The authors would like to thank Professor Peter Tyrer who co-authored the original chapter on which this is based.

Further information

Sampson, M., McCubbin, R., and Tyrer, P. (eds.) (2006). *Personality disorder and community mental health teams, a practitioners guide*. John Wiley & Sons, Ltd., West Sussex. ISBN: 978-0-470-01171-3.

National Institute for Health and Clinical Excellence (UK): http://guidance.nice.org.uk/page.aspx?o=357063&c=91523

References

1. Zanarini, M.C., Krankenburg, F.R., Hennen, J., *et al.* (2003). Longitudinal course of borderline psychopathology: 6 year prospective study. *The American Journal of Psychiatry*, **160**, 274–83.
2. Davidson, K. (2007). *Cognitive therapy for personality disorders: a guide for clinicians* (2nd edn). Routledge, Hove.
3. Verheul, R., van den Bosch, L.M.C., Koeter, W.J., *et al.* (2003). Dialectical behaviour therapy with women with borderline personality disorder. 12-month randomised clinical trial in the Netherlands. *The British Journal of Psychiatry*, **182**, 135–40.
4. Linehan, M.M., Comtois, K.A., Murray, A. M., *et al.* (2006). Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Archives of General Psychiatry*, **63**, 757–66.
5. Feinstein, A. (1970). The pre-therapeutic classification of comorbidity in chronic disease. *Journal of Chronic Diseases*, **23**, 455–62.
6. Tyrer, P. (1996). Comorbidity or consanguinity. *The British Journal of Psychiatry*, **168**, 669–71.
7. Fyer, M.R., Frances, A.J., Sullivan, T., *et al.* (1988). Co-morbidity of borderline personality disorder. *Archives of General Psychiatry*, **45**, 348–52.
8. Coppen, A.L. and Metcalfe, H. (1965). The effect of a depressive illness on MMPI scores. *The British Journal of Psychiatry*, **111**, 236–9.
9. Stuart, S., Simons, A.D., Thase, M.E., *et al.* (1992). Are personality assessments valid in acute major depression? *Journal of Affective Disorders*, **24**, 281–9.
10. Tyrer, P., Coombs, N., Ibrahimi, F., *et al.* (2007). Critical developments in the assessment of personality disorder. *The British Journal of Psychiatry*, **190**(Suppl. 49), s51–9.
11. Adler, G. (1979). The myth of the alliance with borderline patients. *The American Journal of Psychiatry*, **136**, 642–5.
12. Tyrer, P., Mitchard, S., Methuen, C., *et al.* (2003). Treatment-rejecting and treatment-seeking personality disorders: Type R and Type S. *Journal of Personality Disorders*, **17**, 265–70.
13. Soloff, P.H., George, A., Nathan, R.S., *et al.* (1986). Progress in pharmacotherapy of personality disorders: a double blind study of amitriptyline, haloperidol and placebo. *Archives of General Psychiatry*, **43**, 691–7.
14. Seivewright, H., Tyrer, P., and Johnson, T. (2002). Changes in personality status in neurotic disorder. *Lancet*, **359**, 2253–4.
15. Davidson, K., Norrie, J., Tyrer, P., *et al.* (2006). The effectiveness of cognitive behaviour therapy for borderline personality disorder: results from the BOSCOT trial. *Journal of Personality Disorders*, **20**, 450–65.
16. Linehan, M.M., Comtois, K.A., Murray, A.M., *et al.* (2006). Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Archives of General Psychiatry*, **63**, 757–66.
17. Ryle, A. (1997). The structure and development of borderline personality disorder: a proposed model. *The British Journal of Psychiatry*, **170**, 82–7.
18. Ryle, A. and Kerr, I.B. (2002). *Introducing cognitive analytic therapy: principles and practice*. John Wiley, Chichester.
19. Davidson, K. (2007). *Cognitive therapy for personality disorders: a guide for clinicians* (2nd edn). Routledge, London.
20. Beck, A.T. and Freeman, A. (1990). *Cognitive therapy of personality disorders*. Guilford Press, New York.
21. Young, J.E., Klosko, J.S., and Weishaar, M.E. (2003). *Schema therapy: a practitioner's guide*. The Guilford Press, New York.
22. Linehan, M.M. (1992). *Cognitive therapy for borderline personality disorder*. Guilford Press, New York.
23. Alexander, F. (1930). The neurotic character. *International Journal of Psycho-analysis*, **11**, 291–311.
24. Blum, N., Pfohl, B., St. John, D., *et al.* (2002). STEPPS: a cognitive behavioral systems-based group treatment for outpatients with borderline personality disorder-apreliminary report. *Comprehensive Psychiatry*, **43**, 301–10.
25. Tyrer, P., Sensky, T., and Mitchard, S. (2003). The principles of nidotherapy in the treatment of persistent mental and personality disorders. *Psychotherapy and Psychosomatics*, **72**, 350–6.
26. Tyrer, P. and Bajaj, P. (2005). Nidotherapy: making the environment do the therapeutic work. *Advances in Psychiatric Treatment*, **11**, 232–8.
27. Tyrer, P. and Tyrer, S. (2008). Other treatments for persistent disturbances of behaviour. In *Cambridge textbook of effective treatments in psychiatry* (ed. P. Tyrer and K.R. Silk). Cambridge University Press, Cambridge.
28. Giesen-Bloo, J., van Dyck, R., Spinhoven, P., *et al.* (2006). Outpatient psychotherapy for borderline personality disorder: randomized trial of schema-focused therapy vs transference-focused psychotherapy. *Archives of General Psychiatry*, **63**, 649–58.
29. Verheul, R., van den Bosch, L.M.C., Koeter, W.J., *et al.* (2003). Dialectical behaviour therapy with women with borderline personality disorder. 12-month randomised clinical trial in the Netherlands. *The British Journal of Psychiatry*, **182**, 135–40.
30. Linehan, M., Dimeff, L., Reynolds, S., *et al.* (2002). Dialectical behavior therapy versus comprehensive validation therapy plus 12-step for the treatment of opioid dependent women meeting criteria for borderline personality disorder. *Drug and Alcohol Dependence*, **67**, 13–26.

31. Koons, C.R., Robins, C.J., Tweed, J.L., *et al.* (2001). Efficacy of dialectical behavior therapy in women veterans with borderline personality disorder. *Behavior Therapy*, **32**, 371–90.
32. Leichsenring, F. and Leibing, E. (2003). The effectiveness of psychodynamic therapy and cognitive behavior therapy in the treatment of personality disorders: a meta-analysis. *The American Journal of Psychiatry*, **160**, 1223–32.
33. Emmelkamp, P.M.G., Benner, A., Kuipers, A., *et al.* (2006). Comparison of brief dynamic and cognitive-behavioural therapies in avoidant personality disorder. *The British Journal of Psychiatry*, **189**, 60–4.
34. Jones, M. (1953). The therapeutic community: a new treatment method in psychiatry. Basic Books, New York.
35. Campling, P. (2001). Therapeutic communities. *Advances in Psychiatric Treatment*, **7**, 365–72.
36. Rutter, D. and Tyrer, P. (2003). The value of therapeutic communities in the treatment of personality disorder: a suitable place for treatment? *Journal of Psychiatric Practice*, **9**, 291–302.
37. Soloff, P.H., George, A., Nathan, R.S., *et al.* (1986). Progress in pharmacotherapy of personality disorders: a double blind study of amitriptyline, haloperidol and placebo. *Archives of General Psychiatry*, **43**, 691–7.
38. Coccaro, E.F. and Kavoussi, R.J. (1997). Fluoxetine and impulsive-aggressive behaviour in personality disordered subjects. *Archives of General Psychiatry*, **54**, 1081–8.
39. Rinne, T., van den Brink, W., Wouters, L., *et al.* (2002). SSRI treatment of borderline personality disorder: a randomized, placebo-controlled clinical trial for female patients with borderline personality disorder. *The American Journal of Psychiatry*, **159**, 2048–54.
40. Soloff, P.H., George, A., Nathan, S., *et al.* (1993). Efficacy of phenelzine and haloperidol in borderline personality disorder. *Archives of General Psychiatry*, **50**, 377–85.
41. Bogenschutz, M.P. and George Nurnberg, H. (2004). Olanzapine versus placebo in the treatment of borderline personality disorder. *The Journal of Clinical Psychiatry*, **65**, 104–9.
42. Zanarini, M.C. and Frankenburg, F.R. (2001). Olanzapine treatment of female borderline personality disorder patients: a double-blind, placebo-controlled pilot study. *The Journal of Clinical Psychiatry*, **62**, 849–54.
43. Nickel, M.K., Muehlbacher, M., Nickel, C., *et al.* (2006). Aripiprazole in the treatment of patients with borderline personality disorder: a double-blind, placebo-controlled study. *The American Journal of Psychiatry*, **163**, 833–8.
44. De La Fuente, J.M. and Lotstra, F. (1994). A trial of carbamazepine in borderline personality-disorder. *European Neuropsychopharmacology*, **4**, 479–86.
45. Frankenburg, F.R. and Zanarini, M.C. (2002). Divalproex sodium treatment of women with borderline personality disorder and bipolar II disorder: a double-blind placebo-controlled pilot study. *The Journal of Clinical Psychiatry*, **63**, 442–6.
46. Hollander, E., Tracy, K.A., Swann, A.C., *et al.* (2003). Divalproex in the treatment of impulsive aggression: efficacy in cluster B personality disorders. *Neuropsychopharmacology*, **28**, 1186–97.
47. Tritt, K., Nickel, C., Lahmann, C., *et al.* (2005). Lamotrigine treatment of aggression in female borderline-patients: a randomized, double-blind, placebo-controlled study. *Journal of Psychopharmacology*, **19**, 287–91.
48. Nickel, M.K., Nickel, C., Kaplan, P., *et al.* (2005). Treatment of aggression with topiramate in male borderline patients: a double-blind, placebo-controlled study. *Biological Psychiatry*, **57**, 498–9.
49. American Psychiatric Association. (2001). Practice guideline for the treatment of patients with borderline personality disorder. *The American Journal of Psychiatry*, **158**(Suppl.), 1–52.
50. Cornelius, J., Soloff, P., Perel, J., *et al.* (1993). Continuation pharmacotherapy of borderline personality disorder with haloperidol and phenelzine. *The American Journal of Psychiatry*, **150**, 1843–8.
51. Loew, T.H., Nickel, M.K., Muehlbacher, M., *et al.* (2006). Topiramate treatment for women with borderline personality disorder—a double-blind, placebo-controlled study. *Journal of Clinical Psychopharmacology*, **26**, 61–6.
52. McGlashan, T.H. (2002). The borderline personality disorder practice guidelines: the good, the bad, and the realistic. *Journal of Personality Disorders*, **16**, 119–21.
53. Sanderson, C., Swenson, C., and Bohus, M. (2002). A critique of the American Psychiatric Practice Guideline for the treatment of patients with borderline personality disorder. *Journal of Personality Disorders*, **16**, 122–9.
54. Tyrer, P. (2002). Practice guideline for the treatment of personality disorders: a bridge too far. *Journal of Personality Disorders*, **16**, 113–8.
55. Weinberg, I., Gunderson, J.G., Hennen, J., *et al.* (2006). Manual assisted cognitive treatment for deliberate self-harm in borderline personality disorder. *Journal of Personality Disorders*, **20**, 467–82.
56. Zanarini, M.C. (2004). Update on pharmacotherapy of borderline personality disorder. *Current Psychiatry Reports*, **6**, 66–70.
57. Sheard, M.H., Marini, J.L., Bridges, C.I., *et al.* (1976). The effect of lithium on impulsive aggressive behavior in man. *The American Journal of Psychiatry*, **133**, 1409–13.
58. Tyrer, S.P., Walsh, A., Edwards, D.E., *et al.* (1984). Factors associated with a good response to lithium in aggressive mentally handicapped subjects. *Progress in Neuropsychopharmacology & Biological Psychiatry*, **8**, 751–5.
59. Cowdry, R.W. and Gardner, D.L. (1988). Pharmacotherapy of borderline personality disorder: alprazolam, carbamazepine, trifluoperazine and tranylcypromine. *Archives of General Psychiatry*, **45**, 111–9.
60. Deltito, J.A. and Stam, M. (1989). Psychopharmacological treatment of avoidant personality disorder. *Comprehensive Psychiatry*, **30**, 498–504.
61. Ansseau, M., Troisfontaines, B., Papart, P., *et al.* (1991). Compulsive personality as predictor of response to serotonergic antidepressants. *British Medical Journal*, **303**, 760–1.
62. Campbell, M., Fitzpatrick, R., Haines, A., *et al.* (2000). A framework for the design and evaluation of complex interventions to improve health. *British Medical Journal*, **321**, 694–6.
63. Bateman, A.W. and Tyrer, P. (2004). Services for personality disorder: organization for inclusion. *Advances in Psychiatric Treatment*, **10**, 425–33.
64. Tyrer, P., Manley, C., Van Horn, E., *et al.* (2000). Personality abnormality in severe mental illness and its influence on outcome of intensive and standard case management: a randomised controlled trial. *European Psychiatry*, **15**(Suppl. 1), 7–10.
65. National Institute of Mental Health (England). (2003). *Personality disorder: no longer a diagnosis of exclusion*. Department of Health, London.
66. Geller, J.L., Fisher, W.H., McDermeit, M., *et al.* (1998). The effects of public managed care on patterns of intensive use of inpatient psychiatric services. *Psychiatric Services*, **49**, 327–32.
67. Davidson, K.M., Tyrer, P., Gumley, A., *et al.* (2006). Rationale, description, and sample characteristics of a randomised controlled trial of cognitive therapy for borderline personality disorder: the BOSCOT study. *Journal of Personality Disorders*, **20**, 431–49.
68. Huband, N., Duggan, C., McMurran, M., *et al.* Social problem-solving plus psychoeducation for adults with personality disorder: a pragmatic randomised clinical trial. *The British Journal of Psychiatry*, **190**, 307–13.

69. Linehan, M.M., Armstrong, H.E., Suarez, A., *et al.* (1991). Cognitive–behavioral treatment of chronically parasuicidal borderline patients. *Archives of General Psychiatry*, **48**, 1060–4.
70. Zanarini, M.C. and Frankenburg, F.R. (2003). Omega-3 fatty acid treatment of women with borderline personality disorder: a double-blind, placebo-controlled pilot study. *The American Journal of Psychiatry*, **160**, 167–9.
71. Ansseau, M., Troisfontaines, B., Papart, P., *et al.* (1991). Compulsive personality as predictor of response to serotonergic antidepressants. *British Medical Journal*, **303**, 760–1.
72. Crawford, M., Price, K., Weaver, T., *et al.* (2007). Learning the lessons: an evaluation of pilot community services for adults with personality disorder. Department of Health, London.
73. Breslow, R.E., Klinger, B.I., and Erickson, B.J. (1996). Acute intoxication and substance abuse among patients presenting to a psychiatric emergency service. *General Hospital Psychiatry*, **18**, 183–91.
74. North, C.S., Thompson, S.J., Pollio, D.E., *et al.* (1997). *A diagnostic comparison of homeless and nonhomeless patients in an urban mental health clinic. Social Psychiatry and Psychiatric Epidemiology*, **32**, 236–40.
75. Kent, S., Fogarty, M., and Yellowlees, P. (1995). Heavy utilization of inpatient and outpatient services in a public mental health service. *Psychiatric Services*, **46**, 1254–7.

4.13

Habit and impulse control disorder

Contents

4.13.1 Impulse control disorders

Susan L. McElroy and Paul E. Keck, Jr.

This chapter first defines impulse control disorders, and then summarizes available research on the clinical features, epidemiology, psychiatric comorbidity, family studies, psychobiology, and treatment response of the most common of these conditions (except for pathological gambling, which is reviewed in Chapter 4.13.2).

Definitions of impulse control disorders

Historically, impulse control disorders have been broadly defined as harmful behaviours performed in response to irresistible impulses.[1] In DSM-IV, an impulse control disorder is defined as the failure to resist an impulse, drive, or temptation to commit an act that is harmful to the individual or to others.[2] DSM-IV also stipulates that for most impulse control disorders, the individual feels an increasing sense of tension or arousal before committing the act and then experiences pleasure, gratification, or relief at the time of committing the act. In the text describing these disorders, DSM-IV states that after the act, there may or may not be genuine regret, self-reproach, or guilt. In ICD-10,[3] these conditions are classified as habit and impulse disorders and defined as repeated acts that have no clear rational motivation, cannot be controlled, and that generally harm the patient's own interests and those of other people. ICD-10 further states that the behaviour is associated with impulses to action.

In DSM-IV, impulse control disorders are listed in a residual category, 'Impulse control disorders not elsewhere classified', which includes intermittent explosive disorder (IED), kleptomania, pyromania, pathological gambling, trichotill16omania, and impulse control disorders not otherwise specified (NOS). Examples of impulse control disorders NOS are compulsive buying disorder, repetitive self-mutilation, pathological skin picking, and onychophagia (severe nail-biting).[1] In ICD-10, habit and impulse disorders are also listed as a residual category. Similar to DSM-IV, it includes pathological gambling, pathological fire-setting (pyromania), pathological stealing (kleptomania), trichotillomania, other habit and impulse disorders (which includes IED), and habit and impulse disorder, unspecified.

It should be noted that with mounting research, the impulse control disorders are increasingly viewed as complex conditions sharing, in addition to irresistible impulses to perform harmful behaviours, features of trait impulsivity, trait compulsivity, and mood dysregulation, as well as obsessive compulsive mood, and addictive disorders.[1]

Intermittent explosive disorder

Definition and clinical features

Intermittent explosive disorder (IED) is defined in DSM-IV as several discrete episodes of failure to resist aggressive impulses that result in serious assaultive acts or destruction of property (criterion A). Also, the degree of aggression expressed during an episode is grossly out of proportion to any precipitating psychosocial stressors (criterion B) and the explosive episodes are not better accounted for by another mental disorder or due to the direct physiological effects of a substance or a general medical condition (criterion C). Varying definitions of IED based on the DSM-IV criteria have been proposed and used.[4,5] One important set of research criteria for IED, for example, allows for less severe aggressive episodes, such as recurrent verbal outbursts against others without physical aggression, but requires that the aggressive episodes be recurrent and associated with distress or dysfunction.[4] Although ICD-10 lists IED under 'Other habit and impulse disorders', it does not provide specific criteria for its diagnosis.

Regarding phenomenology, persons with IED describe their aggressive episodes as explosive, uncontrollable, unpremeditated, and brief; often provoked by minor stimuli; and associated with various psychological and physical symptoms, especially changes in mood, awareness, and autonomic arousal.[4,6] The frequency of

episodes depends in part on how the disorder is defined. In the National Comorbidity Survey Replication (NCS-R), where the DSM-IV A criteria of 'several' episodes was operationalized to be three or more lifetime attacks, persons with IED had a mean of 43 lifetime attacks.

Many persons with IED describe problems with chronic or trait anger and frequent 'subthreshold' aggressive episodes in which they manage to resist enacting aggressive impulses or express them with less destructive behaviours (e.g. screaming rather than assault).[4,6] These subthreshold episodes are similar to the anger attacks (sudden episodes of intense anger with autonomic arousal) often described in patients with mood (bipolar and depressive) disorders.[7]

Of note, the relationship between IED specifically and impulsive aggression in general remains unclear and the two phrases are not necessarily synonymous. In particular, like other impulse control disorders, the aggression of IED usually involves elements of lack of control (e.g. compulsivity) and affect dysregulation (extreme anger, irritability and/or mood instability) as well as impulsivity. Thus, IED may be one form of impulsive aggression.

Epidemiology and course

Once thought to be rare, recent research has shown that IED is common in both clinical and general population samples. In the most rigorous study to date, the NCS-R, lifetime and 12-month prevalence estimates of DSM-IV IED in the general population were 7.3 per cent and 3.9 per cent, respectively.[5] The lifetime and 12-month prevalences of more narrowly defined IED (in which three episodes in the same year were required for diagnosis) were 5.4 per cent and 3.5 per cent, respectively. The disorder is also likely to be common in forensic populations but data are not available.

IED is probably more common in males than females. In the NCS-R, 9.3 per cent of men versus 5.6 per cent of women met lifetime DSM-IV criteria for the disorder. IED begins in childhood or adolescence; follows a chronic or episodic course; and is associated with distress, morbidity (e.g. injuries), and social and occupational impairment.[4-6] For example, in the NCS-R, IED had a mean age of onset of 14 years, was persistent over the life course (with averages of 6.2–11.8 years with attacks), and was associated with substantial role impairment.[5] However, the prevalence of the disorder was significantly lower among persons 60 years and older (2.1 per cent).

Associated psychopathology and comorbidity

IED often co-occurs with other psychiatric disorders.[4–6] In the NCS-R, 81.8 per cent of respondents with lifetime DSM-IV IED met criteria for at least one other lifetime DSM-IV disorder. Specifically, IED was significantly comorbid with all DSM-IV depressive, anxiety, and substance use disorders assessed after controlling for age, sex, and race. It was also significantly comorbid with oppositional defiant disorder, conduct disorder, and attention-deficit/hyperactivity disorder.

Importantly, the boundaries between IED and other conditions characterized by episodic and/or impulsive aggression have not been clearly delineated. Indeed, the comorbidity between IED and both bipolar disorders and Axis II disorders was not assessed in the NCS-R.[5] Comorbidity with Axis II disorders was not determined because the prevalence of these disorders was not evaluated. Comorbidity with bipolar disorders was not assessed because of how IED was defined; cases of IED with lifetime mania or hypomania were excluded from analysis (number not specified) so that the prevalence of IED was not overestimated by cases of bipolar disorder with anger attacks. Of note, although the relationships between IED and both bipolar and cluster B personality disorders remain unclear, clinical studies suggest that patients with IED have high rates of these conditions.[4,6]

Family studies

Family studies suggest that relatives of probands with IED have high rates of impulsive violent behaviour, substance abuse, and possibly mood and other impulse control disorders.[4,6,8,9] In a family history study of patients with temper outbursts meeting the first two DSM-III criteria for IED, non-adopted patients were significantly more likely than adopted patients to have a family history of temper outbursts.[9] Of 25 subjects with DSM-IV IED evaluated via the family history method, 8 (32 per cent) of subjects had a first-degree relative with probable IED, 20 (80 per cent) had at least one first-degree relative with a substance use disorder, 14 (56 per cent) a mood disorder, and 14 (56 per cent) an impulse control disorder.[6] A blinded, controlled family history study using broadly-defined IED criteria found a significantly increased morbid risk of the condition in relatives of affected probands (26 per cent) compared with relatives of control probands (8 per cent).[4]

Psychobiology

Persons with impulsive aggression have been consistently found to have abnormalities in serotonergic function.[4] Although most studies included subjects with impulsive aggression and personality disorders, a few included persons with IED or possible IED. Thus, in a study of 58 violent offenders and impulsive fire-setters, 33 (57 per cent) of whom had DSM-III IED, lower cerebrospinal fluid (CSF) concentrations of 5-hydroxyindoleacetic acid (5-HIAA) were found in the impulsive offenders and fire-setters than in the non-impulsive offenders and normal control subjects.[10]

In a functional magnetic resonance imaging (MRI) study of response to social threat, 10 subjects with IED showed exaggerated amygdala reactivity and diminished orbitofrontal cortex activation to faces expressing anger compared with controls.[11] The authors noted these findings were similar to other disorders characterized by impulsive aggression, including borderline personality and bipolar disorders, and that they supported a link between a dysfunctional frontal-limbic network and aggression.

Treatment response

Clinical experience suggests that IED may be less responsive to insight-oriented and more responsive to cognitive behavioural therapies, particularly those stressing anger management.[4,6,12] Medications reported effective in definite or probable IED, some in controlled trials, include antiepileptics (e.g. phenytoin, carbamazepine, oxcarbazepine), antidepressants (e.g. tricyclics, serotonin reuptake inhibitors), mood stabilizers (e.g. lithium, valproate), β-blockers, psychostimulants, and even antiandrogens. Mood stabilizer monotherapy and antidepressant augmentation of mood stabilizers have both been reported to successfully treat IED and/or anger attacks in patients with bipolar disorders.[6,7] Antidepressants have been reported to be effective in anger attacks associated with

major depression.[12] Finally, serotonin reuptake inhibitors, mood stabilizers, antiepileptics, and antipsychotics may be effective for impulsive-aggressive behaviour in personality-disordered patients.[13]

Kleptomania

Definition and clinical features

Kleptomania is defined in DSM-IV as follows:

- recurrent failure to resist impulses to steal objects that are not needed for personal use or for their monetary value (criterion A);
- increasing sense of tension immediately before committing the theft (criterion B);
- pleasure, gratification, or relief at the time of committing the theft (criterion C);
- the stealing is not committed to express anger or vengeance and is not in response to a delusion or a hallucination (criterion D);
- the stealing is not better accounted for by conduct disorder, a manic episode, or antisocial personality disorder (criterion E).

In ICD-10, kleptomania (or pathological stealing) is defined as the repeated failure to resist impulses to steal objects that are not acquired for personal use or monetary gain.

An increasing number of studies have systematically examined the phenomenology of groups of people with DSM-defined kleptomania.[14–17] In these studies, most subjects described irresistible impulses or urges to steal, tension with the impulses, and tension relief either during or shortly after the act of theft (as required by the DSM criteria). Many subjects described the impulses as senseless, intrusive, uncomfortable, and uncontrollable. Many tried to resist the impulses with varying degrees of success. Some reported pleasurable feelings during the act of theft, often described as 'a rush,' 'a high,' or 'a thrill.' Most patients reported instances of impulsive stealing, but some also described premeditated stealing, the aim of which was sometimes to relieve the impulses to steal. Many subjects reported that they had lied to conceal their stealing. Some subjects developed rules for their stealing behaviour—for instance, stealing only from work or from certain types of shops (e.g. drug stores but not department stores), or stealing certain items but not others (e.g. jewellery but not clothing). Many subjects considered their stealing to be wrong, and many, but not all, reported guilt or remorse after stealing. Subjects who had been arrested for shoplifting reported that it had varying effects on their symptoms—some stopped stealing completely, some stopped for a limited amount of time, while others reported that their stealing was unaffected. Some stated they continued to steal once incarcerated.

These studies have also found that some subjects with apparent kleptomania report varying degrees of amnesia surrounding the act of stealing.[17] Many of these subjects deny impulses, tension, or relief with their thefts. Other subjects who are not amnesiac for their stealing episodes may also deny experiencing impulses, tension, relief, and/or pleasure. For these subjects, stealing appears to have become automatic or habit-like.[17]

Epidemiology and course

Kleptomania is presumed to be rare but its prevalence is unknown. Available studies suggest that only a small portion of shoplifters (from none to 8 per cent) represent true cases of kleptomania.[14] However, it has been argued that these rates may be spuriously low because psychiatric evaluations may not have always been sufficiently thorough, operational diagnostic criteria were rarely used, and kleptomania may have been under-represented in the samples due to selection bias (i.e. people with repeated apprehensions were more likely to be legally rather than psychiatrically referred). Also, kleptomania may be relatively common in clinical populations; it was the second most common lifetime impulse control disorder in a group of adult psychiatric inpatients assessed with a structured interview, present in 9.3 per cent of the sample.[18]

Kleptomania is probably more common in women than in men.[14–17] Many cases begin in adolescence or early adulthood, and often follow an episodic or a chronic course.

Associated psychopathology and comorbidity

Clinical studies show that kleptomania often co-occurs with other Axis I psychiatric disorders, including mood, anxiety, substance use, eating, and impulse control disorders.[14–18] In the only controlled study,[16] 10 patients with kleptomania had significantly higher rates of comorbid psychiatric disorders, particularly mood disorders, other impulse control disorders, and substance abuse or dependence (mainly nicotine dependence), than 29 psychiatric comparison patients and 60 patients with alcohol abuse or dependence. Several studies, including the one controlled study, found especially high rates of bipolar disorders.[15–17] Conversely, high rates of kleptomania have been found in women with eating disorders[14] and patients with depressive disorders.[19]

Preliminary data suggest patients with kleptomania may also have high rates of certain Axis II disorders.[17] However, the relationship between kleptomania and antisocial personality disorder is not understood.

Family studies

Uncontrolled studies suggest kleptomania may be associated with increased familial rates of mood, substance use, anxiety, and possibly impulse control disorders.[17] For example, of 103 first-degree relatives of 20 patients with DSM-IIIR-defined kleptomania evaluated blindly by the family history method, 22 (21 per cent) had a major mood disorder, 21 (20 per cent) had a substance use disorder, and 13 (13 per cent) had an anxiety disorder, including seven (7 per cent) with obsessive compulsive disorder[15] Also, two (2 per cent) had apparent kleptomania. However, a controlled family study found similar rates of kleptomania in first-degree relatives of probands with obsessive compulsive disorder and those of control probands.[20]

Psychobiology

Preliminary research suggests kleptomania may be associated with serotonergic and frontal lobe dysfunction. In one study, the number of platelet serotonin transporters, evaluated via [3H] paroxetine binding, was lower in 20 patients with obsessive-compulsive related disorders, including five patients with kleptomania, than in 20 healthy control subjects.[17] In another study, 10 females with DSM-IV kleptomania were more likely than controls to have decreased white matter microstructural integrity in inferior frontal brain regions when evaluated with diffusion tensor imaging.[21]

Treatment response

Although no controlled psychological treatment studies of kleptomania have been published, various types of cognitive behavioural therapy may be effective.[12,17] There are also successful reports of the use of psychodynamic psychotherapies, but there are negative reports as well.[12,15]

Medical treatments with antidepressant, mood-stabilizing, or anxiolytic properties have been reported to be effective in kleptomania, primarily in case reports and case series. These treatments include tricyclics, serotonin reuptake inhibitors, trazodone, lithium, valproate, electroconvulsive therapy, and benzodiazepines.[12,15,17] There are also reports of patients with kleptomania responding to the opioid antagonist naltrexone and the antiglutamatergic agent topiramate.[12,17]

However, in the only controlled pharmacotherapy study of kleptomania published to date, an open-label trial of escitalopram treatment in 24 subjects followed by double-blind discontinuation in 15 of 19 responders, there was no difference in response rate (defined as greater than a 50 per cent decrease in theft episodes per week) between subjects receiving escitalopram (3 [43 per cent]) and those receiving placebo (4 [50 per cent]).[22]

Pyromania

Definition and clinical features

Pyromania is defined in DSM-IV as follows: deliberate and purposeful fire-setting on more than one occasion (criterion A) that is associated with tension or affective arousal before the act (criterion B), fascination with, interest in, curiosity about, or attraction to fire and its situational contexts (criterion C), and pleasure, gratification, or relief when setting fires, or when witnessing or participating in their aftermath (criterion D). Also, the fire-setting is not done for monetary gain, as an expression of sociopolitical ideology, to conceal criminal activity, to express anger or vengeance, to improve one's living circumstances, in response to a delusion or hallucination, or as a result of impaired judgement (criterion E), and is not better accounted for by conduct disorder, a manic episode, or antisocial personality disorder (criterion F). In ICD-10, pyromania (or pathological fire-setting) is defined as multiple acts of, or attempts at, setting fire to property or other objects, without apparent motive, and by a persistent preoccupation with subjects related to fire and burning. The essential features are as follows:

- repeated fire-setting without any obvious motive such as monetary gain, revenge, or political extremism
- an intense interest in watching fires burn
- reported feelings of increasing tension before the act, and intense excitement immediately after it has been carried out.

Although the authors were unable to locate any systematic reports of a group of people with pyromania by either of the above criteria sets, there are numerous case reports and case series of people with repetitive fire-setting behaviour who would probably meet these criteria for pyromania. For example, in what is still probably the largest study of pathological fire-setting, in 1951, Lewis and Yarnell[23] evaluated 1 145 of 2 000 American case records of males 16 years of age and older from the National Board of Fire Underwriters (selection criteria were otherwise not clearly specified). They concluded that 688 of these males were best classified as 'pyromaniacs' as 'they set fires for no practical reason and received no material profit from the act, their only motive being to obtain some sort of sensual satisfaction'. Lewis and Yarnell did not provide quantitative data summarizing these 688 cases, but stated that 50 of the subjects 'approached true pyromania', in that they were able to give a 'classical description of the irresistible impulse'. Specifically, before they set fires, these subjects described 'mounting tension; . . . restlessness; the urge for motion; . . . conversion symptoms such as headaches, palpitations, ringing in the ears, and the gradual merging of their identity into a state of unreality'.

Epidemiology and course

The prevalence of pyromania is unknown.[19] Although there are numerous studies of fire-setting behaviour, few of these studies systematically assessed pyromania in their subjects. Those that did use variable definitions of pyromania and reported widely discrepant rates. For example, in their 1951 study of 1 145 adult males with pathological fire-setting, Lewis and Yarnell[23] reported that 688 (60 per cent) could be classified as having broadly defined pyromania, but only 50 (4 per cent) as having the 'true' disorder.

Pyromania is probably more common in males than females and usually begins in adolescence or early adulthood.[19,23] How often childhood fire-setting represents pyromania is unknown. Clinical descriptions indicate that the course of pyromania may be episodic or chronic, but its course into old age is unknown.

Associated psychopathology and comorbidity

There are no studies of the psychiatric comorbidity of a group of people with well-defined pyromania. However, there are case reports of people with apparent pyromania who have comorbid mood, obsessive compulsive, eating, paraphilic, and possibly psychotic disorders.[23] In addition, impulsive fire-setters have been reported to have high rates of mood, substance use, and cluster B personality disorders, as well as suicide attempts.[10]

Family studies

There are no family history studies of pyromania, but studies of impulsive fire-setters suggest elevated familial rates of substance use disorders.[8]

Biological studies

There are no biological studies of pyromania, but studies of impulsive fire-setters suggest they have abnormalities in central serotonergic neurotransmission.[10]

Treatment response

There are no systematic treatment studies of pyromania.[12,19] There is one case of an 18-year-old male with DSM-IV pyromania and a chief complaint of 'feeling addicted to setting fires' who responded to the combination of topiramate and cognitive behaviour therapy.[24] There is also a case report of two men in both of whom pyromania appeared to be part of a paraphilia and responded to antiandrogen medication.[25]

Clinical reports on the treatment of pathological fire-setting in general stress use of various psychological interventions (e.g. cognitive behavioural, psychoeducational, supportive, and insight-oriented) to address the fire-setting behaviour, and appropriate

treatment of any comorbid psychiatric disorders. Preliminary data suggest combined psychosocial treatment approaches may be helpful in reducing further fire-setting behaviour, at least in juveniles.[19]

Trichotillomania

Definition and clinical features

Trichotillomania is defined in DSM-IV as follows:

- recurrent pulling out of one's hair resulting in noticeable hair loss
- an increasing sense of tension immediately before pulling out the hair or when attempting to resist the behaviour
- pleasure, gratification, or relief when pulling out the hair
- the disturbance is not better accounted for by another mental disorder and is not due to a general medical condition (e.g. a dermatological condition)
- the disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.

In ICD-10, trichotillomania is defined as noticeable hair-loss due to a recurrent failure to resist impulses to pull out hairs. ICD-10 further states the hair-pulling is usually preceded by mounting tension and is followed by a sense of relief or gratification and that the diagnosis should not be made if there is a pre-existing inflammation of the skin, or if the hair pulling is in response to a delusion or a hallucination.

Hair is most often pulled from the scalp but also from the eyelashes, eyebrows, face, axilla, arms, legs, abdomen, and pubis.[26,27] Extracted hair may be chewed or swallowed. Medical complications include trichobezoars (hairballs that form in the stomach) and, uncommonly, obstruction or perforation of the stomach or bowel.

Some authorities have argued that both the DSM-IV and ICD-10 criteria for trichotillomania are too narrow, noting that patients with distressing hair pulling behaviour, especially children, may not always experience impulses and/or tension before hair pulling or relief with or after hair pulling.[26–28] Indeed, for some persons, the hair pulling may be automatic or habit-like and not associated with urges, tension, or relief. In addition, the hair loss may not be noticeable. For this reason and because hair pulling is a self-grooming behaviour, some authorities have argued that trichotillomania should be grouped with other self-grooming behaviours that may become problematic (e.g. skin picking and nail biting, which are discussed later) into a family of grooming disorders or body-focused impulse control disorders.[20,26,27]

Epidemiology and course

The prevalence of DSM-IV or ICD-10 defined trichotillomania is unknown, but survey studies suggest between 0.5 per cent—3.5 per cent of college students report problematic hair pulling.[26,27]

Clinical studies indicate trichotillomania is more common in females than in males; may begin in childhood, adolescence, or adulthood; and may have an episodic or chronic course.[26,27] Spontaneous remissions may occur, particularly in children with recent onset of the disorder.

Associated psychopathology and comorbidity

Trichotillomania often co-occurs with mood, anxiety, eating, substance use, and other impulse control disorders in clinical samples of adults.[26,27] It may also co-occur with various personality disorders, with histrionic, borderline, and obsessive compulsive commonly being cited.[26,27] In paediatric samples, trichotillomania is similarly associated with mood, anxiety, and substance use disorders.[26]

Family Studies

Preliminary family research suggests that trichotillomania may be associated with increased rates of obsessive-compulsive and grooming disorders among first-degree relatives.[26] A controlled family study of 22 probands with compulsive hair pulling, 17 (77 per cent) of whom met DSM-III-R criteria for trichotillomania, found that depression, alcoholism, drug abuse, obsessive compulsive disorder, and antisocial personality disorder were significantly more frequent among the first degree relatives of hair pullers than relatives of control probands.[28] Conversely, another controlled family study found significantly higher rates of grooming disorders (trichotillomania as well as pathological skin picking, pathological nail biting, and impulse control disorder NOS) in first-degree relatives of probands with obsessive compulsive disorder than in first-degree relatives of control probands.[20]

Psychobiology

Neuroimaging studies in subjects with trichotillomania have shown hyperactivity in the cerebellum and right superior parietal lobe as well as possible structural abnormalities of the left putamen, left inferior frontal gyrus, right cuneal cortex, and cerebellum.[26,27,29] Neuropsychological abnormalities found in trichotillomania patients have included increased error rates in spatial processing, divided attention, nonverbal memory, and executive functioning.[26,29]

Genetic studies suggest trichotillomania may be associated with sequence variants in the slit and trk-like 1 (SLITRK1) gene and the T10ZC polymorphism of the serotonin receptor 2A gene.[30,31]

Treatment response

There are a few published reports of the successful treatment of trichotillomania with insight-oriented psychotherapies, but there are many successful reports with various behavioural-based therapies.[12,26,27] Indeed, to date, four randomized, controlled studies supporting the effectiveness of behavioural treatments in adult trichotillomania have been published. In these studies, habit reversal was more effective than negative practice training (N=34); CBT was more effective than clomipramine and placebo (N=16); behaviour therapy was superior to fluoxetine and a wait-list control (N=43); and acceptance and commitment therapy plus habit reversal was superior to wait-list control (N=25).[12,26,27] In addition, one randomized, controlled trial found a cognitive behavioural therapy package of awareness training, stimulus control, and habit-reversal training was superior to minimal attention control in paediatric trichotillomania.[27]

Although many case reports and open trials describe successful treatment of trichotillomania with various serotonin reuptake inhibitors, controlled studies have yielded mixed results.[12,26,27] Two small (N=13 and N=12) double-blind, crossover trials found

clomipramine was superior to desipramine and equivalent to fluoxetine (which had beneficial effects), respectively, in reducing hair-pulling symptoms. In contrast, 2 slightly larger (N=21 and N=23) placebo-controlled, double-blind crossover studies of fluoxetine (both up to 80 mg/day) in adult chronic hair pullers found fluoxetine was not superior to placebo in suppressing hair-pulling symptoms.[12,26,27] In addition, as noted earlier, in the two controlled comparisons which found cognitive behaviour therapy or behaviour therapy superior to clomipramine and fluoxetine in 16 and 43 patients with trichotillomania, respectively, neither drug was superior to placebo in reducing trichotillomanic symptoms.[12]

Despite these negative studies, thymoleptics other than serotonin reuptake inhibitors have been reported to be effective for trichotillomania in case reports and case series. These include antidepressants such as imipramine, amitriptyline, isocarboxazid, trazodone, mianserin, and bupropion; the mood stabilizers lithium, olanzapine, and quetiapine; and various antimanic antipsychotics—used as either monotherapy or adjunctively with serotonin reuptake inhibitors.[12,26,27]

Other medications described as being effective for trichotillomania, primarily in case reports or case series, are buspirone, fenfluramine, topiramate, inositol, the progestin levonorgestrel, and naltrexone.[12,26,27,32,33] Naltrexone was reported superior to placebo in one small controlled trial which has only been presented in abstract form.[12] Case reports also describe the successful use of topical steroid ointments in combination with psychotropics when skin is irritated.[12]

Compulsive buying disorder

History and clinical description

Although compulsive buying disorder (also called compulsive shopping, buying mania, and oniomania) is not classified in DSM-IV or ICD-10 as a mental disorder, diagnostic criteria have been proposed.[34,35] These include being frequently preoccupied with buying or subject to irresistible, intrusive, and/or senseless impulses to buy; frequently buying unneeded items or more than can be afforded; shopping for periods longer than intended; and experiencing adverse consequences, such as marked distress, impaired social or occupational functioning, and/or financial problems.[35] Persons with compulsive buying disorder often report irresistible or uncontrollable impulses to buy or shop; mounting tension or anxiety with the impulses; and relief of tension and/or pleasurable feelings (e.g. 'a high', 'a buzz', or 'a rush') with the act of buying or shopping. The disorder is associated with distress, financial and legal difficulties, and impairment in social and vocational functioning.[34,35]

Epidemiology and course

Compulsive buying behaviour is thought to be common, with an estimated lifetime prevalence of 5.8 per cent in the United States general adult population.[36] Indeed, in a recent study of the prevalence of various impulse control disorders in a psychiatric inpatient population, compulsive buying disorder was the most common current (9.3 per cent) and lifetime (9.3 per cent) impulse control disorder diagnosis.[18]

Compulsive buying disorder is probably more common in women than men.[34,35] It may begin in adolescence or adulthood and usually has either an episodic or a chronic course. The course of compulsive buying disorder into old age is unknown.

Associated psychopathology and comorbidity

Compulsive buying disorder often co-occurs with mood, anxiety, substance use, eating, and other impulse control disorders in clinical samples.[34,35] It may also be associated with certain personality disorders, but this comorbidity has received less systematic attention.[34]

Family studies

Preliminary research, including one controlled study, suggests compulsive buying disorder is associated with increased familial rates of mood, substance use, and possibly impulse control disorders.[34,35]

Psychobiology

In a molecular genetic study, no association was found between two serotonin transporter gene polymorphisms and compulsive buying disorder.[34]

Treatment response

Isolated reports of psychoanalytic and insight-oriented psychotherapy in compulsive buying disorder have mostly been unsuccessful. Cognitive behavioural therapy may hold promise. Two patients with compulsive buying disorder each responding to cue exposure plus response prevention after failing clomipramine treatment have been described.[12] In addition, a randomized study found group cognitive behavioural therapy (N=28) superior to wait-list control (N=11) in reducing compulsive buying episodes, time spent buying, and scores on buying symptom measures in 39 patients with compulsive buying disorder.[37] Support groups patterned after Alcoholics Anonymous, such as Debtors Anonymous, self-help books, and financial counselling are available, but their effectiveness has not been formally evaluated.[12,34]

Various antidepressant medications have been reported to be effective for compulsive buying in case reports and open trials, but two randomized, placebo-controlled, double blind studies of fluvoxamine in a total of 54 patients with compulsive buying (N=17 and N=37, respectively) failed to show separation between drug and placebo.[12,34] Both studies, however, were limited by high placebo response rates, possibly due to the use of diaries to record buying behaviour. In a 7-week, open-label trial of citalopram (N=24) followed by a 9-week double-blind, placebo-controlled continuation trial that omitted use of shopping diaries for the 15 subjects who met responder criteria, none of 7 patients randomized to remain on citalopram relapsed as compared with 5 (63 per cent) of 8 patients randomized to receive placebo (P=0.019)[12,34].

Patients with compulsive buying have also been reported to respond to mood stabilizers, naltrexone (at 100 mg/day but not 50 mg/day), and topiramate.[12,34]

Repetitive self-mutilation

Clinical description

Repetitive self-mutilation, also called impulsive deliberate self-harm, parasuicide, or self-injurious behaviour, is the repeated,

direct destruction of body tissue without suicidal intent.[38,39] Examples include skin cutting, skin burning, self-hitting, severe skin scratching, and even bone breaking. A wide range of body parts are mutilated, such as arms, legs, abdomen, head, chest, and genitals.

Numerous clinical studies suggest that this syndrome often meets the DSM-IV and ICD-10 definitions of impulse control disorders.[38] Specifically, repetitive self-mutilation is characterized by intrusive, recurrent, and irresistible impulses to harm oneself without suicidal intent that are associated with increasing tension, anxiety, anger, or other dysphoric affective states, along with relief of the uncomfortable affect with or shortly after the act of self-harm. In addition, the act of self-harm is often not associated with pain (i.e. associated with analgesia) and performed privately.[38,39]

Epidemiology and course

The prevalence of narrowly defined repetitive self-mutilation is unknown.[38,39] Clinical studies, however, suggest that the condition is more common in females than males, usually begins early in life (e.g. late childhood, adolescence, and early adulthood), and may persist for 10 to 15 years.[38,39]

Associated psychopathology and comorbidity

Repetitive self-mutilation often co-occurs with other Axis I and II psychiatric disorders.[38,39] These include mood, substance use, eating, psychotic, dissociative, and borderline personality disorders. Repetitive self mutilation may also co-occur with suicide attempts and adverse childhood experiences in patients with certain pathologies, especially borderline personality disorder.[39] Of note, although deliberate self injury is a core feature of borderline personality disorder, not all patients with repetitive self mutilation have borderline personality disorder.[38]

Family history

No family history studies of repetitive self-mutilation have been conducted.

Psychobiology

Although studies have consistently found an association between low central CSF 5-H1AA levels with both impulsive aggression and violent suicide, the results of such studies in patients with repetitive self mutilation have been mixed—with some, but not all, finding similar reductions.[38,39] One study found that broadly-defined self harm was associated with allelic variation in the tryptophan hydroxylase gene (TPH A779C), but not with polymorphisms of five other serotonergic genes.[40]

Studies have found increased pain thresholds in borderline personality patients with repetitive self mutilation who are analgesic to the pain of their self injurious behaviour, suggesting dysfunction the endogenous opioid system.[38,39] In support of this possibility, one study found elevated plasma metenkephalin studies in a small group of analgesic self-injuring persons. Another study, however, found pretreatment with naltrexone did not reduce the anesthesia (as evaluated by the cold pressor test) of a similar group of subjects.[39]

Treatment response

There are no controlled treatment studies of narrowly-defined repetitive self-mutilation. However, two psychological treatments, dialectal behaviour therapy and psychoanalytically-oriented partial hospitalization, have each been shown superior to 'treatment as usual' in decreasing chronic parasuicidal behaviour in women with borderline personality disorder.[39] In addition, a 1998 meta-analysis of 20 treatment trials of broadly-defined deliberate self harm indicated significantly reduced repetition of self harm for problem solving therapy and provision of an emergency contract card in addition to standard care.[41]

Regarding medical treatments, agents with antidepressant, mood-stabilizing, antipsychotic, anticonvulsant, and anti-opiate properties have been reported to be effective in case reports or case series.[12,33,39] In the above noted 1998 meta-analysis, a significantly reduced rate of further self harm was observed for depot flupenthixol versus placebo in the one study of antipsychotic medication that was located and analyzed.[41] The two studies of antidepressants evaluated, however, showed no benefit.

Pathological skin picking

Clinical description

Pathological skin picking (also called neurotic or psychogenic excoriation, compulsive skin picking, dermatotillomania, and *acné excorié*) is excessive scratching, picking, gouging, or squeezing of the skin sometimes in response to an itch or other skin sensation or to remove a lesion on the skin.[42,43] Most patients use fingernails to excoriate the skin, but the teeth and instruments (for example, tweezers, nail files, pins, or knives) are also used. Pathological skin picking causes substantial distress in patients, with most experiencing functional impairment and many reporting medical complications, some severe enough to warrant surgery.

Although not recognized as a distinct DSM-IV or ICD-10 disorder, pathological skin picking resembles an impulse control disorder in that patients sometimes experience an increase in tension prior to picking with transient relief or pleasure with picking or immediately afterwards. Many patients find themselves acting automatically. It also has compulsive features, in that it is repetitive, ritualistic, anxiety reducing, often resisted, and egodystonic. Moreover, some patients describe obsessions about an irregularity on the skin or preoccupations with having smooth skin.

Epidemiology and course

Pathological skin picking may occur in about 2 per cent of dermatology clinic patients, predominately in women, and up to 3.8 per cent of college students.[42,43] The disorder may begin in adolescence or adulthood, and the mean duration of symptoms has ranged from 5 to 18 years, with a better prognosis for patients who have had the symptoms for less than 1 year.[42]

Associated psychopathology and comorbidity

Pathological skin picking often co-occurs with mood, anxiety, and somatoform disorders.[42] It is especially common in body dysmorphic disorder.[43] The comorbidity of pathological skin picking and personality disorders has not been systematically studied.

Family studies

Preliminary family history data suggest first-degree relatives of probands with pathological skin picking may have elevated rates of mood and substance use disorders.[42]

Treatment response

Various behavioural treatments (e.g. habit reversal) may be effective in pathological skin picking.[12] One small placebo-controlled trial (N=21) found that fluoxetine (mean dose of 55 mg/day) may be beneficial[12]. Other serotonin reuptake inhibitors, the tricyclic doxepin (which has been hypothesized to have antipruritic properties due to its antihistaminic effects), inositol, the glutamate-modulating agent riluzole, and certain direct skin treatments (dermatologic and surgical) may also be effective.[12,32,42,44,45]

Onychophagia

Clinical description

Onychophagia is repetitive and excessive nail-biting.[46] The cuticles and skin around the nails are also frequently bitten, picked, or clipped. Onychotillomania, the excessive picking, clipping, or tearing of the nail, may be a variant.

Although not classified as a psychiatric disorder in DSM-IV or ICD-10, onychophagia resembles an impulse control disorder in that the behaviour is often irresistible, automatic, and associated with an increase of tension before and relief or pleasure during or immediately after its enactment.[46] It also has compulsive features in that it is repetitive, resisted, and associated with relief of anxiety.

Epidemiology and course

Nail-biting is more common in children than adults, and may affect 5 to 10 per cent of adults over the age of 30 years.[46] Boys and girls are affected equally until after the age of 10 years, when nail-biting becomes more common in boys.

Associated psychopathology and comorbidity

Onychophagia may be associated with mood, anxiety, and substance use disorders. Of 25 adult subjects who underwent a medication trial for onychophagia, 17 (68 per cent) had a lifetime Axis I psychiatric disorder—despite the exclusion of subjects with obsessive–compulsive disorder, a primary major affective disorder, current substance abuse, or psychosis.[46] Four subjects (16 per cent) had at least one personality disorder.

Family studies

Severe nail biting may be familial.[46,47] Of 112 family members of 25 subjects entering a medication trial for onychophagia, seven (6 per cent) had severe nail-damaging behaviour, four (4 per cent) were severe nail-biters, and three (3 per cent) picked or chewed their hands or feet.[46] In addition, twin studies have found higher concordance rates of nail biting in monozygotic compared with dizygotic twins.[47]

Treatment response

Various cognitive behavioural therapies (especially habit reversal, but also self-monitoring, use of bitter tasting substances, competing responses, and negative practice training) are probably effective in onychophagia.[12] In the only controlled pharmacotherapy study of onychophagia, clomipramine (mean dose 120 mg/day) was superior to desipramine (mean dose 135 mg/day) in eliminating nail-biting, reducing nail-biting severity and impairment, and in improving overall clinical progress.[46]

Conclusion

Growing research shows that the impulse control disorders are much more common than once thought to be. The consistency of the 'structure' of the irresistible impulse (a core disturbance of impulsivity and compulsivity) together with increasing research showing that it responds to certain treatments, especially cognitive-behavioural psychotherapies and medical treatments with thymoleptic or anticraving properties, regardless of its 'content' (the specific impulse experienced), strongly suggest that it is an important psychopathological symptom, and that impulse control disorders are legitimate mental disorders that are in fact likely to be related despite their apparent differences.

Further information

Hollander, E. and Stein, D.J. (eds.) (2006). *Clinical Manual of Impulse-Control Disorders.* American Psychiatric Publishing, Inc., Arlington, VA.

Gabbard, G.O. (ed.) (2007). *Gabbard's Treatments of Psychiatric Disorders* (4th ed.). American Psychiatric Publishing, Inc., Arlington, VA.

References

1. Hollander, E. and Stein, D.J, (eds.) (2006). *Clinical manual of impulse control disorders*. American Psychiatric Publishing, Arlington, VA.
2. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). American Psychiatric Association, Washington, DC.
3. World Health Organization. (1992). *International statistical classifi cation of diseases and related health problems*, 10th revision. WHO, Geneva
4. Coccaro, E.F. and Danehy, M. (2006). Intermittent explosive disorder. In *Clinical manual of impulse control disorders* (eds. E. Hollander and D.J. Stein,), pp. 19–37. American Psychiatric Publishing, Arlington, VA.
5. Kessler, R.C., Coccaro, E.F., Fava, M., *et al.* (2006). The prevalence and correlates of DSM-IV intermittent explosive disorder in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **63**, 669–78.
6. McElroy, S.L., Soutullo, C.A., Beckman, D.A., *et al.* (1998). DSM-IV intermittent explosive disorder: a report of 27 cases. *Journal of Clinical Psychiatry*, **59**, 203–10.
7. Perlis, R.H., Smoller, J.W., Fava, M., *et al.* (2004). The prevalence and clinical correlates of anger attacks during depressive episodes in bipolar disorder. *Journal of Affective Disorders,* **79**, 291–5.
8. Linnoila, M., DeJong, J. and Virkkunen, M. (1989). Family history of alcoholism in violent offenders and impulsive fire setters. *Archives of General Psychiatry*, **46**, 613–16.
9. Mattes, J.A. and Fink, M. (1990). A controlled family study of adopted patients with temper outbursts. *Journal of Nervous and Mental Disease*, **178**, 138–9.
10. Virkkunen, M., DeJong, J., Bartko, J., *et al.* (1989). Psychobiological concomitants of history of suicide attempts among violent offenders and impulsive fire setters. *Archives of General Psychiatry*, **46**, 604–6.
11. Coccaro, E.F., McCloskey, M.S., Fitzgerald, D.A., *et al.* (2007). Amygdala and orbitofrontal reactivity to social threat in individuals with impulsive aggression. Biological Psychiatry, **62,** 168–71
12. McElroy, S.L. and Keck, P.E. Jr (2007). Impulse control disorders. In *Gabbard's treatments of psychiatric disorders,* (4th edn),(ed. G.O. Gabbard), pp. 877–88. American Psychiatric Publishing, Arlington, VA.
13. Goedhard, L.E., Stolker, J.J., Heerdink, E.R. *et al.* (2006). Pharmacotherapy for the treatment of aggressive behavior in general adult psychiatry: a systematic review. *Journal of Clinical Psychiatry,* **67**, 1013–24.

14. McElroy, S.L., Keck, P.E. Jr, Pope, H.G.J., *et al.* (1991). Kleptomania: clinical characteristics and associated psychopathology. *Psychological Medicine*, **21**, 93–108.
15. McElroy, S.L., Pope, H.G. Jr, Hudson, J.I., *et al.* (1991). Kleptomania: a report of 20 cases. *American Journal of Psychiatry*, **148**, 652–7.
16. Bayle, F.J., Caci, H., Millet, B., *et al.* (2003) Psychopathology and comorbidity of psychiatric disorders in patients with kleptomania. *American Journal of Psychiatry*, **160**, 1509–13.
17. Grant, J.E., (2006). Kleptomania. In *Clinical manual of impulse control disorders* (eds. E. Hollander and D.J. Stein), pp. 175–201. American Psychiatric Publishing, Arlington, VA.
18. Grant, J.E., Levine, L., Kim, D., *et al.* (2005). Impulse control disorders in adult psychiatric inpatients. *American Journal of Psychiatry*, **162**, 2184–8.
19. Lejoyeux, M., McLoughlin, M. and Adès, J. (2006). Pyromania. In *Clinical manual of impulse control disorders* (eds. E. Hollander and D.J. Stein), pp. 229–50. American Psychiatric Publishing, Arlington VA.
20. Bienvenu, O.J., Samuels, J.F., Riddle, M.A., *et al.* (2000). The relationship of obsessive-compulsive disorder to possible spectrum disorders: results from a family study. *Biological Psychiatry*, **15**, 287–93.
21. Grant, J.E., Correia, S. and Brennan-Krohn, T. (2006).White matter integrity in kleptomania: A pilot study. Psychiatry Research, **147**, 233–7.
22. Koran, L.M., Aboujaoude, E.N. and Gamel, N.N. (2007). Escitalopram treatment of kleptomania: an open-label trial followed by double-blind discontinuation. *Journal of Clinical Psychiatry*, **68**, 422–7.
23. Lewis, N.D.C. and Yarnell, H. (1951). Pathological firesetting (pyromania). Nervous and mental disease monograph 82. Coolidge Foundation, New York.
24. Grant, J.E. (2006). SPECT imaging and treatment of pyromania. *Journal of Clinical Psychiatry,* **67**, 998.
25. Bourget, D. and Bradford, J. (1987). Fire fetishism, diagnostic and clinical implications: a review of two cases. *Canadian Journal of Psychiatry*, **32**, 459–62.
26. Franklin, M.E., Tolin, D.F. and Diefenbach, G.J. (2006). Trichotillomania. In *Clinical manual of impulse control disorders* (eds. E. Hollander and D.J. Stein), pp. 1149–73. American Psychiatric Publishing, Arlington, VA.
27. Woods, D.W., Flessner, C., Franklin, M.E., *et al.* (2006). Understanding and treating trichotillomania: what we know and what we don't know. *Psychiatric Clinics of North America*, **29**, 487–501.
28. Schlosser, S., Black, D.W., Blum, N., *et al.* (1994). The demography, phenomenology, and family history of 22 persons with compulsive hair pulling. *Annals of Clinical Psychiatry*, **6**, 147–52.
29. Keuthen, N.J., Makris, N., Schlerf, J.E., *et al.* (2007). Evidence for reduced cerebellar volumes in trichotillomania. *Biological Psychiatry,* **61**, 374–81.
30. Zuchner, S., Cuccaro, M.L., Tran-Viet, K.N., *et al.* (2006). SLITRK1 mutations in trichotillomania. *Molecular Psychiatry*, **11**, 887–9.
31. Hemmings, S.M., Kinnear, C.J., Lochner, C., *et al.* (2006). Genetic correlates in trichotillomania – a case-control association study in the South African Caucasian population. *Israel Journal of Psychiatry and Related Sciences*, **43**, 93–101.
32. Lochner, C., Seedat, S., Niehaus, D.J., *et al.* (2006). Topiramate in the treatment of trichotillomania: an open-label pilot study. *International Clinical Psychopharmacology*, **21**, 255–9.
33. Seedat, S., Stein, D.J. and Harvey, B.H. (2001). Inositol in the treatment of trichotillomania and compulsive skin picking. *Journal of Clinical Psychiatry*, **62**, 60–1.
34. Black, D.W. (2006). Compulsive shopping. In *Clinical manual of impulse control disorders* (eds. E. Hollander and D.J. Stein), pp. 203–28. American Psychiatric Publishing, Arlington, VA.
35. McElroy, S.L., Keck, P.E. Jr., Pope, H.F. Jr., *et al.* (1994). Compulsive buying: a report of 20 cases. *Journal of Clinical Psychiatry*, **55**, 242–8.
36. Koran, L.M., Faber, R.J., Aboujaoude, E., *et al.* (2006). Estimated prevalence of compulsive buying behavior in the United States. *American Journal of Psychiatry*, **163**, 1806–12.
37. Mitchell, J.E., Burgard, M., Faber, R., *et al.* (2006). Cognitive behavioral therapy for compulsive buying disorder. *Behavior Research and Therapy*, **44**, 1859–65.
38. Favazza, A.R. (1998). The coming of age of self mutilation. *Journal of Nervous and Mental Disease*, **186**, 259–68.
39. Simeon, D. (2006). Self-injurious behaviors. In *Clinical manual of impulse control disorders* (eds. E. Hollander and D.J. Stein), pp. 63–81. American Psychiatric Publishing, Arlington, VA.
40. Pooley, E.C., Houston, K., Hawton, K., *et al.* (2003). Deliberate selfharm is associated with allelic variation in the tryptophan hydroxylase gene (TPH A779C), but not with polymorphisms in five other serotonergic genes. *Psychological Medicine*, **33**, 775–83.
41. Hawton, K., Arensman, E, Townsend, E., *et al.* (1998). Deliberate self harm: systematic review of effi cacy of psychosocial and pharmacological treatments in preventing repetition. *British Medical Journal*, **327**, 441–7.
42. Arnold, L.M., Auchenbach, M.B. and McElroy, S.L. (2001). Psychogenic excoriation. Clinical features, proposed diagnostic criteria, epidemiology and approaches to treatment. *CNS Drugs*, **15**, 351–9.
43. Grant, J.E., Menard, W. and Phillips, K.A. (2006). Pathological skin picking in individuals with body dysmorphic disorder. *General Hospital Psychiatry*, **28**, 487–93.
44. Sasso, D.A., Kalanithi, P.S., Trueblood, K.V., *et al.* (2006). Beneficial effects of the glutamate-modulating agent riluzole on disordered eating and pathological skin-picking behaviors. *Journal of Clinical Psychopharmacology*, **26**, 685–7.
45. Bowes, L.E. and Alster, T.S. (2004). Treatment of facial scarring and ulceration resulting from ***acné excorié*** with 585-nm pulsed dye laser irradiation and cognitive psychotherapy. *Dermatologic Surgery*, **30**, 934–8.
46. Leonard, H.L., Lenane, M.C., Swedo, S.E., *et al.* (1991). A double-blind comparison of clomipramine and desipramine treatment of severe onychophagia (nail biting). *Archives of General Psychiatry*, **48**, 821–7.
47. Ooki, S. (2005). Genetic and environmental influences on finger-sucking and nail-biting in Japanese twin children. *Twin Research and Human Genetics*, **8**, 320–7.

4.13.2 Special psychiatric problems relating to gambling

Emanuel Moran

Introduction

Gambling is an activity with the following elements:

- A contract between two or more people, which is based on a forecast of the outcome of an uncertain event involving random processes.
- Property, referred to as the stake, is transferred between those taking part, so that some gain at the expense of others.
- The property transfer depends on the outcome or result of the uncertain event, which has been forecast.
- Participation is voluntary and not necessarily related to gaining the property, but used to obtain an experience.

Clinical features

Gambling misuse is a behavioural disorder that can usually be recognized by the presence of any of the following features:

- Excessive gambling either in terms of the money spent or the time devoted.
- Intermittent or continuous preoccupation with gambling and the development of tolerance and craving for it.
- Loss of control over gambling and 'chasing of losses', despite the realization that damage is resulting.
- Disorder affecting the person who is gambling and the family:
  - financial disturbances, such as debt and shortage;
  - social disturbances, such as loss of employment and friends, running away from home, eviction, marital problems, divorce, behaviour disorders in the children of the family, criminality and imprisonment;
  - psychological disturbances, such as depression and attempted suicide.

Classification

In the past, this syndrome has been referred to as compulsive gambling. However, it is not a true obsessive–compulsive state but a heterogeneous group of conditions, characterized by excessive gambling resulting in disturbance for those involved. The term 'pathological gambling' is more appropriate, since it is not based on any assumptions regarding the underlying processes.[1]

ICD-10[2] describes pathological gambling as a form of behaviour under 'habit and impulse disorders'. On the other hand, DSM-IV[3] implies a homogeneous disease entity and provides criteria for its recognition under 'impulse-control disorders not elsewhere classified'. The ICD-10 approach is preferable since it emphasizes the fact that the condition is a behavioural disorder resulting from faulty habits.

Five varieties of pathological gambling can be recognized[4,5]:

- Subcultural gambling arises out of the person's background, which is one of socially accepted heavy gambling.
- Impulsive gambling is characterized by loss of control for varying periods and the tendency to be associated with tolerance, craving, and dependence on the activity.
- Neurotic gambling occurs as a response to an emotional problem, particularly in a disturbed relationship in marriage or during adolescence.
- Symptomatic gambling occurs in mental illness, usually depression, which is the primary disorder.
- Psychopathic gambling is part of the generalized disturbance of behaviour that characterizes antisocial personality disorder.

Diagnosis

For various social reasons, pathological gambling is most easily recognized in men since they tend to patronize those types of gambling that have a high turnover of money so that excess is more likely to become apparent.

Women have tended to gamble on lotteries, bingo, and football pools. These may not involve such large sums of money and excess often presents with disturbances in the social sphere rather than through the accumulation of large debts. However, the greater general acceptance of gambling and the advent of remote gambling via the Internet, television, and mobile devices is changing the situation considerably.

While pathological gambling is seen in all age groups, an increasing number of children and young people are presenting with the condition as a result of gambling on slot/gaming machines. Also, in recent years, remote gambling among children and young people has led to increasing problems. This is in spite of the fact that most jurisdictions treat gambling as an adult activity. Pathological gambling in adulthood frequently has its origins in heavy gambling in childhood and adolescence.

Aetiology and epidemiology

The nature of gambling

The experience of risk provides amusement, thrill, and excitement and is therefore pleasurable. These experiences make gambling attractive and the stake money is used to purchase them, with winnings as an occasional bonus. A few, who gamble professionally, are also able to win money regularly because they have sources of information that reduce the uncertainty, as in betting on horses and dogs. Their gambling is planned and deliberate.

Gambling is usually organized commercially with the odds in favour of the provider. There is therefore an in-built financial disadvantage to those who use the facilities. In slot/gaming machines where the provider is at a distance from the gambling event, this is often not apparent to those who take part.

Commercial gambling involves large sums of money, and has traditionally been confined to licensed premises. Those present have gone there because they have decided to take part in the gambling. However, developments in technology have made it possible to provide gambling facilities on a remote basis via the Internet, television, and mobile devices.

A number of features inherent in the activity of gambling have effects that make it difficult for a person to stop.

Psychological effects

- Underlying all gambling activity is operant conditioning with intermittent variable ratio reinforcement.[6] This is a most effective schedule for habit-formation and produces a stable, persistent response. Consequently, the long-term net gain or loss to those who gamble is almost irrelevant to the continuation of the activity. It is most dramatically seen in slot/gaming machines, which consequently are the main source of profit for the gambling industry.
- Rapid turnover gambling as in casinos restricts the ability of those who gamble to apply any considered judgement. Inevitably, gambling becomes more impulsive, easily leading to excessive participation.
- The assessment of probability of winning (psychological probability) in the gambling situation differs from the mathematical probability. At low probabilities, it is higher than the mathematical probability and at moderate and high probabilities, it is lower. This even occurs in people who are mathematically knowledgeable.

- In a gambling situation involving only random processes, where the outcome of successive events is completely independent, there is usually the irrational belief that a string of losses makes a win more likely. This is the negative recency effect, which is also referred to as the'Monte Carlo Fallacy' since it forms the basis of many spurious gambling systems, especially in roulette. Paradoxically, it is associated with the belief that a string of wins is likely to continue ('a lucky streak'). Also, a 'near win' generally tends to be treated as a prelude to a win. These illogical ways of thinking encourage continuous gambling and are exploited by slot/gaming machines and lottery scratch cards called 'heart stoppers'.
- Large prizes, even at very low probabilities, entice the gambler because of the *possibility* of winning. The stimulant effect of roll-overs in lotteries illustrates this.
- Skill in gambling is usually overrated and often implies an unrealistic ability to control the uncertain event that is the subject of the gamble. Thus, in dog and horse race betting, punters tend to place their bets just before 'the off' in the fantasy that this will affect the result.
- There is a tendency to lose track of time during a gambling session.

These psychological effects have been increased as a result of recent developments in commercial gambling.

- Loyalty cards providing rewards for money spent on gambling in a particular facility.
- Remote gambling on the Internet, television, and mobile devices has resulted in the following:
  - The convenience and anonymity of gambling being available in an isolated domestic setting, without the checks and constraints that can be exercised by the presence of others as in licensed premises.
  - Monetary credit in the form of e-cash systems, reducing the likelihood that those gambling will set a limit on the money staked.
  - Behavioural targeting and messages on some online gambling sites encouraging further gambling when an attempt is made to stop.
  - Difficulty in preventing children and young people from having access, especially since the advent of online social networks.

Physical effects

- A gambling loss in normal subjects immediately results in particular localized activity in the medial frontal cortex of the brain. This is then associated with subsequent *more risky* gambling choices. This is consistent with the negative recency effect.
- Disturbances involving the reward pathways in the brain are significantly associated with excessive gambling.
- There is a great range and strength of emotions during gambling decisions associated with cortical responses in the brain to the expectation of winning money.
- Even normal, social levels of drinking alcohol that alter self-control over decision-making, increase the difficulty in deciding at what point to stop, when losing, in a gambling situation.

Predisposing factors

In the presence of available gambling facilities, certain predispositions may increase the likelihood of pathological gambling.

Morbid risk-taking

Since gambling is a type of risk-taking, it lends itself to be used by those who, for reasons related to their personality, have a high need for risk. They spend large sums of money on the intangible commodity of risk, which may easily pass unnoticed because it is fleeting.

This morbid propensity to take risks shows itself in other ways. Thus, the incidence of attempted suicide is high among those whose gambling is pathological.[5]

Other personality factors

Freud's formulation of gambling was that it resembles masturbation, is a substitute for it, and is resorted to in the context of unresolved Oedipal difficulties.[7] Others have pointed out that pathological gambling may be a manifestation of self-punishment, with an unconscious desire to lose, arising from a psychological mechanism referred to as 'psychic masochism'.[8]

Those whose gambling is pathological appear to have other predisposing personality traits. They view their behaviour as being largely determined by factors outside their personal control. They also tend towards greater impulsivity.[5,9]

Learning processes

Apart from the winnings and losses, the gambling situation itself may affect learning. As far as the random processes inherent in gambling are concerned, all participants, even a total failure, stand on an equal footing. This may be the only circumstance in which some people have this experience. Gambling may therefore provide a means of dealing with morbid anxiety in the presence of feelings of inadequacy, leading to a conditioned avoidance reaction.

Mental disorder

Pathological gambling may occur in any mental disorder. However, it is most commonly associated with depression. More usually, a neurotic type of depression occurs after a bout of heavy gambling with large losses. In symptomatic pathological gambling, the depression is primary and the gambling is a response to the tension and feelings of guilt that occur in depression. This latter situation is similar to alcohol misuse and shoplifting, as part of the depressive syndrome. Pathological gambling may also be a manifestation of antisocial personality disorder.

Misuse of alcohol and pathological gambling can occur together; either may be the primary disorder and either may lead to the other.

Constitutional factors and physical disorder

Twin studies have demonstrated that the likelihood of pathological gambling occurring in a person is influenced, to an important degree, by inherited factors and/or experiences shared during childhood.[10]

There also appears to be a significant association between pathological gambling and genetic abnormalities involving the dopamine reward pathways.[11] Disturbances of serotonergic, noradrenergic, and dopaminergic neurotransmitter systems have

all been implicated in the aetiology of pathological gambling. This is particularly so in relation to the arousal, behavioural initiation, behavioural disinhibition, and reward/reinforcement mechanisms that are evident in this condition.[12]

There have been reports of pathological gambling associated with dopamine agonist administration for Parkinsonism.[13]

Course and prognosis

The natural history of pathological gambling is one characterized by exacerbations and remissions, often related to life events. Important elements in this are relationships within the family, especially with the spouse/partner. An example of this is the not infrequent sequence of an exacerbation of heavy gambling in the husband, at the time of the wife's first pregnancy.

The outlook in pathological gambling is usually determined by the integrity of the underlying personality. In those in whom the condition appears as a symptom of a neurotic disorder or depression, the prognosis depends on that of the underlying disorder.

Management and treatment

Pathological gambling involves a whole way of life, which has many ramifications. If its management is to be successful, there need to be major changes in the lifestyle of the person concerned. It is best dealt with by a team approach involving at least a psychiatrist, psychologist, and social worker and must include the spouse/partner. Recently, counselling services have been set up but their efficacy has yet to be established.

Assessment of the problem

The following aspects are important:

- An appraisal of the extent and amount of present gambling.
- A history of the development of the gambling from its early beginnings, which is best done if the person being assessed provides the information by means of a detailed written narrative.
- A discussion of this written narrative.
- An indication of the person's motivation, since many who seek help for pathological gambling readily admit that they enjoy it and only want assistance for the problems that have resulted.
- At least initially, an immediate period of total abstinence from gambling.

Supervision of the finances

Excessive gambling is usually associated with a disturbed appreciation of the value of money. In view of this and the continued temptation to gamble, the family finances should be dealt with as follows:

- All monies should be controlled, at least for some time, by the spouse/partner or some trusted person.
- Regular income from wages/salaries should be paid into a bank account over which the spouse/partner or trusted person has sole control.
- A detailed statement should be drawn up of all the outstanding debts, as well as an inventory of the income and outgoings of the person-seeking help and his or her family.
- The person whose gambling has been pathological should discuss the matter with all creditors and agree a repayment plan. This should be consistent with the person's regular income and circumstances to avoid a situation where there would be the temptation to gamble in order to maintain repayments. Since debts are often considerable, these may have to continue over many years.
- After a period of abstinence from gambling, the person whose gambling has been pathological needs to become gradually involved in working jointly with whoever controls the finances.

Counselling

On the basis of information obtained during the course of the initial assessment, the following aspects need further consideration:

- The features inherent in gambling that affect people so that they find it difficult to stop should be highlighted and discussed.
- Social relationships of the person whose gambling has become pathological and the spouse/partner should be reviewed, especially if there have been serious marital problems predating the pathological gambling.
- The way spare time is spent, what friends are cultivated, and what interests are pursued should be reviewed. Since incitement to gamble will have occurred in the past within specific settings, arrangements need to be made to avoid these or, at least, to be prepared for them.
- A joint contract to be reviewed regularly spelling out in detail those types of behaviour to be avoided as well as those to be encouraged may be found helpful.

Gamblers Anonymous

This form of self-help for pathological gambling is organized in regular local groups. As well as meetings for those who have a gambling disorder, there are also separate ones for their spouses/partners. Quite apart from the valuable work done in the group setting, Gamblers Anonymous provides a useful means of establishing alternative social contacts from those that were associated with gambling. Indeed, for some people, Gamblers Anonymous may be the vehicle through which all the necessary help can be provided. Even if this is not the case, Gamblers Anonymous still provides a valuable form of support for the individual and the family.

Psychological treatments

A variety of psychological treatments have been advocated but, in general, their long-term efficacy has not been established. A good outcome has been reported after a cognitive behavioural approach.[14] Also, controlled gambling, rather than permanent abstinence, has led to a reported successful outcome after behavioural treatment.[15]

Psychiatric treatments

Specialist treatment from a psychiatrist and/or a psychotherapist for a neurotic disorder or severe depression may be required, if these clearly underlie the pathological gambling.

Prevention

In view of the nature of gambling and the importance of the social causation of pathological gambling, it is vital that it should be seen as an activity that requires moderation. Unfortunately, the recent increasing reliance of governments and states on gambling for revenue purposes is resulting in a vast growth in the availability of gambling facilities and the incitements to participate.

This has been associated with public policies that actively promote gambling and also claim to encourage moderation. The inconsistency in trying to do both inevitably has a harmful effect on any educational attempt to provide a sensible attitude to gambling. It also undermines any help for those whose gambling has become excessive.

Further information

Journal of Gambling Studies. Springer. http://www.springer.com/west/home/social+sciences/sociology?SGWID=4-40440-70-35680327-0 Online version available http://www.springerlink.com/content/1050-5350

International Gambling Studies. Routledge. http://www.informaworld.com/smpp/title~content=t713701604~tab=sample?action=view&db=all

References

1. Moran, E. (1970). Gambling as a form of dependence. *British Journal of Addiction*, **64**, 419–27.
2. World Health Organization. (1992). *International statistical classification of diseases and related health problems*, 10th revision. WHO, Geneva.
3. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
4. Moran, E. (1970). Varieties of pathological gambling. *The British Journal of Psychiatry*, **116**, 593–7.
5. Moran, E. (1970). Clinical and social aspects of risk-taking. *Proceedings of the Royal Society of Medicine*, **63**, 1273–7.
6. Skinner, B.F. (1966). *Science and human behaviour*. Macmillan, New York.
7. Freud, S. (1961). Dostoevsky and parricide. In *Standard edition of the complete psychological works of Sigmund Freud*, Vol. 21 (ed. J. Strachey), p. 177. Hogarth Press, London.
8. Bergler, E. (1970). *The psychology of gambling*. International Universities Press, New York.
9. Steel, Z. and Blaszczynski, A. (1998). Impulsivity, personality disorders and pathological gambling severity. *Addiction*, **93**, 895–905.
10. Eisen, S.A., Lin, N., Lyons, M.J., *et al.* (1998). Familial influences on gambling behaviour: an analysis of 3359 twin pairs. *Addiction*, **93**, 1375–84.
11. Cummings, D.E. (1998). The molecular genetics of pathological gambling. *CNS Spectrums: International Journal of Neuropsychiatric Medicine*, **3**, 20–37.
12. DeCaria, C.M., Begaz, T., and Hollander, E. (1998). Serotonergenic and noradrenergic function in pathological gambling. *CNS Spectrum: International Journal of Neuropsychiatric Medicine*, **3**, 38–47.
13. O'Sullivan, S.S. and Lees, A.J. (2007). Pathological gambling in Parkinson's disease. *Lancet Neurology*, **6**, 384–6
14. Sylvain, C., Ladouceur, R., and Boisvert, J.M. (1997). Cognitive and behavioural treatment of pathological gambling: a controlled study. *Journal of Consulting and Clinical Psychology*, **65**, 727–32.
15. Blaszczynski, A., McConaghy, N., and Frankova, A. (1991). Control versus abstinence in the treatment of pathological gambling: a two to nine year follow-up. *British Journal of Addiction*, **86**, 299–306.

4.14

Sleep–wake disorders

Contents

4.14.1 Basic aspects of sleep–wake disorders

Gregory Stores

Introduction

A sound working knowledge of the diagnosis, significance, and treatment of sleep disorders is essential in all branches of clinical psychiatry. Unfortunately, however, psychiatrists and psychologists share with other specialties and disciplines an apparently universal neglect of sleep and its disorders in their training. Surveys in the United States and Europe point to the consistently meagre coverage of these topics in their courses at both undergraduate and postgraduate levels.

The following account is an introductory overview of normal sleep, the effects of sleep disturbance, sleep disorders and the risk of failure to recognize them in psychiatric practice, assessment of sleep disturbance, and the various forms of treatment that are available. The aim is to provide a background for the other chapters in this section.

The close links between the field of sleep disorders and psychiatry which make it essential that psychiatrists are familiar with the field are as follows:

- Sleep disturbance is an almost invariable feature and complication of psychiatric disorders from childhood to old age, with the risk of further reducing the individual's capacity to cope with their difficulties (see Table 4.14.1.3 for further details).
- Sleep disturbance can presage psychiatric disorder.
- Some psychotropic medications produce significant sleep disturbance.
- Of importance to liason psychiatry is the fact that many general medical or paediatric disorders disturb sleep sufficiently to contribute to psychological or psychiatric problems.
- Because of lack of familiarity with sleep disorders and their various manifestations, such disorders may well be misinterpreted as primary psychiatric disorders (or, indeed, other clinical conditions) with the result that effective treatments for the sleep disorder are unwittingly withheld (see later).

Some of these points will be amplified in later sections of this chapter.

Basic features of normal sleep

The scientific study of sleep and its disorders is largely confined to the last several decades. Essentially interdisciplinary advances have displaced earlier speculative accounts including those in psychiatry concerning the significance of dreams, for example. They are well described in recent textbooks of sleep disorders medicine (see recommended sources). Only general points are mentioned here, with special reference to psychiatry where possible.

The nature of sleep

Sleep has characteristic physiological features which distinguish it from other states of relative inactivity. Two distinct sleep states have been defined, that is non-rapid eye movement (**NREM**) sleep and rapid eye movement (**REM**) sleep. The onset of sleep is not simply the shutdown of wakefulness but also the switching between wakefulness, NREM and REM sleep involve complicated active neurochemical mechanisms in different parts of the brain.

The functions of sleep

Debate continues about the various theories concerning sleep, each of which has emphasized physical and psychological restoration and recovery, energy conservation, memory consolidation, discharge of emotions, brain growth and various other biological functions including somatic growth and repair, and maintenance of immune

systems. No one theory accounts for all the complexities of sleep and it seems likely that sleep serves multiple purposes.

From the practical point of view, the most obvious observation is that both physical and psychological impairment follows persistent sleep disturbance. Animals totally deprived of sleep for a long periods die with loss of temperature regulation and multiple system failure. As described later, the adverse effects of chronic sleep loss (considered to be common in modern society) on mood, behaviour, and cognitive function can be substantial, with various consequences for personal, social, occupational, educational, and family functioning.

Sleep stages

Conventionally, standard criteria are used to identify different sleep stages according to their characteristic physiological features especially in the electroencephalogram (**EEG**), electrooculogram (**EOG**), and electromyogram (**EMG**).

NREM sleep is divided into four stages of increasing depth. **Stage I** occurs at sleep onset or following arousal from another stage of sleep. This stage represents 4–16 per cent of the main sleep period. **Stage II** contains more slow EEG activity but is still relatively light sleep. It accounts for 45–55 per cent of overnight sleep. **Stage III** (4–6 per cent of total sleep time) contains yet more slow EEG activity. **Stage IV** is characterized by the slowest activity and constitutes 12–15 per cent of sleep. The combination of stages III and IV is called slow wave sleep (**SWS**) or delta sleep and is considered to be the deepest form of sleep from which awakening is particularly difficult. The arousal disorders such as sleepwalking arise from SWS.

REM sleep is physiologically very different. Brain metabolism is highest in this stage of sleep. Spontaneous rapid eye movements are seen and the skeletal musculature is effectively paralysed. Heart rate, blood pressure, and respiration are all variable, body temperature regulation ceases temporarily, and penile and clitoral tumescence occurs. REM sleep usually takes up 20–25 per cent of total sleep time. Most dreams, including nightmares, occur in REM sleep.

Sleep architecture

NREM and REM sleep alternate cyclically throughout the night starting with NREM sleep lasting about 80 min followed by about 10 min of REM sleep. This 90 min sleep cycle is repeated three to six times each night. Each REM period typically ends with a brief arousal or transition into light NREM sleep.

In successive cycles the amount of NREM sleep decreases and the amount of REM sleep increases. SWS is usually confined to the first two sleep cycles. The diagrammatic representation of overnight sleep is known as a **hypnogram**, a simplified form of which is shown in Fig. 4.14.1.1.

In addition to this conventional sleep staging, there has been increasing interest in the microstructural fragmentation of sleep by frequent, brief arousals (seen mainly in the EEG) lasting a matter of seconds without obvious clinical accompaniments. This subtle type of sleep disruption, overlooked by conventional sleep staging, is increasingly associated with impairment of daytime performance, mood, and behaviour.

Circadian sleep–wake rhythms

The timing of sleep (but not its amount) is regulated by a circadian 'clock' in the suprachiasmatic nucleus (**SCN**) of the hypothalamus.

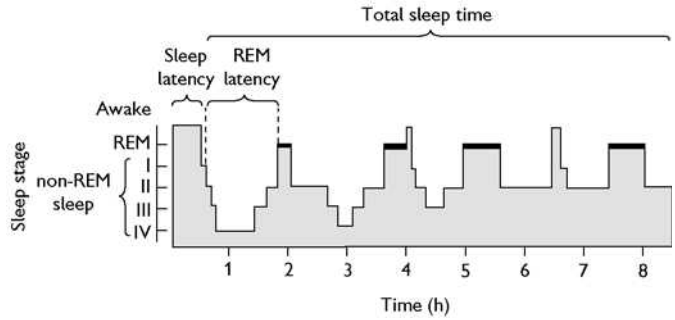


Fig. 4.14.1.1 Diagram of an overnight hypnogram in a young adult.

The intrinsic circadian sleep–wake rhythm is close to 24 h in human adults. Other species are different, an extreme example being dolphins and some other creatures which shut down one cerebral hemisphere at a time ('unihemispheric sleep'), allowing them to be constantly alert. From an early age the individual sleep–wake rhythm has to synchronize with the 24-h day–night rhythm. The main *zeitgeber* by which this is achieved ('entrainment') is sunlight but social cues, such as mealtimes and social activities, are also important.

The SCN also controls other biological rhythms including body temperature and cortisol production with which the sleep–wake rhythm is normally synchronized. In contrast, growth hormone is locked to the sleep–wake cycle and is released with the onset of SWS, whatever its timing.

Melatonin is related to the light–dark cycle rather than the sleep–wake cycle. It is secreted by the pineal gland during darkness and suppressed by exposure to bright light ('the hormone of darkness'). It influences circadian rhythms via the SCN pacemaker which in turn, regulates melatonin secretion by relaying light information to the pineal gland. The widespread popularity of melatonin as a sleep-promoting agent is not justified by what little is known about its action and clinical effectiveness.

Changes with age

Changes in basic aspects of sleep are prominent from birth to old age, although individual differences are seen at all ages. Changes of clinical significance include the following:

- **Total sleep time** decreases with age. Average daily values are as follows: newborns 6–18 h; young children 10 h; adolescents 9 h (although often they obtain significantly less than this); adults 7.5–8 h, including possibly the same in elderly people. The total amount of sleep includes daytime napping in children up to the age of about 3 years.
- **SWS** is particularly prominent in prepubertal children who sleep very soundly. Its decline begins in early adolescence and continues throughout childhood.
- The proportion of **REM sleep** declines from 50 per cent or more of total sleep time in the newborn (more than this in premature babies) to 20–25 per cent by 2 years. This figure remains fairly constant throughout the rest of life. The high level of REM sleep in very early life suggests a role in cerebral maturation but the reason for its persistence throughout life is unclear. Memory processing appears to depend on sleep. However, people deprived

of REM sleep, experimentally or pathologically, can be relatively unaffected either emotionally or cognitively. Deep sleep decreases in the elderly.

- **NREM–REM sleep cycles** occur at intervals of 50–60 min in infants who often enter REM at the start of their sleep period. This interval between sleep cycles remains until adolescence when the periodicity changes to 90–100 min, which persists into adult life. The amounts of NREM and REM in each sleep cycle is about equal in early infancy. Afterwards, NREM sleep (especially SWS) predominates in the earlier cycles and REM sleep in the later cycles.
- **Continuity of sleep** is greatest in pre-pubertal children (as mentioned previously) and least at the extremes of age. Infants are easily awakened and so are the elderly who also wake spontaneously more often. Fragmentation of sleep by brief arousals, or very brief awakenings, is particularly common in old age.
- **Circadian sleep–wake rhythms** change considerably in early development. Full-term neomates show 3–4-h sleep–wake cycles. Sleep periods have largely shifted to the night and wakefulness to daytime by 12 months, except for napping which gradually diminishes and has usually stopped by about 3 years of age. However, a physiological tendency towards an afternoon nap remains throughout the rest of the life. Although repeated brief waking at night is more common in infancy and early childhood than later, it remains a normal occurrence throughout life, increasing in frequency again in old age. The clinical problem arises when there is difficulty returning to sleep after such awakenings.

Psychological effects of sleep disturbance

There is extensive clinical and experimental evidence that sustained sleep disturbance can have serious adverse psychological effects.[1,2] The term sleep disturbance covers the following:

- Loss of sleep (i.e. shortened duration).
- Impaired quality of sleep (repeated disruption of sleep architecture).
- Inappropriate timing of the sleep period in relation to day–night rhythms (as in the various circadian sleep rhythm disorders such as jet lag, shift work, or the more frequently encountered forms seen in clinical practice, as discussed later).

Experimental studies of **total sleep loss** demonstrate a progressive deterioration in cognitive function, mood, and behaviour related to length of sleep loss. However, inter- and also intra-individual differences in susceptibility are seen, reflecting such factors as motivation, personality, and usual sleep requirements. Task characteristics (e.g. brief or prolonged and monotonous tasks), timing of the task in relation to the circadian sleep–wake rhythm, and physical environmental factors such as noise and other distracting stimuli, are also important.

Variations for similar reasons are important in **partial sleep deprivation** experiments which (like those concerning fragmentation of sleep) are much closer to real-life sleep disturbance caused by social activities, job demands, and other aspects of modern lifestyle. These studies raise the issues of how much sleep is needed for optimal daytime functioning and whether these requirements are not being met. It has been argued that there is 'national sleep debt' in the United States and other western countries, and that by sleeping longer than they do habitually, many people would increase their performance and improve their well-being during the day.

The usual subjective effects of sleep disturbance are irritability, fatigue, poor concentration, and depression. More dramatic effects are described with prolonged and severe sleep disturbance, such as disorientation, illusions, hallucinations, persecutory ideas, and inappropriate behaviour with impaired awareness ('automatic behaviour') caused by frequent microsleeps. Psychometric studies have shown that sleep disturbance can produce a range of cognitive impairments, again depending on its duration and individual susceptibility. Sustained attention (vigilance) is particularly vulnerable and possibility abstract thinking and divergent intelligence or creativity.

The experimental findings are in keeping with the results from studies of various occupational groups including junior hospital doctors and drivers of various types of vehicle, in which reduced performance or accidents are associated with sleep disturbance. The common and increasing practice of night-shift work (as part of the '24 h society') is contrary to the fundamental biorhythm of sleeping at night and being awake during the day, and is often accompanied by a reduction in total sleep time and poor quality sleep. It is not surprising that working shifts commonly results in loss of well-being, physical complaints, and impaired productivity and safety, as well as physical disorders. Similarly, the distribution over the 24 h period of road accidents (especially those not involving other vehicles) and other mishaps at work, correspond to that of the levels of sleep tendency assessed objectively. Even industrial and engineering disasters have been attributed to sleep loss and impaired performance on the part of key personnel.

Additional evidence that sleep disturbance affects daytime function comes from neuropsychological studies of certain sleep disorders. Impaired performance on prolonged and complex tasks of subjects with narcolepsy has been shown to be secondary to the effects of their daytime sleepiness rather than an intrinsic neurological deficit. In the many adult patients with obstructive sleep apnoea, attention memory impairment (like depression and irritability commonly reported by these patients) are also largely attributable to daytime sleepiness. There is some evidence that deficits in more complicated 'executive functions' (formulating goals, planning, and carrying out plans effectively) are not necessarily reversed when their sleepiness is relieved by treatment. This might be the result of irreversible anoxic brain changes in the later stages of the condition. Clearly, early detection and treatment of this condition is essential to prevent this happening.

When return to normal sleep is possible, recovery from short periods of sleep disturbance occurs after much less sleep than that originally lost, for example, after one night's sleep following sleep loss over several days and nights. Reversal of the effects of long sleep disturbance in real-life is likely to be complicated, for example by emotional consequences of the disturbance.

Many of the above observations about the psychological effects of sleep disturbance (and their reversibility) have been made on young adult subjects or patients. The area is largely unexplored in other age groups but there is no reason why the general principle should not apply to children and the elderly including demented patients in whom sleep disturbance is particularly prominent.

Another group on whom further research is particularly required are people with learning disabilities (intellectual disability).

The available literature provides good reason to believe that the sleep disorders, especially in the more severely disabled groups, not only affects the majority but also are unusually severe and long-lasting because of lack of appropriate advice and treatment. The sleep disturbance is a problem in its own right and is often associated with various cognitive and behavioural abnormalities which might, at least partly, be the consequence of the sleep disturbance. Sleep disturbance in the duration or the quality of sleep may be one of the few ways of improving to some extent the psychological well-being of people with learning disabilities or dementia (and that of their carers) whose basic condition itself cannot be improved. In the case of the learning disabled, contrary to the common supposition by both professionals and relatives, success can usually be achieved (even in severe and long-standing problems), given an accurate diagnosis of the type of sleep disorder which may be predominantly behavioural or physical in type depending on the cause of the learning disability.

Sleep disturbance in the aetiology of psychiatric illness

Various 'psychotic' and other abnormal psychological phenomena were mentioned earlier resulting from prolonged and severe sleep disturbance, but these are reversed when normal sleep is restored. It remains an open question how often sleep disturbance is a primary cause of psychiatric illness. Evidence is patchy, tentative, and still in need of clarification.

- Over a wide age range, patients with a prior history of insomnia have been found consistently to be at significantly increased risk for severe depression. This could be interpreted in different ways including that sleep disturbance and the depression have a common underlying pathology, or that the sleep problems are an early sign of depression.
- A less fundamental role (but again implying preventative possibilities) is the suggestion that sleep deprivation late in pregnancy and in labour and childbirth at night might trigger post-natal depression.
- Abnormal circadian sleep–wake rhythms have been implicated in various depressive disorders including seasonal affective disorder (SAD). Light therapy has been used to correct the abnormality and relieve the depression and other symptoms.
- Disordered REM sleep mechanisms have (questionably) been considered as fundamental in the development of post-traumatic stress disorder symptoms.
- Some forms of attention-deficit hyperactivity disorder in children are attributed to persistent sleep disturbance.
- In a proportion of patients with schizophrenia, narcolepsy has been reported as the cause of their psychotic symptoms.

Disorders of sleep

Sleep complaints

The starting point for the clinician is the patient's sleep complaint. They are of three basic types:

- Not enough sleep, or unrefreshing sleep (**insomnia**).
- Sleeping too much (**excessive daytime sleepiness** or **hypersomnia**).
- Disturbed episodes during or otherwise related to sleep (**parasomnias**).

The detailed accounts later in this section are organized in relation to these main types of sleep complaint: insomnias (Chapter 4.14.2); excessive daytime sleepiness (Chapter 4.14.3); and parasomnias (Chapter 4.14.4). Sleep problems in childhood and adolescence are discussed in Chapter 9.2.9.

Whatever the clinical setting in which sleep complaints are investigated, the essential aim is to identify the specific sleep disorder from the many other conditions that can give rise to such complaints. Some sleep disorders may cause more than one type of complaint, and a patient may have more than one sleep disorder. The question arises how best to classify the many sleep disorders that have been described.

International classification of sleep disorders—second edition 2005 (ICSD-2)

This system, derived from wide international consultation, is the latest attempt to organize rationally the many ways in which sleep can be disturbed. ICSD-2 replaced the ICSD-Revised scheme outlined in the first edition of this textbook. More than 90 different sleep disorders are grouped as shown in Table 4.14.1.1. The grouping reflects the fact that knowledge about individual sleep disorders is very varied. The basic pathophysiology of some is quite well documented; in others little is known beyond their manifestations, and even they are subject to change as clinical observations improve. As a result, the ICSD-2 groupings are a mixture of those based on a common complaint (e.g. insomnia or hypersomnia), others on presumed aetiology (circadian rhythm sleep–wake disorders), and yet others are grouped according to the organ system from which the problems arise (such as sleep-related breathing disorders). Two additional groups in the system reflect current uncertainty about their status as disorders, or about their true nature.

Each sleep disorder is described in a standardized fashion using a series of sub-headings which include clinical features, demographics, pathology, and differential diagnosis. Treatments are not covered. Some publications are recommended concerning each sleep disorder. An attempt has been made throughout ICSD-2 to highlight aspects of sleep disorders of particular relevance to children.

In all, ICSD-2 provides a concise, easily accessible and up-to-date source of information for consultation by the clinician. Without being over-technical, it is more comprehensive and informative than current ICSD-10 and DSM-IV systems.

Sleep disorders mistaken for primarily psychological or psychiatric conditions

Sleep disorders manifest themselves in many ways. Failure to realize this can result in the misdiagnosis of primary sleep disorders as other types of clinical conditions of psychiatric, neurological, or otherwise medical conditions especially if there is limited familiarity with the sleep disorders field (which, unfortunately, is generally the case). Clearly such mistakes compromise patient care. The following are some of the main examples of this problem. Further details, with examples, are available elsewhere.[3]

Table 4.14.1.1 ICSD-2 groups of sleep disorders

I *Insomnias* The many psychological and physical causes of difficulty getting off to sleep, not staying asleep, early morning wakening, and feeling un-refreshed by sleep are included here including stress, poor sleep habits, and various mental and medical conditions.
II *Sleep-related breathing disorders* This group includes the common condition of obstructive sleep apnoea in adults and in children which often causes daytime sleepiness and other serious effects including changes in mood and behaviour. Central apnoea and various types of hypoventilation/hypoxaemic syndromes are also part of this group.
III *Hypersomnias of central origin not due to a circadian rhythm sleep disorder, sleep-related breathing disorder, or other cause of disturbed nocturnal sleep* Included here are narcolepsy and the causes of intermittent or recurrent hypersomnia such as the Kleine–Levin syndrome.
IV *Circadian rhythm sleep disorders* These disorders are characterized by a mistiming (and often disruption) of the sleep period, resulting in insomnia and/or hypersomnia. A prominent example is the delayed sleep phase syndrome common in adolescence. The advanced sleep phase syndrome can be seen in the elderly in which the sleep period begins in the evening with waking early when sleep requirements have been met. Irregular sleep–wake rhythms may be the result of an ill-organized way of life and substance abuse. Jet lag and nightshift work disorder are further examples of sleep problems caused by disturbance of the biological clock controlling the sleep–wake cycle.
V *Parasomnias* These are abnormal behaviours or sensations during or otherwise closely related to sleep. Many can be categorized according to the stage of sleep with which they are usually associated for example NREM sleep (sleepwalking) and REM sleep (nightmares, REM sleep behaviour disorder). Other parasomnias of particular psychiatric interest include sleep-related dissociative disorders and sleep-related eating disorders.
VI Sleep-related movement disorders These include the restless leg syndrome and periodic limb movement disorder.
VII Isolated symptoms, apparently normal variants and unresolved issues
VIII Other sleep disorders The classificatory scheme also includes an appendix on sleep disorders associated with conditions classifiable elsewhere such as sleep-related epilepsy, headaches, gastro-oesophageal reflux. A further appendix is concerned with other psychiatric and behavioural disorders frequently encountered in the differential diagnosis of sleep disorders. This appendix covers mood disorders, anxiety disorders, somatoform disorders, schizophrenia and other psychiatric disorders, personality disorders, and disorders of a psychiatric or behavioural type first diagnosed in infancy, childhood, or adolescence.

- Persistently not obtaining enough sleep, or having poor quality sleep because of interruptions by frequent subclinical arousals (as in obstructive sleep apnoea), is likely to cause tiredness, fatigue, irritability, poor concentration, impaired performance (possibly causing injuries or accidents at work or while driving), or depression. Out of the various possible explanations for such changes of behaviour, sleep disturbance may well be overlooked with failure to appreciate that, with improvement in sleep (which is usually possible with the right advice), such problems can be resolved. Occupational groups at special risk of sleep disturbance and its harmful effects include some clinicians.
- Excessive sleepiness, whatever its cause out of the many possibilities including physical conditions, is often misjudged as laziness, disinterest, daydreaming, lack of motivation, depression, intellectual inadequacy, or a number of other unwelcome states of mind. Sometimes, in very sleepy states, periods of 'automatic behaviour' occur, i.e. prolonged, complex, and possibly inappropriate behaviour with impaired awareness of events and, therefore, amnesia for them. Such episodes, the result of repeated 'microsleeps', can easily be misconstrued as reprehensible or dissociative features, misbehaviour, or prolonged seizure states. The paradoxical effect in young children of sleepiness causing over-activity has sometimes led to a diagnosis of attention-deficit hyperactivity disorder (ADHD) inappropriately treated with stimulant drugs instead of treatment for the sleep disorder.

A number of specific sleep disorders are at particular risk of being misinterpreted.

- In so-called delayed sleep phase syndrome, in which there is difficulty getting to sleep until very late, and problems getting up in the morning because of a shift in this timing of the sleep phase, is considered to be particularly common in adolescents who may be mistakenly thought to be awkward, lazy, irresponsible, or indulging in school refusal of the more usual type. In fact, this sleep disorder at that age is the result of a combination of normal pubertal biological body clock changes and alterations in lifestyle involving staying up late for social reasons or for study.
- It is not generally appreciated that even very complicated behaviour is possible whilst a person is still asleep as in sleepwalking episodes. Those with agitated sleepwalking or sleep terrors may appear to be very fearful and distressed and rush about and cry out as if escaping from danger. Other sleepwalkers develop an eating disorder with excessive weight gain due to the amount of food they consume while they are still asleep at night. Yet others behave in an aggressive or destructive way causing injury to themselves or other people and, at times sexual or other serious offences have been committed during a sleepwalking episode. If it is not known that such complicated actions are compatible with still being asleep, it is likely to be assumed that the person was awake at the time and aware of what he or she was doing, and, therefore, responsible for what had happened.
- Obstructive sleep apnoea is another case in point where this essentially physical disorder may be mistaken for being something very different from its true nature. The impairment of sleep quality, which characterizes this condition, can cause excessive daytime sleepiness, changes of personality, as well as adverse affects on social life and performance at work, as well as intellectual deterioration to the extent that dementia is suspected.
- Narcolepsy/cataplexy is also at serious risk of being misdiagnosed, sometimes for many years. Neurosis or depression is commonly mistaken for the narcolepsy symptoms.
- In REM sleep behaviour disorder there is a pathological retention of muscle tone during REM sleep so that a person can act out their dreams and behave violently if they have violent dreams. The dramatic behaviour that may result, including attacks on the sleeping partner, is easily misconstrued as intentional aggression.

- The complicated behaviour (far removed from that seen in other seizure states) that characterizes nocturnal frontal lobe epilepsy is often mistakenly thought to be evidence of a psychiatric disorder.
- In addition, sleep paralysis may be misconstrued as a psychotic disorder when (not uncommonly) accompanied by hallucinatory experiences.

In addition to sleep disorders being mistaken for psychiatric disorder, the opposite problem arises occasionally, that is some patients simulate excessive daytime sleepiness in order to avoid a psychologically troubling situation. Similarly, apparent parasomnias during sleep have sometimes been shown by polysomnography to occur when the patient is actually awake.

Detection and assessment of sleep disorders

As suggested earlier, evidence of a sleep disturbance should be actively sought in members of the general population and the various groups, including psychiatric patients, who are at special risk of sleep disorders. Otherwise, many instances of even severe sleep disturbance will continue to go unrecognized and untreated.

Sleep history

Routinely, all patients should be asked the following screening questions:

- Do you sleep long enough or well enough?
- Are you very sleepy during the day?
- Do you do unusual things or have strange experiences at night?

Ideally, their partner or other relatives should also be asked the same questions about the patient because the existence or severity of some forms of sleep disturbance are not known to the patient. In the case of children, parents are the main source of information but teachers' observations about daytime sleepiness or disturbance are also important.

If the answer to any of these enquiries is positive, a detailed sleep history is required. As traditional clinical history-taking schedules pay little attention to sleep, additional sleep-related enquiries will need to be made, covering the following points about the sleep problem.

- Precise nature of the sleep complaint, its onset, development, and current patterns.
- Medical or psychological factors at the onset of the sleep problem or which might have maintained it.
- Patterns of occurrence of the symptoms that is factors making them better or worse, weekdays compared to weekends, or work compared with holiday periods.
- Effect on mood, behaviour, work, social life, other family members.
- Past and present treatments for sleep problems and their effects.
- Past and present medication or other treatments for other illness or disorder.

In addition, detailed information is needed concerning the following:

- The patient's typical 24-h sleep–wake schedule. This can usually start with the evening meal, followed by preparation for and timing of bedtime, time and process of getting to sleep, events during the night, time and ease of waking up and getting up, level of alertness, and mental state and behaviour during the day.
- An attempt should be made to establish the duration, continuity, and timing of the patient's overnight sleep as these are the most important aspects of sleep for daytime functioning. It is also important to identify events of particular diagnostic significance, for example loud snoring.
- Sleep hygiene.

Compilation of a sleep history can be aided by the use of a preliminary sleep questionnaire (e.g. ref.[4]).

Sleep diary

Systematic recording in a booklet each day over 2 weeks or more, using a standardized format, avoids the bias or distortion of retrospective generalizations.

Other histories

Medical and psychiatric histories should include past and current treatment details (in view of the wide range of illnesses or disorders and their treatment with which sleep disturbance is associated). **Social history** should include occupational, marital, and recreational factors (drinking, smoking, drug use), which may affect sleep. **Family history** may be positive, for example in sleepwalking and associated arousal disorders.

Review of systems

Breathing difficulties and nocturia, for example, are associated with sleep disturbance. Severe obstructive sleep apnoea can cause cardio-respiratory and other cardiovascular complications.

Physical and mental state examination

Particular attention should be paid to the following:

- Evidence of any systemic illness including cardio-respiratory disease or neurological disorder (such as Parkinson's disease or stroke) which may disturb sleep.
- Obesity, oral and pharyngeal abnormalities, retrognathia, or mid-face deformity (predisposing to upper airway obstruction).
- Depression or other psychiatric disorder.
- Intellectual impairment, especially features of intellectual disability or dementia (including specific retardation syndromes) in view of their strong association with sleep disturbance.

Audio–video recording and actigraphy

Audio–video recordings can be very instructive in the parasomnias, sometimes revealing a very different picture than that provided in the clinic. Home video systems can be used where admission to hospital is not feasible. Similarly, monitoring of body movements via means of wrist-watch-like devices (**actigraphy**) can be used at home, if necessary, to quantify basic circadian sleep–wake patterns.

Polysomnography (PSG)

Physiological sleep studies are necessary for diagnosis in only the minority of sleep disorders. Traditionally this has entailed admission to a sleep laboratory. However, especially where such facilities are difficult to obtain, where the laboratory situation is unacceptable to patients (including some children or patients who are psychiatrically disturbed), **home PSG**, using portable systems, is useful, although the procedure has yet to be fully standardized and for some disorders is best seen as only a screening procedure. Recording in the home environment has the further advantage that, if the patient is allowed to adapt to the recording procedure before bedtime, the results from a single night's recording can be representative of the patient's habitual sleep. In contrast 'first night effects' are prominent in laboratory recordings and more than one night of polysomnography is required to allow adaptation to take place.

Basic polysomnography entails the recording of EEG, EOG, and EMG. This allows sleep to be staged and a hypnogram to be compiled as illustrated in Fig. 4.14.1.1. Usually the recording is made overnight but it may be continued during the day if required. Basic measures obtained from this information are as follows: total sleep time, time awake and number of awakenings, amount of REM and NREM sleep and their distribution overnight.

PSG can be extended to additional physiological measures, especially the following:

- Respiratory variables and audio–video recordings for sleep-related breathing problems.
- Additional EEG channels (combined with video) if nocturnal epilepsy is suspected.
- Anterior tibialis EMG for the detection of period limb movements in sleep.

Main indications for PSG are:

- The investigation of excessive daytime sleepiness, including the diagnosis of sleep apnoea, narcolepsy, or PLMS.
- The diagnosis of parasomnias where their nature is unclear from the clinical details, where PSG findings contribute essentially to the diagnosis (e.g. REM sleep behaviour disorder), where the possibility of epilepsy exists, or whether there may be more than one parasomnia present.
- As an objective check on the accuracy of the sleep complaint, or response to treatment.

Other investigations

Further laboratory investigations may be appropriate depending on the nature of the sleep problem and the purpose of the assessment.

- **A multiple sleep latency test** (**MSLT**), involving the recording of basic PSG variables, quantifies the degree of daytime sleepiness by measuring the time a patient takes to fall asleep during five opportunities to do so during the day. In adults, a mean sleep latency of 5 min or less indicates pathological sleepiness (usually of organic origin); 5 to 10 min is a grey area which usually includes excessive sleepiness associated with primary psychiatric disorder; while longer than 10 min is normal. These values do not apply in children whose sleep tendency varies with age.
- **HLA typing**, and possibly **CSF hypocretin (orexin) levels** for the investigation of narcolepsy, and **nocturnal penile tumescence monitoring** in the differential diagnosis of organic versus psychogenic impotence, are examples of other specific tests that may be appropriate depending on the clinical problem.

Treatment approaches for sleep disorders

In clinical practice, the pharmacological approach to treatment of sleep problems (especially insomnia) is generally overemphasized, especially in the use of hypnotic-sedative drugs. Table 4.14.1.2 provides some indication of the wide range of available types of treatment, as well as general principles of management, for adults and children. They are roughly arranged in order of the frequency of their use in the comprehensive management of sleep disorders in general. An appropriate choice from this range requires an accurate diagnosis of the underlying sleep disorder. Further details are provided in later contributions to this section on sleep disorders. Claims for the effectiveness of these various measures are based on widespread clinical experience and reports. Few randomized controlled trials have not been published, as yet.

Certain aspects of treatment with special relevance to psychiatric practice are included in Table 4.14.1.3.

Clinical sleep disorders in psychiatric conditions

As mentioned earlier, sleep disturbance is a feature of many psychiatric disorders at all ages. Sometimes a sleep disturbance is profound and constitutes one of the defining characteristics of the psychiatric condition. This aspect of the psychiatric state requires careful definition and quite possibly treatment in its own right, alongside primarily psychiatric help, in order to facilitate recovery.

Table 4.14.1.3 outlines the main types of sleep problem or disorder associated with various psychiatric conditions. Emphasis has been placed on clinical sleep problems rather than PSG abnormalities which are seen in some of the conditions which (although interesting and of potential pathophysiological significance) are

Table 4.14.1.2 Examples of treatment approaches for sleep disorders in adults

General principles
Explain the problem, reassure where appropriate and provide support
Encourage good sleep hygiene (see Chapter 4.14.2)
Where possible treat the underlying cause of the sleep disturbance (e.g. medical or psychiatric disorder)
Take safety or protective measures (e.g. for hazardous parasomnias)
Specific behavioural treatments for insomnia (Chapter 4.14.2)
Chronotherapy (for circadian sleep-wake rhythm disorders)
Sleep phase retiming
Light therapy
Medication
Hypnotics (very selective and short-term)
Stimulants (narcolepsy)
Melatonin (for some circadian sleep–wake rhythm disorders)
Physical measures (e.g. for obstructive sleep apnoea)
Continuous positive airway pressure (CPAP)
Surgery (selected cases)

Table 4.14.1.3 Clinical sleep disturbance in psychiatric disorders

Psychiatric condition	Likely/possible sleep problem/disorder	Possible treatment issues/principles (see also text for general points)
Depression		*General* Emphasis on treatment of depression Possible sleep complications of ADs: drowsiness, insomnia (including SSRIs), RLS, PLMS, RBD
	Insomnia	Early treatment required to prevent worsening of depression Sedating ADs may be helpful
	EDS (minority)	More stimulating ADs may be appropriate Light therapy for SAD
Mania	Profound insomnia	Vigorous treatment required Mood-stimulating drugs may cause arousal disorders
Anxiety disorders	Insomnia and disrupted (poor quality) sleep PTSD nightmares Nocturnal panic attacks	In addition to behavioural treatments; larger dose of antipsychotics at bedtime may be helpful
Eating disorders	Insomnia (especially anorexia nervosa) EDS (especially bulimia nervosa) Sleepwalking with eating behaviours	General principles for insomnia (see Chapter 4.14.2) Possibly SSRIs See Chapter 4.14.4 for management of sleepwalking
Schizophrenia	Insomnia	Most neuroleptics promote sleep As in other psychotic states, behavioural sleep treatments and sleep hygiene are important
Alcohol and other substance abuse (including withdrawal states)	Insomnia including poor qualitysleep Disrupted sleep-wake cycle	Avoid sedative-hypnotic drugs Sleep hygiene principles (Chapter 4.14) and chronotherapy for sleep–wake cycle abnormalities
Alzheimer's disease and other dementing disorders	Progressive insomnia with fragmented sleep and sleep–wake cycle disorders including nocturnal agitated wandering	General principles for insomnia treatments; Sleep problems with some anti-dementia drugs Low-dose antipsychotics, if necessary Promote day–night cues including regular experience of daylight; discourage daytime napping
	RBD	Clonazepam
Child and adult ADHD	Insomnia	Stimulants can add to the sleep problem

KEY:

ADs = Antidepressants
ADHD = Attention-deficit hyperactivity disorder
EDS = Excessive daytime sleepiness
PLMS = Periodic limb movements in sleep
RBD = REM sleep behaviour disorder
RLS = Restless legs syndrome
SAD = Seasonal affective disorder
SSRI = Selective serotonin reuptake inhibitors

not necessarily accompanied by clinical manifestations. A prime example of this is the reduced time between onset of sleep and the start of the first REM sleep period ('REM latency') in certain forms of severe depression. This can be viewed as a type of biological marker which may persist even after the depression itself has lifted, and which may also be seen in relatives of a depressed person with this PSG finding without they themselves suffering from depression. However, the basic significance of reduced REM latency is, as yet, obscure and not helped by the fact that it has also been described in other psychiatric conditions, narcolepsy, following withdrawal from REM sleep-suppressing substances, during recovery from sleep deprivation, and a number of other diverse circumstances.

Another aspect of the pathophysiological relationship between sleep and mood which awaits clarification is a paradox that, while depression is usually associated with sleep disturbance and loss, sleep deprivation is reported to have an antidepressant effect in some patients although the effect does not persist beyond the period of depression.

Clinical sleep disorders associated with neurological and other medical illness

Of obvious relevance to psychiatric practice in general, but liaison psychiatry in particular, is the fact that disturbed sleep (often severe) is associated with many medical conditions. Main examples of this are shown in Table 4.14.1.4 together with mention of treatment issues in relation to each type of illness. Additional general considerations regarding treatment are as follows:

- The main emphasis in the management of sleep disorders in this context should be placed on the treatment of the medical

Table 4.14.1.4 Clinical sleep disturbance in neurological and other medical illness

Medical condition	Likely/possible sleep problem/disorder	Possible treatment issues/principles (see also text for general points)
Parkinson's disease and/or related syndromes	Progressive insomnia including poor quality sleep	Mainly behavioural treatment and sleep hygiene measures as in other insomnias (see Chapter 4.14.2)
	Parasomnias (e.g. nightmares, nocturnal hallucinations)	Consider possible effects of anti-Parkinson's medication on these and other sleep problems
	EDS including sleep attacks	Possibly daytime stimulants
	PLMS	Usual treatments for PLMS, if severe
	RBD	Clonazepam usually effective
Epilepsy	Disrupted sleep depending on type, cause, and severity	Sleep generally improves with seizure control
	EDS	Sedating AEDs can be a factor
Stroke	Insomnia	Usual measures for insomnias
	OSA	
Head injury	Effects depend on type and severity	
	Poor quality sleep	Avoid respiratory depressant substances
	OSA, EDS	Usual measures for OSA (see Chapter 4.14.3)
Neuromuscular disease	OSA, EDS	As above
Tourette syndrome	Disrupted sleep	Treatment of tic disorder should improve sleep
	Sleepwalking	
Cardiovascular disease		
Congestive heart failure	Central sleep apnoea, orthopnoea	Some antihypertensive, hypolipidaemic and antiarrythmic drugs can cause insomnia
Coronary heart disease	Nocturnal angina	Beta blockers can produce insomnia and nightmares
Respiratory disease		
COPD	Nocturnal dyspnoea	Hypnotics contraindicated in respiratory disease
Asthma	Nocturnal awakenings, EDS	Theophylline can cause insomnia
Gastrointestinal disorders (peptic ulcer, reflux oesophagitis)	Nocturnal awakenings	
Rheumatological disorders and other chronic pain conditions (including cancer)	Insomnia, disrupted sleep, EDS	Cortico-steroids and NSAI agents can disturb sleep Major analgesics have sedative effects Various anti-cancer drugs can disrupt sleep
Iron deficiency	PLMS	Treat underlying condition
	RLS	
Endocrine disease		
Diabetes	Awakenings from nocturnia or painful peripheral neuropathy	
	OSA	
Hyperthyroidism	Insomnia	
Myxoedema	OSA	
Chronic renal failure	Sleep disruption	Improvement with haemodialysis
	EDS	
	OSA	
	RLS and PLMS	
ICU patients	Severe sleep disruption and deprivation including sleep–wake cycle disorders causing confusional and psychotic states	Cause of reason for intensive care relevant, plus sleep disrupting effect of ICU environment, procedures and medications
Obesity	OSA	Complications of obesity (e.g. diabetes, joint diseases) affect sleep

KEY

AED = Anti-epileptic drugs
COPD = Chronic obstructive pulmonary disease
EDS = Excessive daytime sleepiness
ICU = Intensive care unit
NSAI = Non-steroidal anti-inflammatory
OSA = Obstructive sleep apnoea
PLMS = Periodic limb movements in sleep
RBD = REM sleep behaviour disorder
RLS = Restless legs syndrome

condition itself, although additional treatment for the sleep disorder may well be required.

- Treatment for other, co-morbid conditions may well be needed, in particular anxiety and depression which will have their own adverse effects on sleep.
- Hypnotic drugs should be used very sparingly (especially in the elderly) because of their potential complications as described in Chapter 6.2.2. In view of their respiratory-depressand effects, benzodiazepines in particular are contraindicated in sleep apnoea and in the presence of other severe respiratory disease.
- The possible effects on sleep of over-the-counter medications also need to be considered. Nasal decongestants and anorectics are stimulants and can cause insomnia. The same is true of caffeine-containing drinks used as 'energy boosters'. Non-prescribed sleeping preparations (which usually contain anti-histamines) may well cause daytime drowsiness.

As mentioned earlier, the link between medical illness and sleep disturbance may operate in the opposite direction. For example, chronic sleep loss or disruption is associated with the development of such chronic health problems as coronary artery disease or diabetes.

Further information

Kryger, M.H., Roth, T., and Dement, W.C. (eds.) (2005). *Principles and practice of sleep medicine* (4th edn). Elsevier Saunders, Philadelphia, PA.

American Academy of Sleep Medicine. (2005). *The international classification of sleep disorders: diagnostic and coding manual* (2nd edn). American Academy of Sleep Medicine, Westchester, IL.

Colton, H.R. and Altevogt, B.M. (eds.) (2006). *Sleep disorders and sleep deprivation: an unmet public health problem*. National Academies Press, Washington, DC. Available from http.//www.nap.edu/catalog/11617 html

Shapiro, C. and McCall Smith, A. (eds.) (1997). *Forensic aspects of sleep*. Wiley, Chichester.

References

1. Bonnet, M.H. (2005). Acute sleep deprivation. In *Principles and practice of sleep medicine* (4th edn) (eds. M.H. Kryger, T. Roth, and W.C. Dement), pp. 51–6. Saunders, Philadelphia, PA.
2. Dinges, D.F., Rogers, N.L., and Baynard, M.D. (2005). Chronic sleep deprivation. In *Principles and practice of sleep medicine* (4th edn) (eds. M.H. Kryger, T. Roth, and W.C. Dement), pp. 67–76. W.B. Saunders, Philadelphia, PA.
3. Stores, G. (2007). Clinical diagnosis and misdiagnosis of sleep disorders (Review). *Journal of Neurology, Neurosurgery, and Psychiatry*, **78**, 1293–7.
4. Buysse, D.J., Reynolds, C.F., Monk, T.H., *et al.* (1989) The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatric Research*, **28**, 193–213.

4.14.2 Insomnias

Colin A. Espie and Delwyn J. Bartlett

Introduction

Most people's experiences of poor sleep are memorable, because sleeplessness and its daytime consequences are unpleasant. There are those, however, for whom insomnia is the norm. Persistent and severe sleep disturbance affects at least one in 10 adults and one in five older adults, thus representing a considerable public health concern. Sleep disruption is central to a number of medical and psychiatric disorders, and insomnia is usually treated by general practitioners. Therefore differential diagnosis is important, and respiratory physicians, neurologists, psychiatrists, and clinical psychologists need to be involved. The purpose of this chapter is to summarize current understanding of the insomnias, their appraisal, and treatment. Particular emphasis will be placed upon evidence-based practical management.

Clinical features

Insomnia often remains unreported, and finally presents when a poor sleep pattern is well established. Alcohol has long been a first-line self-administered sleep aid, and recent years have seen an increasing use of 'over-the-counter' preparations and 'self-help' strategies. The clinical presentation is commonly of a frustrated patient, trapped in a vicious circle of anxiety and poor sleep, who reports having 'tried everything'.[1,2]

There may be concern about the pattern of sleep. This is the most quantifiable aspect of self-report relating to, for example, length of time taken to fall asleep, frequency and duration of wakenings, or total amount of sleep. A poorly established sleep pattern can lead to unpredictability of what sleep will be like on any given night. Patients often report poor quality of sleep, and subjective perceived quality can be a critically important outcome variable. Typical reports relate to light sleep and sleep felt to be unrestorative. Although it may be unclear how such complaints relate to EEG sleep architecture, the clinician should not overlook qualitative report as it may reflect underlying pathophysiology. Concerns are normally expressed also about the daytime effects of poor sleep. These can be cognitive effects, such as fatigue, sleepiness, inattention, and some impairments in performance, or emotional effects, such as irritability and anxiety.[2]

Classification

The *International Classification of Sleep Disorders* (second edition: **ICSD-2**)[3] was published in 2005 and provides the most comprehensive account of sleep disorders, both for descriptive purposes and for differential diagnosis (see Chapter 4.14.1). ICSD-2 describes insomnias as disorders of initiating and maintaining sleep. Patients may have either sleep-onset problems or wakenings from sleep, or both of these. Table 4.14.2.1 summarizes the principal classifications that relate to the insomnias, along with some other sleep disorders where patients commonly present with insomnia symptoms. As can be seen, concomitant symptomatology, potential aetiological factors, and sleeplessness require careful assessment in order to reach a valid diagnosis.

Table 4.14.2.1 The classification and differential diagnosis of the insomnias within ICSD-2

Classification	Sleep disorder	Essential features, complaint of insomnia plus
Insomnias	Psychophysiological insomnia	Learned sleep preventing associations, conditioned arousal, 'racing mind' phenomenon
	Paradoxical insomnia	Complaint of poor sleep disproportionate to sleep pattern and sleep duration
	Idiopathic insomnia	Insomnia typically begins in childhood or from birth
	Insomnia due to a mental disorder	Course of sleep disturbance concurrent with mental disorder
	Inadequate sleep hygiene	Daily living activities inconsistent with maintaining good-quality sleep
	Insomnia due to a medical disorder	Course of sleep disturbance concurrent with mental disorder
	Insomnia due to drug or substance	Sleep disruption caused by prescription medication, recreational drug, caffeine, alcohol or foodstuff
	Adjustment insomnia	Presence of identifiable stressor, insomnia resolves or is expected to resolve when stressor is removed
Sleep-related breathing disorders	Obstructive sleep apnoea syndrome	Excessive sleepiness, obstructed breathing in sleep, associated symptoms include snoring and a dry mouth
	Periodic limb movement disorder	Episodes of repetitive, highly stereotyped limb movements occurring in sleep
	Restless legs syndrome	Strong, nearly irresistible urge to move legs relieved by walking
Circadian rhythm sleep disorders	Delayed sleep phase type	Phase delay of major sleep episode, initial insomnia, excessive sleepiness in morning
	Advanced sleep-phase type	Phase advance of major sleep episode, inability to stay awake in evening, early wakening

Diagnosis and differential diagnosis

Severity of insomnia is judged along dimensions of frequency, intensity, and duration, as well as impact on daytime functioning and quality of life. Generally, the criteria for severe and chronic insomnia are a minimum duration of 6 months with problems presenting three or more nights per week. Restlessness, irritability, anxiety, daytime fatigue, and tiredness commonly accompany such presentations.[2] Mild and moderate insomnia may be diagnosed where problems are less intrusive.

Most patients presenting with insomnia have psychophysiological difficulty initiating and/or maintaining sleep. Usually marked functional effects and somatized tension associated with sleep are evident. The patient reports extreme tiredness while being unable to sleep satisfactorily and appears preoccupied with sleep and its consequences. This contrasts, for example, with the circadian disorders where, in delayed sleep-phase type, the patient may not feel sleepy until late in the normal sleep period, and in advanced sleep-phase type, may waken early and be unable to return to sleep. Taking a history, incorporating screening questions on restlessness, limb movements, and breathing can help to diagnose obstructive sleep apnoea syndrome, periodic limb movement disorder, and restless legs syndrome, although full polysomnographic evaluation may also be required.[4] However, polysomnography is not essential for the diagnosis of insomnia, for which sleep diary monitoring (see Chapter 4.14.1) is usually the most useful form of assessment.[2] Wrist actigraphy is an inexpensive objective evaluation, which estimates sleep/wakefulness based upon body movement.[5] Continuous recordings can be made over 5 to 10 consecutive 24 h periods. It is useful in identifying paradoxical insomnia, and charted data can be inspected for circadian anomalies.

Other causes of insomnia are reported in Table 4.14.2.1 and should not be overlooked. In particular, insomnias due to a drug or substance can include hypnotic-dependent sleep disorder, associated most commonly with benzodiazepine (BZ) drugs where withdrawal leads to exacerbation of the primary problem.[6] This can be mistaken for a severe underlying insomnia and hence reinforce hypnotic dependency. Likewise, a wide range of psychiatric conditions, particularly affective disorders, has associated sleep symptomatology (see Chapter 4.14.1). A primary diagnosis of psychophysiological insomnia cannot be made where diagnostic criteria for DSM-IV Axis I or Axis II disorders are fulfilled. However, it is very important to note that sleep disturbance often precedes depression. The bulk of the psychiatric epidemiological data indicate that insomnia is an independent risk factor for first episode depressive illness, and for recurrence of depression, in adults of all ages.[7,8] Insomnia should not be assumed to be simply a symptom of underlying depression, even when depression is present. Unless the illness courses clearly co-vary it is best to make a diagnosis of co-morbid insomnia. Similar caveats apply to insomnia associated with medical disorders, both in terms of identifying a primary illness, and concluding that insomnia has *the* status of an associated/ co-morbid disorder (see Chapter 4.14.1).

Epidemiology

Insomnia affects one-third of adults occasionally, and 9 to 12 per cent on a chronic basis. It is more common in women, in shift workers, and in patients with medical and psychiatric disorders. Prevalence amongst older adults has been estimated at up to 25 per cent and sleepiness and hypnotic drugs are risk factors for injury and fracture.[9] The decline in prescription of anxiolytics has been greater than the rate of decline for hypnotics [taking BZ and benzodiazepine receptor agonists (BzRAs) together]. Furthermore, there is increasing use of (off-label) sedative antidepressants primarily to treat insomnia.

Aetiology

Many patients report having been marginal light sleepers before developing insomnia.[4] Sleep disturbance often arises during life change or stress, and such adjustment sleep disorder may represent a normal transient disruption of sleep. However, secondary factors, such as anxiety over sleep and faulty sleep–wake conditioning, may

exacerbate and maintain the insomnia as a chronic problem when sleep itself becomes a focus for concern. People with insomnia may be hyperaroused relative to normal sleepers, for example having higher levels of cortisol and ACTH, and also find it difficult to 'down-regulate' their arousal at bedtime.[2,10,11]

Course and prognosis

There has been little research on the natural course of insomnia. However, untreated psychophysiological insomnia can last for decades, and may gradually worsen over time. Indeed, there is a developmental trend for sleep pattern to deteriorate, with increasing age. On the other hand, delayed sleep-phase syndrome and insufficient sleep hygiene can be associated with lifestyle problems and may ameliorate as these are resolved. Although certain insomnias *tend* to persist if untreated, prognosis with effective treatment can be very good.

Treatment

A review of the evidence

(a) Drug therapy

Traditionally, insomnia has been treated pharmacologically. Barbiturates were superseded by BZ compounds during the 1960s and 1970s. These drugs were safer in overdose, were thought to have fewer side effects, and to be less *addictive*. Controlled studies have demonstrated that a considerable number of BZ, of short to intermediate half-life, are effective hypnotic agents. However, from the mid-1970s potential problems became apparent, both during administration and withdrawal. Longer-acting hypnotics were prone to carry-over effects of morning lethargy, and shorter-acting drugs to 'rebound insomnia'.[6] Furthermore, tolerance develops, leading either to increased dosing or switching to alternative medication. Although BZs used for short periods/intermittently can maintain effectiveness, these are not the treatment of choice in chronic insomnia,[12] and are contraindicated in older adults and where insomnia may involve sleep-related breathing disorder because of their potentially depressant effects on respiration. A number of BZ compounds have been removed from the market in the United Kingdom, United States, and elsewhere.

Contemporary hypnotic therapy has extended to include BzRAs (often referred to as the 'z' drugs), and more recently melatonin receptor agonists (MeRAs) have been introduced. Whereas the place in therapeutics of MeRAs has yet to become established, the BzRAs are often thought to offer more sustained benefit for insomnia, and to have fewer adverse effects. Nevertheless, there remains uncertainty about the effectiveness of BzRAs in chronic insomnia.[13]

(b) Psychological therapy

Psychological treatment for chronic insomnia, primarily in the form of cognitive behavioural therapy (CBT), has been extensively investigated in over 100 controlled studies during the past 20 years. Five meta-analyses and numerous systematic reviews have demonstrated that CBT is associated with large effect size changes (measured in standardized z scores) in the primary symptom measures of sleep latency (difficulty getting to sleep) and wake time after sleep-onset (difficulty remaining asleep).[14,15] Around 70 per cent of patients with persistent sleep problems appear to benefit from CBT and effects are maintained to long-term follow-up. It is thought that CBT achieves these outcomes because it tackles directly the dysfunctional thoughts and maladaptive behaviours that otherwise maintain insomnia. Recent controlled studies have shown that CBT may be effective in general practice settings with nurses delivering the intervention according to a standard protocol.[16,17] Despite the superior efficacy of CBT relative to medication for insomnia, and these recent demonstrations of CBT working in real-world settings, practical problems remain in making CBT widely available.

Within the CBT model, a number of strategies have strong empirical support. Behavioural procedures such as stimulus control and sleep restriction, and cognitive strategies such as paradoxical intention and thought restructuring have been extensively investigated[2,14,15] and are outlined briefly below.

(c) Melatonin, light therapy, and exercise

The pineal hormone melatonin has been the subject of highly publicized claims. However, scientific research has been limited. Several controlled studies support its sleep-promoting effects, but the use of melatonin continues to be controversial. At best it may be useful as a chronobiotic for reducing sleep latency.[18] Several MeRA products are currently under formal evaluation, so more data may be available soon.

Bright light is a potent marker for human circadian rhythms, and has been known for some time to enable the resetting of such rhythms in advanced sleep-phase syndrome and delayed sleep-phase syndrome.[19] The results of studies investigating the efficacy of bright light against psychological treatments for psychophysiological insomnia are awaited. A limiting factor to the value of light therapy is that continued treatment may be required to maintain therapeutic effects.

Athletic people sleep well, although this may be more to do with behavioural patterning than aerobic fitness. Nevertheless, there is evidence that exercise can have positive effects upon sleep quality, particularly if taken late afternoon or early evening, and in otherwise relatively fit individuals.[20] Morning exercise can also be an effective modality to encourage the same waking time and early morning light exposure; which help to reset sleep patterns on a daily basis.

Advice about management

(a) General perspective

Non-pharmacological treatment using CBT procedures should be preferred over pharmacological treatment, in cases of severe persistent insomnia. Hypnotic agents should be recommended mainly for short-term or occasional use, although longer-term trial data are now becoming available. The practitioner should be aware of morning-after effects, and potential problems of withdrawal and dependency, not only with BZs but also possibly with BzRAs. Psychological intervention may also facilitate reduction or discontinuation of medication in hypnotic-dependent person with insomnias.[21] There is limited support for the use of melatonin or exercise as treatments of choice, although light therapy seems effective for circadian disorders.

(b) Using cognitive behavioural therapies

Brief descriptions of effective management strategies are presented in Tables 4.14.2.2 and 4.14.2.3. The following text provides

Table 4.14.2.2 Summary description of sleep hygiene and education components for the treatment of chronic insomnia

Components of sleep education
The need for sleep and its functions
Sleep patterns across the lifespan
Sleep as a process with stages/phases
Factors adversely affecting steep
The effects of sleep loss
The concept of insomnia
Measuring sleep and sleep problems
Components of sleep hygiene treatment
Bedroom comfortable for sleep
Regular exercise, timing, and fitness
Stable and appropriate diet
Undesirable effects of caffeine and other stimulants
Moderation of alcohol consumption
Other common 'self-help' strategies

explanation of underlying psychological models and further information on implementation.

(i) Sleep education and sleep hygiene

The simple provision of information ameliorates the sense of being out of control. Inaccurate attributions are challenged and misunderstandings corrected by understanding what sleep is, how ommon insomnia can be, how sleep changes with age, good sleep hygiene practices, and some facts about sleep loss. Similarly, sleep hygiene provides patients with a starting point for self-management. These techniques are best construed as introductory but they will not of themselves treat insomnia effectively.

Table 4.14.2.3 Summary description of cognitive behavioural components for the treatment of chronic insomnia

Components of stimulus control and sleep restriction treatment
Define individual sleep requirements
Establish parameters for bedtime period (threshold time and rising time)
Eliminate daytime napping
Differentiate rest from sleep
Schedule sleep periods with respect to needs
Establish 7 day per week compliance
Remove incompatible activity from bedroom environment
Rise from bed if wakeful (>20 min)
Avoid recovery sleep as 'compensation'
Establish stability from night to night
Adjust the sleep period as sleep efficiency improves
Components of cognitive intervention
Identify thought patterns and content that intrude
Address (mis)attributions connecting sleep and waking life
Establish rehearsal/planning time in early evening
Relaxation and imagery training
Distraction and thought blocking
Develop accurate beliefs/attributions about sleep and sleep loss
Challenge negative and invalid thoughts
Eliminate 'effort' to control sleep
Motivate to maintain behaviour and cognitive change
Utilize relapse-prevention techniques

(ii) Stimulus control treatment

Stimulus control increases the bedroom's cueing potential for sleep. For good sleepers, the pre-bedtime period and the stimulus environment trigger positive associations of sleepiness and sleep. For the poor sleeper, however, the bedroom triggers associations with restlessness and lengthy night-time wakening via a stimulus-response relationship, thereby continuing to promote wakefulness and arousal. The model is similar to phobic conditions where a conditioned stimulus precipitates an anxiety response.

Treatment involves removing from the bedroom all stimuli which are potentially sleep-incompatible. Reading and watching television, for example, are confined to living rooms. Sleeping is excluded from living areas and from daytime, and wakefulness is excluded from the bedroom. The individual is instructed to get up if not asleep within 15–20 min or if wakeful during the night. Conceptually, stimulus control is a reconditioning treatment which forces discrimination between daytime and sleeping environments.

(iii) Sleep restriction therapy

Sleep restriction restricts sleep to the length of time which the person is likely to sleep. This may be equivalent to promoting 'core sleep' at the expense of 'optional sleep'. Sleep restriction primarily aims to improve sleep efficiency. Since sleep efficiency is the ratio of time asleep to time in bed, it can be improved either by increasing the numerator (time spent asleep) or by reducing the denominator (time spent in bed). People with insomnia generally seek the former, but this may not be necessary, either biologically or psychologically. Sleep restriction first involves recording in a sleep diary and calculating average nightly sleep duration. The aim, then, is to obtain this average each night. This is achieved by setting rising time as an 'anchor' each day and delaying going to bed until a 'threshold time' which permits this designated amount of sleep. Thus, the sleep period is compressed and sleep efficiency is likely to increase. The permitted 'sleep window' can then be titrated week-by-week in 15 increments in response to sleep efficiency improvements.

(iv) Cognitive control

This technique aims to deal with thought material in advance of bedtime and to reduce intrusive bedtime thinking. The person with insomnia is asked to set aside 15 to 20 min in the early evening to rehearse the day and to plan ahead for tomorrow; thus putting the day to rest. It is a technique for dealing with unfinished business and may be most effective for rehearsal, planning, and self-evaluative thoughts which are important to the individual and which, if not dealt with, may intrude during the sleep-onset period.

(v) Thought suppression

Thought-stopping and articulatory suppression attempt to interrupt the flow of thoughts. No attempt is made to deal with thought material *per se* but rather to attenuate thinking. With articulatory suppression the patient is instructed to repeat, subvocally, the word 'the' every 3 s. This procedure is derived from the experimental psychology literature. Articulatory suppression is thought to occupy the short-term memory store used in the processing of

information. The type of material most likely to respond is repetitive but non-affect-laden thoughts, not powerful enough to demand attention. Additionally, this technique may be useful during the night to enable rapid return to sleep.

(vi) Imagery and relaxation

There is a wide range of relaxation methods including progressive relaxation, imagery training, biofeedback, meditation, hypnosis, and autogenic training, but little evidence to indicate superiority of any one approach. Furthermore, there is little evidence to support either the presumption that people with insomnia are hyperaroused in physiological terms, or that relaxation has its effect through autonomic change. At the cognitive level, these techniques may act through distraction and the promotion of mastery. During relaxation, the mind focuses upon alternative themes such as visualized images or physiological responses. In meditation the focus is upon a 'mantra' and in self-hypnosis upon positive self-statements. Relaxation may be effective for thought processes that are anxiety-based, confused, and which flit from topic-to-topic.

(vii) Cognitive restructuring

Cognitive restructuring challenges faulty beliefs which maintain wakefulness and the helplessness which many people with insomnia report. It appears to work through appraisal by testing the validity of assumptions against evidence and real-life experience. As an evaluative technique, it may be effective with beliefs that are irrational but compelling. If such thoughts, for example 'I am going to be incapable at work tomorrow', are not challenged, they will create high levels of preoccupation and anxiety and sleep is unlikely to occur. With cognitive restructuring, the person with insomnia learns alternative responses to replace inaccurate thinking.

(viii) Paradoxical intention

Finally, the technique of paradoxical intention is useful in situations where performance anxiety has developed, that is, where the effort to produce a response inhibits that response itself. The paradoxical instruction is to allow sleep to occur naturally through passively attempting to remain quietly wakeful rather than attempting to fall asleep. Paradox may be regarded as a decatastrophizing technique since it appears to act upon the ultimate anxious thought (of remaining awake indefinitely) initially by focusing on and enhancing this thought (a habituation model) and then subjecting it to appraisal through rationalization and experience. By intending to remain awake, and failing to do so, the strength of the sleep drive is re-established, and performance effort is reduced.

Possibilities for prevention

There is insufficient knowledge of the natural course of transient sleep disorders. Mention has been made of adjustment sleep disorder and of the association of life events and stressors with the onset of insomnia. Systematic research is required to establish the 'setting conditions' for the secondary maintenance of insomnia beyond an initial normative reaction to events. Perhaps there is an interaction with a predisposing tendency to light sleep, or with introspection and worry. The instinct to increase opportunity to sleep (spend longer in bed to catch up) when insomnia symptoms develop should probably be resisted. If anything it may be better to advise patients to limit sleep opportunity so that their pattern knits together again more quickly.

The establishment and maintenance of a regular 'tight' routine, both pre-bedtime and in terms of sleep schedule, seem to be important preventive factors. Such chronobehavioural functioning can be at risk of disruption by, for example, jet lag, shift work, weekend patterns differing from weekday, adolescent lifestyle, and retirement. Adherence to, and/or reinstatement of, an adaptive pattern seems crucial.

It is important not to underestimate the importance of attitudes and beliefs in the presentation of insomnia. Exaggerated or emotionally and mentally arousing thoughts should be dealt with promptly. Sleep loss can be distressing, but patients should be reminded that nature seeks to restore equilibrium. What they need to do is to provide the conditions under which sleep can occur rather than attempt directly to control the sleep process. Expectations are important also, since it is the breach of these which generally give rise to anxiety and dysfunctional beliefs about sleep requirements. More often than not sleep-related expectations are unrealistic and require reappraisal, even more so in older adults.

Finally, prevention should be extended to the known extrinsic causes of certain sleep disorders. Where alcohol, stimulants, or proprietary drugs interfere with sleep and the recovery of the normal sleep process, attention should be paid to these factors. Better still, patients should be encouraged to seek advice early rather than go down the path of self-administered treatment. Avoiding the use of hypnotic agents, both in general practice and during acute admissions to hospital, would substantially reduce the number of iatrogenic cases of chronic insomnia.

Further information

Espie, C.A. (2006). *Overcoming insomnia and sleep problems: a self-help guide using cognitive behavioral techniques*. Constable & Robinson Ltd, London [ISBN13: 978-184529-070-2/ ISBN10: 1-84529-070-4].

Perlis, M.L. and Lichstein, K.L. (2003). *Treating sleep disorders: principles and practice of behavioral sleep medicine*. John Wiley & Sons, Inc. Hoboken, NJ [ISBN0-47-44343-3].

Morin, C.M., Bootzin, R.R., Buysse, D.J., *et al.* (2006). Psychological and behavioural treatment of insomnia. Update of the recent evidence (1998–2004) prepared by a Task Force of the American Academy of Sleep Medicine. *Sleep*, **29**, 1398–414.

Morin, C.M. and Espie, C.A. (2003). *Insomnia: a clinical guide to assessment and treatment*. Kluwer Academic/ Plenum Publishers, New York [ISBN 0-306-47750-5].

References

1. Espie, C.A. (1991). *The psychological treatment of insomnia*. Wiley, Chichester.
2. Morin, C.M. and Espie, C.A. (2003). *Insomnia: a clinical guide to assessment and treatment*. Kluwer Academic/Plenum Publishers, New York.
3. American Academy of Sleep Medicine. (2005). *International classification of sleep disorders: diagnostic and coding manual* (2nd edn). AASM, Westchester, IL.
4. Reite, M., Buysse, D., Reynolds, C., *et al.* (1995). The use of polysomnography in the evaluation of insomnia. *Sleep*, **18**, 58–70.
5. Ancoli-Israel, S., Cole, R., Alessi, C., *et al.* (2003). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*, **26**, 342–92.
6. Kripke, D. (2000). Hypnotic drugs: deadly risks, doubtful benefits. *Sleep Medicine Reviews*, 2000, **4**, 5–20.

7. Cole, M.G. and Dendukuri, N. (2003). Risk factors for depression among elderly community subjects: a systematic review and meta-analysis. *The American Journal of Psychiatry*, **160**, 1147–56.
8. Riemann, D. and Voderholzer, U. (2003). Primary insomnia: a risk factor to develop depression? *Journal of Affective Disorders*, **76**, 255–9.
9. Lichstein, K.L., Durrence, H.H., Reidel, B.W., *et al.* (2004). *The epidemiology of sleep: age, gender and ethnicity*. Lawrence Erlbaum Associates, Mahwah, NJ.
10. Espie, C.A. (2002). Insomnia: conceptual issues in the development, persistence and treatment of sleep disorder in adults. *Annual Review of Psychology*, **53**, 215–43.
11. Perlis, M.L., Pigeon, W., and Smith, M.T. (2005). Etiology and pathophysiology of insomnia. In *The principles and practice of sleep medicine* (4th edn) (eds. M.H. Kryger, T. Roth, and W.C. Dement), pp. 714–25. W.B. Saunders, Philadelphia.
12. NIH. (2005). *State-of-the-science conference statement on manifestations and management of chronic insomnia in adults*. National Institutes of Health, Vol. 22, Number 2. Natcher Conference Center, Bethesda Maryland, USA.
13. NICE. (2004). *Guidance on the use of zaleplon, zolpidem and zopiclone for the short-term management of insomnia*. Technology Appraisal Guidance No.77. National Institute for Clinical Excellence, London.
14. Smith, M.T., Perlis, M.L., Park, A., *et al.* (2002). Behavioral treatment vs pharmacotherapy for insomnia—a comparative meta-analysis. *The American Journal of Psychiatry*, **159**, 5–11.
15. Morin, C.M., Bootzin, R.R., Buysse, D.J., *et al.* (2006). Psychological and behavioural treatment of insomnia. Update of the recent evidence (1998–2004) prepared by a Task Force of the American Academy of Sleep Medicine. *Sleep*, **29**, 1398–414.
16. Espie, C.A., Inglis, S.J., Tessier, S., *et al.* (2001). The clinical effectiveness of cognitive behaviour therapy for chronic insomnia: implementation and evaluation of a Sleep Clinic in general medical practice. *Behaviour Research and Therapy*, **39**, 45–60.
17. Espie, C.A., MacMahon, K.M.A., and Kelly, H.L. (2007). Randomised clinical effectiveness trial of nurse-administered small group CBT for persistent insomnia in general practice. *Sleep*, **30**, 574–84.
18. Mendelson, W.B. (1997). A critical evaluation of the hypnotic efficacy of melatonin. *Sleep*, **20**, 916–19.
19. Czeisler, C.A., Johnson, M.P., Duffy, J.F., *et al.* (1990). Exposure to bright light and darkness to treat physiologic maladaptation to night work. *The New England Journal of Medicine*, **322**, 1253–8.
20. Singh, N.A., Clements, K.M., and Fiatarone, M.A. (1997). A randomized controlled trial of the effect of exercise on sleep. *Sleep*, **20**, 95–101.
21. Morgan, K., Dixon, S., Mathers, M., *et al.* (2003). Psychological treatment for insomnia in the management of long-term hypnotic drug use: a pragmatic randomised controlled trial. *The British Journal of General Practice*, **53**, 923–8.

4.14.3 Excessive sleepiness

Michel Billiard

Introduction

Excessive sleepiness is not an homogeneous concept. It can manifest itself as bouts of sleepiness, irresistible and refreshing sleep episodes, abnormal lengthening of night sleep with a major difficulty waking up in the morning or at the end of a nap or even periods of a week or so of almost continuous sleep recurring at several months' intervals.

According to the recent second edition of the *International Classification of Sleep Disorders* (*ICSD*-2),[1] disorders of excessive sleepiness are distributed within three chapters: sleep-related breathing disorders, hypersomnias of central origin not due to a circadian rhythm sleep disorder, sleep-related breathing disorders, or other cause of disturbed nocturnal sleep, and circadian rhythm sleep disorders.

However in this volume aimed at psychiatrists, the presentation of disorders of excessive sleepiness will obey another logic. Following "Generalities" including epidemiology, morbidity, clinical work-up, and laboratory tests, the various aetiologies will be presented according to the following six subchapters:

- Hypersomnia not due to substance or known physiological condition (non-organic hypersomnia or psychiatric hypersomnia)
- Hypersomnia due to drug or substance
- Behaviourally induced insufficient sleep syndrome
- Hypersomnia in the context of sleep-related breathing disorders
- Hypersomnias of central origin
- And the special case of delayed sleep phase syndrome.

Epidemiology

Contrary to common thinking, excessive sleepiness is neither exceptional nor rare. Epidemiological surveys generally agree on a figure of severe sleepiness (daily and embarrassing) in 5 per cent of the general population and of moderate sleepiness (occasional) in another 15 per cent.[2] Interestingly, only a fraction of these subjects are aware of their condition, due to the fact that they progressively lose reference to a normal state of alertness. As a consequence many subjects will not consult their physician for excessive sleepiness but will be brought to him by the spouse, worried about his or her falling asleep repeatedly in the middle of the day, or even referred by the company's doctor due to unexplained car accidents or poor work efficiency.

Morbidity

Excessive sleepiness has a severe impact on the life of patients. Nearly half of the patients with excessive sleepiness report automobile accidents. Many have lost jobs because of their sleepiness. In addition, sleepiness is disruptive of family life. Cognitive function is also impaired by sleepiness. In children excessive sleepiness has been associated with learning disability and in adults memory problems are frequent.

Clinical work-up and laboratory tests

Whatever the circumstance of the first visit, the patient should be interviewed on the history of excessive sleepiness, the type and severity of it, the associated symptoms, the familial and occupational consequences, the past and current treatments, and the personal and familial medical past-history.

In addition, the subject will complete a self-administered behavioural scale, the Epworth sleepiness scale. This scale asks the subject to rate the probability of dozing from 0 (would never doze) to

3 (high chance of dozing) in eight more or less soporific daily situations. A score of over 10 is taken to indicate abnormal sleepiness.

The subject will then undergo a physical and psychological examination.

Laboratory tests will be chosen according to the clinical impression.

The most frequently used test is the *multiple sleep latency test* (MSLT). The test was developed on the basis of the following principle. The sleepier the subject, the faster he falls asleep. The test is based on 20 min polygraphic recordings (EEG, EOG, EMG) repeated every 2 h (four or five times a day) starting 2 h about after morning awakening. The global sleepiness index is provided by the mean latency to sleep in the four or five tests. A sleep laboratory of less than 5 min indicates pathologic sleepiness, a sleep latency from 10 to 20 min is considerd as normal, and latencies falling between the normal and the pathological values are considered as a diagnostic grey area.

Another test, the *maintenance of wakefulness test* (MWT), is a variant of the MSLT. It was designed to evaluate treatment efficiency in patients with excessive sleepiness. The major difference with the MSLT is in the instruction given to the test subject. The subject being tested is told to attempt to remain awake. The subject is seated in comfortable position in bed, as opposed to lying down in the MSLT, with low lighting behind him (7.5 W at 1 m). Specific recommendations include using a four-trial, 40 min version of the MWT. A mean sleep latency of less than 8 min on the 40 min MWT is abnormal and scores between 8 and 40 min are of uncertain significance.

Prolonged polysomnographic recordings, obtained by either traditional laboratory polysomnographic monitoring or ambulatory recordings, provide a good picture of the actual time asleep within the 24 h period. However, this procedure is neither validated nor standardized.

In addition, whenever there is some doubt about the possibility of hypersomnia associated with a psychiatric disorder, a *psychometric/psychiatric evaluation* will be performed.

Aetiology and treatment

Hypersomnia not due to substance or known physiological condition

It explains about 5 to 7 per cent of cases of hypersomnia seen in sleep disorders centres. Women are more susceptible than men.

Excessive daytime sleepiness is reported. Subjects show an elevated score on the Epworth sleepiness scale. Night sleep is perceived as non-restorative and generally of poor quality. Patients are often intensely focused on their hypersomnia, and psychiatric symptoms typically become apparent only after prolonged interview or psychometric testing. Poor work attendance, abruptly leaving work because of a perceived need to sleep are common. Polysomnography typically shows a prolonged sleep latency, an increased wake time after sleep onset, and a low sleep efficiency. REM latency may be shortened in the case of bipolar disorder. Contrasting with the elevated score on the Epworth sleepiness scale, sleep latency on the MSLT is often within normal limits. A 24 h continuous sleep recording typically shows considerable time spent in bed during day and night, a behaviour referred to as clinophilia, from the Greek κλινη (bed) and φιλεω (love).

Psychiatric interview is essential to diagnose the underlying condition. Causative psychiatric conditions include bipolar type II disorder, dysthymic disorder, undifferentiated somatoform disorder, adjustment disorder, or personality disorder.

Conventional drugs such as antidepressants or anxiolytics are often insufficient. Modafinil, an awakening drug given at a daily dose of 100 to 200 mg, is usually active.

In the group of psychiatric disorders a separate place should be reserved to seasonal affective disorder remarkable for episodes of major depression occurring only during the winter months, associated with fatigue, loss of concentration, increased appetite for carbohydrates, weight gain, and increased sleep duration. Morning bright light treatment (2500 lux for 2 h) is efficient.

Hypersomnia due to drug or substance

A wide spectrum of medications used in psychiatry may be responsible for excessive sleepiness.

(a) Anxiolytics and hypnotics

Benzodiazepines have sedative effects, but these effects vary with dose, administration (single or repeated dose), age, and state of the subject (normal, anxious, or depressed). Non-benzodiazepines usually induce limited sleepiness only.

(b) Antidepressants

Tricyclic antidepressants have sedative properties depending on the molecule, dose, and the subject to whom they are administered. SSRI can also induce sleepiness with high within-patient variability. Venlafaxin, a serotonine, and norepinephrine reuptake blocker may induce excessive sleepiness.

(c) Neuroleptics

The degree of sedation varies widely from subject to subject. Empirically, three-fourth of the patients treated with neuroleptic phenothiazines experience sleepiness in a dependent manner. Among the newer agents clozapine is the most sedating drug, followed by olanzapine and quetiapine. Risperidone and sertindole are less sedating drugs.

Behaviourally induced insufficient sleep syndrome

According to ICSD-2, this syndrome occurs when an individual persistently fails to obtain the amount of sleep required to maintain normal levels of alertness. Behaviourally induced insufficient sleep syndrome is likely the most common cause of daytime sleepiness. In a population-based study conducted in Japan among 3030 subjects aged 20 years and older, 29 per cent slept less than 6 h per night, and 23 per cent reported having insufficient sleep.[3] The syndrome is likely to be widespread in truck drivers, working mothers, family doctors, executives, and students. The main symptoms are excessive sleepiness in the afternoon or early evening, decrease of diurnal performances, and, of interest to the psychiatrist, irritability, nervousness, and depression. Diagnosis of the syndrome is relatively easy provided that a thorough interview is conducted. The most rational treatment is an increase of daily total sleep time, either by spending more time in bed at night, or by taking one or two naps per day.

Hypersomnia in the context of sleep-related breathing disorders

The most frequent condition among these disorders is the obstructive sleep apnoea syndrome. This syndrome was first described by

Guilleminault *et al.*[4] It is most frequent in 50-year-old males. According to Young *et al.*[5] the prevalence of obstructive sleep apnoeas accompanied by excessive daytime sleepiness in North America is 4 per cent in men and 2 per cent in women.

Clinical features include night-time and daytime symptoms. Night-time symptoms are represented by loud snoring, apnoeic episodes ending with sonorous breathing resumption, nocturia, severe fatigue upon awakening, and sometimes headache. Daytime symptoms are dominated by excessive sleepiness, which varies in intensity among patients. Other symptoms include irritability, negligence, loss of concentration, loss of libido, impotence, and sometimes depression.

Patients are often obese or mildly obese. High blood pressure is a frequent feature. The ear, nose, and throat examination usually reveals a narrow upper airway due to close-set posterior tonsillar pillars, an abnormally long and hypotonic soft palate, a hypertrophic uvula, or macroglossia.

The positive diagnosis rests on polysomnography allowing the observation of nocturnal disrupted sleep and the identification of apnoeas and of their type (obstructive, central, or mixed) as well as their consequences on heart rate, oxygen desaturation, and degree of somnolence.

Of note, some subjects do not have apnoeas or hypopnoeas but increasing respiratory effort resulting in respiratory effort-related arousals (RERAs) and are believed to be as much at risk for complications. This state is most accurately identified with a quantative measurement of airflow and oesophageal manometry.

Obstructive sleep apnoea patients are at risk for systemic hypertension, occasional arrhythmias and conduction disturbances, cardiac or cerebral ischaemia, functional cognitive impairment, and depression. In a multicentre telephone survey carried-out in 1994–1999 in five European countries, among 18 980 subjects aged 15 to 100 years, 18 per cent of the individuals receiving a diagnosis of major depression also had a sleep-related breathing disorder and 17.6 per cent of the individuals receiving a diagnosis of sleep-related breathing disorder also had a diagnosis of major depression.[6]

There are several possible approaches to treatment:

- Weight loss may be beneficial. Avoidance of alcohol and sedatives should be recommended in all cases.
- Continuous positive airway pressure (CPAP) at night is the most widely used treatment. Good compliance requires proper preparation for the patient and an adaptation period.
- Oral appliances are suitable in mild obstructive sleep apnoea syndrome.
- Surgery consists mainly in nasal reconstruction in case of symptomatic airway blockade caused by bony, cartilaginous, or hypertrophied tissues that interfere with nasal breathing during sleep.

Hypersomnias of central origin

(a) Narcolepsy

First described in 1877[7] and given its name by Gelineau in 1880,[8] narcolepsy is now distinguished into three entities, narcolepsy with cataplexy, narcolepsy without cataplexy, and narcolepsy due to medical condition.

(i) Narcolepsy with cataplexy

Narcolepsy with cataplexy is not an exceptional condition. According to most recent evaluations its prevalence is 0.20 to 0.40 per 1000, i.e. slightly less than the prevalence of multiple sleep sclerosis.

Narcolepsy with cataplexy is characterized by two cardinal symptoms, excessive daytime sleepiness/irresistible sleep episodes and cataplexy, and other clinical features that are not necessarily part of the actual clinical picture.

Excessive sleepiness occurs daily. It comes in waves of varying degree of severity, depending on the individual, building up into irresistible and refreshing episodes of sleep. Excessive sleepiness is brought on by passive situations such as watching TV, being a passenger in a car, or attending a lecture. However it can also awkwardly occur in unexpected situations such as eating, walking, swimming, or driving a car. Excessive sleepiness may lead to automatic behaviour, such as saying totally inappropriate words or sentences, arranging objects in unlikely places, or driving a vehicle to an unintended destination.

Cataplexy is the most specific symptom. It consists of a sudden bilateral loss of voluntary muscle tone with preserved consciousness. It is triggered by environmental factors which are usually positive, such as a fit of laughter, receiving a compliment, humour expressed by the subject himself, the sight of prey for the hunter, the perception of a fish biting at the hook for the angler, a well-caught ball at tennis, or by anger, but almost never by stress or fear. All the striated muscles may be affected except the extra-ocular and respiratory musculature, leading to the progressive slackening of the whole body. More often however the attack is partial, involving certain muscles only, for example the jaw muscles, producing sudden difficulty in articulating words; the facial muscles, responsible for a grimace; or the thigh muscles, causing a brief unlocking of the knees. Consciousness is maintained during the episode. Cataplexy varies in duration from a split second to several minutes. Attack frequency varies from only a few per year, or even less, to several per day. In rare cases, especially after abrupt withdrawal of antidepressant medication, episodes of cataplexy may repeat themselves for minutes or hours, a state referred to as "status cataplecticus".

The other clinical features are deemed accessory to the extent that they are not indispensable for diagnosis.

Hallucinations, whether hypnagogic (at the onset of sleep) or hypnopompic (on awakening), are vivid perceptual experiences, not dreams, either visual, or auditory or kinetic. Patients may perceive someone entering and walking in the room, find themselves flying through the air or falling from a skyscraper. In some cases of unrecognized narcolepsy with daytime hypnagogic/hypnopompic hallucinations, the patient may be mistakenly diagnosed as having a delusional psychosis.

Sleep paralysis is a sudden inability to move during the transition from sleep to wakefulness or vice versa, while the subject is conscious. It is often accompanied by hypnagogic hallucinations. It is very unpleasant. In narcoleptic patients sleep paralysis may last up to 10 min.

Nocturnal sleep disruption occurs in approximately 50 per cent of cases, depending on age and the time elapsed since the onset of condition. The patient typically falls asleep as soon as he gets to bed, but his sleep is disturbed by recurring awakenings. He may complain of disturbing dreams.

REM-sleep behaviour disorder is frequent, either as a mere polysomnographic finding, excess of muscle tone or phasic EMG twitching activity during REM sleep, or as clinically significant complaint manifesting itself as an attempted enactment of distinctly altered, unpleasant, action-filled, and violent dreams in which the individual is being confronted, attacked, or chased by unfamiliar people. Periodic limb movements in sleep are more frequent than in normal subjects.

Physical examination is normal except for a frequently increased body mass index, especially in children at the onset of the condition. Noteworthy is the abolition of deep tendon reflexes during a cataplectic attack.

In a majority of cases, excessive daytime sleepiness is the first symptom to appear. In some patients, cataplexy occurs at the same time as excessive daytime sleepiness, but more often is delayed by one or several years. It is extremely rare for cataplexy to be the first clinical manifestation of narcolepsy with cataplexy.

The clinical diagnosis is based on clinical features. However, additional tests are highly recommended to confirm the diagnosis. The first test is polysomnography followed by an MSLT. At night a sleep onset REM period (SOREMP) is highly specific. It is observed in 25 to 50 per cent of cases. An increase in the amount of stage 1 NREM sleep and repeated awakenings are frequent findings. The MSLT shows a mean sleep latency of less than 8 min and two or more SOREMPs. However, some patients, especially in middle or old age, with clear-cut excessive daytime sleepiness and cataplexy, may have only one SOREMP, or even none during the MSLT procedure.

Today a CSF level of hypocretin-1 below 110 pg/mL is recognized as highly specific and sensitive for narcolepsy with cataplexy. However, up to 10 per cent of narcolepsy with cataplexy patients have normal levels of hypocretin-1.

There is no set pattern of evolution across patients. Excessive daytime sleepiness and irresistible episodes of sleep persist throughout life, but may diminish with age in some subjects. Cataplexy may spontaneously diminish with time or even totally disappear in some patients.

Narcolepsy is a very incapacitating disease. It interferes with every aspect of life. Education, performance at school and workplace, driving capability, recreational activities, interpersonal relationships, sexual life, and self-esteem. Depression is a frequent consequence of narcolepsy.

The mainstay of treatment is pharmacological although taking naps alleviates excessive sleepiness and refraining from emotion may prevent cataplexy.

Modafinil is the first-line treatment of excessive daytime sleepiness and irresistible episodes of sleep. Methylphenidate is less used today. Based on several large randomized controlled trials showing the activity of sodium oxybate on excessive daytime sleepiness and irresistible episodes of sleep, there is a growing practice in the United States to use it for the later indications. However, adverse effects including nausea, nocturnal enuresis, confusional arousals, malaise, and headache are not exceptional. Sodium oxybate is the only registered treatment for cataplexy. Other treatments are antidepressants, including tricyclics, selective serotonion reuptake inhibitors (SSRIS) and more recently new antidepressants such as venlafaxine or atomoxetine, the later being increasingly used despite few or non-randomized placebo-controlled clinical trials. As for disturbed nocturnal sleep the most widely used treatment is still hypnotics. However, the same randomized controlled trials have shown the activity of sodium oxybate against disturbed nocturnal sleep.

(ii) Narcolepsy without cataplexy

Narcolepsy without cataplexy is described as excessive daytime sleepiness and irresistible episodes of sleep unaccompanied by cataplexy. Automatic behaviour may be present as may hypnagogic hallucinations or sleep paralysis. Nocturnal sleep is usually less disturbed than in narcolepsy with cataplexy. When the subject is young, cataplexy may develop later in the course of the disorder.

Given the risk of overdiagnosing, a positive diagnosis of narcolepsy without cataplexy cannot be done without an all-night polysomnography followed by an MSLT documenting a mean sleep latency of less than 8 min and two or more SOREMPs.

(iii) Narcolepsy due to medical condition

As in the two previous categories, the patient has a complaint of excessive daytime sleepiness occurring almost daily for at least 3 months. However, the distinct feature of this subtype of narcolepsy is the existence of a significant underlying medical or neurological disorder accounting for the excessive daytime sleepiness and/or cataplexy.

According to the *ICSD*-2 the diagnosis of narcolepsy due to medical condition can be made only if one of the following is observed:

- a definite history of cataplexy
- a polysomnographic monitoring followed by an MSLT demonstrating a mean sleep latency of less than 8 min with two or more SOREMPs
- hypocretin-1 levels in the CSF lower than 110 pg/ml, provided the patient is not comatose.

In addition, a consistent chronological link with the presumed underlying disease must be established.

Narcolepsy due to a medical condition is extremely rare.

(b) Idiopathic hypersomnia

In comparison with narcolepsy, which is characterized by well-defined clinical, polysomnographic, and biochemical features, idiopathic hypersomnia is not as well delineated.

Its first description dates back to 1976,[9] almost a century after that of narcolepsy. According to various sleep disorders centres series populations the ratio of idiopathic hypersomnia to narcolepsy with cataplexy would be around 15 per cent.

Idiopathic hypersomnia includes two forms referred to as idiopathic hypersomnia with long sleep time and idiopathic hypersomnia without long sleep time. The first one is remarkable for three symptoms: a complaint of constant or recurrent excessive sleepiness and unwanted naps, usually longer and less irresitible than in narcolepsy, and non-refreshing irrespective of their duration; night sleep is sound, uninterrupted, and prolonged; morning or nap awakening is laborious. Subjects do not awaken to the ringing of a clock, of a telephone, and often rely on their family members who must use vigorous and repeated procedures to wake them up. Even then patients may remain confused, unable to react adequately to external stimuli, a state referred to as "sleep drunkenness" In contrast, idiopathic hypersomnia without long sleep time stands as isolated excessive daytime sleepiness.

The diagnosis of idiopathic hypersomnia is mainly based on clinical features. However, laboratory tests are necessary to rule out other hypersomniac conditions. Polysomnographic monitoring of nocturnal sleep demonstrates normal sleep, except for its prolonged duration in the case of idiopathic hypersomnia with long sleep time. NREM and REM sleep are in normal proportions. There is no SOREMP. Sleep apnoeas and periodic limb movements should theoretically be absent, but may be acceptable in the case of an early onset of idiopathic hypersomnia and of their late occurrence. The MSLT usually demonstrates a mean sleep latency less than 8 min, but is typically longer than in narcolepsy. Fewer than two SOREMPs are present. In cases with normal MSLT findings, several authors have suggested the use of long-term sleep monitoring to document the excessive amount of sleep.

In contrast with narcolepsy, onset of idiopathic hypersomnia is much more progressive. The disorder is generally stable and long lasting. Complications are mostly social and professional.

Treatment of idiopathic hypersomnia relies mainly on modafinil. However, awakening difficulties are hardly improved by this treatment.

(c) Recurrent hypersomnia

The most classic form of recurrent hypersomnia is the Kleine–Levin syndrome.[10] This is an uncommon disorder with roughly 300 cases published in the world literature. Adolescent males are most commonly affected.

The syndrome is characterized by recurrent episodes of hypersomnia associated with behavioural disorders including binge eating (rapid consumption of a large amount of food on a compulsory manner), oversexuality in the form of sexual advances, shamlessly expressed fantasies or masturbation in public, irritability, odd behaviours (like standing on the head, singing loudly, talking in a childish manner), and cognitive disorders, feeling of unreality, confusion, visual or auditory hallucinations. Simultaneous occurrence of all these symptoms is more the exception than the rule however. During the episode the patient may sleep as long as 14 to 18 h per day, waking or getting up only to eat and void.

Body weight gain of a few kilograms can be observed during the episode.

Amnesia of the episode, transient dysphoria, or elation with insomnia for 1 or 2 days, may follow the episode itself. During asymptomatic intervals patients sleep normally and do not experience behavioural or cognitive disorders.

Diagnosis of the Kleine–Levin syndrome is essentially clinical and laboratory tests merely serve to exclude the possibility of rare recurrent hypersomnias of secondary origin, organic, or psychiatric. Of note due to disordered behavioural features, especially hypersexuality, it is not rare that patients are first hospitalized on psychiatric wards and given antipsychotic drugs.

The course is usually benign with episodes lessening in frequency, duration, and severity. Complications are mainly social and occupational.

Treatment of the Kleine–Levin syndrome is not well codified. The effects of stimulant drugs such as modafinil or methylphenidate on the hypersomniac episodes are difficult to assess given that the methods of evaluation are purely subjective and that the episodes vanish spontaneously within a few days. A prophylactic treatment based on mood stabilizers (carbamazepine, lithium carbonate, and more recently valproic acid) deserves to be prescribed in severe cases. Positive results, that is absence of recurrence of the symptoms throughout the period of administration and recurrence with discontinuation, has been reported.

(d) Hypersomnias associated with various medical disorders

(i) Associated with neurological diseases

Hypersomnia may occur in any intracranial pressure syndrome, but it may also result from tumours affecting the diencephalon, specially the ventro-lateral posterior part of the hypothalamus or the peduncular region, with no associated intracranial hypertension. Cases of narcolepsy secondary to brain tumours affecting the hypothalamus or the midbrain region have been reported.

Uni or bilateral paramedian thalamic infarcts and paramedian pedunculo thalamic infarcts are the most typical causes of hypersomnia of vascular origin.

A non-negligible fraction of patients with Parkinson's disease present excessive sleepiness. This is even more the case of patients with multiple system atrophy.

Normal pressure hydrocephalus, Arnold–Chiari malformation, myotonic dystrophy, may also lead to excessive sleepiness.

(ii) Associated with infectious diseases

Intense fatigue and severe excessive sleepiness may develop in the month following Epstein–Barr disease. The same holds true of atypical viral pneumonia, hepatitis B viral infection and the Guillain–Barré syndrome. Hypersomnia tends to go into gradual remission after several months or years.

Disorders of alertness and/or consciousness are found in virtually all patients affected by viral encephalitis.

Human African Trypanosomiasis (sleeping sickness) is a subacute or chronic parasitic disease caused by the inoculation of a protozoan, *Trypanosoma brucei* transmitted by the tsetse fly. It is endemic to certain regions of tropical Africa. The form found in West and Central Africa is due to *Trypanosoma gambiense*. The invasion of the central nervous system is characterized by meningoencephalitis with abnormal sleepiness, headache, trembling, dyskenesia, choreoathetosis, and mood changes. Polysomnography has shown episodes of sleep occurring randomly day and night and SOREMPs.

(iii) Associated with endocrine diseases

Hypothyroidism and acromegaly are the two main sources of hypersomnia, usually due to obstructive sleep apnoea syndrome.

(iv) Post-traumatic hypersomnia

Excessive daytime sleepiness appearing during the year following a head injury may be considered *a priori* as post-traumatic. This typically presents as extended night sleep and episodes of daytime sleep. Brain imaging may reveal lesions affecting the hypothalamic region or brainstem, midbrain or pontine tegmentum, rarely hydrocephalia, or more often the absence of any significant lesions.

The delayed sleep phase syndrome

The major sleep episode is usually delayed 3 to 6 hours relative to conventional sleep-wake times. Affected individuals complain of great difficulty falling asleep at a socially acceptable time, but once sleep ensues, sleep is reported to be normal. Enforced conventional

wake time usually results in morning excessive sleepiness. The disorder has been associated with mental disorders, particularly in adolescents. Schizoid and avoidant personality features are frequently associated with this disorder.

Treatment rests on various approaches including chronotherapy, light exposure, and melatonin therapy.

Conclusion

Hypersomnia deserves to be taken into consideration by the psychiatrist. It may be the consequence of a psychiatric disorder or of its treatment, but it may also be non psychiatric in nature and result in psychiatric symptoms.

Further information

Nofzinger, E.A., Thase, M.E., Reynolds, C.F., *et al.* (1991). Hypersomnia in bipolar depression: a comparison with narcolepsy using the multiple sleep latency test. *The American Journal of Psychiatry*, **148**, 1177–81.

American Academy of Sleep Medicine. (1999). Sleep related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. *Sleep*, **22**, 667–89.

Dauvilliers, Y., Arnulf, I., and Mignot, E. (2007). Narcolepsy with cataplexy. *Lancet*, **369**, 499–511.

Billiard, M. and Dauvilliers, Y. (2007). Idiopathic hypersomnia. In *Narcolepsy and hypersomnia* (eds. C.L. Bassetti, M. Billiard, and E. Mignot), pp. 77–87. Informa Healthcare, New York.

Arnulf, I., Zeitzer, J.M., File, J., *et al.* (2005). Kleine-Levin syndrome: a systematic review of 186 cases in the literature. *Brain*, **128**, 2763–76.

Regenstein, Q. and Monk, T. (1995). Delayed sleep phase syndrome: a review of its clinical aspects. *The American Journal of Psychiatry*, **152**, 602–8.

References

1. American Academy of Sleep Medicine. (2005). *International classification of sleep disorders: diagnostic and coding manual* (2nd edn). American Academy of Sleep Medicine, Westchester, IL.
2. Ohayon, M.M., Caulet, M., Philip, P., *et al.* (1997). How sleep and mental disorders are related to complaints of daytime sleepiness. *Archives of Internal Medicine*, **157**, 2645–52.
3. Liu, X., Uchiyama, M., Kim, K., *et al.* (2000). Sleep loss and daytime sleepiness in the general adult population of Japan. *Psychiatric Research*, **93**, 1–11.
4. Guilleminault, C., Tilkian, A., and Dement, W.C. (1976). The sleep apnea syndromes. *Annual Review of Medicine*, **27**, 465–84.
5. Young, T., Palta, M., Dempsey, J., *et al.* (1993). The occurrence of sleep disordered breathing among middle-aged adults. *The New England Journal of Medicine*, **328**, 1230–5.
6. Ohayon, M.M. (2003). The effects of breathing-related sleep disorders on mood disturbances in the general population. *Journal of Clinical Psychiatry*, **64**, 1195–2000.
7. Westphal, C. (1877). Eigentümliche mit Einschlafen verbundene Anfälle. *Archives für Psychiatrie und Nervenkrankheiten*, **7**, 631–5.
8. Gelineau, J. (1880). De la narcolepsie. *Gazette des Hôpitaux* (Paris), **55**, 626–8; 635–7.
9. Roth, B. (1976). Narcolepsy and hypersomnia. *Schweizer Archiv für Neurologie und Psychiatrie*, **119**, 31–41.
10. Critchley, M. (1962). Periodic hypersomnia and megaphagia in adolescent males. *Brain*, **85**, 627–56.

4.14.4 Parasomnias

Carlos H. Schenck, and Mark W. Mahowald

In all of us, even in good men, there is a lawless, wild-beast nature which peers out in sleep. –Plato, *The Republic*

Relevance of parasomnias to psychiatrists

Parasomnias are defined as undesirable physical and/or experiential phenomena accompanying sleep that involve skeletal muscle activity (movements, behaviours), autonomic nervous system changes, and/or emotional-perceptual events.[1] Parasomnias can emerge during entry into sleep, within sleep, or during arousals from any stage of sleep; therefore, all of sleep carries a vulnerability for parasomnias.[1] Parasomnias can be objectively diagnosed by means of polysomnography (i.e. the physiologic monitoring of sleep—figures 4.14.4.1, 4.14.4.2), and can be successfully treated in the majority of cases.[2–5] Understanding of the parasomnias, based on polysomnographic documentation, has expanded greatly over the past two decades, as new disorders have been identified, and as known disorders have been recognized to occur more frequently, across a broader age group, and with more serious consequences than previously understood.[1–10] Parasomnias demonstrate how our instinctual behaviours, such as locomotion, feeding, sex, and aggression, can be released during sleep, itself a basic instinct. There are at least eight reasons why parasomnias should be of interest and importance to psychiatrists:

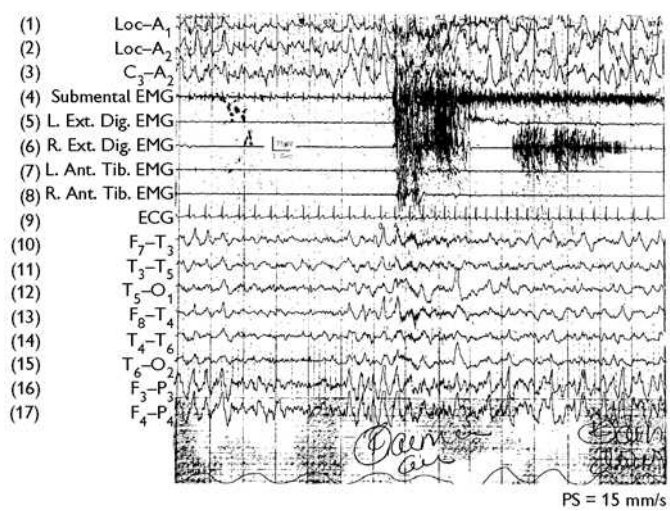


Fig. 4.14.4.1 Polysomnogram of a disordered arousal, with the persistence of sleep, in a 23-year-old man with a history of sleepwalking and sleep terrors. After a behavioural arousal from slow-wave sleep (with arm lifted up and then down), the EEG shows irregular delta and theta activity and superimposed faster frequencies. Immediately preceding the arousal, there is a cluster of three high-amplitude delta waves (channel 3). Electro-occulogram, channels 1. 2; EEG, channels 3, 10–17; EMG, electromyogram. (Reproduced from C.H. Schenck *et al.* Analysis of polysomnographic events surrounding 252 slow-wave sleep arousals in thirty-eight adults with injurious sleepwalking and sleep terrors. *Journal of Clinical Neurophysiology*, **15**, 159–66, copyright 1998, American Clinical Neurophysiology Society)

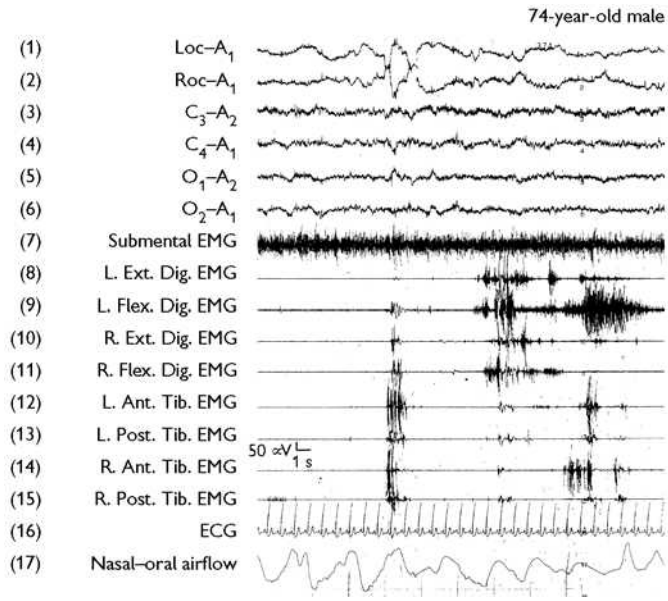


Fig. 4.14.4.2 Polysomnogram of disordered rapid eye movement (REM) sleep in a man with REM sleep behaviour disorder who eventually developed Parkinson's disease. There is complete loss of 'REM-atonia', as the submental electromyogram (EMG) shows continuous muscle tone (channel 7). The appearance of a rapid eye movement (channels 1, 2) signals the onset of excessive muscle twitching in the upper/lower extremity EMGs (channels 8–15). The EEG (channels 3–6) shows the typical low-voltage fast-frequency desynchronized activity of REM sleep. ECG rate (channel 16) remains constant despite generalized muscle twitching, which is a common finding in REM sleep behaviour disorder. Elecro-oculogram: channels 1,2.

1 Parasomnias can be misdiagnosed and inappropriately treated as a psychiatric disorder.

2 Parasomnias can be a direct manifestation of a psychiatric disorder, e.g. dissociative disorder, nocturnal bulimia nervosa.

3 The emergence and/or recurrence of a parasomnia can be triggered by stress.

4 Psychotropic medications can induce the initial emergence of a parasomnia, or aggravate a preexisting parasomnia.

5 Parasomnias can cause psychological distress or can induce or reactivate a psychiatric disorder in the patient or bed partner on account of repeated loss of self-control during sleep and sleep-related injuries.

6 Familiarity with the parasomnias will allow psychiatrists to be more fully aware of the various medical and neurological disorders, and their therapies, that can be associated with disturbed (sleep-related) behaviour and disturbed dreaming.

7 Parasomnias present a special opportunity for interlinking animal basic science research (including parasomnia animal models) with human (sleep) behavioural disorders.

8 Parasomnias carry forensic implications, as exemplified by the newly-recognized entity of 'Parasomnia Pseudo-suicide.' Also, psychiatrists are often asked to render an expert opinion in medicolegal cases pertaining to sleep-related violence.

Classification of parasomnias

Parasomnias can be classified according to whether the signs or symptoms are primary phenomena of sleep itself, or whether they are secondary phenomena derived from various underlying disorders, with sleep facilitating the nocturnal manifestation of these disorders.[6] Table 4.14.4.1 contains such a classification, and provides a context (along with other current sources[1,11]) for the parasomnias to be discussed in this chapter. Parasomnias demonstrate how sleep and wakefulness are not mutually exclusive states. Features of rapid eye movement (REM) sleep, non-REM sleep and wakefulness can occur simultaneously, and with rapid oscillations.[12] Status dissociatus represents the most extreme form of state dissociation.[13]

Clinical evaluation of parasomnias

The evaluation of complex and violent nocturnal behaviours at our centre (Minnesota Regional Sleep Disorders Center) includes the following:[2,14]

1 Clinical sleep-wake interview, with review of medical records, and review of a patient questionnaire that covers sleep-wake, medical, psychiatric and alcohol/chemical use and abuse history, review of systems, family history, and past or current physical, sexual, and emotional abuse.

2 Psychiatric and neurologic interviews and examinations, including psychometric testing.

Table 4.14.4.1 Classification of parasomnias: primary and secondary sleep phenomena

Primary sleep phenomena
Non-REM sleep
Disorders of arousal: sleepwalking/sleep terrors/confusional arousal
REM sleep
REM sleep behaviour disorder (RBD)
Dream anxiety attacks (nightmares)
Miscellaneous (including mixed non-REM/REM sleep)
Parasomnia overlap disorder (sleepwalking/sleep terrors/RBD)
Sleep related eating disorder
Restless legs syndrome/periodic limb movements in sleep
Obstructive sleep apnoea-related parasomnias
Rhythmic movement disorders
Status dissociatus
Bruxism
Secondary sleep phenomena
Central nervous system
Seizures (conventional, unconventional)
Headaches
Psychiatric
Nocturnal dissociative disorders
Nocturnal panic attacks
Nocturnal bulimia nervosa
Post-traumatic stress disorder
Cardiopulmonary (arrhythmias, asthma, etc.)
Gastrointestinal (gastro-oesophageal reflux etc)
Malingering

Modified from Mahowald and Ettinger.[7]

3 Extensive overnight polysomnographic monitoring at a hospital sleep laboratory, with continuous audio-visual recording. Figures 4.14.4.1, 4.14.4.2 depict the polysomnographic montage that includes the electro-oculogram, EEG, chin and four-limb electromyograms, ECG, and nasal-oral airflow (with full respiratory monitoring whenever indicated). Polysomnographic recording speeds of 15 to 30 mm/s are employed in order to detect epileptiform activity. Urine toxicology screening is performed whenever indicated.

4 Daytime multiple sleep latency testing, if there is a complaint or suspicion of excessive daytime sleepiness or fatigue (methods discussed in Narcolepsy chapter).

Causes of sleep related injury

A report on a series of 100 consecutive adults presenting to a multidisciplinary sleep disorders centre on account of sleep-related injury identified five causes:[2]

1 Sleepwalking/sleep terrors (M:F 3:2; mean age of onset, 5 years);

2 REM sleep behaviour disorder (predominantly male; mean age of onset, 57 years)

3 Dissociative disorders (Mostly female mean age of onset, 21 years)

4 Nocturnal seizures (uncommon)

5 Obstructive sleep apnoea/periodic limb movements (uncommon).

The sleep-related injuries included ecchymoses, lacerations, and fractures in 95 per cent, 30 per cent, and 9 per cent of patients respectively.

Non-REM sleep parasomnias: sleepwalking and sleep terrors

The polysomnographic correlates of sleepwalking and sleep terrors were first identified in the 1960s and 1970s by Gastaut and Broughton,[15, 16] Kales *et al.*,[17] and Fisher *et al.*[18] from France, Canada, and the United States. Sleepwalking and sleep terrors are classified as 'disorders of arousal,' and typically arise from delta non-REM (slow wave) sleep and usually affect children, but adults can also be afflicted, and suffer from sleep-related injuries and adverse social consequences.[2,9,10,19–24]

Clinical findings

Sleepwalking (SW) is characterized by complex, automatic behaviours, such as aimlessly wandering about, nonsensically carrying objects from one place to another, rearranging furniture, eating inappropriately, urinating in cupboards, going outdoors, and on rare occasion, driving a car.[1,25,26] The eyes are usually wide open and have a glassy stare, and there may be some mumbling. However, communication with a sleepwalker is usually poor or impossible. Frenzied or aggressive behaviour, the wielding of weapons (knives, guns), or the calm suspension of judgement (e.g. leaving via a bedroom window, wandering far outdoors) can result in inadvertent injury or death to self or others. Homicidal sleepwalking can occur.[25]

Sleepwalking episodes usually emerge 15–120 minutes after sleep onset, but can occur throughout the entire sleep period in adults. The duration of each episode can vary widely. The following is a wife's description.[2]

'He seems to have the strength of 10 men and shoots straight up from bed onto his feet in one motion. He's landed clear across the room on many occasions and has pulled down curtains (bending the rods), upset lamps, and so forth. He's grabbed me and pulled on me, hurting my arms, because he's usually dreaming that he's getting me out of danger . . . He's landed on the floor so hard that he's injured his own body . . . There are low windows right beside our bed and I'm afraid he'll go through them some night.'

Sleep terrors (ST) are characterized by sudden, loud, terrified screaming and prominent autonomic nervous system activation (tachycardia, tachypnea, diaphoresis, mydriasis) that usually appears early in the sleep period, although episodes in adults can occur at any time of the night. The individual may sit up rapidly while screaming, and engage in frenzied activity, such as bolting out of bed, and becoming injured.

Childhood sleepwalking and sleep terrors are characterized by complete amnesia for the events. In adult SW and ST, there can be subsequent recall of the behavioural episode, and also recall of dream-like mentation that usually involves being threatened by imminent danger.[2,27] The distinction between ST and agitated SW in adults is often blurred, with both states being admixed in response to a perceived threat.[24]

The prevalence of SW has been estimated to be as high as 17 per cent in childhood (peaking at age 4–8), and recent data indicate a higher prevalence in adults (4 per cent) than previously recognized.[28–30] The prevalence of ST in children can be greater than 6 per cent and greater than 2 per cent in adults.[1] A familial-genetic basis for SW and ST has been well-established.[30,31] Non-injurious SW does not have a gender preference,[1] although injurious SW appears to be more male-predominant.[2] Sleep terrors do not have a gender preference.[1] 'Confusional arousals' comprise another category of 'disorder of arousal,' and represent partial manifestations of sleepwalking and sleep terrors in which aggression and sexual impulses can be released.[1,10]

Polysomnographic findings

Sleepwalking/sleep terrors episodes arise abruptly during arousals from delta non-REM sleep.[1,15–18] In a systematic study of 38 adults with injurious SW/ST,[24] three postarousal EEG patterns were detected: diffuse, rhythmic delta activity; diffuse delta and theta activity intermixed with alpha and beta activity; and prominent alpha and beta activity. Thus, the postarousal EEG can show the complete persistence of sleep, the admixture of sleep and wakefulness, or complete wakefulness. Figure 4.14.4.1 shows the polysomnogram of a disordered arousal from slow-wave sleep.

Although the sleep architecture (i.e. distribution of sleep stages) is usually normal in SW and ST, the 'micro-structure' of non-REM sleep in adult SW/ST can be perturbed, with increased micro-arousals and increased rate of the 'cyclic alternating EEG pattern'.[1,2,21,23]

Association with medical and psychiatric risk factors

Sleepwalking and sleep terrors may be triggered by sleep deprivation, febrile illness, alcohol use or abuse, pregnancy, menstruation, obstructive sleep apnea, periodic limb movements, nocturnal

seizures, medical and neurological disorders, and psychotropic medications—especially zolpidem, lithium carbonate and anticholinergic agents.[1]

A strong association between sleepwalking/sleep terrors and psychopathology in adults was suggested by an early literature, but polysomnograpic monitoring was not conducted, and there were considerable methodological problems. The recent literature reporting PSG-confirmed cases has indicated that most adult cases are not closely associated with a psychiatric disorder, although stress can play a promoting role[2,19,29,22,32]. Genetic-constitutional factors appear to be predominant in adult and childhood sleepwalking/sleep terrors.[1]

Treatment

Treatment (especially in childhood SW/ST) is usually not necessary, other than identifying and minimizing any identified risk factors, and safety measure to avoid accidental injury. In cases involving sleep-related injury, pharmacological treatment is necessary and can be life-saving. A benzodiazepine, such as clonazepam (0.25–1.5 mg) taken 1 hour before sleep onset is usually effective. Alprazolam, diazepam, imipramine, and paroxetine can also be used. Teaching a patient self-hypnosis can be effective in milder cases of either adult or childhood SW/ST.[33] Treatment of any concurrent psychiatric disorder does not usually control sleepwalking/sleep terrors.[2,22] Attempts to waken the patient may cause confusion and distress.

REM sleep behaviour disorder (RBD)

Although various aspects of RBD have been identified by European, Japanese and American investigators since 1966,[34] RBD was not formally recognized and named until 1986–1987,[35,36] and it is incorporated within the international classification of sleep disorders.[1] A typical clinical presentation of RBD is contained in the description of the index case:[35]

> 'A 67-year-old dextral man was referred because of violent behavior during sleep . . . 4 years before referral . . . he experienced the first 'physically moving dream' several hours after sleep onset; he found himself out of bed attempting to carry out a dream. This episode signaled the onset of an increasingly frequent and progressively severe sleep disorder; he would punch and kick his wife, fall out of bed, stagger about the room, crash into objects, and injure himself . . . his wife began to sleep in another room 2 years before referral. They remain happily married, believing that these nocturnal behaviors are out of his control and discordant with his waking personality.'

Mammalian REM sleep, REM atonia, and paradox lost

REM sleep in mammals involves a highly energized state of brain activity, with both tonic (i.e. continuous) and phasic (i.e. intermittent) activations occurring across a spectrum of physiologic parameters.[36] REM sleep has two major synonyms: 'active sleep,' because of the high level of brain activity during REM sleep, and 'paradoxical sleep,' because of the nearly complete suppression of muscle tone despite the high level of brain activity. This generalized skeletal muscle atonia ('REM-atonia') is one of the three defining features of mammalian REM sleep, besides rapid eye movements and a desynchronized EEG. The loss of the customary paradox of REM sleep in RBD bears serious clinical consequences: 'paradox lost' means loss of safe sleep.[9]

Animal model of RBD

In 1965, Jouvet and Delorme reported from France that experimental pontine lesions in cats caused permanent loss of REM-atonia, and the cats displayed attack and exploratory behaviours during REM sleep that resembled dream-enactment (oneirism). This experiment established the first animal model of RBD.[34]

Clinical and polysomnographic findings

Between 1982 and 1991, 96 patients at our centre were diagnosed with chronic RBD. (There is an acute form of RBD that can emerge during withdrawal from ethanol or sedative-hypnotic abuse and with anti-cholinergic and other drug intoxication states.[36] Data on this series[37] are contained in Table 2, and are concordant with the published world literature.[34,36] The older male predominance in RBD is striking, although females and virtually all age groups are represented. Approximately half of RBD cases are closely associated with neurological disorders, predominantly neurodegenerative disorders (especially parkinsonism), narcolepsy and stroke. In fact, RBD may be the first sign of a parkinsonian disorder whose other (classic) manifestations may not emerge until several years or even decades after the onset of RBD.[36,38,39] Thus, routine neurological evaluations are indicated in the long-term management of RBD. The prevalence of RBD is estimated to be from 0.38 per cent to 0.5 per cent.[1] The course is usually progressive; spontaneous

Table 4.14.4.2 Findings in 96 patients with chronic REM sleep behaviour disorder (RBD)

Categories	Percentage (*N*)	Comments
Gender		
Male	87.5 (84)	Mean age of RBD onset (*N* = 90): 52 years (range 9–81)
Female	12.5 (12)	Mean age at polysomnography: 58 years (range 10–83)
Chief complaint		
Sleep injury	79.2 (76)	Ecchymoses (76); lacerations (32); fractures (7)
Sleep disruption	20.8 (20)	
Altered dream process and content	87.5 (84)	More vivid, intense, action filled, violent (reported as severe nightmares)
Dream-enacting behaviours	87.5 (84)	Talking, laughing, yelling, swearing, gesturing, reaching, grabbing, arm flailing, punching, kicking, sitting, jumping out of bed, crawling, running
Clonazepam treatment		
Efficacy (*N* = 67)		
Complete	79.1 (44/67)	Rapid control of problematic sleep behaviours *and* altered dreams, sustained for up to 9 years
Partial	11.9 (8/67)	
Total	91.0 (61/67)	

Modified from Schenck *et al.*[36]

remissions are very rare. Patients with RBD usually have calm and pleasant personalities, and do not display irritability or anger while awake, even though their dreams are highly aggressive.[40] Figure 4.14.4.2 depicts a typical polysomnogram of RBD with attempted dream-enactment.

Association of RBD with psychiatric disorders and Stress

Psychiatric disorders are rarely associated with RBD.[34,41] Fluoxetine treatment of obsessive compulsive disorder,[42] or cessation of use or abuse of REM-suppressing agents (viz. ethanol, amphetamine, cocaine, imipramine) can trigger RBD.[36,43] Tricyclic antidepressants, selective serotonin reuptake inhibitors, venlafaxine, mirtazapine, and monoamine oxidase inhibitors can induce or aggravate RBD. In four cases, major stress involving divorce, automobile accident, sea disaster, or public humiliation triggered RBD.[2,34,43]

Diagnosis of RBD

The diagnostic criteria of RBD are as follows:[1,36]

1 Polysomnographic abnormality during REM sleep: elevated submental electromyographic tone and/or excessive phasic submental and/or limb electromyographic twitching.

2 Documentation of abnormal REM sleep behaviours during polysomnographic studies, or a history of injurious or disruptive sleep behaviours.

3 Absence of EEG epileptiform activity during REM sleep.

Treatment of RBD

Clonazepam is remarkably effective in controlling both the behavioural and the dream-disordered components of RBD, at a usual bedtime dose of 0.5 to 1.0 mg. The long-term efficacy and safety of chronic, nightly clonazepam treatment of RBD has been established.[3] Other treatments include melatonin, pramipexole, etc.[36] Maximizing the safety of the sleeping environment should always be encouraged.

Parasomnia overlap disorder: sleepwalking/ sleep terrors/RBD

A group of 33 patients has been reported with polysomnogram-documented sleepwalking, sleep terrors, and RBD.[8] The mean age was 34 years, the mean age of parasomnia onset was 15 years (range: 1–66), and 70 per cent were male. An idiopathic subgroup (N=22) had a significantly earlier mean age of parasomnia onset (9 years) than a symptomatic subgroup (27) whose parasomnia began with either a neurologic disorder (N=6), nocturnal paroxysmal atrial fibrillation (N=1), post-traumatic stress disorder/ major depression (N=1), chronic ethanol/amphetamine abuse and withdrawal (N=1), or mixed disorders (schizophrenia; brain trauma; substance abuse [N=2]). The rate of psychiatric disorders was not elevated; group scores on various psychometric tests were not elevated. Forty-five percent (N=15) had previously received psychologic or psychiatric therapy for their parasomnia, without benefit. Treatment, usually with clonazepam, was effective for most patients. The natural history of this disorder is not yet known. Other cases of Parasomnia Overlap Disorder have been reported.[34,36]

Sleep related eating disorder

The 'night-eating syndrome' was first reported in 1955, and featured abnormal eating during full wakefulness in insomniac patients. Over the next 35 years, abnormal nocturnal eating received scant attention in the literature, until a series of 19 cases (expanded to 38 cases) with polysomnographic data was published on the new entity of sleep related eating disorder,[1,4,44] which featured abnormal eating during partial arousals from sleep in patients with other parasomnias or those who were idiopathic. Reports from various countries have now been published.[45–49]

Clinical findings

Pathological sleep-related eating has characteristic features, as identified in two separate series of adult cases.[4,44,48] Neither daytime binge-eating nor obsessive-compulsive disorder was diagnosed in any patient. Sleep-related eating was not associated with either the onset or the course of a psychiatric disorder. Patients do not usually experience hunger or thirst during their 'driven' nocturnal eating and drinking. A typical behavioural sequence consists of 'automatic' arising from bed, and going straight to the kitchen with a compulsive 'out of control' urge to eat. Alcohol is almost never consumed, even in former alcohol abusers. Thick substances, such as milkshakes, peanut butter, and brownies, are preferentially consumed at night. Purging does not take place, either at night or in the morning. Sleep-related eating is usually invariant, and is not influenced by day of the week or sleeping away from home. More than 40 per cent of patients in one series were overweight from the nocturnal eating, according to standardized body mass index criteria.[4]

Sleep related eating is most commonly associated with sleepwalking, but also with restless legs syndrome, obstructive sleep apnoea, narcolepsy, zolpidem/triazolam/midazolam/other psychotropic medication use or abuse, cessation of cigarette smoking, cessation of alcohol, opiate and cocaine abuse; cessation of cigarette smoking; stress (particularly involving separation anxiety), and various medical/neurological disorders (e.g. autoimmune hepatitis, encephalitis, migraines). Rarely, anorexia or bulimia nervosa or a dissociative disorder can be associated with sleep-related eating.[45]

Prescribing a monoamine oxidase inhibitor to a patient with sleep related eating disorder can be hazardous, since indiscriminate food consumption at night could jeopardize the mandatory dietary restrictions of MAOI treatment.

Treatment

Treatment, with one notable exception, is primarily directed at controlling the underlying sleep disorder. For cases associated with restless legs syndrome, therapy with dopaminergics (at times supplemented with an opiate such as codeine) is usually effective; likewise, therapy of obstructive sleep apnoea with nasal continuous positive airway pressure is often effective in controlling the sleep related eating. In contrast, in cases that are idiopathic and in those cases associated with sleepwalking, therapy with the anticonvulsant topiramate (that suppresses overeating urges) was effective in two-thirds of cases in two reported series.[5,50]

Sleep related eating disorder shares many features in common with the 'night eating syndrome' in adults,[49,51] although there are usually major differences in regards to level of consciousness during nocturnal eating, association with other sleep disorders,

and response to therapy. It is likely that these two conditions sit at opposite poles of a spectrum of abnormal nocturnal eating.

Sleep related dissociative disorders

A nocturnal dissociative disorder with polysomnographic monitoring was first reported in 1976 by Rice and Fisher in a man with daytime and night-time fugues that began after his father's death.[52] Another polysomnogram-documented case was described in a woman with a history of being physically and sexually assaulted.[1] A series of eight cases was reported from our centre in 1989.[53]

Clinical manifestations

The onset may be sudden or gradual, and the course is chronic. There usually is a history of repeated physical and sexual abuse in childhood and/or adulthood. In a series of eight cases, seven were female who also had daytime states of dissociation with self-mutilating behaviours, such as genital cutting, self-burning, and punching through windows.[53] One male patient had exclusively nocturnal episodes in which he acted like a jungle cat. Patients may report sleep phobia as a consequence of bed-related and sleep-related sexual and physical abuse.

A typical spell during polysomnographic monitoring involves complex and lengthy behaviours that emerge during well-established EEG wakefulness, after a prior episode of sleep.[53] The nocturnal episodes appear to be re-enactments of previous assaults.

Treatment

Treatment involves long-term therapy of the dissociative disorders and of associated psychiatric disorders, usually initiated in a specialized in-patient setting. Bedtime administration of benzodiazepines may aggravate a nocturnal dissociative disorder.

Restless legs syndrome (RLS)

RLS, first described in 1945 by Ekbom from Sweden, is a chronic disorder that often results in severe insomnia, and can be incapacitating.[1,54] It is characterized by unpleasant and at times painful sensations in the lower extremities that emerge during periods of inactivity, particularly during the transition from wakefulness to sleep. These abnormal sensations are relieved by movement or stimulation of the legs, such as walking, stomping the feet, rubbing or squeezing the legs, taking hot showers, or applying hot packs or ointments to the legs.

Both the RLS and the therapeutic interventions just described are incompatible with successful entry into sleep. The more severely affected individuals cannot easily sit still while watching television or a film, or during protracted plane or train journeys. RLS is quite common, affecting up to 10 per cent of the general population, and tends to become more prevalent with increasing age. It affects females more than males. Childhood cases, though rare, at times may masquerade as 'attention deficit disorder with hyperactivity,' or as 'growing pains.'

Most RLS cases are idiopathic with a strong familial basis. Caffeine, fatigue, or stress may worsen the symptoms. There is no evidence that RLS is related to psychiatric disorders, although the symptoms of RLS may suggest a primary anxiety disorder and through impairment of the duration and quality of sleep it can affect daytime mood and behaviour. Also, neuroleptic-induced akathisia can mimic RLS. Secondary RLS can be associated with peripheral neuropathies, renal disease, and psychotropic medications (especially the SSRIs, venlafaxine, and tricyclic antidepressants).

Nearly all patients with RLS have 'periodic limb movements' (PLMs) of non-REM sleep (formerly called 'nocturnal myoclonus'). PLMs are characterized by periodic (every 15–40 sec) movements of the legs, viz. slow dorsi-flexion, which may or may not be associated with arousals. Excessive arousals during sleep from any cause, including PLMs, may result in daytime fatigue or sleepiness.

RLS can usually be diagnosed by history alone, and polysomnographic monitoring is usually not indicated. The syndrome is one of the major organic causes of insomnia, and commonly responds to treatment that includes use of dopaminergics (e.g. pramipexole, ropinerole, levodopa), opiates, and benzodiazepines (e.g. clonazepam). Combinations of these drugs are often necessary. Since RLS is a chronic disorder that often worsens with advancing age, chronic long-term treatment is usually warranted.

Differential diagnosis of dream-enacting behaviours and other parasomnias

A history of dream-enacting behaviours does not automatically implicate RBD. Other diagnoses include sleepwalking/sleep terrors;[2,9] obstructive sleep apnoea, with apnoea-induced arousals from REM sleep being associated with violent behaviours and vivid REM-related dreams;[2,55] nocturnal complex seizures, with the 'dreams' being the seizure equivalents;[34] sleep related dissociative disorders, with the 'dreams' being wakeful memories of past abuse;[53] intoxication states; and malingering.

Other parasomnias of interest to psychiatrists include sexual parasomnias ('sleepsex', 'sexsomnia'),[10] nocturnal panic attacks, which arise from Stages II and III non-REM sleep,[56] sleep related trichotillomania, nocturnal frontal lobe epilepsy,[11] and rhythmic movement disorders of non-REM and REM sleep, including head-banging and body rocking. Finally, expanded knowledge on the parasomnias has allowed for the recognition of a new medical-legal entity called 'parasomnia pseudo-suicide'.[58] This term refers to the unfortunate, but unintentional, fatal consequence of complex, sleep-related behaviours that may be erroneously attributed to suicide.[1]

Forensic guidelines for psychiatrists

The legal implications of automatic behaviour have long been discussed and debated. With regard to sleep-related automatic behaviours, the objective identification of a sleep disorder does not establish causality for any given deed. Guidelines have been developed to assist in determining the role of a sleep disorder in an act of violence.[57]

- There should be reason (by history or by formal sleep laboratory evaluation) to suspect a bona fide sleep disorder. Similar episodes, with benign or morbid outcome, should have occurred previously. (Note that disorders of arousal may begin in adulthood).
- The duration of the action is usually brief (minutes).
- The behaviour is usually abrupt, immediate, impulsive, and senseless—without apparent motivation. Although ostensibly

purposeful, it is completely inappropriate to the total situation, out of (waking) character for the individual, and without evidence of premeditation.

- The victim is someone who merely happened to be present and who may have been the stimulus for the arousal.
- Immediately after return to consciousness, there is perplexity or horror, without an attempt to escape, conceal, or cover up the action. There is evidence of lack of awareness on the part of the individual during the event.
- There is usually some degree of amnesia for the event, however, this amnesia may not be complete.
- In the case of ST or SW or sleep drunkenness, the act may (a) occur on awakening, (rarely immediately upon falling asleep) and usually at least 1 hour after sleep onset; (b) occur on attempts to awaken the subject; and (c) have been potentiated by alcohol ingestion, sedative or hypnotic administration or prior sleep deprivation.

The American Academy of Neurology and other professional organizations have developed guidelines for expert witnesses, with the most stringent being as follows.

- The expert should be willing to submit his or her testimony for peer review.
- The expert must be impartial and be willing to prepare his or her testimony for use by either or both the plaintiff or defendant. Ideally, the expert should assume the role of 'friend of the court.'

Further information

Schenck, C.H., Mahowald, M.W. Parasommias (2008). Associated with sleep-disordered breathing and its therphy, including sexsomnia as a recently recognized parasomnia. *Somnology*, **12**, 38–49.

Howell, M.J., Schenck, C.H., Crow, S.J. (2008). A review of night-time eating disorders. Sleep Medicine Reviews; doi: 10.1016/j.smrv.2008.07.005.

Schenck, C.H. (2005). *Paradox lost: midnight in the battleground of sleep and dreams*. Extreme-Nights, LLC, Minneapolis, Minnesota www.parasomnias-rbd.com

"Sleep Runners: The Stories Behind Everyday Paramnias—Deluxe Academic Edition" (3 DVDs, including 50 minutes of indexed sleep lab parasomnia footage; and CD-ROM), Slow-Wave Films, LLC, St. Paul, Minnesota, 2007. www.sleeprunners.com

www.parasomnias-rbd.com

References

1 American Academy of Sleep Medicine (2005). *International classifi cation of sleep disorders:Diagnostic and coding manual* (2nd edn.). Westchester, Ill:*American Academy of Sleep Medicine*.

2. Schenck, C.H., Milner, D.M., Hurwitz, T.D., *et al.* (1989). A polysomnographic and clinical report on sleep-related injury in 100 adult patients. *American Journal of Psychiatry*, **146**, 1166–73.
3. Schenck, C.H., and Mahowald, M.W. (1996b). Long-term, nightly benzodiazepine treatment of injurious parasomnias and other disorders of disrupted nocturnal sleep in 170 adults. *American Journal of Medicine*, **100**, 548–54.
4. Schenck, C.H., Hurwitz, T.D., O'Connor, K.A., *et al.* (1993). Additional categories of sleep-related eating disorders and the current status of treatment. *Sleep*, **16**, 457–66.
5. Winkelman J.W. (2006). Efficacy and tolerability of open-label topiramate in the treatment of sleep-related eating disorders: an open-label, retrospective case series. *Journal of Clinical Psychiatry*, **67**(11), 1729–34.
6. Mahowald, M.W., and Ettinger, M.G. (1990). Things that go bump in the night: the parasomnias revisited. *Journal of Clinical Neurophysiology*, **7**, 119–43.
7. Ohayon M.M., *et al.* (1997) Violent behavior during sleep. *Journal of Clinical Psychiatry*, **58,** 369–76.
8. Schenck, C.H., Boyd, J.L., and Mahowald, M.W. (1997). A parasomnia overlap disorder involving sleepwalking, sleep terrors, and REM sleep behavior disorder in 33 polysomnographically confirmed cases. *Sleep,* **20**, 972–81.
9. Schenck, C.H. (2005). *Paradox Lost: Midnight In The Battleground Of Sleep And Dreams*. Minneapolis, Minnesota, Extreme-Nights, LLC.
10. Schenck, C.H., Arnulf, I., Mahowald, M.W. (2007). Sleep and sex: what can go wrong? A review of the literature on sleep disorders and abnormal sexual behaviors and experiences. *Sleep*, **30**, 683–702.
11. Kryger, M.H., Roth, T., and Dement, W.C. (2005). *Principles and Practice of Sleep Medicine* (4th edn.). Elsevier: Philadelphia, PA.
12. Mahowald, M.W. and Schenck, C.H. (1992). Dissociated states of wakefulness and sleep. *Neurology*, **42**(Suppl. 6), 44.
13. Mahowald, M.W. and Schenck, C.H. (1992). Status dissociatus—a perspective on states of being. *Sleep*, **14,** 69.
14. Schenck, C.H., Mahowald, M.W. (2005). Rapid Eye Movement Sleep Parasomnias. *Neurologic Clinics*, **23**: 1107–26.
15. Gastaut, H. and Broughton, R. (1965). A clinical and polygraphic study of episodic phenomena during sleep. *Recent Advances in Biological Psychiatry*, **7**, 197–222.
16. Broughton, R. (1968). Disorders of sleep: disorders of arousal? Science, **59**, 1070–8.
17. Kales, A., Jacobson, A., Paulson, J., *et al.* (1966). Somnambulism: psychophysiological correlates. I. All-night EEG studies. *Archives of General Psychiatry*, **14**, 586–94.
18. Fisher, C., Kahn, E., Edwards, A., *et al.* (1973). A psychophysiological study of nightmares and night terrors. I. Physiological aspects of the stage 4 night terror. *Journal of Nervous and Mental Disease*, **157**, 75–98.
19. Kavey, N.B., Whyte, J., Resor, S.R., *et al.* (1990). Somnambulism in adults. *Neurology*, **49**, 749–52.
20. Crisp, A.H., Matthews, B.M., Oakey, M., *et al.* (1990). Sleepwalking, night terrors, and consciousness. *British Medical Journal*, **300**, 360–2.
21. Blatt, I., Peled, R., Gadoth, N., *et al.* (1991). The value of sleep recording in evaluating somnambulism in young adults. *Electroencephalography and Clinical Neurophysiology*, **78**, 407–12.
22. Llorente, M.D., Currier, M.B., Norman, S.E., *et al.* (1992). Night terrors in adults: phenomenology and relationship to psychopathology. *Journal of Clinical Psychiatry*, **53**, 392–4.
23. Zucconi M, *et al.* (1995) Arousal fl uctuations in non-rapid eye movement parasomnias: the role of cyclic alternating pattern as a measure of sleep instability. *Journal of Clinical Neurophysiology*, **12**:147–54.
24. Schenck, C.H., Pareja, J.A., Patterson, A.L., *et al.* (1998). Analysis of polysomnographic events surrounding 252 slow-wave sleep arousals in thirty-eight adults with injurious sleepwalking and sleep terrors. *Journal of Clinical Neurophysiology*, **15**, 159–66.
25. Broughton, R., Billings, R., Cartwright, R., *et al.* (1994). Homicidal somnambulism: a case report. *Sleep*, **17**, 253–64.
26. Schenck, C.H., and Mahowald, M.W. (1995). A polysomnographically documented case of adult somnambulism with long-distance automobile driving and frequent nocturnal violence: parasomnia with continuing danger as a non-insane automatism? *Sleep*, **18**, 765–72.
27. Kavey, N.B. and Whyte, J. (1993). Somnambulism associated with hallucinations.

28. Bixler, E.O., Kales, A., Soldatos, C.R., *et al.* (1979). Prevalence of sleep disorders in the Los Angeles metropolitan area. *American Journal of Psychiatry*, **136**, 1257–62.
29. Klackenberg, G. (1982). Somnambulism in childhood—prevalence, course and behavior correlates. A prospective longitudinal study (6–16 years). *Acta Paediatrica*, **71**, 495–9.
30. Hublin, C., Kaprio, J., Partinen, M., *et al.* (1997). Prevalence and genetics of sleepwalking: a population-based twin study. *Neurology*, **48**, 177–81.
31. Kales, A., Soldatos, C.R., Bixler, E.O., *et al.* (1980). Hereditary factors in sleepwalking and night terrors. *British Journal of Psychiatry*, **137**, 111–18.
32. Crisp. A.H. (1996). The sleepwalking/night terrors syndrome in adults. *Postgraduate Medical Journal*, **72**, 599–604.
33. Hurwitz, T.D., Mahowald, M.W., Schenck, C.H., *et al.* (1991). A retrospective outcome study and review of hypnosis as treatment of adults with sleepwalking and sleep terror. *Journal of Nervous and Mental Disease*, **179**, 228–33.
34. Schenck, C.H., Mahowald, M.W. (2002). REM Sleep Behaviour Disorder: Clinical, Developmental, and Neuroscience Perspectives 16 Years After Its Formal Identification in Sleep, *Sleep,* **25**,120–38.
35. Schenck, C.H., Bundlie, S.R., Ettinger, M.G., *et al.* (1986). Chronic behavioral disorders of human REM sleep: a new category of parasomnia. *Sleep*, **9**, 293–308.
36. Mahowald, M.W., Schenck, C.H. (2005). REM Sleep Parasomnias. In *Principles and Practice of Sleep Medicine* (4th edn.) (eds. M.H. Kryger, T. Roth, W.C. Dement,), pp. 897–916, Elsevier Saunders, Philadelphia, Pennsylvania.
37. Schenck, C.H., Hurwitz, T.D., and Mahowald, M.W. (1993). REM sleep behaviour disorder: an update on a series of 96 patients and a review of the world literature. *Journal of Sleep Research*, **2**, 224.
38. Gagnon J-F, Postuma R.B., Mazza, S., *et al.* (2006). Rapid-eye-movement sleep behavior disorder and neurodegenerative diseases. *Lancet Neurology*, **5**, 424–32.
39. Iranzo, A., Molinuevo, J.L., Santamaria, J., *et al.* (2006) Rapid-eye-movement sleep behavior disorder as an early marker for a neurodegenerative disorder: a descriptive study. *Lancet Neurology*, DOI: 10.1016/S1474-4422(06)70476–8.
40. Fantini, M.L., Corona, A., Clerici, S., *et al.* (2005) Aggressive dream content without daytime aggressiveness in REM sleep behavior disorder. *Neurology*, **65**, 1010-15.
41. Schenck, C.H., and Mahowald, M.W. (1990). A polysomnographic, neurologic, psychiatric and clinical outcome report on 70 consecutive cases with REM sleep behavior disorder (RBD): sustained clonazepam efficacy in 89.5 per cent of 57 treated patients. *Cleveland Clinical Journal of Medical*, **57**(Suppl), 10–24.
42. Schenck, C.H., Mahowald, M.W., Kim, S.W., *et al.* (1992). Prominent eye movements during NREM sleep and REM sleep behavior disorder associated with fl uoxetine treatment of depression and obsessive-compulsive disorder. *Sleep*, **15**:226–35.
43. Schenck, C.H., Hurwitz, T.D., Mahowald M.W. (1988). REM sleep behaviour disorder. *American Journal of Psychiatry*, **145,** 652.
44. Schenck, C.H., Hurwitz, T.D., Bundlie S.R., *et al.* (1991). Sleep-related eating disorders: polysomnographic correlates of a heterogeneous syndrome distinct from daytime eating disorders. *Sleep,* **14**, 419–31.
45. Schenck, C.H., and Mahowald, M.W. (1994). Review of nocturnal sleep-related eating disorders. *International Journal of Eating Disorders*, **15**, 343–56.
46. Spaggiari, M.C., Granella, F., Parrino, L., (1994). Nocturnal eating syndrome in adults. *Sleep*, **17**, 339–44.
47. Manni, R., Ratti, M.T., and Tartara, A. (1997). Nocturnal eating: prevalence and features in 120 insomniac referrals. *Sleep*, **20**, 734–8.
48. Winkelman, J.W. (1998). Clinical and polysomnographic features of sleep-related eating disorder. *Journal of Clinical Psychiatry*, **59**, 14–19.
49. Vetrugno, R., Manconi, M., Ferini-Strambi, L., *et al.* (2006). Nocturnal eating: sleep-related eating disorder or night eating syndrome? A videopolysomnographic study. *Sleep*, **29**: 949–54.
50. Schenck, C.H., and Mahowald, M.W. (2006). Topiramate therapy of sleep related eating disorder. *Sleep*, **29**:A268.
51. Schenck, C.H. (2006). Journal Search And Commentary: A Study of Circadian Eating and Sleeping Patterns In Night Eating Syndrome (NES) Points The Way To Future Studies On NES And Sleep-Related Eating Disorder. *Sleep Medicine*, **7**, 653–6.
52. Rice, E. and Fisher, C. (1976). Fugue states in sleep and wakefulness: a psychophysiological study. *Journal of Nervous and Mental Disease*, **163**, 79–87.
53. Schenck, C.H., Milner, D.M., Hurwitz, T.D., *et al.* (1989b). Dissociative disorders presenting as somnambulism: polysomnographic, video and clinical documentation (8 cases). *Dissociation*, **2**, 194–204.
54. Montplaisir, J., Allen, R.P., Walters, A.S., *et al.* (2005). Restless legs syndrome and periodic limb movements during sleep. In *Principles and Practice of Sleep Medicine* (4th edn.), (eds. M.H. Kryger, T. Roth, W.C. Dement), pp. 839–849. Elsevier Saunders, Philadelphia, Pennsylvania.
55. Iranzo, A. and Santamaria, J. (2005). Severe obstructive sleep apnea/hypopnea mimicking REM sleep behavior disorder. *Sleep*, **28**: 203–06.
56. Shapiro, C.M. and Sloan, E.P. (1998). Nocturnal panic—an underrecognized entity. *Journal of Psychosomatic Research*, **44**, 21–3.
57. Mahowald, M.W. and Schenck, C.H. (1999). Sleep-related violence and forensic medicine issues. In Sleep disorders medicine: basic science, technical considerations, and clinical aspects (2nd edn) (ed. S. Chokroverty), pp. 729–39. Butterworth Heinemann, Boston, MA.
58. Mahowald, M.W., Schenck, C.H., Goldner, M., *et al.* (2003). Parasomnia Pseudo- Suicide. *Journal of Forensic Sciences*, **48**: 1158–62.

4.15

Suicide

Contents

4.15.1 Epidemiology and causes of suicide

Jouko K. Lonnqvist

Definition of suicide and the reliability of suicide statistics

Suicidal behaviour or suicidality can be conceptualized as a continuum ranging from suicidal ideation and communications to suicide attempts and completed suicide. A developmental process which leads to suicidal ideation, suicidal communication, self-destructive behaviour, in some cases even to suicide, and its consequences to the survivors is often referred to as a suicidal process. There is no single unanimously accepted definition of suicide, although in most proposed definitions it is considered as a fatal act of self-injury (self-harm) undertaken with more or less conscious self-destructive intent, however vague and ambiguous. Since the deceased cannot testify as to his or her intent, the conclusions about this must be drawn by inference. The evidence required for this inference depends on many factors, for example the mode of death, the use of autopsy, age, gender, social and occupational status, and the social stigma of suicide in the person's culture. The assessment of suicide intent is always based on a balance of probabilities.

Besides the conceptual problems, there are differences in operational definitions of suicidal behaviour which may lead to lack of uniformity of case definition and difficulties in comparing suicide statistics. The reliability of suicide statistics is influenced by whether suicide is ascertained by legal officials as in the United Kingdom and Ireland, or by medical examiners as elsewhere in Europe. In general, suicides tend to be undercounted, whereas non-suicidal deaths are very rarely misidentified as suicides. Most misclassified suicides fall into the category of undetermined deaths and are more like suicides than accidents. Underestimation is reasoned to be less than 10 per cent in the more developed countries, which allows rate comparisons between countries and over time. Despite problems in the recording of suicide, reports on suicide rates among different cultures or people suggest a true variation in suicide mortality.[1,2]

The suicide process and the act of suicide

Suicide is a mode of death usually consequent to a complex and multifaceted behaviour pattern. It is typically seen as the fatal outcome of a long-term process shaped by a number of interacting cultural, social, situational, psychological, and biological factors. Suicide is a rare, shocking, and very individual final act, which often leaves the survivors helpless. The suicide process model is used to organize and clarify the complexity of factors associated with suicide (Fig. 4.15.1.1).

Suicide is usually preceded by years of suicidal behaviour or feelings, and plans and warnings. In about half of all suicides, a previous attempt is found in the person's history, which offers, in theory, an opportunity for suicide intervention wherever suicide attempts occur. Male suicide attempts are more violent and the first attempt more likely to end in death. Successful suicide prevention calls for sensitive understanding of suicidal intent and active early intervention.[3]

Various risk or protective factors underlie suicidal behaviour. An appearance of suicidality means either an intensified effect of risk factors or a weakened effect of protective factors. For example, a separation from someone close may precipitate a suicidal imbalance in a vulnerable person due to the adverse life event as a stressor and the broken social network as a loss of social support.

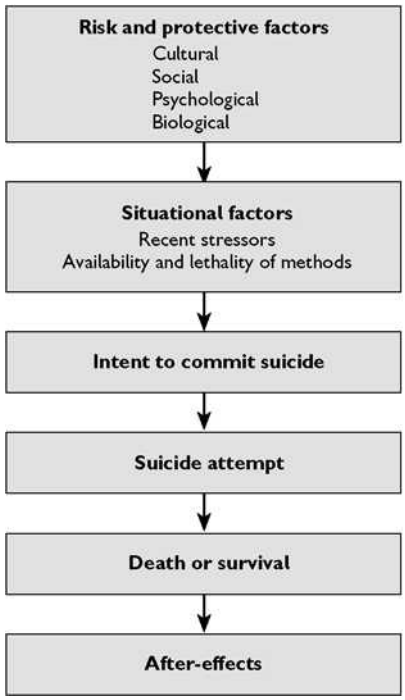


Fig. 4.15.1.1 The suicide process model.

The treating personnel and relatives of the suicide victim tend to overemphasize the meaning of the most recent events in the course of the suicide process. A precipitating factor may well be decisive in explaining the precise timing of suicide in the long course of a person's suicide process. Often, however, it also allows a simple and rational explanation in the face of the complexity of suicide.

The choice of a specific method takes place at the very end of the suicide process and represents the last possibility to intervene. Hanging is universally available and it is the most common suicide method globally. In many countries the ready access to firearms makes them potentially dangerous, especially among male adolescents and young adults. Restricting access to handguns might be expected to reduce the suicide rate of young people. Previously, domestic gas was frequently used as a suicide method, and detoxification resulted in a significant decrease in suicide rates. Nowadays, the increasing suicide rate in many Asian societies has been largely linked with pesticides and other poisons. Restrictive availability of lethal measures may also be important in the clinical treatment of individual suicidal patients. Restriction in availability of dangerous means is a strategy based on the fact that suicidal crises are often brief, suicidal acts are often impulsive, and the long-term suicide rate of serious suicide attempts is remarkably low.[4–6]

Firearms, carbon monoxide, and hanging are active suicide methods with the highest potential to cause death. Jumping from a height or leaping in the front of a moving vehicle are more passive ways, but are also highly damaging in nature. Poisoning, drowning, or wrist cutting are typically methods which leave more time for help seeking and intervention.

Imitation means learning the use of a specific suicide method from a model which is overtly available in a culture, community, institution, or mass media. Imitation may have a significant effect on the choice of a suicide method, especially at schools, in psychiatric hospital wards, and in the general population of young people. The most famous example of imitation is the effect of Goethe's novel *The Sorrows of Young Werther*, which was widely read in Europe about 200 years ago. The suicide of the hero was imitated to such an extent that authorities in several European countries banned the novel. The 'Werther effect' also appeared after the death of Marilyn Monroe; the suicide rate rose about 10 per cent over the next 10 days. Recommendations for reporting of suicide have encouraged avoidance of repetitive and excessive reports, descriptions of technical details, simplified explanations for suicide, presenting suicide as means of coping with personal problems, or glorifying suicide victims.[7]

Most suicides are solitary and private, but a few result from a pact between people to die together. Suicide pacts are exceptional, accounting for less than 1 per cent of all suicides.

Suicide always has a major impact on the survivors. Suicide is a threatening event not only among close family members, but also in the surrounding population, including treating personnel and the people at the victim's workplace. The major challenges after a suicide, in addition to a normal mourning process, are dealing with shame and guilt feelings, and the crisis of survivors. Sharing of the traumatic experience and social support should be arranged immediately and continued, if necessary, at least some months after the suicide.

Epidemiology and public health aspects of suicide

Every year one million people commit suicide, accounting for 1 to 2 per cent of total global mortality. Suicide is a leading cause of premature death, especially among young adults. It is the fifth highest cause of years of life lost in the developed world. In many westernized countries, suicide is a more frequent cause of death than traffic accidents. According to World Health Organization (**WHO**) statistics, the annual world-wide incidence of completed suicide was 16 per 100 000 persons in 2000. This means that globally one person commits suicide every minute. Suicide is estimated to represent two per cent of the total global burden of disease.

The long-term trend in suicide mortality has been increasing at least during the last 50 years. The rank order of suicide mortality in the European region in 2001 to 2003 shows that most of the countries with high suicide mortality are located in Eastern Europe (Table 4.15.1.1). Outside this region, suicide mortality has been high in Japan and China. Everywhere, the male suicide rate is clearly higher than the female rate; China is the only exception with a very high female suicide rate.

The suicide rate of elderly people has been higher than in the younger age groups almost universally. However, in many Western countries, the suicide rate for people aged 65 years and over has been declining for decades. This change is associated with the growth of the general well being and the better social and health services.

Traditionally the incidence of suicide has been low in the younger age group (15–24), but during the past 40 years the suicide rate has been rising in many Western countries, especially among young males.

Table 4.15.1.1 Suicide rates per 100 000 by country, 2001–2003

European legion	Males	Females
Lithuania	74.3	13.9
Russian Federation	69.3	11.9
Belarus	63.3	10.3
Kazakhstan	50.2	8.8
Estonia	47.7	9.8
Ukraine	46.7	8.4
Latvia	45.0	9.7
Hungary	44.9	12.0
Finland	31.9	9.8
Republic of Moldova	30.6	4.8
Czech Republic	27.5	6.8
Austria	27.1	9.3
France	26.6	9.1
Poland	26.6	5.0
Switzerland	26.5	10.6
Romania	23.9	4.7
Slovakia	23.6	3.6
Ireland	21.4	4.1
Bulgaria	21.0	7.3
Germany	20.4	7.0
Iceland	19.6	5.6
Sweden	18.9	8.1
Portugal	18.9	4.9
Luxembourg	18.5	3.5
Norway	16.1	5.8
Netherlands	12.7	5.9
Spain	12.6	3.9
Italy	11.1	3.3
United Kingdom	10.8	3.1
TFYR Macedonia	9.5	4.0
Uzbekistan	9.3	3.1
Malta	8.6	1.5
Albania	4.7	3.3
Greece	4.7	1.2
Georgia	3.4	1.1
Armenia	3.2	0.5
Azerbaijan	1.8	0.5
Other		
Japan	35.2	12.8
Republic of Korea	24.7	11.2
China (Hong Kong SAR)	20.7	10.2
Australia	20.1	5.3
Canada	18.7	5.2
United States of America	17.6	4.1
Thailand	12.0	3.8
Singapore	11.4	7.6
Kuwait	2.5	1.4

A long list of major public health concerns in the field of suicidology has emerged:

- suicidal ideation and suicide attempts are surprisingly common in the general population
- the high rate of suicides among adolescents and young adults
- unemployment as a major risk factor for suicide
- easy access to lethal suicide methods such as psychotropic or analgesic drugs, guns, and motor vehicles
- high alcohol consumption and increasing substance misuse
- undertreatment of major psychiatric disorders such as depression and schizophrenia
- suicide models projected by the mass media.

These findings indicate that rapid growth and continuous changes in society are simultaneously causing instability and disturbing the development of integration. Some regions and groups of people are inevitably affected negatively by this development, and large numbers of people are thus moving towards a greater risk of suicide.

Determinants of suicide

Usually, suicide has no single cause. It is the endpoint of an individual process, in which several interacting determinants or risk factors can be identified (Tables 4.15.1.2 and 4.15.1.3). Risk factors are by their nature cultural, social, situational, psychological, biological, and even genetic.[8]

(a) Cultural factors in suicide

Culture defines basic attitudes towards life and death, and also towards suicide in society. We still have stigma against suicide. A hundred years ago, suicide was illegal in many European countries. Similarly, most churches overtly opposed suicide and allowed suicide victims to be buried only outside the cemetery. Religion was also a major integrating force between individuals and the community. In a modern secularized society, religion is still a

Table 4.15.1.2 Risk factors for suicide: sociodemographic variables

Gender	Male
Age	Elderly
Social status	Low
Educational status	Low
Marital status	Unmarried, separated, divorced, widowed
Residential status	Living alone
Employment status	Unemployed, retired, insecure employment
Economic status	Weak (males)
Profession	Farmer, female doctor, student, sailor
Special subpopulations	Students, prisoners, immigrants, refugees, religious sects
Special institutions	Hospitals, prisons, army
Region	Uneven distribution locally by urban–rural, residential, or subcultural area
Season and time	Spring and autumn, weekend, evening, anniversary
Life events	Adverse life events such as losses and separations, criminal charges
Social support	Low
Social integration	Lacking

Table 4.15.1.3 Clinical determinants of suicide

Family history	Suicidal behaviour, mental disorders
Mental disorders	Any disorder, depression, substance use disorders, personality disorders, schizophrenia
Contact with psychiatric services	Any contacts, recent contacts, post-discharge period, psychotropic drugs
Psychiatric symptoms	Hopeless, helpless, depressive, psychotic, delirious, anxious, aggressive, introversive
Suicidal behaviour	Previous suicide attempts, suicidal ideations, death wishes, indirect gestures
Physical health	Severe physical illness such as cancer, AIDS, stroke, and epilepsy; permanent sickness
Availability of suicide methods	Easy access to lethal methods

meaningful and protective factor for many individuals in a suicidal crisis. Western culture has had a tendency to emphasize the individuals's free will and the shouldering of responsibility for one's life, while egoistic and anomic trends in society have intensified and altruism has almost disappeared. Such changes may have increased the incidence of suicide in society. The cultural background of suicide is a deep structure inherited over generations. Cultural factors also prevent rapid changes in suicidal behaviour, which is evident among immigrants, whose mode and rate of suicide usually lie somewhere between the original and the host cultures.[9]

(b) Sociological theories on suicide

Classic sociology views suicide as a social, not an individual, phenomenon. The suicide victim's moral predisposition to commit suicide, not his or her individual experiences, is felt to be the crucial factor. Moral predisposition means the degree to which the victim is involved in more or less integrated groups and in the values of those groups. Suicides are seen as a disturbance or symptom of a relationship between society and individuals. In 1897, Durkheim published his famous work on suicide and described four basic types. Anomic suicide reflects a situation where an individual is no longer guided by the society due to its weakness, like the suicide of an unemployed and rejected alcoholic without any support from society. Altruistic suicide is illustrated by a society which can exert a strong influence on an individual's decision to sacrifice his or her life, as did the captain of the *Titanic*, for example. Egoistic suicide is an individualistic decision of a person no longer dependent on others' control or opinion such as a person who has arranged an assisted suicide. Fatalistic suicide is seen as a result of strict rules in a society which have proved decisive for the destiny of an individual, for example the suicide of a person held as a slave.[10] There are also newer social theories of suicide which stress more the joint effects of social factors. The concept of social isolation has been clinically useful in understanding the socio-ecological and social-psychiatric background of suicide.[11] Some sociologists have underlined the individual meanings associated with suicide.

(c) Life events and social support

The life situation preceding suicide is typically characterized by an excess of adverse life events and recent stressors. Usually, the sum effect of events is overwhelming and more important than a single life event. Job problems, family discord, somatic illness, financial trouble, unemployment, separation, and death and illness in the family are the most common life events preceding suicide. Somatic illness and retirement are age-specifically connected with the suicides of elderly men, while separation, financial troubles, job problems, and unemployment are more common among younger men. Severely disabling somatic illness is a very important risk factor for suicide in elderly male patients. In general, suicide among men is more often related to recent stressors than it is among women.

In most cases, life events are not accidental, but are usually also dependent on the individual's own behaviour. Personality features, even mental disorders, often explain the difficulties the victim has had. Among male alcoholics, life stress is connected with family discord and separations in all age groups. Other sources of stress in alcoholic male suicides are unemployment and financial troubles, whereas in depressive non-alcoholic male victims life stress is associated more with somatic illness. Among alcoholic males, adverse life events and living alone clearly have an enhancing effect on suicidality. Among females, life events as psychosocial stress are less strongly connected with suicide. Depression and adverse interpersonal life events are more frequent contributors to female than male suicides.[12]

(d) Psychology of suicide

Early psychological theories of suicide focussed on the concept of the self. A classical example is Freud's theory assuming that self-destructive behaviour in depression represents aggression directed against a part of the self that has incorporated a loss or rejection of a love object. In his later theory of suicide, Freud presented the construction of the dual instincts, where Eros is a life-sustaining and life-enhancing drive in constant interaction with Thanatos, the aggressive death instinct.

Later psychodynamic thinking on suicide focussed more on the self in relation to others. Failures in the developmental and adaptational processes are reflected in negative self-images and distorted cognitive schemas, leading to such feelings as depression, hopelessness, rage, shame, guilt, and anxiety. It is widely held that psychological pain is found as a common element at the core of all suicidal behaviours; suicide occurs when the individual can no longer endure the pain. Most recent psychological theories of suicide accept a multifactorial causation of suicide resulting from an interaction of predisposing and precipitating factors.[13] A person moves towards a suicidal crisis depending on the stressors and presence or absence of protective factors in his or her life.

(e) Neurobiological determinants of suicide

Suicidal behaviour is highly familial. Relatives of patients who commit suicide are themselves more likely to commit suicide than relatives of patients who do not commit suicide. Liability to suicidal behaviour may be a familially transmitted trait which is partly independent on the specific mental disorders. Since the heritability of liability to suicidal behaviour appears to be on the order of 30–50 per cent, interactions with environmental factors must also be significant.

Results of adoption studies suggest that genetic factors rather than familial environmental factors are the determinants of familial concordance for suicidal behaviour. Among biological relatives

of adopted suicide victims there is a higher incidence of suicide than among the relatives of non-suicide controls or among the adoptive relatives of the suicide victims. Also identical twins have a higher concordance for suicide, attempted suicide, and suicide ideation compared with non-identical twins.[14]

The findings that genetic factors contribute to suicidal behaviour has stimulated studies aimed at identifying susceptibility genes. So far molecular genetic studies have focussed on the serontonergic pathway. Two genes have emerged as being involved in the vulnerability to suicidality: the tryptophan hydroxylase 1 (TPH1) gene, as a quantitative risk factor for suicidal behaviour, and serotonin transporter gene (5-HTTLPR), which is consistently associated with impulsive-aggressive personality traits.[15] Patients who have seriously attempted suicide by violent means have low levels of the serotonin metabolite 5-hydroxyindole acetic acid in their cerebrospinal fluid.[16] These and other neurobiological changes are discussed in Chapter 4.15.3.

Basic characteristics of the suicide victim

Persons at greatest risk of suicide are usually middle-aged or older, non-married men with poor social and economic position, and a family history of mental disorders and suicidal behaviour. Usually they are living alone, and often unemployed or with insecure employment. They also typically have marked recent life stress and a weak social network. Most suffer from depression, and feel hopeless and many have a comorbid substance abuse or personality disorder. Almost all elderly victims have comorbid physical illness or are permanently disabled. Most have previously contacted health care, and communicated their suicidal tendencies at least indirectly, although usually without receiving adequate psychiatric treatment. Half of them have previously attempted suicide.

Mental disorders and suicide

Virtually all mental disorders carry an increased risk of suicide.[17] The suicide risk in functional mental disorders is double that in substance use disorders, which in turn carry double the risk of suicide compared to organic disorders. The greatest risk of suicide among all clinical states is in attempted suicide, which carries about 40 times the expected value (Table 4.15.1.4). In anorexia nervosa and major depression, the risk is about 20-fold, and in other mood disorders and psychoses about 10 to 15 times higher than expected. In anxiety, personality, and substance use disorders the suicide risk is at lower levels, but about 5 to 10 times higher than the expected value. In subtance disorders the risk is dependent on the type of disorder, being clearly lowest in alcohol, cannabis, and nicotine abusers.[17]

Psychological autopsy studies have been used to construct an overall view of suicide by collecting all available relevant information on the victim's life preceding his or her death. In psychological autopsy studies, mental disorders of suicide victims have been assessed using DSM-diagnoses and large unselected samples. In two recent meta-analyses[18,19] the victim received at least one diagnosis on Axis I in 87 to 90 per cent of the suicides. In all studies, depressive disorders (43 per cent) and substance use disorders (26 per cent), personality disorders (16 per cent) and psychoses (9 per cent) were frequent and comorbidity was common.

Table 4.15.1.4 Rank order of suicide in mental disorders

Suicide attempt
Anorexia nervosa
Major depression
Mood disorders not otherwise specified
Reactive psychoses
Bipolar disorder
Dysthymia
Schizophrenia
Anxiety disorders
Personality disorders
Substance use disorders

In two European psychological autopsy studies[20,21] from Finland and Northern Ireland, the distribution of the principal diagnoses was similar (Table 4.15.1.5). The most common psychiatric diagnoses in suicide were major depression and alcohol dependence. Major mood disorders together comprised 42 to 36 per cent and substance use disorders, 19 to 30 per cent of all suicides. Comorbidity was a major finding in both samples, most commonly substance use disorder with major depression. A recent European psychological autopsy study gave similar results underlining, however, the role of personality disorders as a risk factor for suicide.[22]

Table 4.15.1.5 Principal diagnoses of suicide victims in Finland and Northern Ireland

Diagnosis	Finland (%) (n = 229)	Northern Ireland (%) (n = 118)
Major depression	30	31
Depressive disorder not otherwise specified	9	—
Dysthymia	—	1
Bipolar disorder	3	4
Alcohol dependence	17	24
Alcohol misuse	2	4
Other substance use disorders	—	2
Schizophrenia	7	6
Schizoaffective disorder	3	1
Other psychoses	3	4
Anxiety disorders	1	5
Adjustment disorder	3	3
Organic mental disorders	2	1
Other Axis I disorders	2	2
Personality disorder	9	3
No diagnosis	2	10
Insufficient information for assessment	7	—

The mortality risk for suicide in major depression is 20 times that expected, and 15-to 20-fold in all affective disorders. Every sixth death among depressive people treated as psychiatric patients is by suicide.[23] The risk of suicide varies across the subclasses of depression, and is related to the selection of suicidal patients for the various types of treatment. The risk is highest for depressive inpatients, even during the postdischarge period, and much lower among psychiatric outpatients, although clearly lowest for those treated for depression in primary care.[24] A meta-analysis found a hierarchy of life-time suicide prevalence: eight per cent in people ever admitted for suicidality, four per cent in patients admitted with affective disorder but not for suicidality, and two per cent in mixed inpatient and outpatient populations.[24]

Depression of suicide victims differ qualitatively from that of living controls; it seems to be more severe and accompanied more often by insomnia, weight or appetite loss, feelings of worthlessness, inappropriate guilt, and thoughts of death or suicidal ideation.[26] In addition, impulsive and aggressive behaviour, alcohol and drug abuse and dependence, and cluster B personality disorders increase the risk of suicide in individuals with major depression.[27] Inadequate and inefficient antidepressant treatment of depressed suicide victims has been a persistent finding in several studies. Less than half of suicide victims with major depression have been in contact with psychiatric care at the time of suicide. However, there is some evidence that good monitoring and maintenance treatment in high-risk groups of patients may be able to decrease their suicide rates.[28]

Alcohol and drugs, often combined, are a major risk or a precipitating factor for suicide. They may intensify the suicidal intent, offer a constantly available suicide method, worsen the somatic status of the victim, and increase the risk of complications after the attempt. Alcohol and drugs impair judgement and lower the threshold to suicide. Alcohol is detected in about every third case at the moment of suicide.[29] The lifetime risk of suicide has been estimated at 7 per cent for alcohol dependence, with only slight variation over the life.[30] The suicide rate in heavy drinking is 3.5 times and in alcohol use disorders ten times higher than that in the general population.[31] In drug dependence or abuse it is 15 times higher than expected.[17,31] The role of substance use disorders varies greatly by country. In a recent study from Estonia, alcohol dependence was found in a half of suicide victims.[32]

The suicide risk in schizophrenia appears to be almost 10 times higher than in the general population.[17] The lifetime risk of suicide in schizophrenia is estimated to be 5 per cent.[30,33] The great majority of schizophrenic patients commit suicide in the active phase of the disorder after having suffered depressive symptoms. Suicide in schizophrenia is thus less of a surprise; it is typically preceded by a previous attempt, and suicidal intent has been communicated at least as often as in non-schizophrenic suicides.[34] Schizophrenic suicide victims differ from other schizophrenic patients by having suicidal thoughts and previous suicide attempts, being more depressive, and having more positive symptoms.[35] Undertreatment, comorbidity, treatment non-compliance, and a high frequency of non-responders are also common problems among schizophrenic suicide victims. Adequacy of comprehensive care is crucial for suicide prevention in schizophrenia, especially among actively psychotic patients with recent suicidal behaviour and depressive symptoms.[36]

Personality disorders are tightly connected with suicide. Most of the suicide victims with personality disorder, especially with borderline and emotionally unstable personality disorder, have also comorbid depressive disorder or substance abuse. They often suffer from impulsive and aggressive behaviour.[20,22,27,38] This kind of comorbidity is very frequent among the young suicide victims.

Mental disorders, particularly depressive disorders, substance abuse, and antisocial behaviour have an important role in the adolescent suicides. The diagnostic distribution of mental disorders among them is surprisingly similar to that of the young and even middle-aged adults.[39]

References

1. Lönnqvist, J. (1977). Suicide in Helsinki: an epidemiological and socialpsychiatric study of suicides in Helsinki in 1960–61 and 1970–71. *Monographs of Psychiatria Fennica*, No. 8.
2. Sainsbury, P. and Jenkins, J.S. (1982). The accuracy of officially reported suicide statistics for purposes of epidemiological research. *Journal of Epidemiology and Community Health*, **36**, 43–8.
3. Isometsä, E.T. and Lönnqvist, J.K. (1998). Suicide attempts preceding completed suicide. *British Journal of Psychiatry*, **173**, 531–6.
4. Ohberg, A., Lonnqvist, J., Sarna, S., *et al.* (1995). Trends and availability of suicide methods in Finland: proposals for restrictive measures. *British Journal of Psychiatry*, **166**, 35–43.
5. Kreitman, N. (1976). The coal gas history: UK suicide rates 1960–1971. *British Journal of Preventive and Social Medicine*, **30**, 83–90..
6. Lewis, G., Hawton, K., and Jones, P. (1997). Strategies for preventing suicide. *British Journal of Psychiatry*, **171**, 351–4.
7. Hawton, K., Fagg, J., Simkin, S., *et al.* (1998). Methods used for suicide by farmers in England and Wales. *British Journal of Psychiatry*, **173**, 320–4.
8. Maris, R.W. (2002). Suicide. *Lancet*, **360**, 319-26.
9. Neeleman, J., Halpern, D., Leon, D., *et al.* (1997). Tolerance of suicide, religion and suicide rates: an ecological and individual study in 19 Western countries. *Psychological Medicine*, **27**, 1165–71.
10. Durkheim, E. (1897). *Le suicide, etude de sociologie*. Alcan, Paris. (English translation: *Suicide*. Routledge and Kegan Paul, London, 1952.)
11. Maris, R.W. (1981). *Pathways to suicide*. John Hopkins Univesrity Press, Baltimore, MD.
12. Heikkinen, M.E., Isometsä, E.T., Marttunen, M.J., *et al.* (1995). Social factors in suicide. *British Journal of Psychiatry*, **167**, 747–53.
13. Farberow, N.L. (1997). The psychology of suicide: past and present. In *Suicide: biopsychosocial approaches* (ed. A.J. Botsis, C.R. Soldatos, and C.N. Stefanis), pp. 147–63. Elsevier, Amsterdam.
14. Brent, D.A., and Mann, J.J. (2005). Family genetic studies, suicide, and suicidal behaviour. *American Journal of Medical Genetics (Semin. Med. Genet.)*, **133**(1), 13–24.
15. Bondy B., Buettner A., and Zill P. (2006). Genetics of suicide. *Molecular Psychiatry*, **11**, 336–351.
16. Mann, J.J. (2003). Neurobiology of suicidal behaviour. *Nature Reviews Neuroscience*, **4**(10), 819–828.
17. Harris, E.C. and Barraclough, B. (1997). Suicide as an outcome for mental disorders. A meta-analysis. *British Journal of Psychiatry*, **170**, 205–28.
18. Cavanagh, J.T.O., Carson, A.J., Sharpe, M. *et al* (2003). Psychological autopsy studies of suicide: a systematic review. *Psychological Medicine*, **33**, 395–405.
19. Arsenault-Lapierre, G., Kim C., and Turecki G. (2004). Psychiatric diagnoses in 3275 suicides: a meta-analysis. *BMC Psychiatry*, **4**, 37.

20. Henriksson, M.H., Aro, H.A., Marttunen, M.J., *et al.* (1993). Mental disorders and comorbidity in suicide. *American Journal of Psychiatry*, **150**, 935–40.
21. Foster, T., Gillespie, K., and McClelland, R. (1997). Mental disorders and suicide in Northern Ireland. *British Journal of Psychiatry*, **170**, 447–52.
22. Schneider, B., Wetterling, T., Sargk, D., *et al.* (2006). Axis I disorders and personality disorders as risk factors for suicide. *European Archives of Psychiatry and Clinical Neuroscience*, **256**, 17–27.
23. Wulsin, L.R., Vaillant, G.E., and Wells, V.E. (1999). A systematic review of the mortality of depression. *Psychosomatic Medicine*, **61**, 6–17.
24. Simon, G.E. and VonKorff, M. (1998). Suicide mortality among patients treated for depression in an insured population. *American Journal of Epidemiology*, **147**, 155–60.
25. Bostwick, J.M. and Pankratz, V.S. (2000). Affective disorders and suicide risk: a reexamination. *American Journal of Psychiatry*, **157**, 1925–32.
26. McGirr, A., Renaud, J., Seguin, M., *et al.* (2007). An examination of DSM-IV depressive symptoms and risk for suicide completion in major depressive disorder: A psychological autopsy study. *Journal of Affective Disorders*, **97**, 203–9.
27. Dumais, A., Lesage, A.D, Alda, M., *et al.* (2005). Risk factors for suicide completion in major depression: a case-control study of impulsive and aggressive behaviors in men. *American Journal of Psychiatry*, **162**, 2116–24.
28. Isometsä, E.T., Henriksson, M.M., Aro, H.M., *et al.* (1994). Suicide in major depression. *American Journal of Psychiatry*, **151**, 530–6.
29. Ohberg, A., Vuori, E., Ojanperä, I., *et al.* (1996). Alcohol and drugs in suicides. *British Journal of Psychiatry*, **169**, 75–80.
30. Inskip, H.M., Harris, E.C., and Barraclough, B. (1998). Lifetime risk of suicide for affective disorder, alcoholism and schizophrenia. *British Journal of Psychiatry*, **172**, 35–7.
31. Willcox, H.C., Conner, K.R., and Caine, E.D. (2004). Association of alcohol and drug use disorders and completed suicide: an empirical review of cohort studies. *Drug and Alcohol Dependence*, **76S**, S11-S19.
32. Kõlves, K., Värnik, A., Tooding, L-M., *et al.* (2006). The role of alcohol in suicide: a case-control psychological autopsy study. *Psychological Medicine*, **36**, 923–930.
33. Palmer, B.A., Pankratz, V.S., and Bostwick, J.M. (2005). The lifetime risk of suicide in schizophrenia. *Archives of General Psychiatry*, **62**, 247–253.
34. Heila, H., Isometsä, E., Henriksson, M.H., *et al.* (1997). Suicide and schizophrenia: a nationwide psychological autopsy study on age- and sex-specific clinical characteristics of 92 suicide victims with schizophrenia. *American Journal of Psychiatry*, **154**, 1235–42.
35. Kelly, D.L., Shim, J-C., Feldman, S.M, *et al.* (2004). Lifetime psychiatric symptoms in persons with schizophrenia who died by suicide compared to other means of death. *Journal of Psychiatric Research*, **38**, 531–6.
36. Heilä, H., Isometsä, E., Henriksson, M.H., *et al.* (1999). Suicide victims with schizophrenia in different treatment phases and adequacy of antipsychotic medication. *Journal of Clinical Psychiatry*, **60**, 200–8.
37. Zouk, H., Tousignant, M., Sequin, M., *et al.* (2006). Characterization of impulsivity in suicide completers: Clinical, behavioral and psychosocial dimensions. *Journal of Affective Disorders*, **92**, 195–204.
38. Brezo, J., Paris, J., and Turecki, G. (2006). Personality traits as correlates of suicidal ideation, suicide attempts, and suicide completions: a systematic review. *Acta Psychiatrica Scandinavica*, **113**, 180–206.
39. Marttunen, M.J., Aro, H.M., Henriksson, M.M., *et al.* (1991). Mental disorders in adolescent suicide: DSM-III-R axes I and II diagnoses in suicides among 13- to 19-year-olds in Finland. *Archives of General Psychiatry*, **48**, 834–9.

4.15.2 Deliberate self-harm: epidemiology and risk factors

Ella Arensman and Ad J. F. M. Kerkhof

Introduction

Deliberate self-harm (DSH) refers to behaviour through which people deliberately inflict acute harm upon themselves, poison themselves, or try do so, with non-fatal outcome. These behaviours are somehow linked to, but do not result in, death. Common to these behaviours is that they occur in conditions of emotional turmoil. In former days these behaviours were often regarded as failed suicides. However, this view did not appear to be correct, and the great majority of patients in fact do not try to kill themselves. Therefore, the term deliberate self-harm was introduced to describe the behaviour without implying any specific motive.[1] But this too has some disadvantages because there is a temporal association between non-fatal and fatal suicidal behaviour; many people who die by suicide have engaged in DSH before. Thus, Kreitman *et al.*[2] introduced the concept of parasuicide to describe behaviour that, mostly without the intention to kill oneself, communicates a degree of suicidal intent. However, both terms, deliberate self-harm and parasuicide, are still somewhat confusing, because in practice they include people who really have the intent of killing themselves but survive the attempt. The difficulty of finding a good terminology for these behaviours is reflected in differences in research populations in empirical studies: some studies are limited to self-poisoning only (overdose), a few studies are restricted to self-injury (wrist cutting) only, some to self-poisoning and self-injury combined, and some studies include behaviours in which, due to last-moment intervention from others, there was no actual self-harm inflicted at all. In recent years the term self-harm is being used in the United Kingdom and North America since the adjective 'deliberate' is not favoured by patients, particularly those who repeatedly engage in acts of self-harm.[3]

In this chapter, we will use the term deliberate self-harm interchangeably with attempted suicide to refer to non-fatal suicidal behaviours in which there may have been an intention to die, however ambiguous this intention may have been, and irrespective of other intentions that may have been operating at the same time. It should be stressed that in deliberate self-harm many motives may play a role simultaneously, even contradictory motives such as the hope of being rescued and the wish to continue living. Intentions may vary from attention seeking or communication of despair, appeal for help, to a means for stress reduction. Common to these behaviours is that they are motivated by change: people want to bring about changes in their present situation through the actual or intended harm or unconsciousness inflicted upon the body. Deliberate self-harm may be defined as follows.[4]

> An act with non-fatal outcome, in which an individual deliberately initiates a non-habitual behaviour that, without intervention from others, will cause self-harm, or deliberately ingests a substance in excess of the prescribed or generally recognised therapeutic dosage, and which is aimed at realising changes which the subject desired via the actual or expected physical consequences.

This definition covers deliberate non-fatal suicidal behaviours. Not included are accidental cases of self-poisoning, accidental overdoses of opiates, or self-harmful acts by persons who do not anticipate the consequences of their actions. It does not include automutilation, which is an habitual, often obsessive act of inflicting (minor) self-harm, mostly without a conscious intent of changing the present situation, as with certain persons with learning disability.

Clinical features

Deliberate self-harm can have very different motivations, varying from an intention to die to a cry for help. These behaviours may be well prepared or carried out impulsively, and may have different physical consequences. The degree of lethality and the degree of medical seriousness of the consequences thus depend upon intention, preparation, knowledge and expectations of the method chosen, and sometimes upon coincidental factors such as intervention from others.

It is often difficult to assess the true intent of DSH. Because of fear for consequences, such as admission to a psychiatric hospital, or because of psychological defence mechanisms, people sometimes deny or conceal their intention to die. They also may exaggerate their intention to die in order to receive help. Sometimes people engage in potentially highly lethal self-harming behaviour without any wish to die, for example when they do not have adequate knowledge of the medication used. People who present at a general hospital with minor self-injury or minor self-poisoning may have had strong intentions to die but had insufficient knowledge of the lethality of the method. Therefore, one cannot always reliably infer what the precise meaning of the behaviour was, either from its overt characteristics or from the person's self-report. Among a large sample of adolescents aged 15 and 16 years, Rodham *et al.* found that adolescents who took an overdose more often expressed a wish to die compared to those who engaged in self-cutting.[5] Motives associated with self-cutting were self-punishment and interruption, i.e. trying to get relief from a terrible state of mind.

Epidemiology

In the 1960s and 1970s, there was a sharp increase in the number of people treated in hospitals in Europe, the United States and Australia because of intentional overdoses or self-injury. In the 1980s several studies showed a stabilization.[6,7] In the early 1990s these numbers increased further in some regions.[8,9] The absolute number of persons treated for deliberate self-harm in general hospitals, however, does not adequately reflect the size of the problem. These numbers should be calculated against the size and the characteristics of the population in the areas that are being served by the hospitals. Furthermore, in some countries DSH patients are treated by general practitioners when there is no need for hospital admission. In many instances emergency attendance for overdosing is not even registered. Except for Ireland, where a National Registry of Deliberate Self-Harm has been established,[10] there are no national registries that reliably monitor trends in DSH treated in general hospitals. Even though DSH is considered a major problem in the United States, clinical epidemiological research into DSH is uncommon.[11] Also, few epidemiological studies on DSH originate from other parts of the world.

Changes over time

In Edinburgh and Oxford, in the United Kingdom, there has been continuous monitoring of deliberate self-harm over a long period of time, where characteristics of persons engaging in DSH have been related to the corresponding population.[7,12,13] In these two cities trends in DSH rates have been documented reliably. After a period of stabilization in the 1980s a marked increase was observed. Between 1985 and 1995 the rates of DSH in Oxford increased by 62 per cent in males and 42 per cent in females. The increase in DSH has been most marked among young males. A similar trend has been observed in North Worcestershire where hospital referred cases of DSH were monitored over a period of 20 years (1981–2000).[14]

In Canada, the DSH rate was estimated to be around 304 per 100 000.[15] In the United States National Institute of Mental Health's Epidemiological Catchment Area study (1980–1985) it was found that 2.9 per cent of the respondents had engaged in DSH at some point of time.[16]

So far only one international multicentre study into deliberate self-harm has been conducted taking into consideration the methodological pitfalls outlined above. The World Health Organization (WHO) initiated a collaborative multicentre study in 16 regions in Europe using the same methodology, definition, and case-finding criteria.[9,17] The findings were related to the size and characteristics of the corresponding general population in order to investigate rates, trends, risk factors, and social indicators. Most of the epidemiological data presented here have been drawn from that study.[9,18]

Differences between countries and regions

There is widespread variation between countries with regard to rates of deliberate self-harm. Based on the latest available data for the years 1995–1999, overall, DSH rates (person-based) were highest in the United Kingdom (Oxford), Belgium (Ghent), Hungary (Pecs) and Finland (Helsinki), with female rates per 100 000 ranging from 83 in Padova (Italy) to 433 in Oxford. Male DSH rates per 100 000 ranged from 53 in Umea (Sweden) to 337 in Oxford.[9] Looking at trends over time, an average decrease for male DSH rates of 13 per cent was found comparing average person-based rates for 1989/1993 to the period 1995/1999, with the greatest reduction (70 per cent) in Innsbruck (Austria). For female DSH rates the average decrease in the same period was lower (4 per cent), with the greatest reduction (31 per cent) in Sor-Trondelag (Norway). In addition to medically referred cases of deliberate self-harm, community-based studies show that an even higher proportion of DSH appears to be 'hidden' from health care services.[19]

Differences between catchment areas in deliberate self-harm rates in the WHO/EURO study have been studied in relation to socio-economic characteristics of these areas.[9,20] No correlations were found with most of the social and economical factors supposedly related to DSH, such as population density, urban–rural distribution, proportion working in agriculture forestry or fishery, sex ratio, percentage aged 40 and over, number of people per household, percentage people living alone, percentage single parent families, per capita income, unemployment rate, life expectancy, mortality rate, infant mortality, crimes per year per 1000, and per capita alcohol consumption. Only two characteristics of the catchment areas seemed to be related to DSH rates: the percentage

of divorced people in the area and the percentage receiving social security. Family stability and the percentage of the population relying on welfare both seem to be related to the frequency of DSH, but the interpretation of these findings is difficult because one would expect the other related social indicators of societal cohesion to covary as well.

It is important, however, to realize that the characteristics mentioned above relate to regions or countries, and do not relate to individuals. At individual level, characteristics such as unemployment play an important role, but this does not mean that unemployment rates do explain high DSH rates in a region.[21,22] This relationship holds only for some regions and not for others, as is documented repeatedly.[23] The effect of exposure to risks factors may be due to contextual effects, which arise if individuals' risks of suicidal behaviour depends not only on their personal exposure to risk or protective factors, but also on how these are distributed in their social, cultural or economic environments.[24,25] In a small area study in South East London, Neeleman *et al.* found that the DSH rate of minority groups relative to the white group was low in some areas and high in other areas.

Cultural variation in DSH has been documented from India,[26] Sri Lanka,[27] and Pakistan,[28] and from ethnic groups within Western societies, such as the Inuit in Canada.[29] Neeleman *et al.*[30] studied ethnic differences in DSH in Camberwell, London, and found considerable differences between the DSH rates for white people and for British-born Indian females and African-Caribbeans.[30] Indian females had a particularly high rate, 7.8 times that of white females. Marriage problems seem to be related to DSH in Asian countries such as India, Pakistan, Sri Lanka, and China. Young married women may have serious difficulties after moving in with their husbands' extended families. Dowry problems and problems with in-laws are thought to be precipitants of attempted suicide among young married women. In Asian countries the methods used in DSH reflect differences in accessibility. Self-poisoning with organophosphate pesticides and other household poisons is prevalent. As in the Western world DSH appears to reflect feelings of hopelessness and helplessness in adverse living conditions with no prospect of improvement. Women tend to be more powerless to bring about changes in their living conditions. In Sri Lanka, the continuous warfare, poverty, and the lack of opportunities at home and abroad frustrates the young who are relatively well educated.[27]

Sex and age

In all but one centre (Helsinki) of the WHO Multicentre Study on Suicidal Behaviour the female DSH rates were higher than the male rates. Across the participating regions, on average, the rates for females were 1.5 times higher than those for men. DSH rates were consistently higher among those in the young age groups, with the highest person-based male DSH rates in the age group 25–34 years, whereas for females in most centres the highest rates were found in the age group 15–24 years.[9]

Sociodemographic characteristics

Single and divorced people were over-represented among people who engaged in DSH in the WHO/EURO study.[9,18] Nearly half of the males and 38 per cent of the females were never married. An interaction effect was found for age in that the proportion of single persons among those engaging in deliberate self-harm reduced with increasing age, whereas the proportion of divorced, separated, and widowed people increased with age. Among deliberate self-harm patients who were economically active, a high percentage was unemployed. Based on average DSH rates over the period 1995–1999, 26 per cent of the males and 14 per cent of the females were unemployed.[9]

These findings are consistent with outcomes of earlier research conducted in the United Kingdom, where socio-economic deprivation (low social class and unemployment) repeatedly appear as characteristics of the DSH populations.[22]

These findings indicate that DSH patients disproportionately have had low education, and have high levels of unemployment, poverty, and divorce. The findings may be partly related to underlying common causes, such as the presence of psychiatric disorders, but they also suggest the influence of sociological factors impacting on a relatively economically deprived group in society with a greater share of adversity.[24] Socio-economic deprivation is a well-established determinant of psychiatric morbidity and DSH.[32,33] In contrast with completed suicide, where the presence of psychiatric disorders is well documented (up to 95 per cent of suicides may have suffered from a psychiatric disorder), psychiatric disorders are much less frequent among those who deliberately harm themselves. Among those who engage in DSH for the first time in their lives, the prevalence of psychiatric disorders may be rather low; among repeaters, psychiatric morbidity is considerable.[34,35]

Methods

Methods used in deliberate self-harm are mostly 'non-violent'. In the WHO Multicentre Study, 65 per cent of males and 82 per cent of females took an overdose, based on average DSH rates for the period 1995–1999. Cutting, mostly wrist cutting, was employed in 16 per cent of male cases and 9 per cent of female cases. There are some differences between European countries in the use of particular methods. Based on the years 1995–1999, a relatively high percentage of self-cutting was found in Tallinn (Estonia, 50 per cent), Ljubljana (Slovenia, 30 per cent), and Innsbruck (Austria, 26 per cent). In Szeged (Hungary), 19 per cent of males and 15 per cent of females used poisoning with pesticides, herbicides, or other toxic agricultural chemicals, whereas in other regions this ranged from 0–3 per cent.[9] In Sor-Trondelag, Norway, higher percentages engaged in DSH by deliberate alcohol overdose (6 per cent of males and 5 per cent of females). In general, somewhat older men used the method of jumping or jumping in front of a moving object. In the Oxford studies between 1985 and 1995, 88 per cent of all episodes involved self-poisoning, 8 per cent involved self-injury, and 4 per cent involved both. There was an increase in the use of paracetamol from 31 per cent of poisoning cases in 1985 to 50 per cent in 1995.[6] There was also an increase in antidepressant overdoses and a decrease in overdoses of minor tranquillizers and sedatives. Comparing the early 1990s with the late 1990s, a slight increase was observed in overdose by medication.[9] For all regions the methods used in DSH acts did not covary significantly with age.

The differences in methods between countries may be related to differences in the accessibility of certain methods. Until 1998 paracetamol was available in large quantities in the United Kingdom, unlike other European countries.[36,37] The ingestion of

alcohol during or before the act sometimes can be considered to be a part of the actual method of DSH (when used to bring about unconsciousness, or to increase the risk of a fatal outcome), as part of the preparation (to lower the threshold for engaging in an act of DSH, because of disinhibition), or as a long-term risk factor. Hawton *et al.*[38,39] found that 22 to 26 per cent of DSH patients had consumed alcohol at the time of the act (males more frequent than females), and that 44 to 50 per cent had consumed alcohol during the 6 h before the DSH acts, this again being more common in males than in females. About 28 per cent of DSH patients in Oxford appeared to be substance misusers (alcohol and drugs).

General population self-report surveys

General population epidemiological surveys of adolescents indicate that DSH acts occur more frequently than suggested by hospital statistics.[40,41] A number of surveys have been conducted to estimate the prevalence of DSH. Most of these surveys concerned adolescents and were administered anonymously. Most questionnaire studies revealed that between 1 and 20 per cent of respondents had engaged in DSH at some point in time.[41–44] However, the methodology used in the various studies varies considerably and therefore limits comparison of the outcomes. In the study by Hawton *et al.*[41] a minority (12.6 per cent) of adolescents who had engaged in DSH had presented to hospital.

Lifetime prevalence

Based upon the rates from the WHO/EURO study the lifetime prevalence of deliberate self-harm should be around 3 per cent for females and 2 per cent for males, with some variations between countries and regions.[9] DSH acts that did not lead to medical treatment at a hospital or general practitioner's surgery are very difficult to study, because of the limited validity of self-report data. Based on two large community-based surveys, the lifetime prevalence of suicidal ideation varied from 2.6 to 25.4 per cent and for deliberate self-harm this varied from 0.4 to 4.2 per cent.[45,46]

Classification

As previously mentioned, there is a considerable variety of behaviours within the broad category of non-fatal suicidal behaviour. A review of classification studies[34] revealed three types of DSH patients: a 'mild' type, a 'severe' type, and a 'mixed' type in between.

The mild type of DSH encompasses mostly relatively non-violent methods followed by non-serious physical injury. Young age, living together, few precautions to prevent discovery, low level of suicidal preoccupation, low suicidal intent, interpersonal motivation are all characteristics associated with mild forms of attempted suicide/deliberate self-harm. The severe category consists mostly of relatively hard methods followed by serious physical consequences. Older age (over 40), many precautions to prevent discovery, high level of suicidal preoccupation, high suicidal intent, self-directed motivation, often relocated, previous attempted suicides, depression, drug dependence, a high degree of overall dysfunctioning, poor physical health, and previous psychiatric treatment are all characteristics associated with the concept of 'severe' deliberate self-harm. The risk of repetition is greater in the severe type. In between, in the mixed type of DSH, the DSH acts and patients involved show mixed characteristics, which makes this type harder to identify in medical practice.

In order to further refine the classification of deliberate self-harm, Arensman[47] included psychological and personal history variables and these characteristics were studied in relation to recurrent DSH in a follow-up period of 1 year. **The mild DSH type** was validated, approximately 40 per cent of the total sample, as being predominantly younger than 30 years of age, single, living alone or with parents, and having minor injuries because of the index attempt. The mean number of previous DSH acts was 3.7. The repetition rate in the follow-up period for this group was 27 per cent. In the older age group, two groups were distinguished: a moderate group and a group with an extremely high risk for non-fatal repetition were identified. The high-risk group, consisting of approximately 28 per cent of the total sample, suffered more physical injury as a consequence of their deliberate self-harm.

The high-risk group consisted predominantly of females in the age group 30 to 49 years who were divorced or separated, living alone, and who were economically inactive. Most of them had engaged in previous DSH acts (mean number: 5). They had histories of traumatic life events that mostly started early in life. The high-risk group showed the highest scores on depression, hopelessness and expression of state-anger, and two-thirds were diagnosed as having borderline personality disorder. In the follow-up at least 75 per cent engaged in repeated DSH.

The moderate group was characterized by low levels of physical injury following their DSH, they were predominantly aged at least 30 years and married, and scored intermediate on measures of depression, hopelessness, and anger. The mean number of previous attempts was 2.3, and 33 per cent made one or more repeated DSH acts in the follow-up. Surprisingly, this classification into three types of non-fatal suicidal behaviours did not correspond to the levels of reported suicide intent nor to the levels of the different motivations reported (to die, to appeal, to lose consciousness, revenge), underlining the difficulty of classifying DSH according to intentions. Rodham *et al.*[5] found evidence for different subgroups of DSH patients based on DSH methods and motives. For example, they identified a subgroup of female adolescents who engaged in self-cutting and who reported *self-punishment* as the primary motive, followed by *trying to get relief from a terrible state of mind*. Furthermore, *wish to die* as a motive was significantly more often reported by those who took an overdose compared with those who engaged in self-cutting.

Aetiology

The last psychological step towards deliberate self-harm is always set in conditions of emotional turmoil, an emotional crisis. Essential in crisis is the absence of any positive outlook towards the future. People completing suicide do not expect any improvement of their situation in the near or distant future. People who engage in deliberate self-harm indicate that their future is hopeless, but they still seem to have a faint hope, however ambiguous this may be, that the future might improve. In this way deliberate self-harm may be conceived as a self-invented form of crisis intervention. Studies into the cognitive functioning of DSH patients indeed show a global and stable form of negative anticipations and absence of positive anticipations towards the future, probably as a consequence of disturbances in their autobiographical memory, i.e. an

overgeneral memory.[48,49] Whenever these anticipations remain overgenerally negative after an act of DSH, it is likely that hopelessness will increase and that this behaviour will occur again.[50]

Precipitants

Difficulties or conflicts that may bring the person to believe that his or her future is without hope can trigger the psychological crisis resulting in deliberate self-harm. DSH is often precipitated by disharmony with key figures, work-related problems, financial difficulties, or physical illnesses. Long-standing relationship problems or feelings of loneliness are especially common. People who engage in DSH have a weak social support system,[48] and they report relational difficulties as major problems in life. They show deficits in interpersonal problem solving, and their future holds no promise. Their emotional status can best be described as a state of learned helplessness, a situation of a blocked escape, in which no solution exists for a perceived insurmountable adversity.[51] This leads to the question as to why these persons have developed such helpless attitudes.

Long-term vulnerability factors

The conflicts experienced by DSH patients in the days before an act of DSH are not different from the same conflicts they have experienced over and over again. Not only recent life events, but also the life events that occurred in their past are important.[47] Many DSH patients, males as well as females, have had traumatic childhood experiences, including physical and emotional neglect, broken homes, other unstable parental conditions, violence, sexual and physical abuse, incest, parents who had psychiatric treatment, who were alcoholics and/or addicted to opiates. Women who have been abused have a much greater probability of becoming a repeater later. In addition they often develop poor relationships, lack self-esteem, and experience overwhelming feelings of helplessness and hopelessness. Any trigger, for example an argument with a friend, may be sufficient to provoke suicidal ideation and behaviour.

DSH patients not only suffer from helplessness with regard to interpersonal conflicts, they also tend to be powerless in other domains of life. The DSH population disproportionately consists of unemployed persons, from low social classes, with low educational levels, economically deprived, divorced, disabled, addicted, incarcerated, and/or lonely. Many have received in- or outpatient psychiatric treatment. These findings are somewhat complicated by the fact that many of these vulnerability factors are strongly interrelated. Unemployment, addiction, and unstable partnership relations all may be caused by psychiatric diseases. For example, unemployment and DSH may both be a consequence of addiction. However, it is not fair to assume that most of the economic deprivation of suicidal patients is explained by their psychiatric condition. The considerable differences between nations in the prevalence of DSH support the importance of socio-economic conditions.

Course and prognosis

Repetition is one of the core characteristics of suicidal behaviour. Among those who die by suicide up to 40 per cent have a history of previous DSH acts.[52] Among DSH patients 'repeaters' are more common than 'first-evers'. Between 30 and 60 per cent of DSH patients engaged in previous acts, and between 15 and 25 per cent did so within the last year.[3,7,53,54]

Risk of suicide after deliberate self-harm

Prospectively, DSH patients have a high risk of dying by suicide. Between 10 and 15 per cent eventually die because of suicide.[11,53] The connection between DSH and suicide lies between 0.5 and 2 per cent after 1 year and above 5 per cent after 9 years.[11] Mortality by suicide is higher among DSH patients who have engaged in previous acts of DSH.[55,56] The risk of suicide after deliberate self-harm for males is nearly twice the female risk, the risk being particularly high in the first year.[57,58] Alcohol and drug abuse and related social deterioration are risk factors for subsequent suicide,[59] as are psychiatric diagnosis (affective disorders, schizophrenia, personality disorders), and a highly lethal non-impulsive index act of DSH. In a large European study focusing on young deliberate self-harm patients, positive correlations were found between rates of DSH and suicide for both males and females, with a statistically significant association among males aged 15–24.[60]

Repetition of deliberate self-harm

Risk of repeated DSH is highest during the first year after an act of DSH, and especially within the first 3 to 6 months.[55,57,61] In the WHO/EURO Multicentre Study on Suicidal Behaviour it was found that at least 54 per cent of DSH patients had engaged in a DSH act before, 30 per cent at least twice. Prospectively, 30 per cent of DSH patients made at least one repeated attempt in a 1-year follow-up.[62,63]

It is hoped for that knowledge of antecedents or risk factors may foster early identification of persons at risk, and improvement of treatment. Many studies have tried to identify risk factors or antecedents and some of these by now are well known. Sociodemographic risk factors associated with repetition are belonging to the age group of 25 to 49 years, being divorced, unemployed, and coming from low social class. Psychosocial characteristics of repeaters are substance abuse, depression, hopelessness, personality disorders, unstable living conditions/living alone, criminal records, previous psychiatric treatment, and a history of stressful traumatic life events, including broken homes and family violence, especially physical and mental maltreatment by partners. Prospectively, a history of previous attempts is one of the most powerful predictors of future non-fatal suicide attempts.[34,51,64,65]

Conclusions

Deliberate self-harm is a major problem in many contemporary societies. DSH seems to reflect the degree of powerlessness and hopelessness of young people with low education, low income, unemployment, and difficulties in coping with life stress. As such, non-fatal suicidal behaviour should be a major concern for politicians. There are substantial differences between communities in the prevalence of deliberate self-harm. This suggests that some communities better meet the needs of their underprivileged youngsters than others do, but we barely understand the differences between communities and nations. Preventive action therefore is difficult to design. There is a need for a better nationwide continuous registration of DSH and related socio-economic conditions. There is also a need for better mental health care management of DSH patients, and for experimental studies on the prevention of

repetition. Although we know that persons who engage in DSH are at high risk for future fatal and non-fatal suicidal behaviour, development of effective intervention, and prevention programmes is a key priority.

Further information

Hawton, K. (ed.) (2005). *Prevention and treatment of suicidal behaviour: from science to practice*. Oxford University Press, Oxford.

De Leo, D., Bille-Brahe, U., Kerkhof, A., and Schmidtke, A. (eds.) (2004). *Suicidal behaviour: theories and research findings*. Hogrefe and Huber, Göttingen, Germany.

Van Heeringen, K. (ed.) (2001). *Understanding suicidal behaviour. The suicidal process approach to research, treatment and prevention*. John Wiley & Sons Ltd., Chichester, UK.

Hawton, K. and Van Heeringen, K. (eds.) (2000). *The international handbook of suicide and attempted suicide*. John Wiley & Sons, Chichester, UK.

Website of the International Association for Suicide Prevention (IASP): http://www.med.uio.no/iasp/

References

1. Morgan, H.G., Barton, J., Pottle, S., *et al.* (1976). Deliberate self-harm: a follow-up study of 279 patients. *The British Journal of Psychiatry*, **128**, 361–8.
2. Kreitman, N., Philip, A.E., Greer, S., *et al.* (1969). Parasuicide. *The British Journal of Psychiatry*, **115**, 746–7.
3. Skegg, K. (2005). Self-harm. *Lancet*, **366**, 1471–83.
4. Platt, S., Bille-Brahe, U., Kerkhof, A., *et al.* (1992). Parasuicide in Europe: the WHO/EURO multicentre study on parasuicide. I. Introduction and preliminary analysis for 1989. *Acta Psychiatrica Scandinavica*, **85**, 97–104.
5. Rodham, K., Hawton, K., and Evans, E. (2004). Reasons for deliberate self-harm: comparison of self-poisoners and self-cutters in a community sample of adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, **43**, 80–7.
6. Hawton, K. and Fagg, J. (1992). Trends in deliberate self-poisoning and self-injury in Oxford, 1976–1990. *British Medical Journal*, **304**, 1409–11.
7. Platt, S., Hawton, K., Kreitman, N., *et al.* (1988). Recent clinical and epidemiological trends in parasuicide in Edinburgh and Oxford: a tale of two cities. *Psychological Medicine*, **18**, 405–18.
8. Schmidtke, A., Bille Brahe, U., De Leo, D., *et al.* (1996). Attempted suicide in Europe: rates, trends and sociodemographic characteristics of suicide attempters during the period 1989–1992. Results of the WHO/EURO multicentre study on parasuicide. *Acta Psychiatrica Scandinavica*, **93**, 327–38.
9. Schmidtke, A., Bille-Brahe, U., De Leo, D., *et al.* (2004). *Suicidal behaviour in Europe: results from the WHO/EURO multicentre study on suicidal behaviour*. Hogrefe & Huber, Göttingen, Germany.
10. National Suicide Research Foundation. (2007). *National registry of deliberate self harm report—2006*. National Suicide Research Foundation, Cork.
11. Owens, D., Horrocks, J., and House, A. (2002). Fatal and non-fatal repetition of self-harm. Systematic review. *The British Journal of Psychiatry*, **181**, 193–9.
12. Kreitman, N. (1977). *Parasuicide*. Wiley, London.
13. Hawton, K., Fagg, J., Simkin, S., *et al.* (1997). Trends in deliberate self-harm in Oxford, 1985–1995. *The British Journal of Psychiatry*, **171**, 556–60.
14. O'Loughlin, S. and Sherwood, J. (2005). A 20-year review of trends in deliberate self-harm in a British town, 1981–2000. *Social Psychiatr y and Psychiatric Epidemiology*, **40**, 446–53.
15. Sakinofsky, I. (1996). The epidemiology of suicide in Canada. In *Suicide in Canada* (eds. A.A. Leenaars, S. Wenckstern, I. Sakinovsky, D. Dyck, M. Kral, and R. Bland). University of Toronto Press, Toronto.
16. Moscicki, E.K., O'Carroll, P., Rae, D.S., *et al.* (1988). Suicide attempts in the epidemiologic catchment area study. *The Yale Journal of Biology and Medicine*, **61**, 259–68.
17. Kerkhof, A.J.J.M., Schmidtke, A., Bille Brahe, U., *et al.* (1994). *Attempted suicide in Europe*. DSWO-Press/World Health Organization, Leiden and Copenhagen.
18. Schmidtke, A., Bille Brahe, U., De Leo, D., *et al.* (1996). Attempted suicide in Europe: rates, trends and sociodemographic characteristics of suicide attempters during the period 1989–1992. Results of the WHO/EURO multicentre study on parasuicide. *Acta Psychiatrica Scandinavica*, **93**, 327–38.
19. Evans, E., Hawton, K., and Rodham, K. (2004). Factors associated with suicidal phenomena in adolescents: a systematic review of population-based studies. *Clinical Psychology Review*, **24**, 957–79.
20. Bille-Brahe, U., Andersen, K., Wasserman, D., *et al.* (1996). The WHO-EURO multicentre study: risk of parasuicide and the comparability of the areas under study. *Crisis*, **17**, 32–42.
21. Platt, S. (1984). Unemployment and suicidal behaviour: a review of the literature. *Social Science & Medicine*, **19**, 93–115.
22. Platt, S. and Dyer, J. (1987). Psychological correlates of unemployment among male parasuicides in Edinburgh. *The British Journal of Psychiatry*, **151**, 27–32.
23. Adam, K.S. (1990). Environmental, psychosocial, and psychoanalytic aspects of suicidal behavior. In *Suicide over the life cycle* (eds. S.J. Blumenthal and D.J. Kupfer), pp. 39–96. American Psychiatric Press, Washington, DC.
24. Neeleman, J., Wilson-Jones, C., and Wessely, S. (2001). Ethnic density and deliberate self harm: a small area study in south east London. *Journal of Epidemiology and Community Health*, **55**, 85–90.
25. Neeleman, J. (2002). Beyond risk theory: suicidal behaviour in its social and epidemiological context. *Crisis*, **23**, 114–20.
26. Latha, K.S., Bhat, S.M., and D'Souza, P. (1996). Suicide attempters in a general hospital unit in India: their socio-demographic and clinical profile—emphasis on cross-cultural aspects. *Acta Psychiatrica Scandinavica*, **94**, 26–30.
27. Eddleston, M., Rezvi Sheriff, M.H., and Hawton, K. (1998). Deliberate self harm in Sri Lanka: an overlooked tragedy in the developing world. *British Medical Journal*, **317**, 133–5.
28. Khan, M.M., Islam, S., and Kundi, A.K. (1996). Parasuicide in Pakistan: experience at a university hospital. *Acta Psychiatrica Scandinavica*, **94**, 264–7.
29. Kirmayer, L.J., Malus, M., and Boothroyd, L.J. (1996). Suicide attempts among Inuit youth: a community survey of prevalence and risk factors. *Acta Psychiatrica Scandinavica*, **94**, 8–17.
30. Neeleman, J., Jones, P., van Os, J., *et al.* (1996). Parasuicide in Camberwell—ethnic differences. *Social Psychiatry and Psychiatric Epidemiology*, **31**, 284–7.
31. Hawton, K. and Catalan, J. (1982). *Attempted suicide. A practical guide to its nature and management*. Oxford University Press, Oxford.
32. Gunnell, D.J., Peters, T.J., Kammerling, R.M., *et al.* (1995). Relation between parasuicide, suicide, psychiatric admissions, and socioeconomic deprivation. *British Medical Journal*, **311**, 226–30.
33. Congdon, P. (1996). Suicide and parasuicide in London; a small-area study. *Urban Studies*, **1**, 137–58.
34. Arensman, E. and Kerkhof, A.J.F.M. (1996). Classification of attempted suicide: a review of empirical studies, 1963–1993. *Suicide & Life-threatening Behavior*, **26**, 46–67.
35. De Leo, D., Bille-Brahe, U., Kerkhof, A.J.F.M., *et al.* (2004). *Suicidal behaviour: theories and findings*. Hogrefe & Huber, Göttingen, Germany.

36. Hawton, K., Simkin, S., Deeks, J., *et al.* (2004). UK legislation on analgesic packs: before and after study of long-term effect on poisonings. *British Medical Journal*, **329**, 1076.
37. Gunnell, D., Hawton, K., Murray, V., *et al.* (1997). Use of paracetamol for suicide and non-fatal poisoning in the UK and France: are restrictions on availability justified? *Journal of Epidemiology and Community Health*, **51**, 175–9.
38. Hawton, K., Fagg, J., Simkin, S., *et al.* (1997, 1998). *Deliberate self-harm in Oxford, 1996, 1997. Reports from the Oxford monitoring system for attempted suicide*. Warneford Hospital, Oxford.
39. Hawton, K., Simkin, S., and Fagg, J. (1997). Deliberate self-harm in alcohol and drug misusers: patient characteristics and patterns of clinical care. *Drug and Alcohol Review*, **16**, 123–9.
40. Choquet, M. and Ledoux, S. (1994). *Adolescents*: *enquête nationale*. Inserm, Villejuif Cedex.
41. Hawton, K., Rodham, K., Evans, E., *et al.* (2002). Deliberate self harm in adolescents: self-report survey in schools in England. *British Medical Journal*, **325**, 1207–11.
42. Paykel, E.S., Myers, J.K., Lindentall, J.J., *et al.* (1974). Suicidal feelings in the general population: a prevalence study. *The British Journal of Psychiatry*, **124**, 460–9.
43. Kienhorst, C.W.M., de Wilde, E.J., Diekstra, R.F.W., *et al.* (1991). Construction of an index for predicting suicide attempts in depressed adolescents. *The British Journal of Psychiatry*, **159**, 676–82.
44. Rubinstein, J.L., Heeren, T., Housman, D., *et al.* (1989). Suicidal behavior in 'normal' adolescents: risk and protective factors. *The American Journal of Orthopsychiatry*, **59**, 59–71.
45. Bertelote, J.M, Fleischmann, A., De Leo, D., *et al.* (2005). Suicide attempts, plans, and ideation in culturally diverse sites: the WHO SUPRE-MISS community survey. *Psychological Medicine*, **35**, 1–9.
46. Bernal, M., Haro, J.M., Bernert, S., *et al.* (in press). Risk factors for suicidality in Europe: results from the ESEMED study. *Journal of Affective Disorders*.
47. Arensman, E. (1997). Attempted suicide: epidemiology and classification. Unpublished PhD. Dissertation, University of Leiden, The Netherlands.
48. Williams, J.M.G. (1997). *The cry of pain*. Harmondsworth, Penguin.
49. Sinclair, J.M., Crane, C., Hawton, *et al.* (in press). The role of autobiographical memory specificity in deliberate self-harm: correlates and consequences. *Journal of Affective Disorders*.
50. MacLeod, A.K., Rose, G.S., and Williams, J.M.G. (1993). Components of hopelessness about the future in parasuicide. *Cognitive Therapy and Research*, **17**, 441–55.
51. McAuliffe, C., Corcoran, P., Keeley, H.S., *et al.* (2006). Problem-solving ability and repetition of deliberate self-harm: a multicentre study. *Psychological Medicine*, **36**, 45–55.
52. Maris, R.W. (1992). The relationship of nonfatal suicide attempts to completed suicide. In *Assessment and prediction of suicide* (eds. R.W. Maris, A.L. Berman, J.T. Maltsberger, and R.I. Yufit), pp. 362–80. Guilford Press, New York.
53. Kreitman, N. and Casey, P. (1988). Repetition of parasuicide: an epidemiological and clinical study. *The British Journal of Psychiatry*, **153**, 792–800.
54. Hawton, K. and Fagg, J. (1995). Repetition of attempted suicide: the performance of the Edinburgh predictive scales in patients in Oxford. *Archives of Suicide Research*, **1**, 261–72.
55. Hawton, K. and Catalan, J. (1981). Psychiatric management of attempted suicide patients. *British Journal of Hospital Medicine*, **26**, 365–8.
56. Hawton, K. and Fagg, J. (1988). Suicide and other causes of death, following attempted suicide. *The British Journal of Psychiatry*, **152**, 359–66.
57. Nordstrom, P., Samuelsson, M., and Asberg, M. (1995). Survival analysis of suicide risk after attempted suicide. *Acta Psychiatrica Scandinavica*, **91**, 336–40.
58. Suokas, J. and Lonnqvist, J. (1991). Outcome of attempted suicide and psychiatric consultation: risk factors and suicide mortality during a five year follow-up. *Acta Psychiatrica Scandinavica*, **84**, 545–9.
59. Cullberg, J., Wasserman, D., and Stefansson, C.G. (1988). Who commits suicide after a suicide attempt? An 8 to 10 year follow-up in a suburban catchment area. *Acta Psychiatrica Scandinavica*, **77**, 598–603.
60. Hawton, K., Arensman, E., Wasserman, D., *et al.* (1998). Relation between attempted suicide and suicide rates among young people in Europe. *Journal of Epidemiology and Community Health*, **52**, 191–4.
61. Goldacre, M. and Hawton, K. (1985). Repetition of self-poisoning and subsequent death in adolescents who take overdoses. *The British Journal of Psychiatry*, **146**, 395–8.
62. Kerkhof, A.J.F.M. and Arensman, E. (2004). Repetition of attempted suicide: frequent, but hard to predict. In *Suicidal behaviour: theories and findings* (eds. D. De Leo, U. Bille-Brahe, A.J.F.M. Kerkhof, and A. Schmidtke). Hogrefe and Huber, Göttingen.
63. Arensman, E., Kerkhof, A., Dirkzwager, A., *et al.* (1999). Prevalence and risk factors for repeated suicidal behaviour: results from the WHO/EURO multicentre study on parasuicide, 1989–1992 Report University of Leiden, The Netherlands.
64. Van Egmond, M. and Diekstra, R.F.W. (1989). The predictability of suicidal behaviour: the results of a meta-analysis of published studies. In *Suicide and its prevention* (eds. R.F.W. Diekstra, R. Maris, S. Platt, A. Schmidtke, and G. Sonneck), pp. 37–61. Brill, Leiden.
65. Kreitman, N. and Foster, J. (1991). Construction and selection of predictive scales, with special reference to parasuicide. *The British Journal of Psychiatry*, **159**, 185–92.

4.15.3 Biological aspects of suicidal behaviour

J. John Mann and Dianne Currier

Modelling suicidal behaviours

To understand the biological underpinnings of multi-determined behaviours such as suicide and attempted suicide it is necessary to situate them within an explanatory model that can elaborate the causal pathways and interrelations between biological, clinical, genetic, and environmental factors that all play a role in suicidal behaviour. Where possible, such a model should be clinically explanatory, incorporate biological correlates, be testable in both clinical and biological studies, and have some utility in identifying high-risk individuals.

We have proposed a stress–diathesis model of suicidal behaviour wherein exposure to a stressor precipitates a suicidal act in those with the diathesis, or propensity, for suicidal behaviour.[1] Stressors are generally state-dependent factors such as an episode of major depression or adverse life event. The diathesis, we have hypothesized, comprises trait characteristics such as impulsive aggression, and pessimism.[1] Uncovering the biological mechanisms relevant to the stress and the diathesis dimensions of suicidal behaviour will facilitate the identification of both enduring and proximal markers of risk, as well as potential targets for treatment.

One biological correlate of the diathesis for suicidal behaviour appears to be low serotonergic activity. Abnormal serotonergic function may be the result of numerous factors including genetics, early life experience, chronic medical illness, alcoholism or substance use disorder, many of which have been correlated with increased risk for suicidal behaviour. Moreover, serotonergic dysfunction may underlie recurrent mood disorders or behavioural traits that characterize the diathesis, such as aggression and impulsivity. In terms of stress response, the noradrenergic and HPA axis have been the focus of biological studies in suicidal behaviour. This chapter gives an overview of the major neurobiological findings in suicide and attempted suicide, as well as emerging findings from studies of genes related to those systems.

Serotonergic system

Serotonin is involved in brain development, behavioural regulation, modulation of sleep, mood, anxiety, cognition, and memory and is shown to be disturbed in various psychiatric disorders. Serotonergic function is under genetic control and, moreover, deficits in functioning have been shown to be enduring, marking it as a biological trait. The serotonergic system became a target for investigation in relation to suicide when, more than 30 years ago, Asberg and colleagues observed that depressed individuals who had either attempted suicide by violent means or subsequently died by suicide in the study follow-up period were more likely to have lower CSF 5-HIAA levels.[2] Since that time the function of the serotonergic system in suicide and attempted has been examined in many paradigms, and while not all studies agree, there is substantial consensus that individuals who die by suicide, or make serious non-fatal suicide attempts, exhibit a deficiency in CNS serotonin neurotransmission.

Evidence of hypofunction comes from *cerebrospinal fluid* and postmortem studies. 5-hydroxyindoleacetic acid (5-HIAA) is the major metabolite of serotonin and level of CSF 5-HIAA is a guide to serotonin activity in parts of the brain including the prefrontal cortex. There have been over 20 studies of CSF 5-HIAA and suicidal behaviour in mood disorders, and a meta-analysis of prospective studies of 5-HIAA found that in mood disorders lower CSF 5-HIAA increased the chance of death by suicide over fourfold over follow-up periods of 1–14 years.[3]

Multiple *postmortem studies* of suicide, report lower brainstem levels of 5-HIAA and serotonin (5-hydroxytryptamine, 5-HT) (see Mann *et al.* for a review[4]). These deficits in 5-HT or 5-HIAA are observable across diagnostic groups[5] and, despite early reports to the contrary, independent of suicide method. This abnormality appears to be largely specific to the brainstem, and multiple studies have reported no differences between suicides and controls in 5-HT level in other brain regions including the hippocampus, the occipital cortex, the frontal cortex, the temporal cortex, the caudate, the striatum, or the hypothalamus.[4] Serotonin neurone cell bodies are in the brainstem raphe nuclei, while their axons innervate most of the brain including the ventral prefrontal cortex. Morphological analysis of stained serotonin neurones in the brainstem of depressed suicides and non-suicides observed greater cell density in the dorsal raphe nucleus in the suicides[6] suggesting that reduction in serotonin activity is associated with dysfunctional neurones and not with fewer neurones.

Neuroendocrine challenge studies using fenfluramine provide further evidence of anomalous serotonergic function associated with suicidal behaviour. Fenfluramine is a serotonin-releasing drug and a reuptake inhibitor that may also directly stimulate postsynaptic 5-HT receptors. The release of serotonin by fenfluramine causes a measurable increase in serum prolactin levels that is an indirect index of central serotonergic responsiveness. In depressed patients, those with a history of suicide attempts have a more blunted prolactin response to fenfluramine challenge than non-attempters with some evidence that the effect is more strongly related to seriousness of past attempt.[7]

Studies of receptors suggest lower serotonergic transmission in the central nervous system may be accompanied by a compensatory upregulation of some serotonergic postsynaptic receptors such as the 5-HT_{1A} and 5-HT_{2A}, and a decrease in the number of serotonin reuptake sites.[4] There is a reported increase in the concentration of the postsynaptic 5-HT_{2A} receptors in the prefrontal cortex of suicides compared with non-suicides.[8] This increased binding is reflected in more protein and may be due to elevated gene expression in youth suicide.[9] Elevated 5-HT_{2A} binding has also been reported in the amygdala in depressed suicides. In depressed and non-depressed suicides there is evidence that 5-HT_{2A} receptors are upregulated in the dorsal prefrontal cortex but not the rostral prefrontal cortex.[8]

Platelet studies examine 5-HT_{2A} in living subjects with respect to non-fatal suicide attempt. 5-HT_{2A} receptors, serotonin reuptake sites, and serotonin second messenger systems are present in blood platelets, and changes in these platelet measurements may reflect similar changes in the CNS. Multiple studies have reported higher platelet 5-HT_{2A} receptor numbers in suicide attempters compared with non-attempters and healthy controls.[11]

Studies of second messengers indicate impaired 5-HT_{2A} receptor mediated signal transduction in the prefrontal cortex of suicides, [12] and in platelets 5-HT_{2A} receptor responsivity is significantly blunted in patients with major depression who have made high-lethality suicide attempts compared to depressed patients who have made low-lethality suicide attempts.[13] The implications of such a defect in signal transduction, if present in the brain, would be that although there may be greater density of 5-HT_{2A} receptors, the signal transduced by 5-HT_{2A} receptor activation may be blunted, which would compound deficient serotonergic input as seen in the lower levels of brainstem serotonin and/or 5-HIAA in suicide victims.

Some *postmortem studies* of the postsynaptic 5-HT_{1A} receptor report higher binding in prefrontal cortex and more rostral segments of raphe nuclei, and lower binding in more caudal raphe nuclei, hippocampus, prefrontal cortex, and temporal cortex.[7] Less 5-HT_{1A} autoreceptor gene expression is also reported in the dorsal raphe[14] and would favour higher serotonin neurone firing rates.

Postmortem studies of depressed suicides report fewer 5-HT transporters in prefrontal cortex, hypothalamus, occipital cortex, and brainstem.[15] Moreover, in suicides this deficit appears localized to the ventromedial prefrontal cortex, whereas depressed individuals who died of other causes had lower binding throughout the prefrontal cortex.[16]

The emerging picture from postmortem studies of greater 5-HT_{2A} receptor binding in the frontal cortex of depressed individuals who die by suicide, fewer brainstem 5-HT_{1A} autoreceptors, and fewer serotonin transporters in the cortex, as well as findings of greater tryptophan hydroxylase (the rate-limiting step in serotonin synthesis) immunoreactivity in serotonin nuclei in the brainstem[17] all point to homeostatic changes designed to increase deficient serotonergic transmission evidenced by low 5-HIAA in CSF and

brain, low 5-HT and 5-HIAA in brainstem, and blunted prolactin response to fenfluramine challenge.

Serotonergic dysfunction and suicide endophenotypes. Increased aggression has been associated with suicide and more highly lethal suicide attempts and impulsivity has shown a stronger relationship to non-fatal suicide attempts.[18] Impulsive aggressive traits are potentially part of the diathesis for suicidal behaviour.[1] Reduced activity of the serotonin system has been implicated in impulsive violence and aggression in studies in a variety of paradigms including: Low CSF 5-HIAA in individuals with a lifetime history of aggressive behaviour with personality and other psychiatric disorders;[19, 20] a blunted prolactin response to serotonin-releasing agent fenfluramine in personality disorder patients,[21, 22] and; greater platelet 5-HT_{2A} binding correlated with aggressive behaviour in personality and other psychiatric disorder patients.[23, 24] In a postmortem study of aggression, suicidal behaviour, and serotonergic function a positive relationship between lifetime history of aggression scores and 5-HT_{2A} binding in several regions of prefrontal cortex of individuals who had died by suicide was found.[25]

Positron emission tomography (PET) studies have shown a deficient response to serotonergic challenge in the orbitofrontal cortex, medial frontal, and cingulate regions in individuals with impulsive aggression compared to controls[26, 27] and lower serotonin transporter binding in the anterior cingulate cortex in impulsive aggressive individuals compared to healthy controls.[28] The prefrontal cortex is important in the inhibitory control of behaviour, including impulsive and aggressive behaviour.[29] Thus aggressive/impulsive traits, related to serotonergic dysfunction, are potentially an aspect of the diathesis for suicidal behaviour, whereby aggressive/suicidal behaviours is manifested in response to stressful circumstances or powerful emotions. This tendency might be conceived of as a diminution in natural inhibitory circuits, or as a volatile cognitive decision style.

Noradrenergic system

Within the stress–diathesis model of suicidal behaviour, it is the confluence of stressful events with the diathesis that is thought to precipitate a suicidal act. Thus, investigating the functioning of stress response systems in suicidal individuals is important for elucidating neurobiological concomitants of suicidal behaviour and identifying targets for preventative intervention. The noradrenergic system and the HPA axis are two key stress response systems.

The majority of norepinephrine neurones in the brain are located in the brainstem locus coeruleus. Postmortem studies of suicides have documented fewer noradrenergic neurones in the locus coeruleus.[30] There are also indications of cortical noradrenergic overactivity including lower alpha and high-affinity $beta_1$-adrenergic receptor binding,[31] and lower β-adrenoceptor density and $alpha_2$–adrenergic binding in the prefrontal cortex in individuals who died by suicide.[23] There is some, but not unanimous, evidence from prospective studies of lower levels of CSF 3-methoxy-4-hydroxphenylglycol (MHPG), a metabolite of noradrenaline, in future suicides,[33] although not in those making non-fatal suicide attempts.[34]

Fewer noradrenergic neurones observed in depressed suicides may indicate a lower functional reserve of the noradrenergic system, which if accompanied by an exaggerated stress response with greater release of noradrenaline may result in norepinephrine depletion leading to depression and hopelessness, both of which are contributory factors to suicidal behaviour.

Noradrenergic and HPA axis responses to stress in adulthood appears to be greater in those reporting an abusive experience in childhood.[35] Such individuals are potentially at greater risk in adulthood for major depression and suicidal behaviour. Childhood abuse may be associated with increased risk for depression and suicidal behaviour because of a dysfunctional stress response both via the noradrenergic system and the HPA axis, and secondary effects of norepinephrine depletion and elevated cortisol levels. There is interaction between the noradrenergic system and the stress response activity of the HPA axis with reciprocal neural connections between corticotropin-releasing hormone neurones in the hypothalamic paraventricular nucleus and noradrenergic neurones in human brainstem and the locus coeruleus.[36]

The hypothalamic-pituitary-adrenal (HPA) axis

The hypothalamic-pituitary-adrenal axis is a major stress response system. Major depression is associated with hyperactivity of the HPA axis,[37] and suicidal patients in diagnostically heterogeneous populations exhibit HPA axis abnormalities, most commonly failure to suppress cortisol normally after dexamethasone.[33] We found most future suicides were dexamethasone suppression test (DST) non-suppressors.[33] In mood disorders, DST non-suppressors had a 4.5-fold greater risk of dying by suicide compared with suppressors.[3] Moreover, non-suppression may be characteristic of more serious attempts that result in greater medical damage[38, 39] or the use of violent method in the suicide attempt.[40] In other indices of HPA axis function suicide attempters had attenuated plasma cortisol responses to fenfluramine although that may indicate less serotonin release and not an HPA abnormality,[41, 42] and lower CSF corticotropin-releasing hormone (CRH) compared to non-attempters,[43] though not all studies agree.

Larger pituitary and larger adrenal gland volumes are reported in depressed suicides,[44, 45] and fewer CRH-binding sites in the prefrontal cortex of depressed suicide victims which may mean receptor downregulation due to elevated CRF release.[46]

As with the noradrenergic system, early life adversity appears to have lasting effects on stress response in the HPA axis in adulthood. Abnormalities in HPA axis function have been implicated in poor response to antidepressant treatment, and greater likelihood of relapse in major depression, both of which increase the risk for suicidal acts.[33] Increased anxiety and agitation are another potential pathway whereby abnormal stress response, in both the noradrenergic and HPA axis, contributes to risk for suicidal behaviour.

Other biologic systems

Abnormality in the dopaminergic system has been reported in depressive disorders,[47] however studies of dopaminergic function and suicidal behaviour are relatively few and inconclusive.[48] Low dihydroxyphenylacetic acid levels, indicative of reduced dopamine turnover, in the caudate, putamen, and nucleus accumbens are reported in depressed suicides,[49] although the same group of investigators found no difference in number or affinity of the dopamine

transporters.[50] Accordingly, it is unlikely that the reduced dopamine turnover initially observed in depressed suicides is a result of decreased dopaminergic innervation of those regions. Prospective studies disagree as to whether CSF HVA predicts suicidal behaviour.[51–53]

There is a well-documented relationship between thyroid dysfunction and depression[54] and some studies link thyroid function and suicide. Abnormal thyroid-stimulating hormone (TSH) response to thyrotropin-releasing hormone (TRH) has been observed in individuals who died by suicide in a follow-up study. [55] Abnormal TSH response to challenge tests has also been associated with poor response to antidepressant treatment and a higher relapse rate, which may increase risk for suicidal behaviour.[56]

Neurotrophins are involved in brain development and growth, neuronal functioning, and synaptic plasticity. Lower protein levels and gene expression of brain-derived neurotrophic factor (BDNF) in the prefrontal cortex and hippocampus,[57,58] and less mRNA of nerve growth factor, neurotrophin 3 and neurotrophin 4/5 in the hippocampus[59] are reported postmortem in suicides. Lower plasma BDNF has been reported in MDD suicide attempters compared to MDD non-attempters and healthy controls.[60]

Suicide is more common in groups with very low cholesterol levels or after cholesterol lowering by diet (see[61] for a review). This relationship between cholesterol and suicide may be mediated by serotonergic function, as studies of non-human primates on a low-fat diet found lower serotonergic activity and increased aggressive behaviours.[62] Long chain polyunsaturated fatty acids, particularly omega-3, may also be a mediating factor in the relationship between low cholesterol and increased risk for depression and suicide.[63] Lower docosahexaenoic acid percentage of total plasma polyunsaturated fatty acids and a higher omega-6/omega-3 ratio predicted depressed individuals who made a suicide attempt during a 2-year follow-up,[64] and lower eicosapentaenoic acid is found in red blood cells of suicide attempters compared to controls.[65]

Neurobiology, genetics, and suicidal behaviour

Family, twin, and adoption studies support a genetic contribution to suicidal behaviour independent of psychiatric disorder (see Brent and Mann for a review),[66] and genetic studies have sought to determine the responsible genes for suicide and suicide attempt though linkage and SNP association studies. Candidate genes for most studies were selected based on evidence from neurobiological studies in suicide, as a result of which the serotonergic system has been most extensively investigated. A tri-allelic polymorphism in the serotonin transporter promotor has two alleles with lower transcriptional activity and fewer transporters. In varied psychiatric populations, despite some negative findings, the S, or more common lower-expressing allele, has been associated with suicide and with suicide attempts, particularly violent or high-lethality attempts.[67] Functional MRI studies find greater amygdala activation in individuals with the SS genotype when they are exposed to negative stimuli such as angry or fearful faces, negative words, or aversive pictures (see Brown and Hariri 2006 for a review).[68] The amygdala is densely innervated by serotonergic neurones and 5-HT receptors are abundant, and plays a central role in emotional regulation and memory. Excessive responses to emotionally negative events such as abuse, may be over-encoded and contribute to stress-sensitivity in adulthood and thereby to major depression after stress and even suicidal behaviour.

Other genetic studies of the serotonergic system including the 5-HT_{1A}, 5-HT_{2A}, 5-HT_{1B} and other serotonin receptors have largely reported negative results, although there have been some positive findings for the 5-HT_{2A} 102C allele and attempted suicide or suicidal ideation.[67] For tryptophan hydroxylase (TPH1 and TPH2 are two forms of TPH with TPH1 only expressed in the brain during development), associations are reported with suicide and suicide attempt and TPH1 SNPs, however multiple negative findings have also been reported,[67] while haplotype and SNP studies suggest the involvement of the TPH2 gene in suicide and suicide attempt, however again not all studies agree.[67] Monoamine oxidase (MAO-A) plays a key role in metabolism of amines. Low MAO activity results in elevated levels of serotonin, norepinephrine, and dopamine in the brain. The MAO-A gene has functional variable number tandem repeat however no association has been found between this uVNTR and suicidal behaviour, although there is some indication that it may be related to aggression[67] and it is linked to the impact of adversity in childhood on adult antisocial behaviour and trait impulsiveness.[69,70]

Genetic studies in dopaminergic system, noradrenergic system, BDNF, and GABA are few and generally negative,[67] although there are reports of positive association of the catechol-*O*-methyltransferase (COMT), a major catecholamine-catabolic enzyme, gene in Finnish and Caucasian suicide attempters[71] and in Japanese suicides.[72] Inconsistent findings in genetic studies of suicidal behaviour may be due to the complexity of the suicide phenotype, gene–gene interactions, the presence of multiple psychiatric disorders, population racial differences, possible epigenetic effects, and the influence of gene/environment interactions. Nonetheless, new microarray technologies that test expression of thousands of genes simultaneously allowing better gene coverage, and haplotype mapping approaches offer promise for future investigation. Other options include examining more basic endophenotypes such as mood regulation and decision-making.

Genes and environment

Early life stress in conjunction with genetic vulnerability can have enduring effects into adulthood and affect psychopathology and the functioning of biological systems including the serotonergic and stress-response systems (see Mann and Currier 2006).[73] For example monkeys exposed to maternal deprivation in infancy and having the 5-HTTLPR lower expressing S allele in the serotonin transporter gene manifest a lowering of CSF 5-HIAA that persists into adulthood.[74] In 6-month-old macaque monkeys exposed to social stress, those with the S allele had a higher ACTH response, an HPA axis hormone related to stress response, compared with those without that allele and to S allele animals who were maternally reared.[75] Thus the low-expressing S allele not only increased vulnerability to stress in development, but early life stress may further interact with genotype to lower serotonergic function and to increase sensitivity to stressful events later in life, both of which are risk factors for suicidal behaviour.

In human studies, individuals who had experienced childhood maltreatment, those with the low-expressing S allele were at risk for

suicidal ideation and suicide attempt,[76] and those with a lower expressing variant of the MAO-A gene were more likely to manifest antisocial behaviour and more impulsivity as adults.[69, 70]

Future directions

There is much still to be learned about the biologic aetiology of suicidal behaviour and the pathways and mechanisms through which biologic dysfunction is involved in suicidal acts. New techniques for imaging the brain, identification of basic intermediate phenotypes and denser gene markers will contribute to elucidating the biological factors and mechanisms involved in suicide and attempted suicide, and identifying potential targets for prevention.

Further information

Mann, J.J. (2003). Neurobiology of suicidal behaviour. *Nature Reviews Neuroscience*, **4**(10), 819–28.

Pandey, G.N. (1997). Altered serotonin function in suicide. Evidence from platelet and neuroendocrine studies. *Annals of the New York Academy of Sciences*, **836**, 182–200.

Mann, J.J. and Currier, D. (2007). A review of prospective studies of biologic predictors of suicidal behavior in mood disorders. *Archives of Suicide Research: Official Journal of the International Academy for Suicide Research*, **11**(1), 3–16.

Bondy, B., Buettner, A., and Zill, P. (2006). Genetics of suicide. *Molecular Psychiatry*, **11**(4), 336–51.

References

1. Mann, J.J., Waternaux, C., Haas, G.L., *et al.* (1999). Toward a clinical model of suicidal behavior in psychiatric patients. *The American Journal of Psychiatry*, **156**(2), 181–9.
2. Åsberg, M., Thorén, P., Träskman, L., Bertilsson, L., and Ringberger, V. (1976). "Serotonin depression"—A biochemical subgroup within the affective disorders? *Science*, **191**, 478–80.
3. Mann, J.J., Currier, D., Stanley, B., *et al.* (2006). Can biological tests assist prediction of suicide in mood disorders? *The International Journal of Neuropsychopharmacology*, **9**(4), 465–74.
4. Mann, J.J., Underwood, M.D., and Arango, V. (1996). Postmortem studies of suicide victims. In *Biology of schizophrenia and affective disease* (ed. S.J. Watson), pp. 197–220. American Psychiatric Press, Washington, DC.
5. Mann, J.J. (1998). The neurobiology of suicide. *Nature Medicine*, **4**(1), 25–30.
6. Arango, V., Underwood, M.D., Gubbi, A.V., *et al.* (1995). Localized alterations in pre- and postsynaptic serotonin binding sites in the ventrolateral prefrontal cortex of suicide victims. *Brain Research*, **688**(1–2), 121–33.
7. Kamali, M., Oquendo, M.A., and Mann, J.J. (2001). Understanding the neurobiology of suicidal behavior. *Depression and Anxiety*, **14**(3), 164–76.
8. Stockmeier, C.A. (2003). Involvement of serotonin in depression: evidence from postmortem and imaging studies of serotonin receptors and the serotonin transporter. *Journal of Psychiatric Research*, **37**(5), 357–73.
9. Pandey, G.N., Dwivedi, Y., Rizavi, H.S., *et al.* (2002). Higher expression of serotonin 5-HT(2A) receptors in the postmortem brains of teenage suicide victims. *The American Journal of Psychiatry*, **159**(3), 419–29.
10. Hrdina, P.D., Demeter, E., Vu, T.B., *et al.* (1993). 5-HT uptake sites and 5-HT_2 receptors in brain of antidepressant-free suicide victims/depressives: increase in 5-HT_2 sites in cortex and amygdala. *Brain Research*, **614**, 37–44.
11. Pandey, G.N. (1997). Altered serotonin function in suicide. Evidence from platelet and neuroendocrine studies. *Annals of the New York Academy of Sciences*, **836**, 182–200.
12. Pandey, G.N., Dwivedi, Y., Pandey, S.C., *et al.* (1999). Low phosphoinositide-specific phospholipase C activity and expression of phospholipase C beta1 protein in the prefrontal cortex of teenage suicide subjects. *The American Journal of Psychiatry*, **156**(12), 1895–901.
13. Malone, K.M., Ellis, S.P., Currier, D., *et al.* (2007). Platelet 5-HT2A receptor subresponsivity and lethality of attempted suicide in depressed in-patients. *The International Journal of Neuropsychopharmacology*, **10**(3): 335–430.
14. Arango, V., Underwood, M.D., Boldrini, M., *et al.* (2001). Serotonin 1A receptors, serotonin transporter binding and serotonin transporter mRNA expression in the brainstem of depressed suicide victims. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, **25**(6), 892–903.
15. Purselle, D.C. and Nemeroff, C.B. (2003). Serotonin transporter: a potential substrate in the biology of suicide. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, **28**(4), 613–19.
16. Mann, J.J., Huang, Y.Y., Underwood, M.D., *et al.* (2000). A serotonin transporter gene promoter polymorphism (5-HTTLPR) and prefrontal cortical binding in major depression and suicide. *Archives of General Psychiatry*, **57**(8), 729–38.
17. Boldrini, M., Underwood, M.D., Mann, J.J., *et al.* (2005). More tryptophan hydroxylase in the brainstem dorsal raphe nucleus in depressed suicides. *Brain Research*, **1041**(1), 19–28.
18. Oquendo, M.A., Galfalvy, H., Russo, S., *et al.* (2004). Prospective study of clinical predictors of suicidal acts after a major depressive episode in patients with major depressive disorder or bipolar disorder. *The American Journal of Psychiatry*, **161**(8), 1433–41.
19. Brown, G.L. and Goodwin, F.K. (1986). Cerebrospinal fluid correlates of suicide attempts and aggression. *Annals of the New York Academy of Sciences*, **487**, 175–88.
20. Stanley, B., Molcho, A., Stanley, M., *et al.* (2000). Association of aggressive behavior with altered serotonergic function in patients who are not suicidal. *The American Journal of Psychiatry*, **157**(4), 609–14.
21. Coccaro, E.F., Siever, L.J., Klar, H.M., *et al.* (1989). Serotonergic studies in patients with affective and personality disorders. Correlates with suicidal and impulsive aggressive behavior. *Archives of General Psychiatry*, **46**, 587–99.
22. New, A.S., Trestman, R.F., Mitropoulou, V., *et al.* (2004). Low prolactin response to fenfluramine in impulsive aggression. *Journal of Psychiatric Research*, **38**(3), 223–30.
23. Coccaro, E.F., Kavoussi, R.J., Sheline, Y.I., *et al.* (1997). Impulsive aggression in personality disorder correlates with platelet 5-HT_{2A} receptor binding. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, **16**(3), 211–16.
24. McBride, P.A., Brown, R.P., DeMeo, M., *et al.* (1994). The relationship of platelet 5-HT_2 receptor indices to major depressive disorder, personality traits, and suicidal behavior. *Biological Psychiatry*, **35**, 295–308.
25. Oquendo, M.A., Russo, S.A., Underwood, M.D., *et al.* (2006). Higher postmortem prefrontal 5-HT2A receptor binding correlates with lifetime aggression in suicide. *Biological Psychiatry*, **59**(3), 235–43.
26. Siever, L.J., Buchsbaum, M.S., New, A.S., *et al.* (1999). D,1-Fenfluramine response in impulsive personality disorder assessed with [^{18}F]fluorodeoxyglucose positron emission tomography. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, **20**(5), 413–23.
27. New, A.S., Hazlett, E.A., Buchsbaum, M.S., *et al.* (2002). Blunted prefrontal cortical 18fluorodeoxyglucose positron emission tomography response to meta-chlorophenylpiperazine in impulsive aggression. *Archives of General Psychiatry*, **59**(7), 621–9.

28. Frankle, W.G., Lombardo, I., New, A.S., *et al.* (2005). Brain serotonin transporter distribution in subjects with impulsive aggressivity: a positron emission study with [^{11}C]McN 5652. *The American Journal of Psychiatry*, **162**(5), 915–23.
29. de Almeida, R.M., Rosa, M.M., Santos, D.M., *et al.* (2006). 5-HT(1B) receptors, ventral orbitofrontal cortex, and aggressive behavior in mice. *Psychopharmacology*, **185**(4), 441–50.
30. Arango, V., Underwood, M.D., and Mann, J.J. (1996). Fewer pigmented locus coeruleus neurons in suicide victims: preliminary results. *Biological Psychiatry*, **39**, 112–20.
31. Arango, V., Ernsberger, P., Sved, A.F., *et al.* (1993). Quantitative autoradiography of a_1- and a_2-adrenergic receptors in the cerebral cortex of controls and suicide victims. *Brain Research*, **630**, 271–82.
32. De Paermentier, F., Cheetham, S.C., Crompton, M.R., *et al.* (1990). Brain b-adrenoceptor binding sites in antidepressant-free depressed suicide victims. *Brain Research*, **525**, 71–7.
33. Mann, J.J. and Currier, D. (2007). A review of prospective studies of biologic predictors of suicidal behavior in mood disorders. *Archives of Suicide Research: Official Journal of the International Academy for Suicide Research*, **11**(1), 3–16.
34. Lester, D. (1995). The concentration of neurotransmitter metabolites in the cerebrospinal fluid of suicidal individuals: a meta-analysis. *Pharmacopsychiatry*, **28**(2), 45–50.
35. Heim, C. and Nemeroff, C.B. (2001). The role of childhood trauma in the neurobiology of mood and anxiety disorders: preclinical and clinical studies. *Biological Psychiatry*, **49**(12), 1023–39.
36. Austin, M.C., Rice, P.M., Mann, J.J., *et al.* (1995). Localization of corticotropin-releasing hormone in the human locus coeruleus and pedunculopontine tegmental nucleus: an immunocytochemical and in situ hybridization study. *Neuroscience*, **64**(3), 713–27.
37. Carroll, B.J., Feinberg, M., Greden, J.F., *et al.* (1981). A specific laboratory test for the diagnosis of melancholia. Standardization, validation, and clinical utility. *Archives of General Psychiatry*, **38**, 15–22.
38. Norman, W.H., Brown, W.A., Miller, I.W., *et al.* (1990). The dexamethasone suppression test and completed suicide. *Acta Psychiatrica Scandinavica*, **81**(2), 120–5.
39. Coryell, W. (1990). DST abnormality as a predictor of course in major depression. *Journal of Affective Disorders*, **19**(3), 163–9.
40. Roy, A. (1992). Hypothalamic-pituitary-adrenal axis function and suicidal behavior in depression. *Biological Psychiatry*, **32**, 812–16.
41. Duval, F., Mokrani, M.C., Correa, H., *et al.* (2001). Lack of effect of HPA axis hyperactivity on hormonal responses to d- fenfluramine in major depressed patients: implications for pathogenesis of suicidal behaviour. *Psychoneuroendocrinology*, **26**(5), 521–37.
42. Malone, K.M., Corbitt, E.M., Li, S., *et al.* (1996). Prolactin response to fenfluramine and suicide attempt lethality in major depression. *The British Journal of Psychiatry: The Journal of Mental Science*, **168**, 324–9.
43. Brunner, J., Stalla, G.K., Stalla, J., *et al.* (2001). Decreased corticotropin-releasing hormone (CRH) concentrations in the cerebrospinal fluid of eucortisolemic suicide attempters. *Journal of Psychiatric Research*, **35**(1), 1–9.
44. Szigethy, E., Conwell, Y., Forbes, N.T., *et al.* (1994). Adrenal weight and morphology in victims of completed suicide. *Biological Psychiatry*, **36**(6), 374–80.
45. Dumser, T., Barocka, A., and Schubert, E. (1998). Weight of adrenal glands may be increased in persons who commit suicide. *The American Journal of Forensic Medicine and Pathology: Official Publication of the National Association of Medical Examiners*, **19**(1), 72–6.
46. Nemeroff, C.B., Owens, M.J., Bissette, G., *et al.* (1988). Reduced corticotropin releasing factor binding sites in the frontal cortex of suicide victims. *Archives of General Psychiatry*, **45**, 577–9.
47. Dailly, E., Chenu, F., Renard, C.E., *et al.* (2004). Dopamine, depression and antidepressants. *Fundamental and Clinical Pharmacology*, **18**(6), 601–7.
48. Mann, J.J. (2003). Neurobiology of suicidal behaviour. *Nature Reviews Neuroscience*, **4**(10), 819–28.
49. Bowden, C., Cheetham, S.C., Lowther, S., *et al.* (1997). Reduced dopamine turnover in the basal ganglia of depressed suicides. *Brain Research*, **769**(1), 135–40.
50. Bowden, C., Theodorou, A.E., Cheetham, S.C., *et al.* (1997). Dopamine D_1 and D_2 receptor binding sites in brain samples from depressed suicides and controls. *Brain Research*, **752**, 227–33.
51. Roy, A., De Jong, J., and Linnoila, M. (1989). Cerebrospinal fluid monoamine metabolites and suicidal behavior in depressed patients. A 5-year follow-up study. *Archives of General Psychiatry*, **46**, 609–12.
52. Engstrom, G., Alling, C., Blennow, K., *et al.* (1999). Reduced cerebrospinal HVA concentrations and HVA/5-HIAA ratios in suicide attempters. Monoamine metabolites in 120 suicide attempters and 47 controls. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, **9**(5), 399–405.
53. Placidi, G.P., Oquendo, M.A., Malone, K.M., *et al.* (2001). Aggressivity, suicide attempts, and depression: relationship to cerebrospinal fluid monoamine metabolite levels. *Biological Psychiatry*, **50**(10), 783–91.
54. Jackson, I.M. (1998). The thyroid axis and depression. *Thyroid: Official Journal of the American Thyroid Association*, **8**(10), 951–6.
55. Linkowski, P., Van Wettere. J.P., Kerkhofs, M., *et al.* (1984). Violent suicidal behavior and the thyrotropin-releasing hormone-thyroid-stimulating hormone test: a clinical outcome study. *Neuropsychobiology*, **12**(1), 19–22.
56. Targum, S.D. (1984). Persistent neuroendocrine dysregulation in major depressive disorder: a marker for early relapse. *Biological Psychiatry*, **19**(3), 305–18.
57. Dwivedi, Y., Rizavi, H.S., Conley, R.R., *et al.* (2003). Altered gene expression of brain-derived neurotrophic factor and receptor tyrosine kinase B in postmortem brain of suicide subjects. *Archives of General Psychiatry*, **60**(8), 804–15.
58. Karege, F., Vaudan, G., Schwald, M., *et al.* (2005). Neurotrophin levels in postmortem brains of suicide victims and the effects of antemortem diagnosis and psychotropic drugs. *Brain Research Molecular Brain Research*, **136**(1–2), 29–37.
59. Dwivedi, Y., Mondal, A.C., Rizavi, H.S., *et al.* (2005). Suicide brain is associated with decreased expression of neurotrophins. *Biological Psychiatry*, **58**(4), 315–24.
60. Kim, Y.K., Lee, H.P., Won, S.D., *et al.* (2007). Low plasma BDNF is associated with suicidal behavior in major depression. *Progress in Neuropsychopharmacology and Biological Psychiatry*, **31**(1), 78–85.
61. Golomb, B.A. (1998). Cholesterol and violence: is there a connection? *Annals of Internal Medicine*, **128**, 478–87.
62. Muldoon, M.F., Rossouw, J.E., Manuck, S.B., *et al.* (1993). Low or lowered cholesterol and risk of death from suicide and trauma. *Metabolism: Clinical and Experimental*, **42**(Suppl. 1), 45–56.
63. Brunner, J., Parhofer, K.G., Schwandt, P., *et al.* (2002). Cholesterol, essential fatty acids, and suicide. *Pharmacopsychiatry*, **35**(1), 1–5.
64. Sublette, M.E., Hibbeln, J.R., Galfalvy, H., *et al.* (2006). Omega-3 polyunsaturated essential fatty acid status as a predictor of future suicide risk. *American Journal of Psychiatry*, **163**(6), 1100–2.
65. Huan, M., Hamazaki, K., Sun, Y., *et al.* (2004). Suicide attempt and *n*–3 fatty acid levels in red blood cells: a case control study in China. *Biological Psychiatry*, **56**(7), 490–6.
66. Brent, D.A. and Mann, J.J. (2005). Family genetic studies, suicide, and suicidal behavior. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics*, **133**(1), 13–24.
67. Bondy, B., Buettner, A., and Zill, P. (2006). Genetics of suicide. *Molecular Psychiatry*, **11**(4), 336–51.

68. Brown, S.M. and Hariri, A.R. (2006). Neuroimaging studies of serotonin gene polymorphisms: exploring the interplay of genes, brain, and behavior. *Cognitive Affective and Behavioral Neuroscience*, **6**(1), 44–52.
69. Caspi, A., McClay, J., Moffitt, T.E., *et al.* (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, **297**(5582), 851–4.
70. Huang, Y.Y., Cate, S.P., Battistuzzi, C., *et al.* (2004). An association between a functional polymorphism in the monoamine oxidase a gene promoter, impulsive traits and early abuse experiences. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, **29**(8), 1498–505.
71. Nolan, K.A., Volavka, J., Czobor, P., *et al.* (2000). Suicidal behavior in patients with schizophrenia is related to COMT polymorphism. *Psychiatric Genetics*, **10**(3), 117–24.
72. Ono, H., Shirakawa, O., Nushida, H., *et al.* (2004). Association between catechol-*O*-methyltransferase functional polymorphism and male suicide completers. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, **29**(7), 1374–7.
73. Mann, J.J. and Currier, D. (2006). Effects of genes and stress on the neurobiology of depression. *International Review of Neurobiology*, **73**, 153–89.
74. Bennett, A.J., Lesch, K.P., Heils, A., *et al.* (2002). Early experience and serotonin transporter gene variation interact to influence primate CNS function. *Molecular Psychiatry*, **7**(1), 118–22.
75. Barr, C.S., Newman, T.K., Shannon, C., *et al.* (2004). Rearing condition and rh5-HTTLPR interact to influence limbic-hypothalamic-pituitary-adrenal axis response to stress in infant macaques. *Biological Psychiatry*, **55**(7), 733–8.
76. Caspi, A., Sugden, K., Moffitt, T.E., *et al.* (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*, **301**(5631), 386–9.

4.15.4 Treatment of suicide attempters and prevention of suicide and attempted suicide

Keith Hawton and Tatiana Taylor

Introduction

In considering the treatment and prevention of suicidal behaviour account should be taken of recent trends in suicide and attempted suicide, particularly in individual countries. These have been reviewed in other chapters in this section. The term 'attempted suicide' is used in this chapter and includes any act of non-fatal self-poisoning or self-injury, irrespective of motive or intention.

Suicide prevention has been incorporated within the World Health Organization Health for All strategy[1] and has received substantial support from the United Nations.[2] Furthermore, in recent years several countries have developed national suicide prevention programmes. Increased suicide rates in young people have probably acted as a stimulus behind this trend. However, suicide rates in most countries remain higher in older populations and prevention programmes must include this increasingly larger sector of society.

Treatment of suicide attempters

Suicide attempts occur for a wide range of reasons. In many cases the primary aim is not death but some other outcome, such as demonstrating distress to other people, seeking a change in other people's behaviour or temporary escape.[3] This means that a broad range of treatments are required since the needs of individual patients will vary widely (Table 4.15.4.1).

Factors relevant to treatment needs in suicide attempts

(a) Repetition of attempts and risk of suicide

Repetition of attempts is common, with 15 to 25 per cent repeating suicidal acts within a year, and is associated with a greater risk of eventual suicide.[4] The frequency of suicide following attempted suicide varies from country to country,[5] depending on the overall characteristics of the patient population and the rate of suicide in the general population. Prevention of repetition of suicidal behaviour and especially of suicide is a major aim in treating suicide attempters.

(b) Psychiatric and personality disorders

A range of psychiatric disorders are found in suicide attempters.[6] Depression and alcohol abuse are particularly common. In addition, substantial proportions of patients have personality disorders. While treatment directed at the underlying causes of such disorders, where possible, will be important in managing attempted suicide patients, often the disorders themselves will require specific treatment.

(c) Life events and difficulties

Certain problems are particularly common in suicide attempters, including difficulties in interpersonal relationships, especially with partners and with other family members, employment problems, particularly in males, and financial difficulties. Life events, especially disruption in a relationship with a partner, frequently precede suicidal acts.[7]

(d) Poor problem-solving skills

Many suicide attempters have difficulties in problem-solving, particularly in dealing with difficulties in interpersonal relationships.[8] These difficulties are more marked in suicide attempters than in patients with psychiatric disorders who have not carried out a suicidal act.

Table 4.15.4.1 Factors relevant to treatment needs in suicide attempters

Risk of repetition of attempts
Risk of suicide
Psychiatric disorder (especially depression and substance abuse)
Personality disorders
Life events and difficulties
Poor problem-solving skills
Impulsivity and aggression
Hopelessness
Low self-esteem
Motivational problems and poor compliance with treatment

(e) Impulsivity and aggression

There is a strong link between suicidal behaviour and both impulsivity and aggression. There is also accumulating evidence that hypofunction of brain serotonergic systems is linked to aggression (and possibly impulsivity) and also to suicidal behaviour[9] (see Chapter 4.15.3). It is unclear whether this represents a state phenomenon associated with psychiatric disturbance or a trait phenomenon, but current evidence points towards the latter.

(f) Hopelessness and low self-esteem

Hopelessness, or pessimism about the future, which has been shown to be a key factor linking depression with suicidal acts, is an important predictor of repetition of suicidal behaviour, and a risk factor for eventual suicide.[10] Low self-esteem is another important characteristic associated with suicidal behaviour. There is likely to be a link between low self-esteem and a tendency to experience hopelessness when faced by adverse circumstances.

(g) Motivational problems and poor compliance with treatment

Management of suicide attempters is complicated by the fact that some patients appear to be poorly motivated to engage in aftercare. This is also likely to affect compliance with treatments. The style of organization of general hospital psychiatric services (including continuity of care) and the attitudes of clinical staff may be important factors determining whether patients engage in aftercare.

General overview of treatments

Treatments for suicide attempters include both psychosocial and pharmacological approaches. While these are considered separately below, in some patients both will be appropriate. This might be the case, for example, if a patient suffers from depression with biological features in the setting of employment and financial difficulties, when treatment with an antidepressant might be combined with problem-solving therapy.

Psychosocial treatments

A range of psychosocial therapies have been evaluated in suicide attempters in randomized controlled clinical trials. The efficacy of these approaches has been examined in a systematic review of the worldwide literature.[11] The findings from this review and some further studies are summarized below.

(a) Problem-solving

Meta-analysis of the results of trials that have been conducted so far to evaluate the effectiveness of brief problem-solving therapy (see Chapter 6.3.1) compared with treatment as usual indicates a trend towards reduction of repetition of self-harm episodes, but the total numbers of subjects and trials have precluded a definitive result. However, evidence of other positive outcomes, such as reduced levels of depression and hopelessness and improvement in problems, has been convincingly demonstrated in these studies.[12] This approach is useful, either used alone or in the context of other treatment. It is reasonably easily taught and can be used by clinicians from different professional backgrounds.

(b) Psychotherapy

Cognitive behaviour therapy, combined with care management has recently been shown to be effective in reducing frequency of suicide attempts and in producing other positive outcomes.[13] A brief psychological intervention combined with provision of a treatment manual seems to be less effective in the treatment of patients with repeated attempts.[14]

Two trials have been conducted in which an intensive form of psychological treatment known as dialectical behaviour therapy was evaluated.[15,16] Female patients with borderline personality disorders who had a history of repeated self-harm were offered a year of individual and group cognitive behavioural therapy aimed at addressing the patients' problems of motivation and strengthening their behavioural skills, particularly in relation to interpersonal difficulties. Compared with routine care this approach seems to result in a reduction in repetition of self-harm as well as a number of other positive outcomes. Further evaluation of this approach is required to determine if it is effective in male patients and in adolescents, and whether it can be delivered in an abbreviated form. While it is a labour-intensive approach, it appears to be helpful for what is a particularly difficult group of patients.

(c) Outreach

Several trials have been conducted to assess the impact of community outreach, either for all patients or for those that have not attended treatment sessions. Some of these studies have included relatively intensive treatment programmes. Overall these studies indicate that some form of outreach may improve outcome in terms of reducing repetition of attempted suicide. In one study, nurse home visits to encourage non-attending participants to attend outpatient appointments resulted in a significantly greater number of appointments attended as compared to the control group and there was a near significant reduction in the rate of repetition of suicide attempts during the year after study entry.[17] In other studies, telephone contact,[18] and contacting patients regularly by post[19] have also produced promising results. Outreach combined with specific treatment may be useful, perhaps reserved for those who are poorly compliant with aftercare.

(d) Provision of emergency cards

In the United Kingdom there has recently been interest in providing suicide attempters with cards which indicate how they might get emergency help at times of crisis. Two initial, relatively small, studies of this approach, one involving adults and the other young adolescents, produced encouraging results but a larger evaluation did not show the cards to be effective.[20] Provision of emergency cards requires there to be a 24 h service to deal with emergency calls. They might be thought helpful in a minority of cases, but there needs to be careful selection of patients who are offered this facility because of risk of it possibly being abused.

Pharmacological treatments

There have been relatively few treatment trials evaluating the effectiveness of pharmacological agents in suicide attempters. This perhaps reflects the problems of compliance with therapy, which were noted earlier, and risk of overdose.

(a) Antidepressants

A trial in the Netherlands in which paroxetine was compared with placebo in patients who were all repeaters of self-harm but who did not suffer from current depressive disorder showed apparent benefits for a subgroup of patients who received paroxetine, namely

those who had a history of between one and four episodes of self-harm. Patients with a history of five or more episodes did not seem to benefit.[21] The findings of this study are clearly of interest (although *post hoc* subgroup analyses of this kind must be treated with caution). Recently there has been much attention to the risk of antidepressants increasing suicidal ideation and acts in adolescents.[22] Also, it has become clear that there is increased risk with all types of antidepressants during the initial period of treatment.[23] These findings have highlighted the need to be cautious in the use of antidepressants, to provide early follow-up after initiating therapy, and to consider combining antidepressant treatment with other therapies, especially for adolescents (in whom only fluoxetine is currently recommended for the treatment of depression).

(b) Neuroleptics

A trial in which the depot neuroleptic flupenthixol was administered monthly in a dose of 20 mg for 6 months to repeaters of self-harm and compared with placebo in similar patients appeared to show that the active drug was effective in reducing the recurrence of self-harm.[24] While this type of study requires replication, perhaps using one of the atypical oral neuroleptics in patients who frequently repeat self-harm may be worth trying.

(c) Lithium

A systematic review of trials of lithium therapy versus a range of other drugs and placebo in patients with affective disorders has shown convincing evidence that lithium may prevent suicide.[25] It is not known if it may be anti-suicidal in other groups of patients.

Management in clinical practice

Before a treatment plan can be formulated a careful assessment must be carried out. In conducting the assessment the clinician needs to try and establish good rapport with the patient and be sensitive to the patient's preferences. The key factors that should be covered during the assessment are listed in Table 4.15.4.2. For the purpose of formulating a management plan it is particularly useful to draw up a problem list which summarizes the patient's current difficulties. This should be done in active collaboration with the patient as far as possible. Qualitative studies have shown that patients appreciate clinicians and other staff keeping them well informed of their mental and physical status and including them in decision-making in regard to their care.[26]

(a) Assessment

During the assessment it is crucial to estimate the risk of suicide or another non-fatal attempt. However, accurate assessment is far from easy. Risk factors for suicide following attempted suicide are shown in Table 4.15.4.3. Because suicide is uncommon, the predictive value of the items is limited. One predictor of suicide risk is the degree of suicidal intent involved in the current attempt (see Table 4.15.4.4). Clinicians should consider the use of the valuable Beck Suicide Intent scale.[27] Factors known to be associated with risk of a further attempt are listed in Table 4.15.4.5. It should be noted that while individuals who score positive on several of these factors will have considerably increased risk of repetition, a substantial proportion of those who repeat will not show these characteristics, i.e. the predictive value of scales to predict repetition is modest.

Table 4.15.4.2 The assessment of attempted-suicide patients

Factors that should be covered
Life events that preceded the attempt
Motives for the act, including suicidal intent and other reasons
Problems faced by the patient
Psychiatric disorder
Personality traits and disorder
Alcohol and drug misuse
Family and personal history
Current circumstances
Social (e.g. extent of social relationships)
Domestic (e.g. living alone or with others)
Occupation (e.g. whether employed)
Psychiatric history, including previous suicide attempts
Assessments that should be made
Risk of a further attempt
Risk of suicide
Coping resources and supports
What treatment is appropriate to the patient's needs
Motivation of patient (and significant others where appropriate, to engage in treatment)

Table 4.15.4.3 Factors associated with risk of suicide after attempted suicide

Older age
Male gender
Unemployed or retired
Separated, divorced or widowed
Living alone
Poor physical health
Psychiatric disorder (particularly depression, alcoholism, schizophrenia, and 'sociopathic' personality disorder)
High suicidal intent in current episode
Violent method involved in current attempt (e.g. attempted hanging, shooting, jumping)
Leaving a note
Previous attempt(s) (including repetitive self-injury)

Table 4.15.4.4 Factors that suggest high suicidal intent

Act carried out in isolation
Act timed so that intervention unlikely
Precautions taken to avoid discovery
Preparations made in anticipation of death (e.g. making will, organizing insurance)
Preparations made for the act (e.g. purchasing means, saving up tablets)
Communicating intent to others beforehand
Extensive premeditation
Leaving a note
Note alerting potential helpers after the act
Subsequent admission of suicidal intent

Table 4.15.4.5 Factors associated with risk of repetition of attempted suicide

Previous attempt(s)
Depression
Personality disorder
Alcohol or drug abuse
Previous psychiatric treatment
Unemployment
Lower socio-economic status
Criminal record
History of violence
Age 25–54 years
Single, divorced, or separated

(b) Treatment

The treatment plan should be drawn up on the basis of the patient's needs and risks. Inpatient psychiatric treatment will usually be indicated for patients with severe psychiatric disorders, especially where immediate risk of suicide appears to be high.

(i) Psychiatric disorders

Major psychiatric disorders should be treated in the usual way, but with particular care about use of medication which might be toxic in overdose. Specific treatment should be provided for alcohol and drug abuse; indeed, a suicide attempt is sometimes the first occasion that abuse may come to clinical attention.

(ii) Community therapy

Most patients can be managed in the community. Brief psychological therapy, with a focus on problem-solving will be appropriate for those patients who have clear problems, such as in interpersonal relationships, employment, or social adjustment. Some form of outreach (e.g. home visiting, telephone contact) may be helpful to increase the proportion of patients who engage in treatment, but is not necessary for most patients. Outreach may be essential in the treatment of patients in remote rural areas in developing countries.

If possible there should be continuity of therapy in terms of the same person who saw the patient in hospital after their attempt providing aftercare as this is likely to result in better compliance with therapy. Longer term cognitive behavioural therapy or dynamic psychotherapy may be required for patients whose attempts are related to traumas, such as sexual abuse, or to personality disorder. People who are repeaters of suicide attempts may also require more intensive treatment, especially those who frequently repeat. If resources permit, the use of a programme based on dialectical behaviour therapy, possibly using a group format for at least part of treatment, might be considered.

(iii) Adolescents

Family therapy may be required for young adolescents, and also for patients with difficulties in relation to children. Group therapy may be helpful for adolescents who are repeaters of self-harm.[28]

Prevention of suicide and attempted suicide

A widely diverse group of individuals are at risk of suicidal behaviour and it occurs in relation to a wide range of problems and situations. For example, suicide may occur in the context of long-term difficulties extending back to childhood, acute severe life events or, and perhaps most importantly, acute or long-term and relapsing mental illness. Because of this the range of potential prevention strategies is also considerable.

Suicide prevention programmes have been established in many countries. This is to be welcomed, not only because of the potential benefits in terms of suicide prevention, but also because of the likely benefits for the broader population of individuals with mental health problems. When considering prevention strategies, it is important to be aware of and sensitive towards issues relating to culture and ethnicity. For example, while suicide rates are generally relatively low in young females in the United Kingdom, this is not the case in young Asian females of the Hindu faith, in which rates appear to be relatively high and greater than those of their male peers.[29] The issues surrounding such deaths are often related to cultural clashes regarding values and expectations between young Asian females and their parents.

While ethical issues in relation to suicide prevention are not dealt with in detail in this chapter, they are none the less highly important.[30] Opinions will vary, for example, about whether suicide should always be prevented. This particularly relates to suicides occurring in the context of terminal and/or painful physical illnesses, and relapsing and debilitating mental illnesses. The ethics of suicide prevention overlap those of assisted suicide and euthanasia. Psychiatrists are increasingly likely to be drawn into debate and controversy about the ethical aspects of these issues, particularly in relation to severe and chronic mental illness, and mental health aspects of assisted suicide and euthanasia in people with severe physical illnesses.

General principles of prevention

Broadly there are two approaches to suicide prevention.[31,32] As described by Rose[33] in the context of prevention of health problems in general, one can distinguish between population approaches, which aim to decrease risk in the population as a whole, and high-risk group strategies, in which specific groups that are at increased risk are targeted. High-risk group strategies often appear more attractive and realistic. However, risk factors for many disorders are widely spread in the population and so the high-risk strategy tends to exclude a large number of people at moderate risk and is often ineffective in reducing the burden of a disease at the population level. Conversely, population strategies may appear more difficult to achieve but are more likely to be effective in reducing population levels of disease (see also Chapter 7.4). The main population and high-risk group strategies in the prevention of suicide and attempted suicide which are considered here are shown in Table 4.15.4.6.

It is unclear if national suicide prevention programmes are effective, although evidence of effectiveness for specific components of such strategies is emerging.[34] The most impressive programme, developed in Finland, was based on information from a detailed national study of all suicides in 1 year and includes a wide range of elements.[35] A decline in the Finnish suicide rate has been attributed to the programme.[36] In England a national suicide prevention strategy with a suicide target was introduced in 2002.[37] Strategies have also been introduced in Scotland[38] and Ireland.[39] While prevention strategies are difficult to evaluate[40] there are indications that programmes for prevention of suicide on a national scale may be effective.

Table 4.15.4.6 Examples of strategies for prevention of suicide and attempted suicide

Population strategies
Reducing availability of means for suicide
Educating primary care physicians
Influencing media portrayal of suicide
Educating the public about mental illness and its treatment
Educational approaches in schools
Befriending agencies and telephone helplines
Addressing the economic factors associated with suicidal behaviour
High-risk strategies
Prevention of suicide in:
Patients with psychiatric disorders
The elderly
Suicide attempters
High-risk occupational groups
Prisoners

Population strategies

(a) Reducing availability of means for suicide

This is the most widely discussed population strategy.[41] It is based on evidence that if the availability and/or danger of a popular method for suicide changes then this tends to have an impact on suicide rates. The general principles of prevention through reducing availability of means are, first, that many suicidal acts occur impulsively and therefore if a dangerous means is available this is more likely to result in death and, secondly, that the eventual suicide rate in survivors of serious attempts is remarkably low. Also the common adage that if people are intent on committing suicide they will find a means is not necessarily correct (see below).

(i) Coal gas

The most cited evidence for the effectiveness of this approach is the reduction in suicides in the United Kingdom which occurred in the 1960s and early 1970s when toxic coal gas supplies were gradually replaced with non-toxic North Sea gas.[42] Prior to this time coal gas poisoning through people placing their head in a gas oven was the most common method of suicide in the United Kingdom. As North Sea gas was gradually introduced the suicide rate dropped steadily, eventually being reduced by approximately a third. It is estimated that as many as 6000 deaths may have been prevented by this change. The effect also illustrates the point that when one method of suicide is no longer available people do not automatically turn to another, or if they do it may be to one that is less likely to cause death. Thus, it was some years before the suicide rate rose again, this being related to an increase in deaths from poisoning with carbon monoxide from car exhausts. Another factor that may have been relevant to the decline in suicides was the reduction in prescribing of barbiturates, these being replaced by far less toxic benzodiazepines.

(ii) Carbon monoxide

Suicide by carbon monoxide poisoning from car exhausts has become less common because cars are now fitted with catalytic converters. This has resulted in a decline in suicide rates in countries where this method of suicide had become more common, particularly in young males.[43]

(iii) Firearms

The widespread availability of guns in certain countries, particularly the United States, has been proposed as an important reason for their relatively high suicide rates. Guns are used in more than half of all suicides in the United States and their use for suicide correlates with the holding of gun licences in households.[44] Some controversy surrounds the question of whether restricting availability of guns leads to a reduction in suicide rates, but the weight of evidence seems to indicate that it does.[45]

(iv) Antidepressants

Given the very strong link between suicide and depression, and the risk of death from overdose of some of the older antidepressants, there has been much debate about whether more extensive use of newer, less toxic antidepressants would prevent suicides. This is not a simple question, as some patients respond better to the older tricyclic antidepressants. Another consideration concerns the cost of the newer antidepressants compared with the older varieties. Also it is very important to remember that most people who are taking antidepressants do not kill themselves with their antidepressants but use other methods. This and the probable selective prescribing of SSRIs to people judged to be at risk may account for the finding that suicide rates were higher in patients taking fluoxetine than patients taking other and in some cases more toxic antidepressants.[46] Nevertheless, common sense dictates that patients known to be at risk, and especially those with a history of suicidal behaviour, should be prescribed the less toxic preparations.

(v) Analgesics

In the United Kingdom and some other countries there has been particular concern about deaths from self-poisoning with paracetamol. Due to evidence that countries which have fewer tablets per pack seem to have a lower rate of mortality from paracetamol self-poisoning and because overdoses of paracetamol are often taken impulsively and involve household supplies, legislation was introduced in the United Kingdom in 1998 to reduce in the number of tablets of paracetamol (and aspirin) available per pack. This resulted in fewer overdoses, decreased cases of hepatotoxicity due to paracetamol toxicity, and a reduced number of deaths from both paracetamol and aspirin.[47]

(vi) Safety measures

Much attention has been paid to improving safety at popular sites for suicide. This includes erecting suicide barriers on bridges, multi-storey car parks, and other sites. If environmental changes are made such that a popular suicide site becomes safer, this does not mean that people at risk automatically move to using another site. For example, erection of barriers on the Clifton Suspension Bridge in Bristol, a popular site for suicide, has resulted in far fewer deaths by jumping.[48]

Clinicians involved in the development of suicide prevention strategies should look very carefully for local patterns which might provide clues about potentially effective measures for reducing access to methods. This could include, for example, ensuring that psychiatric inpatient units are free of hooks, pipes, and other objects or structures from which patients could hang themselves, and that all bed rails are collapsible (compulsory in the United Kingdom), secure fencing of railway lines or waterways close to psychiatric hospitals, and making local popular sites for suicide safer (e.g. suicide barriers on bridges). In addition, attention should

be paid to common dangerous methods of self-poisoning. Specific strategies may be required depending on local patterns. For example, the high rates of suicide in rural areas of developing countries due to self-poisoning with pesticides might be reversed with safe-storage programmes.

(b) Education of primary care physicians

Much of the attention regarding improved detection of individuals at risk has concerned the detection and management of depression in general practice. This was stimulated by findings that showed many people who died by suicide or who attempted suicide had seen their general practitioners shortly before these acts. Evidence that an intensive educational programme for general practitioners might be effective in influencing suicide rates comes from a study conducted on the Swedish island of Gotland.[49] In the year following this programme the suicide rate dropped significantly, prescribing of antidepressants by general practitioners increased, referrals to psychiatry, especially for depression, decreased, the amount of time lost from work for depression decreased, as did psychiatric admissions. Unfortunately this effect was fairly short-lived in that suicide rates rose again in subsequent years, which the authors attributed to some of the general practitioners having left the island. They also suggested that such programmes need to be repeated.[50] It is also important to note that the suicide rate only declined in females. The evidence in this study was based on relatively small numbers, at least with reference to suicide, although the effects on the management and outcome of depression were perhaps more impressive.

While the Gotland study has generated a lot of debate about suicide prevention in primary care, detection of people most at risk in general practice is extremely difficult because a large number of patients share risk factors and because suicide is a rare event. The most pragmatic view is that effective detection and treatment of depression (and other psychiatric disorders) in primary care are extremely important aims in their own rights and that they might also have benefits in terms of preventing some suicides.

Psychiatrists involved in designing suicide prevention strategies might ensure that there are effective local educational programmes for clinicians in primary care and other settings regarding detection and treatment of people with mental disorders.

(c) Influencing media portrayal of suicidal behaviour

Dramatic reporting and portrayal of suicidal behaviour by the media can facilitate suicidal acts in other people. This has been shown in a wide range of studies of both newspaper and television reporting of suicides and fictional presentations of suicidal behaviours in films and television dramas. The impact of media presentations appears to be greatest where the method used in the suicidal act is described in detail, where details of the deceased and/or the site of the act is provided, and for deaths by suicide of celebrities. The largest impact of media influence is on young people, although there is also influence on older people.[51]

In each country, consideration should be paid to the development of consensus statements about media policies in relation to reporting and portrayal of suicide,[52] which could be produced by joint working parties including representatives of the press, clinical and voluntary agencies, and experts in the field of suicidal behaviour. More difficult is the potentially valuable task of encouraging a policy whereby the media can be used to portray effective coping strategies for people in distress. Such a strategy will need to encompass local cultural factors. Psychiatrists and other professionals developing suicide prevention strategies might examine the practices of their local media with regard to reporting of suicides and, if necessary, hold meetings with media producers to explain the dangers of dramatic and extensive reporting, and also to explore how the media might help in prevention.

Concern is growing about influences on suicidal behaviour through the internet, especially web sites providing instructions on methods of suicide and chat rooms whereby individuals can instruct. Some sites seem to be intended to promote suicide, such as those which initiate meetings between suicidal individuals. Attempts can be made to regulate such sites, but this impossible for sites from other countries. Also, internet providers can be encouraged to ensure that sites offering positive help appear before less desirable sites.

(d) Education of the public about mental illness and its treatment

In view of the very strong link between suicide and mental illness, effective treatment of psychiatric disorders must be a central theme in suicide prevention. However, detection of people with disorders will depend on the awareness that they and those around them have regarding the signs and symptoms of disorder, and their willingness to seek appropriate help.[53] These important stages in receiving effective help will depend on the general public's attitudes towards mental illness and knowledge of its nature and the feasibility of treatment. In several countries, programmes to encourage education of the public about psychiatric disorders (especially depression) and to tackle stigmatization of those who are ill have been established. At this stage, evidence is lacking as to whether or not they have been successful. Psychiatrists and their colleagues might consider similar campaigns where these are not already in place, although the method of delivery of messages (e.g. media presentations, leaflets, workshops, articles) will clearly depend on local factors.

(e) Educational approaches in schools

There have been three broad approaches in trying to prevent suicide through school-based programmes.[54] The first of these includes teaching about the facts of suicide. Worrying evidence from the United States that such a programme appeared to lead to a small increase in pupils' ratings of the acceptability of suicide as an option compared with the ratings of pupils who did not receive the programme suggests that this is not a wise approach.

Suicidal behaviour in young people often appears to be related to depression, anxiety, low self-esteem, difficulties during upbringing (e.g. abuse, deprivation), life events (especially break-up of relationships, family problems, and bullying), and poor problem-solving skills.[54] Also, troubled and suicidal young people most often seek help from their peers. A second school-based strategy has been the development of educational programmes in schools about recognition of psychological distress in individuals and their peers, problem-solving, and peer support. Given the early age at which suicidal behaviour begins, such programmes should probably be targeted at extremely young school children, with later sessions for adolescents.

A third approach is to screen adolescents with questionnaires to detect children and adolescents at risk of psychiatric disorder and

possible suicidal behaviour. Pupils that are so detected will then need referral to an appropriate agency for further assessment and possible treatment. While there is some evidence to support such an approach it is not without drawbacks. (Suicide in children and adolescents is considered further in Chapter 9.2.10).

For psychiatrists and others involved in developing local prevention strategies it is important to recognize that school-based approaches to prevention represents is a highly sensitive area and one where the most effective (and least risky) approach is at present unclear. Another important aspect of suicidal behaviour in school pupils is the management of the aftermath of suicides and its impact on other pupils and how to tackle outbreaks of self-harm.[54]

(f) Befriending agencies and telephone helplines

A very important component of suicide prevention policy in many countries is the support provided by largely volunteer staffed befriending agencies and especially telephone helplines. The best known of these is Samaritans. A key principle on which such services are based is that people in distress and at risk of suicide will benefit from being able to discuss their problems with someone entirely confidentially. Recently, more assertive outreach programmes, in which volunteers meet up with distressed individuals such as in prisons and in remote areas, have been added to the traditional telephone service. In the United Kingdom and elsewhere counselling by e-mail and text messaging is being used extensively.

The effectiveness of these approaches is largely unknown. Conducting controlled trials to examine their efficacy is very difficult. Naturalistic studies have produced conflicting evidence about the effectiveness of the Samaritans in the United Kingdom.[55,56] An examination of changes in suicide rates in areas with and without crisis intervention services in the United States suggested that suicide rates in young white females may have been reduced in areas where such services were developed.[57] Given the large numbers of contacts made with Samaritans in the United Kingdom (nearly 5 million in 2005), it appears that the service is valued by people in distress.

Volunteer-run telephone helplines and similar services may benefit greatly from the support and advice of local clinicians, who should regard them as a potentially valuable element in a local suicide prevention strategy.

Addressing the economic factors associated with suicidal behaviour

The association between suicidal behaviour and unemployment and poverty suggests that in order for suicide rates to change markedly these important socio-economic factors must be modified. The big increase in suicide rates during the economic depression of the late 1920s and early 1930s and the statistical association between suicide risk and unemployment would support this. Clearly such factors are increasingly a reflection of the global economic situation, but the strategies of individual governments, particularly in relation to the employment prospects for young people, may be influential. The main role of psychiatrists may be in highlighting these factors. The considerable evidence that changes in the economic environment can exert a powerful influence on suicide rates indicates that governments with serious intentions to reduce suicide rates should address these issues.[32]

Strategies for high-risk groups

There are a wide range of possible prevention strategies which can be targeted at high-risk groups. Here are some of the more important examples of such groups and relevant strategies will be discussed.

(a) Patients with psychiatric disorders

(i) Risk identification

One approach to preventing suicide in people with known psychiatric disorders is to try and use recognized risk factors for suicide in each disorder to identify high-risk patients. The main psychiatric disorders in series of people who have died by suicide are depression (approximately two-thirds), severe alcohol abuse (approximately 15 per cent), and schizophrenia (5–10 per cent).[58] The main suicide risk factors identified in these three disorders include, for example, previous attempts, family history of suicidal behaviour, and living alone. Comorbidity of disorders (e.g. depression and alcohol abuse) and of personality and psychiatric disorders increases risk.

One difficulty in using a risk-identification approach, however, is that the risk factors identified from studies of groups of individuals who have died from suicide are often misleading when applied to individual patients. Also when applying such factors a relatively large number of individuals will appear to be at risk when they may not in fact be so. In clinical practice it is important to be aware of patients who, because of their individual characteristics, are at long-term high risk. Clinicians must also be aware of acute situations which may temporarily increase the risk in patients, be they ones at long-term risk or not.

The most pragmatic approach, therefore, is to ensure that proven effective treatments for patients with these conditions are available and also to be particularly cautious at times of obvious high risk. There are particular periods of risk of suicide for patients with psychiatric disorders. One of these is during the first few weeks after discharge from psychiatric hospital.[59] This emphasizes the necessity for continuity of care at this critical time. Other risk times may be following the break-up of a relationship or other significant loss, during periods of marked hopelessness, shortly after discharge from hospital, and following recent suicidal behaviour by another patient or someone else close to the individual.

(ii) Preventative strategies

Prevention of suicide in patients with psychiatric disorders must be a major element in any suicide prevention strategy.[60] Important strategies in preventing suicide in patients with affective disorders include active treatment of individual episodes of illness, psychological therapy to improve compliance with treatment and assisting individuals to manage their disorder, use of lithium and other mood stabilizers for patients with recurrent bipolar disorders, and use of long-term antidepressants in patients with frequent relapses of depressive disorders. Risk is often greatest during the early stages of a disorder.[53]

The risk factors in schizophrenia indicate that risk tends to be highest between episodes of acute illness when patients may have insight and feel hopeless about their circumstances and prospects. Risk is related more to affective symptoms than core features of the disorders.[61] Continuity of care is likely to be a particularly important factor in preventing suicides in such patients at risk, with care

being continued energetically during periods of remission. Community psychiatric nurses have a very important role with such patients. The use of the newer atypical neuroleptics may also be beneficial.[62]

Direct treatment of abuse is likely to be the best preventive strategy for patients with substance abuse disorders, with care taken to manage episodes of depression. The particularly high risk in the weeks following a break-up of a relationship for patients with severe alcohol abuse[63] again points to the need for continuity of support in the community.

The prevention of suicide in patients with comorbid disorders, especially the combination of depression with alcohol abuse and/or personality disorder, is a challenging task, particularly as compliance with treatment is often less good than in patients with single disorders. Effective prevention is likely to depend on close integration of care between different statutory care agencies.

Another important element in prevention in this population is education in suicide risk assessment and management procedures for clinical staff at all levels of seniority. These should be incorporated in educational programmes for risk assessment in general.

(b) Elderly people

In planning suicide prevention in the elderly population, account must be taken of the relative immobility of many older people. In a region of Italy, introduction of a telephone service to provide support and access to emergency help for elderly persons at risk has been associated with an encouraging decline in elderly suicides in the area.[64] This might serve as a model for other countries.

(c) Suicide attempters

In view of the clear association between non-fatal suicidal behaviour and subsequent suicide, establishment of adequate services for suicide attempters, including the provision of careful assessments of patients in the general hospital and offering treatments for which at least some indicators of benefit are available (see above), is an important element in any national suicide prevention strategy. There is good evidence that well-trained, non-medical psychiatric staff can effectively carry out assessments and arrange aftercare. Models for ideal services exist, such as those published by the National Institute for Clinical Excellence[65] in the United Kingdom.

(d) High-risk occupational groups

Certain occupational groups are known to be at relatively high risk of suicide. In the United Kingdom these include farmers, veterinary surgeons, dental practitioners, medical practitioners, pharmacists, and female nurses. It is interesting to note that all these groups have relatively easy access to dangerous methods for suicide. Since prevention through detection of those most at risk encounters the usual difficulties of prevention of relatively rare behaviour using rather crude risk factors, it is probably more important to have general strategies for improving care in individual groups. In doctors, for example, there are some particular difficulties about confidentiality and therefore providing easy means of doctors getting confidential help is important. In farmers, improving the knowledge and attitudes of farming communities towards psychiatric disorder, and removing access to firearms at times of risk, are likely to be important.

(e) Prisoners

There are relatively high suicide rates in prisoners,[66] especially young males held on remand. While one aspect of prevention is through ensuring that prisons and police cells are safe in terms of absence of structures from which inmates can hang themselves, there are a range of other potentially useful and humane strategies. These include careful assessments of new inmates using risk-assessment procedures, training of staff with regard to both assessment skills and attitudes towards mental health problems and suicide prevention, in-reach programmes by befriending organizations such as the Samaritans, and ready access to psychiatric and psychological services. Clinicians involved in local suicide prevention programmes should include prisons in their considerations.

Further information

Hawton, K. (2005). *Prevention and treatment of suicidal behaviour: from science to practice*. Oxford University Press, Oxford.

Hawton, K. and Van Heeringen, K. (2000). *The international handbook of suicide and attempted suicide*. Wiley, Chichester.

National Collaborating Centre for Mental Health. (2004). *Self-harm: the short-term physical and psychological management and secondary prevention of self-harm in primary and secondary care (full guideline) clinical guideline 16*. National Institute for Clinical Excellence, London.

References

1. World Health Organization. (1994). *Ninth general programme of work covering the period 1996–2000*. World Health Organization, Geneva.
2. United Nations. (1996). *Prevention of suicide: guidelines for the formulation and implementation of national strategies*. Department for Policy Coordination and Sustainable Development, New York.
3. Hjelmeland, H., Hawton, K., Nordvik, H., *et al.* (2002). Why people engage in parasuicide: a cross-cultural study of intentions. *Suicide & Life Threatening Behavior*, **32**, 380–93.
4. Zahl, D. and Hawton, K. (2004). Repetition of deliberate self-harm and subsequent suicide risk: long-term follow-up study in 11,583 patients. *The British Journal of Psychiatry*, **185**, 70–5.
5. Owens, D., Horrocks, J., and House, A. (2002). Fatal and non-fatal repetition of self-harm systematic review. *The British Journal of Psychiatry*, **181**, 193–9.
6. Haw, C., Hawton, K., Houston, K., *et al.* (2001). Psychiatric and personality disorders in deliberate self-harm patients. *The British Journal of Psychiatry*, **178**, 48–54.
7. Paykel, E.S., Prusoff, B.A., and Myers, J.K. (1975). Suicide attempts and recent life events: a controlled comparison. *Archives of General Psychiatry*, **32**, 327–33.
8. Williams, J.M.G., Crane, C., Barnhofer, T., *et al.* (2005). Psychology and suicidal behaviour: elaborating the entrapment model. In *Prevention and treatment of suicidal behaviour: from science to practice* (ed. K. Hawton), pp. 71–90. Oxford University Press, Oxford.
9. Mann, J.J. (2003). Neurobiology of suicidal behaviour. *Nature Reviews Neuroscience*, **4**, 819–28.
10. Beck, A.T., Steer, R.A., Kovacs, M., *et al.* (1985). Hopelessness and eventual suicide: a 10 year prospective study of patients hospitalised with suicidal ideation. *The American Journal of Psychiatry*, **145**, 559–63.
11. Hawton, K., Townsend, E., Arensman, E., *et al.* (2005). Psychosocial and pharmacological treatments for deliberate self harm *Cochrane Database of Systematic Reviews*, Issue: 4, Art. No.: CD001764. DOI: 10.1002/14651858.CD001764.

12. Townsend, E., Hawton, K., Altman, D.G., *et al.* (2001). The efficacy of problem-solving treatments after deliberate self-harm: meta-analysis of randomised controlled trials with respect to depression, hopelessness and improvement in problems. *Psychological Medicine*, **31**, 979–88.
13. Brown, G.K., Have, T.T., Henriques, G.R., *et al.* (2005). Cognitive therapy for the prevention of suicide attempts: a randomized controlled trial. *The Journal of the American Medical Association*, **294**, 563–70.
14. Tyrer, P., Thompson, S., Schmidt, U., *et al.* (2003). Randomized controlled trial of brief cognitive behaviour therapy versus treatment as usual in recurrent deliberate self-harm: the POPMACT study. *Psychological Medicine*, **33**, 969–76.
15. Linehan, M.M., Armstrong, H.E., Suarez, A., *et al.* (1991). Cognitive-behavioral treatment of chronically parasuicidal borderline patients. *Archives of General Psychiatry*, **48**, 1060–4.
16. Linehan, M.M., Comtois, K.A., Murray, A.M., *et al.* (2006). Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Archives of General Psychiatry*, **63**, 757–66.
17. Van Heeringen, C., Jannes, S., Buylaert, W., *et al.* (1995). The management of non-compliance with referral to out-patient after-care among attempted suicide patients: a controlled intervention study. *Psychological Medicine*, **25**, 963–70.
18. Vaiva, G., Ducrocq, F., Meyer, P., *et al.* (2006). Effect of telephone contact on further suicide attempts in patients discharged from an emergency department: randomised controlled study. *British Medical Journal*, **332**, 1241–5.
19. Carter, G.L., Clover, K., Whyte, I.M., *et al.* (2005). Postcards from the EDge project: randomised controlled trial of an intervention using postcards to reduce repetition of hospital treated deliberate self poisoning. *British Medical Journal*, **331**, 805–9.
20. Evans, J., Evans, M., Morgan, H.G., *et al.* (2005). Crisis card following self-harm: 12-month follow-up of a randomised controlled trial. *The British Journal of Psychiatry*, **187**, 186–7.
21. Verkes, R.J., Van-der-Mast, R.C., Hengeveld, M.W., *et al.* (1998). Reduction by paroxetine of suicidal behavior in patients with repeated suicide attempts but not major depression. *The American Journal of Psychiatry*, **155**, 543–7.
22. Committee on Safety of Medicines. (2003). Selective serotonin reuptake inhibitors (SSRIs): overview of regulatory status and CSM advice relating to major depressive disorder (MDD) in children and adolescents including a summary of available safety and efficacy data. Medicines and Healthcare products Regulatory Agency, London.
23. Jick, H., Kaye, J.A., and Jick, S.S. (2004). Antidepressants and the risk of suicidal behaviors. *The Journal of the American Medical Association*, **292**, 338–43.
24. Montgomery, S.A., Montgomery, D.B., Jayanthi-Rani, S., *et al.* (1979). Maintenance therapy in repeat suicidal behaviour: a placebo controlled trial. Proceedings of the 10th International Congress for Suicide Prevention & Crisis Intervention. International Association for Suicide Prevention, pp. 227–9. Ottawa, Canada.
25. Cipriani, A., Pretty, H., Hawton, K., *et al.* (2005). Lithium in the prevention of suicidal behaviour and all-cause mortality in patients with mood disorders: a systematic review of randomised trials. *The American Journal of Psychiatry*, **162**, 1805–19.
26. Horrocks, J., Hughes, J., Martin, C., *et al.* (2005). *Patient experiences of hospital care following self-harm—a qualitative study*. Academic Unit of Psychiatry and Behavioural Sciences, Leeds.
27. Beck, A., Schuyler, D., and Herman, J. (1974). Development of suicidal intent scales. In *The prediction of suicide* (eds. A. Beck, H. Resnik, and D.J. Lettieri), pp. 45–56. Charles Press, Maryland.
28. Wood, A., Trainor, G., Rothwell, J., *et al.* (2001). Randomized trial of group therapy for repeated deliberate self-harm in adolescents. *Journal of American Academy of Child and Adolescent Psychiatry*, **40**, 1246–53.
29. Soni Raleigh, V. and Balarajan, R. (1992). Suicide and self-burning among Indians and West Indians in England and Wales. *The British Journal Psychiatry*, **161**, 365–8.
30. Burgess, S. and Hawton, K. (1998). Suicide, euthanasia and the psychiatrist. *Philosophy, Psychiatry and Psychology*, **5**, 113–26.
31. Hawton, K. (2005). *Prevention and treatment of suicidal behaviour: from science to practice*. Oxford University Press, Oxford.
32. Lewis, G., Hawton, K., and Jones, P. (1997). Strategies for preventing suicide. *The British Journal Psychiatry*, **171**, 351–4.
33. Rose, G. (1992). *The strategy of preventive medicine*. Oxford University Press, Oxford.
34. Mann, J.J., Apter, A., Bertolote, J., *et al.* (2005). Suicide prevention strategies. A systematic review. *The Journal of the American Medical Association*, **294**, 2064–74.
35. Upanne, M. (1998). Implementation of the suicide prevention strategy in Finland: first follow-up. In *Suicide prevention: a holistic approach* (eds. D. De Leo, A. Schmidtke, and R.F.W. Diekstra), pp. 219–23. Kluwer Academic Publishers, Dordrecht.
36. Ohberg, A. and Lönnqvist, J. (1997). Suicide trends in Finland 1980–1995. *Psychiatrica Fennica*, **28**, 11–23.
37. Department of Health. (2002). *National suicide prevention strategy for England*. Department of Health, London.
38. Scottish Executive. (2002). Choose life: a national strategy and action plan to prevent suicide in Scotland. The Stationery Office, Edinburgh.
39. Health Service Executive, the National Suicide Review Group, Department of Health and Children. (2005). *Reach out: national strategy for action on suicide prevention 2005–2014*. Health Service Executive, Dublin.
40. Goldney, R.D. (1998). Suicide prevention is possible: a review of recent studies. *Archives of Suicide Research*, **4**, 329–39.
41. Hawton, K. (2005). Restriction of access to methods of suicide as a means of suicide prevention. In *Prevention and treatment of suicidal behaviour: from science to practice* (ed. K. Hawton), pp. 279–91. Oxford University Press, Oxford.
42. Kreitman, N. (1976). The coal gas story: United Kingdom suicide rates 1960–1971. *British Journal of Preventive and Social Medicine*, **30**, 86–93.
43. Amos, T., Appleby, L., and Kiernan, K. (2001). Changes in rates of suicide by car exhaust asphyxiation in England and Wales. *Psychological Medicine*, **31**, 935–9.
44. Kellermann, A.L., Rivara, F.P., Somes, G., *et al.* (1992). Suicide in the home in relation to gun ownership. *New England Journal of Medicine*, **327**, 467–72.
45. Youth Suicide by Firearms Task Force. (1998). Consensus statement on youth suicide by firearms. *Archives of Suicide Research*, **4**, 89–94.
46. Jick, S.S., Dean, A.D., and Jick, H. (1995). Antidepressants and suicide. *British Medical Journal*, **310**, 215–8.
47. Hawton, K., Simkin, S., Deeks, J., *et al.* (2004). UK legislation on analgesic packs: before and after study of long term effect on poisonings. *British Medical Journal*, **329**, 1076–9.
48. Bennewith, O., Nowers, M., and Gunnell, D. (2007). Effect of barriers on the Clifton suspension bridge, England, on local patterns of suicide: implications for prevention. *The British Journal Psychiatry*, **190**, 266–7.
49. Rutz, W., von Knorring, L., and Walinder, J. (1989). Frequency of suicide on Gotland after systematic postgraduate education of general practitioners. *Acta Psychiatrica Scandinavica*, **80**, 151–4.
50. Rutz, W., von Knorring, L., and Walinder, J. (1992). Long-term effects of an educational program for general practitioners given by the Swedish committee for the prevention and treatment of depression. *Acta Psychiatrica Scandinavica*, **85**, 83–8.
51. Hawton, K. and Williams, K. (2005). Media influences on suicidal behaviour: evidence and prevention. In *Prevention and treatment of suicidal behaviour: from science to practice* (ed. K. Hawton), pp. 293–306. Oxford University Press, Oxford.

52. Pirkis, J., Blood, R.W., Beautrais, A., *et al.* (2006). Media guidelines on the reporting of suicide. *Crisis*, **27**, 82–7.
53. Jamison, K.R. and Hawton, K. (2005). The burden of suicide and clinical suggestions for prevention. In *Prevention and treatment of suicidal behaviour: from science to practice* (ed. K. Hawton), pp. 183–96. Oxford University Press, Oxford.
54. Hawton, K. and Rodham, K. (2006). *By their own young hand: deliberate self-harm and suicidal ideas in adolescents*. Jessica Kingsley Publishers, London.
55. Bagley, C.R. (1968). The evaluation of a suicide prevention scheme by an ecological method. *Social Science & Medicine*, **2**, 1–14.
56. Jennings, C., Barraclough, B.M., and Moss, J.R. (1978). Have the Samaritans lowered the suicide rate? A controlled study. *Psychological Medicine*, **8**, 413–22.
57. Lester, D. (1974). Effect of suicide prevention centers on suicide rates in the United States. *Health Service Reports*, **89**, 37–9.
58. Cavanagh, J.T.O., Carson, A.J., Sharpe, M., *et al.* (2003). Psychological autopsy studies of suicide: a systematic review. *Psychological Medicine*, **33**, 395–405.
59. Goldacre, M., Seagroatt, V., and Hawton, K. (1993). Suicide after discharge from psychiatric inpatient care. *Lancet*, **342**, 283–6.
60. University of Manchester. (2006). Avoidable deaths. Five year report of the National Confidential Inquiry into suicide and homicide by people with mental illness. National Confidential Inquiry into Suicide and Homicide by People with Mental Illness, Manchester.
61. Hawton, K., Sutton, L., Haw, C., *et al.* (2005). Schizophrenia and suicide: a systematic review of risk factors. *The British Journal of Psychiatry*, **187**, 9–20.
62. Meltzer, H.Y., Alphs, L., Green, A.I., *et al.* (2003). Clozapine treatment for suicidality in schizophrenia. *Archives of General Psychiatry*, **60**, 82–91.
63. Murphy, G.E., Armstrong, J.W. Jr, Hermele, S.L., *et al.* (1979). Suicide and alcoholism: interpersonal loss confirmed as a predictor. *Archives of General Psychiatry*, **36**, 65–9.
64. De Leo, D., Buono, M.D., and Dwyer, J. (2002). Suicide among the elderly: the long-term impact of a telephone support and assessment intervention in northern Italy. *The British Journal Psychiatry*, **181**, 226–9.
65. National Institute for Clinical Excellence. (2004). Clinical guideline 16. Self-harm: the short-term physical and psychological management and secondary prevention of self-harm in primary and secondary care. National Institute for Clinical Excellence, London.
66. Fazel, S., Benning, R., and Danesh, J. (2005). Suicides in male prisoners in England and Wales, 1978–2003. *Lancet*, **366**, 1301–2.

4.16

Culture-related specific psychiatric syndromes

Wen-Shing Tseng

The concept of culture-related specific (psychiatric) syndromes

In certain ways, all psychiatric disorders are more or less influenced by cultural factors, in addition to biological and psychological factors, for their occurrence and manifestation. 'Major' psychiatric disorders (such as schizophrenia or bipolar disorders) are more determined by biological factors and relatively less by psychological and cultural factors, but 'minor' psychiatric disorders (such as anxiety disorders, conversion disorders, or adjustment disorders) are more subject to psychological causes as well as cultural factors. In addition to this, there are groups of psychiatric disorders that are heavily related to and influenced by cultural factors, and therefore addressed as culture-related specific psychiatric syndromes.

Culture-related specific psychiatric syndromes, also called culture-bound syndromes[1] or culture-specific disorders,[2] refer to mental conditions or psychiatric syndromes whose occurrence or manifestations are closely related to cultural factors and thus warrant understanding and management primarily from a cultural perspective. Because the presentation is usually unique, with special clinical manifestations, the disorder is called a culture-related specific psychiatric syndrome.[3] From a phenomenological point of view, such a condition is not easily categorized according to existing psychiatric classifications, which are based on clinical experiences of commonly observed psychiatric disorders in western societies, without adequate orientation towards less frequently encountered psychiatric conditions and diverse cultures worldwide.

Around the turn of the twentieth century, during a period of colonization by western societies, western ministers, physicians, and others visited faraway countries, where they encountered behaviours and unique psychiatric conditions that they had never experienced at home. Most of these conditions were known to the local people by folk names, such as *latah*, *amok*, *koro*, *susto*, and so on, and were described by westerners as exotic, rare, uncommon, extraordinary mental disorders, mental illnesses peculiar to certain cultures, or culture-bound syndromes. The latter term implies that such syndromes are bound to a particular cultural region.[4]

Recently, however, cultural psychiatrists have realized that such psychiatric manifestations are not necessarily confined to particular ethnic-cultural groups. For instance, epidemic occurrences of *koro* (penis-shrinking panic) occur among Thai or Indian people, not only among the Southern Chinese as originally claimed; and sporadic occurrences of *amok* attacks (mass, indiscriminate homicidal acts) are observed in the Philippines, Thailand, Papua New Guinea, and in epidemic proportions in many places in South Asia,[5] in addition to Malaysia where it was believed to most commonly occur. Terrifying examples of *amok* have recently occurred with frequency on school campuses and in workplaces in the United States.

Therefore, the term culture-bound does not seem to apply, and it has been suggested that culture-related specific psychiatric syndrome would be more accurate to describe a syndrome that is closely **related** to certain *cultural traits* or *cultural features* rather than **bound** specifically to any one *cultural system* or *cultural region*.[4] Accordingly, the definition has been modified to a collection of signs and symptoms that are restricted to a limited number of cultures, primarily by reason of certain of their psychosocial features,[6] even though it is recognized that every psychopathology is influenced by culture to a certain degree.

Subgrouping of culture-related specific syndromes

In order to organize and categorize the various culture-related syndromes, several different subgroup systems have been proposed by different scholars in the past, such as by cardinal symptoms or by 'taxon' (according to a common factor). However, instead of focusing on the clinical manifestation descriptively, it will be more meaningful to subgroup the syndromes according to how they might be affected by cultural factors.

It has been recognized that there are different ways culture contributes to psychopathology. Namely: *pathogenetic effect* (culture has causative effect), *pathoselective effect* (culture selects the nature and type of psychopathology), *pathoplastic effect* (culture contributes to the manifestation of psychopathology), *pathoelaborating effect* (culture elaborates and reinforces certain types of manifestations), *pathofacilitating effect* (culture contributes to the frequent occurrence of particular psychopathologies), or *pathoreactive effect* (culture determines the reaction to psychopathology). Furthermore, culture impacts differently on different types of psychopathology.

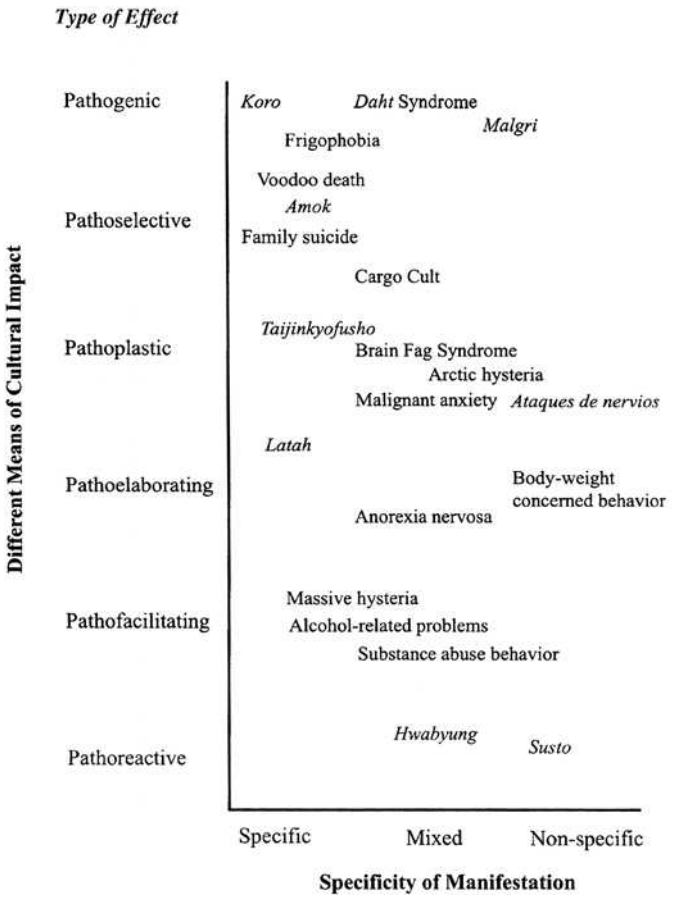


Fig. 4.16.1 Position of culture-related syndromes according to two parameters (This article was published in Handbook of Cultural Psychiatry, Tsend, W.S., copyright Elsevier (2001).)

If the psychopathology is divided into that which is more biologically determined (such as organic mental disorders or major psychiatric disorders), psychologically determined (such as minor psychiatric disorders), or socio-culturally determined (such as culture-related specific syndromes or epidemic mental disorders), it can be said that a pathoreactive effect is observed for all types of psychiatric disorders while a psychogenetic effect is observed mainly for those psychiatric disorders which are more socio-culturally determined, such as culture-related specific syndromes or epidemic mental disorders.

Academically identified culture-related specific syndromes can be compared according to two parameters, namely, the different means of cultural impact and the specificity of the manifestation. It will be useful to determine the basic ways culture contributes to the psychopathology, namely: pathogenic, pathoselective, pathoplastic, pathoelaborating, pathofacilitating, or pathoreactive effects; and to what extent the manifested syndrome is: specific, mixed, or non-specific (Fig. 4.16.1). For example, the primary cultural contribution to *koro* is a pathogenic effect with specific manifestation; to *latah* a pathoelaborating effect with specific manifestation; and to *susto* a pathoreactive effect with non-specific manifestation.

Culture-related beliefs as causes of the occurrence (pathogenetic effects)

(a) Koro (genital-retraction anxiety disorder)

(i) Definition

Koro is a Malay term which means the head of a turtle, symbolizing the male sexual organ which can 'shrink'. Clinically *koro* refers to the psychiatric condition in which the patient is morbidly concerned that his penis is shrinking excessively and subsequently dangerous consequences (such as death) might occur. The manifested symptoms may vary from simple excessive concern to obsessive/hypochondriac concern, intense anxiety, or a panic condition related to the shrinking of the penis. Clinically, this is usually a benign (non-psychotic) condition that occurs in individual, as a sporadic case, but occasionally it may grow to epidemic proportions, so that several hundreds or a thousand people may develop this disorder in a panic manner within a limited time of several weeks or several months. The majority of cases are young males who fear that their penises are shrinking. However, the organ concerned may be any protruding part of the body, such as the nose or ear (particularly when patients are prepuberty children) or the nipples or labia (in females).

Sporadic occurrences of female *koro* cases have never been reported. However, in *koro* epidemics, a small portion of the victims may be female.[7,8] In those cases, the female patients demonstrate slightly different clinical pictures, mainly focusing on the retraction of the nipples and some on the labia. The clinical condition is characterized by a more or less hysterical panic, associated with multiple somatic symptoms, a bewitched feeling, or the misinterpretation or accusation by others of being bewitched.

(ii) Geographic and ethnic group distribution

Cultural psychiatrists originally considered *koro* to be a culture-bound disorder related only to the Chinese. Most Chinese investigators have taken the view that the disorder is related to the Chinese cultural belief in *suoyang*, literally means shrinkage of yang organ due to excess loss of Yang (male) element. It has been speculated further that the occurrence of *koro* among people in South Asian countries, such as Malaysia and Indonesia, was the result of Chinese migrants. However, this cultural-diffusion view is doubted now, since *koro* epidemics have been reported in Thailand and India as well, involving non-Chinese victims.

As a result of the dissemination of knowledge about *koro* as a culture-bound syndrome, there is increased literature reporting so-called *koro* cases from various ethnic groups around the world, such as from the United Kingdom, Canada, and Israel. However, it was pointed out that among the cases reported, all suffered primarily from many psychiatric conditions: affective disorders, schizophrenia and anxiety disorders, drug abuse, or organic brain disorders. Therefore, they were referred as having *koro*-like symptoms, not exactly the same as the *koro* syndrome. It is necessary for clinicians to recognize that *koro* is referred to on different levels as: a symptom, a syndrome, or an epidemic disorder.

(iii) Epidemic of koro

Koro attack may occur occasionally as endemic or epidemic, involving many victims. Epidemic *koro* has been observed in several areas: Guangdong area of China, Singapore, Thailand, and India. As an epidemic, it is manifested as a panic state rather than an anxiety or obsessive state. The epidemic *koro* has tended to occur when there

was socio-political stress within the epidemic areas, and an outburst of *koro* by a group of people may be interpreted as a way to deal with the social stress that they encountered.

(iv) Diagnostic issue

Clinicians habitually try to fit pathologies into certain diagnostic categories. Because the patient, based on his misinterpretation, is morbidly preoccupied with the idea that certain ill-effects may occur due to the excessive retraction of his genital organ, the condition may, in a broad sense, be classified as a hypochondriacal disorder as defined in DSM-IV. The condition is also similar to a body dysmorphic disorder, as the patient is preoccupied with a culturally induced, imagined defect in his physical condition. If the focus is on how the patient reacts emotionally, how he responds to the culture-genic stress, with fear, anxiety, or a panic state, anxiety disorder may be considered. However, when we try to categorize culture-related specific syndromes according to the existing nosologically oriented classification system, their meaning and purpose are lost.

(v) Therapy

As for therapy of sporadic individual cases, assurance may be provided or medical knowledge offered in the form of educational counselling to eliminate the patient's concern about impending death. This supportive therapy may work in many cases, but, for someone who firmly believes the *koro* concept, it may not. In general, a young, unmarried male, who lacks adequate psychosexual knowledge and experience, will respond favourably to therapy. If necessary, it is desirable to work on issues such as the patient's self-image, self-confidence, or his masculinity.

(b) Dhat syndrome (semen-loss anxiety)

Very closely related to the genital-retraction anxiety disorder (*koro*) is the semen-loss or semen-leaking anxiety disorder, or spermatorrhea, also known by its Indian folk name, *dhat* syndrome. The *dhat* syndrome refers to the clinical condition in which the patient is morbidly preoccupied with the excessive loss of semen from an improper form of leaking, such as nocturnal emissions, masturbation, or urination. The underlying anxiety is based on the cultural belief that excessive semen loss will result in illness. Therefore, it is a pathogenically induced psychological disorder.[9,10] The medical term spermatorrhea is a misnomer, as there is no actual problem of sperm leakage from a urological point of view.

From a clinical point of view, the patients are predominantly young males who present vague, multiple somatic symptoms such as fatigue, weakness, anxiety, loss of appetite, and feelings of guilt (about having indulged in sexual acts such as masturbation or having sex with prostitutes). Some also complain of sexual dysfunction (impotence or premature ejaculation). The chief complaint is often that the urine is opaque, which is attributed to the presence of semen. The patient attributes the passing of semen in the urine to his excessive indulgence in masturbation or other socially defined sexual improprieties.[11] The syndrome is also widespread in Nepal, Sri Lanka (where it is referred to as *prameha* disease), Bangladesh, and Pakistan.

(c) Sorcery fear and voodoo death (magic-fear-induced death)

The peculiar phenomenon of voodoo death refers to the sudden occurrence of death associated with taboo-breaking or curse fear. It is based on the belief in witchcraft, the putative power to bring about misfortune, disability, and even death through spiritual mechanisms.[12] A severe fear reaction may result from such beliefs, which may actually end in death. From a psychosomatic point of view, it would be psychogenically induced death. From a cultural psychiatric perspective, it is another example of culture-induced morbid fear reaction.

Medically it has been recognized that sudden death was related to psychological stress occurring during experiences of acute grief, the threat of the loss of a close person, personal danger, or other stressful situations. It has been speculated that the cause of death was possibly from natural causes; the possible use of poisons due to sorcery or witchcraft; excessive fear reaction; or death from dehydration or existing physical illness due to old age.

Culture-patterned specific coping reactions (pathoselective effect)

(a) Amok (indiscriminate mass homicide attacks)

(i) The nature of the behaviour

Amok is a Malay term that means to engage furiously in battle. Clinically *amok* refer to an acute outburst of unrestrained violence associated with (indiscriminate) homicidal attacks, preceded by a period of brooding and ending with exhaustion and amnesia.[13] *Amok* homicides are distinct from other murders: the killer chooses an extremely destructive weapon, a crowded location, and insanely and indiscriminately kills a large number of people.[14]

There has been much speculation as to why *amok* behaviour tends to occur in Malay society. One explanation is its connection to the religious background of the people. Muslims are not permitted to commit suicide, which is considered a most heinous act in the Mohammedan religion. In the past, aggressive-homicidal behaviour influenced by infectious diseases has been considered, along with malaria, dengue, neurosyphilis, epilepsy, and so on, as biological in some cases. From a psychological point of view, an extraordinary sensitivity to hurt and the tendency to blame others for one's own difficulties are considered possible causes for the phenomenon. Loss of social standing, by way of insult, loss of employment, or financial loss, has been posited as a precipitating event for *amok*.

(ii) Amok behaviour in other areas

The outburst of aggressive (mass) homicidal behaviour is not necessarily confined to one cultural area, but can potentially be observed elsewhere, such as the territory of Papua and New Guinea, or many areas in South Asia, such as Thailand, the Philippines, Lao, and Indonesia.[5] It has challenged the previous view that *amok* occurs endemically within a particular society. It has been indicated that *amok* could happen in a fashion by communicability and through transmission from one population to another.[5] During the past decade in the United States, there have been increasing episodes of massive (and aimless) killing of people in neighbourhoods, workplaces, and of teachers/students in schools by deadly military weapons. These are American versions of *amok* attacks.

(b) Family suicide

When adults encounter severe difficulties (such as financial debt or a disgraceful event), there are many ways to deal with such problems. As one of the ways to cope with the difficult situation, Japanese parents, may decide to commit suicide together with their young children. This stress-coping method is based on the cultural

belief that it would be disgraceful to live after a shameful thing had happened, and that the shame would be relieved by ending one's life. This is coupled with the belief that the children, if left as orphans (after parents' committing husband–wife double suicide), would be mistreated by others. 'Blood is thicker than water' is the common saying reflecting the emphasis of blood-related family tie. Therefore, it would be better for them to die with their parents. This unique way of solving problems was often observed in Japan in the past. It is declining now but still observed some times.

(c) Cargo-cult syndromes (millenniary delusions)

Numerous social and behaviour scientists have noted that, historically, there have been occurrences of crisis cults in many different countries. The Taiping (Great Peace) Rebellion in China, Kikuyu maumau in Kenya, and the Ghost Dance of the Plains Indians of North America are some examples. Central to all these cultures are marked feelings of inferiority, conflict, and anxiety among the member-participants after being exposed to other, superior cultures and an attempt to renovate their self-images. Underlying these non-logical, magic-religious endeavours is a strong wish for resolution of their social, economic, and political problems and for a new and better way of life, such as that of the invading, superior cultures.

One kind of crisis cult is the cargo cult that has repeatedly arisen in Melanesia over the past century as a means of obtaining the manufactured articles possessed by European invaders.[15] Cargo is a neo-Melanesian or pidgin word that designates all of the manufactured goods, including canned foods and weapons, possessed by the Europeans, which are greatly desired by the indigenous people. Without knowing how the cargo was manufactured in the home countries of Europe, based on their own folk beliefs, the local people thought that it was given to the white people by their powerful ancestors through the performance of proper rituals. Accordingly, the local people tried to perform the white people's rituals, in the hopes that their ancestors would send them a lot of cargo and their lives would eventually be full of wealth. This behaviour might be individual, or it might involve a group of followers that gave up their normal lives to perform religious rituals, waiting for the arrival of the cargo, not only for several months, but for many years. They would become collectively deluded and led by a cult leader. As a culture-related specific syndrome, it may be understood that culture contributed to the stress that was encountered and also shaped the unique, pathological pattern of coping with it, a combination of pathogenic and pathoselective effects.

Culture-shaped variations of psychopathology (pathoplastic effect)

This category includes a group of disorders that manifests a clinical picture that is considerably different from the ordinal symptomatology of identified disorders described in current psychiatric classifications (of Euro-American origin). It is considered that the uniqueness of the symptomatology may be culturally attributed, i.e. due to cultural pathoplastic effects. Culture affects not only the content of symptoms, but, even more, the total clinical picture by the absence, addition, or variation of symptoms, resulting in considerable change in the manifestations of variations or subtypes of universally recognized psychiatric disorders.

(a) Anthropophobia (interpersonal relation phobia)

(i) Definition and nature of the disorder

Anthropophobia is the English translation for the Japanese term *taijin-kyofu-shio*. In Japanese, *taijin* means interpersonal, *kyofu* means phobic, and *shio* means syndrome or disorder. Therefore, *taijin-kyofu-shio* literally means the disorder with fear of interpersonal relation.[16] *Taijinkyofushio* is said to be prevalent among Japanese and is considered a culture-related psychiatric disorder. According to the clinical study, the onset of illness was as early mostly between 15 and 25 years, more prevalent among males than female. The cardinal symptoms manifested by the patients are: fear of one's bodily odours, fear of flushing, fear of showing odd attitudes towards others, fear of eye contact with others, concern about others' attitudes towards oneself, and fear of body dysmorphia.

(ii) Dynamic interpretation and culture formulation

The characteristic of *taijinkyofushio* is that the fear is induced in the presence of classmates, colleagues, or friends, those who are neither particularly close (such as family members) nor totally strange (such as people in the street). In other words, subjects are concerned with how to relate to people of intermediate familiarity. It is towards these people that a person must exercise delicate social etiquette. This is different from social phobia described in western societies, where patients have fear of socializing with strangers.

Culturally it has been explained that Japan is a situation-oriented society, very much concerned with how others see one's behaviour. It is considered that the act of staring at the person to whom one is talking is quite extraordinary and considered to be rude. Thus, there are cultural characteristics that cause Japanese to be hypersensitive about looking at and being looked at.

Taijinkyofushio is a psychological disorder of the adolescent. It is closely related to the problems associated with psychological development in the area of socialization. The Japanese child is raised in an atmosphere of indulgence and trust. However, when this protected child enters the wider world of junior high school, he or she faces multiple tasks: coping with conflict between biological needs and social restrictions, personal identity problems, and an increasing need for acceptance and love by others in social settings. This intensifies a feeling of unworthiness, making him or her more concerned about others' sensibilities and reactions.

(b) Brain fag syndrome

This syndrome was described originally as a very common minor psychiatric disorder occurring among the students of Southern Nigeria. Clinically, it is characterized by subjective complaints of intellectual impairment, (visual) sensory impairment, and somatic complaints, mostly of pain or a burning sensation in the head and neck. The student-patients often used the term brain fag to complain that they were no longer able to read, grasp what they were reading, or recall what they had just read, basically stressing their difficulty in mental function. Therefore, the term brain fag syndrome was suggested for this distinct clinical mental condition.[17]

In Nigeria, education was often a family affair, in which one of the brighter children was supported financially by family members; the educated member in turn was expected to be responsible for other family members when the need arose. Because of this family aspect of education, the student was burdened by the responsibility of maintaining the family's prestige. Thus, his or her academic

success or failure was associated with great stress. Therefore, it can be said that this syndrome is adjustment disorder with somatic feature and its symptomatology is shaped by culture.

(c) Arctic hysteria (pibloktoq)

Arctic (or polar) hysteria, also known by the local name *pibloktoq*, refers to a unique hysterical attack observed among the polar Eskimo people living in Arctic areas. The clinical condition is characterized by the sudden onset of loss or disturbance of consciousness. During the attack, as the patient may show various abnormal behaviours, such as tearing off his or her clothing, glossolalia, fleeing (nude or clothed), rolling in the snow, throwing anything handy around, performing mimetic acts, convulsion, or other bizarre behaviour. This emotional outburst occurs predominantly in women, but occasionally among men.[18]

No specific precipitating causes are noted. It has been speculated that the reaction is a manifestation of the basic Eskimo personality. Because the reaction is prevalent in winter, it is also thought that it may be related to increased threats of starvation or higher accident rates. Generally, it is suspected that the disorder is due to some basic, underlying anxiety, triggered by severe, culturally typical stresses: fear of certain impending situations, fear of loss, or fear of losing emotional support, including the sense of being on safe, solid, familiar ground.

(d) Malignant anxiety

A special, intensified form of anxiety disorder, termed as malignant anxiety, was reported to occur in Africa.[19] The condition was characterized by intense anxiety, extreme irritability, restlessness, and intense fear, and, therefore, named it malignant anxiety. It was referred to as frenzied anxiety as well. Often, the patient claimed that there was a change in his sense of self and reality, but there was no sign of personality deterioration or disintegration and no latent or overt psychotic symptoms. However, patients often suffered from intense feelings of anger that led to homicidal behaviour. The condition usually occurred in sporadic cases, but occasionally as an endemic. The disorder was thought to be situation related, associated with problems of adaptation to new and stressful life situations. Very often the patients were culturally marginal persons, who were in the process of renouncing their age-old cultures, but had failed to assimilate the new.

Culturally elaborated unique behaviour (pathoelaborating effect)

(a) Latah (startle-induced dissociative reaction)

(i) Defining the condition

Latah is a Malay word referring to the condition in which a person, after being startled by external stimuli, such as being tickled, suddenly experiences an altered consciousness and falls into a transient, dissociated state, exhibiting unusual behaviour (such as echolalia, echopraxia, or command automatism), including explosive verbal outbursts, usually of erotic words that are not ordinarily acceptable.[20] Beside Malaysia, the phenomenon has been observed in other places around the world and has been given various folk names: in Burma (where it is called *yaun*), Indonesia, Thailand (*bah-tsche*), the Philippines (*mali-mali*), indigenous tribes in Siberia, Russia (*myriachit*), and among the Ainu in Japan (who call it *imu*).

The *latah* reaction is found predominantly among women, although men may occasionally be involved. It was found that *latah* tends to run in families. In the past, it was primarily young women who were involved in *latah*. Most of the subjects found now are beyond middle age. Most cases are found in rural areas. Some develop *latah* reactions insidiously, without any precipitating events, whereas others symptoms occur after they endure psychologically stressful events. The loss of a significant person usually occurs shortly before the first experience of *latah* reaction. Once the reaction is experienced, it becomes habitual, and, thereafter, any sudden stimulation may provoke it. Hearing a sudden noise or being suddenly touched or poked by others may cause *latah*. Throwing a rope or other snake-like object in front of the person or simply shouting snake! will sufficiently startle the person to start a *latah* reaction.

The condition may last for several minutes or several hours if the person is continuously provoked. After the dissociative reaction is over, the subject usually claims amnesia and is puzzled about what had happened. Often, the subject is very apologetic and embarrassed for the (socially) inappropriate things he or she may have said (sexually coloured, dirty words) or done (such as touching men) during the attack.

(ii) Aetiological speculations

Anthropologists have tried to understand the *latah* condition from a cultural point of view. It has been pointed out that the *latah* reaction is remarkably congruent with the cultural themes emphasized, but in a paradoxical way. It is a peculiarly appropriate means of communicating marginality to others. The *latah* subjects are engaging in a performance, a role, and theatre, a culture-specific idiom expressing marginality while simultaneously reaffirming normative boundaries. The traditional polygamous Malay extended family structure is male dominated. Within this cultural system, *latah* is socially accepted as a female attention-seeking response, one of the few permissible overt, excitable, aggressive, and/or sexual demonstrations. In other words, *latah* is a culturally sanctioned emotional outlet for females. It is also a culturally elaborated unique behaviour.[21]

Cultural influence of prevalent occurrence of disorders (pathofacilitating effects)

This category includes several conditions that are commonly known as psychiatric disorders. There is nothing particular about them in terms of clinical manifestation (thus, they are not *specific* or *unique* disorders). However, their prevalence is influenced strongly enough by cultural factors that they may be viewed as heavily culture-related syndromes, rather than merely *ordinary* psychiatric disorders. Among this group of disorders, massive hysteria, group suicide, alcohol-related problems, or substance abuse are some of the examples.

Cultural interpretation of certain mental conditions (pathoreactive effect)

(a) Ataques de nervios (attack of nerves)

The folk name, *ataques de nervios*, literally meaning attack of nerves, refers to a stress-induced, culturally shaped unique emotional reaction with mixed anxiety-hysterical features.[22] This is an illness category used frequently by Hispanic people.

Initially observed among Puerto Rican army recruits, it was also labelled Puerto Rican syndrome. This condition typically occurs at funerals, in accidents, or in family conflicts, and calls forth family or other social supports. Commonly experienced symptoms include shaking, palpitations, a sense of heat rising to the head, and numbness, symptoms resembling a panic attack. The individual may shout, swear, and strike out at others, and finally fall to the ground, manifesting convulsion-like movements.

Based on clinical study, it has been reported that most of the patients (about 80 per cent) were female. From a clinical, diagnostic point of view, according to DSM-III-R criteria, the condition belongs to many subtypes of disorders, including panic disorder, recurrent major depression, generalized anxiety disorder, non-specific anxiety disorder, and others. Because the clinical picture is of a mixed, rather than a specific, nature, it may be interpreted simply as a folk label for an emotional reaction based on psychoreactive effect. It may be understood as an acute episode of social and psychological distress related to upsetting or frightening events in the family sphere. Focusing on symptoms alone misses what is most salient and meaningful about illness categories.

(b) Hwabyung (fire sickness)

Hwa-byung in the Korean language literally means fire (*hwa*) sickness (*byung*). Based on a traditional Chinese medical concept that is still prevalent in Korea, that an imbalance among the five elements within the body (metal, wood, water, fire, and earth) may cause physical disorders, laypeople in Korea use the folk term fire sickness to describe ill conditions. This is a folk idiom of distress characterized by a wide range of somatic and emotional symptoms. About three-fourths of the patients that complained of *hwabyung* were women, who linked their conditions to anger provoked by domestic problems, such as their husbands' extramarital affairs and strained in-law relationships. Culturally, it has been explained that male chauvinism has always been dominant in Korean society, and women tend to suffer from their vulnerable status. When a housewife is mistreated by her husband or is having troubles with her in-laws, she has to suppress her emotional reactions so that there will be no disturbance in the stability of the family. As a woman, she is taught to accept defeat, bear frustration, and suppress her hatred. As a result, accumulated resentment (*hahn* in Korean) becomes a major issue for some women. This is the core dynamic for understanding *hwabyung*.[23]

Cultural factors may indirectly contribute to the occurrence of particular psychological problems that are encountered by Korean women, but through pathoreactive effect, emotional reaction was labelled by laymen as an indication of fire sickness (*hwabyung*).

(c) Susto (soul loss)

Susto is a Spanish word that literally means fright. The term is widely used by people in Latin America to refer to the condition of loss of soul.[24] It is based on the folk belief that every individual possesses a soul, but, through certain experiences, such as being frightened or startled, a person's soul may depart from the body. As a result, the soul-lost person will manifest certain morbid mental conditions and illness behaviour. The remedy for such a condition is to recapture the soul through certain rituals. The concept of loss of soul as a cause for sickness is prevalent around the world, and that terms similar, or equivalent, to *susto* are found widely distributed across many different cultural groups, such as *el miedo* (fright) in Bolivia, *lanti* in the Philippines, or *mogo laya* in Papua New Guinea.

It should be pointed out that, although the cause is attributed to spiritual-psychological reasons relating mostly to a frightening experience or misfortune, from a clinical point of view, the manifested syndrome is quite heterogeneous, without a commonly shared syndrome. The victim may manifest loss of appetite, sleep disturbance, reduced strength, absentmindedness, headache, dizziness, or other somatic symptoms, as well as emotional symptoms of depression, anxiety, or irritability. Therefore, strictly speaking, it is not a culture-related specific syndrome derived from psychogenic or psychoplastic effects. It is culture-related *only* in the sense that the morbid condition is interpreted after the fact according to folk concepts of aetiology and certain ways of regaining the lost soul, such as rituals, are offered. Therefore, the role of culture is interpretation of and reaction to the illness.

Final comments

Diagnosis and classification issues

Associated with the increased awareness of the impact of culture on psychiatric classifications, there is controversy regarding how to deal with culture-related specific syndromes from a diagnostic point of view.[12] Some clinicians feel strongly that various known culture-bound syndromes (such as *koro* or *hwabyung*) should be officially recognized and included in the present classification system.

However, it needs to be pointed out that the present DSM classification system is based on the descriptive approach, categorizing psychiatric disorders by certain sets of manifested symptomatology. If we try to fit culture-related specific syndromes into the categories of the existing classification system or try to create new categories of disorders, they will be classified as NOS (not otherwise specified) or, at best, as variations of presently recognized disorders. Most importantly, by squeezing the culture-related specific syndromes into the descriptive-oriented classification system, we will lose the unique meaning of the syndromes from a cultural perspective.[25]

Culture-related specific syndromes in western societies

Another point that must be made is that, by definition, culture-related specific syndromes should be able to be discovered everywhere, as every society, no matter East or West, has its own culture. However, the trend has been to consider that most culture-related specific syndromes (such as *koro*, *amok*, or *dhat* syndromes) occur in non-western societies. This is because they were considered peculiar phenomena observed in areas previously colonized by western people and they simply did not fit the classification system developed for Euro-American populations. This trend is now changing. There is an increased interest in recognizing syndromes in our own western cultures that are heavily culture-related.[26] Several psychiatric disorders have been suggested by various scholars for consideration as western culture-related syndromes. These include: anorexia nervosa, obesity, drug-induced dissociated states, and multiple personality, disorders that are seldom observed or concerned in non-western societies. Because these conditions are already recognized in the existing western nosological system, they are, in a sense, not viewed as specific syndromes. However,

they can be viewed as culture-related psychiatric conditions that are influenced by the pathoelaborative, pathofacilitative, or pathoplastic effects of western culture.

Clinical implications for general psychiatry

Even though the encounter of culture-related specific psychiatric disorder in our daily psychiatric practice is relatively rare, the purpose of examining such specific syndromes has its significant purpose and implications. Through such unique examples, it helps us to appreciate the cultural attribution to the stress formation, reaction pattern, symptom manifestation, occurrence of frequency of disorders, and reaction to the disorders. It also concerns how to work on therapy for the disorder by complying patient's cultural background.

Further information

Tseng, W.S. (2001). D: 13. Culture-related specific syndromes. In *Handbook of cultural psychiatry* (ed. W.S. Tseng), pp. 211–63. Academic Press, San Diego.

References

1. Yap, P.M. (1967). Classification of the culture-bound reactive syndromes. *The Australian and New Zealand Journal of Psychiatry*, **1**, 172–9.
2. Jilek, W.G. and Jilek-Aall, L. (2001). Culture-specific mental disorders. In *Contemporary psychiatry. Psychiatry in special situations*, Vol. 2 (eds. F. Henn, N. Sartorius, H. Helmchen, *et al.*), pp. 219–45. Springer, Berlin.
3. Tseng, W.S. (2001). *Handbook of cultural psychiatry*. Academic Press, San Diego.
4. Tseng, W.S. (2006). From peculiar psychiatric disorders through culture-bound syndromes to culture-related specific syndromes. *Transcultural Psychiatry*, **43**, 554–76.
5. Westermeyer, J. (1973). On the epidemicy of *amok* violence. *Archives of General Psychiatry*, **28**, 873–6.
6. Prince, R. and Tcheng-Laroche, F. (1987). Culture-bound syndromes and international disease classifications. *Culture, Medicine and Psychiatry*, **11**, 3–19.
7. Chowdhury, A.N. (1994). Koro in females: an analysis of 48 cases. *Transcultural Psychiatric Research Review*, **31**, 369–80.
8. Tseng, W.S., Mo, G.M., Hsu, J., *et al.* (1988). A sociocultural study of koro epidemics in Guandong, China. *The American Journal of Psychiatry*, **145**, 1538–43.
9. Malhotra, H.K. and Wig, N.N. (1975). Dhat syndrome: a culture-bound sex neurosis of the orient. *Archives of Sexual Behavior*, **4**, 519–28.
10. Neki, J.S. (1973). Psychiatry in south-east Asia. *The British Journal of Psychiatry*, **123**, 256–69.
11. Bhatia, M.S. and Malik, S.C. (1991). Dhat syndrome—a useful diagnostic entity in Indian culture. *The British Journal of Psychiatry*, **159**, 691–5.
12. Hughes, C.C. (1996). The culture-bound syndromes and psychiatric diagnosis. In *Culture and psychiatric diagnosis*: *a DSM-IV perspective* (eds. J.E. Mezzich, A. Kleinman, H. Fabrega, Jr, *et al.*), pp. 289–307. American Psychiatric Press, Washington, DC.
13. Yap, P.M. (1951). Mental diseases peculiar to certain cultures: a survey of comparative psychiatry. *The Journal of Mental Science*, **97**, 313–27.
14. Westermeyer, J. (1972). A comparison of *amok* and other homicide in Laos. *The American Journal of Psychiatry*, **129**, 703–9.
15. Burton-Bradley, B.G. (1975). Cargo cult. In *Stone age crisis: a psychiatric appraisal* (ed. B.G. Burton-Bradley), pp. 10–31. Vanderbilt University Press, Nashville, TN.
16. Kitanishi, K. and Mori, A. (1995). Morita therapy: 1919 to 1995. *Psychiatry and Clinical Neurosciences*, **13**, 31–7.
17. Prince, R. (1960). The brain fag syndrome in Nigerian students. *The Journal of Mental Science*, **104**, 559–70.
18. Foulks, E.F. (1972). The Arctic hysterias of the North Alaskan Eskimo. In *Anthropological studies* (No. 10) (ed. D.H. Maybury Lewis). American Anthropological Association, Washington, DC.
19. Lambo, T.A. (1962). Malignant anxiety: a syndrome associated with criminal conduct in Africans. *The Journal of Mental Science*, **108**, 256–64.
20. Simons, R.C. (1980). The resolution of the latah paradox. *The Journal of Nervous and Mental Disease*, **168**, 195–206.
21. Simons, R.C. (1996). *Boo!—culture, experience, and the startle reflex*. Oxford University Press, New York.
22. Guarnaccia, P.J. (1993). Ataques de nervios in Puerto Rico: culture-bound syndrome or popular illness? *Medical Anthropology*, **15**, 157–70.
23. Min, S.W.K. (2006). Hwabyung: a culture-related chronic anger syndrome of Korea. The First World Congress of Cultural Psychiatry Proceeding. (S-III-23).
24. Rubel, A.J. (1964). The epidemiology of a folk illness: *Susto* in Hispanic America. *Ethology*, **3**, 268–83.
25. Pfeiffer, W.M. (1982). Culture-bound syndromes. In *Culture and psychopathology* (ed. I. Al-Issa). University Park Press, Baltimore.
26. Littlewood, R. and Lipsedge, M. (1986). The culture-bound syndromes of the dominant culture: culture, psychopathology and biomedicine. In *Transcultural psychiatry* (ed. J.L. Cox), pp. 253–73. Croom Helm, London.

This page intentionally left blank

New Oxford Textbook of

Psychiatry

This page intentionally left blank

VOLUME 2

New Oxford Textbook of Psychiatry

SECOND EDITION

Edited by

Michael G. Gelder
Emeritus Professor of Psychiatry,
Warneford Hospital, University of Oxford,
Oxford, UK

Nancy C. Andreasen
Director, Mental Health Clinical Research Centre,
University of Iowa Hospital and Clinic,
Iowa City, USA

Juan J. López-Ibor Jr.
Professor of Psychiatry,
Complutense University, Madrid, Spain

and

John R. Geddes
Professor of Epidemiological Psychiatry
University of Oxford, Warneford Hospital,
Oxford, UK

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford ox2 6dp

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First edition published 2000
Reprinted 2003

This edition published 2009

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloguing in Publication Data

Data available

Typeset in Cepha Imaging Pvt. Ltd., Bangalore, India
Printed in Italy on acid-free paper by Rotolito Lombarda SpA

ISBN 978-0-19-920669-8
ISBN 978-0-19-920669-8 (set)
ISBN 978-0-19-955992-3 (Volume 1)
ISBN 978-0-19-955993-0 (Volume 2)

10 9 8 7 6 5 4 3 2 1

Preface to the second edition

This new edition, like the first, aims to present a comprehensive account of clinical psychiatry with reference to its scientific basis and to the ill person's perspective. As in the first edition, the authors are drawn from many countries, including the UK, the USA, 12 countries in continental Europe, and Australasia. The favourable reception of the first edition has led us to invite many of the original authors to revise their chapters for this second edition but 50 chapters are the work of new authors, many concerned with subjects that appeared in the first edition, while others are completely new. The forensic psychiatry section has the most new chapters, followed by the section on psychology as a scientific basis of psychiatry.

The overall plan of the book resembles that of the first edition (see preface to the 1st edition, reprinted on pages vii and viii). One important feature is that information about treatment appears in more than one place. The commonly used physical and psychological treatments are described in Section 6. Their use in the treatment of any particular disorder is considered in the chapter concerned with that disorder and the account is in two parts. The first part is a review of evidence about the effects of each of the treatments when used for that disorder. The second part, called Management, combines evidence from clinical trials with accumulated clinical experience to produce practical advice about the day to day care of people with the disorder.

Although much information can now be obtained from internet searches, textbooks are still needed to provide the comprehensive account of established knowledge into which new information can be fitted and against which recent findings can be evaluated. As well as seeking to provide an authoritative account of essential knowledge, each chapter in the new edition includes a brief list of sources of further information, including where appropriate, regularly updated web sites.

An essential component of good practice is the need to be aware of patients' perspectives, to respect their wishes, and to work with them, and often their families, as partners. The book opens with an important chapter on the experience of being a patient, and there are chapters on stigma, ethics, and the developing topic of values-based practice.

We are grateful to the following who advised us about parts of the book; Professor John Bancroft (Psychosexual Disorders), Professor Tom Burns (Social and Community Psychiatry), Professor William Fraser (Intellectual Disability), Professor Keith Hawton (Suicide and Deliberate Self Harm), Professor Susan Iversen (Psychology), Professor Robin Jacoby (Old Age Psychiatry), Professor Paul Mullen (Forensic Psychiatry), Sir Michael Rutter (Child and Adolescent Psychiatry), and Professor Gregory Stores (Sleep Disorders).

The editors

This page intentionally left blank

Preface to the first edition

Three themes can be discerned in contemporary psychiatry: the growing unity of the subject, the pace of scientific advance, and the growth of practice in the community. We have sought to reflect these themes in the *New Oxford Textbook of Psychiatry* and to present the state of psychiatry at the start of the new millennium. The book is written for psychiatrists engaged in continuous education and recertification; the previous, shorter, *Oxford Textbook of Psychiatry* remains available for psychiatrists in training. The book is intended to be suitable also as a work of reference for psychiatrists of all levels of experience, and for other professionals whose work involves them in the problems of psychiatry.

The growing unity of psychiatry

The growing unity in psychiatry is evident in several ways. Biological and psychosocial approaches have been largely reconciled with a general recognition that genetic and environmental factors interact, and that psychological processes are based in and can influence neurobiological mechanisms. At the same time, the common ground between the different psychodynamic theories has been recognized, and is widely accepted as more valuable than the differences between them.

The practice of psychiatry is increasingly similar in different countries, with the remaining variations related more to differences between national systems of health care and the resources available to clinicians, than to differences in the aims of the psychiatrists working in these countries. This unity of approach is reflected in this book whose authors practise in many different countries and yet present a common approach. In this respect this textbook differs importantly from others which present the views of authors drawn predominantly from a single country or region.

Greater agreement about diagnosis and nosology has led to a better understanding of how different treatment approaches are effective in different disorders. The relative specificity of psychopharmacological treatments is being matched increasingly by the specificity of some of the recently developed psychological treatments, so that psychological treatment should no longer be applied without reference to diagnosis, as was sometimes done in the past.

The pace of scientific advance

Advances in genetics and in the neurosciences have already increased knowledge of the basic mechanisms of the brain and are beginning to uncover the neurobiological mechanisms involved in psychiatric disorder. Striking progress has been achieved in the understanding of Alzheimer's disease, for example, and there are indications that similar progress will follow in uncovering the causes of mood disorder, schizophrenia, and autism. Knowledge of genetics and the neurosciences is so extensive and the pace of change is so rapid that it is difficult to present a complete account within the limited space available in a textbook of clinical psychiatry. We have selected aspects of these sciences that seem, to us and the authors, to have contributed significantly to psychiatry or to be likely to do so before long.

Psychological and social sciences and epidemiology are essential methods of investigation in psychiatry. Although the pace of advance in these sciences may not be as great as in the neurosciences, the findings generally have a more direct relation to clinical phenomena. Moreover, the mechanisms by which psychological and social factors interact with genetic, biochemical, and structural ones will continue to be important however great the progress in these other sciences. Among the advances in the psychological and social sciences that are relevant to clinical phenomena, we have included accounts of memory, psychological development, research on life events, and the effects of culture. Epidemiological studies continue to be crucial for defining psychiatric disorders, following their course, and identifying their causes.

Psychiatry in the community

In most countries, psychiatry is now practised in the community rather than in institutions, and where this change has yet be completed, it is generally recognized that it should take place. The change has done much more than transfer the locus of care; it has converted patients from passive recipients of care to active participants with individual needs and preferences. Psychiatrists are now involved in the planning, provision, and evaluation of services for whole communities, which may include members of ethnic minorities, homeless people, and refugees. Responsibility for a community has underlined the importance of the prevention as well as the treatment of mental disorder and of the role of agencies other than health services in both. Care in the community has also drawn attention to the many people with psychiatric disorder who are treated in primary care, and has led to new ways of working between psychiatrists and physicians. At the same time, psychiatrists have

worked more in general hospitals, helping patients with both medical and psychiatric problems. Others have provided care for offenders.

The organization of the book

In most ways, the organization of this book is along conventional lines. However, some matters require explanation.

Part 1 contains a variety of diverse topics brought together under the general heading of the subject matter and approach to psychiatry. Phenomenology, assessment, classification, and ethical problems are included, together with the role of the psychiatrist as educator and as manager. Public health aspects of psychiatry are considered together with public attitudes to psychiatry and to psychiatric patients. Part 1 ends with a chapter on the links between science and practice. It begins with a topic that is central to good practice—the understanding of the experience of becoming a psychiatric patient.

Part 2 is concerned with the scientific foundations of psychiatry grouped under the headings neurosciences, genetics, psychological sciences, social sciences, and epidemiology. The chapters contain general information about these sciences; findings specific to a particular disorder are described in the chapter on that disorder. Brain imaging techniques are discussed here because they link basic sciences with clinical research. As explained above, the chapters are selective and, in some, readers who wish to study the subjects in greater detail will find suggestions for further reading.

Part 3 is concerned with dynamic approaches to psychiatry. The principal schools of thought are presented as alternative ways of understanding the influence of life experience on personality and on responses to stressful events and to illness. Some reference is made to dynamic psychotherapy in these accounts, but the main account of these treatments is in Part 6. This arrangement separates the chapters on the practice of dynamic psychotherapy from those on psychodynamic theory, but we consider that this disadvantage is outweighed by the benefit of considering together the commonly used forms of psychotherapy.

Part 4 is long, with chapters on the clinical syndromes of adult psychiatry, with the exception of somatoform disorders which appear in Part 5, Psychiatry and Medicine. This latter contains more than a traditional account of psychosomatic medicine. It also includes a review of psychiatric disorders that may cause medical symptoms unexplained by physical pathology, the medical, surgical, gynaecological, and obstetric conditions most often associated with psychiatric disorder, health psychology, and the treatment of psychiatric disorder in medically ill patients.

Information about treatment appears in more than one part of the book. Part 6 contains descriptions of the physical and psychological treatments in common use in psychiatry. Dynamic psychotherapy and psychoanalysis are described alongside counselling and cognitive behavioural techniques. This part of the book contains general descriptions of the treatments; their use for a particular disorder is considered in the chapter on that disorder. In the latter, the account is generally in two parts: a review of evidence about the efficacy of the treatment, followed by advice on management in which available evidence is supplemented, where necessary, with clinical experience. Treatment methods designed specially for children and adolescents, for people with mental retardation (learning disability), and for patients within the forensic services are considered in Parts 9, 10, and 11 respectively.

Social psychiatry and service provision are described in Part 7. Public policy issues, as well as the planning, delivery, and evaluation of services, are discussed here. Psychiatry in primary care is an important topic in this part of the book. There are chapters on the special problems of members of ethnic minorities, homeless people, and refugees, and the effects of culture on the provision and uptake of services.

Child and adolescent psychiatry, old age psychiatry, and mental retardation are described in Parts 8, 9, and 10. These accounts are less detailed than might be found in textbooks intended for specialists working exclusively in the relevant subspecialty. Rather, they are written for readers experienced in another branch of psychiatry who wish to improve their knowledge of the special subject. We are aware of the controversy surrounding our choice of the title of Part 10. We have selected the term 'mental retardation' because it is used in both ICD-10 and DSM-IV. In some countries this term has been replaced by another that is thought to be less stigmatizing and more acceptable to patients and families. For example, in the United Kingdom the preferred term is 'learning disability'. While we sympathize with the aims of those who adopt this and other alternative terms, the book is intended for an international readership and it seems best to use the term chosen by the World Health Organization as most generally understood. Thus the term mental retardation is used unless there is a special reason to use another.

In Part 11, Forensic Psychiatry, it has been especially difficult to present a general account of the subject that is not tied to practice in a single country. This is because systems of law differ between countries and the practice of forensic psychiatry has to conform with the local legal system. Although many of the examples in this part of the book may at first seem restricted in their relevance because they are described in the context of English law, we hope that readers will be able to transfer the principles described in these chapters to the legal tradition in which they work.

Finally, readers should note that the history of psychiatry is presented in more than one part of the book. The history of psychiatry as a medical specialty is described in Part 1. The history of ideas about the various psychiatric disorders appears, where relevant, in the chapters on these disorders, where they can be considered in relation to present-day concepts. The history of ideas about aetiology is considered in Part 2, which covers the scientific basis of psychiatric aetiology, while the historical development of dynamic psychiatry is described in Part 3.

Michael Gelder
Juan López-Ibor
Nancy Andreasen

Acknowledgements from the first edition

We are grateful to the many colleagues who have advised us about certain parts of the book.

The following helped us to plan specialized parts of the book: Dr Jeremy Holmes (Section 3, Psychodynamic Contributions to Psychiatry); Professor Richard Mayou (Section 5, Psychiatry and Medicine); Professor Robin Jacoby (Section 8, Psychiatry of Old Age); Sir Michael Rutter (Section 9, Child and Adolescent Psychiatry); Professor William Fraser (Section 10, Intellectual Disablity); Professor Robert Bluglass (Section 11, Forensic Psychiatry).

The following helped us to plan certain sections within Section 4, General Psychiatry: Professor Alwyn Lishman (delirium, dementia, amnestic syndrome, and other cognitive disorders); Professor Griffith Edwards (alcohol use disorders); Dr Philip Robson (other substance use disorders); Professor Guy Goodwin (mood disorders); Professor John Bancroft (sexuality, gender identity, and their disorders); Professor Gregory Stores (sleep–wake disorders); Professor Keith Hawton (suicide and attempted suicide). In Section 6, Professor Philip Cowen advised about somatic treatments, Dr Jeremy Holmes about psychodynamic treatments, and Professor David Clark about cognitive behavioural therapy. Dr Max Marshall provided helpful advice about forensic issues for Section 7. We also thank the many other colleagues whose helpful suggestions about specific problems aided the planning of the book.

Finally, we record our special gratitude to the authors and to the staff of Oxford University Press.

This page intentionally left blank

Contents
Volume 1

Contents
Volume 2

This page intentionally left blank

Contributors list

S.E. Abbey Associate Professor of Psychiatry, University of Toronto, Toronto, Canada
Chapter 5.3.6

Henry David Abraham Distinguished Life Fellow, American Psychiatric Association
Chapter 4.2.3.3

Clive E. Adams Cochrane Schizophrenia Group, University of Oxford Department of Psychiatry, Warneford Hospital, Oxford, UK
Chapter 6.1.1

Hagop S. Akiskal Professor of Psychiatry and Director of the International Mood Center, University of California at San Diego, California, USA
Chapter 4.5.9

Nancy C. Andreasen Dept of Psychiatry, University of Iowa Hospitals & Clinics, Iowa City, USA
Chapter 4.3.1

Adrian Angold Associate Professor of Child and Adolescent Psychiatry, Duke University Medical Center, Durham, North Carolina, USA
Chapter 9.1.2

Jules Angst Emeritus Professor of Psychiatry, Zurich University
Chapter 4.5.7

Ella Arensman Director of Research, National Suicide Research Foundation, Ireland
Chapter 4.15.2

J.K. Aronson Reader in Clinical Pharmacology, University Department of Primary Health Care, Headington, Oxford
Chapter 6.2.1

José-Luis Ayuso-Mateos Chairman, Department of Psychiatry, Universidad Autónoma de Madrid, Hospital Universitario de la Princesa, Spain
Chapter 5.3.5

Susan Bailey Consultant Child and Adolescent Forensic Psychiatrist, Salford NHS Trust and Maudsley NHS Trust; Senior Research Fellow, University of Manchester, UK
Chapter 11.7

Robert Baldwin Consultant, Old Age Psychiatrist, and Honorary Senior Lecturer, Manchester Royal Infirmary, UK
Chapter 8.5.4

James C. Ballenger Retired Professor and Chairman, Department of Psychiatry and Behavioral Sciences and Director, Institute of Psychiatry, Medical University of South Carolina
Chapter 4.7.3

John Bancroft, The Kinsey Institute for Research in Sex, Gender, & Reproduction and Department of Psychiatry, University of Oxford
Chapter 4.11.2

David H. Barlow Center for Anxiety and Related Disorders at Boston University, Massachusetts, USA
Chapter 4.7.1

Delwyn J. Bartlett Woolcock Institute of Medical Research, Sydney, Australia
Chapter 4.14.2

Christopher Bass Consultant in Liaison Psychiatry, John Radcliffe Hospital, Oxford, UK
Chapters 5.2.4 and 5.2.9

Antony W. Bateman Halliwick Psychotherapy Dept, St Ann's Hospital, London, UK
Chapter 4.12.6

Per Bech Professor of Psychiatry and Head of Psychiatric Research Unit, WHO Collaborating Centre, Frederiksborg General Hospital, Hillerød, Denmark
Chapter 4.5.2

Sidney Benjamin Senior Lecturer, University of Manchester, UK
Chapter 5.2.6

Thomas P. Berney Consultant Developmental Psychiatrist Honorary Research Associate, University of Newcastle upon Tyne
Chapter 10.6

Jose M. Bertolote Chief, Mental Disorders Control Unit, World Health Organization, Geneva; Associate Professor, Department of Psychogeriatrics, University of Lausanne, Switzerland
Chapter 7.4

Zubin Bhagwagar CT Mental Health Center, Yale University, New Haven CT, USA
Chapter 6.2.3

Mary Beth Connolly Gibbons Assistant Professor of Psychology in Psychiatry Department of Psychiatry, University of Pennsylvania, Pennsylvania, USA
Chapter 6.1.2

Dinesh Bhugra Professor of Mental Health and Cultural Diversity, King's College London, Institute of Psychiatry, London, UK
Chapter 7.10.3

Michel Billiard Professor of Neurology, School of Medicine, Guide Chauliac Hospital, Montpellier, France
Chapter 4.14.3

Max Birchwood Director, Early Intervention Service, Northern Birmingham Mental Health Trust, and University of Birmingham, UK
Chapter 6.3.2.4

Boris Birmaher UPMC Western Psychiatric Institute, Pittsburgh, USA
Chapter 9.2.7

Stella Bitran, Center for Anxiety and Related Disorders, Boston University, Beacon, MA
Chapter 4.7.1

Alan H. Bittles Centre for Comparative Genomics, Murdoch University, Perth, Australia
Chapter 8.1

Dora Black Honorary Consultant, Child and Adolescent Psychiatry, Traumatic Stress Clinic, London; Honorary Lecturer, University of London, UK
Chapters 9.3.2 and 9.3.7

Michelle A. Blackmore, Doctoral Student of Clinical Psychology Adult Anxiety Clinic at Temple University, Philadelphia, Pennsylvania, USA
Chapter 4.7.2

Carlos Blanco New York State Psychiatric Institute, New York,
Chapter 6.3.3

Sidney Bloch Professor of Psychiatry, University of Melbourne; Senior Psychiatrist, St Vincent's Hospital, Melbourne, Australia
Chapters 1.5 and 6.3.8

William V. Bobo Assistant Professor of Psychiatry, Vanderbilt University School of Medicine Nashville, Tennessee (USA)
Chapter 6.2.5

Bernhard Bogerts Department of Psychiatry, University of Magdeburg, Germany
Chapter 2.3.5

Jeff Bostic School of Psychiatry, Harvard Medical School, Cambridge MA, USA
Chapter 9.1.3

Nick Bouras Professor, Institute of Psychiatry - King's College London MHiLD - York Clinic, London, UK
Chapter 10.9

Christopher R. Bowie Department of Psychiatry, Mount Sinai School of Medicine, New York, USA
Chapter 4.3.3

Colin Bradbury Department of Psychological Medicine, Institute of Psychiatry, De Crespigny Park, London, UK
Chapter 4.2.4

Brian P. Brennan Instructor in Psychiatry, Harvard Medical School and Associate Director for Translational Neuroscience Research, Biological Psychiatry Laboratory, McLean Hospital, Belmont, MA
Chapter 6.2.6

David Brent Dept of Psychiatry, University of Pittsburgh Medical School, Pittsburgh PA, USA
Chapter 9.2.7

Chris R. Brewin Research Dept of Clinical, Educational & Health Psychology, University College London, UK
Chapter 4.6.3

Elaine Brohan Institute of Psychiatry, David Goldberg Centre, De Crespigny Park, London, UK
Chapter 1.2

Ian Brockington Professor of Psychiatry, University of Birmingham, UK
Chapter 5.4

George W. Brown Professor of Sociology, Academic Department of Psychiatry, St Thomas's Hospital, London, UK
Chapter 2.6.1

Arne Brun Professor of Neuropathology. Department of Pathology, Lund University Hospital, Lund, Sweden
Chapter 4.1.3

Richard A. Bryant School of Psychology, University of New South Wales, Sydney NSW, Australia
Chapter 4.6.1

E.T. Bullmore Institute of Psychiatry, King's College London, UK
Chapters 2.3.7 and 2.3.8

Tom Burns Professor of social psychiatry, Dept of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK
Chapter 7.5

José Luis Carrasco Professor of Psychiatry, Hospital Fundacion Jimenez Diaz, Universidad Autonoma, Madrid, Spain
Chapter 4.12.3

D.J. Castle University of Western Australia, Fremantle, Australia
Chapter 4.3.6.1

Jonathan Chick Consultant Psychiatrist, NHS Lothian, and Senior Lecturer, Department of Psychiatry, University of Edinburgh
Chapter 4.2.2.4

Daniel Chisholm Department of Health System Financing, Health Systems and Services, World Health Organization, Geneva, Switzerland
Chapter 7.7

David Christmas Dept of Psychiatry, University of Dundee, Dundee, UK
Chapter 6.2.10.4

David M. Clarke Consultant Psychiatrist, Lea Castle Centre, Kidderminster DY10 3PP
Chapters 6.3.3.1 and 10.4

L. Clark Dept of Experimental Psychology, University of Cambridge, Cambridge, UK
Chapter 2.5.5

Judy Clegg Lecturer, Speech and language therapist, HPC, RCSLT Department of Human Communication Sciences University of Sheffield, UK
Chapter 9.2.11

C. Robert Cloninger Dept of Psychiatry, Washington University School of Medicine, St Louis MO, USA
Chapter 1.8.2

John Collinge Head of the Department of Neurodegenerative Disease at the Institute of Neurology, University College London and the Director of the UK Medical Research Council's Prion Unit, London, UK
Chapter 4.1.4

Henry O'Connell Consultant Psychiatrist, Co. Tipperary, Ireland
Chapter 8.5.2

Melissa A. Culhane Harvard Program in Refugee Trauma, Department of Psychiatry, Massachusetts General Hospital, Cambridge, USA
Chapter 7.10.1

John E. Cooper Emeritus Professor of Psychiatry, University of Nottingham, UK
Chapter 1.8.1

Sally-Ann Cooper Professor of Learning Disabilities, Division of Community Based Sciences, Faculty of Medicine, University of Glasgow
Chapter 10.2

Zafra Cooper Principal Research Psychologist, Oxford University Department of Psychiatry, Warneford Hospital, Oxford, UK
Chapters 4.10.2 and 6.3.2.2

E. Jane Costello Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Brightleaf Square, Durham NC
Chapter 9.1.2

Elizabeth Coulthard Institute of Neurology, University College London, UK
Chapter 2.5.2

F.-R. Cousin Psychiatrist, Centre Hospitalier Saint-Anne, Paris, France
Chapter 4.3.10

Philip J. Cowen Professor of Psychopharmacology, Department of Psychiatry, University of Oxford
Chapter 6.2.10.2

Tom K.J. Craig Professor of Social Psychiatry, King's College London, Institute of Psychiatry, London UK
Chapters 7.10.2 and 7.10.3

Paul Crits-Christoph Professor of Psychology in Psychiatry Director, Center for Psychotherapy Research Department of Psychiatry, University of Pennsylvania. Pennsylvania, USA
Chapter 6.1.2

Michael Crowe Consultant Psychiatrist, South London and Maudsley NHS Trust; Honorary Senior Lecturer, Institute of Psychiatry, King's College London, UK
Chapter 6.3.7

D.G. Cunningham Owens Reader in Psychiatry, Department of Psychiatry, University of Edinburgh, UK
Chapter 4.3.8

Dianne Currier Division of Molectular Imaging & Neuropathology, Department of Psychiatry, Columbia University
Chapter 4.15.3

Robert Dantzer Integrative Immunology and Behavior Program, University of Illinois at Urbana-Champaign, Edward R. Madigan Laboratory, West Gregory Drive, Urbana, IL, USA
Chapter 2.3.10

Rajan Darjee Division of Psychiatry, University of Edinburgh, Edinburgh, UK
Chapter 11.3.1

Anthony S. David Professor of Cognitive Neuropsychiatry, Institute of Psychiatry, King's College London, UK
Chapters 4.3.4 and 4.9

Kate Davidson Senior Research Psychologist, Department of Psychological Medicine, University of Glasgow, UK
Chapter 4.12.7

Martin Davies Dept of Experimental Psychology, University of Oxford, Oxford, UK
Chapter 2.1

Giovanni de Girolamo Health Care Research Agency, Emilia-Romagna Region, Bologna, Italy
Chapters 4.12.2 and 4.12.4

Shoumitro Deb Clinical Professor of Neuropsychiatry & Intellectual Disability, Division of Neuroscience, University of Birmingham, UK
Chapter 10.4

R.J. Dolan Institute of Neurology, University College London, UK
Chapter 2.5.4

Anton Došen Emeritus Professor of Psychiatric Aspects of Intellectual Disability at the Radboud University, Nijmegen, The Netherlands
Chapter 10.5.2

D. Colin Drummond Professor of Addiction Psychiatry, Section of Alcohol Research, National Addiction Centre, Division of Psychological Medicine and Psychiatry, Institute of Psychiatry, King's College London
Chapter 4.2.2.5

Emma Dunn School of Medicine, Cardiff University, Cardiff, UK
Chapter 11.17

Graham Dunn Professor of Biomedical Statistics, Health Methodology Research Group, School of Community Based Medicine, University of Manchester
Chapter 2.2

Julie Dunsmore Honorary Clinical Associate, SciMHA Unit, University of Western Sydney, Australia
Chapter 4.6.5

Anke Ehlers Department of Psychiatry, University of Oxford, UK
Chapters 4.6.1 and 4.6.2

Timo Erkinjuntti Professor of Neurology, Head of the University Department of Neurological Sciences, University of Helsinki and Head Physician, Department of Neurology and Memory Research Unit, Helsinki University Central Hospital, Finland
Chapter 4.1.8

Brigette A. Erwin Adult Anxiety Clinic of Temple University, Philadelphia, Pennsylvania, USA
Chapter 4.7.2

Colin A. Espie Professor of Clinical Psychology and Head of Department of Psychological Medicine, University of Glasgow, UK
Chapter 4.14.2

Jonathan J. Evans Section of Psychological Medicine, University of Glasgow, Glasgow, UK
Chapter 4.1.14

Christopher G. Fairburn Wellcome Principal Research Fellow and Professor of Psychiatry, University of Oxford, UK
Chapters 4.10.2 and 6.3.2.2

Peter Falkai Professor of Medical Psychology, Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany
Chapter 2.3.5

Stephen V. Faraone Director, Medical Genetics Research, Professor of Psychiatry and of Neuroscience & Physiology, Director, Child and Adolescent Psychiatry Research, SUNY Upstate Medical University, New York
Chapter 4.3.9

Michael Farrell Senior Lecturer and Consultant Psychiatrist, National Addiction Centre, South London and Maudsley NHS Trust, London, UK
Chapter 4.2.3.4

David P. Farrington Professor of Psychological Criminology, University of Cambridge, UK
Chapter 11.2

J. Paul Fedoroff Director, Sexual Behaviors Clinic Royal Ottawa Mental Health Centre and Director of Forensic Research University of Ottawa Institute of Mental Health Research
Chapter 4.11.3

Howard H. Feldman Professor and Head, Division of Neurology, Department of Medicine, University of British Columbia, Vancouver, BC, Canada
Chapter 8.5.1.1

Melanie J.V. Fennell Consultant Clinical Psychologist; Director, Oxford Diploma in Cognitive Therapy, University of Oxford Department of Psychiatry, Warneford Hospital, Oxford, UK
Chapter 6.3.2.3

Max Fink Emeritus Professor of Psychiatry and Neurology, State University of New York at Stony Brook; Professor of Psychiatry, Albert Einstein College of Medicine; Attending Psychiatrist, Long Island Jewish Medical Center, New York, USA
Chapter 6.2.10.1

Michael B. First Columbia University, New York, USA
Chapter 1.9

Per Fink Director, Research Unit for Functional Disorders, Aarhus University Hospital, Risskov, Denmark
Chapter 5.2.3

Philip A. Fisher Research Scientist, Oregon Social Learning Center, Eugene, Oregon, USA
Chapter 9.5.4

Martine F. Flament Chargée de Récherche INSERM, CNRS UMR 7593, Paris, France
Chapter 9.2.8

Simon Fleminger Consultant Neuropsychiatries, Lishman Brain Injury Unit, Maudsley Hospital, London, UK
Chapters 4.1.10 and 4.1.13

Jonathan Flint Wellcome Trust Centre for Human Genetics Roosevelt Drive, Oxford
Chapter 2.4.2

Susan Folstein Professor of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, USA
Chapter 4.1.7

Peter Fonagy Freud Memorial Professor of Psychoanalysis, University College London; Director of Research, Anna Freud Centre, London, UK; Director, Child and Family Center and Clinical Protocols and Outcomes Center, Menninger Clinic, Topeka, Kansas, USA
Chapters 4.12.6, 6.3.5 and 9.5.2

Gillian C. Forrest Consultant Child and Adolescent Psychiatrist
Chapter 9.5.8

Leah Fostick Department of Psychiatry, Chaim Sheba Medical Centre, Tel Hashomer, Israel
Chapter 4.8

W. Fraser Division of Psychological Medicine, University of Wales College of Medicine, Cardiff, UK
Introduction to Section 10

Laura Fratiglioni Aging Research Centre, Karolinska Institute, Stockholm, Sweden
Chapter 8.3

David M. Fresco Adult Anxiety Clinic of Temple University, Philadelphia, Pennsylvania, USA
Chapter 4.7.2

K.W.M. Fulford Professor of Philosophy and Mental Health, University of Warwick; Honorary Consultant Psychiatrist, University of Oxford, UK
Chapter 1.5.2

Robert Fung, Specialist Registrar in Psychiatry, Sheffield Care NHS Trust, UK
Chapter 4.2.3.2

Pier Maria Furlan Director of Department of Mental Health San Luigi Gonzaga Hospital - University of Torino, Italy
Chapter 5.5

Glen O. Gabbard Bessie Walker Callaway Distinguished Professor of Psychoanalysis and Education in the Kansas School of Psychiatry, Menninger Clinic, Topeka; Clinical Professor of Psychiatry of Kansas School of Medicine, Wichita, Kansas, USA
Chapter 3.3

Jean Garrabé Honorary President of L'Evolution psychiatrique, Paris
Chapter 4.3.10

M. Elena Garralda Professor of Child and Adolescent Psychiatry, Imperial College of Medicine, London, UK
Chapter 9.3.4

Ann Gath Formerly of University College London, UK
Chapter 10.8

John R. Geddes Professor of Epidemiological Psychiatry, Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford
Chapters 1.10 and 4.5.1

David Gill Research Fellow, Department of Psychiatry, University of Oxford, UK
Chapters 5.2.9

Kathryn J. Gill MUHB Addictions Unit, McGill University, Montreal QC, Canada
Chapter 4.2.2.1

Julia Gledhill Clinical Research Fellow, Imperial College of Medicine, London, UK
Chapter 9.3.4

David Goldberg Director of Research and Development, Institute of Psychiatry, King's College London, UK
Chapter 7.8

Terry E. Goldberg The Zucker Hillside Hospital, Glen Oaks NY, USA
Chapter 2.5.3

Guy Goodwin Professor, University Department, Warneford Hospital, Oxford
Chapter 4.5.6

Kevin Gournay Emeritus Professor, Institute of Psychiatry, King's College London
Chapter 6.4.2

Simon G. Gowers Professor of Adolescent Psychiatry, University of Liverpool
Chapter 9.5.9

Cynthia A. Graham, Oxford Doctoral Course in Clinical Psychology Warneford Hospital, Oxford and The Kinsey Institute for Research in Sex, Gender, & Reproduction
Chapter 4.11.2

Philip Graham Emeritus Professor of Child Psychiatry, Institute of Child Health, London
Chapter 9.5.3

P.M. Grasby Senior Lecturer, MRC Cyclotron Unit, Hammersmith Hospital, London, UK
Chapter 2.3.6

Richard Green Head, Gender Identity Clinic, and Visiting Professor of Psychiatry, Imperial College of Medicine at Charing Cross Hospital, London, UK; Emeritus Professor of Psychiatry, University of California, Los Angeles, California, USA
Chapters 4.11.4 and 9.2.12

Stephen Green Clinical Professor of Psychiatry, Georgetown University School of Medicine, Washington, D.C.
Chapter 1.5.1

Michael Gunn Professor of Law and Head of Department, Department of Academic Legal Studies, Nottingham Law School, Nottingham Trent University, UK
Chapter 11.1

Lars Gustafson Professor of Geriatric Psychiatry, Lund University Hospital, Lund, Sweden
Chapter 4.1.3

Francesca Guzzetta Bologna, Italy
Chapter 4.12.4

Jennifer Gutstein Department of Child Psychiatry, College of Physicians and Surgeons, Columbia University, New York, USA
Chapter 9.2.10

Sarah Harper Oxford Institute for Aging, University of Oxford, Oxford, UK
Chapter 8.2

Rex Haigh Project Lead, Community of Communities, Centre for Quality Improvement, Royal College of Psychiatrists, London; Consultant Psychiatrist, Berkshire Healthcare NHS Foundation Trust
Chapter 6.3.9

S.A. Hales Psychiatry Fellow, Princess Margaret Hospital, University Health Network, Toronto, Canada
Chapter 5.3.6

John N. Hall Professor of Mental Health, School of Health and Social Care, Oxford Brookes University, Oxford, UK
Chapter 1.8.3

Wayne Hall Professor of Public Health Policy, University of Queensland, Herston, Australia
Chapter 4.2.3.7

Edwin Harari Consultant Psychiatrist, St Vincent's Hospital, Melbourne, Australia
Chapter 6.3.8

Sarah Harper Oxford Institute for Aging, University of Oxford, Oxford, UK
Chapter 8.2

James C. Harris Director Developmental Neuropsychiatry Clinic, Professor of Psychiatry and Behavioral Sciences, Pediatrics, and Mental Hygiene, The Johns Hopkins University School of Medicine
Chapter 9.2.1

Paul J. Harrison Clinical Reader in Psychiatry, University of Oxford Department of Psychiatry, Warneford Hospital, Oxford, UK
Chapter 4.3.6.2

Allison G. Harvey Department of Experimental Psychology, University of Oxford, UK
Chapter 4.6.1

Philip D. Harvey Professor of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Woodruff Memorial Building, Atlanta, GA, USA
Chapter 4.3.3

Keith Hawton Director, Centre for Suicide Research, University Department of Psychiatry, Warneford Hospital, Oxford
Chapter 4.15.4

Richard G. Heimberg Adult Anxiety Clinic of Temple University, Philadelphia, Pennsylvania, USA
Chapter 4.7.2

Scott Henderson Emeritus Professor, The Australian National University, Canberra, Australia
Chapters 2.7 and 8.3

George R. Heninger Professor, Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut, USA
Chapter 6.2.3

Leslie Hicks, University of York, UK
Chapter 9.5.6

A.J. Holland Lecturer, Department of Psychiatry, University of Cambridge, UK
Chapter 10.1

Jimmie C. Holland Wayne E. Chapman Chair in Psychiatric Oncology, Department of Psychiatry and Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, USA
Chapter 5.3.7

Sheila Hollins Professor of Psychiatry of Learning Disability, Department of Psychiatry and Disability, St George's Hospital Medical School, University of London, UK
Chapter 10.7

Jeremy Holmes Consultant Psychiatrist/Psychotherapist, North Devon District Hospital, Barnstaple; Senior Lecturer, University of Bristol, UK
Chapter 3.2

Suzanne Holroyd Professor, Director of Geriatric Psychiatry, Department of Psychiatry and Neurobehavioral Science, University of Virginia, Charlottesville VA
Chapter 8.5.6

Geraldine Holt Honorary Senior Lecturer in Psychiatry at the Institute of Psychiatry, King's College London, UK
Chapter 10.9

Allan House Professor of Liaison Psychiatry, University of Leeds, UK
Chapter 5.3.1

Daniel H. Hovelson The Harvard program in refugee trauma, Massachusetts general hospital, Dept of psychiatry.
Chapter 7.10.1

Jane Hubert Senior Lecturer in Social Anthropology, Department of Psychiatry and Disability, St George's Hospital Medical School, University of London, UK
Chapter 10.7

Stephen Hucker University of Toronto, Toronto, Canada
Chapter 11.8

T. A. Hughes Consultant Psychiatrist, St Mary's Hospital, Leeds, UK
Chapter 4.1.6

Masud Husain Institute of Neurology & Institute of Cognitive Neuroscience, UCL, London and National Hospital for Neurology & Neurosurgery, London
Chapter 2.5.2

Frits J. Huyse Psychiatrist, Consultant integrated care, Department of General Internal Medicine, University Medical Centre Groningen (UMCG), Groningen, The Netherlands
Chapter 5.7

Matti Iivanainen Professor, Department of Child Neurology, University of Helsinki, Finland
Chapter 10.5.3

Leslie Iversen Visting Professor, Department of Pharmacology, University of Oxford, UK
Chapter 6.2.7

Richard Ives National Children's Bureau, London, UK
Chapter 4.2.3.6

Assen Jablensky Professor of Psychiatry, University of Western Australia, Perth, Australia
Chapters 4.3.5 and 4.3.7

Robin Jacoby Clinical Reader in the Psychiatry of Old Age, University of Oxford, UK
Chapters 8.4 and 8.5.7

Claudia Jacova Assistant Professor, Division of Neurology, Department of Medicine, University of British Columbia, Vancouver, BC, Canada
Chapter 8.5.1.1

David V. James Consultant Forensic Psychiatrist, North London Forensic Service and Fixated Threat Assessment Centre
Chapter 11.6

Kay Redfield Jamison Professor of Psychiatry, Johns Hopkins School of Medicine, Baltimore, Maryland, USA
Chapter 1.1

John G.R. Jefferys Department of Neurophysiology, Division of Neuroscience, University of Birmingham, UK
Chapter 2.3.9

Andrew Johns Consultant Forensic Psychiatry and Honorary Senior Lecturer, Maudsley Hospital, London, UK
Chapter 11.3.2.

E.C. Johnstone Professor of Psychiatry and Head, Department of Psychiatry, University of Edinburgh, UK
Chapter 4.3.8

David P.H. Jones Senior Clinical Lecturer in Child Psychiatry, Park Hospital for Children, University of Oxford, UK
Chapter 9.3.3

Peter R. Joyce Professor, Department of Psychological Medicine, Christchurch School of Medicine, Christchurch, New Zealand
Chapter 4.5.4

Elizabeth Juven-Wetzler Department of Psychiatry, Chaim Sheba Medical Centre, Tel Hashomer, Israel
Chapter 4.8

Horst Kachele Universitätsklinik Psychosomatische Medizin and Psychotherapie Universitätsklinik Ulm, Germany
Chapter 6.3.5

Adam Ian Kaplin Assistant Professor, Departments of Psychiatry and Neurology, Johns Hopkins University School of Medicine, Johns Hopkins Hospital, Baltimore, MD
Chapter 1.1

Markus Kaski Director, Rinnekoti Research Foundation, Director and Chief Physician of Rinnekoti Foundation, Espoo, Finland
Chapter 10.3

Aliya Kassam Institute of Psychiatry, David Glodberg Centre, De Crespigny Park, London, UK
Chapter 1.2

Roger G. Kathol, Adjunct Professor of Internal Medicine and Psychiatry, University of Minnesota, President, Cartesian Solutions, Inc. Burnsville, MN, USA
Chapter 5.7

Paul E. Keck Jr. Lindner Center of HOPE, Mason, and Department of Psychiatry, University of Cincinnati College of Medicine, Cincinnati, OH, USA
Chapter 4.13.1

John Kellett St George's Hospital Medical School, London, UK
Chapter 8.5.8

Keith W. Kelley Department of Animal Sciences, University of Illinois, Urbana-Champaign, USA
Chapter 2.3.10

David Kennard Chair of the UK Network of the International Society for the Psychological Treatments of the Schizophrenias and other psychoses (ISPS UK); former Head of Psychology Services, The Retreat, York, UK
Chapter 6.3.9

Ad.J.F.M. Kerkhof Professor of Clinical Psychology, Vrije Universiteit, Amsterdam, The Netherlands
Chapter 4.15.2

Otto F. Kernberg Professor of Psychiatry, Cornell University Medical College, New York; Training and Supervising Analyst, Columbia University Center for Psychoanalytic Training and Research, New York, USA
Chapter 3.1

Falk Kiefer Professor of Addiction Research, Deputy Director, Department of Addictive Behaviour and Addiction Medicine, Central Institute of Mental Health CIMH, University of Heidelberg, Mannheim, Germany
Chapter 4.2.2.3

Arthur Kleinman Presley Professor of Anthropology and Psychiatry, Harvard University; Chair, Department of Social Medicine, Harvard Medical School, Cambridge, Massachusetts, USA
Chapter 2.6.2

Ami Klin Yale University, New Haven, Connecticut, USA
Chapter 9.2.3

Kimberly Klipstein Department of Psychiatry, Mount Sinai School of Medicine, New York, USA
Chapter 4.6.4

Martin Knapp Institute of Psychiatry, King's College London; London School of Economics and Political Science, University of London, UK
Chapter 7.7

Michael D. Kopelman Professor of Neuropsychiatry at King's College London, Institute of Psychiatry, UK
Chapter 4.1.12

Malcolm Lader Emeritus Professor of Clinical Psychopharmacology, King's College London, Institute of Psychiatry, Denmark Hill, London, UK
Chapter 6.2.2

Fergus D. Law Honorary Senior Registrar and Clinical Lecturer, Psychopharmacology Unit, University of Bristol, UK
Chapters 4.2.1 and 6.2.8

Brian Lawlor Conolly Norman Professor of Old Age Psychiatry, St. James's Hospital & Trinity College, Dublin, Ireland
Chapter 8.5.2

Dusica Lecic-Tosevski Professor of Psychiatry, Institute of Mental Health, School of Medicine, University of Belgrade, Belgrade, Serbia
Chapter 4.12.3

Julian Leff Emeritus Professor, Department of Psychological Medicine, Institute of Psychiatry, King's College London, UK
Chapter 1.3.2

R.J. Levin Department of Biomedical Science, University of Sheffield, UK
Chapter 4.11.1

Rhohel Lenroot Child Psychiatry Branch, NIMH, Bethesda MD, USA
Chapter 9.1.4

Peter F. Liddle Professor of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada
Chapter 4.3.2

Felice Lieh Mak Emeritus Professor, Department of Psychiatry, University of Hong Kong, Hong Kong
Chapter 5.2.10

James Lindesay Professor of Psychiatry for the Elderly, University of Leicester, UK
Chapters 8.5.1 and 8.5.5

Jouko K. Lonnqvist Professor, National Public Health Institute, Helsinki, Finland
Chapter 4.15.1

Juan J. López-Ibor Jr. Chairman, Department of Psychiatry, San Carlos University Hospital, Complutense University, Madrid, Spain
Chapters 1.6 and 4.12.1

Mª Inés López-Ibor Madrid, Spain
Chapter 4.2.3.8

Simon Lovestone Professor of Old Age Psychiatry, NIHR Biomedical Research Centre for Mental Health, MRC Centre for Neurodegeneration Research, Departments of Psychological Medicine and Neuroscience, King's College London, Institute of Psychiatry, London, UK
Chapter 4.1.2

Leanne Magee Temple University, Philadelphia, Pennsylvania, USA
Chapter 4.7.2

Andrew McCulloch The Mental Health Foundation, London, UK
Chapter 7.1

Jane McCarthy Division of Mental Health, St George's Hospital, London, UK
Chapter 10.8

Susan L. McElroy Lindner Center of HOPE, Mason, and Department of Psychiatry, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA
Chapter 4.13.1

Peter McGuffin Director and Professor of Psychiatric Genetics, Institute of Psychiatry, King's College London, UK
Chapter 2.4.1

I.G. McKeith Clinical Director, Institute for Ageing and Health, Newcastle University, Newcastle Upon Tyne, UK
Chapter 4.1.5

Kwame McKenzie Centre for Addictions and Mental Health, Toronto, Canada; University of Toronto, Canada; University of Central Lancashire, UK
Chapter 7.3

Declan McLoughlin Institute of Psychiatry, King's College London, UK
Chapter 6.2.10.3

Mark W. Mahowald Director, Minnesota Regional Sleep Disorders Center, Hennepin County Medical Center; Professor of Neurology, University of Minnesota Medical School, Minneapolis, Minnesota, USA
Chapter 4.14.4

Mario Maj Institute of Psychiatry, University of Naples, Italy
Chapter 4.1.9

Ulrik Fredrik Malt Professor of Psychiatry (Psychosomatic Medicine), National Hospital, University of Oslo, Norway
Chapter 5.3.8

J. John Mann Vice Chair for Research Scientific Director, Kreitchman PET Center, Columbia University and Chief, Division of Molecular Imaging & Neuropathology, New York State Psychiatric Institute, USA
Chapter 4.15.3

Karl F. Mann Professor and Chair in Addiction Research, Deputy Director Central Institute of Mental Health (CIMH), University of Heidelberg, Mannheim, Germany
Chapter 4.2.2.3

Russell L. Margolis Professor of Psychiatry and Neurology Director, Johns Hopkins Schizophrenia Program Director, Laboratory of Genetic Neurobiology Division of Neurobiology, Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, USA
Chapter 4.1.7

John C. Markowitz Associate Professor of Psychiatry, Weill Medical College of Cornell University; Director, Psychotherapy Clinic, Payne Whitney Clinic, New York Presbyterian Hospital, New York, USA
Chapter 6.3.3

John Marsden Lecturer, Institute of Psychiatry, King's College London, UK
Chapter 4.2.4

Jane Marshall Senior Lecturer in the Addictions, National Addiction Centre, Institute of Psychiatry, King's College London, UK
Chapters 4.1.11 and 4.2.2.2

Andrés Martin Professor of Child Psychiatry, Child Study Center Yale University School of Medicine, New Haven, Connecticut, USA
Chapter 9.1.3

Keith Matthews Dept of Psychiatry, University of Dundee, Dundee, UK
Chapter 6.2.10.4

Barbara Maughan MRC Child Psychiatry Unit, Institute of Psychiatry, King's College London, UK
Chapter 9.3.1

Soraya Mayet National Addiction Centre, Institute of Psychiatry, King's College London, UK
Chapter 4.2.3.1

Richard Mayou Emeritus Professor of Psychiatry, University of Oxford, UK
Chapter 5.2.1

Nick Medford Institute of Psychiatry, King's College London, UK
Chapter 4.9

David Meagher Dept of Adult Psychiatry, Midwestern Regional Hospital, Limerick, Ireland
Chapter 4.1.1

Thomas W. Meeks Division of Geriatric Psychiatry, University of California San Diego, La Jolla CA, USA
Chapter 8.5.3

Pamela S. Melding Honorary Senior Lecturer, Department of Psychological Medicine, University of Auckland, New Zealand and Consultant in Psychiatry of Old Age, Mental Health Serviced, North Shore Hospital, Waitemata District Health Board, Takapuna, North Shore City, Auckland, New Zealand
Chapter 8.7

Herbert Y. Meltzer Bixler/May/Johnaon Professor of Psychiatry, Professor of Pharmacoloqy Vanderbilt University School of Medicine, Nashville, Tennessee, USA
Chapter 6.2.5

Julien Mendlewicz Department of Psychiatry, University Clinics of Brussels, Erasme Hospital, Brussels, Belgium
Chapter 4.5.5

Andreas Meyer-Lindenberg Dept of Psychiatry, Central Institute of Mental Health, Mannheim, Germany
Chapter 2.5.3

Gillian C. Mezey Consultant and Senior Lecturer in Forensic Psychiatry, Traumatic Stress Service, St George's Hospital Medical School, London, UK
Chapters 11.12 and 11.13

R.H.S. Mindham Emeritus Professor of Psychiatry, University of Leeds, UK
Chapter 4.1.6

Andrew Mogg Institute of Psychiatry, King's College London, UK
Chapter 6.2.10.3

Richard F. Mollica Director, Harvard Program in Refugee Trauma; Associate Professor of Psychiatry, Harvard Medical School and Harvard School of Public Health, Cambridge, Massachusetts, USA
Chapter 7.10.1

Emanuel Moran Consultant Psychiatrist, Grovelands Priory Hospital, London, UK
Chapter 4.13.2

Stella Morris Dept of Psychological Medicine, Hull Royal Infirmary, Hull, UK
Chapter 5.2.6

Matt Muijen WHO Regional Office for Europe, Copenhagen, Denmark
Chapter 7.1

Paul E. Mullen Professor of Forensic Psychiatry, Monash University; Clinical Director, Victorian Institute of Forensic Mental Health, Monash University, Melbourne, Australia
Chapters 11.10, 11.11 and 11.14

Alistair Munro Emeritus Professor of Psychiatry, Dalhousie University, Halifax, Nova Scotia, Canada
Chapter 4.4

Rebecca Murphy Research Psychologist, Oxford University Department of Psychiatry, Warneford Hospital, Oxford, UK
Chapters 4.10.2 and 6.3.2.2

Lynne Murray Winnicott Research Unit, University of Reading, Reading, UK
Chapter 9.3.6

R.M. Murray Institute of Psychiatry, King's College London, UK
Chapter 4.3.6.1

Norbert Nedopil Professor of Forensic Psychiatry, Head of the Department of Forensic Psychiatry at the Psychiatric Hospital of the University of Munich, Munich, Germany
Chapter 11.3.3

Juan C. Negrete Professor and Head, Addictions Psychiatry Program, University of Toronto, Canada
Chapter 4.2.2.1

Gretchen N. Neigh Dept of Psychiatry and Behavioral Sciences, Emory University, Atlanta GA, USA
Chapter 2.3.3

Charles B. Nemeroff Reunette W. Harris Professor and Chairman, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia, USA
Chapter 2.3.3

Giles Newton-Howes Division of Neurosciences and Mental Health, Imperial College School of Medicine, London, UL
Chapter 4.12.7

Jeffrey Newcorm Mount Sinai School of Medicine, New York, USA
Chapter 4.6.4

Russell Noyes Jr. Department of Psychiatry, University of Iowa College of Medicine, Iowa City, Iowa, USA
Chapter 5.2.5

David J. Nutt Professor of Psychopharmacology and Head of Clinical Medicine, University of Bristol, UK
Chapters 4.2.1 and 6.2.8

Margaret Oates Senior Lecturer in Psychiatry, University of Nottingham, UK
Chapter 1.8.1

James R.P. Ogloff Victorian Institute of Forensic Mental Health, Thomas Embling Hospital, Fairfield VIC, Australia
Chapters 11.4 ,11.14, and 11.16

John O'Grady Knowle Hospital, Fareham, UK
Chapter 11.15

Catherine Oppenheimer Consultant Psychiatrist, Warneford Hospital, Oxford, UK
Chapter 8.5.8 and 8.6

Luca Ostacoli Liaison Psychiatry and Psychosomatic Unit, Department of Mental Health, San Luigi Gonzaga Hospital - University of Torino, Italy
Chapter 5.5

Pierre Oswald Dept of Psychiatry, ULB Erasme, Brussels, Belgium
Chapter 4.5.5

Barton W. Palmer Veterans Affairs Medical Center, University of California, San Diego CA, USA
Chapter 8.5.3

Gordon Parker Professor, University of New South Wales; and Executive Director, Black Dog Institute, Australia
Chapter 4.5.3

E.S. Paykel Emeritus Professor of Psychiatry, Department of Psychiatry, University of Cambridge, UK
Chapter 4.5.8

John B. Pearce Emeritus Professor of Child and Adolescent Psychiatry, University of Nottingham, UK
Chapters 9.4 and 9.5.1

†**R.C.A. Pearson** Department of Biomedical Science, University of Sheffield, UK
Chapter 2.3.1

Keith J. Petrie Associate Professor, School of Medicine,University of Auckland, New Zealand
Chapter 5.6

Cynthia R. Pfeffer Weill Medical College of Cornell University, New York Presbyterian Hospital-Westchester Division, White Plains, New York, USA
Chapter 9.2.10

Katharine A. Phillips Professor of Psychiatry and Human Behavior, The Warren Alpert Medical School of Brown University; Director, Body Dysmorphic Disorder Program, Butler Hospital, Providence, USA
Chapter 5.2.8

Pierre Pichot Académie Nationale de Médecine, Paris, France
Chapter 1.4

Harold Alan Pincus Columbia University, New York, USA
Chapter 1.9

Vanessa Pinfold 'Rethink', London, UK
Chapter 7.9

Daniel S. Pine Division of Intramural Research Programs, National Institutes of Health, Bethesda, USA
Chapter 9.2.6

Malcolm Pines Founding Member, Institute of Group Analysis, London, UK
Chapter 6.3.6

Harrison G. Pope Jr. Professor of Psychiatry, Harvard Medical School, Boston; Chief, Biological Psychiatry Laboratory, McClean Hospital, Belmont, Massachusetts, USA
Chapter 6.2.6

Robert M. Post Chief, Biological Psychiatry Branch, National Institute of Mental Health, Bethesda, Maryland, USA
Chapter 6.2.4

Graham E. Powell Psychology Services, Powell Campbell Edelmann, London, UK
Chapter 1.8.3

Herschel Prins Professor, Midlands Centre for Criminology and Criminal Justice, University of Loughborough, UK
Chapter 11.9

Paul Ramchandani Dept of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK
Chapter 9.3.6

Shulamit Ramon Professor of Interprofessional Health and Social Studies, Anglia Polytechnic University, Cambridge, UK
Chapter 6.4.3

Beverley Raphael University of Western Sydney Medical School, Sydney NSW, Australia
Chapter 4.6.5

James Reich Clinical Professor of Psychiatry, University of California, San Francisco Medical School and Adjunct Associate Professor of Psychiatry, Stanford School of Medicine
Chapter 4.12.2

Helmut Remschmidt Director, Department of Child and Adolescent Psychiatry, Philipps Universität, Marburg, Germany
Chapter 9.2.2

Philippe Robaey Institute of Mental Health Research, Royal Ottawa Hospital, Ottawa, Canada
Chapter 9.2.8

Ian Robbins Consultant Clinical Psychologist, St George's Hospital, London, UK
Chapter 11.13

T.W. Robbins Section of Forensic Psychiatry, St George's Hospital Medical School, London, UK
Chapter 2.5.5

G.M. Rodin Professor of Psychiatry, University of Toronto, Toronto, Canada
Chapter 5.3.6

Maria A. Ron Professor of Neuropsychiatry, Institute of Neurology, University College London, UK
Chapter 5.3.2

Robin Room Professor, School of Population Health, University of Melbourne; and Director, AER Centre for Alcohol Policy Research, Turning Point Alcohol and Drug Centre, Fitzroy, Victoria, Australia
Chapter 4.2.2.6

W. Rössler Professor of Clinical Psychiatry and Psychology, University of Zürich, Switzerland
Chapter 6.4.1

James R. Rundell Department of Psychiatry and Psychology, Mayo Clinic Professor of Psychiatry, Mayo Clinic College of Medicine
Chapter 5.3.4

Gerald Russell Emeritus Professor of Psychiatry, Director of the Eating Disorders Unit, Hayes Grove Priory Hospital, Hayes, Kent, UK
Chapter 4.10.1

B.J. Sahakian Dept of Psychiatry, University of Cambridge, Cambridge, UK
Chapter 2.5.5

Diana Sanders Chartered Counselling Psychologist, working in Psychological Medicine in Oxford, UK
Chapter 6.3.1

Paramala J. Santosh Great Ormond Street Hospital for Sick Children, London, UK
Chapter 9.5.5

Benedetto Saraceno Director of Department of Mental Health and Substance Abuse, World Health Organization WHO
Chapter 1.3.1

Gauri N. Savla, Veterans Affairs Medical Center, University of California, San Diego CA, USA
Chapter 8.5.3

Carlos H. Schenck Staff Psychiatrist, Minnesota Regional Sleep Disorders Center, Hennepin County Medical Center; Associate Professor of Psychiatry, University of Minnesota Medical School, Minneapolis, Minnesota, USA
Chapter 4.14.4

John Schlapobersky Consultant Psychotherapist, Trumatic Stress Clinic Middlesex/University College Hospital, formerly also of The Medical Foundation for the Care of Victims of Torture London, UK
Chapter 6.3.6

Fabrizio Schiffano, Chair in Clinical Pharmacology and Therapeutics Associate Dean, Postgraduate Medical School, Hon Consultant Psychiatrist Addictions, University of Hertfordshire, School of Pharmacy, College Lane Campus, Hatfield, UK
Chapter 4.2.3.5

Gerd Schulte-Körne Director of the Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University of Munich, Pettenkoferstr, München/Germany
Chapter 9.2.2

J. Scott Professor of Psychological Medicine, University of Newcastle & Honorary Professor, Psychological Treatments Research, Institute of Psychiatry, London and University Department of Psychiatry, Leazes Wing, Royal Victoria Infirmary, Newcastle upon Tyne, England
Chapter 4.5.8

Stephen Scott Professor of Child Health & Behaviour, King's College London, Institute of Psychiatry, and Director of Research National Academy for Parenting Practitioners, London, UK
Chapters 9.1.1 and 9.2.5

Nicholas Seivewright Consultant Psychiatrist in Substance Misuse, Community Health Sheffield NHS Trust, Sheffield, UK
Chapter 4.2.3.2

David Shaffer Department of Child Psychiatry, College of Physicians and Surgeons, Columbia University, New York, USA
Chapter 9.2.10

Trevor Sharp Dept of Pharmacology, University of Oxford, Oxford, UK
Chapter 2.3.4

Michael Sharpe Professor of Psychological Medicine & Symptoms Research, University of Edinburgh, UK
Chapters 5.1 and 5.2.7

Jennifer Shaw Centre for Suicide Prevention, The School of Medicine, University of Manchester, UK
Chapter 11.5

Mauricio Sierra Institute of Psychiatry, King's College London, UK
Chapter 4.9

Gregory Simon Investigator, Center for Health Studies, Group Health Cooperative, Seattle, Washington, USA
Chapter 5.2.2

Andrew Sims Professor of Psychiatry, University of Leeds, UK
Chapter 1.7

Ian Sinclair Professor of Social Work, University of York, UK
Chapter 9.5.6

Mike Slade Health Service and Population Research Department and Institute of Psychiatry, King's College London, UK
Chapter 7.2

Elita Smiley Consultant Psychiatrist and Clinical Senior Lecturer, Division of Community Based Sciences, Faculty of Medicine, University of Glasgow, UK
Chapter 10.2

Wolfgang Söllner Department of Psychosomatic Medicine and Psychotherapy General Hospital Nuremberg, Prof.Ernst-Nathan-Str. 1, Nürnberg, Germany
Chapter 5.7

Daniel Souery Department of Psychiatry, University Clinics of Brussels, Erasme Hospital, Brussels, Belgium
Chapter 4.5.5

Elizabeth Spencer Senior Clinical Medical Officer, Early Intervention Service, Northern Birmingham Mental Health Trust, Birmingham, UK
Chapter 6.3.2.4

David A. Spiegel Center for Anxiety and Related Disorders at Boston University, Boston, Massachusetts, USA
Chapter 4.7.1

Costas Stefanis Honorary Professor of Psychiatry, University of Athens, Greece
Chapter 1.6

Alan Stein Royal Free and University College Medical School, University College London, and Tavistock Clinic, London, UK
Chapter 9.3.6

Jessica Stiles Department of Psychiatry and Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, USA
Chapter 5.3.7

William S. Stone Assistant Professor of Psychology, Director of Neuropsychology Training and Clinical Services, Department of Psychiatry, Harvard Medical School, Massachusetts Mental Health Center Public Psychiatry, Division of the Beth Israel Deaconess Medical Center, Boston, USA
Chapter 4.3.9

Gregory Stores Emeritus Professor of Developmental Neuropsychiatry, University of Oxford, UK
Chapters 4.14.1 and 9.2.9

Elizabeth A. Stormshak Assistant Professor, University of Oregon, Eugene, Oregon, USA
Chapter 9.5.4

James J. Strain Professor/Director, Behavioral Medicine and Consultation Psychiatry, Mount Sinai School of Medicine, New York, USA
Chapter 4.6.4

John Strang National Addiction Centre, Institute of Psychiatry, King's College London, UK
Chapters 4.2.3.1 and 4.2.4

J. Suckling Brain Mapping Unit, Department of Psychiatry, University of Cambridge, Addenbrookes Hospital, Cambridge, UK
Chapters 2.3.7 and 2.3.8

Nicola Swinson Centre for Sucide Prevention, The School of Medicine, University of Manchester, UK
Chapter 11.5

Michele Tansella Professor of Psychiatry and Chairman, Department of Medicine and Public Health, Section of Psychiatry, University of Verona, Italy
Chapters 7.2 and 7.6

Mary Target Senior Lecturer in Psychoanalysis, Psychoanalysis Unit, University College London; Deputy Director of Research, Anna Freud Centre, London, UK
Chapter 9.5.2

Eric Taylor Head of Department, Child & Adolescent Psychiatry, King's College London, Institute of Psychiatry
Chapter 9.2.4

John-Paul Taylor Academic Specialist Registrar, Institute for Ageing and Health Newcastle University, Campus for Ageing and Vitality, Newcastle upon Tyne, UK
Chapter 4.1.13

Pamela J. Taylor School of Medicine, Cardiff University, Cardiff, UK
Chapter 11.17

Tatiana Taylor Dept of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK
*Chapter 4.15.*4

Mary Teasdale 'Rethink', London, UK
Chapter 7.9

Anita Thapar Department of Psychological Medicine, School of Medicine, Cardiff University, UK
Chapter 2.4.1

June Thoburn Emeritus Professor of Social Work, University of East Anglia, Norwich, UK
Chapter 9.3.5

Sian Thomas Chester Young People's Centre, Shester, UK
Chapter 9.5.9

Lindsay Thomson Division of Psychiatry, University of Edinburgh, Edinburgh, UK
Chapter 11.3.1

Anne E. Thompson Emeritus Professor Child and Adolescent Psychiatry, University of Nottingham, UK
Chapter 9.4

Graham Thornicroft Professor of Community Psychiatry, Institute of Psychiatry, King's College London, UK
Chapters 1.2, 7.2 and 7.6

Adolf Tobeña Professor of Psychiatry, Director of the Dept. of Psychiatry and Forensic Medicine, Autonomous University of Barcelona, Bellaterra (Barcelona), Spain
Chapter 4.12.5

Bruce J. Tonge Head Monash University School of Psychology Psychiatry & Psychological Medicine, Monash Medical Centre, Clayton, Victoria, Australia
Chapter 10.5.1

Brian Toone Consultant, Maudsley Hospital; Honorary Senior Lecturer, Institute of Psychiatry, King's College London, UK
Chapter 5.3.3

David Trickey Leicester Royal Infirmary, Leicestershire Partnership NHS Trust, UK
Chapters 9.3.2 and 9.3.7

Paula Trzepacz Eli Lilly & Co, USA
Chapter 4.1.1

Wen-Shing Tseng Professor at Department of Psychiatry, University of Hawaii School of Medicine, USA
Chapters 4.16 and 6.5

Ming T. Tsuang Behavioral Genomics Endowed Chair and University Professor, University of California; Distinguished Professor of Psychiatry and Director, Center for Behavioral Genomics, Department of Psychiatry, University of California, San Diego, CA, USA
Chapter 4.3.9

André Tylee Director, Royal College of General Practitioners Unit for Mental Health Education in Primary Care, Institute of Psychiatry, King's College London, UK
Chapter 7.8

Amy M. Ursano Department of Psychiatry, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, North Carolina, USA
Chapter 6.3.4

Robert J. Ursano Professor and Chairman, Department of Psychiatry, Uniformed Services University of the Health Sciences, F. Edward Herbert School of Medicine, Bethesda, Maryland, USA
Chapter 6.3.4

Jim van Os Professor of Psychiatric Epidemiology, Maastricht University, Maastricht, The Netherlands and Visiting Professor of Psychiatric Epidemiology Institute of Psychiatry, London, UK
Chapter 7.3

Fred R. Volkmar Yale University, New Haven, Connecticut, USA
Chapter 9.2.3

Jane Walker Clinical Lecturer and Honorary Specialist Registrar in Liaison Psychiatry, Psychological Medicine & Symptoms Research Group, School of Molecular & Clinical Medicine, University of Edinburgh, UK
Chapter 5.1

Diane Waller Professor of Art Psychotherapy, Goldsmiths, University of London
Chapter 6.4.4

Paul Walters MRC Fellow & Specialist Psychiatrist , Programme Leader MSc in Mental Health Services Research, Section of Primary Care Mental Health, Health Service and Population Research Department, David Goldberg Centre, Institute of Psychiatry, London, UK
Chapter 7.8

John Weinmann Professor of Psychology as applied to Medicine, Institute of Psychiatry, King's College London, UK
Chapter 5.6

Myrna M. Weissman Professor of Epidemiology in Psychiatry, College of Physicians and Surgeons of Columbia University; Chief, Division of Clinical and Genetic Epidemiology, New York State Psychiatric Institute, New York, USA
Chapter 6.3.3

Sarah Welch Gloucestershire Partnership NHS Foundation Trust, UK
Chapter 4.2.3.4

Ursula Werneke Consultant Psychiatrist, Norrkoping, Sweden
Chapter 6.2.9

Simon Wessely Professor of Epidemiological and Liaison Psychiatry, Institute of Psychiatry, King's College London, UK
Chapter 5.2.7

Kay Wheat Senior Lecturer in Law, Department of Academic Legal Studies, Nottingham Law School, Nottingham Trent University, UK
Chapter 11.1

Adam R. Winstock Senior Staff Specialist, Drug Health Services, Conjoint Senior Lecturer, National Drug and Alcohol Research Centre, UNSW
Chapters 4.2.3.1 and 4.2.3.5

Sally Wooding Senior Research Fellow, SciMHA Unit, University of Western Sydney, Australia
Chapter 4.6.5

Matt Woolgar Institute of Psychiatry, King's College London, UK
Chapter 2.5.1

Miranda Wolpert Director of Child and Adolescent Mental Health Services, Evidence Based Practice Unit, University College London and Anna Freud Centre, UK
Chapter 9.5.7

Lawson Wulsin Professor of Psychiatry and Family Medicine, University of Cincinnati, OH, USA
Chapter 5.7

Richard Jed Wyatt† National Institutes of Mental Health, Bethesda, Maryland, USA
Chapter 1.1

William Yule Professor of Applied Child Psychology, Institute of Psychiatry, King's College London, UK
Chapter 2.5.1

Karl Zilles Professor, Institute of Neuroscience and Biophysics, INB-3 Research Centre, Jülich and C.&O. Vogt Institute of Brain Research, University Düsseldorf, Germany
Chapter 2.3.2

Joseph Zohar Psychiatric Medical Center, Sheba Medical Center, Tel Hashomer and Sackler School of Medicine, Tel Aviv University, Israel
Chapter 4.8

SECTION 5

Psychiatry and Medicine

This page intentionally left blank

5.1

Mind–body dualism, psychiatry, and medicine

Michael Sharpe and Jane Walker

Introduction

Patients usually attend doctors because they are concerned about symptoms. When these symptoms are associated with persistent distress or disability we refer to the patient as having an illness. When assessing the patient's illness the doctor aims to make a diagnosis, on the basis of which management can be planned and prognosis made. The diagnoses available to doctors are conventionally defined as either 'medical' or 'psychiatric'. This division of illness into two types is such an accepted feature of current medical practice that we tend to take it for granted. But is it really the best way to think about patients' illnesses and to plan their care?

In order to answer this question we will examine what is meant by 'medical' and 'psychiatric' diagnoses and the assumptions underpinning this division. The disadvantages of this dualistic approach will be considered and solutions proposed.

Diagnosis

Medical diagnosis

A medical diagnosis is a label for a condition that is: (a) conventionally treated by medical doctors and (b) listed in the classifications of medical conditions such as ICD-10. Most medical diagnoses are based on identifiable bodily pathology (abnormal structure and/or function). Therefore, to make a medical diagnosis (such as cancer) doctors will seek specific bodily symptoms before confirming the presence of bodily pathology with physical signs and biological investigations (such as X-rays).

Psychiatric diagnosis

Similarly a psychiatric diagnosis is a label for a condition that is: (a) conventionally treated by psychiatrists and (b) defined in the psychiatric diagnostic classifications of ICD and DSM. Psychiatric diagnoses are not based not on bodily pathology. They are however associated with the idea of 'psychopathology', that is proposed abnormalities of the mind. Unlike bodily pathology these abnormalities of the mind cannot be objectively identified and have to be inferred from the patient's mental symptoms and their behaviour. Investigations play little or no role in diagnosis. Psychiatric diagnoses are therefore defined on the basis of symptoms and syndromes.

When is an illness psychiatric?

Why are some illnesses regarded as 'mental' or 'psychiatric' as opposed to 'medical'? Examination of the criteria for diagnoses listed as psychiatric reveals that readily observable factors common to most 'psychiatric' illnesses are:

- an absence of known bodily pathology
- an abnormal mental state as inferred by the patient's report
- a presentation with disturbed behaviour

Mind–body dualism

The underlying assumption of this dichotomous view is that it is both valid and useful to divide human illnesses into those of the body and those of the mind.[1] This idea of mind–body dualism is commonly attributed to the writings of the philosopher Descartes. So-called Cartesian dualism has exerted a profound influence on Western medical thinking and still shapes our thinking, training, and service provision.

However, dualism is at best an oversimplification and at worst a source of serious theoretical and practical problems. It may be argued that there is no such thing as a purely 'bodily' or purely 'mental' illness and that all illnesses have mental and bodily aspects.[2] Furthermore, the assumption that bodily symptoms indicate bodily pathology and that mental symptoms indicate psychopathology gives rise to specific problems: (a) when bodily symptoms occur without bodily pathology and (b) when mental symptoms occur together with bodily pathology (see Table 5.1.1).

Bodily symptoms with no bodily pathology: somatization

When patients present with bodily symptoms and bodily pathology is confirmed they are given a medical diagnosis. When patients have bodily symptoms but there is no evidence of bodily pathology the terms 'somatization' or 'somatoform disorder' are used to describe their illness. It is unclear, however, whether these illnesses are properly regarded as 'psychiatric' or as 'medical' as they do not clearly fulfil criteria for either. One solution to this dilemma is to allocate these illnesses to psychiatry. The assumption is made that their somatic symptoms are really explained by psychopathology. The absence of mental symptoms, from which psychopathology

Table 5.1.1 Diagnoses symptoms and bodily pathology

Symptoms	Bodily pathology	Diagnosis
Bodily symptoms	Present Absent	Medical diagnosis Somatization
Mental symptoms	Present Absent	Comorbidity Psychiatric diagnosis

can be inferred, is explained by the idea that the psychopathology is hidden and 'converted' into bodily symptoms by a process called 'somatization' (literally making the mental somatic). Clearly these are questionable assumptions.[3]

A second solution is to assume that the patients really do have bodily pathology in some form (even though is it unknown) and to give them a medical diagnosis of a so-called 'functional disorder' such as fibromyalgia.[4] As with somatization this approach is based on questionable assumptions.

A third, and all too common solution, is for the patient to be rejected as 'not really ill' by both psychiatry and medicine. They then end up in a no-man's land between specialities. The inadequacy of all three solutions has been particularly well illustrated by the controversy and conflict surrounding the condition called Chronic Fatigue Syndrome (CFS) or Myalgic Encephalomyelitis (ME).[5]

Mental symptoms and bodily pathology: comorbidity

When a patient has both bodily pathology and mental symptoms they are given both a medical diagnosis (based on the bodily pathology) and a psychiatric diagnosis (based on presumed psychopathology). This idea of 'comorbidity' gives rise to both theoretical and practical problems, however.

The theoretical problem concerns the psychiatric diagnosis. To make this diagnosis the doctor must identify symptoms, which are considered to be evidence of psychopathology. However, some symptoms may be considered as evidence of both psychopathology and bodily pathology. For example, a patient has mental symptoms of low mood and worthlessness, and a medical diagnosis of cancer, based on bodily pathology. Should the patient's weight loss be counted toward a psychiatric diagnosis of depression or regarded as a symptom of his cancer? There is no generally agreed answer to this conundrum (although a variety of ways of addressing it have been proposed[6]), probably because it is a manifestation of the fundamentally flawed dualistic assumption.

The main practical problem that results from making two diagnoses is a failure to adequately treat the patient. Depressive disorder comorbid with a chronic medical condition is a major cause of morbidity. However, the patient's need for psychiatric treatment often goes unmet[7] because the patient is considered to have two illnesses, each requiring diagnosis and treatment by a different speciality and the treatment of the medical condition takes precedence.

Solutions to dualism

Theoretical solutions

New scientific knowledge, such as the demonstration of a bodily (neural) basis to many 'mental' symptoms is increasingly rendering crude dualistic thinking theoretically untenable.[8] Mind and brain are coming to be regarded as two sides of the same coin—the mind/brain. This paradigm shift implies that 'psychiatric' illnesses are no more distinct from 'medical conditions' than the nervous system is separate from the rest of the body. Hence, there is a need for psychiatry to become less 'brain-less' and for medicine to become less 'mind-less'.[9] According to this new way of thinking, all symptoms, whether previously regarded as 'bodily' or 'mental' are in fact products of the mind/brain's integration of bodily, psychological, and social information. Therefore to speak of 'medical' and 'psychiatric' symptoms makes no sense. Symptoms are just symptoms.

If this paradigm shift is to be fully translated into clinical practice a new unified classification system is needed that would be used by both medicine and psychiatry. One way of achieving this might be to create a multi-axial system as is currently used by DSM-IV. However, rather than using separate axes for psychiatric and medical diagnoses, separate axes would be used for symptoms (not distinguishing between medical and psychiatric) and bodily pathology.[10] Other axes could be added to ensure that other important information is included. An example is shown in Table 5.1.2.

Practical solutions

For the present we must accept that dualism continues to shape our every day thinking, practice, and service organization. It is important therefore, that the psychiatrist is aware of the practical problems that result and is equipped with ways of addressing them. In this regard the psychiatrist is especially well placed to make a major contribution to the care of all patients by ensuring that biological, psychological, and social aspects of illness are considered in every case. This so-called 'biopsychosocial' formulation was first proposed by Engel.[11] A further enhancement of this formulation is to divide the aetiological factors into those that predisposed the patient to the illness, those that precipitated or triggered it, and those that are perpetuating it. The last group of causes is a target for treatment and the first two for prevention. A useful diagram that lists factors to consider in a biopsychosocial formulation is shown in Table 5.1.3.

Service solutions

Finally, the consequence of the professional and organizational separation of medicine and psychiatry has been a major obstacle to the integrated care of patients, especially those with comorbidity and somatoform disorders. One service solution has been the establishment of so-called liaison (linking) psychiatry services to general hospital inpatient units. Another is the increasing integration of psychological management into chronic illness management programmes.[12] However truly integrated care remains the exception rather than the rule.

Table 5.1.2 A proposed multi-axial diagnostic system for use by both psychiatry and medicine

Axis 1	Symptoms or syndrome, e.g. chronic fatigue or depression
Axis 2	Bodily pathology, e.g. cancer
Axis 3	Biological factors, e.g. autonomic arousal
Axis 4	Psychological factors, e.g. beliefs
Axis 5	Social and situational factors, e.g. bereavement

Table 5.1.3 A biopsychosocial formulation

Main factors	Subfactors	Predisposing	Precipitating	Perpetuating
Biological	Disease physiology			
Psychological	Cognition mood behaviour			
Social	Interpersonal social and occupational health care system			

Conclusion

It has been taken for granted that it is appropriate and desirable to separate patients' illnesses into medical and psychiatric types. Such an approach has had advantages in allowing specialization of training and service planning but has also created obstacles to effective patient care. It is important that practising psychiatrists are aware of these obstacles and ways of overcoming them. It also seems increasingly likely that in time better understanding of neuroscience will make dualism increasingly theoretically untenable and that a better understanding of chronic illness management will make it practically redundant. Only then will psychiatry become fully reintegrated with the rest of medicine.

Further information

White, P.D. (2005). *Biopsychosocial medicine*. Oxford University Press, Oxford.

Damasio, A.R. (1994) *Descartes' error*. GP Putnam's Sons, New York.

References

1. Miresco, M.J. and Kirmayer, L.J. (2006). The persistence of mind-brain dualism in psychiatric reasoning about clinical scenarios. *The American Journal of Psychiatry*, **163**, 913–8.
2. Wade, D.T. and Halligan, P.W. (2004). Do biomedical models of illness make for good healthcare systems? *British Medical Journal*, **329**, 1398–401.
3. DeGucht, V. and Fischler, B. (2002). Somatization: a critical review of conceptual and methodological issues. *Psychosomatics*, **43**, 1–9.
4. Wessely, S., Nimnuan, C., and Sharpe M. (1999). Functional somatic syndromes: one or many? *Lancet*, **354**, 936–9.
5. Sharpe, M. (2002). The English Chief Medical Officer's Working Parties' report on the management of CFS/ME: significant breakthrough or unsatisfactory compromise? *Journal of Psychosomatic Research*, **52**, 437–8.
6. Cohen-Cole, S.A., Brown, F.W., and McDaniel, J.S. (1993). Diagnostic assessment of depression in the medically ill. In *Psychiatric care of the medical patient* (eds. A. Stoudemire and B. Fogel), pp. 53–70. Oxford University Press, New York.
7. Moussavi, S., Chatterji, S., Verdes, E., *et al*. (2007). Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet*, **370**, 851–8.
8. Kendler, K.S. (2001). A psychiatric dialogue on the mind-body problem. *The American Journal of Psychiatry*, **158**, 989–1000.
9. Eisenberg, L. (1986). Mindless and brainless in psychiatry. *British Journal of Psychiatry*, **148**, 497–508.
10. Sharpe, M., Mayou, R., and Walker, J. (2006). Bodily symptoms: new approaches to classification. *Journal of Psychosomatic Research*, **60**, 353–6.
11. Engel, G.L. (1977). The need for a new medical model: a challenge for biomedicine. *Science*, **196**, 129–96.
12. Von Korff, M., Glasgow, R.E., and Sharpe, M. (2002). Organising care for chronic illness. *British Medical Journal*, **325**, 92–4.

5.2

Somatoform disorders and other causes of medically unexplained symptoms

Contents

5.2.1 Somatoform disorders and functional symptoms

Richard Mayou

Non-specific symptoms that are not explained by organic pathology are extremely frequent in the general population[1] and in all medical settings. Most are transient, but a substantial minority is persistent, disabling, and often associated with frequent consultation. They are likely, especially when there are multiple unexplained symptoms, to be associated with psychiatric disorder (see Chapter 5.2.3). They are widely regarded as difficult to treat but only a very small proportion is seen by psychiatrists and psychologists.

This chapter covers general issues relating to functional symptoms and syndromes and their psychiatric associations. The following chapters provide more detail about the more specific forms of somatoform disorder and about functional syndromes (pain, chronic fatigue).

Terminology of functional symptoms

The terminology is unsatisfactory.[2] These symptoms are often referred to as 'medically unexplained symptoms'. This usage has the advantage of describing the clinical problem without assumptions of aetiology, but it is unsatisfactory in that it wrongly implies that there is no medical explanation. Other generally used terms include somatization, somatoform symptoms, and functional overlay. It is perhaps most satisfactory to refer to functional symptoms and functional syndromes.

This chapter is concerned with functional symptoms whether or not they are associated with psychiatric disorder.

Aetiology

A traditional Western dualist view of aetiology as being either physical or psychological, continues to influence clinical practice and current psychiatric classifications (see Chapter 5.1). In western countries, this view has resulted in great problems in psychiatric and lay understanding, in taxonomy, and in the treatment of 'unexplained' symptoms. It has also caused bewilderment in cultures that do not share this dualist approach.

An increasingly widely held alternative view, for which there is compelling evidence, is that functional symptoms result from the interaction of physiological, pathological, and psychosocial variables.[2] A primary bodily sensation or concern (Table 5.2.1.1) is then attributed or interpreted as being of sinister significance with resulting subjective symptoms, disability, and behavioural and emotional consequences. For example, awareness of normal heart rate increase due to excitement or anxiety can result in, on the one hand, panic and, on the other, worry about heart disease, restriction of daily activities, and repeated consultation to seek investigation

Table 5.2.1.1 Causes of bodily sensations

Major pathology
Minor pathology
Physiological processes, for example:
Sinus tachycardia and benign minor arrhythmias
Effects of fatigue
Hangover
Effects of overeating
Effects of prolonged inactivity
Autonomic effects of anxiety
Lack of sleep

and reassurance. The role of these factors may vary over time during the course of any individual clinical problem.

There is considerable evidence on the ways in which psychological processes affect the interpretation of physical symptoms, whatever the underlying (major or minor) pathology or physiological processes. Cognitive-behavioural formulations emphasize the central significance of health anxiety and suggest that feedback of the physiological, cognitive, affective, and behavioural consequences of this anxiety can reinforce the physical symptoms as well as their effects on everyday life.

The process of interpretation of a bodily sensation or fear is affected by several sets of factors:

- the individual's medical experience and beliefs
- social circumstances (Table 5.2.1.2)
- personality and mental state

Once symptoms have developed they may be maintained by behavioural and psychological factors and also by the reactions of others. As with other forms of anxiety, neurobiological mechanisms may perpetuate and complicate the initial presentation.

Simple reassurance is often ineffective especially in those who, by reason of personality, are inclined to worry about their health. Misconceptions are frequently reinforced and maintained by the lack of any medical explanation for worrying symptoms or by ambiguous or contradictory advice.

The association with psychiatric disorder

The majority of functional symptoms in general populations are short lived and not associated with psychiatric disorder. There is now considerable evidence both from smaller local studies and international collaborative research that the more severe and disabling functional symptoms are associated with anxiety and depressive disorder, and that this relationship is strongest for those who have the greatest number of 'unexplained' symptoms. This is so for all ethnic groups and cultures studied.[1] There are also associations with the somatoform disorders as described below.

Table 5.2.1.2 Illness experience, which may affect the interpretation of bodily sensations and concern

Childhood illness
Family illness and consultation in childhood
Childhood consultation and school absence
Physical illness in adult life
Experience and satisfaction with medical consultation
Illness in family and friends
Publicity in television, newspapers, etc.
Knowledge of illness and its treatment

Classification of unexplained symptoms

The classification of persistent and disabling functional symptoms has taken two parallel approaches.

(a) Medical descriptive syndromes

These are very numerous, clinical patterns and terms overlap and some include assumptions about aetiology. There are cultural differences in the definition and naming. There is little evidence for the validity of separate syndromes. Lay pressure groups have increasingly claimed specific syndromes, such as alleged sensitivity to dental amalgam and many 'food allergies', which are more likely to be due to their own predicaments and the apparent lack of success of conventional medicine.[3] A small number of syndromes have now received operational diagnostic criteria which have proved valuable in clinical understanding and in planning treatment, for example the criteria for chronic fatigue (Chapter 5.2.7).

(b) Psychiatric classification

This covers both well established categories, such as anxiety and depressive disorders, and the new concept, first introduced in DSM-III, of somatoform disorder.

(i) Somatoform disorder

Somatoform disorders (Table 5.2.1.3) were seen as speculative and provisional in DSM-III. The defining feature was '*physical symptoms suggesting a physical disorder for which there are no demonstrable organic findings on known physiological mechanisms, and for which there is strong evidence, or a strong presumption, that the symptoms are linked to psychological factors or conflicts*'.

The original DSM-III classification was relatively narrow but subsequent revisions of DSM and ICD-10 have incorporated non-specific categories which have turned out to be much more prevalent in all settings.

Table 5.2.1.3 Categories of somatoform disorders in ICD-10 and DSM-IV

ICD-10	DSM-IV
Somatization disorder	Somatization disorder
Undifferentiated somatoform disorder	Undifferentiated somatoform disorder
Hypochondriacal disorder	Hypochondriasis
Somatoform autonomic dysfunction	—
Persistent pain disorder	Pain disorder associated with psychological factors (and a general medical condition)
Other somatoform disorders	Somatoform disorders not otherwise specified
—	Body dysmorphic disorder
—	Conversion disorders
Neurasthenia	—

It is important to recall that somatoform disorder remains a provisional grouping for statistical purposes rather than a grouping of categories that satisfy the normal requirements of disease entities. It nevertheless indicates a substantial clinical problem associated with considerable use of health care provisions.[3]

(ii) Factitious disorder

DSM-III also introduced another new category of 'factitious disorder' for self-inflicted physical problems. These are described in Chapter 5.2.9 and should be distinguished from deliberate falsification for external gain—*malingering*. It must be remembered that patients with factitious disorder may also suffer from unexplained symptoms attributable to somatoform or other psychiatric disorders and, indeed, not uncommonly also report symptoms of undoubted physical illness.

Somatoform disorders in DSM and ICD

There are substantial differences between the use of subcategories in DSM and ICD.[4] Neurasthenia is included in ICD-10 but is not used in any section of DSM-IV; conversion disorder is a somatoform disorder in DSM-IV but not in ICD. Both classifications include both relatively specific categories (e.g. *somatization disorder* and *hypochondriasis*) and also several very vaguely defined non-specific categories. These latter include *Undifferentiated Somatoform Disorder*, *Somatoform Autonomic Dysfunction* (ICD-10 only), and *Other Somatoform Disorders*. Although these latter have attracted less clinical and research attention, they are by far the most common forms of somatoform disorder in all epidemiological studies. So broad are the criteria that it is possible to use these categories for almost all persistent unexplained physical symptoms. Epidemiological comparisons of ICD and DSM show that the use of their rather different criteria results in substantially different prevalences of somatoform disorder in community and primary care populations.

Problems in the definition of somatoform disorder

It is widely recognized that there are serious problems[2] in the overall concept of somatoform disorders and in the definition of subcategories:

- There is no unifying theoretical basis for the whole category; it is a disparate group of problems that are not easily fitted into other parts of the classifcations.
- Comorbidity is very common, especially with anxiety disorder, depressive disorder, and personality disorder.
- Some types of somatoform disorder could be more satisfactorily reassigned to other parts of the classification (for example hypochondriasis might be renamed health anxiety and moved to anxiety disorder).
- The definitions of the less specific categories (pain disorder, Undifferentiated Somatoform Disorder) do not include any psychological criteria. Instead they rely on the description and the number of physical symptoms, the same symptoms that are used to make accompanying Axis III diagnoses. Somatization disorder has attracted disproportionate attention, but appears to be no more than an uncommon, arbitrary and unreliable extreme of the spectrum of multiple physical symptoms.
- Criteria have little meaning for cultures that do not share the western presumption of the separation of body and mind.

It has become apparent that the present classifications have no value in guiding treatment and that they are both confusing and unpopular with patients and with those who treat them.

DSM-V and ICD-11 can be expected to make large changes which will depend on the resolution of conceptual arguments and substantial further research. It is hoped that the new classification and terminology will be more reliable and valid and also be much more meaningful and acceptable.[2]

It is likely that the more specific subcategories (hypochondriasis, body dysmorphic disorder) will be reassigned to other parts of the classification and that there will be modifications in their criteria. The greatest problems relate to the much more prevalent non-specific categories. There is a consensus that a more rational operational approach is required to categorize multiple symptoms which should result in either a renamed grouping on Axis I or more logically a transfer to Axis III. It should be possible to give much greater prominence both in criteria and in accompanying text to underlying psychological and behavioural abnormalities.

Classification in clinical practice

The following chapters in this section of the book describe syndromes that have proved to have some administrative value despite the acknowledged lack of validity. Anxiety and depression are considered fully elsewhere. The recognition of anxiety and depression is important because of the therapeutic implications.

In everyday clinical practice it is rarely necessary (or helpful) to attach a somatoform label. It is more useful to be able to provide brief descriptions of the clinical problem, since these can be used as a basis for formulating treatment:

- acute or chronic
- number of physical symptoms
- the nature and pattern of symptoms (i.e. clinical syndromes such as fatigue)
- association with anxiety disorder, depressive disorder, or other specific psychiatric disorder
- beliefs about cause

Assessment and treatment

The majority of those presenting unexplained symptoms in primary care require no more than medically appropriate assessment and reassurance (Table 5.2.1.4). The latter should convey to the patient that the symptoms are accepted as real and provide an explanation for their origin as well as answering the patient's worries. It is also necessary to discuss the results of any negative investigations fully.

Symptoms that persist or recur despite reassurance are generally regarded as difficult to treat. Continuing symptoms without any specific medical explanation are likely to confirm and maintain worries about serious illness, which may be further exacerbated by secondary anxiety and behavioural consequences. Therefore

Table 5.2.1.4 General principles of assessment

Consider psychological factors from the outset
Use appropriate physical investigation to exclude physical cause
Clarify psychological and physical complaints
Clarify previous personality and concerns about physical illness
Understand patient's beliefs and expectations
Identify depression or other psychiatric disorder
Identify psychosocial problems

effective treatment depends upon sympathetic treatment that meets the needs of both the patient and the family. A multi-causal view of aetiology leads to conclusions about treatment and avoids psychiatric diagnoses that may be unacceptable to the patient. Much can be done by general practitioners or non-specialists.

The general principles of treatment (Table 5.2.1.5) are similar for all forms of unexplained symptoms, single or multiple, but individual treatment plans must take account of psychiatric diagnoses of anxiety or depression and the particular type of physical symptoms.

The treatments for particular forms of somatoform disorder are discussed in later chapters. It is important to be aware that the commonest type of somatoform disorder, Undifferentiated Somatoform Disorder is not discussed separately; treatment follows the general principles described in this and other chapters. The treatments of other functional syndromes such as irritable bowel syndrome, chronic fatigue, and atypical facial pain all depend on the therapist being familiar with these syndromes and being able to provide an appropriate combination of treatment methods. For example, the management of physical de-conditioning is central to the treatment of chronic fatigue, whereas antidepressant medication has a major role in the treatment of atypical facial pain. The chapter on chronic fatigue (Chapter 5.2.7) is an example of a functional syndrome.

Much can be achieved by components of good non-specialist care, such as the following:

- discussion and explanation of the aetiology
- treatment of any minor underlying physical problem
- anxiety management (including tapes and handouts)
- advice on diary monitoring and graded return to full activities
- specific self-help programmes (e.g. chronic fatigue, irritable bowel syndrome)
- including relatives in the assessment, discussion of the nature of the problems, and explanations of the treatment

Table 5.2.1.5 General principles of treatment

Emphasize that symptoms are real and familiar and that medical care is appropriate
Minimize and control physical care
Offer an explanation and discuss
Allow patients and families to ask questions
Discuss the role of psychological factors in all medical care
Treat any primary psychiatric disorder
Agree a treatment plan

However, chronic and recurrent problems may need specialist treatment:

- psychotropic medication (antidepressants, anxiolytics)
- cognitive behavioural therapy
- interpretative psychotherapy (individual and group)
- specific psychiatric treatment for associated psychiatric and social problems
- programme to co-ordinate and control all medical care

There is a lot of evidence on the effectiveness of a range of treatments in specialist care,[5,6] but there is much less evidence about simple routine measures. The outlook for simpler syndromes of relatively recent onset is good, but the prognosis for very prolonged chronic, multiple, or recurrent syndromes (e.g. somatization disorder) is not as good. In these circumstances the control of medical care and the prevention of further iatrogenic disability may be more realistic than cure.

Further information

Mayou, R., Kirmayer, L.J., Simon, G., *et al.* (2005). Somatoform disorders: time for a new approach. *American Journal of Psychiatry*, **162**, 847–55.

Maj, M., Akiskal, H.S., Mezzich, J., *et al.* (eds.) *Somatoform disorders*. Wiley, Chichester.

References

1. Üstün, T.B. and Sartorius, N. (1995). *World health organization. Mental illness in general health care. An international study*. Wiley, Chichester.
2. Mayou, R., Kirmayer, L.J., Simon, G., *et al.* (2005). Somatoform disorders: time for a new approach. *American Journal of Psychiatry*, **162**, 847–55.
3. Barsky, A.J., Orav, E.J., and Bates, D.W. (2005). Somatization increases medical utilization and costs independent of psychiatric and medical comorbidity *Archives of General Psychiatry*, **62**, 903–10.
4. Fink, P., Hansen, M.S., and Oxhoj, M.L. (2004). The prevalence of somatoform disorders among internal medicine patients. *Journal of Psychosomatic Research*, **56**, 413–18.
5. Kroenke, K. and Swindle, R. (2000). Cognitive-behavioral therapy for somatization and symptom syndromes; a critical review of controlled clinical trials. *Psychotherapy and Psychosomatics*, **69**, 205–15.
6. O'Malley, P.G., Jackson, J.L., Santoro, J., *et al.* (2005) Antidepressant therapy for unexplained symptoms and symptom syndromes. *The Journal of Family Practice*, **48**, 980–90.

5.2.2 Epidemiology of somatoform disorders and other causes of unexplained medical symptoms

Gregory Simon

While nearly every psychiatric syndrome may include some somatic signs or symptoms, a specific group of syndromes has been traditionally defined as somatoform. This group of disorders is distinguished by certain key features: prominent reporting of somatic

symptoms, concern about medical illness, and frequent presentation to general medical providers. As in other categories of mental disorder, the boundaries between individual syndromes are more distinct in our systems of classification than they are in nature. Understanding that various somatoform disorders often overlap, this review is organized according to the major categories of somatoform disorder described in the ICD and DSM classification systems.

Somatization disorders

Phenomenology

The term somatization has been used to refer to a variety of clinical phenomena. One traditional view defines somatization as an inability or unwillingness to express emotional distress,[1] so that somatic symptoms are an alternative 'idiom of distress'. An alternative view defines somatization as the presentation of somatic complaints to medical providers in the presence of an occult anxiety or depressive disorder.[2] A third view defines somatization as somatic symptoms, which have no clear medical explanation.[3] While these definitions appear closely related, they identify somewhat different groups of patients. The third definition (presentation of unexplained somatic symptoms) is used by official systems of classification and by most epidemiological studies, so this review will focus on that phenomenon.

Both the ICD and DSM classification systems define somatization disorder as a chronic condition characterized by the reporting of numerous unexplained somatic symptoms.[4, 5] Recent versions of both classification systems identify a core syndrome of somatization (a persistent tendency to report multiple unexplained somatic symptoms) using a simplified set of diagnostic criteria.

Prevalence

The reported prevalence of well-defined somatization disorder appears to depend significantly on the method used for assessment. Community and primary care surveys have typically relied on structured interviews to assess the lifetime prevalence of unexplained somatic symptoms. Community surveys in North America[6] and Western Europe[7, 8] have found prevalence rates of less than 2 per cent with primary care surveys finding only slightly higher prevalence rates.[9] Data from the World Health Organization (**WHO**) multicentre primary care survey indicate that recall during structured interviews may significantly underestimate the lifetime prevalence of somatization symptoms.[10] More accurate recall of lifetime symptoms (by either repeated assessments or the use of medical records) might yield significantly higher prevalence rates.

Correlates

The prevalence of somatization disorder and unexplained somatic symptoms is typically twice as high in women as in men,[11, 12] and this difference appears at time of menarche.[13] Community and primary care surveys demonstrate a substantial overlap between somatization disorder and anxiety and depressive disorders.[7, 14, 15] Anxiety and depressive disorders also predict the subsequent onset of somatization disorder.[16]

Available data show a mixed picture regarding cross-national or cross-cultural differences in the prevalence of somatization. Studies of clinical samples find that somatic symptoms are a common accompaniment of depressive and anxiety disorders worldwide.[17–19] The WHO primary care survey documented large differences in the prevalence of unexplained somatic symptoms with a markedly higher prevalence in South America than in Europe or the United States.[9] That same study, however, found that the association between unexplained symptoms and symptoms of depression or anxiety was similar across a wide range of cultures and levels of economic development.[14] One explanation for these apparently disparate findings is that the prevalence of unexplained somatic symptoms (like the prevalence of anxiety or depressive disorder) varies widely across nations and cultures, but the association between somatic and psychological distress is universal. Countries or cultures with higher rates of anxiety or depressive disorders would be expected to have higher prevalence of somatization disorder and other somatization syndromes. Given the consistent overlap between somatization disorders and other common mental disorders, some have questioned whether these conditions actually belong in a distinct category.[20, 21]

Controversies and questions

Available data do not support a specific diagnostic threshold based on the number or distribution of unexplained somatic symptoms. An increasing number of somatic symptoms is consistently associated with increases in comorbid mood or anxiety disorder, functional impairment, and use of health services.[14,15] Mindful of this continuum, both Escobar *et al.*[22] and Kroenke *et al.*[23] have described less restrictive somatization syndromes, which, despite their higher prevalence, are strongly associated with impairment and the use of health services. Both the ICD and DSM classification systems describe subthreshold or less extreme forms of this condition characterized by a smaller number of medically unexplained symptoms.[5, 24]

Longitudinal data raise questions about the presumed stability or chronicity of somatization disorder or medically unexplained somatic symptoms. Traditional descriptions of somatization disorder emphasize its stability and chronicity. Data from the WHO primary care survey, however, suggest that individual somatization symptoms vary considerably over time.[9] While the syndrome of somatization seemed somewhat more stable than anxiety and depressive disorders (typically regarded as episodic), only half of the primary care patients, satisfying Escobar's criteria for somatization syndrome at the baseline assessment, continued to meet the criteria one year later.

Hypochondriacal disorders

Phenomenology

Both the ICD and DSM classification systems define hypochondriasis by the triad of disease conviction, functional impairment, and refusal to accept appropriate reassurance.

Prevalence

Attempts to estimate the prevalence of hypochondriasis have been limited by the absence of proven standardized methods for standardized assessment. Community surveys find prevalence rates of 1 per cent or less,[25] while primary care surveys typically find rates of approximately 5 per cent,[26, 27] while the WHO multicentre primary care survey[28] found an overall prevalence of only 0.8 per cent.

In reviewing data from the WHO survey, Gureje *et al.*[28] found that a less restrictive definition more than doubled the prevalence rate (to 2.2 per cent). Cases added by this relaxed definition did not differ significantly from those satisfying CIDI/ICD criteria, suggesting that CIDI/ICD criteria may be somewhat too restrictive.

Correlates

Despite the variation in prevalence, primary care surveys yield similar results regarding demographic correlates of hypochondriasis. The prevalence of hypochondriasis is 1.5 to 2 times as great in women as men but does not appear to vary significantly with age.[28]

Controversies and questions

While the ICD and DSM classification systems suggest that hypochondriasis is distinct from anxiety and depressive disorders, available data suggests considerable overlap. In every sample examined, hypochondriasis is strongly associated with major depression, panic disorder, and generalized anxiety disorder.[26,28–30] Among those with hypochondriasis, clinical features do not clearly distinguish those with and without a comorbid psychiatric diagnosis.[30] In addition, changes over time in anxiety or depression are consistently associated with parallel changes in symptoms of hypochondriasis.[31] As with somatization disorders, some have recently argued that hypochondriasis be re-classified as a form of anxiety disorder.[20,21]

Pain syndromes

Phenomenology

While the ICD and DSM classification systems both define somatoform pain disorders, the two systems differ in their descriptions of the clinical features. In the ICD diagnostic system, somatoform pain disorder is defined as persistent pain without clear medical explanation.[5] The DSM system[4] specifies that 'psychological factors are judged to have an important role in the onset, severity, exacerbation, or maintenance of the pain'. Both definitions are somewhat problematic. Basic research on neural changes associated with persistent pain raise doubts about the distinction between pain with and without a biomedical explanation.[32] As discussed below, longitudinal research supports the view that persistent pain causes psychological disorder as much as it supports the DSM view that pain results from psychological distress. Recent epidemiological research has attempted to avoid questions of aetiology and has examined the prevalence and correlates of persistent pain.

Prevalence

Epidemiological studies consistently find that pain syndromes are among the most common problems presented to general medical providers. Population surveys indicate that over 25 per cent of community residents suffer from recurrent or persistent pain symptoms and that 2 to 3 per cent experience disabling pain syndromes.[33] The recent WHO primary care survey[34] found that approximately 20 per cent of primary care patients suffered from persistent pain (one or more pain symptoms present for most of the last 6 months). Pain syndromes are approximately twice as prevalent in women as in men.[15]

The limited data available do not allow definite conclusions about cross-cultural or cross-national variability in pain syndromes. Some studies have documented cross-national or cross-cultural differences based on small samples of patients treated for pain syndromes—often in specialist pain clinics.[35] The WHO primary care survey[34] included both the largest number of patients with pain syndromes as well as the broadest range of cultures and levels of economic development. In that study, both the prevalence and correlates of persistent pain varied widely across sites. No clear pattern (e.g. a higher prevalence in developing or non-Western countries) was evident.

Correlates

All available data indicate that pain symptoms are strongly associated with anxiety and depressive disorders. This relationship has been consistently demonstrated in both community[36] and primary care[34] studies across a broad range of cultural and socioeconomic divides. Psychological distress is most strongly associated with pain occurring at multiple sites and pain associated with functional impairment.[37, 38] While epidemiological studies strongly support an association between pain complaints and psychological distress, this does not necessarily imply that pain is a consequence of psychological distress. Some studies find that the presence of psychological distress predicts the onset of pain syndromes,[39,40] while others support the opposite relationship—that persistent pain predicts subsequent psychological disorder.[41]

Other somatoform conditions

Specific somatoform syndromes

A number of specific somatic syndromes have been described over the last several decades. These specific syndromes are sometimes defined by the particular somatic symptoms experienced (e.g. fibromyalgia, irritable bowel syndrome, chronic fatigue syndrome) and sometimes by particular beliefs about aetiology (e.g. multiple chemical sensitivity, systemic candidiasis, electrical allergy). In every case, controversy persists about whether the somatic symptoms should be considered 'medically unexplained' (that is to say a somatoform disorder). Community surveys suggest that non-specific symptoms (such as fatigue or diffuse musculoskeletal pain) are common, but that the prevalence of strictly defined syndromes (such as fibromyalgia or chronic fatigue syndrome) varies considerably with the criteria applied.[42, 43] Most of these syndromes appear more often in women than in men.[44] Both community[43, 44] and primary care surveys[45, 46] have found several of these syndromes to be associated with anxiety and depressive symptoms. However, two studies of chronic fatigue[47] and irritable bowel[48] symptoms found that psychological distress was associated with seeking care for somatic symptoms rather than the presence or nature of the somatic symptoms themselves.

Body dysmorphic disorder or monosymptomatic hypochondriasis

Body dysmorphic disorder has recently been identified as a distinct clinical entity. The limited epidemiological data available suggest significant overlap with anxiety and depressive disorders.[8, 50] Prevalence estimates range from less than 1 per cent among unselected primary care patients,[8] to as high as 5 to 10 per cent among patients seeking cosmetic surgery[49] or outpatients with anxiety

Table 5.2.2.1 Summary of prevalence rates of specific somatoform disorders in community and primary care studies

	Community studies	Primary care studies	Notes
Somatization disorder	1%	1–2%	Diagnostic interviews probably under-count past symptoms, records reviews tend to find higher rates
Multisomatoform disorder or subthreshold somatization	5%	8–10%	Less extreme form of somatization, but still strongly associated with disability and symptoms of depression and anxiety
Hypochondriasis	1%	1–5%	Strongly associated with anxiety disorders
Persistent pain syndromes	2–3%	15–20%	Bi-directional relationship with depressive and anxiety disorders
Body dysmorphic disorder	1%	1–5%	Significantly higher in certain medical settings (dermatology, cosmetic surgery) and in people with anxiety disorders

disorders.[50] Some questionnaire studies find prevalence rates as high as 5 per cent among university students.[51]

Conversion disorders

Limited epidemiological data are available concerning conversion-type somatoform disorders. As with other varieties of unexplained somatic symptoms, these disorders appear to be more common among women[52] and are associated with increased prevalence of depressive and anxiety disorders.[53] Some evidence suggests that these conditions have declined in prevalence.[52] While many sources report that these conditions are more common in non-Western or developing countries, available epidemiological data do not necessarily support this view.[9]

Factitious disorders

No systematic data are available regarding the epidemiology of factitious disorders. The available data include numerous case reports and a small number of case series—typically drawn from medical inpatient or medical specialty settings. Because the syndrome is defined by deception, it is likely that a large proportion of cases go undetected.

Further information

Creed, F. and Barsky, A. (2004). A systematic review of the epidemiology of somatisation disorder and hypochondriasis. *Journal of Psychosomatic Research*, **56**, 391–408.

Ustun, T.B. and Sartorius, N. (eds.). (1995). *Mental illness in general health care*. John Wiley & Sons, Chichester, England.

References

1. Kleinman, A. (1977). Depression, somatization and the "new cross-cultural psychiatry". *Social Science & Medicine*, **11**, 3–10.
2. Goldberg, D. and Bridges, K. (1988). Somatic presentations of psychiatric illness in primary care. *Journal of Psychosomatic Research*, **32**, 137–44.
3. Barsky, A. (1992). Amplification, somatization, and the somatoform disorders. *Psychosomatics*, **33**, 28–34.
4. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
5. World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders. Clinical descriptions and diagnostic guidelines*. World Health Organization, Geneva.
6. Swartz, M., Blazer, D., George, L., *et al.* (1986). Somatization disorder in a community population. *American Journal of Psychiatry*, **143**, 1403–8.
7. Grabe, H., Meyer, C., Hapke, U., *et al.* (2003). Specific somatoform disorders in the general population. *Psychosomatics*, **44**, 304–11.
8. Faravelli, C., Salvatori, S., Galassi, F., *et al.* (1997). Epidemiology of somatoform disorders: a community survey in Florence. *Social Psychiatry and Psychiatric Epidemiology*, **32**, 24–9.
9. Gureje, O., Simon, G., Ustun, T., *et al.* (1997). Somatization in cross-cultural perspective: a World Health Organization study in primary care. *American Journal of Psychiatry*, **154**, 989–95.
10. Simon, G. and Gureje, O. (1999). Stability of somatization disorder and somatization symptoms among primary care patients. *Archives of General Psychiatry*, **56**, 90–5.
11. Kroenke, K. and Spitzer, R. (1998). Gender differences in the reporting of physical and somatoform symptoms. *Psychosomatcic Medicine*, **60**, 150–5.
12. Ladwig, K., Marten-Mittag, B., Erazo, N., *et al.* (2001). Identifying somatization disorder in a population-based health examination survey: psychosocial burden and gender differences. *Psychosomatics*, **42**, 511–18.
13. LeResche, L., Mancl, L., Drangsholt, M., *et al.* (2005). Relationship of pain and symptoms to pubertal development in adolescents. *Pain*, **118**, 201–9.
14. Simon, G.E., VonKorff, M., Piccinelli, M., *et al.* (1999). An international study of the relation between somatic symptoms and depression. *New England Journal of Medicine*, **341**, 1329–35.
15. Kroenke, K., Spitzer, R.L., Williams, J.B.W., *et al.* (1994). Physical symptoms in primary care: predictors of psychiatric disorders and functional impairment. *Archieves of Family Medicine*, **3**, 774–9.
16. Gureje, O. and Simon, G. (1999). The natural history of somatisation in primary care. *Psychological Medicine*, **29**, 669–76.
17. Ulasahin, A., Basoglu, M., and Paykel, E. (1994). A cross-cultural comparative study of depressive symptoms in British and Turkish clinical samples. *Social Psychiatry and Psychiatric Epidemiology*, **29**, 31–9.
18. Escobar, J., Gomez, J., and Tuason, V. (1983). Depressive phenomenology in North and South American patients. *American Journal of Psychiatry*, **140**, 47–51.
19. Ebert, D. and Martus, P. (1994). Somatization as a core symptom of melancholic type depression. Evidence from a cross-cultural study. *Journal of Affective Disorder*, **32**, 253–6.
20. Creed, F. and Barsky, A. (2004). A systematic review of the epidemiology of somatisation disorder and hypochondriasis. *Journal of Psychosomatic Research*, **56**, 391–408.
21. Mayou, R., Kirmayer, L., Simon, G., *et al.* (2005). Somatoform disorder: time for a new approach in DSM-V. *American Journal of Psychiatry*, **162**, 847–55.
22. Escobar, J.I., Burnam, M.A., Karno, M., *et al.* (1987). Somatization in the community. *Archives of General Psychiatry*, **44**, 713–18.
23. Kroenke, K., Spitzer, R., deGruy, F., *et al.* (1997). Multisomatoform disorder. An alternative to undifferentiated somatoform disorder for the somatizing patient in primary care. *Archives of General Psychiatry*, **54**, 352–8.

24. American Psychiatric Association. (1995). *Diagnostic and statistical manual of mental disorders* (DSM-IV) (4th edn), *primary care version*. American Psychiatric Press, Washington.
25. Looper, K.J. and Kirmayer L.J. (2001). Hypochondriacal concerns in a community population. *Psychological Medicine*, **31**, 577–84.
26. Escobar, J., Gara, M., Waitzkin, H., *et al.* (1998). DSM-IV hypochondriasis in primary care. *General Hospital Psychiatry*, **20**, 155–9.
27. Barsky, A., Wyshak, G., Klerman, G., *et al.* (1990). The prevalence of hypochondriasis in medical outpatients. *Social Psychiatry and Psychiatric Epidemiology*, **25**, 89–94.
28. Gureje, O., Ustun, T., and Simon, G. (1997). The syndrome of hypochondriasis: a cross-national study in primary care. *Psychological Medicine*, **27**, 1001–10.
29. Noyes, R.J. (1999). The relationship of hypochondriasis to anxiety disorders. *General Hospital Psychiatry*, **21**, 8–17.
30. Barsky, A., Wyshak, G., and Klerman, G. (1992). Psychiatric comorbidity in DSM-IIIR hypochondriasis. *Archives of General Psychiatry*, **49**, 101–8.
31. Simon, G., Gureje, O., and Fullerton, C. (2001). Course of hypochondriasis in an international primary care study. *General Hospital Psychiatry*, **23**, 51–5.
32. Coderre, T., Katz, J., Vaccarine, A.L., *et al.* (1993). Contribution of central neuroplasticity to pathological pain: review of clinical and experimental evidence. *Pain*, **52**, 259–85.
33. Von Korff, M., Dworkin, S.F., and Le Resche, L. (1990). Graded chronic pain status: an epidemiologic evaluation. *Pain*, **40(3)**, 279–91.
34. Gureje, O., Von Korff, M., Simon, G.E., *et al.* (1998). Persistent pain and well-being. A World Health Organization study in primary care. *Journal of American Medical Association*, **280**, 147–51.
35. Sanders, S., Brena, S., Spier, C., *et al.* (1992). Chronic low back pain patients around the world: cross-cultural similarities and differences. *Clinical Journal of Pain*, **8**, 317–23.
36. McWilliams, L., Cox, B., and Enns, M. (2003). Mood and anxiety disorders associated with chronic pain: an examination in a nationally representative sample. *Pain*, **106**, 127–33.
37. Dworkin, S.F., Von Korff, M., and LeResche, L. (1990). Multiple pains and psychiatric disturbance: an epidemiologic investigation. *Archives of General Psychiatry*, **47**, 239–44.
38. Benjamin, S., Morris, S., McBeth, J., *et al.* (2000). The association between chronic widespread pain and mental disorder: a population-based study. *Arthritis and Rheumatism*, **43**, 561–7.
39. Von Korff, M., Le Resche, L., and Dworkin, S.F. (1993). First onset of common pain symptoms: a prospective study of depression as a risk factor. *Pain*, **55**, 251–8.
40. Croft, P., Papegeorgiou, A., Ferry, S., *et al.* (1995). Psychologic distress and low back pain. Evidence from a prospective study in the general population. *Spine*, **15**, 2731–7.
41. VonKorff, M. and Simon, G.E. (1996). The relationship between pain and depression. *British Journal of Psychiatry*, **168**(Suppl. 30), 101–8.
42. Wessely, S., Chalder, T., Hirsch, S., *et al.* (1997). The prevalence and morbidity of chronic fatigue and chronic fatigue syndrome: a prospective primary care study. *American Journal of Public Health*, **87**, 1449–55.
43. Wolfe, F., Ross, K., Anderson,J., *et al.* (1995). The prevalence and characteristics of fibromyalgia in the general population. *Arthritis and Rheumatism*, **38**, 19–28.
44. Pawlikowska, T., Chalder, T., Hirsch, S., *et al.* (1994). Population-based study of fatigue and psychological distress. *British Medical Journal*, **308**, 763–6.
45. Wessely, S., Chalder, T., Hirsch, S., *et al.* (1996). Psychological symptoms, somatic symptoms, and psychiatric disorder in chronic fatigue and chronic fatigue syndrome: a prospective study in the primary care setting. *American Journal of Psychiatry*, **153**, 1050–9.
46. Henningsen, P., Zimmermann, T., and Sattel, H. (2003). Medically unexplained physical symptoms, anxiety, and depression: a meta-analytic review. *Psychosomatic Medicine*, **65**, 528–33.
47. Lawrie, S., Manders, D., Geddes, J., *et al.* (1997). A population-based incidence study of chronic fatigue. *Psychological Medicine*, **27**, 343–53.
48. Whitehead, W., Bosmajian, L., Zonderman, A., *et al.* (1988). Symptoms of psychologic distress associated with irritable bowel syndrome. Comparison of community and medical clinic samples. *Gastroenterology*, **95**, 709–14.
49. Sarwer, D., Wadden, T., Pertschuk, M., *et al.* (1998). Body image dissatisfaction and body dysmorphic disorder in 100 cosmetic surgery patients. *Plastic and Reconstructive Surgery*, **101**, 1644–9.
50. Simeon, D., Hollander, E., Stein, D., *et al.* (1995). Body dysmorphic disorder in the DSM-IV field trial for obesessive-compulsive disorder. *American Journal of Psychiatry*, **152**, 1207–9.
51. Bohne, A., Wilhelm, S., Keuthen, N., *et al.* (2002). Prevalence of body dysmorphic disorder in a German college student sample. *Psychiatry Research*, **109**, 101–4.
52. Singh, S. and Lee, A. (1997). Conversion disorders in Nottingham: alive, but not kicking. *Journal of Psychosomatic Research*, **43**, 425–30.
53. Sar, V., Akyuz, G., Kundakci, T., *et al.* (2004). Childhood trauma, dissociation, and psychiatric comorbidity in patients with conversion disorder. *American Journal of Psychiatry*, **161**, 2271–6.

5.2.3 Somatization disorder and related disorders

Per Fink

Introduction

The essential feature of somatization disorder and related disorders is that the patient presents multiple, medically unexplained symptoms or functional somatic symptoms. These physical complaints are not consistent with the clinical picture of known, verifiable, conventionally defined diseases, and are unsupported by clinical or paraclinical findings. The phenomenon of medically unexplained symptoms cannot simply be classified into one or a few diagnostic categories, but must be regarded as an expression of a basic mechanism by which people may respond to stressors as in the cases of depression and anxiety.[1–3] Somatization disorder and related disorders must thus be considered to possess a spectrum of severity.[3, 4] In this chapter, the focus will be on the chronic and multisymptomatic forms.

The somatization disorder diagnosis has its origin in the concept of hysteria. It was introduced in DSM-III in 1980 and was based on the criteria for 'Briquet's syndrome', a syndrome described in the early 1960s by Perley and Guze.[5] They listed 59 physical and psychological symptoms distributed in 10 groups: 25 of the symptoms from nine groups were required to qualify for the diagnosis of somatization disorder. All psychological symptoms were eliminated in the DSM-III modification to avoid overlapping with other diagnoses.

The diagnostic criteria for DSM somatization disorder varied until the introduction of the current DSM-IV. The diagnosis was included in ICD-10 in 1992, but the ICD-10 criteria list different

symptoms, and require a different number of symptoms compared with the corresponding DSM criteria.

The somatization disorder diagnosis has been criticized for being too rigid for clinical use. Only the most severe cases with a specific predefined symptom profile fulfil the diagnostic criteria, and the majority of those with multiple symptoms fall into one of the residual categories of 'undifferentiated' or 'not otherwise specified' somatoform disorders.[6]

To increase the sensitivity, Escobar *et al.*[7] introduced an abridged somatization index. This required 4 symptoms for males and 6 symptoms for females out of the 37 somatic symptoms listed in the DSM-III, compared with 12 and 14 symptoms respectively for the full DSM-III somatization disorder diagnosis. Kroenke *et al.*[8] have suggested a diagnosis of 'multisomatoform disorder', defined as three or more medically unexplained physical symptoms from a 15-symptom checklist along with at least a 2-year history of medically unexplained symptoms.

However, these abridged versions share the same basic problem as the original ones, namely that the chosen number of symptoms to qualify for the individual diagnoses is arbitrary and not empirically based. Recently a new empirically based construct was introduced, and this may have solved the problem. This new 'bodily distress disorder' diagnosis is based on positive criteria and not solely on the exclusion of all organic possibilities.[3] (Table 5.2.3.1)

This chapter will not differentiate between the different subcategories of somatoform and related disorders that are present along with somatic symptoms.

Table 5.2.3.1 Symptoms of and diagnostic criteria for bodily distress disorder

Yes	No	Symptom groups
		≥ 3 Cardiopulmonal/autonomic arousal Palpitations, heart pounding, precordial discomfort, breathlessness without exertion, hyperventilation, hot or cold sweats, trembling or shaking, dry mouth, churning in stomach, "butterflies", flushing or blushing
		≥ 3 Gastrointestinal arousal Frequent loose bowel movements, abdominal pains, feeling bloated, full of gas, distended, heavy in the stomach, regurgitations, constipation, nausea, vomiting, burning sensation in chest or epigastrium
		≥ 3 Musculoskeletal tension Pains in arms or legs, muscular aches or pains, feelings of paresis or localized weakness, back ache, pain moving from one place to another, unpleasant numbness or tingling sensations
		≥ 3 General symptoms Concentration difficulties, impairment of memory, feeling tired, headache, memory loss, dizziness
		≥ 4 symptoms from one of the above groups
Diagnostic criteria:		
1–3: 'yes': Moderate 'bodily distress disorder'		
4–5: 'yes': Severe 'bodily distress disorder'		

Clinical features

Physical symptoms and complaints

Patients with somatization and related disorders may complain of any medically unexplained non-verifiable subjective physical symptoms, and the symptoms may refer to any part or system of the body.

Complaints can be divided into:

- *Subjective symptoms*, which are sensations and other complaints that cannot be verified by another individual or by general methods of examination (e.g. pains and paraesthesia).
- *Objective symptoms*, which are complaints that can be verified if present at the time of examination (e.g. haematuria, icterus, etc.).

Findings can be divided into:

- *Provoked findings*, i.e. symptoms or signs (such as soreness resulting from pressure or sensory impairment), which the patient is unaware of until these are provoked during the physical examination
- *Certain findings*, which include objective symptoms that are verified and phenomena unnoticed by the patient but found during the physical examination (such as an abdominal tumor)

The symptom complaints in patients with somatization disorder and related disorders are dominated by subjective symptoms and provoked findings, whereas objective symptoms and positive certain findings and paraclinical findings are unusual.

Subjective symptoms may be considered to be psychological phenomena arising from personal experiences, which others cannot judge or measure, despite the fact that these symptoms could be fully explained by the presence of organic pathology. This means that there are considerable inter-individual, cultural, and historical variations in the symptom presentation, which are determined by the patient's life experience and sociocultural background, and the setting in which the patient is seen also plays a role.[9] However, the patients may also present verifiable symptoms and signs due to a physical disease or defect, which they exaggerate and incorporate into their illness. Incidental inborn errors or degenerative changes, which are asymptomatic in most individuals, may tenuously be assigned clinical importance by the doctor or the patient. For instance, degenerative changes in the spinal column are seen in most individuals when they become older, and most do not have any pain, but a patient's backache may be attributed to those changes. Furthermore, over time, the patient with a chronic condition is likely to have undergone multiple tests, invasive procedures, operations, and received medications for treatment or diagnostic purposes, and this may cause not only iatrogenic harm but also physical complications.[10] Finally, the patient may have a concurrent physical disease. The presented symptoms may thus be a difficult mixture of complaints of both organic and non-organic origin.[11]

The patients typically complain of *multiple physical symptoms*, but the number of symptoms reported by the patients vary considerably from one patient to another and over time in the same individual. The patients may complain of multiple, medically unexplained symptoms in numerous bodily systems at presentation, but sometimes the complaints are concentrated on one

symptom pattern at one time (e.g. a gastrointestinal illness) and on a different symptom pattern at another (e.g. a cardiopulmonary illness).[11] This single-organ illness picture may be due to physicians being inclined to focus their attention and investigation on the organ of their own specialty—especially in a multisymptomatic, complex patient, i.e. a gastroenterologist will focus on gastrointestinal symptoms and may ignore musculoskeletal symptoms. A new set of symptoms from another organ system may come to attention when diagnoses and treatment options have been ruled out for the current complaint. Iatrogenic factors may thus contribute significantly to the presented symptom pattern and changes in symptom pattern. Patients with somatization disorder are often inconsistent historians. They may supply incorrect information about previous episodes of their illness, minimizing or ignoring earlier instances of illness and exclusively focusing on the current symptom pattern. This may be because the patients find it difficult to account for their complicated medical history, or because they do not want to confuse the doctor with what they believe is irrelevant information. Therefore, the full clinical picture often only becomes evident after a full medical history has been obtained and the patient has been followed for some time.

Symptoms and findings are not idiosyncratic but need clarification or specification before becoming meaningful clinically. For instance, a patient complains of chest pain. There are multiple causes for chest pain, so for the doctor this is not very informative. A clarification or specification is needed. A retrosternal-localized pain of a pressing nature offers the doctor a very different diagnostic association than a chest pain that is described as being stabbing and located in the left side of the chest. In somatization disorder and related disorders, the patient usually presents a *vague illness picture* with symptoms of *non-specific* character and of low diagnostic value, i.e. symptoms that are common in the general population and which are found in many different mental and physical disorders (symptoms like fatigue, nausea, headache, dizziness).

The presentation of medically unexplained symptoms is *atypical*, that is to say the symptoms lie outside what is usual in an authentic physical disease.[11] However, the patients may have 'learned' the typical symptom presentation from different sources. For example, a patient with atypical asthma-like attacks shared a room with a patient with genuine asthma during her third hospital stay; subsequently, her attacks took on a more 'authentic' appearance.[11]

Descriptions of *symptoms are usually vague, imprecise, and inconsistent*, and the patients often have difficulties giving further details about their illness and symptoms, i.e. describing the quality, intensity, and chronology. The symptoms are described as being of maximum intensity all the time, but if the patients keep a diary of symptoms or if information is gathered from other sources like relatives or the family physician, a considerable variation in the symptom intensity from day to day or from year to year often surfaces. The patients may have difficulties in the chronology of symptoms, mixing current and past symptoms and illness episodes in a disorganized and confusing manner. It is difficult for the patients to identify relieving factors or behaviour and to identify triggering events or things that make them worse, or these are multiple, or vague and unspecific. This is in contrast to patients with physical disease, who usually describe their symptoms in a consistent and precise manner (Table 5.2.3.2).

Typically, there is a marked *discrepancy* between a patient's subjective complaints and reports on his or her functioning when this is compared with the way the patient is observed to act, move

Table 5.2.3.2 Characteristic differences in symptom description and other characteristics of well-defined somatic and related disorders including functional somatic syndromes

	Somatization disorder and related disorders	Physical disease
Symptom description		
Location	Vague, diffuse, alternating	Well-defined, constant
Intensity	Vague, indistinctly defined intensity, few variations, often at maximum at all times	Well-defined changes and levels of intensity
Periodicity	Diffuse, difficult to define, are often denied	Typically well-defined periods with aggravation or improvement
Relieving or aggravating factors	Vague, indistinct, numerous	Well-defined, few
Number of symptoms	Numerous, vague	Few, well-defined, clearly described
The nature of symptoms	Unspecific	Specific
The character of the symptoms	Uncharacteristic	Characteristic
Iatropic symptoms and main complaints	Vague, difficult to identify	Can be identified and delimited from comorbid symptoms
Description	Affective, emotional, interpreting	Clear and descriptive
Other characteristics		
Treatment and medication	Difficult to assess the effect, transitory	Level of effect well-defined
Previous treatment	Unclear what treatment the patient has undergone. Diagnostic tests are often interpreted as treatment	The patient can account for previous treatments
Emotionality	Inadequate, e.g. exaggerated or marked unaffectedness ('la belle indifférence')	Adequate - empathic

and perform during the examination, or compared with information from other sources like family. For instance, the patient moves and sits completely freely despite complaining of severe back pain or gives detailed information despite complaints of severe memory impairment.

There may be *emotional discrepancy* in which the patient shows a lack of concern about the nature and implications of the symptoms despite presenting severe symptoms that are threatening the patient's future functioning and quality of life. Other patients may in turn be very affected and emotional in their description of the symptoms and illness, describing in a colourful, dramatic, and strikingly graphic manner.

The patient's centre of attention is typically on the suffering, on the psychosocial consequences, and the restrictions that the symptoms impose on their life. On the contrary, patients with well-defined physical disease are concerned or worried about the implication of their disease for their future health, i.e. will they recover or will they die from the disease. This emotional or *psychosocial communication* among patients with somatization disorder may put pressure on the doctor to do something.

Patients with somatization disorder and related disorders usually attribute their symptoms to a physical disease, and in some cases they persistently may *refuse to accept medical reassurance* despite appropriate medical evaluations. The ICD-10 criteria for somatoform disorders include this refusal to accept medical reassurance, but recent research indicates that many patients are unsure what is wrong with them, and they do not necessarily refuse non-medical explanations if they are presented in a meaningful and understandable way.[12] Although the patients may recognize that their physical symptoms are caused by, e.g. stressful events, this does not make the symptoms disappear, and they still need treatment. The weight of the symptom (refusing to accept medical reassurance) in the medical literature therefore seems out of proportion. However, in the most severe cases, patients may be involved in patient organizations fighting for their illness to be recognized as an 'authentic' physical disease or fighting for a particular causality of their illness as, e.g. whiplash-associated disorder or hypersensitivity to electricity or chemicals (multiple chemical sensitivity). The patients may also fight for disability pension or financial compensation. This is often more a question of getting their illness recognized than receiving financial compensation. Some patients may be preoccupied with the idea that they have been mistreated or neglected by doctors, and this group will often become involved in conflicts with doctors and in legal disputes.

In some cases, there is a sudden onset of the disorder in connection with a medical condition or a trauma in a previously normal and healthy individual. It could be a whiplash trauma, a fracture, an infection, or an acute intoxication. The symptoms persist despite the original disease being cured according to a biomedical or a surgical judgement. Instead, the illness worsens and more symptoms may emerge. Our knowledge about such disorders with abrupt onset is sparse.

Psychological symptoms and comorbidity

At examination, the patients may deny emotional symptoms or conflicts, and when they do report them, they often blame them on their physical affliction. Patients may also be reluctant to display emotional difficulties because of bad experiences of doing so. They may have experienced that doctors did not believe them or accused them of making up their symptoms and have consequently felt that their physical problems are not taken seriously. However, sooner or later, most patients will exhibit emotional difficulties, and if the patients feel understood by the physician, emotional problems may as well be presented. Patients may present many different types of emotional symptoms, often unspecific, but prominent anxiety and depressive symptoms are prevalent. Although the symptoms may be as marked as in affective and anxiety disorders, they are usually more transient, changing from one day to another and especially related to specific events. At times, the psychological symptoms may fulfil the criteria for a mood or anxiety disorder; but it is characteristic that the illness picture shows variations in both bodily and emotional symptoms.

Suicidal attempts are unusual but may occur especially among severe cases, but suicide is rare. Substance abuse is frequent, whether or not this is iatrogenic sanctioned.

The way the patients present themselves is inextricably linked to personal style and possible personality disorder. As a broad spectrum of personality disorders or traits[13] is associated with somatization disorder, the presenting style varies greatly from one patient to another. Characteristically, three broad patterns of personality style may be found in these patients, especially in chronic cases: dramatic–emotional type, paranoid–hostile type, and passive–aggressive-dependent type. The same patient may show all three patterns. In less severe cases it is often observed clinically that the patients have previously been very active and hard working, have conformed socially, and had a strong social network with many responsibilities. The patients often display perfectionist traits and prefer to be in control of a situation.

Illness behaviour

Typically, patients with somatization disorder persistently exhibit consulting behaviour which results in an excessive use of medical services and alternative therapies. In chronic cases, they have often been subject to a large number of futile examinations, surgery, and medical/surgical attempts at treatment.[10,14,15] However, some patients realize quite early in the illness course that the doctors cannot help them, or they are well managed by their family physician, so they do not necessarily display this consultation behaviour.

Due to negative results of medical check-up and treatment attempts and the patients' persistent belief that they must have a physical disease, the patients may consult different physicians. The patients may have been, or may feel, mistreated or neglected by doctors and therefore want to get a second opinion, or they want to find a doctor who can help them. Sometimes this behaviour, together with the patients' personality, can result in disagreement and a mutual hostility between the patients and their doctors.

Furthermore, the different illness patterns at different times, combined with the patients' seductive, demanding personality style, may result in disagreements between the different health care professionals involved in their care, which may complicate their care.

In chronic cases, all aspects of the patients' social and family life may be centered around their illness, so that the whole of their family life is adjusted to the patients' demands ('illness as a way of living').[16]

Classification

The diagnostic criteria for somatization disorder and related disorders have varied, with different permutations of the diagnostic terminology reflecting difficulties in classification and in establishing valid criteria. The distinction between the individual somatoform disorders is unclear, which means that the majority of patients will exhibit clinical characteristics from different diagnostic categories.[6]

Except for hypochondriasis or health anxiety, the somatoform disorder categories are primarily based on the number or specificity of bodily symptoms and on the duration of illness.[17] The disorders can be divided into acute and chronic forms, into a multisymptomatic form, and into a form in which the patients only present few symptoms or symptoms mainly referring to a single organ system. The somatization disorder diagnosis includes the most chronic multisymptomatic cases lasting for 2 years or more.

The ICD-10 criteria require the following:

1 at least 2 years of multiple and variable physical symptoms for which no adequate physical explanation has been found;

2 persistent refusal to accept the advice or reassurance of several doctors that there is no physical explanation for the symptoms;

3 some degree of impairment of social and family functioning attributable to the nature of the symptoms and resulting behaviour.

The ICD-10 criteria require 'multiple physical symptoms' to include at least 6 out of 14 predefined symptoms, involving at least 2 of the following: gastrointestinal, cardiovascular, urogenital, or skin or pain symptoms. In contrast, the DSM-IV criteria demand four pain symptoms, two gastrointestinal symptoms, one sexual symptom, and one pseudoneurological symptom that are not fully explained by a medical condition. No specific symptoms are listed, but examples are given. Consequently, there is only poor to moderate agreement between the DSM-IV and ICD-10 diagnostic criteria.[16] Cases lasting less than 2 years are classified as undifferentiated somatization disorder. In ICD-10, multiple symptoms are required, which is not the case in DSM-IV.

Functional or somatoform diagnoses defined mainly by the number or specificity of bodily symptoms include, besides somatization disorder (F45.0 and 300.81), undifferentiated SD (F45.1 and 300.81), persistent somatoform pain disorder (F45.4 and 307.80), and somatoform disorder unspecified/NOS (F45.9 and 300.82). In ICD-10, this also includes somatoform autonomic dysfunction.[3–5] There may be reasons for also including neurasthenia (F48.0) into the group. The somatoform disorder concept has never been accepted among non-psychiatrists, which has led to the introduction of many different functional somatic syndromes, e.g. chronic fatigue syndrome (CFS), fibromyalgia, irritable bowel syndrome (IBS), and chronic benign pain syndrome, and new syndromes are intermittently introduced.[18]

The newly introduced diagnosis of bodily distress syndrome or disorder may be a solution to this classification problem, although it has not yet been sufficiently tested in daily clinical practice.[3] The suggested diagnosis is based on an analysis of a large sample of patients from different medical settings, and it seems to encompass the various functional syndromes advanced by different medical specialties as well as somatization disorder and related diagnoses of the psychiatric classification. The disorder may have different manifestations, i.e. GI, MS, or CP syndromes as shown in Table 5.2.3.1.

Diagnosis

A somatization disorder should be suspected in any individual with a vague or complicated medical history or unaccountable non-responsiveness to therapy. Patients with somatization disorder may not have or may deny emotional symptoms or conflicts, so the absence of significant emotional symptoms at the general psychiatric interview and history taking will not exclude the diagnosis. But the presence of a previous or current emotional disturbance does support the diagnosis, as do previous episodes of medically unexplained bodily symptoms. Taken at face value, the physical symptoms are only of modest diagnostic importance, whereas unspecific or atypical symptoms in several bodily systems, or a very unusual presentation, speak in favour of the diagnosis. Multiple fluctuating symptoms of obscure origin, and their onset before the age of 30, strongly support the diagnosis. The diagnostic criteria displayed in Table 5.2.3.1 may be used, or the criteria listed in the DSM-IV or ICD-10 diagnostic criteria for somatoform disorders may be used.

Differential diagnosis

Mental and somatoform disorders

In **malingering**, the patient feigns illness with a conscious motivation to avoid responsibility or to gain an advantage. In **factitious disorder**, the symptoms are intentionally produced and the patient may self-inflict or induce diseases and lesions. In contrast to malingering, there is no external incentive for producing the symptom(s), and the motive is unconscious and only understandable in a psychopathological context.[19, 20] In somatoform disorders, both the symptom-producing behaviour and the motive are believed to be unconscious. However, factitious or malingering symptoms, mixed with other non-intentional symptoms, may occur in somatization disorder and related disorders.[11]

Hypochondriasis or health anxiety is mainly defined in cognitive terms with the emphasis on a preoccupation with physical appearance or the fears of harbouring or developing a serious physical disease. The other categories of somatoform disorders put more emphasis on bodily symptoms. In **dissociative or conversion disorder** the patients usually present fewer symptoms, but these are almost exclusively pseudoneurological symptoms. The onset is sudden, and closely associated in time with traumatic events, insoluble and intolerable problems, or disturbed relationships. The symptoms are transient and often remit suddenly after a few days, although they may persist for longer, but seldom for more than a few months. Episodes of dissociative or conversion disorders frequently occur in patients with other somatoform disorders.

In **pain disorder**, the predominating complaints are of medically unexplained pain in one or more anatomical sites. Various aches and pains are common in somatization disorder but are more fluctuating and not so dominating in the clinical presentation since they merge with other complaints.

In **ICD-10 somatoform autonomic dysfunction**, the patients complain of symptoms associated with a specific system or organ

that is largely or completely under autonomic innervation and control. For example, the patient refers the symptoms to the heart and cardiovascular system, the gastrointestinal tract, respiratory system, genitourinary system, etc.

In most **other mental disorders** physical symptoms are prominent; it is the rule, rather than the exception, that patients with mental disorders consult their family physician because of physical and not emotional symptoms.[2, 21] The symptoms may be misinterpreted by the patient and the doctor as being caused by a physical disease. However, in these cases of 'presenting' or 'facultative' somatizing, the patient will accept the diagnosis of a psychiatric disorder when it is established and will accept that the symptoms are attributable to a psychic rather than a physical affliction.[2]

Autonomic bodily symptoms are prominent in panic disorder and generalized anxiety disorders, but the emotional component of the disorders is unmistakable and the patients will not, or only temporarily, attribute their symptoms to a physical disease.

In psychoses, particular schizophrenic physical symptoms and hypochondriacal belief are common. However, the complaints are of a paranoid quality and other psychotic symptoms are prominent, except in the case of hypochondriacal paranoia. Psychotic episodes may occur in somatization disorder, but these are usually transient.[22]

In depressive disorders, the patient's complaints are mood congruent, and the symptoms will disappear if the depression is treated. In a few cases, however, a depression may be expressed in mainly bodily complaints (i.e. masked depression), and if left untreated may result in a prolonged course.

In obsessive-compulsive disorder (OCD), the patient may fear contracting a disease from an outside source (e.g. dirt, germs, viruses, etc.). The OCD disorders share many similarities with hypochondriasis, i.e. cognitive and emotional complaints are the main problem, whereas bodily symptoms seldom are dominating.

Medical conditions

The onset of multiple physical symptoms after the age of 40, in a previously physically and mentally healthy individual, suggests a general medical condition. However, multiple vague physical complaints may be prominent in mental disorders with late onset such as mood disorders, dementia, and withdrawal states.

Only a limited number of general medical conditions present vague, non-specific, and multiple somatic symptoms (e.g. hyperparathyroidism, hyperthyroidism, acute intermittent porphyria, myasthenia gravis, AIDS, multiple sclerosis, systemic lupus erythematosus, lyme disease, and connective tissues disease). In most of these medical conditions there will be positive paraclinical findings.

In genuine physical disorders, the key symptoms will usually be characteristic from one patient to another and across illness episodes. In contrast, the constellation of symptoms in somatization disorder and related disorders will usually be incompatible with any known, authentic physical disease, and the symptom picture will be of a more fluctuating nature.

Epidemiology

The prevalence of somatization disorder is about 1 per cent in the general population and 1 to 6 per cent in primary care and in inpatient medical settings, but it is much higher if less restrictive, abridged criteria are used.[3, 7, 8, 21, 23–28] Both among primary care patients, newly referred neurological in and outpatients, and among internal medicine inpatients, the prevalence of severe bodily distress disorder is 3.3 per cent, whereas the prevalence of modest bodily distress is 25.3 per cent.[3] Regardless of the criteria used, females predominate with a male-to-female ratio of 1:2 to 1:6, and the prevalence is rather constant as to age.[21, 24–29]

The age of onset is usually before 30 to 40 years.[14] This patient group poses a considerable financial burden to health and social service provision, a loss to industry, and a dependence on invalidity benefits for long periods.[14, 24, 30, 31]

Somatization disorder and related disorders are associated with a wide spectrum of heterogeneous psychiatric disorders, including personality disorders and mental retardation. The comorbidity rates are highest in the most chronic cases, in which rates of 50 per cent or higher are reported.[3, 13, 22, 24] Cultural and ethnic factors may affect the prevalence rate. This may be due to the influence of these factors on the likelihood of presentation to health care rather than real prevalence differences in the general population.[7, 21, 23] The typical presentation of physical complaints has varied throughout history and with the sociocultural environment of the patient.[32]

Aetiology

The aetiology of somatization disorder is unknown, but it is most likely multifactorial including biological, physiological, psychological, social, cultural, and iatrogenic factors. Different factors may have different importance at different times in the natural course of the illness. For instance, it may be a psychological trauma that precipitates the illness—but iatrogenic factors that maintain the illness. Whether the behavioural, cognitive, and other typical disturbances in somatization disorder and related disorders develop as a consequence of a basic biological defect in interaction with the patient's life experiences, or vice versa, or whether the mentioned disturbances and other predisposing factors have an independent impact is, as yet, unresolved.

Viewing functional somatic symptoms as a common reaction of human beings to stressors, like in the case of depressed mood and anxiety, results in the conclusion that different unspecific predisposing factors common for different mental disorders are involved.[33] Relatively specific predisposing factors are physical or sexual abuse and parental complaints of poor physical health and medically unexplained symptoms during the patient's childhood, whereas neither parental nor childhood well-defined physical illnesses seem to be predisposing factors.[34, 35]

The reported family transmission in somatization disorder may be due to sociocultural learning. However, there is some support for genetic transmission in somatization disorder, although twin studies have been inconclusive.[36]

Psychological theories

In the classical psychodynamic drive theory, medically unexplained physical symptoms are believed to develop as a reaction to the repression of unacceptable wishes or instinctual impulses and internal psychic conflicts.[37]

According to the theory of self-psychology, the anxiety connected with a threatening defragmentation or disintegration of the self is

the most profound form of anxiety that a person can experience.[37] In a defence against the feeling of emptiness, the individual becomes directed on the outside world and on physical stimuli. This process has been called 'stimulus entrapment' meaning that in somatization disorder, the individual becomes addicted to stimuli to his or her body.[38]

Individuals with alexithymia have a poorly developed language of emotions, and it has been suggested that instead they might respond with bodily symptoms. It is, however, unlikely that alexithymia has a specific aetiological role in somatization disorder.

The cognitive theory endorses the importance of the patients' misinterpretation of benign symptoms and normal physical sensations that they erroneously attribute to a physical disease.[39]

Biological factors

It is beyond doubt that an important biological component is involved, although the specific nature has still to be discovered. A neurophysiological dysfunction in the attention process has been demonstrated in somatization disorder, which may be explained by a reduced corticofugal inhibition in the diencephalon and the brainstem of afferent bodily stimuli, resulting in insufficient filtering of irrelevant bodily stimuli. A dysfunction of the secondary somatosensory area in the brain, a hypersensitivity of the limbic system towards bodily stimuli, or other dysfunctions may also be aetiologically involved.[40–42]

Other factors

In a few cases, a simple compensation claim may be of aetiological importance.

Physicians are primarily trained in a biomedical illness model and may have insufficient knowledge about diagnosing and managing somatization disorder and related disorders. Physicians may thus have a tendency to pursue organic possibilities and feel compelled to evaluate and treat all symptoms, and consequently a considerable iatrogenic reinforcement of physical symptoms is often involved.

Course and prognosis

Somatization disorder and related disorders have a spectrum of severity ranging from cases that may be difficult to delimit from normality to severely ill patients.[17]

In severe cases, the patients are chronically ill for most of their lives, but there may be periods of partial, but seldom full, remission. Some patients are able to work, others are severely disabled and are chair or bed bound, and their families have to provide virtually all aspects of physical care. Patients with somatization disorder and related disorders are often subjectively more functionally handicapped than patients who have a comparable, yet fully explained medical condition.[31]

Assessment

The purposes of the initial assessment are to (a) establish the diagnosis and rule out differential diagnoses, (b) examine which specific management or treatment strategies are possible and best for the patient, and (c) engage the patient in therapy.[15]

A scheme for the initial assessment by the psychiatric specialist is given in Table 5.2.3.3. For primary care physicians and in general hospitals this may be too comprehensive, and step one in the TERM model (Table 5.2.3.5) or another more simple model may be used.[43]

Table 5.2.3.3 Clinical assessment of patients with somatoform and related disorders including functional somatic syndromes

Before the meeting with the patient	
Review medical records and other relevant material	
At the examination	
Attitude towards the referral and the treatment	
Physical complaints	Chronology, intensity, provoking / relieving factors etc.
Triggering factors	Physical trauma or disease
	Psychosocial stressors
Current and previous emotional and behavioural complaints	
Social, funtional level, strain and coping	
The patient's illness belief and perception of symptoms	
Expectations to treatment and investigation	Physical Psychiatric
Past medical, surgical, and psychological history	
Dispositions	
Physical examination	
Paraclinical tests: Obtain focused diagnostic tests if not already done	

Before meeting the patient

Before meeting the patient it is important to review medical case notes and to gather information from other sources, e.g. the primary care physicians or the family, as patients with somatization disorder and related disorders may be inconsistent historians as a result of their complex medical history. The aim is to get an overview of the patient's medical history, the illness picture and complaints, examinations and diagnostic tests, and treatments and the outcome of these. It must also be assessed whether the patient is sufficiently examined for relevant differential diagnoses. The review may furthermore impart important information about psychological and social issues.

The examination

(a) Attitudes towards the referral and treatment

Patients with somatization disorder and related disorders may be sceptical about seeing a psychiatrist as they believe their problem to be of a physical nature and not psychiatric. This should be addressed directly by asking the patients what they have been told by the referring doctor, their reasons for coming to see the psychiatrist, and their feelings about it. It is important that such thoughts are brought to light to avoid misunderstandings and misconceptions, and to help the patients feel understood and in safe hands. In acknowledging the patients' fears of being stigmatized, it may be helpful to discuss negative public attitudes to psychiatry.

(b) Current physical complaints

The patients' physical complaints should be reviewed in detail for diagnostic purposes and to make the patients feel understood and taken seriously, which is a precondition for a good rapport between doctor and patient. Each symptom is explored to establish their characteristics, location, intensity, chronology and variation in intensity, onset, duration, and impact on daily life. Besides, provoking/relieving factors, previous treatments and the outcome of these are explored. It may be helpful together with the patient to write down the medical history on a time axis as patients may have difficulties with the chronology of their illness.

(c) Triggering factors

It is explored if onset of the illness is associated with physical trauma or diseases or with exposure to psychosocial stress or trauma.

(d) Current and previous emotional and behavioural complaints

As to a high comorbidity between somatization disorder and related disorders and other mental disorders, it is necessary to methodically clarify if the patients have a depression, anxiety disorder, or another mental illness.

(e) Illness beliefs and perception of symptoms

The patients' illness beliefs and perception of symptoms are of paramount importance for their illness and functioning as the illness behaviour has its origin in those beliefs and in illness attitudes. To change an inappropriate behaviour, it is necessary to identify such dysfunctional beliefs. Also, the attitudes and behaviour of the family may be a crucial factor in understanding the presentation and in planning the intervention.

(f) Expectations to treatments and investigations

The patients are questioned about wishes for and expectations to treatment and diagnostics tests and about what they believe may help them. Patients with somatization disorder and related disorders often have an unrealistic expectation to the impact of diagnostic tests and the effect of medical or surgical treatments. This may result in an intensive use of consultations. The patients must be helped to face the limits of what medicine can do and to acknowledge that continued medical consultations will be fruitless.

(g) Past medical, surgical, and psychological history

The patients' history is reviewed and related to the information gathered during the medical case note review. Furthermore, it is attempted to elucidate the patients' premorbid psyche and personality.

(h) Dispostions

Somatoform and related disorders often run in families and are associated with other mental disorders, and hence dispositions are explored.

(i) Physical examination

If not already made, a clinical examination ought to be carried out. Besides excluding organic possibilities, it also has a psychological purpose in making the patients feel that their physical complaints are taken seriously and that the psychiatrist is not exclusively focusing on the psychological part of the problem.

Routine laboratory test battery may be; complete blood count, electrolytes, blood urea nitrogen, creatinine, glucose, calcium, phosphate, liver function test, total protein, thyroid-stimulation hormone, erythrocyte sedimentation rate or CPR, urinalysis, and if indicated by symptoms or history, serological tests for Epstein–Barr virus, lyme disease, and immunological function test. Other tests may be relevant depending on the patient's illness picture.

Feedback of the results

Feedback of the psychiatric examination results to the referring doctor must be done in a way that is intelligible to a doctor who may not be psychologically minded. Statements like 'no formal psychiatric disorders are found', which unfortunately are frequent in consultation notes, may be somewhat useful when dealing with a patient referred for functional somatic symptoms. Such a statement just proves that the psychiatrist is unfamiliar with somatization disorder and related disorders.

The psychiatrist must be careful not to become involved in criticism of medical colleagues and in the divisions that these patients sometimes try to create between different therapists.

Treatment and management

Evidence-based treatment

Many different therapies have been used in patients with somatization disorder and related disorders like family therapy, physical therapy, biofeedback, relaxation therapies, hypnotherapy, psychodynamic psychotherapy, cognitive-behavioural therapy etc. The focus in the management varies a lot from (a) focus on the patients (organ-oriented approach or cognitive interpersonal approach, i.e. pattern of bodily and emotional symptoms over time, focus on dysfunction of central processing and context factors, interventions aimed at sensations, cognitions, affects, behaviours, and restoring overall function), (b) focus on the doctor (early recognition, communication skills, avoidance of iatrogenic harm), and (c) focus on context factors (doctor reimbursement system, patient compensation schemes, health care system, workplace characteristics, cultural belief).[44]

It must be concluded that the evidence for an effective treatment of patients with multiple functional symptoms is unsystematic. The use of multiple treatment methods and outcome measurement makes it difficult to compare studies. However, there seems to be substantial evidence that a specialized assessment with discharge letter, CBT, brief psychodynamic psychotherapy, and antidepressants have some effect on one or more outcome parameters.[44–47]

Treatment setting

Functional somatic symptoms and functional somatic syndromes are common in all medical settings. The patients often believe that they have a physical and not a psychological problem and will primarily seek medical and not psychological care. Hence, the management of somatization disorder and related disorders is not only an issue for psychiatrists but for all settings within the health care system.

The management must follow a stepped-care model, in which it is defined at which level of specialization each patient should be treated and, for each step, who is responsible for which parts of the treatment. For example, the mild and uncomplicated cases are treated by the primary care physician, modest to severe cases mainly by the primary care physician but in collaboration with a specialist, whereas the severe and complicated cases are managed in specialized care.

Besides the severity of the disorder, defining the steps in the model includes considerations about feasibility both as to available treatment resources and what is acceptable to the patient and the skills and knowledge of the primary care physician. Patients with a chronic somatization disorder may be well cared for by their primary care physician, provided the latter has the necessary skills and knowledge.

Inpatient care may be appropriate in a few cases, but patients with somatization disorder and related disorders are difficult to treat in ordinary psychiatric wards, where they are often met with considerable resistance from the psychiatric staff. Specialized inpatient units only exist in a few places in the world, and this treatment has not been documented.

Non-specialiszed treatment and management

(a) General hospital departments and non-psychiatric specialists

Because of the prevalence and the risk of iatrogenic harm, it is important that non-psychiatric specialists know about somatoform and related disorders and know how to identify and diagnose them. If the assessing physician exclusively focuses on symptoms that he finds relevant for his own specialty, there is a great risk of pursuing a wrong diagnosis. The fear of overlooking a definite physical disease, as an explanation of the physical symptoms, is deeply rooted in doctors, and this may, together with a poor knowledge about somatoform disorders, result in the doctor attempting to rule out even the rarest physical causes before a somatization disorder is even considered. However, there is little evidence that important medical diagnoses are missed more often in patients with somatoform disorders than in patients with other disorders.[48–50] Unnecessary procedures and diagnostic tests are not only unpleasant and potentially risky for the patient but may also delay or hinder sufficient treatment resulting in an aggravation of the disorder or perhaps chronicity. However, it must always be borne in mind that patients with a somatoform disorder may also have or acquire a concurrent physical disease. Instead of viewing somatization disorder and related disorders as a diagnosis of exclusion, all diagnostic possibilities ought to be included in the diagnostic consideration and the examination plan from the initial contact in the same way as when it is a question about two well-defined organic diseases. Diagnostic tests should be on medical indication and not on patient demand.

The primary role of the non-psychiatric specialists in the treatment of somatization disorder and related disorders is to:

- Exclude physical disease or trauma that can be treated medically or surgically
- In an empathic way, make it clear to the patient that he or she does not have the physical disease he or she fears, and that there is no indication of any other physical disease or defect that needs medical attention.
- That there is no medical indication for further diagnostic tests or examinations
- Coordinate the management with the primary care physician and other doctors that the patient may be in contact with.
- Consider a referral to a psychiatrist for examination or treatment
- In chronic cases, follow the advice given in Table 5.2.3.4.

Table 5.2.3.4 General advice on the non-specialized management of chronic somatization

Physical
1. Make a brief physical examination focusing on the organ system from which the patient has (new) complaints. - Look for signs of disease instead of symptoms. - Avoid tests and procedures unless indicated by signs of disease or a well-defined (new) clinical illness picture.
2. Reduce unnecessary drugs. Do not use on demand prescriptions and avoid dependence-forming medication.
Psychological
3. Make the diagnosis and inform the patient that the disorder is known and has a name.
4. Acknowledge the reality of the symptoms.
5. Be direct and honest with the patient about the areas you agree on, those you do not agree on, but be careful as not to make the patient feel ignorant or not respected.
6. Be stoical; do not expect rapid change or cure.
7. Reduce expectations of cure and accept the patient as being chronically ill. Aim at containment and (iatrogenic) damage limitation, i.e. use the management rather than treatment.
8. Perceive worsening of/or new symptoms as emotional communication rather than as a manifestation of a new disease.
9. Apply a specific therapeutic technique if you master it and consider referral to specialist treatment.
Psychopharmacological treatment
10. Consider treatment with psychoactive mediation (primarily antidepressant).
11. Choose non-habit forming medication and, if possible, choose medication that can be serum monitored.
12. Start with a smaller dosage than usual and increase slowly. Be stoical about side effects.
13. Take regular serum values to compliance issued and for validating complaints of adverse effects.
14. Treat any co-existing psychiatric disorders according to usual guidelines.
Administrative
15. Be proactive rather than reactive. Agree on a course with fixed, scheduled appointments with 2–6-weeks intervals and avoid consultations on patient demand (if needed, accept on demand a maximum of 1 phone appointment per week).
16. If the patient has a job, avoid giving him sick leave if at all possible.
17. Try to become the patient's only physician and minimize the patient's contact to other health care professionals, doctors on call, and alternative therapists.
18. Inform your colleagues of your management plans and develop contingency plans for when you are not accessible.
19. Inform the patient's nearest relative and try to co-opt a relative as a therapeutic ally.
20. If necessary, arrange support/supervision for yourself.
21. If necessary, motivate the patient to receive psychiatric treatment.

As previously mentioned, patients with somatization disorder are often resistant to psychiatric referral as they believe they have a physical and not a psychological problem. They may construe the referral as a sign that the physicians are not taking their symptoms seriously. It is important that the referring physicians avoid giving the message that the patients are not genuinely ill, that they trouble the doctors unnecessarily, or that they are 'mad'.[51] Instead, the

physician must try to meet the patients' wishes about knowing the cause of their illness, being taken seriously, getting explanations, information, advice, and reassurance.[52] A close liaison with medical colleagues guiding them in making a psychiatric referral in an acceptable way is important for engaging the patient in treatment.

Primary care

The TERM model (the extend reattribution model) is a simple cognitive-behavioural orientated treatment method to improve the primary care physicians' detection and management of patients presenting with medically unexplained symptoms. The method can effectively be taught to primary care physicians and will improve the outcome of their patients' treatment.[43] The TERM model is one of several different models that have been developed on the basis of the original reattribution model.[53,54]

The first stage of the TERM model (Table 5.2.3.5) is called 'understanding', as the important point, besides assessment of the patient, is that the patient feels understood and taken seriously by the doctor. The second stage is called 'the physician's expertise and acknowledgement of illness', in which the physician feeds back the results of his examination, but at the same time acknowledges the reality of the symptoms. The third stage is called 'reframing', in which a new model of understanding of the patient's problem is negotiated between patient and doctor. As a fourth stage, the model includes techniques for negotiating further treatment.[43]

Finally, the model includes principles for management of chronic, somatizing patients (Table 5.2.3.4). In chronic cases, damage limitation is a more realistic therapeutic goal than cure, and management is thus a more realistic aim than treatment.[51,55,56] The main aim is to stop the pathological cycle of interventions and consultations and the consequential somatic 'overtreatment' (i.e. treatment on obscure indication), and then, if possible, gradually to motivate the patient to accept specialized care if available in the area. Management according to these principles has shown to be effective in randomized controlled studies and should therefore always be implemented either solely or combined with one of the treatments described below.[57]

Table 5.2.3.5 The TERM model

1. **Understanding**
Take a full history of the symptoms
Explore emotional clues
Inquire directly about symptoms of anxiety and depression
Explore life events, stress, and other external factors
Explore functional level
Explore the patient's health beliefs
Explore the patient's expectations to treatment
Make a brief, focused physical examination
2. **The physician's expertise and acknowledgement of illness**
Provide feedback on the results of the physical examination
Acknowledge the reality of the symptoms
Make it clear that there is no (or that there is indeed) indication for further examination of nonpsychiatric treatment
3. **Negotiating a new model of understanding (reframing)**
Simple explanations
Physical symptoms are common reactions to, for example, stress and strain/nervousness
Depression lowers the threshold of pain
Muscular tension in anxiety and nervousness causes pain
Demonstrations
Practical (hyperventilation, muscular tension)
Establish the association between physical discomfort, emotional reactions, and life events
"Here and now" (nervous about consulting the physician)
Severe cases
Known phenomenon with a name: somatization
Basically the cause is unknown, but nothing indicates a hidden physical disease
Biological explanation: some are bodily sensitive than others, which explains their more intense symptoms
Individual symptom coping and reactions determine one's future well-being
4. **Negotiating further treatment**
Sum up agreements made during the consultation
Agree on specific objectives, contents, and form for the future course
Acute cases: no further appointments
Subacute cases: therapy sessions, regular scheduled appointments
Chronic: consider status consultation, regular scheduled appointments
Consider referral to psychiatrist, psychologist, or specialist service
5. **Chronic cases**
See Table 5.2.3.4

Specialized treatment and management

Cognitive-behavioural therapy is the best documented and most widely used therapy and is hence the focus in this section along with pharmacotherapy. The general principles of CBT are described elsewhere, so this section concentrates on the general techniques used in somatization disorder and related disorders, and how they differ from the CBT techniques applied in other disorders.

(a) Goal setting

Early in the therapy, the goals for the therapy are established. It is important to set up goals that are realistic in the light of the patient's illness and the framework of the therapy.

(b) Engagement and motivation

Treatment of somatization disorder and related disorders differs from the treatment of other mental disorders on several points, one being that it is very important to work systematically by engaging the patients in therapy. As the patients believe they have a medical condition, they may have very low belief in psychological treatment. It may be helpful to discuss the idea that all illnesses have an emotional component and that a psychological treatment focusing on the emotional component is often helpful in reducing suffering. The motivation and engagement should preferably be established during the assessment interview.

(c) Psychoeducation

The patient needs to be taught about somatoform and related disorders and about the body's normal reactions to stress and how stress may be expressed in physical symptoms. It is important that the patient learns about the possible biological and physiological basis in the CNS for the symptoms. This will make the reality of the symptoms clear to the patient and emphasize that the illness is not imaginary or made up. The education may be supported by written information. Parts of the information ought to be repeated during the therapy when it, in a relevant way, can be linked to the patient's personal experience.

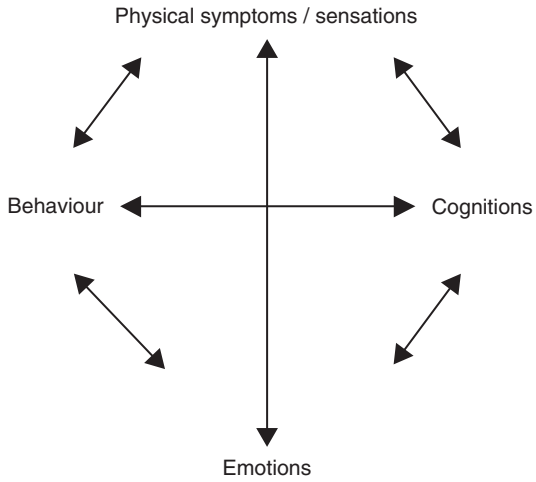


Fig. 5.2.3.1 Basic model in cognitive behavioural therapy.

(d) Physical symptoms and symptom attribution

Figure 5.2.3.1 illustrates the relationship between symptoms, cognitions, emotions, and behaviour in the cognitive-behavioural model. The first step in the therapy is to clarify the patient's dysfunctional automatic thoughts and basic beliefs about illness and symptoms, i.e. cause, consequences for future health and function, treatment, etc. Those thoughts are related to feelings and illness behaviour.

In the next step, the patient's disease model and symptom attributions are challenged. The therapist asks the patient to consider alternative possibilities, in which process even the most unlikely explanations are welcome. The therapist and the patient then explore which explanation is the most likely by investigating pros and cons for each possible explanation.

Finally, it is clarified how the patient would behave and react to a particular understanding or belief and which feelings this would produce.

(e) Behaviour and coping

Even if the patient does not want to or cannot work with his or her illness perception and symptom attribution, it would still be possible to work with the way the patient copes with illness. Illness behaviour and coping with symptoms and illness are scrutinized, and the behaviour is linked to the underlying thoughts and feelings if possible. When an example of a behaviour or coping strategy is elucidated, it is listed with arguments for and against a particular behaviour or coping strategy. A brainstorm on alternative possibilities is advisable, and pros and cons for each of these are also scrutinized. In a negotiation with the therapist, the patient chooses an alternative possibility, which is tested by the patient as homework, and at next session the effect of this is explored.

It is important to go slowly for the patient to experience success. The therapist is obliged to make sure that the patient sets up realistic goals with a good chance of success.

(f) Links between symptoms and stressors

The cause of somatization disorder and related disorder may be viewed as a combination of personal vulnerability and the stress and strain an individual is exposed to. The patients are often unaware of their patterns of reaction, but this can be established by careful registration of variations in symptom intensity and then relating them to what the patients are doing or thinking at that time. Based on variation in symptom intensity and stressors, potential important stressors in everyday life may be identified. This may provide a focus for intervention.

(g) Family and social network

The patient's health beliefs are often shared with the family and social network. Therefore it is necessary to create an alliance with the family to make sure they support the patient and not counteract the therapy. Family members are invited to a consultation and informed about the nature of the disorder and about the planned treatment. Misunderstandings and prejudgements about somatization disorder are eliminated.

(h) Treatment and help seeking behaviour and the physicians' handling

For patients with somatization disorder, social and family life is often centered on their illness, and an objective of the therapy is to reduce the importance of the illness and try to build up other interests and aspects of life. The patient's consulting behaviour may result in multiple fruitless diagnostic tests, referrals, and treatment attempts, which expose the patient to iatrogenic harm. Often, the patient grows tired of the doctors responding to their questions merely by referring or prescribing medication. In therapy, the patients are taught how to present their problems to the physicians in a way that prevents referrals and medication.[58] A patient may for instance tell the doctor that he or she is worried about some new symptoms, because he or she does not know whether the symptoms are just part of his or her somatoform disorder or something else, and that the doctor's expertise is required in order to clarify this.

Medication

Coexisting mood, anxiety, or other mental disorders are as effectively treated with psychotropic medication in patients with a somatoform disorder as in patients without.[59] Patients without coexisting mental disorders also seem to benefit from psychopharmacological treatment, antidepressants being the first choice.[44] In some cases, medication with peripheral action may be helpful as symptomatic treatment, for example in case of gastrointestinal symptoms like IBS.[44] Tricyclic antidepressants seem most effective, but due to side effects SSRI or SNRI ought to be the first choice. A useful strategy in the psychopharmacological treatment of somatizing patients is to start on a lower dosage than is usually recommended and to increase the dosage only gradually in order to avoid side effects. Antidepressants with the fewest interactions should be chosen as polypharmacy is common in these patients. The use of benzodiazepines and other dependence-producing drugs should be avoided. Stronger painkillers usually only produce a partial and temporary improvement, but they may result in misuse and should be avoided. Mild painkillers and tricyclic antidepressants are to be preferred.

In general, a degree of stoicism is required on the part of the clinician as patients may have varying symptom intensity and motivation and because patients may complain of side effects as a result of increased sensitivity to bodily sensations. Drugs that can be serum-monitored are preferred so that compliance and the likelihood of side effects can be assessed.

Possibilities for prevention

Patients with medically unexplained symptoms are often unpopular both with general psychiatrists and other doctors. The patients are seen as neither physically nor mentally ill, but simply as individuals complaining in order to avoid normal life responsibilities. Overcoming these negative attitudes is a matter of proper training of doctors and education of medical students.

Since an early diagnosis is important for preventing physical fixation, iatrogenic harm, and chronicity, doctors must be taught to view somatization disorder and related disorders with the same seriousness as well-defined physical diseases. The dominating dualistic way of thinking in medicine must be counteracted.

Further information

Mayou, R., Sharpe, M., and Carson, A.J. (eds.) (2004). *ABC of psychological medicine*. BMJ Books, London.

Woolfolk, R.L. and Allen, L.A. (2006). *Treating somatization. A cognitive-behavioral approach*. The Guilford Press.

Bass, C. (1990). *Somatization: physical symptoms & psychological illness* (1st edn). Blackwell Scientific Publications Oxford.

Mayou, R., Bass, C., and Sharpe, M. (1995). *Treatment of functional somatic symptoms.* Oxford University Press, Oxford.

References

1. Goldberg, D.P. and Bridges, K. (1988). Somatic presentations of psychiatric illness in primary care setting. *Journal of Psychosomatic Research*, **32**(2), 137–44.
2. Simon, G., Gater, R., Kisely, S., *et al.* (1996). Somatic symptoms of distress: an international primary care study. *Psychosomatic Medicine*, **58**(5), 481–8.
3. Fink, P., Toft, T., Hansen, M.S., *et al.* (2007). Symptoms and syndromes of bodily distress: an exploratory study of 978 internal medical, neurological, and primary care patients. *Psychosomatic Medicine*, **69**(1), 30–9.
4. Katon, W., Lin, E., Von Korff, M., *et al.* (1991). Somatization: a spectrum of severity. *American Journal of Psychiatry*, **148**(1), 34–40.
5. Perley, M.J. and Guze, S.B. (1962). Hysteria-the stability and usefulness of clinical criteria. A quantitative study based on a follow-up period of six to eight years in 39 patients. *The New England Journal of Medicine*, **266**, 421–6.
6. Murphy, M.R. (1990). Classification of the somatoform disorders. In *Somatization: physical symptoms & psychological illness* (ed. C. Bass), pp. 10–39. Blackwell Scientific Publications, Oxford.
7. Escobar, J.I., Rubio-Stipec, M., Canino, G., *et al.* (March 1989). Somatic symptom index (SSI): a new and abridged somatization construct. Prevalence and epidemiological correlates in two large community samples. *The Journal of Nervous and Mental Disease*, **177**(3), 140–6.
8. Kroenke, K., Spitzer, R.L., deGruy, F.V., *et al.* (1997). Multisomatoform disorder. An alternative to undifferentiated somatoform disorder for the somatizing patient in primary care. *Archives of General Psychiatry*, **54**(4), 352–8.
9. Kirmayer, L.J., Groleau, D., Looper, K.J., *et al.* (2004). Explaining medically unexplained symptoms. *Canadian Journal of Psychiatry*, **49**(10), 663–72.
10. Fink, P. (1992). Surgery and medical treatment in persistent somatizing patients. *Journal of Psychosomatic Research*, **36**, 439–47.
11. Fink, P. (1992). Physical complaints and symptoms of somatizing patients. *Journal of Psychosomatic Research*, **36**, 125–36.
12. Salmon, P., Peters, S., and Stanley, I. (1999). Patients' perceptions of medical explanations for somatisation disorders: qualitative analysis. *British Medical Journal*, **318**(7180), 372–6.
13. Stern, J., Murphy, M., and Bass, C. (1993). Personality disorders in patients with somatisation disorder. A controlled study. *British Journal of Psychiatry*, **163**, 785–9.
14. Fink, P. (1992). The use of hospitalizations by persistent somatizing patients. *Psychological Medicine*, **22**, 173–80.
15. House, A. (1995). The patient with medically unexplained symptoms: making the initial psychiatric contact. In *Treatment of functional somatic symptoms* (eds. R. Mayou, C. Bass, and M. Sharpe), pp. 89–102. Oxford University Press, Oxford.
16. Yutzy, S.H., Cloninger, C.R., Guze, S.B., *et al.* (1995). DSM-IV field trial: testing a new proposal for somatization disorder. *The American Journal of Psychiatry*, **152**(1), 97–101.
17. Fink, P., Rosendal, M., and Olesen, F. (2005). Classification of somatization and functional somatic symptoms in primary care. *Australian and New Zealand Journal of Psychiatry*, **39**(9), 772–81.
18. Wiesmuller, G.A., Ebel, H., Hornberg, C., *et al.* (2003). Are syndromes in environmental medicine variants of somatoform disorders? *Medical Hypotheses*, **61**(4), 419–30.
19. Feldman, M.D. and Ford, C.V. (1994). *Patient or pretender. Inside the strange world of factitious disorders*. John Wiley & Sons, Inc, NY, USA.
20. Eisendrath, S.J. (1984). Factitious illness: A clarification. *Psychosomatics*, **25**, 110–17.
21. Üstün, T.B. and Sartorius, N. (1995). *Mental illness in general health care, an international study*. John Wiley & Sons, Chichester, UK.
22. Fink, P. (1995). Psychiatric illness in patients with persistent somatisation. *British Journal of Psychiatry*, **166**(1), 93–9.
23. Gureje, O., Simon, G.E., Ustun, T.B., *et al.* (1997). Somatization in cross-cultural perspective: a World Health Organization study in primary care. *The American Journal of Psychiatry*, **154**(7), 989–95.
24. Creed, F. and Barsky, A. (2004). A systematic review of the epidemiology of somatisation disorder and hypochondriasis. *Journal of Psychosomatic Research*, **56**(4), 391–408.
25. Fink, P., Steen, H.M., and Sondergaard, L. (2005). Somatoform disorders among first-time referrals to a neurology service. *Psychosomatics*, **46**(6), 540–8.
26. Fink, P., Hansen, M.S., and Oxhoj, M.L. (2004). The prevalence of somatoform disorders among internal medical inpatients. *Journal of Psychosomatic Research*, **56**(4), 413–18.
27. Toft, T., Fink, P., Oernboel, E., *et al.* (2005). Mental disorders in primary care: prevalence and co-morbidity among disorders. Results from the functional illness in primary care (FIP) study. *Psychological Medicine*, **35**(8), 1175–84.
28. De Waal, M.W., Arnold, I.A., Eekhof, J.A., *et al.* (2004). Somatoform disorders in general practice: prevalence, functional impairment and comorbidity with anxiety and depressive disorders. *British Journal of Psychiatry*, **184**, 470–6.
29. Kroenke, K. and Spitzer, R.L. (1998). Gender differences in the reporting of physical and somatoform symptoms. *Psychosomatic Medicine*, **60**(2), 150–5.
30. Barsky, A.J., Orav, E.J., and Bates, D.W. (2005). Somatization increases medical utilization and costs independent of psychiatric and medical comorbidity. *Archives of General Psychiatry*, **62**(8), 903–10.
31. Hansen, M.S., Fink, P., Frydenberg, M., *et al.* (2002). Use of health services, mental illness, and self-rated disability and health in medical inpatients. *Psychosomatic Medicine*, **64**(4), 668–75.
32. Shorter, E. (1992). *From paralysis to fatigue. A history of psychosomatic illness in the modern era*. The Free Press, Macmillan Inc. New York.
33. Mayou, R., Bass, C., and Sharpe, M. (1995). Overview of epidemiology, classification, and aetiology. In *Treatment of functional somatic symptoms* (eds. R. Mayou, C. Bass, and M. Sharpe), pp. 42–65. Oxford University Press, Oxford.
34. Hotopf, M., Carr, S., Mayou, R., *et al.* (1998). Why do children have chronic abdominal pain, and what happens to them when they grow

up? Population based cohort study. *British Medical Journal*, **316**(7139), 1196–200.
35. Leserman, J., Drossman, D.A., Li, Z., *et al.* (1996). Sexual and physical abuse history in gastroenterology practice: how types of abuse impact health status. *Psychosomatic Medicine*, **58**(1), 4–15.
36. Guze, S.B. (1993). Genetics of Briquet's syndrome and somatization disorder. A review of family, adoption, and twin studies. *Annals of Clinical Psychiatry*, **5**, 225–30.
37. Rodin, G.M. (1991). Somatization: a perspective from self psychology. *Journal of American Academy Psychoanalysis*, **19**, 367–84.
38. Meares, R. (1997). Stimulus entrapment: on a common basis of somatization. *Psychoanalytic Inquiry*, **17**(2), 223–34.
39. Sharpe, M. (1995). Cognitive behavioural therapies in the treatment of functional somatic symptoms. In *Treatment of functional somatic symptoms* (eds. R. Mayou, C. Bass, and M. Sharpe), pp. 122–43. Oxford University Press, Oxford.
40. Miller, L. (1984). Neuropsychological concepts of somatoform disorders. *International Journal of Psychiatry in Medicine*, **14**(1), 31–46.
41. Rief, W. and Barsky, A.J. (2005). Psychobiological perspectives on somatoform disorders. *Psychoneuroendocrinology*, **30**, (10), 996–1002.
42. Mertz, H. (2002). Role of the brain and sensory pathways in gastrointestinal sensory disorders in humans. *Gut*, **51**(Suppl. 1), i29–i33.
43. Fink, P., Rosendal, M., and Toft,T. (2002). Assessment and treatment of functional disorders in general practice: the extended reattribution and management model—an advanced educational program for nonpsychiatric doctors. *Psychosomatics*, **43**(2), 93–131.
44. Henningsen, P., Zipfel, S., and Herzog, W. (2007). Management of functional somatic syndromes. *Lancet*, **369**(9565), 946–55.
45. O'Malley, P.G., Jackson, J.L., Santoro, J., *et al.* (1999). Antidepressant therapy for unexplained symptoms and symptom syndromes. *Journal of Family Practice*, **48**(12), 980–90.
46. Kroenke, K. (2007). Efficacy of treatment for somatoform disorders. A review of randomized controlled trials. *Psychosomatic Medicine*, in press.
47. Kroenke, K. and Swindle, R. (2000). Cognitive-behavioural therapy for somatization and symptom syndromes: a critical review of controlled clinical trials. *Psychothererapy and Psychosomatics*, **69**(4), 205–15.
48. Fink, P. (1997). *Persistent somatization*. Thesis; faculty of health sciences. University of Aarhus, Denmark.
49. Crimlisk, H.L., Bhatia, K., Cope, H., *et al.* (1998). Slater revisited: 6 year follow up study of patients with medically unexplained motor symptoms. *British Medical Journal*, **316**(7131), 582–6.
50. Stone, J., Wojcik, W., Durrance, D., *et al.* (2002). What should we say to patients with symptoms unexplained by disease? The 'number needed to offend'. *British Medical Journal*, **325**(7378), 1449–50.
51. Bass, C. (1990). Assessment and management of patients with functional somatic symptoms. In *Somatization: physical symptoms & psychological illness* (ed. C. Bass), pp. 40–72. Blackwell Scientific Publications, Oxford.
52. Price, J. and Leaver, L. (2002). ABC of psychological medicine: beginning treatment. *British Medical Journal*, **325**(7354), 33–5.
53. Gask, L. (1995). Management in primary care. In *Treatment of functional somatic symptoms* (eds. R. Mayou, C. Bass, and M. Sharpe), pp. 391–409. Oxford University Press, Oxford.
54. Morriss, R., Gask, L., Ronalds, C., *et al.* (1998). Cost-effectiveness of a new treatment for somatized mental disorder taught to GPs. *Family Practice*, **15**(2), 119–25.
55. Smith, G.R. Jr. (1995). Treatment of patients with multiple symptoms. In *Treatment of functional somatic symptoms* (eds. R. Mayou, C. Bass, and M. Sharpe), pp.175–87. Oxford University Press, Oxford.
56. Bass, C., Sharpe, M., and Mayou, R. (1995). The management of patients with functional somatic symptoms in the general hospital. In *Treatment of functional somatic symptoms* (eds. R. Mayou. C. Bass, and S. Sharpe), pp. 410–27. Oxford University Press, Oxford.
57. Smith, G.R. Jr., Rost, K., and Kashner, T.M. (1995). A trial of the effect of a standardized psychiatric consultation on health outcomes and costs in somatizing patients. *Archives of General Psychiatry*, **52**(3), 238–43.
58. Salmon, P., Humphris, G.M., Ring, A., *et al.* (2006). Why do primary care physicians propose medical care to patients with medically unexplained symptoms? A new method of sequence analysis to test theories of patient pressure. *Psychosomatic Medicine*, **68**(4), 570–7.
59. Katon, W. and Sullivan, M. (1995). Antidepressant treatment of functional somatic symptoms. In *Treatment of functional somatic symptoms* (eds. R. Mayou, C. Bass, and M. Sharpe), pp. 103–21. Oxford University Press, Oxford.

5.2.4 Conversion and dissociation disorders

Christopher Bass

Introduction

Of all the disorders characterized by symptoms in the absence of disease, conversion disorders are perhaps the most difficult to explain. How, for example, can one explain functional blindness or a loss of function of both legs in the absence of conspicuous organic disease? The ancient Greeks recognized that if we suffer emotional disturbance as a result of some serious stress (such as personal injury or bereavement), this causes a change in the nervous system which leads in turn to symptoms in different parts of the body according to the underlying pathophysiology. Nineteenth century neurologists made significant advances when they identified specific ideas at the root of the symptoms. In the early nineteenth century Collie[1] also observed that the significance of, and attention to, a symptom or set of symptoms may depend more on what they mean (or their value) to the individual than on the biological underpinnings of the symptom itself.

Spence has recently argued that the problem in hysterical motor disorders is not the voluntary motor system *per se*: rather, *it is in the way that the motor system is utilized in the performance (or non-performance) of certain willed, chosen, actions.*[2] This model invokes a consciousness that acts upon the body and the world. By contrast, the psychodynamic ('conversion') model, which Freud introduced and which held sway for most of the twentieth century, invokes an unconscious mechanism 'acting' independently of consciousness, to interfere with voluntary movement. Spence has further argued that hysterical paralyses are maintained not by unconscious mechanisms, but by conscious processes. The maintenance of these symptoms requires the patient's attention, a characteristic of higher motor acts; the paralyses break down when the subject is distracted, consciousness is obtunded, or when it (the 'paralyses') is circumvented by reflexive motor routines. Hysterical paralyses, Spence avers, are quintessentially disorders of action (or inactions), which the patient disavows, when faced with some overwhelming situation, which threatens the identity of the self.[2]

One regrettable development of psychiatry's adoption of Freudian theory was the fracture in communication between the disciplines of psychiatry and neurology, which has only recently been restored by the sort of collaborative research currently being carried out by neurologists and psychiatrists.[3] In the last decade there have also been exciting advances in neuroimaging, which have stimulated research into the neurophysiology of hysteria, and these will be described later. This chapter will also emphasize contemporary approaches to management of these difficult clinical problems.

Problems with definition

There are a number of problems with the definition of the conversion disorders (CD). First, physical disorder must be excluded, but neurological co-morbidity is known to be high in patients with CD,[4] and distinguishing which symptoms are accounted for by organic disease and which are not can be difficult. Second, it is stated that a temporal association between a psychological stressor and the onset on the disorder should be identified, but in practice this is often impossible to establish and depends to a large extent on the skill of the interviewing doctor. Finally, by definition (according to the glossaries ICD-10 and DSM-IV)[5, 6]; the process should be unconsciously mediated, but it is difficult (some would say impossible) to distinguish between symptoms that are not consciously produced and those that are intentionally manufactured. The DSM-IV provides no criteria to distinguish conscious from unconscious intent, and many authors have argued that the criteria for whether the patients are consciously aware of producing these symptoms should be dropped from the diagnosis of CD.[7]

In clinical practice it is often difficult for a physician, faced with a patient in a hospital bed unable to use his or her legs despite normal tests and clinical findings, to differentiate between conversion disorder, factitious or fabricated disorder, or frank malingering. What the clinician is being asked to do is to determine whether or not the symptoms are being produced intentionally or not; and what the motives are. Table 5.2.4.1 attempts to provide a framework, but it highlights the shortcomings of psychiatric glossaries, which in turn expose the limitations of the medical model, which forces doctors to place patients in categories without taking into account the normal moral capacity of many individuals to exercise choice and determine (at least to some extent) their actions.[8] These medical conundrums have been explored in more detail in the chapter on factitious disorders and malingering (Chapter 5.2.9).

The role of volition

Central to recent debates about hysteria and conversion disorders is the extent to which a person's illness presentation is considered a product of free will and hence social deviance or the result of psychopathology and/or psychosocial influences beyond the volitional control of the subject.[9] The proposal that voluntary processes are involved in some way has a very long history: something prevents a specific voluntary behaviour from being executed through a 'negative', lack of movement (as in paralysis), or a 'positive', abnormality of movement (as in psychogenic tremor). If 'will' is regarded as a conscious capacity that humans possess to choose what to do or refrain from doing, then the problem in CD appears to be that the will fails to produce normal action.[10] Hence, the diagnostic importance is placed on the patient's veracity: if we believe him when he says that he cannot act normally we conclude that his will is impeded pathologically; if we do not believe him we conclude instead that his will is deployed to deceive us. This is the distinction required by the diagnostic systems.

Table 5.2.4.1 Relationships between conversion hysteria, factious disorder, and malingering

	Subject insight		Target of deception		Perceived outcome	Motivation/ reason
	Aware	Unaware	Conscious self	Other		
Hysterical conversion		+	+		Sick and disabled role	Care/ dependency
Factious disorder	+			+	Sick and disabled role	Care/ dependency
Malingering	+			+	Sick and disabled role	Personal benefit, e.g. financial, avoiding prison

(Reproduced from Halligan, P. Bass, C. and Oakley, D. Wilful deception as illness behaviour. In *Malingering and illness deception* (eds. P. Halligan, C. Bass, and D. Oakley), pp. 3–28. Copyright 2003, with permission from Oxford University Press).

Conversion and dissociation

The word *conversion*is conventionally applied to somatic symptoms whereas if the symptom is psychological (e.g. a loss of memory or an external hallucination) rather than bodily (e.g. a loss of power) it is regarded as dissociative. *Dissociation* has attracted considerable recent interest, and it has been argued that the available evidence is more consistent with a model that identifies at least two distinct categories of dissociative phenomena—'detachment' and 'compartmentalization'—that have different definitions, mechanisms, and treatment implications.[11] These have been referred to as Type 1 (compartmentalization) and Type 2 (detachment), respectively (see Table 5.2.4.2).

Compartmentalization phenomena are characterized by impairment in the ability to control processes or actions that would usually be amenable to such control and which are otherwise functioning normally. This category encompasses unexplained neurological symptoms (including dissociative amnesia) and benign phenomena such as those produced by hypnotic suggestion. By contrast, *detachment phenomena* are characterized by an altered state of consciousness associated with a sense of separation from the self, the body, or the world. Depersonalization, derealization and out-of-body experiences constitute archetypal examples of detachment in this account. Evidence suggests that these phenomena are generated by a common pathophysiological mechanism involving the top-down inhibition of limbic emotional processing by frontal brain systems. Although these two types of dissociation are typically conflated, evidence suggests that different pathological mechanisms may be operating in each case.

Table 5.2.4.2 Classification of two types of pathological dissociation

Type 1 dissociation (compartmentalisation)	Type 2 dissociation (detachment)
Conversion disorders	Depersonalization/derealization
Dissociative amnesia	Peri-traumatic dissociation
Dissociative fugue	Out of body experiences
Dissociative identity disorder	Autoscopy (?)

(Reproduced with permission, from R. Brown (2002))

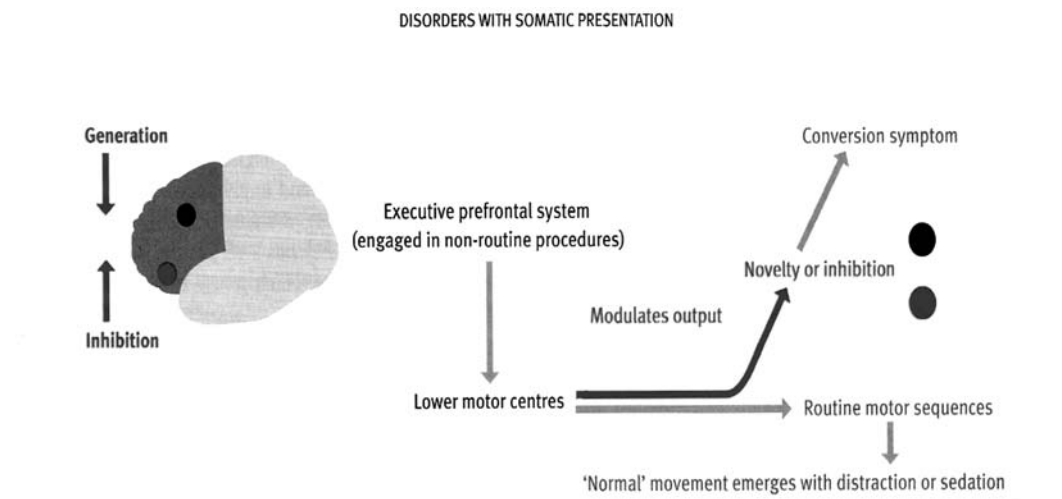
Support for the compartmentalization model comes from psychophysiological research, which suggests that psychogenic illness is associated with a deficit in attentional, conscious processing and the preservation of preattentive, preconscious processes. According to Brown[12] there is very little difference between 'negative' symptoms such as sensory loss, paralysis, etc. and 'positive' symptoms such as tremor, dystonia, etc. in terms of basic underlying mechanisms. By this view, all symptoms result from a loss of normal high-level attentional control over low-level processing systems; in this sense, all symptoms can be thought of as involving a form of compartmentalization.

Pathophysiology

There has been considerable progress in cognitive neuroscience and functional imaging over the last decade, which has provided a conceptual and empirically based platform for developing a neuroscience of not only hysterical symptoms but also free will.[13,14]

Recent functional neuroimaging data suggest that neural circuits linking volition, movement, and perception are disrupted in CD.[13] There are many studies examining the role of specific prefrontal regions in action generation (particularly the dorsolateral prefrontal and supplementary motor areas) and action suppression (especially the orbitofrontal cortices). These 'higher' executive centres supervene only when a change of behaviour is required: inappropriate behaviour must be suppressed or difficult procedures attended to, as when concentration is necessary. Hence, if the problem in hysteria is one of the will, and of abnormality emerging only when subjects attend to their actions, then this suggests the hypothesis that the prefrontal cortex is pivotal to the conversion process (see Fig. 5.2.4.1).

Further evidence that the prefrontal cortices play a key role in the control of action comes from a study of a woman with a left-sided conversion disorder affecting her leg. Marshall *et al.*[15] demonstrated that her attempt to move her paralysed leg was associated with increased activation of orbitofrontal (inhibitory) prefrontal regions, in the absence of motor cortical activity. They argued for an inhibition of motor behaviour by higher centres. Spence and colleagues[16] demonstrated that in three men with conversion symptoms affecting their upper limbs, hypokinetic movement was associated with reduced activation of dorsolateral (action-generation) areas of prefrontal cortex. Moreover, these areas of hypoactivity differed from those exhibited by four healthy men who were asked to feign the same motor impairments (see Fig. 5.2.4.2). It is possible that the application of functional neuroimaging techniques might allow clinicians to distinguish conversion from feigning on objective, empirical grounds.



The executive supervenes on lower motor centres when there is need for novel action generation (black circle) or suppression of inappropriate action (red circle), each implicating specific prefrontal regions: the dorsolateral and orbitofrontal cortices, respectively. Under conditions where the executive may be hypothesized to be disengaged (by distraction or sedation) normal movements emerge (as would be expected of routine actions). Conversion movements seem to require attention, and hence the engagement of the executive. There are 2 mechanisms by which conversion might emerge: failure to generate new actions, consequent upon dorsolateral hypofunction (black circle) or suppression of ongoing motor action, secondary to orbitofrontal activation (red circle). Other mechanisms may also operate. Acknowledgement: Mrs Jean Woodhead.

Fig. 5.2.4.1 Schematic diagram illustrating the role of prefrontal executive in modulating lower motor systems, and its hypothesized involvement in conversion disorder. (Reproduced with permission from Spence, S. (2006). Hysteria: a new look. In *Psychiatry*, **5**(2), pp. 56–60, Elsevier Ltd.)

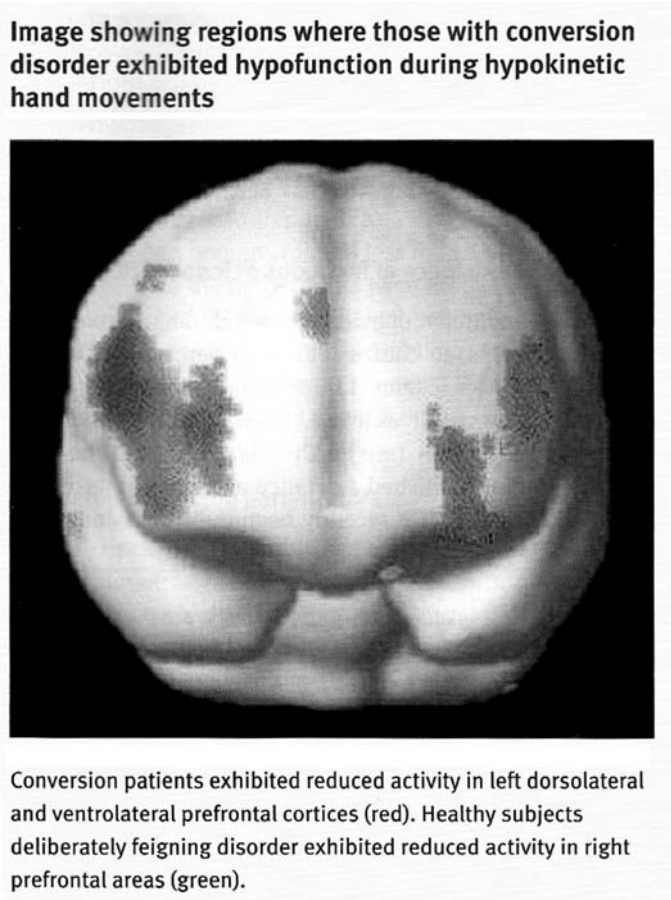


Fig. 5.2.4.2 Image showing regions where those with conversion disorder exhibited hypofunction during hypokinetic hand movements.
(Reproduced with permission (sought) from Spence, S. (2006). Hysteria: a new look. In *Psychiatry*, **5**(2), pp. 56–60, Elsevier Ltd.

In another recent case report using fMRI a patient with right-sided paralysis was asked to recall traumatic memories using a standard life event schedule: cued recall of the event was associated with regional brain activities characteristic of emotional arousal, including the amygdala and right inferior frontal lobe. Such recall was also associated with reduced motor activity in the area corresponding to the subjective paralysed limb.[17] This case study provides neuroimaging evidence for a connection between traumatic events and ongoing neurological symptoms (see Problems with definition, above).

Epidemiology

It is ironic that these research advances have occurred at a time when social historians have confidently asserted that hysteria has disappeared from clinical practice:

The most consequential development in the history of hysteria in the last century was the rapid decline in the medically recorded incidence of the disorder . . . . . .Hysteria—considered variously as a term, a theory, and a behaviour—is vanishing.[18]

It is extraordinary that this was written in 2001, at a time when symptoms considered 'functional, psychogenic, medically unexplained or hysterical' account for up to a third of new referrals to neurology outpatient departments.[3] and up to 9 per cent of admissions to a UK neurology in patient ward.[19] Akagi and House[20] concluded that the lowest prevalence figures suggested a rate of about 50 per 100 000 for cases of CD known to health services at any one time, with perhaps twice that number affected over a 1-year period. These figures suggest that hysteria is as common as other disabling conditions such as multiple sclerosis and schizophrenia. Furthermore, the burden of disability associated with chronic hysteria is far higher than a typical practising psychiatrist might expect, or than is reflected in standard textbooks of psychiatry or clinical neurology.[21]

It is regrettable therefore that it receives none of the resources or media attention that these disorders attract.

Clinical features

Conversion disorder: motor symptoms

The most typical motor symptoms are paralyses, functional weakness, gait disturbances, fits resembling epilepsy, and abnormal movements.

In the last decade diagnostic procedures have improved and the availability of non-invasive, accurate imaging has drastically reduced the rates of undetected organic pathology in patients with diagnoses of hysteria. Indeed, several recent studies have reported rates of misdiagnosis of between 0 and 4 per cent in regional and tertiary neurological centres,[22] which suggests that a diagnosis of CD can be made relatively confidently and accurately. In the following section the process of diagnosis will be briefly outlined through the history, examination, and investigation.

(a) The history

The onset, temporal sequence, and character of the presenting complaint may not be typical of a neurological disorder, and a number of other features may emerge, especially after an interview with a family member or a review of the hospital and general practitioner notes. If the patient is admitted to a general hospital bed the psychiatrist should routinely telephone the patient's primary care doctor and request a recent print out of his/her medical records (with consent). These often reveal key information about life events and/or antecedent illnesses, investigations, etc.

(i) Age of onset and sex

The average age of onset is the mid-30s, and patients with functional paralysis are less likely to be female than patients with pseudo seizures.

(ii) Mode of onset/recent life events or difficulties

An increased number of life events in the year preceding symptom onset have been recorded in small controlled studies of unexplained motor symptoms[23,24]. When patients are interviewed carefully, some report symptoms of panic just before the onset of, for example, functional weakness.[25] Judicious questions about sensations of sweating, dizziness, and difficulty breathing may reveal these somatic symptoms of anxiety, which may also be reported before the onset of sensory symptoms (see below). There is also an important literature describing unilateral somatic symptoms (which may present with sudden onset of functional weakness in a limb) following hyperventilation/panic.[26,27] These patients may present acutely in the A and E department of a general hospital, where they

may be admitted to the hospital stroke unit, or be sent to the general hospital as an emergency with an accompanying letter from the patient's GP describing the patient as 'off legs-please see and investigate.' The following clinical vignette is typical:

A 40-year-old woman was admitted as an emergency to the Stroke unit of the general hospital after collapsing at home. She had been involved in a dispute with her employers for weeks, and on the morning of the referral had an argument with her mother which made her very upset, and during the course of this she became distressed and developed paraesthesiae down the left side of her body, slurred speech, and collapsed, losing consciousness for 30 s. An ambulance was called and she was admitted to the A and E department, where she was noted to be hyperventilating and agitated, and to have weakness of the left leg, unintelligible speech, and claimed not to be able to see. All neurological investigations were normal, and after 12 days in hospital she made a gradual recovery, but was left with functional weakness in the left leg (requiring a wheelchair), and her speech was intermittently 'child-like'. History revealed a similar episode 3 years previously, a recent 2-year history of treatment for irritable bowel syndrome, and considerable work and domestic stress.

She was followed up by the liaison service and community mental health team, but despite the efforts of a community physiotherapist, psychologist, and nursing support her limb weakness continued and she remained disabled; 1 year after admission she was in receipt of disability benefits.

(iii) Previous unexplained symptoms

Evidence is accumulating that the more unexplained symptoms the patient has, the more likely the primary symptom is to be unexplained by disease.[28] In a recent study of patients with medically unexplained motor symptoms, additional unexplained symptoms including paraesthesia (65 per cent), pseudo epileptic seizures (23 per cent), and memory impairment (20 per cent) were reported.[4] It is often useful therefore, when CD is being considered as a diagnosis, to obtain a print out of the patient's past history from the primary care doctor (having obtained the patient's consent). This may reveal repeated presentations to different specialists as well as a history of repeated surgical procedures, particularly without clear evidence of pathology. It is also worth noting that patients with a diagnosis of somatization disorder (what used to be referred to as Briquet's syndrome) have high rates of conversion disorders, which punctuate their illness careers, often after a life event or physical injury/procedure.[29]

(iv) Psychiatric co-morbidity

Rates of depression (38 to 50 per cent) and anxiety (10 to 16 per cent) have been identified in a number of studies. In a recent small prospective controlled study there was a fourfold increase in depression in comparison with matched control with similar organic disability.[30]

(v) Neurological disease and other physical factors

A diagnosis of functional paralysis can be made in a patient who already has some paralysis from another cause, for example, the 'disproportionate disability' in a patient with multiple sclerosis. In a recent study of patients with unexplained motor symptoms 42 per cent had a co-morbid neurological disease and half of these had a peripheral origin.[4] Epilepsy is thought to coexist in a significant percentage of patients with non-epileptic attacks.[31]

It is also important for the physician or surgeon to be aware of the diverse ways in which conversion symptoms can present in the general hospital. In the last decade the author has seen many patients with conversion disorders after surgical procedures, investigations, and operations such as hysterectomy, minor injuries in the workplace, and after (often trivial) road traffic accidents. Numerous case reports implicating accidents, minor surgical procedures, and general anaesthetics as initiating factors have been described, for patients with both non-epileptic seizures and functional neurological syndromes.[32–34] These presentations are more likely to be seen in a medico legal setting, where the symptoms are shaped by the prospect of financial gain.

(vi) Secondary gain/litigation

This is a complex issue but impending litigation has been described in a number of studies of patients with unexplained motor symptoms and tremor.[4,35] One group of patients who may develop abnormal movements are those with reflex sympathetic dystrophy (RSD) which has been renamed Complex Regional Pain Syndrome Type I (CRPS I). It has been reported that patients with CRPS I with abnormal movements typically exhibit pseudo neurological (non-organic) signs, and in some cases malingering has been documented by secret surveillance.[36] The authors concluded that abnormal movements in CRPS I are a key clinical feature that differentiates CRPS I from CRPS II. Psychiatrists will often be asked to express opinions on patients with chronic painful extremities (often labelled as 'CRPS') in which abnormal movements have developed, especially in a medico legal setting. Pearce[37] has remarked that CRPS is best construed as a reaction to injury, or to excessive, often iatrogenic, immobilization after injury; but should not be seen as an independent disease. He asserts that the diagnosis of CRPS groups together ill-defined symptoms under a convenient, but medically untestable label, and that patients, lawyers, and support groups commonly deny psychogenesis, with the sadly mistaken notion that this implies a bogus or spurious cause.

(vii) Laterality of the symptoms

The idea that left-sided symptoms are more common than right has a long history but a recent systematic review found no evidence to support this view.[38]

(viii) History from relative/informant

There is a considerable amount of evidence to suggest that the observations and attitudes of carers may be important in the perpetuation of medically unexplained symptoms, especially motor conversion symptoms. For example, Davison *et al.*[39] found carers to be ill-informed and dissatisfied with the advice they had received from doctors about their relatives' diagnosis and disabilities. The education of carers and relatives is essential and will be dealt with in the section on management.

(b) The examination and diagnostic discrepancies

There is often a discrepancy between the patient's concept of the symptoms and the physician's knowledge of the anatomy and physiology. The way in which a patient moves or undresses may indicate a global affection that is incompatible with a specific nerve lesion or with a hemiplegia.

When considering any sign of functional weakness it is important to remember the following caveats:

1 Any sign that depends on inconsistency does not distinguish 'hysterical' from 'malingered' weakness.

2 The presence of a positive sign of functional weakness does not exclude the possibility that the patient also has an organic disease as well.

3 All physical signs, whether for organic or non-organic signs, have a limited reliability and inter-rater reliability.[22]

Another myth wedded to the concept of hysteria (I have already referred to the mistaken but common belief that it has 'disappeared' and that symptoms are more common on the left) is that the patients exhibit 'la belle indifference' or are inappropriately under concerned about their symptoms. In a recent systematic review Stone *et al.*[40] found that the median frequency of la belle indifference was 21 per cent (range 0–54 per cent) in 356 patients with conversion symptoms and 29 per cent (range 0–60 per cent) in 157 patients with organic disease. Indifference to symptoms is more likely to be noted in patients with factitious disorder (see Chapter 5.2.9).

Give way weakness is often used as a diagnostic test of hysterical paralyses, but it is unreliable. Unilateral functional weakness of a leg, if severe, tends to produce a characteristic gait in which the leg is dragged behind the body as a single unit, like a sack of potatoes. The hip is either held in external or internal rotation so that the foot points inwards or outwards. The most impressive quantitative discrimination to date between hysterical and neurological weakness is reported in a study of Hoover's sign—the involuntary extension of hysterically paralysed leg when the 'good leg' is flexing against resistance. Ziv and colleagues[41] demonstrated a clear difference in the pattern of response between neurological and psychogenic patient groups. It should be borne in mind however that the patient may have both a functional and organic disorder.[4]

Individual symptoms

(a) Paralyses

Paralyses may affect one or more limbs, or one side of the face. They may be flaccid or occur with contractures. In hysterical spasm both arm and leg are contracted on the same side of the body, the hand is closed tightly, the knee is flexed, and perhaps the leg and the foot are drawn up. Paralysis with contractures is one of the most extreme examples of disability caused by hysterical illness.

Hysterical paraplegia has been described,[42] and both spinal and orthopaedic surgeons, as well as rehabilitation specialists and neurologists, should be alert to the development of this disorder in their patients.[43] These patients have the potential to use considerable health care resources.[21]

(b) Abnormal movements

Psychogenic movement disorders are thought to account for 1 in 30 of all patients attending a movement disorder clinic[35] and have been the subject of a recent book.[44] During the last two decades a number of case series of patients with psychogenic dystonia have been reported.[45,46] Clinical features that suggest a psychogenic movement disorder are shown in Table 5.2.4.3.

In a recent systematic study of patients with fixed dystonia Schrag *et al.*[47] found that 37 per cent fulfilled criteria for psychogenic dystonia and 29 per cent criteria for *somatization disorder*, which is characterized by chronic, multiple, persistent, medically unexplained symptoms. Despite the fact that many patients fulfilled strict criteria for a somatoform disorder/psychogenic dystonia, in a proportion of patients the diagnosis remained uncertain, and whether the disorder was primarily neurological or psychiatric remains an open question. These patients require the services of a multi-disciplinary team.

Table 5.2.4.3 Features that suggest a psychogenic movement disorder

Abrupt onset
- Inconsistent movements (changing characteristics over time)
- Incongruous movements (movements do not fit with recognized patterns or with normal physiological patterns)
- Presence of additional types of abnormal movements that are not consistent with the basic abnormal movement pattern or are not congruous with a known movement disorder, particularly:
  - Rhythmical shaking
  - Bizarre gait
  - Deliberate slowness in carrying out the requested voluntary movement
  - Bursts of verbal gibberish
  - Excessive startle (bizarre movements in response to sudden, unexpected noise or threatening movement)
- Entrainment of the psychogenic tremor to the rate of the requested rapid successive movement the patient is asked to perform
- Demonstrating exhaustion and fatigue
- Spontaneous remissions
- Movements disappear with distraction
- Response to placebo, suggestion, or psychotherapy
- Presence as a paroxysmal disorder
- Dystonia beginning as a fixed posture

(Adapted with permission, from S. Fahn (1995))

The most common form of psychogenic movement disorder however is *psychogenic tremor*.[48] Almost 75 per cent of presenting patients are female and preceding events include work-related injuries and other accidents. A positive entrainment test (see Table 5.2.4.3), absence of finger tremor, and slowness of voluntary movements are suggestive of psychogenic origin. One-third has co-morbid somatoform disorders and one-fifth is involved in litigation or compensation. Prognosis is relatively poor if the condition has persisted for over 1 year, and in the long-term 80–90 per cent of patients continue to have abnormal movements.

(c) Seizures (psychogenic non-epileptic seizures or PNES)

It is estimated that more than 25 per cent of patients receiving a diagnosis of refractory epilepsy in a chronic epilepsy clinic do not have epilepsy.[49] Although the population incidence of PNES may be only 4 per cent that of epilepsy, PNES comprises a large share of the workload of neurologists and emergency and general physicians. Unlike patients with epilepsy however, those with PNES often do not have designated nurses or health care workers assigned to help with the management of this potentially disabling disorder.

PNES can be distinguished from epileptic seizures: PNES generally occur in the presence of an audience or when one is close at hand. They may be precipitated by stress, but more often seem to occur in response to the social setting. The fall to the ground is not usually abrupt, and movements may follow the fall with clutching, but the characteristic regular tonic–clonic sequence of epilepsy is not found. Tongue biting and incontinence of urine are rare in

hysterical fits, the corneal reflexes are preserved and the plantars are flexor, unless previously abnormal. Firm handling and pressure on the supra orbital nerves to the point of pain may arouse the patient. PNES occur most often among epileptic patients or among others who have seen epileptic fits. A few epileptic patients learn how to induce ictal discharges and can produce extra fits. Although rarely available during a fit, the EEG is generally abnormal in epilepsy and normal during hysterical fits.[31]

If PNES is not diagnosed and managed early, significant iatrogenic harm may occur. The outcome is not always favourable in these patients: in one recent study carried out at a mean of 11.9 years after manifestation and 4.1 years after diagnosis of PNES, 71 per cent of patients continued to have seizures and 56 per cent were dependent on social security. Outcome was better in patients with greater educational attainments, younger onset and diagnosis, attacks with less dramatic features, and fewer additional medically unexplained complaints.[50]

It has recently been reported that patients with PNES have a consistently different psychosocial profile from patients with motor conversion symptoms. In a prospective study of consecutive neurological inpatients with either motor conversion or pseudo seizures of recent onset, patients with PNES were younger, more likely to have both an emotionally unstable personality disorder and a lower perception of parental care, to report incest, and to have reported more life events in the 12 months before symptom onset than patients with motor conversion symptoms.[24] Recently a helpful fact sheet has been produced to help patients with PNES, which explains the nature of the disorder and approaches to management.[51] Although cognitive–behavioural therapy has been shown to be helpful in an open trial of patients with PNES, these findings need replication in a controlled setting.[52]

Sensory symptoms

(a) Sensory disturbance

The clinical detection and localization of sensory dysfunction is probably one of the least reliable areas of the neurological examination. Sensory loss may involve half the entire body from top to toe or from right to left. It may affect the whole of a limb, and characteristically has a glove or stocking distribution on the arms or legs, or both. The sensory loss generally fails to fit in with known anatomical boundaries but conforms more with the patient's concept of physiology and anatomy. Thus hysterical sensory loss is likely to stop sharply at the midline, while non-hysterical sensory change will only approach the midline since at this point segmental nerves overlap by one or two centimetres on each side.

Unfortunately these classical signs are often unreliable. 'Psychogenic' features on sensory examination and diminished vibration sense over the affected part of the forehead have been found in over half of patients with neurological disorders.[53, 54] Rolak also found that 'midline splitting' of sensory function was not helpful in determining whether there was an underlying neurological disorder.[54] These clinical findings should clearly be interpreted with circumspection.

Toth has recently described 34 patients with the 'hemisensory' syndrome, in which patients present with hemisensory disturbance and intermittent blurring of vision in the ipsilateral eye (asthenopia) and sometimes ipsilateral hearing problems as well.[55] Hemisensory symptoms are increasingly recognized in patients with chronic pain and in patients with reflex sympathetic dystrophy.

(b) Visual disturbances

Ophthalmologists have estimated that psychogenic visual disorders account for up to 5 per cent of their practice.[56] Simple observation of visually guided behaviour will sometimes reveal telling inconsistencies, particularly in the case of severe apparent visual loss. A number of reliable optometric techniques are available to support bedside tests and the diagnosis of psychogenic visual loss, field disturbance, or gaze abnormality (for more details see Stone *et al.*[23]). Disabling hysterical blindness presents more difficulties. Evoked potential studies will help to demonstrate intact visual pathways.

The disability associated with chronic conversion disorders

This topic is under-researched, but it is worth noting that, as the prognosis of CD is often poor, (not infrequently because patients are not diagnosed promptly, and even after diagnosis there are no resources to treat the patient). In the experience of the author this clinical conundrum is not improving: with the increase in provision of psychiatric services to the community and those with 'serious mental illness', patients with CD, even if they are profoundly disabled, often do not receive the appropriate treatment. Much of the disability is iatrogenic, and these patients will, not infrequently, be referred to the psychiatric service after having become confined to a wheelchair and/or in receipt of long-term disability benefits.[21,39] By this time the patients are usually entrenched in the sick role and it is very difficult to change the status quo.

Prognosis

The aetiological implications accruing from recent follow-up studies[4, 57] suggest that a short history and young age are held to be predictors of good outcome, while the presence of a personality disorder, chronicity of symptoms, receipt of disability benefits, and involvement with litigation predict poor recovery. As regards social circumstances, a change in marital status, good family functioning, and the elimination of a stressor has been shown to have a positive effect on outcome.[58] There is little chance of improvement once the symptoms have become chronic and enduring.

Patients with chronic motor symptoms, e.g. unilateral functional weakness, as well as those with sensory symptoms, appear to do particularly poorly. In particular, patients with unexplained motor symptoms who are referred to tertiary care centres continue to do very poorly following discharge. Despite the stability of the diagnosis, a pattern of multiple hospital referrals continues for many of these patients once they have been discharged from the tertiary care centre. Interviews of patients conducted on an average of 6 years after their original admission to a tertiary care centre revealed that many continued to be referred to neurologists and other specialists, but that subsequent psychiatric referral was rare.[59] Many changed their primary care doctor after discharge from hospital and a disproportionate number of re-referrals were made by primary care doctors who had known their patients for less than 6 months. Psychological attribution of symptoms was rare, and many patients felt dissatisfied with the treatment they had received. Many were exposed to unnecessary iatrogenic harm. These consistent findings of very poor outcome following discharge from neurological outpatient and inpatient services in patients with both

unexplained motor disorders as well as PNES suggest that without appropriate treatment the prognosis is poor.

Management

It is remarkable that a disorder as common as schizophrenia and multiple sclerosis should have attracted so little research interest or treatment resources. One reason for this is that there have been no randomized controlled treatment studies of CD, and so at the time of writing there is no good evidence about the best intervention for conversion disorder. There has been considerable interest in this topic however, which has been the subject of a recent Cochrane review.[60] All of the studies in this review were of poor methodological quality. On the credit side there is evidence that interest in CD is increasing and attracting more research funding.

Resources

Before any discussion of treatment it is important to consider the resources available to the neurologist to manage these patients. It is anomalous that, unlike disorders such as MS and schizophrenia, which have a similar prevalence, there are no designated resources for these patients. Some neurologists may have no access whatever to mental health resources, whereas others may have close collaborative links with either clinical psychology or psychiatry services. There is no doubt that the successful management of these patients requires the co-operation of a number of clinical specialties, including psychologists, nurses, physiotherapists, and occupational therapists (OTs). Some patients may be so disturbed or disabled (or both) that they may require inpatient admission to a specialized unit with access to both, mental health care and medical nurses, as well as physiotherapists and OTs. In the opinion of this writer every neurology service should have access to a specialist liaison psychiatry service.[61]

Management strategies for the neurologist

First, the diagnosis has to be established by a neurologist after relevant organic disease has been excluded. Second, the neurologist has to not only explain to the patient that there is no serious underlying organic disease but also provide an explanation for the symptoms that is comprehensible to the patient.

It is worth noting at this stage that patients prefer the term 'functional' rather than 'hysterical', when their unexplained weakness, fits, etc. are being referred to.[62] It is also important to avoid verbal landmines—for example using the phrase 'not sinister' instead of 'not serious'; or 'not structural' instead of 'not physical'. In patients who are generally hostile to psychological explanations it is best to use the word 'functional' instead of 'psychological'.

An example of an explanation to a patient with functional weakness and sensory disturbance may be something like: *you have what we call functional weakness. This is a common medical problem. Your nervous system is not damaged—we can see that from the examination and scans, etc. This is why why when you try and send the messages to your limbs they do not move properly. Similarly this is why the sensations from your body are not being felt properly. The most important thing about this condition is that because your nervous system is not damaged, the problem is potentially reversible. All the parts of the nervous system are there but are just not working properly, so that when you try to move your leg it doesn't do it as well as it should. Sometimes stress can cause these symptoms, which are often accompanied by worry and low mood but these are not the cause of the problem. Stress is a common problem and can lead to headache and abdominal pain as well as what we call functional weakness.*

This explanation can be supplemented by giving the patient a **fact sheet** containing information about functional weakness, which contains information about how to become involved with rehabilitation[51] (Fig. 5.2.4.3).

Further management

Ideally the neurologist and psychiatrist should interview the patient together at the bedside, but this is not always possible. At the very least however close collaboration between the two is essential before the patient is reviewed by the psychiatrist and a formulation proposed (and any potential for iatrogenic illness or diagnostic confusion eliminated).

Traditional behavioural approaches to treatment are based on the premise that the symptoms reported by the patient are interpreted as physical but are amenable to recovery. Treatment aims to bring about a gradual increase in function through a combination of physical and occupational therapies. The patient receives rewards and praise for improvement of function, and withdrawal of reinforcement for continuing signs of disability. Avoiding direct confrontation of psychological problems and providing 'face-saving' techniques are also regarded as key components.[63] More recently the approach to patients has moved from a predominantly medical one, to one in which psychological and sociocultural aspects are equally important, and the need for organized specialist rehabilitation services involving a multi-disciplinary team is recognized as essential.

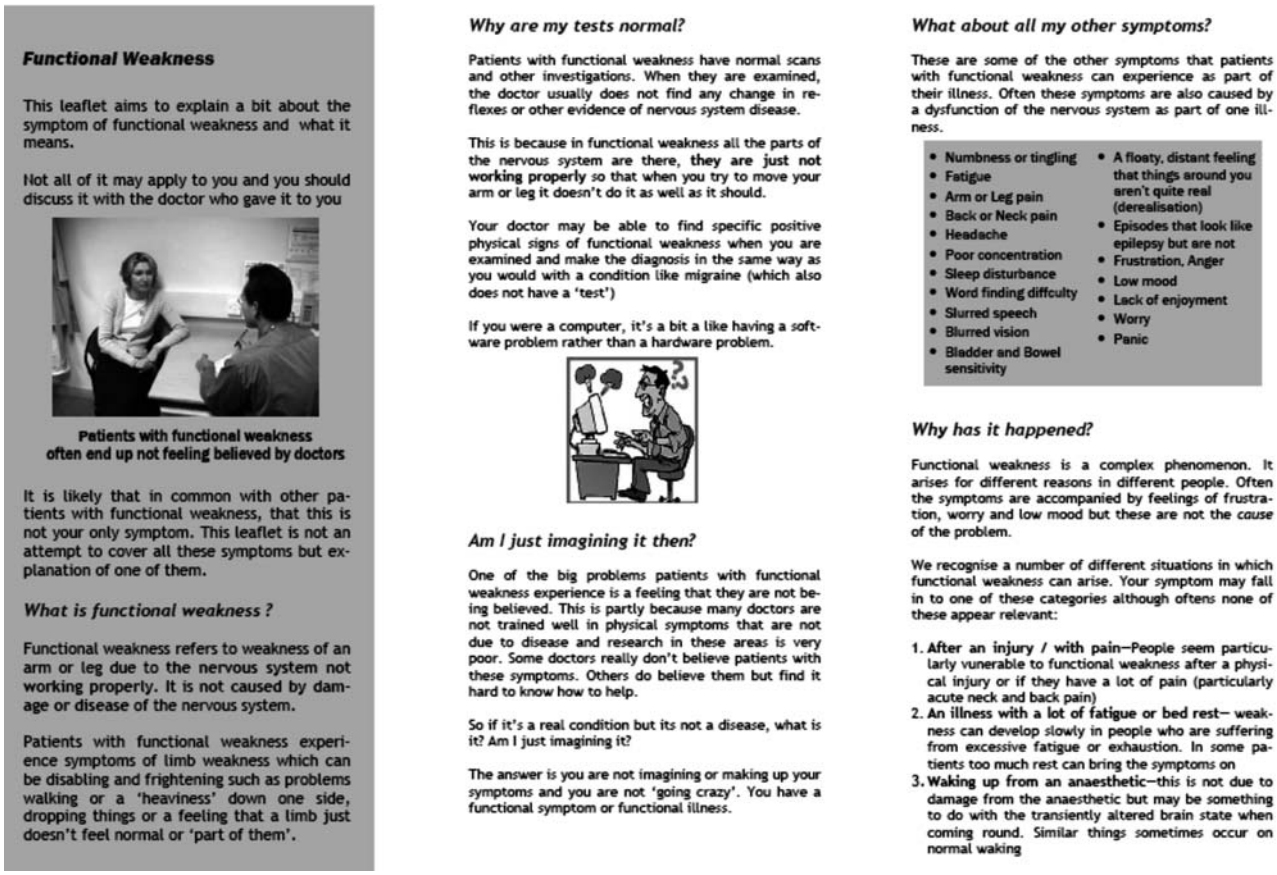
What is the evidence?

With one or two recent exceptions,[64] there are no large, randomized controlled studies of treatment in patients with CDs. Neither is there any good evidence to support the use of one specific intervention, e.g. biofeedback, hypnosis, psychotherapy. Although repeated case series have documented the effectiveness of multidisciplinary inpatient behavioural treatment, there is little controlled research.

In the absence of good experimental evidence a possible framework for future research has been developed which is based on published evidence and described in the WHO ICIDH.[65] This is particularly useful for patients in whom there is a disability that is out of proportion to known disease and signs. The model provides opportunities for intervention, and is well suited to the kind of multi-disciplinary approach that is likely to be successful in these patients.

The model emphasizes that whatever the primary cause of an illness, many factors (both individual and systemic) will have an influence on its manifestations. *For example: a vignette here.*

A 20-year-old woman was referred to the liaison service with functional paraplegia of 12 months duration. Onset was temporally related to a back strain caused by lifting a chair. She was confined to a wheelchair and lived with her parents and 14-year-old brother in an adapted house (specially adapted chair and stair rails) and was in receipt of benefits. A home visit was carried out and the patient denied any current problems or recent life events, although she described a long history of medically unexplained symptoms, multiple food allergies, and previous treatment in an adolescent unit for chronic



Functional Weakness

This leaflet aims to explain a bit about the symptom of functional weakness and what it means.

Not all of it may apply to you and you should discuss it with the doctor who gave it to you

Patients with functional weakness often end up not feeling believed by doctors

It is likely that in common with other patients with functional weakness, that this is not your only symptom. This leaflet is not an attempt to cover all these symptoms but explanation of one of them.

What is functional weakness ?

Functional weakness refers to weakness of an arm or leg due to the nervous system not working properly. It is not caused by damage or disease of the nervous system.

Patients with functional weakness experience symptoms of limb weakness which can be disabling and frightening such as problems walking or a 'heaviness' down one side, dropping things or a feeling that a limb just doesn't feel normal or 'part of them'.

Why are my tests normal?

Patients with functional weakness have normal scans and other investigations. When they are examined, the doctor usually does not find any change in reflexes or other evidence of nervous system disease.

This is because in functional weakness all the parts of the nervous system are there, **they are just not working properly** so that when you try to move your arm or leg it doesn't do it as well as it should.

Your doctor may be able to find specific positive physical signs of functional weakness when you are examined and make the diagnosis in the same way as you would with a condition like migraine (which also does not have a 'test')

If you were a computer, it's a bit a like having a software problem rather than a hardware problem.

Am I just imagining it then?

One of the big problems patients with functional weakness experience is a feeling that they are not being believed. This is partly because many doctors are not trained well in physical symptoms that are not due to disease and research in these areas is very poor. Some doctors really don't believe patients with these symptoms. Others do believe them but find it hard to know how to help.

So if it's a real condition but its not a disease, what is it? Am I just imagining it?

The answer is you are not imagining or making up your symptoms and you are not 'going crazy'. You have a functional symptom or functional illness.

What about all my other symptoms?

These are some of the other symptoms that patients with functional weakness can experience as part of their illness. Often these symptoms are also caused by a dysfunction of the nervous system as part of one illness.

- Numbness or tingling
- Fatigue
- Arm or Leg pain
- Back or Neck pain
- Headache
- Poor concentration
- Sleep disturbance
- Word finding diffculty
- Slurred speech
- Blurred vision
- Bladder and Bowel sensitivity
- A floaty, distant feeling that things around you aren't quite real (derealisation)
- Episodes that look like epilepsy but are not
- Frustration, Anger
- Low mood
- Lack of enjoyment
- Worry
- Panic

Why has it happened?

Functional weakness is a complex phenomenon. It arises for different reasons in different people. Often the symptoms are accompanied by feelings of frustration, worry and low mood but these are not the *cause* of the problem.

We recognise a number of different situations in which functional weakness can arise. Your symptom may fall in to one of these categories although oftens none of these appear relevant:

1. **After an injury / with pain**—People seem particularly vunerable to functional weakness after a physical injury or if they have a lot of pain (particularly acute neck and back pain)
2. **An illness with a lot of fatigue or bed rest**— weakness can develop slowly in people who are suffering from excessive fatigue or exhaustion. In some patients too much rest can bring the symptoms on
3. **Waking up from an anaesthetic**—this is not due to damage from the anaesthetic but may be something to do with the transiently altered brain state when coming round. Similar things sometimes occur on normal waking

Fig. 5.2.4.3 Fact sheet for patient with functional weakness (Reproduced from Stone, J. Carson, A. and Sharpe, M. (2005), Functional symptoms in neurology: management. *Neurology in Practice,* **71**(Suppl. 1), i13–i21. Copyright 2005, BMJ Publishing Group Ltd.)

fatigue syndrome. She was seeing a chiropractor for her symptoms, and had been told that she had nerve damage (not confirmed). Her brother was off school with chronic fatigue and her mother had a long history of emotional problems. At interview, the family were polite but could not identify any link between recent events and her disabling symptoms. Discussion with the GP did not reveal any other relevant information and a follow-up was arranged 1 month later. Before this appointment the patient telephoned the psychiatrist to say that she had not disclosed certain facts at the initial meeting, and revealed that she had confronted her physically abusive step father and asked him to leave the house to live with his mistress (which he duly did). At follow-up there was a great deal of expressed relief, and the patient agreed with our formulation that, to some extent, she had developed weakness in the legs and become wheelchair bound in order to avoid being hit by her step father. She agreed to a brief admission to a rehabilitation unit, was encouraged to mobilize gradually and had regular sessions with a clinical psychologist. She responded well to treatment, commenced a college course, and learned to drive a car. At 3 year follow-up she was symptom free and in gainful employment. [the first vignette in this chapter].

Psychological treatments

Because patients with conversion disorders share features in common with patients with other medically unexplained syndromes, treatments that have been used in these latter disorders may have potential. Most of the evidence-based treatments in this field involve cognitive–behavioural therapy (CBT,[66] or interpersonal therapy (IPT)). These usually have to be undertaken by trained clinical psychologists or other clinicians. However, increasing numbers of specialist nurses are being trained to deliver these treatments, so they should become more widely available.

CBT is concerned mainly with helping the patients overcome identified problems and ascertain specified goals. It discourages 'maintaining factors' such as repeated body self-checking and excessive bedrest, and challenges patients' negative or false beliefs about symptoms. Chalder has described specific CBT based treatment for patients with conversion disorders.[67]

Hypnosis and intravenous sedation

In an inpatient trial of hypnosis both patients with CD and controls improved equally and no extra effect from hypnosis was found.[68] Others have found the use of intravenous sedatives, particularly propofol, helpful in persuading some patients with whom a good relationship has been established, that they can eventually make a recovery.[51]

Pharmacological treatments

There is evidence from randomized controlled trials (RCTs) and systematic reviews that antidepressants (both tricyclics and selective

serotonin reuptake inhibitors (SSRIs)) can be useful in the treatment of patients with medically unexplained symptoms (such as poor sleep and pain), whether or not depression is present.[69]

Further information

Halligan, P., Bass, C., and Marshall, J. (eds.) (2001). *Contemporary approaches to the study of hysteria. Clinical and theoretical perspectives*. Oxford University Press, Oxford.

Hallett, M., Fahn, S., Jancovic, J., *et al.* (eds.) (2005). *Psychogenic movement disorders: psychobiology and treatment of a functional disorder*. Lippincott, Williams and Wilkins.

Trimble, M. (2004). *Somatoform disorders: a medicolegal guide*. Cambridge University Press, Cambridge.

References

1. Collie, J. (1913). *Malingering and feigned sickness*. Edward Arnold, London.
2. Spence, S. (1999). Hysterical paralyses as disorders of action. *Cognitive Neuropsychiatry*, **4**, 203–26.
3. Carson, A., Ringbauer, B., Stone, J., *et al.* (2000). Do medically unexplained symptoms matter? A prospective cohort study of 300 new referrals to neurology outpatient clinics. *Journal of Neurology, Neurosurgery, and Psychiatry*, **68**, 207–11.
4. Crimlisk, H., Bhatia, K., Cope, H., *et al.* (1998). Slater revisited: 6 year follow up study of patients with medically unexplained motor symptoms. *British Medical Journal*, **316**, 582–6.
5. ICD-10 Classification of Mental and Behavioural Disorders (1992). WHO, Geneva.
6. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn.). American Psychiatric Association, Washington, DC.
7. Shapiro, A. and Teasell, R.W. (2004). Behavioural interventions in the rehabilitation of acute v chronic non-organic (conversion/factitious) motor disorders. *British Journal of Psychiatry*, **185**, 140–6.
8. Bass, C. and Halligan, P. (2007). Illness related deception: social or psychiatric problem? *Journal of the Royal Society of Medicine*, **100**, 81–4.
9. Halligan, P., Bass, C., and Oakley, D. (2003). Wilful deception as illness behaviour. In *Malingering and illness deception* (eds. P. Halligan, C. Bass, and D. Oakley), pp. 3–28. Oxford University Press, Oxford.
10. Spence, S., Hunter, M., and Harpin, G. (2002). Neuroscience and the will. *Current Opinion in Psychiatry*, **15**, 519–26.
11. Holmes, E., Brown, R., Mansell, W., *et al.* (2005). Are there two qualitatively distinct forms of dissociation? A review and some clinical implications. *Clinical Psychology Review*, **25**, 1–25.
12. Brown, R.J. (2004). Psychological mechanisms of medically unexplained symptoms: an integrative conceptual model. *Psychological Bulletin*, **130**, 793–812.
13. Broome, M. (2004). A neuroscience of hysteria? *Current Opinion in Psychiatry*, **17**, 465–9.
14. Ghaffar, O., Staines, W., and Feinstein, A. (2006). Unexplained neurologic symptoms: an fMRI study of sensory conversion disorder. *Neurology*, **67**, 2036–8.
15. Marshall, J., Halligan, P., Gink, G., *et al.* (1997). The functional anatomy of a hysterical paralysis. *Cognition*, **64**, B1–8.
16. Spence, S., Crimlisk, H., Cope, H., *et al.* (2000). Discrete neurophysiological correlates in prefrontal cortex during hysterical and feigned disorder of movement. *Lancet*, **355**, 1243–4.
17. Kanaan, R.A., Wessely, S., and David, A. (2007). Imaging repressed memories in motor conversion disorder. *Psychosomatic Medicine*, **69**, 202–5.
18. Micale, M. (2001). Hysteria. In *The Oxford companion to the body* (eds. C. Blakemore and S. Jennett), pp. 382–4. Oxford University Press, Oxford.
19. Parry, A., Murray, B., Hart, Y., *et al.* (2006). Audit of resource use in patients with non-organic disorders admitted to a UK neurology unit. *Journal of Neurology, Neurosurgery and Psychiatry*, **77**, 1200–1.
20. Akagi, H. and House, A. (2002). The clinical epidemiology of hysteria: vanishingly rare, or just vanishing? *Psychological Medicine*, **32**, 191–4.
21. Allanson, J., Wade, D., and Bass, C. (2002). Characteristics of patients with persistent severe disability and medically unexplained neurological symptoms: a pilot study. *Journal of Neurology, Neurosurgery, and Psychiatry*, **73**, 307–9.
22. Stone, J. and Zeman, A. (2001). Hysterical conversion-a view from clinical neurology. In *Contemporary approaches to the science of hysteria* (eds. P. Halligan, C. Bass, and J. C Marshall) pp. 102–25, Oxford University Press, Oxford.
23. Stone, J., Sharpe, M., and Binzer, M. (2004). Motor conversion symptoms and pseudoseizures: a comparison of clinical characteristics. *Psychosomatics*, **45**, 492–9.
24. Binzer, M., Stone, J., and Sharpe, M. (2004). Recent onset pseudoseizures—clues to aetiology. *Seizure*, **13**, 146–55.
25. O'Sullivan, G., Harvey, I., Bass, C., *et al.* (1992). Psychophysiological investigations of patients with unilateral symptoms in the hyperventilation syndrome. *British Journal of Psychiatry*, **160**, 664–7.
26. Tavel, M. (1964). Hyperventilation syndrome with unilateral somatic symptoms. *Journal of the American Medical Association*, **187**, 301–3.
27. Blau, N., Wiles, M., and Solomon, F. (1989). Unilateral somatic symptoms due to hyperventilation. *British Medical Journal*, **286**, 1108.
28. Wessely, S., Nimnuan, C., and Sharpe, M. (1999). Functional somatic syndromes: one or many? *Lancet*, **354**, 36–9.
29. Bhui Hotopf, M. (1997). Somatisation disorder. *British Journal of Hospital Medicine*, **58**, 145–9.
30. Binzer, M., Andersen, P., and Kullgren, G. (1997). Clinical characteristics of patients with motor disability due to conversion disorder: a prospective control group study. *Journal of Neurology, Neurosurgery, and Psychiatry*, **63**, 83–8.
31. Reuber, M. and Elger, C. (2003). Psychogenic non epileptic seizures: review and update. *Epilepsy & Behaviour*, **4**, 205–16.
32. Letonoff, E.J., Williams, T.R., and Sidhu, K. (2002). Hysterical paralysis: a report of three cases and a review of the literature. *Spine*, **27**, E441–5.
33. Lichter, I., Goldstein, L., Toone, B., *et al.* (2004). Nonepileptic seizures following general anaesthetic: a report of 5 cases. *Epilepsy & Behaviour*, **5**, 1005–13.
34. Reuber, M., Howlett, S., Khan, A., *et al.* (2007). Non-epileptic seizures and other functional neurological symptoms: predisposing, precipitating, and perpetuating factors. *Psychosomatics*, **48**, 230–8.
35. Factor, S., Podskalny, R., and Molho, E. (1995). Psychogenic movement disorders: frequency, clinical profile and characteristics. *Journal of Neurology, Neurosurgery and Psychiatry*, **59**, 406–12.
36. Verdugo, R. and Ochoa, J. (2000). Abnormal movements in complex regional pain syndrome: assessment of their nature. *Muscle Nerve*, **23**, 198–205.
37. Pearce, J. (2005). Chronic regional pain and chronic pain syndromes. *Spinal Cord*, **43**, 263–8.
38. Stone, J., Sharpe, M., Carson, A., *et al.* (2002). Are functional motor and sensory symptoms really more frequent on the left? A systematic review. *Journal of Neurology, Neurosurgery, and Psychiatry*, **73**, 578–81.
39. Davison, P., Sharpe, M., Wade, D., *et al.* (1999). "Wheelchair" patients with non organic disease: a psychological enquiry. *Journal of Psychosomatic Research*, **47**, 93–103.
40. Stone, J., Smyth, R., Carson, A., *et al.* (2006). La belle indifference in conversion symptoms and hysteria: systematic review. *British Journal of Psychiatry*, **188**, 204–9.
41. Ziv, I., Djaldetti, R., and Zoldan, Y. (1998). Diagnosis of "non-organic" limb paresis by a novel objective motor assessment: the quantitative Hoover's test. *Journal of Neurology*, **245**, 797–802.
42. Baker, J. and Silver, J. (1987). Hysterical paraplegia. *Journal of Neurology, Neurosurgery and Psychiatry*, **50**, 375–82.

43. Heruti, R., Reznik, J., Adunski, A., *et al.* (2002). Conversion motor paralysis disorder: analysis of 34 consecutive referral. *Spinal Cord*, **430**, 335–40.
44. Hallett, M., Fahn, S., Jankovic, J., *et al.* (eds.) (2005). *Psychogenic movement disorders: psychobiology and treatment of a functional disorder*. Lippincott, Williams and Wilkins, Philadelphia.
45. Fahn, S. and Williams, D. (1988). Psychogenic dystonia. In *Advances of Neurology*, Vol. 50: *Dystonia 2* (eds. S. Fahn, *et al.*), pp. 431–55. Raven Press, New York.
46. Lang, A.E. (1995). Psychogenic dystonia: a review of 18 cases. *Canadian Journal of Neurological Sciences*, **22**, 136–43.
47. Schrag, A., Trimble, M., Quinn, N., *et al.* (2004). The syndrome of fixed dystonia: an evaluation of 103 patients. *Brain*, **127**, 2360–72.
48. Bhatia, K. and Schneider, S. (2007). Psychogenic tremor and related disorders. *Journal of Neurology*, Apr 9; [Epub ahead of print].
49. Smith, D., Defalla, B., and Chadwick, D. (1999). The misdiagnosis of epilepsy and the management of refractory epilepsy in a specialist clinic. *Quarterly Journal of Medicine*, **92**, 15–23.
50. Reuber, M., Pukrop, R., Bauer, J., *et al.* (2003). Outcome in psychogenic nonepileptic seizures: 1 to 10 year follow up in 164 patients. *Annals of Neurology*, **53**, 305–11.
51. Stone, J., Carson, A., and Sharpe, M. (2005). Functional symptoms in neurology: management. *Neurology in Practice*, **71**(Suppl. 1), i13–i21. http://www.jnnp.com.
52. Goldstein, L., Mellors, J., and Toone, B. (2004). An evaluation of cognitive behavioral therapy as a treatment for dissociative seizures: a pilot study. *Cognitive Behavioural Neurology*, **17**, 41–9.
53. Gould, R., Miller, B., Goldberg, M., *et al.* (1986). The validity of hysterical signs and symptoms. *Journal of Nervous and Mental Diseases*, **174**, 593–7.
54. Rolak, L. (1988). Psychogenic sensory loss. *Journal of Nervous and Mental Disease*, **176**, 686–7.
55. Toth, C. (2003). Hemisensory syndrome is associated with a low diagnostic yield and a nearly uniform benign prognosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, **74**, 1113–6.
56. Kathol, R., Cox, T., Corbett, J., *et al.* (1983). Functional visual loss: I. A true psychiatric disorder? *Psychological Medicine*, **13**, 307–14.
57. Stone, J., Sharpe, M., Rothwell, P., *et al.* (2003). The 12-year prognosis of unilateral functional weakness and sensory disturbance. *Journal of Neurology, Neurosurgery, and Psychiatry*, **74**, 591–6.
58. Ron, M. (2001). The prognosis of hysteria/somatisation disorder. In *Contemporary approaches to the study of hysteria* (eds. P. Halligan, C. Bass, and J. Marshall), pp. 271–83, Oxford University Press, Oxford.
59. Crimlisk, H., Bhatia, K., Cope, H., *et al.* (2000). Patterns of referral in patients with medically unexplained motor symptoms. *Journal of Psychosomatic Research*, **49**, 217–9.
60. Ruddy, R. and House, A. (19 October 2005). Psychosocial interventions for conversion disorder. *Cochrane Database Systematic Review*, (4): CD005331.
61. Gotz, M. and House, A. (1998). Prognosis of symptoms that are medically unexplained. *British Medical Journal*, **317**, 536.
62. Stone, J., Wojcik, W., and Durrance, D., (2002). What should we say to patients with symptoms unexplained by disease? The number needed to offend. *British Medical Journal*, **325**, 1449–50.
63. Teasell, R. and Shapiro, A. (1993). Rehabilitation of chronic motor conversion disorder. *Critical Review of Physical and Rehabilitation Medicine*, **5**, 1–13.
64. Moene, F., Spinhoven P, *et al.* (2003). A randomized controlled clinical trial of a hypnosis-based treatment for patients with conversion disorder, motor type. *International Journal of Clinical and Experimental Hypnosis*, **51**, 29–50.
65. WHO. (2001). International classification of functioning, disability and health. World Health Organization, Geneva. http://www3.who.int/icficftemplate.cfm
66. Kroenke, K. and Swindle, R. (2001). Cognitive behavioural therapy for somatisation and symptom syndromes: a critical review of controlled clinical trials. *Psychotherapy and Psychosomatics*, **69**, 205–15.
67 Chalder, T. (2001). Cognitive behavioural therapy as a treatment for conversion disorders. In *Contemporary approaches to the study of hysteria: clinical and theoretical perspectives* (eds. P. Halligan, C. Bass, and J. Marshall), pp. 298–311. Oxford University Press, Oxford.
68. Moene, F., Spinhoven, P., Hoogduin, K., *et al.* (2002). A randomised controlled clinical trial on the additional effect of hypnosis in a comprehensive treatment programme for inpatients with conversion disorder of the motor type. *Psychotherapy and Psychosomatics*, **71**, 66–76.
69. Jackson, J., O'Malley, P., and Kroenke, K. (2006). Antidepressants and cognitive-behavioral therapy for symptom syndromes. *CNS Spectrum*, **11**, 212–22.

5.2.5 Hypochondriasis (health anxiety)

Russell Noyes Jr.

Introduction

Hypochondriasis is a preoccupation with the fear that one has, or may develop, serious disease despite evidence to the contrary. So defined, the disorder affects between 2 and 7 per cent of patients attending general medical clinics and is a cause of physical dysfunction and disability.[1] It is also a reason for increased health care utilization and dissatisfaction with care received. To their physicians, patients with this disorder are an enigma and a source of frustration.

Unfortunately, relatively little is known about hypochondriasis. Primary care physicians have had little interest and psychiatrists see few patients with the condition. It is a pejorative label that, even if entertained, is rarely communicated. And, even if communicated, the diagnosis would not, until very recently, have led to effective treatment.

History

Hypochondria was used by Hippocrates to refer to a region below the cartilage of the ribs. In the second century, Galen linked it to organs in this area as well as humours and animal spirits. The symptom picture was ill-defined and only gradually took on the characteristics recognized today. From earliest times the disorder was associated with melancholia, a temperamental disturbance caused by an excess of black bile. Burton (1621) described hypochondriacal melancholy in terms of vague physical symptoms, disturbances of mood, and fears. In the seventeenth century, Sydenham viewed hypochondria in men as the counterpart of hysteria in women, but the first modern description was published in 1799 by Sims.

By the eighteenth century, hypochondria became part of a fashionable disturbance that Cheyne attributed to the English way of life and environment. However, as notions of aetiology began to shift under the influence of Cartesian dualism, hypochondria was increasingly seen as a weakness and moral failing. Falret (1822) was

Table 5.2.5.1 Essential and associated features of hypochondriasis

Essential features
Fear of disease
Disease conviction
Bodily preoccupation
Somatic symptoms
Reassurance-seeking
Associated features
Fear of aging and death
Overvaluation of health
Low self-esteem
Sense of vulnerability to illness

perhaps the first to identify it as a mental disorder, one of the neuroses. Freud viewed hypochondria as an 'actual neurosis', having a physiological basis and not amenable to psychoanalysis. However, present-day descriptions began with Gillespie,[2] who in 1928 defined hypochondriasis as 'a mental preoccupation with a real or supposititious physical or mental disorder'.

Conceptualizations

Authors disagree about how hypochondriasis should be conceptualized. Some look upon it as a personality trait; its early onset and long-term stability in many patients fit this conception. Others view it as a dimension of psychopathology. They see illness worry as a continuum with hypochondriasis falling on the severe end. For those who take a categorical approach, the issue of whether hypochondriasis is primary or secondary remains unsettled. High rates of comorbidity create doubt about its independent status. Based on existing evidence, some question whether hypochondriasis can be regarded as a discrete psychiatric disorder.[3]

Clinical picture

Essential features

The essential characteristics of hypochondriasis are shown in Table 5.2.5.1. These include fear of serious disease, the consequences of which may include pain, suffering, disability, and death. Such fears take the form of alarming thoughts and images of specific diseases. They also include conviction or belief that the feared disease is already present. This belief is overvalued meaning that it is strongly held despite lack of evidence; it is not delusional.

Bodily preoccupation is perhaps the most important feature.[4] This takes the form of intense interest in, and attention to, what is happening in the body. The focus is upon somatic symptoms which tend to be multiple and diffuse. Attention is also directed to bodily sensations, bodily functions, and minor abnormalities as well as related concerns such as diet, exercise, and environmental exposures. The activities and conversation of patients are dominated by medical concerns. As a consequence of their self-absorption, interest in other people and pursuits is withdrawn.

Reassurance-seeking is the main behavioural feature. Patients repeatedly check their bodies for signs of serious disease. They check their pulse, look for lumps, examine themselves in the mirror, etc. In addition, they search medical sources for the meaning of their symptoms. Such patients also ask friends, family, and medical professionals for reassurance. Their search may lead to excessive utilization of health services.

Associated features

Associated characteristics include fears of aging and death, which appear to be an integral part of hypochondriasis. Overvaluation of health and appearance is another related feature. Hypochondriacal patients may become preoccupied with eating natural foods, achieving physical fitness, and living a healthy lifestyle, activities that reflect their idealized conception of good health.

Patients with hypochondriasis feel unworthy and unlovable.[4] As a consequence of their low self-esteem they have negative expectations of others including medical professionals. In addition, they have a sense of vulnerability to illness.[5] These characteristics have to do with fundamental aspects of the self that the hypochondriacal patient views as deficient.

Subtypes

Hypochondriacal patients are heterogeneous and subtypes may exist. Separate dimensions of disease phobia and disease conviction have consistently been identified; in some patients fears are prominent and in others conviction dominates the picture. Others may resemble patients with obsessive-compulsive disorder or personality disorders of one kind or another.

Classification

Criteria

Hypochondriasis initially appeared in DSM-II as one of the neuroses. In DSM-III, it was moved to the somatoform disorders, and diagnostic criteria were provided. In a revision of the classification (DSM-III-R), a duration of 6 months was added, and patients with delusional beliefs were excluded. The DSM-IV criteria are shown in Table 5.2.5.2. They exclude patients whose symptoms are better explained by other anxiety, depressive, or somatoform disorders.[1] Also, in DSM-IV, specific phobia of illness is separated from hypochondriasis. The illness phobic is said to fear contracting an illness whereas the hypochondriac fears disease already present.

The ICD-10 criteria for hypochondriacal disorder differ from those in DSM-IV. They require a persistent belief about having one or more specifically named serious physical diseases.[6] In addition, they include body dysmorphic disorder. With respect to illness behaviour, the ICD-10 criteria state that hypochondriacal concerns cause persons to seek medical investigation or treatment. They also state that patients may accept reassurance in the short-term, but that in the long run they are not likely to respond.

Table 5.2.5.2 Abbreviated DSM-IV diagnostic criteria for hypochondriasis

(a) Preoccupation with fears of having, or the idea that one has, a serious disease based on misinterpretation of bodily symptoms
(b) The preoccupation persists despite appropriate medical evaluation and reassurance
(c) Belief not of delusional intensity
(d) Preoccupation causes significant distress or impairment
(e) Duration of at least 6 months
(f) Not better accounted for by other anxiety, depressive, or somatoform disorders

The somatoform disorders category to which hypochondriasis belongs is controversial, and many question its inclusion in the classification.[7] They see these disorders as ill-defined, of questionable validity and based more on illness behaviour than on distinctive features. They also view them as creations of Western biomedicine that serve to devalue patients who challenge the theoretical model upon which it is based.[8] According to that model, illness is a response to disease, and the person who is ill without disease, e.g. hypochondriasis, is marginalized.

Were the somatoform disorders to be eliminated, some have proposed moving hypochondriasis to the anxiety disorders (health anxiety) or to a proposed grouping, the obsessive-compulsive spectrum disorders.

Validity

Evidence for the validity and utility of the diagnosis of hypochondriasis remains limited. In studies aimed at demonstrating validity, Barsky *et al.*[9] showed that distinguishing characteristics of the disorder aggregated in some medical outpatients but were less common in others. The same patients had other features of hypochondriasis indicating external validity. Using a structured interview for hypochondriasis, these investigators and others[10,11] observed a positive correlation between interview and physician ratings (concurrent validity). Hypochondriacal patients also had more ancillary features of hypochondriasis than did control patients (external validity). Also, other clinical characteristics distinguished interview positive from interview negative patients, indicating discriminate validity. Follow-up studies have shown a degree of diagnostic stability suggesting predictive validity.[12,13]

Measures

A variety of measures have been developed to screen for hypochondriasis and assess the severity of hypochondriacal concerns.[14] These are shown in Table 5.2.5.3. The Whiteley Index, a self-report instrument based on the observed characteristics of hypochondriacal psychiatric patients, is one of the most widely used.[15] It consists of 14 yes versus no items, but recent work suggests that a 7-item version is satisfactory for screening. The Illness Attitude Scales is a 27-item measure of psychopathology associated with hypochondriasis.[16] A principal components' analysis yielded two factors, one measuring health anxiety and the other illness behaviour. The health anxiety subscale has been used to distinguish hypochondriacal from non-hypochondriacal patients.

Recently, self-assessment measures have been developed to assess the various dimensions of health anxiety and hypochondriasis. The Health Anxiety Inventory contains 47 items covering a range of hypochondriacal features.[17] An advantage of this scale is that it distinguishes patients with high health anxiety from those with physical illness.

The Structured Clinical Interview for DSM-IV (SCID) and the Composite International Diagnostic Interview (CIDI) are comprehensive diagnostic interviews that contain somatoform disorder modules. The CIDI has been used in epidemiologic surveys. Its stem question for hypochondriasis is, 'In the past 12 months, have you had a period of 6 months or more when most of the time you worried about having a serious physical illness or deformity?'

Based on the SCID, Barsky *et al.*[10] developed a structured interview that focuses exclusively on hypochondriasis. It begins with a series of probe questions that, if answered affirmatively, trigger the remaining interview. It is suitable for confirming the diagnosis in a screened population.

Diagnostic assessment remains less than satisfactory because the threshold for caseness has not been established, medical and psychiatric comorbidity make diagnostic decision-making difficult, and independent medical evaluation is rarely part of the process.

Table 5.2.5.3 Measures for the assessment of hypochondriasis

Self-rated questionnaires
Whiteley index
Illness worry scale
Illness attitude scales
Health anxiety questionnaire
Health anxiety inventory
Multidimensional inventory of hypochondriacal traits
Psychiatric diagnostic screening questionnaire
Structured interviews
Structured diagnostic interview for hypochondriasis
Structured clinical interview for DSM-IV
Composite international diagnostic interview
Schedules for clinical assessment in neuropsychiatry

Differential diagnosis

Physical disorders

A few hypochondriacal patients suffer from undetected physical disease. Consequently, it is important to exclude medical conditions that, in their early stages, may cause vague symptoms with few signs or laboratory abnormalities. These include neurological conditions, such as multiple sclerosis or myasthenia gravis; endocrine conditions, such as thyroid or parathyroid disorders; multisystem disease such as systemic lupus erythematosus or occult malignancies. Because of such possibilities, a physical cause warrants continuing consideration even after the initial work-up has been completed.

Psychiatric disorders

Patients with **panic disorder** may be difficult to distinguish from those with hypochondriasis because they commonly have hypochondriacal features. A diagnosis of hypochondriasis should not be made if illness concerns are better accounted for by panic disorder. Patients with hypochondriasis tend to fear the long-term consequences of illness (such as cancer) whereas those with panic fear the immediate consequences of illness events (such as a heart attack); the former fear death, the latter dying. Also, those with hypochondriasis misinterpret a range of bodily sensations, whereas those with panic misinterpret the symptoms of autonomic arousal.

Hypochondriasis must be distinguished from **specific phobia, illness subtype**.[1] Patients with hypochondriasis are preoccupied with a disease they believe is already present, whereas illness phobics fear developing a disease they do not yet have. Illness phobic symptoms are triggered by external as well as internal cues. For instance, exposure to a person with the feared disease may elicit a fear response.

Hypochondriasis must be distinguished from **obsessive-compulsive disorder**. Patients with the latter often have intrusive thoughts about disease or contamination and rituals that involve

checking or reassurance-seeking. They differ from patients with hypochondriasis in having other obsessions and compulsions. Obsessive-compulsive patients tend to regard their ideas as senseless and resist them, whereas those with hypochondriasis regard them with conviction.

Hypochondriasis must also be distinguished from **generalized anxiety disorder** which is characterized by excessive worry about a number of areas. These may include health but other areas are generally involved as well. If worry is confined to illness, then a diagnosis of GAD should not be made. Patients with GAD tend to have health worries that are general, whereas those with hypochondriasis involve specific diseases such as cancer.

Hypochondriasis that develops during an episode of **major depression** and remits with treatment of the mood disturbance may be better accounted for by the depressive disorder. In that case, the patient is likely to focus concern upon the vegetative symptoms of depression and interpret these as irreversible loss of health. On the other hand, a diagnosis of hypochondriasis may be appropriate when hypochondriacal concerns are not confined to an episode of depression and are not focused on symptoms of the mood disorder.

Hypochondriasis and **somatization disorder** are both characterized by somatic symptoms. However, patients with hypochondriasis worry about the meaning of symptoms rather than the symptoms themselves. They are concerned about the consequences of serious illnesses rather than securing the gains of illness (e.g. sick role) as are patients with somatization disorder. Patients with hypochondriasis have an equal sex distribution whereas those with somatization disorder are predominantly women.

Hypochondriacal beliefs of a delusional nature may occur in patients with psychoses, but these patients usually have other psychotic features. However, delusions of disease may be the main or only manifestation of **delusional disorder, somatic type**. Such delusions may be bizarre or unrealistic, whereas the beliefs of patients with hypochondriasis are overvalued.

Epidemiology

Prevalence

The prevalence of hypochondriasis in the **general population** has not been established. Major surveys of psychiatric disorders have either excluded the somatoform disorders or identified few cases. For instance, Looper and Kirmayer[18] found that 6 per cent responded affirmatively to screening for illness worry, but only 0.2 per cent met full criteria for hypochondriasis according to a structured interview. Two studies that focused exclusively on somatoform disorders obtained higher estimates (4.5 and 7.7 per cent).[19, 20] Two other surveys focusing on illness worry found that half the respondents with such worry had the illness they worried about.[18, 21] Among such people it may be difficult to distinguish excessive from normal worry.

The prevalence of hypochondriasis among **primary care outpatients** had been examined in a number of studies. In a cross-national survey, Gureje *et al.*[22] noted that, if the criterion of failure to respond to reassurance were set aside, 2.2 per cent of patients qualified for this diagnosis and were as impaired as those meeting full criteria. In studies based on structured interviews, prevalence estimates have ranged from 2.2 per cent to 9.4 per cent.

Hypochondriasis may be prevalent in **medical specialty populations** where patients with functional disturbances are common. For instance, one survey found the disorder in 13 per cent of otolaryngology clinic patients. Also, hypochondriacal concerns are higher in patients with functional than with organic illnesses. For example, in one study higher hypochondriasis scores were obtained from patients with irritable bowel syndrome than from patients with organic gastrointestinal disease. Hypochondriacal concerns and health anxiety are especially high in patients with chronic pain.

High health anxiety is one of the factors shared by functional somatic syndromes in the general population.[23] However, it is not clear whether this represents a vulnerability factor or a consequence of unexplained symptoms.

Risk factors

Risk factors for unexplained somatic symptoms include female gender, older age, non-white race, less education, and lower income. With respect to hypochondriasis, few of these demographic factors appear important although findings have been inconsistent. The risk for men appears to be equal that for women. Some studies have shown persons with illness worry and hypochondriasis to be older and to have more physical illness. Two studies found them to have less education.

Comorbidity

Hypochondriacal patients in primary care have high levels of psychological as well as somatic symptoms. Strong positive correlations have been observed between hypochondriacal concerns and depressive ($r = 0.58$), anxiety ($r = 0.55$), and somatic symptoms ($r = 0.52$). In one study, the proportions of hypochondriacal and control patients, having one or more comorbid disorder, were 62 and 30 per cent respectively. Anxiety and depressive disorders accounted for most of the excess.

Family and twin studies

Taylor *et al.*[24] used a twin study to examine the genetic and environmental contribution to excessive health anxiety. After controlling for medical morbidity, which may be a source of health anxiety, they found that genetic factors accounted for 37 per cent of the variance in fear of disease and 10 per cent in disease conviction. For both dimensions the remainder of the variance (63 and 90 per cent respectively) was accounted for by non-shared environmental factors. These and other results suggest that some dimensions of health anxiety are moderately heritable. They also suggest that such anxiety is largely a learned phenomenon.

A family study compared the first-degree relatives of probands with and without hypochondriasis obtained from a general medicine clinic.[25] No difference in the frequency of hypochondriasis was found between these groups of family members. However, certain traits and attitudes, such as hostility, low agreeableness, and dissatisfaction with care, were significantly higher among the relatives of hypochondriasis probands. Such traits and attitudes may confer vulnerability to hypochondriasis and/or other somatoform disorders.

Morbidity and service utilization

Hypochondriasis is associated with impairment in physical functioning and work performance. Patients with this disorder view their health as worse, and experience more physical disability as well as impairment in occupational roles than patients without hypochondriasis.[10, 11] They use more medical services yet are less satisfied

with them than non-hypochondriacal patients. This increased utilization includes physician visits, laboratory tests, outpatient costs, and hospitalizations. Hypochondriacal patients tend to feel that their medical problems have not been thoroughly evaluated and as a consequence consult many physicians (i.e. doctor-shopping).

Hypochondriasis and health anxiety tend to be associated with increased symptom reporting and functional impairment, although the findings from various clinical populations have been inconsistent. For instance, hypochondriacal concerns are associated with higher disability and lower quality of life among patients with irritable bowel syndrome, chronic fatigue, and fibromyalgia.[26] One study found hypochondriasis the strongest predictor of pain due to osteoarthritis, and another showed high health anxiety predictive of abdominal pain 1 year later. Hypochondriasis was also a predictor of disability in patients with coronary artery disease. Consistent with these observations, hypochondriasis is associated with increased reporting of, and distress from, medication side effects.

Aetiology and pathogenesis

Personality

Hypochondriacal concerns are strongly related to the major personality dimension of neuroticism or negative emotionality.[27] Positive correlations between neuroticism and hypochondriacal concerns ranging from 0.4 to 0.5 have consistently been observed in non-clinical samples. Neuroticism refers to a tendency to experience and report negative emotions and overreact to stress. Persons high on this dimension are prone to find bodily sensations noxious and interpret them as signs of serious illness. Neuroticism may represent a vulnerability factor for hypochondriasis.

Certain personality traits may have more to do with difficult patient–doctor relationships than with hypochondriasis itself. Patients with hypochondriasis have been described as angry and mistrustful. Such characteristics might reflect negative emotions belonging to the domain of neuroticism or the negative pole of agreeableness, another of the major personality dimensions. They might also reflect obsessive-compulsive or masochistic personality traits observed in some patients.

Developmental factors

Childhood influences appear to be important in the development of hypochondriasis. Reports of **traumatic events during childhood**, including physical and sexual abuse, have been elicited more frequently from hypochondriacal than non-hypochondriacal patients. Although findings are preliminary, they are consistent with a literature linking childhood neglect and abuse to unexplained somatic symptoms in adults.

Childhood experience of illness may contribute to the development of hypochondriasis. For instance, Noyes *et al.*[28] obtained reports of serious illness or injury before age 17 from a third of adults with hypochondriasis. Similar findings from patients with hypochondriasis and somatization have been reported by others. Early illness may create a sense of physical vulnerability in susceptible individuals. Childhood exposure to serious illness or death of a family member or friend may do likewise.

Parental attitudes may also contribute to hypochondriasis. Excessive concern for a child's health or overprotection on the part of a parent may lead to anxiety about health as may special caretaking and rewards for illness. A child may also model exaggerated illness behaviour displayed by a parent. The importance of developmental factors—early adversity, experience of illness, over solicitous parents—suggest that hypochondriasis is in large measure learned behaviour.

Life events

Stressful life events appear to be related to increased reporting of physical symptoms and hypochondriasis, although there have been few studies. Events involving illness and death may have a specific role as the symptoms of hypochondriacal patients sometimes resemble those of family members who have been ill or died. In addition, illness events may give rise to hypochondriacal symptoms; 'cardiac neurosis' following myocardial infarction is an example. Transient hypochondriasis has been observed following medical illness in predisposed individuals.

Cognitive and perceptual factors

According to the cognitive-perceptual model, hypochondriasis is based on misinterpretation of bodily symptoms as signs of serious disease and on the experience of somatic sensations as intense, noxious, and disturbing.[29] In this model, the faulty attribution of innocuous sensations is the central defect. A number of studies have shown that, when symptoms are attributed to pathological processes, they become intensified. Such attribution may focus attention on symptoms thereby amplifying them. Misinterpretation of this kind may arise from cognitive schemata that were established through earlier experience with illness.

The tendency to experience bodily sensations as intense and disturbing has been termed somatosensory amplification. One study that used the Somatosensory Amplification Scale found a positive correlation of 0.56 between amplification and hypochondriasis. This finding suggests that individuals with hypochondriasis have a constitutionally lowered threshold for physical symptoms or that they have a heightened attentional focus and increased physiological arousal.

Evidence of physiological abnormalities was obtained by Gramling *et al.*[30] In a preliminary investigation, they observed physiological reactivity that distinguished women with hypochondriasis from those without. Hypochondriacal subjects had a higher mean heart rate and lower mean hand temperature during a cold pressor test compared to controls. These subjects terminated the test more frequently and rated it as more unpleasant than did controls.

Interpersonal factors

According to the interpersonal model, hypochondriasis is a form of care-eliciting behaviour that finds expression in physical complaints. Through unexplained somatic symptoms and expressions of illness worry, patients with this disorder seek emotional and interpersonal support from family members and physicians. Need for support of this kind arises from insecure attachment that originated in early relationships with caregivers. In a test of this model, Noyes *et al.*[31] found that hypochondriacal concerns among primary care patients were associated with various insecure attachment styles. These concerns were also associated with interpersonal problems and lack of reassurance from medical care.

Social and cultural factors

Social and cultural factors are important determinants of hypochondriasis. Throughout the world physical symptoms are common vehicles for the communication of distress. Somatic distress

gains the attention of **family and community** because it signals impairment in functioning that could alter social roles. Such distress not only calls forth caretaking but also obtains the sick role for those with acute illness. This social role with its privileges and responsibilities protects society from the disruptive effects of illness and promotes the return to health and social functioning of its members. Persons who are socially isolated or lacking in social support are more likely to manifest care-eliciting behaviour such as hypochondriasis.

Physicians play an important role in the development of hypochondriasis. They may make alarming statements or fail to provide reassurance that is based on thorough evaluation. In addition, they may order unnecessary tests, diagnose undetected disease, or treat injudiciously. They may add to concerns by failing to diagnose the psychiatric disturbance—hypochondriasis—telling patients instead that nothing is wrong. In doing this, they challenge and reject patients thereby contributing to suffering and alienation from the health care system.

Cultural attitudes may contribute to hypochondriasis. The American lifestyle, which emphasizes fitness and attractiveness, fosters preoccupation with health and encourages people to see their distress in terms of physical illness. There are, for example, cultural differences in the threshold for pain, pain tolerance, patterns of arousal, and physiological and behavioural responses to pain.

Course and outcome

Course

Hypochondriasis may begin at any age including childhood. The onset may be associated with stressful life events that in some instances involve illness. Some individuals develop hypochondriacal concerns transiently and others lastingly in reaction to physical illness. Among family medicine patients, those who became hypochondriacal a year after initial assessment were found to have had more illness worry and unexplained symptoms and to have rated their health as worse at baseline than non-hypochondriacal patients. Ambiguous symptoms or illness events may contribute to hypochondriacal concerns in patients so predisposed. Hypochondriasis appears to follow a chronic, fluctuating course.

Outcome

Follow-up studies show that, after their initial clinic visit, most patients with hypochondriasis improve. Still, a substantial proportion continue to meet criteria for the disorder and many more have persisting symptoms. For instance, among hypochondriacal general medicine patients, Noyes *et al.*[12] and Barsky *et al.*[13] found that, after 1 to 4 years, two-thirds continued to qualify for the diagnosis and the remaining one-third had persisting symptoms. Thus, despite improvement, the patients continued to be more hypochondriacal, more impaired, and more symptomatic than non-hypochondriacal patients.

Like patients in general, those with hypochondriacal concerns tend to seek care when they are most distressed. Their subsequent improvement may represent a natural fluctuation, a response to physician contact or to non-specific treatment. Some patients report having responded to reassurance. In a few instances, serious medical illness may relieve hypochondriacal concerns by legitimizing symptoms.

Studies indicate that greater severity and longer duration of symptoms are predictive of worse outcome. Failure to remit in one or more follow-up studies was predicted by more severe hypochondriacal concerns and somatic symptoms, longer duration of hypochondriasis, more psychiatric comorbidity, poorer perception of health, and greater neuroticism.

Complications

There is little information concerning complications of hypochondriasis. Because some patients utilize extensive medical care, one might expect complications resulting from repeated or unnecessary evaluations, tests, procedures, or treatments. Such iatrogenic complications have been reported for somatoform disorders but there is little documentation for hypochondriasis. On the other hand, physical illness may be overlooked in patients whose problems are considered psychiatric. There is almost no information on mortality. Suicide is said to be rare in hypochondriasis unless accompanied by severe depression in which case the risk may be increased.

Treatment

Until recently, the treatment of hypochondriasis was regarded with pessimism. It now appears that effective psychological, even pharmacological, interventions are being developed. A variety of approaches have been proposed but controlled trials of cognitive-behavioural therapy have established its efficacy, and preliminary trials of antidepressant medication have shown promise.

Psychological therapies

Most hypochondriacal patients, referred to mental health professionals, receive **psychotherapy** although such treatment has received little study. In one controlled trial a small number of patients with hypochondriasis were randomly assigned to explanatory therapy or a waiting list. The therapy yielded significant improvement in illness behaviour and health care utilization compared to no treatment, and gains were maintained for 6 months. This form of therapy involves repeated physical examinations, reassurance concerning symptoms, and information about psychophysiologic processes. Additional controlled trials of this and other forms of psychotherapy (e.g. psychodynamic, interpersonal) are clearly needed.

Four randomized, controlled trials for patients with hypochondriasis have shown that **cognitive behavioural therapies** are superior to no therapy with benefits sustained for up to 12 months.[32–35] These trials show that psychological treatment is efficacious for referred patients. However, one study showed that behavioural stress management, a non-specific intervention, was effective as well,[33] and another showed that cognitive and behavioural procedures, by themselves, were equally effective.[34]

Cognitive procedures include identifying and challenging dysfunctional thoughts and formulating more realistic beliefs. Behavioural procedures involve exposure *in vivo* with response prevention. These techniques include exposure to feared internal and external stimuli (e.g. physical exercise, visiting sick persons, reading about feared diseases, writing one's obituary) and prevention of checking and reassurance-seeking behaviours.

These trials showed that psychological treatment is effective but leave important questions unanswered. For instance, is psychological

therapy acceptable to most hypochondriacal patients in primary care? Are the techniques specific or do the benefits result from non-specific factors (e.g. therapeutic attention, therapist–patient relationship, credible procedures)? Also, are these treatments cost-effective? One trial involved up to 16 sessions over 4 months, which is expensive in terms of time and resources.

In consideration of these issues, several authors have advocated a **group approach**. For example, one study showed that group treatment is feasible. To improve acceptance, the authors referred to their intervention as a course in stress management and carried it out in a general practice setting.

Pharmacological therapies

There is evidence that patients with secondary hypochondriasis respond to drug therapy for the primary disorder. For example, Noyes *et al.*[36] assessed hypochondriacal concerns in patients receiving pharmacological treatment for panic disorder and agoraphobia. At the completion of treatment, a significant reduction in concerns was observed among those whose anxiety symptoms had improved. Observations of a similar kind have been made in patients with major depression.

No randomized controlled trials of pharmacotherapy for hypochondriasis have yet been completed, but a series of open label studies suggest that medication has promise. For example, Fallon *et al.*[37] reported that 10 or 16 patients with primary hypochondriasis given fluoxetine were very much improved after 12 weeks. And others have reported similar results with paroxetine, fluvoxamine, and nefazodone. Of the more than 50 patients enrolled in these trials, two-thirds responded to an SSRI. In these trials, drugs were relatively well tolerated and few patients dropped out because of side effects. This is noteworthy in view of the sensitivity to adverse effects observed in such patients. Controlled trials are needed to show proof of efficacy in primary hypochondriasis.

Management

Most hypochondriacal patients are best managed by their primary physicians. Few are successfully referred for specialty care because the focus of their concerns is, at least initially, on unexplained somatic symptoms. Although for some the ultimate goal is specific treatment, such treatment is not yet widely available.

Successful management depends upon a trusting relationship with a physician. To establish this, the physician should first **legitimize the patient's symptoms** by listening carefully and completing a thorough evaluation. Respectful treatment and statements to the effect that unexplained symptoms are nonetheless real are often helpful (see Table 5.2.5.4).

The scheduling of **regular visits** is an important strategy. Such visits serve several purposes. First, they reduce the reward for more severe or new symptoms that patients often present at unscheduled calls or visits. Next, they assure patients that the physician has an ongoing interest in their well-being. Finally, they provide reassurance through continued health monitoring.

Physicians should use **restraint in evaluating** hypochondriacal patients. New symptoms must be thoroughly evaluated, but overly aggressive diagnostic evaluation can be counterproductive. Extensive and dramatic tests can generate alarm, and when testing is repeated, it may convey physician uncertainty. Physicians should also avoid making diagnoses simply to have something to treat.

Physicians should also **approach treatment cautiously**. Medications, even when prescribed for benign indications, cause patients to worry about the conditions for which they are given. And too often they result in intolerable side effects and iatrogenic complications.

Hypochondriacal patients need an **explanation for their distress**, one that counters the notion of serious disease. Patients may be told their problem lies in the central nervous system processing of bodily sensations; this means they have a sensitive nervous system that amplifies discomforts and dysfunctions. Such an explanation gives legitimacy to the problem and avoids the stigmatizing label of hypochondriasis. Alternatively, patients may be told that they suffer from excessive health anxiety or worry. An explanation of the role of anxiety in altering attention, amplifying bodily sensations, and generating physiological symptoms may also be acceptable.

The goal of medical management is not to remove symptoms but help patients cope with them. The expectations of patients seeking elimination of symptoms may need to be modified. Reduced dependence on the technical aspects of care (namely, diagnostic testing and corrective intervention) is an important aspect of this overall objective. Patients need assistance in managing their lives so as to minimize continuing symptoms. The **aim is improved functioning**, a greater sense of control, and improved self-esteem. These objectives may accompany a gradual return to work and meaningful activity, and may be enhanced by improvements in exercise, diet, and daily routine.

Developing and **maintaining a therapeutic relationship** with the hypochondriacal patient is often challenging. The patient may be mistrustful and feel that his or her suffering is not understood. Masochistic and obsessional personality traits may contribute to a difficult doctor–patient relationship. A patient with such traits may seek mistreatment and thwart the physician's attempts to be helpful. Yet, a positive relationship is the key to successful management and can be achieved with acceptance, empathy, and understanding.[38]

Table 5.2.5.4 Management strategies for patients with hypochondriasis

Legitimize the patient's symptoms
Establish a regular schedule of visits
Base diagnostic evaluation on objective findings
Approach treatment of physical symptoms cautiously
Provide a plausible explanation for symptoms
Establish a goal of improved functioning

Specific treatment

Many hypochondriacal patients have psychiatric comorbidity, and treatment of comorbid anxiety and depressive disorders may yield significant improvement. If hypochondriasis has arisen during the course of an anxiety or depressive disorder, then successful treatment of the primary disorder may bring remission of hypochondriacal symptoms.

Specific treatments, to be acceptable, must be available in the primary care setting. Treating professionals must let patients know that their concerns are legitimate and their suffering understood. Beyond this, they must place a premium on engaging the patient, techniques for which have been described by a number of authors.

These patients are prone to drug side effects and often discontinue medication. For this reason, initial doses should be small, with gradual increases according to a modifiable schedule. Treating physicians should acknowledge the patient's sensitivity, and indicate that side effects are likely but may be dealt with.

Hypochondriasis is a significant medical condition for which treatment is now available.

Further information

Asmundson, G.J.G., Taylor, S. and Cox, B.J. (eds.) (2001). *Health anxiety: clinical and research perspectives on hypochondriasis and related disorders*. John Wiley, New York.

Lipsitt, D.R. and Starcevic, V. (eds.) (2001). *Hypochondriasis: modern perspectives on an ancient malady*. Oxford University Press, New York.

Taylor, S. and Asmundson, G.J.G. (2004). *Treating health anxiety: a cognitive-behavioral approach*. Guilford Press, New York.

References

1. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
2. Gillespie, R.D. (1928). Hypochondria: its definition, nosology, and psychopathology. *Guy's Hospital Report*, **8**, 408–60.
3. Creed, F. and Barsky, A. (2004). A systematic review of the epidemiology of somatization disorder and hypochondriasis. *Journal of Psychosomatic Research*, **56**, 391–408.
4. Starcevic, V. (2001). Clinical features and diagnosis of hypochondriasis. In *Hypochondriasis: new perspectives on an ancient malady* (eds. V. Starcevic and D.R. Lipsitt), pp. 21–60. Oxford University Press, New York.
5. Barsky, A.J., Ahern, D.K., Bailey, E.D., *et al.* (2001). Hypochondriacal patient's appraisal of health and physical risks. *American Journal of Psychiatry*, **158**, 783–7.
6. World Health Organization. (1993). The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research. World Health Organization, Geneva.
7. Mayou, R., Kirmayer, L.J., Simon, G., *et al.* (2005). Somatoform disorders: time for a new approach in DSM-V. *American Journal of Psychiatry*, **162**, 847–55.
8. Rief, W. and Sharpe, M. (2004). Somatoform disorders—new approaches to classification, conceptualization and treatment. *Journal of Psychosomatic Research*, **56**, 387–90.
9. Barsky, A.J., Wyshak, G., and Klerman, G.L. (1986). Hypochondriasis: an evaluation of the DSM-III criteria in medical outpatients. *Archives of General Psychiatry*, **43**, 493–500.
10. Barsky, A.J., Cleary, P.D., Wyshak, G., *et al.* (1992). A structured diagnostic interview for hypochondriasis: a proposed criterion standard. *Journal of Nervous and Mental Disease*, **180**, 20–7.
11. Noyes, R., Kathol, R.G., Fisher, M.M., *et al.* (1993). The validity of DSM-III-R hypochondriasis. *Archives of General Psychiatry*, **50**, 961–70.
12. Noyes, R., Kathol, R.G., Fisher, M.M., *et al.* (1994). One-year follow-up of medical outpatients with hypochondriasis. *Psychosomatics*, **35**, 533–45.
13. Barsky, A.J., Fama, J.M., Bailey, D., *et al.* (1998). A prospective 4- to 5-year study of DSM-III-R hypochondriasis. *Archives of General Psychiatry*, **55**, 737–44.
14. Speckens, A.E.M. (2001). Assessment of hypochondriasis. In *Hypochondriasis: new perspectives on an ancient malady* (eds. V. Starcevic and D.R. Lipsitt), pp. 61–88. Oxford University Press, New York.
15. Pilowsky, I. and Spence, N.D. (1983). *Manual for the illness behavioural questionnaire* (*IBQ*) (2nd edn.). Department of Psychiatry, University of Adelaide, Adelaide, South Australia.
16. Kellner, R. (1986). *Somatization and hypochondriasis*. Praeger, New York.
17. Salkovskis, P.M., Rimes, K.A., Warwick, H.M.C., *et al.* (2002). The health anxiety inventory: development and validation of scales for the measurement of health anxiety and hypochondriasis. *Psychological Medicine*, **32**, 843–53.
18. Looper, K. and Kirmayer, L.J. (2001). Hypochondriacal concerns in a community population. *Psychological Medicine*, **31**, 577–84.
19. Faravelli, C., Salvatori, S., Galassi, F., *et al.* (1999). Epidemiology of somatoform disorders: a community survey in Florence. *Social Psychiatry and Psychiatric Epidemiology*, **32**, 24–9.
20. Noyes, R., Happel, R.L., and Yagla, S.J. (1999). Correlates of hypochondriasis in a nonclinical population. *Psychosomatics*, **40**, 461–78.
21. Noyes, R., Carney, C.P., Hillis, S.L., *et al.* (2005). Prevalence and correlates of illness worry in the general population. *Psychosomatics*, **46**, 529–39.
22. Gureje, O., Üstün, T.B., and Simon, G.E. (1997). The syndrome of hypochondriasis: a cross-national study in primary care. *Psychological Medicine*, **27**, 1001–10.
23. Aggarwal, V.R., McBeth, J., Zakrzewska, J.M., *et al.* (2006). The epidemiology of chronic syndromes that are frequently unexplained: do they have common associated factors? *International Journal of Epidemiology*, **35**, 468–76.
24. Taylor, S., Thordarson, D.S., Jang, K.L., *et al.* (2006). Genetic and environmental origins of health anxiety: a twin study. *World Psychiatry*, **5**, 47–50.
25. Noyes, R., Holt, C.S., Happel, R.L., *et al.* (1997). A family study of hypochondriasis. *Journal of Nervous and Mental Disease*, **185**, 223–32.
26. Robbins, J.M., Kirmayer, L.J., and Kapusta, M.A. (1990). Illness worry and disability in fibromyalgia syndrome. *International Journal of Psychiatry in Medicine*, **20**, 49–63.
27. McClure, E.B. and Lilienfeld, S.O. (2001). Personality traits and health anxiety. In *Health anxiety: clinical and research perspectives on hypochondriasis and related conditions* (eds. G.J.G. Asmundson, S. Taylor, and B.J. Cox), pp. 65–91. John Wiley, New York.
28. Noyes, R., Stuart, S., Langbehn, D.R., *et al.* (2002). Childhood antecedents of hypochondriasis. *Psychosomatics*, **43**, 282–9.
29. Salkovskis, P.J. and Warwick, H.M.C. (2001). Meaning, misinterpretations, and medicine: a cognitive-behavioural approach to understanding health anxiety and hypochondriasis. In: *Hypochondriasis: modern perspectives on an ancient malady* (eds. V. Starcevic and D.R. Lipsitt), pp. 202–22. Oxford University Press, New York.
30. Gramling, S.E., Clawson, E.P., and McDonald, M.K. (1996). Perceptual and cognitive abnormality model of hypochondriasis: amplification and physiological reactivity in women. *Psychosomatic Medicine*, **58**, 423–31.
31. Noyes, R., Stuart, S. Langbehn, *et al.* (2003). Test of an interpersonal model of hypochondriasis. *Psychosomatic Medicine*, **65**, 292–300.
32. Warwick, H.M.C., Clark, D.M., Cobb, A.M., *et al.* (1996). A controlled trial of cognitive behavioral treatment of hypochondriasis. *British Journal of Psychiatry*, **169**, 189–95.
33. Clark, D.M., Salkovskis, P.M., Hackman, A., *et al.* (1998). Two psychological treatments for hypochondriasis: a randomized controlled trial. *British Journal of Psychiatry*, **173**, 218–25.
34. Visser, S. and Bouman, T.K. (2001). The treatment of hypochondriasis: a randomized controlled trial. *Behavior Research and Therapy*, **39**, 423–42.
35. Barsky, A.J. and Ahern, D.K. (2004). Cognitive behavioral therapy for hypochondriasis: a randomized controlled trial. *Journal of the American Medical Association*, **291**, 1464–70.

36. Noyes, R., Reich, J., Clancy, J., *et al.* (1986). Reduction in hypochondriasis with treatment of panic disorder. *British Journal of Psychiatry*, **149**, 631–5.
37. Fallon, B.A., Liebowitz, M.R., Salman, E., *et al.* (1993). Fluoxetine for hypochondriacal patients without major depression. *Journal of Clinical Psychopharmacology*, **13**, 438–41.
38. Starcevic, V. (2002). Overcoming therapeutic pessimism in hypochondriasis, *American Journal of Psychotherapy*, **56**, 167–77.

5.2.6. Pain disorder

Sidney Benjamin and Stella Morris

Introduction

Persistent somatoform pain disorder is an ICD-10 diagnosis, which is included in the group of somatoform disorders. The term **pain disorder** is used in DSM-IV, and for convenience that is the term used here to refer to both classifications, unless a distinction needs to be made. This chapter aims to clarify the relationship of pain to mental disorders, the diagnosis of pain disorder and its differential diagnosis, and then considers how psychosocial factors contribute to pain, the treatments that stem from them, and the psychiatrist's potential contribution.

Pain has been defined by the International Association for the Study of Pain (**IASP**) as 'an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage'. 'Pain' is used here in this sense; it is not used primarily to indicate mental distress or anguish. As a perception, pain is essentially a subjective experience, and is directly accessible only to the patient. By contrast, tissue damage can be assessed by others, and its relationship with the subjective characteristics of pain have been shown to be variable, modulated by social and cultural experience, as well as within the central and peripheral nervous system.

Pain and the psychiatrist

Psychiatrists are likely to see patients with pain in psychiatric, general hospital, and community settings. Pain is associated with a wide range of mental disorders, and there are different ways in which this relationship may arise.

Pain may contribute to the cause of a mental disorder; for example, when a patient with cancer has pain, which is unrelieved by analgesics, and becomes depressed. This can result in additional distress and disability, and subsequently an exacerbation of pain. Treatment of depression may contribute to the relief from pain and improve the quality of life.

In a general hospital psychiatrists may see patients with **acute pain**, like the patient described above, but more often will see patients with **chronic pain**. Whatever the initial cause, the longer pain persists the more likely is it to result in the development of inappropriate patterns of illness behaviour and to have a profound effect on relationships with the family and other carers, presenting more complex challenges for management and poorer prognosis.

Pain disorder

Diagnostic and clinical features

Persistent somatoform pain disorder in **ICD-10** is the only somatoform disorder that is essentially characterized by pain. The **diagnostic requirements** are as follows:

1 'persistent, severe, and distressing pain';

2 pain 'cannot be explained fully by a physiological process or a physical disorder';

3 'pain occurs in association with emotional conflict or psychosocial problems that are sufficient to allow the conclusion that they are the main causative influences'.

There are also likely to be many of the features that occur in the other somatoform disorders, which have been described in previous chapters. The pain can be localized, as in low back pain, or generalized, as in fibromyalgia.

In **ICD-10**, the diagnosis is excluded if pain, presumed to be mainly psychological in origin, occurs in the course of schizophrenia or depressive disorder, or is believed to be due to psychophysiological mechanisms such as muscle tension. The main **differential diagnosis**, according to ICD-10, is the histrionic elaboration of pain primarily due to organic causes, particularly if this has not yet been diagnosed. In practice, it is uncommon for pain that has been properly investigated, and has persisted for more than 6 months, to be found subsequently to have a specific organic cause.

The **DSM-IV** diagnosis of '**pain disorder**' also needs to be considered because the requirements for diagnosis and the underlying rationale are rather different. This diagnosis is divided into three **subtypes**:

1 '**Pain disorder associated with psychological factors**', in which psychological factors are judged to play the major role, and physical disorders play either no part or only a minor part in its onset or maintenance.

2 '**Pain disorder associated with both psychological factors and a general medical condition**', in which both psychological processes and an organic disorder are judged to make important contributions to causation.

3 '**Pain disorder associated with a general medical condition**', due to an organic disorder and in which psychological factors are judged to make no contribution or to play only a minor role. This subtype is not regarded as a mental disorder but is coded on Axis III.

For the first two subtypes the diagnostic criteria, all of which must be satisfied, are summarized as follows:

(a) Pain, localized or more general, is the predominant symptom and its severity warrants clinical attention

(b) Pain results in distress, and impairment in social, occupational, or other areas of functioning.

(c) Psychological factors are judged to have an important role in the onset, severity, exacerbation, or maintenance of pain.

(d) It is not intentionally produced or feigned (factitious disorder and malingering are specifically excluded).

(e) Pain is not better accounted for by a mood, anxiety, or psychotic disorder and does not meet criteria for dyspareunia.

Pain disorder can also be coded according to whether it is acute or chronic (less or more than 6 months duration).

Comparison of ICD-10 and DSM-IV

The diagnoses of pain disorder in ICD-10 and DSM-IV share a number of characteristics. Pain disorder should be diagnosed as a mental disorder if psychological factors are thought to make a significant contribution to predisposition, precipitation, or maintenance, or to the severity of pain. In ICD-10, there should be evidence that emotional conflict or psychosocial problems are the main 'causative influences', whereas in DSM-IV psychological factors are judged to play either the 'major role' or 'an important role'. In both, the diagnosis can be made even though there may be possible or definite evidence of an organic disorder that contributes to pain (for instance, a prolapsed intervertebral disc), provided that this is judged to be insufficient to account fully for the features of pain. Both classifications stress the severity of pain and the distress caused by it, but only DSM-IV specifically requires a degree of disability as a diagnostic feature. The implication is that diagnosis requires detailed physical and psychiatric evaluation, including an assessment of the family and social context, as well as of disability.

Differential diagnosis of pain disorder

Pain can occur in the setting of virtually any mental disorder. Table 5.2.6.1 lists the ICD-10 diagnoses and their DSM-IV equivalents in which pain may be a predominant feature. The general description of most of these disorders is provided in other chapters of this book and the following account focuses only on aspects relevant to pain.

(a) Organic disorders

Many painful disorders have a well-recognized organic pathology that accounts for the occurrence of pain (for example, angina, sickle cell arthropathy), but psychosocial processes tend to modify the severity of pain and associated disability. Thus, psychological and social interventions may make an important contribution to management, and as pain becomes more chronic, or fails to respond to usually effective physical treatments, psychosocial interventions assume greater significance. These disorders can be diagnosed in ICD-10 within the diagnoses headed '**Psychological interactions with physical disorders**' in Table 5.2.6.1.

(b) Pain syndromes of uncertain origin

There are many disorders characterized by pain, which are essentially syndromes with no known consistent organic pathology (Table 5.2.6.2). Psychological and social factors are thought to contribute to the development and maintenance in many cases,[1] but psychological causes specific to these different syndromes have not been identified. Patients with these pain syndromes tend to have a greater prevalence of non-psychotic mental disorders than is found in the general population. The pain itself can usually be accommodated in ICD-10 within the categories of **somatoform autonomic dysfunction** or **somatoform pain disorder** (see below). The 'diagnoses' listed in Table 5.2.6.2 tend to be used by non-psychiatrists to describe clusters of medically unexplained symptoms and are terms which are likely to be acceptable to patients. Treatments for these disorders generally include physical approaches, often of limited efficacy, as well as a range of psychosocial interventions, which are described below.

(c) Pain and mental disorders

(i) Psychoses

At the beginning of the twentieth century, French psychiatrists described **coenestopathic states** as disorders characterized by unpleasant sensations, particularly pains, thought to be of central origin, but unrelated to organic brain disease.[2] Such disorders were a daily occurrence in psychiatric clinics, commonly associated with the psychoses, and in this setting were related to **somatic hallucinations** and **systematized delusional states**. Such presentations are now described infrequently in Europe and North America.

Patients with any psychosis may complain of pain, sometimes with bizarre descriptions of quality and **delusional attribution**. In practice, it is difficult to differentiate between a **somatic hallucination** and an illusion (arising from physiological or pathological processes). Complaints of pain in psychotic disorders have no psychiatric diagnostic specificity. Pain has been described particularly in association with schizophrenia and depressive psychoses, but may occur in any psychotic disorder. In the course of a psychotic disorder, illusions and delusional interpretations of pain may arise from unrelated organic disorders and therefore require careful **physical assessment**.

(ii) Mood- and anxiety-related disorders

These are by far the most common mental disorders associated with pain in most settings. In the general population, 12 per cent of adults have experienced **chronic widespread pain** (defined according to the criteria of the American College of Rheumatologists) in the previous 3 months and their prevalence of mental disorders is three times that of the pain-free population.[3] Most of these diagnoses are mood and anxiety disorders, with the former being more common in those with chronic pain. In **pain clinic settings**, the prevalence of mental disorders varies according to referral patterns, but about 30 to 40 per cent of patients have depressive disorders, and this is similar in those with and those without a relevant physical disorder.[4] Those without organic disorders tend to have lower ratings for both mood disorders and pain severity. Those with mood disorders report more severe pain.

Diagnosis of mood and anxiety disorders is based on the usual standardized criteria, but may be missed due to the process of **somatization**, particularly where patients attribute their depressed mood to pain and an underlying physical condition (whether present or not) and invite their doctors to share this belief.[5] In the past, pain has been thought of as a proxy for depression, giving rise to the concept of a '**depressive equivalent**' or '**masked depression**'. This has been based mainly on evidence for the psychogenicity of chronic pain rather than a specific relationship to depressive disorders, has received widespread criticism, and has not advanced theoretical knowledge or clinical practice.

(iii) Post-traumatic stress disorder

Many patients with post-traumatic stress disorder (PTSD) have been subjected to actual or threatened physical injury, so it is not surprising that pain is one of the commonest symptoms that they report, the prevalence ranging from 20 to 80 per cent. Further, 10 to 50 per cent of patients with chronic pain satisfy criteria for PTSD, and patients with musculoskeletal pain are four times more likely to develop PTSD than those without it.[6] **Pain disorder and PTSD** can be diagnosed jointly, if criteria for both are satisfied. Mechanisms including shared vulnerability, fear-avoidance, and

Table 5.2.6.1 Mental disorders included in the differential diagnosis of pain disorder

ICD-10		DSM-IV	
Psychotic disorders			
F00–09	Organic mental disorders	290	Dementia 293 Delirium
F20–29	Schizophrenia, schizotypal, and delusional disorders	273	Schizophrenia and other psychotic disorders
Mood- and anxiety-related disorders			
F32/33	Depressive episode	296.2/3	Major depressive disorder
F34.1	Dysthymia	300.4	Dysthymic disorder
F41	Anxiety disorders	300.02	Generalized anxiety disorder
F43.1	Post-traumatic stress disorder	309.81	Post-traumatic stress disorder
F43.2	Adjustment disorders	309	Adjustment disorders
Somatoform disorders			
F44.4	Dissociative (conversion) disorders	300.11	Conversion disorder
F45.0	Somatization disorder	300.81	Somatization disorder
F45.1	Undifferentiated somatoform disorder	300.81	Undifferentiated somatoform disorder
F45.2	Hypochondriacal disorder	300.7	Hypochondriasis
F45.3	Somatoform autonomic dysfunction	300.8	Pain disorder
F45.4	Somatoform pain disorder	300.81	Somatoform disorder NOS
F45.8	Other somatoform disorders		
F45.9	Somatoform disorder, unspecified		
Other neurotic disorders			
F48.0	Neurasthenia		
F48.8	Other specified neurotic disorders (occupational neurosis, e.g. writer's cramp)		
Sexual disorders			
F52.5	Non-organic vaginismus	306.51	Vaginismus
F52.6	Non-organic dyspareunia	302.76	Dyspareunia
Psychological interactions with physical disorders			
F54	Pyschological or behavioural factors associated with disorders or diseases classified elsewhere	316	Psychological factors affecting medical condition
F68.0	Elaboration of physical symptoms for psychological reasons		
Disorders of behaviour			
F68.1	Intentional production or feigning of symptoms	300.19	Factitious disorder
		V65.2	Malingering
Comorbidity of pain disorder			
Any of the above except psychoses and other somatoform disorders			
Substance abuse			
F10	Disorders due to alcohol	291 & 303.9	Alcohol-induced disorders and dependence
F11–13	Disorders due to psychoactive substance abuse	292 & 304	Other substance-induced disorders and dependence
F55	Abuse of non-dependence-producing substances		
Personality disorders			
F60–62	Personality disorders and changes	301	Personality disorders

mutual maintenance have been postulated to account for this comorbidity.[7] This has implications for assessment (described below), and treatment programmes may need to be modified accordingly.

(iv) Somatoform disorders

Somatoform disorders are uncommon in people with chronic pain in the general population.[3] Prevalence varies considerably in clinical samples, but somatoform disorders have been reported in 12 to 52 per cent of patients,[4] so they include highly selected samples.

Complaints of pain occur commonly in each of the somatoform disorders and may be the predominant symptom. Multiple physical complaints, often including pains at different sites, fluctuate from time to time usually for many years, providing a characteristic feature of **somatization disorder**. In **hypochondriacal disorder** pain is a common complaint, and forms the focus for concern and overvalued beliefs about unidentified disease.

The diagnosis of **somatoform autonomic disorder** is based on autonomic arousal (palpitation, sweating, tremor), which must be a prominent feature of the clinical picture, together with physical complaints, often pain, referred to specific organs, systems, or parts of the body. As with other somatoform disorders, the patient will be distressed about the possibility of underlying physical disease and is not reassured by negative findings on appropriate assessment and explanation. This diagnosis is sometimes appropriate for syndromes listed in Table 5.2.6.2.

Pain, as a form of **conversion**, has a traditional place in the literature on **hysteria**, based on the concepts of psychogenicity, the

Table 5.2.6.2 Disorders of uncertain origin, presenting primarily with pain, in which psychosocial factors are thought to contribute to predisposition, precipitation, or course

Generalized
Fibromyalgia
Relatively localized
Tension headache—acute or chronic
Temporomandibular pain and dysfunction syndrome
Atypical facial pain
Atypical (non-cardiac) chest pain
Abdominal pain of psychological origin
Non-ulcer dyspepsia
Irritable bowel syndrome
Chronic pelvic pain
Irritable bladder syndrome
Procatalgia fugax

contribution of stressful experiences with dissociation, and primary gain. In recent years, however, research has focused on other psychological processes, and the concept of conversion as a primary mechanism now seems to be of limited interest. The category of **dissociative (conversion) disorder** in ICD-10 specifically includes sensory loss but excludes pain (sensory amplification), which therefore should not be diagnosed as a dissociative disorder. DSM-IV also excludes pain from the diagnosis of conversion disorder, unless other diagnostic criteria are satisfied.

The uncertain relationship and limited value of the different diagnoses included within the group of somatoform disorders in ICD-10 have been discussed in Chapter 5.2.1, and are well illustrated by the fact that pain may be a prominent feature of each category. Somatoform disorders presenting with pain are usually diagnosed as **somatization disorder or pain disorder**, with the former taking precedence if the diagnostic criteria are satisfied.

Comorbidity

Any physical or mental disorder may be diagnosed in addition to pain disorder. Anxiety and depression are common, and an additional diagnosis of **anxiety disorder** or **mood disorder** can be made if the criteria are satisfied. This dual diagnosis can be useful if, for example, a depressive disorder develops in the presence of a long-standing pain disorder. Any temporal relationship can occur, however, with pain onset preceding, developing simultaneously with, or following the onset of a mood disorder.

Other common comorbid diagnoses include **substance abuse** and dependence, sometimes of iatrogenic origin, and their management is an important component of pain-treatment programmes.[8] **Personality disorders** are an additional category of comorbidity. No single disorder predominates but histrionic, narcissistic, anxious (avoidant), and dependent features are all common in clinical practice, and anankastic traits may feed an inflexible focus on physical illness.

Epidemiology

Although the association of psychiatric symptoms with chronic pain has been studied in the general population, the prevalence of pain disorder, and other mental disorders presenting with pain, is uncertain because large-scale surveys of mental disorders do not include an assessment of pain and of related physical conditions.

Assessment of pain

Clinical assessment

The psychiatric assessment requires a full **psychiatric history** and **mental state examination**, with particular attention to those additional features relevant to pain. **The pain history** should include total duration (often underestimated by the patient), a detailed inquiry about the location and distribution of pain, including direct questions aimed at a total body survey, and the timing of first onset, subsequent periods of relapses and remissions, and their relationship to life events and difficulties. The **family history** should include assessment of severe, chronic or disabling physical disorders, and the patient's involvement with them. The **personal history** should include adverse childhood experiences (discussed below) and the **past history** of physical disorders and disability is particularly important.

Patients who **somatize** will tend to deny concurrent psychosocial events and their significance. For example, one of our patients was consistently unable to recall any distressing events in the year prior to the onset of severe, persistent, and disabling headache. His wife gave an account of the deaths of his father, brother, and closest friend during that year, and moreover described him as so distressed by these bereavements that he felt unable to attend any of the funerals. It is essential to take a **history from other informants**, and this can also provide an opportunity to assess the attitudes, knowledge, and beliefs of carers, and their interaction with the patient.

The patient's **pain beliefs and behaviours** (described below) are key aspects of the **mental state examination**. Patients often attribute chronic pain to an organic disorder and offer diagnoses; it is essential to review their **medical records** to assess the clinical findings and investigations, and the extent to which they support any diagnosis which is offered. Chronic pain associated with an underlying organic disorder may be exacerbated when the patient suffers a stressful life event, so it is important to **avoid assumptions of a dichotomy** of either **'organic' or 'psychogenic' pain**.

Standardized psychometric assessments

Many standardized questionnaires have been developed for the assessment of patients with chronic pain. They can be valuable for identifying mechanisms that contribute to pain, planning treatment, and monitoring changes during and after treatment. The evaluation of pain and associated beliefs and behaviours requires measures developed specifically for this purpose, and these are described below.

Other assessments, for example of **mood**, **illness behaviour**, and **social dysfunction**, have been developed within the field of pain research. Some measures are rather idiosyncratic, with uncertain psychometric properties, aimed at restricted diagnostic groups and clinical settings. This undermines the need to use consistent methods that allow comparison of different groups of patients, with physical, mental, and mixed disorders, at different places and times.

(a) Pain

The **severity** of pain can be assessed[9] using standardized **visual analogue scales** and **numeric analogue scales**. Such scales may have anchor points ranging from 'no pain' to 'the worst possible pain'.

The **quality** of pain can be assessed with **verbal descriptor scales**.[9] Factor analysis has resulted in the emergence of two that have best survived the test of time: an '**affective**' **dimension** (represented by words such as exhausting, terrifying, vicious), and a '**sensory**' **dimension** (e.g. stabbing, crushing, burning). They have been found consistently when administered in different languages and to different cultural groups. Ratings on both these scales are positively correlated with pain severity and mood ratings and, in the presence of mental disorders, contribute little to diagnosis.

The **topographical distribution of pain** can be assessed by using outline drawings of the body (front, back, and sometimes sides), which the patient is asked to shade to indicate the distribution of pain. These can help to identify pain that does not conform to physiological distributions and also widespread pain. Measures of pain intensity, quality, and distribution can be used toget her to capture the rather elusive and entirely subjective experience of pain.

(b) Pain behaviours

Although the experience of pain is entirely personal, it may be communicated to others by a range of verbal and non-verbal behaviours, which in some cases may be maladaptive, and which in turn influence the responses of others. Using a **learning theory model**, Fordyce[10] classified all pain into **'operant' and 'non-operant' pain**. The former includes all pain that is modified by positive or negative reinforcement, whether or not organic pathology is present. Standardized **structured assessments** are available to measure a range of well-defined **behaviours**.[11] These may include complaints of pain, requests for medication, groaning, facial expression, restricted mobility and the use of aids, time spent resting, and postures such as guarding and bracing. Such behaviours have been shown to fluctuate in response to changes in the environment, including different **attitudes and responses of carers**. This has led to the assessment of pain behaviours and their **environmental reinforcers**, and the development of pain-treatment programmes that originally focused on behavioural change by modifying reinforcement. Recent interest has focused on **pain-related fears** (e.g. of exacerbating pain by injury) and the management of consequent **avoidance**.[12]

(c) Pain beliefs

The belief that chronic as well as acute pain signals an underlying physical disease, which requires and should respond to physical intervention, whilst avoiding usual activities and functions, contributes to the development and maintenance of chronic pain and non-adherence to treatment, and the widespread dissatisfaction often expressed by patients and their doctors. Inappropriate beliefs that are relevant to pain assessment fall into three groups[13]:

1 beliefs about the nature of reality—for example, 'life should be pain-free';

2 beliefs in response to challenging circumstances, such as pain—including **locus of control**, **attributional style**, **cognitive errors**, and **coping strategies**;

3 specific ideas about the cause of a pain, appropriate management, and outcome.

The questionnaire assessment of pain-related beliefs has assumed increasing importance in the field of pain research,[13] with the recognition that pain beliefs interact with pain, cognitions, behaviours, affects and disability, and contribute to the prediction of outcome. Thus cognitive approaches to treatment are often integrated with behavioural management.

Psychosocial contributions to the development of pain

The origins of chronic pain are, in several respects, similar to those of somatization and other somatoform disorders. Current models of causation involve the interaction of biological, psychological, and social factors, each contributing to predisposition, precipitation, and maintenance.

The **family and personal histories** of patients with chronic pain include an excess of mood disorders, pain and disability, substance abuse, and personality disorders. Engel[14] described the dynamics of 'the **pain prone patient**' involving abusive childhood experiences, and noted how pain can become a pathway for the expression of guilt and expiation. Recent research[15] has reconsidered the significance of reports of **physical and sexual abuse** and other **adverse childhood experiences**. The relationship between chronic pain in adults and these childhood experiences appears, at least to some extent, to be determined by **selective reporting**, particularly in those with associated mental disorders,[16] but these experiences may make a significant contribution to pain in some individuals.

Precipitation of chronic pain is, in effect, **transition from acute to chronic pain**, and factors associated with this transition[17,18] include current mood and anxiety disorders, negative life events including physical illnesses and trauma, the social support network and dissatisfaction with work. A population based prospective study[19] found that new episodes of chronic widespread pain were predicted by the number of previous non-pain somatic symptoms and by a measure of illness behaviour which assessed numbers of consultations, treatments and perceived disability, and these two measures had an additive effect. Recent research has indicated the potential value of interventions designed to prevent the progression from acute to chronic pain.[17,18]

The psychiatric and psychological management of pain

Treatment of mental disorders

The treatment of chronic pain has presented a challenge to the ingenuity of health professionals, particularly because no single specialty or profession has the range of skills that is required. The treatment of mental disorders,[5] such as depressive or anxiety disorders,[20] is similar in most respects, whether or not pain is a prominent feature. In the presence of pain, however, mental disorders tend to be missed, and when recognized are treated inadequately. Depressive disorders with features indicating a good response to **antidepressants** should be treated with full therapeutic doses, but not with **narcotics**. **Anxiolytic drugs** including **benzodiazepines**, which result in dependence, should not be used in the treatment of these chronic disorders.

The use of antidepressants for pain relief

Antidepressant drugs are often used for the treatment of pain in patients who are not depressed. Randomized controlled trials[21] indicate that antidepressants, in doses within the usual therapeutic

range, provide more effective **analgesia** than placebo preparations in the treatment of diabetic neuropathy, postherpetic neuralgia, and atypical facial pain, as well as chronic non-malignant pain. Different **tricyclic antidepressants** (TCAs) appear to be equally effective and are more effective than **selective serotonin-reuptake inhibitors**. Data on **seretonin noradrenaline-reuptake inhibitors** are increasing and suggest that they may be effective, and preferable to TCAs because of a superior side effect profile.[22] The analgesic effect of antidepressants occurs in patients who are not depressed and is independent of any antidepressant effect.

Psychological treatments

Psychological treatments[5,23] are derived from different theoretical formulations of the aetiology of chronic pain. These include behavioural, cognitive, and psychodynamic approaches. Reviews of randomized controlled trials of **behavioural and cognitive approaches**[24] that have been developed specifically for the treatment of chronic pain illustrate the problems in assessing outcome due to different sampling methods, different types of control groups, non-standardized treatment components, and the different assessments that are included. Despite these limitations, the best studies demonstrate that these treatments are more effective than 'usual' medical treatment, remaining on a waiting list, or exercise programmes, and improvements can be sustained during lengthy follow-up periods.

Other approaches include various forms of 'stress management' including **relaxation techniques**, **biofeedback**, and **hypnosis**. Their value is uncertain; although pain ratings tend to be reduced, this is not a consistent finding on all measures.

Psychological treatments are rarely used in isolation, either from each other or from additional interventions, and integrating different approaches may enhance their effects.[23]

Multi-disciplinary pain management clinics

Pain-treatment centres[25] have been established in many countries and provide a diverse range of professional skills, treatments, and models of service delivery. In some, management is based mainly on anaesthetic techniques and medication, but psychological approaches are provided in others by **clinical psychologists** and **nurse therapists**. The management of problems due to inappropriate medication and **substance abuse** is an essential component of treatment.[8] Many clinics offer **structured programmes** of education and rehabilitation, with increments of **exercise**, to overcome **disability**, rather than aiming primarily at pain relief, and to which **physiotherapists or occupational therapists** may contribute. There is often an emphasis on the patient assuming increasing responsibility, rather than maintaining **dependence on medical services**.

Direct input from **psychiatrists** is variable and some centres specifically exclude the treatment of patients with serious mental disorders because their response to treatment is less certain. Although it is well recognized that social and environmental factors contribute to chronic pain problems, and can undermine progress following treatment, few specialized centres involve **carers** routinely in treatment or offer **family therapy**.

A range of physical, psychological, and social approaches should be offered, based on an **individual structured assessment** of needs. Members of the multi-disciplinary team require specific **training** in the management of pain. The work of the team has to be carefully coordinated, both within the team and with other health professionals, to avoid any ambiguity concerning the methods and goals of treatment.

Many pain clinics provide a treatment package in which cognitive therapy and graded exercise are predominant features. A similar approach is used for a number of other conditions, including **somatization disorder**, **hypochondriacal disorder**, **fibromyalgia**, and **chronic fatigue**, but the extent to which they may have similar origins and outcomes is uncertain.

Effects of treatment

The **outcome of psychological and psychiatric treatment** has been studied extensively,[5,24] but is difficult to evaluate because reports differ with regard to the characteristics of patients and disorders, inclusion criteria, assessments, and treatments as well as details of treatment delivery, attrition rates, choice of control groups, and the duration of follow-up. Many patients with chronic pain are unwilling to accept treatment and others are considered unsuitable. Nevertheless, psychological and rehabilitation treatments can have a sustained effect, based on the range of assessments that have been described. In addition, they can result in reductions in **sickness and benefit payments**,[26] **return to work**[27] and reduced **use and costs of medical services**.[28]

The outcome for patients with different mental disorders has not been assessed systematically. Patients involved in seeking **compensation** tend to have a poorer outcome, even after **litigation** has been concluded, but they can also benefit from treatment. There is some evidence that **secondary prevention** programmes may help to avoid the transition from acute to chronic pain in those who are particularly vulnerable.

Further information

For more information on the topic of this chapter, we have marked with an asterisk (*) those references, which will be of particular interest to the reader.

References

1. *Kellner, R. (1991). *Psychosomatic syndromes and somatic symptoms*. American Psychiatric Press, Washington, DC.
2. Dupré, E. (1913). *Les cénestopathies*. Reprinted as: Coenestopathic states. In *Themes and variations in European psychiatry* (eds. S.R. Hirsch and M. Shepherd), pp. 385–94. John Wright, Bristol, 1974.
3. Benjamin, S., Morris, S., McBeth, J., *et al.* (2000). The association between chronic widespread pain and mental disorder. *Arthritis and Rheumatism*, **43**, 561–7.
4. Benjamin, S., Barnes, D., Berger, S., *et al.* (1988). The relationship of chronic pain, mental illness and organic disorders. *Pain*, **32**, 185–95.
5. Benjamin, S. and Main, C.J. (1995). Psychiatric and psychological approaches to the treatment of chronic pain: concepts and individual treatments. In *Treatment of functional somatic symptoms* (eds. R. Mayou, C. Bass, and M. Sharpe), pp. 188–213. Oxford University Press.
6. Asmundson, J.G., Coons, M.J., Taylor, S., *et al.* (2002). PTSD and the experience of pain; research and clinical implications of shared vulnerability and mutual maintenance models. *Canadian Journal of Psychiatry*, **47**, 930–7.
7. *Otis, J.D., Keane, T.M., and Kerns, R.D. (2003). An examination of the relationship between chronic pain and post-traumatic stress disorder. *Journal of Rehabilitation Research and Development*, **40**, 397–406.

8. British Pain Society. (2006). *Pain and substance misuse: improving the patient experience*. British Pain Society, London.
9. *Williams, A.C.deC. (2004). Assessing chronic pain and its impact. In *Psychosocial aspects of pain: a handbook for healthcare professionals, progress in pain research and management*, Vol. 27 (eds. R.H. Dworkin and W.S. Breitbart). IASP Press, Seattle.
10. Fordyce, W.E. (1985). The behavioural management of chronic pain: a response to critics. *Pain*, **22**, 113–25.
11 Keefe, F.J. and Williams, D.A. (1992). Assessment of pain behaviours. In *Handbook of pain assessment* (eds. D.C. Turk and R. Melzack), pp. 277–92. Guilford Press, New York.
12 De Jong, J.R., Vlaeyen, J.W.S., Onghena, P., *et al.* (2005). Reduction of pain-related fear in complex regional pain syndrome type I: the application of graded exposure in vivo. *Pain*, **116**, 264–75.
13. *DeGood, D.E. and Shutty, M.S. (1992). Assessment of pain beliefs, coping and self-efficacy. In *Handbook of pain assessment* (eds. D.C. Turk and R. Melzack), pp. 214–34. Guilford Press, New York.
14. Engel, G. (1959). 'Psychogenic' pain and the pain prone patient. *The American Journal of Medicine*, **26**, 899–918.
15. Morley, S. (2004). What impact does childhood experience have on the childhood development of chronic pain? In *Psychosocial aspects of pain: a handbook for healthcare professionals, progress in pain research and management*, Vol. 27 (eds. R.H. Dworkin and W.S. Breitbart). IASP Press, Seattle.
16. McBeth, J., Morris, S., Benjamin, S., *et al.* (2001). Associations between adverse events in childhood and chronic widespread pain in adulthood: are they explained by differential recall? *The Journal of Rheumatology*, **28**, 2305–9.
17. *Poleshuck, E.L. and Dworkin, R.H. (2004). Risk factors for chronic pain in patients with acute pain and their implications for prevention. In *Psychosocial aspects of pain: a handbook for healthcare providers, progress in pain research and management*, Vol. 27 (eds. R.H. Dworkin and W.S. Breitbart). IASP Press, Seattle.
18. *Linton, S.J. (2004). Environment and learning factors in the development of chronic pain and disability. In *Psychological methods of pain control: basic science and clinical perspectives, progress in pain research and management*, Vol. 29 (eds. D.D. Price and E.M. Bushnell). IASP Press, Seattle.
19. McBeth, J., Macfarlane, G.J., Benjamin, S., *et al.* (2001). Features of somatization predict the onset of chronic widespread pain. Results of a large population based study. *Arthritis and Rheumatism*, **44**, 940–6.
20. *Gallagher, R.M. and Verma, S. (2004). Mood and anxiety disorders in chronic pain. In *Psychosocial aspects of pain: a handbook for healthcare professionals, progress in pain research and management*, Vol. 27 (eds. R.H. Dworkin and W.S. Breitbart). IASP Press, Seattle.
21. Atkinson, J.H., Meyer, J.M., and Slater, M.A. (2004). Principles of psychopharmacology in pain treatment In *Psychosocial aspects of pain: a handbook for health care professionals, progress in pain research and management*, Vol. 27 (eds. R.H. Dworkin and W.S. Breitbart). IASP Press, Seattle.
22. Sindrup, S., Otto, M., Finnerup, N., *et al.* (2005). Antidepressants in the treatment of neuropathic pain. *Basic and Clinical Pharmacology and Toxicology*, **96**, 399–409.
23. *Waters, S.J, Campbell, L.C., Keefe, F.J., *et al.* (2004). The essence of cognitive-behavioral pain management. In *Psychosocial aspects of pain: a handbook for healthcare professionals, progress in pain research and management*, Vol. 27 (eds. R.H. Dworkin and W.S. Breitbart). IASP Press, Seattle.
24. Morley, S., Eccleston, C., and Williams, A. (1999). Systematic review and meta-analysis of randomised controlled trials of cognitive behaviour therapy and behaviour therapy for chronic pain in adults, excluding headache. *Pain*, **80**, 1–13.
25. Cohen, M.J.M. and Campbell, J.N. (eds.) (1996). *Pain treatment centers at a crossroads: a practical and conceptual reappraisal*. IASP Press, Seattle.
26. Thomsen, A.B., Sorensen, J., and Sjogren, P. (2002). Chronic non-malignant pain patients and health economic consequences. *European Journal of Pain*, **6**, 341–52.
27. Haldorsen, E.M.H., Grasdal, A.L., Skouen, J.S., *et al.* (2002). Is there a right treatment for a particular patient group? Comparison of ordinary treatment, light multidisciplinary treatment, and extensive multidisciplinary treatment for long-term sick-listed employees with musculoskeletal pain. *Pain*, **95**, 49–63.
28. Peters, L., Simon, E.P., Folen, R.A., *et al.* (2000). The COPE program: treatment efficacy and medical utilisation outcome of a chronic pain management program at a major military hospital. *Military Medicine*, **165**, 954–60.

5.2.7 Chronic fatigue syndrome

Michael Sharpe and Simon Wessely

Introduction

Chronic fatigue syndrome is a controversial condition, conflicts about which have frequently burst out of the medical literature into the popular media. Whilst these controversies may initially seem to be of limited interest to those who do not routinely treat such patients, they also exemplify important current issues in medicine. These issues include the nature of symptom-defined illness; patient power versus medical authority; and the uncomfortable but important issues of psychological iatrogenesis.[1,2] The subject is therefore of relevance to all doctors.

Fatigue as a symptom

Fatigue is a subjective feeling of weariness, lack of energy, and exhaustion. Approximately 20 per cent of the general population report significant and persistent fatigue, although relatively few of these people regard themselves as ill and only a small minority seek a medical opinion. Even so, fatigue is a common clinical presentation in primary care.[2]

Fatigue as an illness: chronic fatigue syndrome

When fatigue becomes chronic and associated with disability it is regarded as an illness. Such a syndrome has been recognized at least since the latter half of the last century. Whilst during the Victorian era patients who went to see doctors with this illness often received a diagnosis of neurasthenia, a condition ascribed to the effect of the stresses of modern life on the human nervous system the popularity of this diagnosis waned and by the mid-twentieth century it was rarely diagnosed (although the diagnosis subsequently became popular in the Far East—see Chapter 5.2.1). Although it is possible that the prevalence of chronic fatigue had waned in the population, it is more likely that patients who presented in this way were being given alternative diagnoses. These were mainly the new psychiatric syndromes of depression and anxiety, but also other labels indicating more direct physical explanations, such as chronic brucellosis, spontaneous hypoglycaemia, and latterly chronic Epstein–Barr virus infection.[2]

As well as these sporadic cases of fatiguing illness, epidemics of similar illnesses have been occasionally reported. One which

occurred among staff at the Royal Free Hospital, London in 1955 gave rise to the term myalgic encephalomyelitis (ME), although it should be emphasized that the nature and symptoms of that outbreak are dissimilar to the majority of those now presenting to general practitioners under the same label.

A group of virologists and immunologists proposed the term chronic fatigue syndrome in the late 1980s.[3] This new and aetiologically neutral term was chosen because it was increasingly recognized that many cases of fatigue were often not readily explained either by medical conditions such as Epstein–Barr virus infection or by obvious depression and anxiety disorders. Chronic fatigue syndrome has remained the most commonly used term by researchers. The issue of the name is still not completely resolved however: Neurasthenia remains in the ICD-10 psychiatric classification as a fatigue syndrome unexplained by depressive or anxiety disorder, whilst the equivalent in DSM-IV is undifferentiated somatoform disorder. Myalgic encephalomyelitis or (encephalopathy) is in the neurological section of ICD-10 and is used by some to imply that the illness is neurological as opposed to a psychiatric one. Unfortunately the case descriptions under these different labels make it clear that they all reflect similar symptomatic presentations, adding to confusion. Official UK documents have increasingly adopted the uneasy and probably ultimately unsatisfactory compromise term CFS/ME.[4] In this chapter, we will use the simple term chronic fatigue syndrome (CFS).

Clinical features

Symptoms

Chronic mental and physical fatigue, tiredness, or exhaustion that is typically exacerbated by activity is the core symptom of CFS. Commonly associated symptoms include impaired memory and concentration, muscular and joint pain, unrefreshing sleep, dizziness and breathlessness, headache, tender lymph glands, and sore throat. Patients often describe day-to-day fluctuations in symptoms, irrespective of activity. Periods of almost complete recovery may be followed by relapse, often described as sufficiently severe to make normal daily activity impossible. Depression and anxiety are common, and a proportion of patients suffer panic attacks.

Physical signs

Physical examination is typically unremarkable. Complaints of fever and lymphadenopathy are not confirmed on examination. The presence of definite physical signs (such as objectively measured fever) should not be ascribed to the syndrome and alternative diagnoses should be sought.

Other common characteristics

As well as the symptoms described above patients with CFS commonly have additional clinical characteristics. These are listed in Table 5.2.7.1.

Patients are often worried that remaining active despite fatigue will harm them and consequently avoid activity or oscillate between rest and bursts of activity, which produces fatigue, leading to a return to rest and so on.

Some patients feel strongly that their illness is 'medical' rather than 'psychiatric' and are particularly concerned that a psychiatric diagnosis implies that the illness is their fault, an indication of personal weakness or even an accusation of malingering. Perfectionist and high achieving lifestyles often with low underlying self-esteem are commonly observed in patients referred to hospital clinics.

Table 5.2.7.1 Common characteristics of patients with CFS

Thoughts beliefs and attitudes	Thought that symptoms indicate harm Belief that the illness is purely 'medical' Perfectionist attitudes
Coping behaviours	Avoidance of activities associated with symptoms Reduced activity level Oscillation in overall activity level
Physiology	Poor sleep Physiological deconditioning Effects of inactivity
Interpersonal and social	Dependence on carer Psychological iatrogenesis Occupational difficulties

Although there are no physical signs there may be measurable effects of reduced activity with so-called physiological deconditioning leading to poor tolerance of activity, and in cases where rest has been prolonged other physiological changes such as postural hypotension. Sleep is often unrefreshing and fragmented.

Some patients can become markedly dependent on a carer. Occupational stresses and difficulties are common and it can be difficult to determine if these were contributors to, or are consequence of their illness. Finally many patients have received unhelpful medical attention. Such psychological iatrogenesis includes, on the one hand dismissal of their complaints and on the other over investigation.[5]

Case study

A typical patient is found in the infectious disease department of the general hospital. She is a 30-years-old nurse and her principal complaints are of fatigue, poor concentration, and muscle pain. Her symptoms fluctuate and are made worse by physical and mental exertion. She is no longer able to work and has substantially reduced her daily activities. The history is of an acute onset of symptoms after a 'viral illness'. Enquiry reveals symptoms suggestive of depression or anxiety, but without obvious mood change. The patient strongly believes the illness to be 'medical' rather than 'psychiatric'.

Classification and diagnosis

There are several published case definitions for CFS. The currently most widely used definition is based on an international consensus of researchers is shown in Table 5.2.7.2.[6] A guide on its application has also been published.[7] It should be remembered that this definition represents nothing more than a working definition of a clinical problem, pending further understanding, and as with most psychiatric diagnoses, does not delineate a single disease.

Issues for a definition of chronic fatigue syndrome

The case definition shown in Table 5.2.7.2 has been useful in unifying the field and providing a widely used operational definition. However, it also has significant limitations.

Table 5.2.7.2 International consensus definition of chronic fatigue syndrome

1	Complaint of fatigue
	Of new onset
	Not relieved by rest
	Duration at least 6 months
2	At least four of the following additional symptoms
	Subjective memory impairment
	Sore throat
	Tender lymph nodes
	Muscle pain and joint pain
	Headache
	Unrefreshing sleep
	Post-exertional malaise lasting more than 24 h
3	Impairment of functioning
4	Other conditions that might explain fatigue excluded

(Reproduced from Fukuda, K. Straus, S.E. Hickie, I.B. *et al.* Chronic fatigue syndrome: a comprehensive approach to its definition and management, *Annals of Internal Medicine*, **121**, 953–9. Copyright 1994, The American College of Physicians.)

- It excludes fatigue associated with known organic disease.
- It overlaps with other functional medical diagnoses.
- It overlaps with psychiatric diagnosis.
- The homogeneity of the patient group it identifies is doubtful.

(a) Differentiation from fatigue associated with organic disease

Fatigue is a common symptom of most medical and psychiatric conditions. CFS refers only to fatigue where there is no clear alternative diagnosis (but does not exclude depression and anxiety unless the depression is of melancholic type or a manifestation of a bipolar disorder). It therefore only refers to idiopathic fatigue. This means that the definition highlights an important clinical problem but also means that the interesting equally important and probably informative phenomenon of fatigue in patients with diseases such as multiple sclerosis is excluded from this definition.

(b) Overlap with other medically unexplained syndromes

A number of medical diagnoses are defined only by symptoms. These functional syndromes are medical diagnoses where there is no identifiable pathology. They include chronic pain, fibromyalgia, and irritable bowel syndrome. Although chronic pain syndromes are principally characterized by pain, fibromyalgia by tender points, and irritable bowel syndrome by symptoms of bowel disturbance, all these syndromes are also associated with chronic fatigue, and patients diagnosed with one of these syndromes often meet the diagnostic criteria for CFS.[8]

(c) Overlap with psychiatric syndromes

Most patients who meet criteria for CFS also fulfil criteria for a psychiatric diagnosis. Many meet criteria for anxiety and depressive disorders and others merit diagnoses of somatoform disorder or neurasthenia. This issue is discussed further below.

(i) Depression

If patients with a depressive disorder are asked about a wide range of somatic symptoms including fatigue and/or muscle pain (which they are usually not) they often report these. If the diagnostic criteria for depressive disorders are applied to patients with fatigue a high proportion meet these.[9] Furthermore the prevalence of major depressive disorder in patients referred to hospital with CFS is substantially higher than in patients with chronic disabling medical diseases suggesting that depression is not simply a reaction to disability.[10] In practice, the diagnosis of depression can be difficult in patients presenting with fatigue: depressed mood is often not prominent and anhedonia can be hard to distinguish from the inability to pursue previously enjoyed activities because of fatigue. Finally, whilst there is a strong association between major depressive disorder and CFS, for as many as half of the patients seen in hospital clinics the symptoms cannot be readily given that diagnosis.

(ii) Anxiety disorders

Although less attention has been given to the association between fatigue and anxiety, an examination of diagnostic criteria for anxiety disorders reveals that the typical somatic symptoms of anxiety include fatigue and other symptoms listed as typical of CFS. If sought, generalized anxiety disorder can often be diagnosed in patients with CFS and panic can often be diagnosed in patients with severe episodic symptoms.[11] As with depression, however, anxious mood is rarely obvious and may be hard to distinguish from reasonable concern about consequences of being ill. Likewise, true phobic avoidance may be hard to distinguish from the consequences of fatigue and/or weakness.

(iii) Neurasthenia

ICD-10 differs from DSM-IV in including this diagnosis. It requires that the patient suffers from fatigue which is exacerbated by exertion, as well as several other somatic symptoms, and does not meet the criteria for a depressive or anxiety disorder (see Chapter 5.2.10). One study found that almost all of the referrals to a medical CFS clinic met the criteria for neurasthenia as defined by ICD-10.[12]

(iv) Somatoform disorders

According to DSM-IV patients with severe persistent fatigue who do not meet criteria for anxiety or depressive disorders are assigned to a somatoform disorder diagnosis. These are a controversial group of psychiatric syndromes characterized by medically unexplained symptoms and of presumed psychological origin.[13] There are a number of subcategories:

- Somatization disorder (Briquet's syndrome) is used to describe patients who report multiple, recurrent, medically unexplained symptoms; a minority of patients with CFS will meet the criteria for this disorder.
- Hypochondriasis describes a syndrome in which the patient's main concern is with the possibility that they are suffering from an organic disease. Whilst this diagnosis would seem to be applicable to many patients with CFS, it is problematic when the cause of the illness in question, which is regarded as uncertain by doctors as well as patients.
- Almost all patients with CFS not meeting the criteria for any of the above DSM disorders are likely to fall into the undemanding residual category in DSM-IV of 'undifferentiated somatoform disorder'. This diagnosis is of dubious practical use, and in effect merely confirms that the patient has multiple physical symptoms of unclear aetiology.

(v) Conclusion

Many patients with CFS meet the diagnostic criteria for a depressive or anxiety disorder, although in practice the presentation is often 'atypical'. It is likely that patients who do not meet the criteria for either of these could be diagnosed as having either neurasthenia (ICD-10) or undifferentiated somatoform disorder (DSM-IV).

Should we use the diagnosis of CFS?

From the psychiatrist's perspective it is parsimonious to ask whether a diagnosis of CFS is ever necessary or appropriate when the symptoms can always be described by a psychiatric diagnosis? This unsatisfactory situation is an artefact of parallel medical and psychiatric diagnostic systems for patients with somatic symptoms unexplained by disease. Consequently whether one uses a 'medical' diagnosis of CFS or a 'psychiatric' diagnosis of somatoform disorder is merely a matter of choice. When making that choice the following must be considered:

- A diagnosis of CFS only describes a presenting clinical syndrome, rather than a specific disorder or disease process.
- Pragmatically the relative acceptability of the alternative diagnosis to the patient is important. There is no point in giving a diagnosis that is rejected by the patient and impedes any therapeutic relationship and chances of treatment.
- One approach to overcoming the issue of parallel classification systems is to combine the medical diagnosis of CFS and the psychiatric diagnoses: According to such a scheme CFS would be subclassified into CFS/depression, CFS/anxiety, and CFS without depression or anxiety disorder (i.e. CFS/somatoform or CFS/neurasthenia). Psychiatric diagnoses that have important clinical utility such as major depressive disorder should obviously be made if present. The usefulness of diagnoses such as undifferentiated somatoform disorder is less clear.
- Finally rather than becoming side-tracked by the unanswerable question of whether the patients symptoms are ultimately 'medical' or 'psychiatric' in nature an open-minded and pragmatic approach is required.

Epidemiology

Prevalence

Fatigue is common but CFS is rare. As it can be difficult to differentiate CFS from depressive and anxiety disorders estimates that have attempted to exclude these diagnoses are lower than those that have not. Population studies in the United Kingdom and United States suggest that only approximately 0.5 per cent of the population can be regarded as having CFS.[1] Most of these persons are aged between 20 and 40 with a predominance of females. The syndrome is also seen in children and adolescents but less commonly.

Epidemics

Epidemics of a chronic fatigue-like syndrome have been described from various parts of the globe. This observation is compatible with, but does not establish, an infective cause. It remains unclear whether these were true epidemics and also whether the clinical picture reported is similar to that of cases of sporadic chronic fatigue syndrome.

Aetiology

Limitations in the available data

Although a considerable amount of research has been devoted to investigating the nature and causes of CFS there are few firm conclusions that can be drawn. This partly because many of the studies have had major methodological shortcomings:

- Patients have often been recruited from tertiary care clinics, using various diagnostic criteria inconsistently applied.
- Only a minority of studies have included comparison groups of patients with diagnoses of depression or anxiety disorders.
- Because most studies have used a case–control design, it is often impossible to know whether the findings they report are causes or consequences of the illness (for instance, as the result of reduced activity or sleep disturbance).
- CFS is almost certainly heterogeneous.

Considering these caveats there are a number of areas where positive findings have been reported.

Pathophysiology

Clinical observations of patients with CFS have led to the investigation of a number of hypotheses about the underlying pathophysiological mechanisms.

(a) Genetics

Twin studies suggest that CFS is moderately heritable.[1] Preliminary studies suggesting the involvement of specific genes require replication. Gene expression studies are likewise in their infancy—one problem being that the number of genes studied usually exceeds by several orders of magnitude the numbers of patients studied. Another problem is confounders—a large Swedish twin registry study for example suggested that genetic factors contributed to the risk of CFS both directly and via personality type.[14]

(b) Cardiovascular and respiratory abnormalities

Several investigators have reported abnormalities in the cardiorespiratory systems that may underpin the exercise intolerance. Hyperventilation has been suggested as a mechanism of symptom production, but only a minority of patients have biochemically confirmed hyperventilation. Low blood pressure has long been associated with the symptom of fatigue, and in some parts of Europe unexplained fatigue is confidently ascribed to this. Postural hypotension has been noted in some patients and whilst this may be a cause of fatigue it may also be a consequence of inactivity. Finally, various abnormalities in cardiac function have also been reported but are of uncertain significance.

(c) Infection

Perhaps because patients commonly describe their illness as beginning with 'flu-like symptoms' and because of the apparent epidemics, many investigators have sought objective evidence of an initiating or ongoing viral infection. Viral infection can probably initiate CFS. A prospective follow-up of people with positive evidence of acute Epstein–Barr infection did find that some patients went on to develop a fatigue syndrome.[15] Other infectious agents that may trigger CFS include Q fever and viral meningitis. If viruses do play a role in precipitating CFS, it would appear that it is only when certain types of viruses infect vulnerable persons. If CFS can be

precipitated by viral infection does persistence of the virus cause the ongoing symptoms? On current evidence the answer to this question seems to be no.

(d) Immune dysfunction

The evidence for an association between immunological abnormalities and CFS is more consistent than that for infective agents, with several studies suggesting abnormalities in lymphocyte numbers and function. However, similar changes can be found in patients with depressive disorders, and, although some studies have attempted to control for emotional disorder, both the specificity and causal importance of these observations remain unclear.[16]

(e) Sleep

Unrefreshing sleep is an almost ubiquitous complaint of people suffering from CFS. While studies have identified major sleep disorders such as sleep apnoea and narcolepsy as alternative diagnoses for small number of patients with daytime fatigue, simple disruption of slow-wave sleep is a much more common observation. While inefficient sleep could contribute to the daytime fatigue reported in both conditions, its specificity and aetiological role are uncertain.[16]

(f) Neuroendocrine and neurotransmitter abnormalities

The prominent fatigue of Addison's disease has led to the hypothesis that adrenal function is impaired in patients with CFS. In support of this suggestion there is reasonable evidence that patients with chronic fatigue and fibromyalgia have both low levels of cortisol, a point of difference from major depression. However it remains unclear if this apparent abnormality is cause or effect of the illness and associated inactivity. Patients with CFS also have evidence of abnormal functioning of cerebral serotonergic systems, which differ from those found in patients with depression. Like the abnormalities in adrenal function these findings are preliminary but of potential interest.[17]

(g) Brain imaging

Finally, a variety of techniques have been used to examine both the function and structure of the brain in patients with CFS. Cerebral perfusion studies have shown abnormalities, although similar, if not identical, abnormalities are also found in patients with depression. Possible white-matter changes reported on magnetic resonance scans are more controversial, and harder to interpret.[1,16]

(h) Conclusions

Despite a considerable research effort, so far no single pathophysiological process has been conclusively identified as causal of CFS. There is some evidence for a loss of physical fitness and possibly for abnormalities of neuroendocrine function. Viral infection can play a role as precipitating agent, although its importance as a perpetuating factor is less certain. Immunological abnormalities are common but of uncertain specificity, and appear not to be related to chronic symptoms. The current attention on neuroendocrine function takes the focus of investigation closer to those features known to be associated with depressive states. However, the evidence suggests that the changes in neurotransmitter and neuroendocrine function in patients with CFS may differ from those commonly observed in patients diagnosed with depressive disorder. Further studies are needed to confirm all these abnormalities and to clarify whether they are causal or merely epiphenomena.

Psychopathology

If there are no substantial biological abnormalities are there psychological ones? The initial psychiatric explanation of CFS was that it was misdiagnosed depressive disorder. However, whilst such misdiagnosis does occur, more complex explanations are required to adequately explain many cases of CFS.

(a) Somatization

It has been hypothesized that depression may still be 'behind' CFS even if not apparent. It is argued that a process referred to as 'somatization' (making the mental physical) is responsible. Whist the idea that the somatic symptoms of CFS are readily understandable as part of an emotional disturbance is a parsimonious alternative to some of the more elaborate mechanisms outlined above, as there is no 'marker' for somatization this hypothesis is hard to prove.

(b) Attribution

Patients attending specialist clinics with CFS typically attribute their illness to organic disease even when no evidence of this can be found by their physicians. Perhaps more importantly, some strongly resist psychological explanations for their illness, although most take a more mixed view. Whether these patients are biased in their views about illness or simply wiser than their physicians is unclear. However, strong and exclusively physical disease attributions may be a marker for an important illness-perpetuating process in CFS as they predict a poorer clinical outcome.

(c) Perceptual processes

Patients with CFS report a greater sense of effort in response to both psychological and physical demands than is explicable from objectively measured impairments. This observation raises the possibility that they are especially sensitive to bodily sensations such as effort, that is they 'amplify' or focus on them. Thus, one may hypothesize that as in panic disorder, the patients' beliefs about their symptoms may lead them to focus attention on to bodily sensations. Although plausible there is so far only limited evidence that this process is important in patients with CFS.[1]

(d) Coping behaviour

A tendency to avoid activities that exacerbate symptoms has been shown to occur in patients with CFS. The avoidance may be persistent or episodic in response to exacerbations of symptoms resulting in a 'boom and bust' pattern. Avoidance is associated with persistent disability, and has been suggested as the mechanism by which disease attributions for symptoms predicts poor outcome.[1]

(e) Personality characteristics

Both research studies and clinical experience suggest that many persons with CFS have a tendency towards hard driving, perfectionist, or obsessive–compulsive personalities, and associated overactive lifestyles. Evidence from the UK 1946 birth cohort for example indicates that ratings of physical activity in early life predict later CFS. In a large prospective Swedish twin registry measure of stress and emotionality are consistently associated with subsequent CFS. Those CFS sufferers may be predisposed to becoming physically and emotionally exhausted, and biased towards presenting emotional distress in a somatic form.[14]

(f) Stigma, misinformation, and iatrogenesis

Psychiatric diagnoses are stigmatized in the popular mind as indicating weakness or even unreal illness. Patients with CFS may be

susceptible to those social pressures and consequently prefer a medical diagnosis for their distress and inability to function. It has also been suggested that CFS may serve as culturally defined function of social communication, allowing a socially acceptable and hence 'non-psychiatric' expression of distress and protest about intolerable occupational and personal pressures. Much the same was said of neurasthenia in the People's Republic of China (see Chapter 5.2.1).

Another potentially important social factor is the controversy and the often misleading information about the illness that patients are exposed to. Self-help books the media and some doctors have frequently given the impression that the medial profession is more divided than it actually is in its understanding of CFS and have emphasized 'medical' explanations such as myalgic encephalopathy (ME) as the only appropriate diagnosis.

(g) Conclusions

Psychological and social factors are important perpetuating factors and include focusing of attention on symptoms avoidance of activity and a strong and exclusive medical disease attribution.

A comprehensive view of the cause of CFS

It now seems clear that rather than regarding pathophysiological and psychopathological studies as separate and competing approaches to the problem, it is more useful to consider a formulation of CFS that combines these factors. Table 5.2.7.3 summarizes relevant aetiological factors.

According to this integrated scheme causal factors are divided into those that may predispose to the illness, those that precipitate it, and those that perpetuate an established illness. Predisposing factors include previous episodes of major depressive disorder, and perhaps also certain personality characteristics, particularly achievement orientation and perfectionism associated with chronic stress, especially occupational stress. The precipitation of CFS by a viral infection is clinically plausible and proven in certain circumstances, whilst life stresses also seem to be important. Perpetuating factors may include neuroendocrine dysfunction, emotional disorder, and physical disease attributions, as well as coping by avoidance, chronic unresolved personal difficulties, and misinformation about the illness. They may be effectively combined in a cognitive–behavioural model of the illness that provides a basis for treatment with CBT.[1,18]

Course and prognosis of CFS

Anecdotal reports of the prognosis of CFS make gloomy reading. What is more, systematic studies are hardly more encouraging, suggesting that the commonest outcome of those attending a specialist CFS clinic is continuing ill health, up to and beyond 5 years.[19] However, these observations need elaboration. The rather dispiriting prognostic studies all refer to patients seen in specialist centres. Nearly all had several years of illness prior to referral, and it is unsurprising to find that chronicity predicts chronicity. Patients seen in specialist clinics often have strong views about illness aetiology and illness management that may negatively influence their acceptance of and adherence to potentially effective treatment. Primary care and community samples and patients appear to have a better outcome, as do children and adolescents. Finally, the current generation of outcome studies refer to the situation without treatment as potentially effective treatment was rarely given. Later sections of this chapter suggest that this view requires revision.

Table 5.2.7.3 An aetiological formulation of CFS

	Predisposing	Precipitating	Perpetuating
Biological	Genetics Previous depression	Infection	Effects of inactivity CNS dysfunction Reduced HPA activity
Psychological	Personality (perfectionism)	Response to stressor	Focus on symptoms Disease attribution Avoidant coping
Social	Chronic stress	Social/occupation stress	Life conflicts Psychological iatrogenesis

Evidence for treatments

Drug treatment

Many pharmacological treatments have been suggested for CFS. To date, none are of proven efficacy and several are potentially harmful.[4, 20] The evidence for antidepressant agents is mixed. Of available agents none is clearly superior for this patient group, although clinical experience suggests that the selective serotonin-reuptake inhibitor antidepressants may be better tolerated.

Graded activity (exercise) therapy

Well-conducted randomized controlled trials suggest that graded increases in physical activity is helpful in improving function and relieving symptoms.[4, 20]

Cognitive behaviour therapy

Systematic reviews have concluded that the strongest evidence for efficacy is for a rehabilitative type of cognitive–behaviour therapy (CBT).[4, 20]

Practical management

Assessment

Both a medical and psychiatric assessment is required in every case of suspected CFS.[21]

(a) Excluding organic disease

A small minority of those patients who present with severe chronic fatigue will be found to have occult organic disease. How frequently organic disease is found will depend on how thorough an assessment the patient has already received. The differential diagnosis is listed in Table 5.2.7.4.

There are no specific diagnostic tests and no characteristic abnormalities on laboratory investigations in CFS. Tests are conducted purely to exclude other diseases. All patients should have a full blood count, erythrocyte sedimentation rate or C-reactive

Table 5.2.7.4 Conditions to be considered in the differential diagnosis of chronic fatigue syndrome

Nature of symptoms	Possible condition
General	Occult malignancy Autoimmune disease Endocrine disease Cardiac, respiratory, or renal failure
Gastroenterological Neurological	Malabsorption including celiac disease Disseminated sclerosis Myasthenia gravis Parkinson's disease Early dementia Cerebrovascular disease
Infectious disease	Chronic active hepatitis (B or C) Lyme borreliosis HIV Tuberculosis
Respiratory disease	Nocturnal asthma Obstructive sleep apnoea
Chronic toxicity	Alcohol Solvents Heavy metals Irradiation
Psychiatric	Major depressive disorder Dysthymia Anxiety and panic disorder Somatoform disorder

protein, basic biochemistry screen, creatine kinase, random blood glucose, urine analysis, thyroid function, and possibly antinuclear antibody tests. Further investigation depends on the clinical findings and differential diagnoses under consideration. In our clinical experience unusual clinical features, such as weight loss, an absence of mental fatigue/fatigability, or a history of recent foreign travel, should all increases suspicion of alternative diagnoses.

(b) Excluding psychiatric diagnoses

All patients should have a psychiatric history taken and their mental state examined. The assessment should seek evidence of major depression, anxiety, and panic disorder, and also evaluate any suicidal intent. The psychiatric assessment should be systematic, as hidden distress is common and casual estimates of the patient's degree of distress may be misleading.

Making the diagnosis

As explained above the choice of diagnosis should be pragmatic; there is little merit in giving a diagnosis of CFS if the patient's symptoms are clearly those of depression or anxiety. In other cases, a diagnosis of CFS may be the most appropriate and useful; it offers the patient a coherent label for their symptoms and will therefore lessen the risk that they will embark on a fruitless search for a 'better' explanation. Above all, it is most important that neither the physician nor the patient stops at this diagnosis, but goes on to explain what it does and does not mean.

Making a formulation

An adequate individual patient assessment must identify all the important obstacles to recovery. It often needs to go beyond diagnosis to include a systematic individualized description of the aetiological factors in each case. These should include those factors listed in Table 5.2.7.1.

(a) Case formulation

A multidimensional description of the patient's illness provides a comprehensive picture of the factors that may be relevant to the patient's illness and is an important supplement to diagnosis. Its use can be illustrated by returning to the case example described above.

Case Study

Assessment of the patient described earlier revealed that she believed that her symptoms were by an ongoing virus infection and associated immune disturbance and that she should beware of exacerbating them. She consequently avoided activity and had been profoundly inactive for over a year, often lying in bed and sleeping for long periods. Therefore she had become physiologically deconditioned. She was frustrated with her inability to do things and sometimes felt low in mood about her predicament. Her previous job had been very stressful, but since becoming ill she had been unable to work. She had now lost her job and was cared for by her mother who also believed she had a permanent disability. Her doctor said that the best thing was rest. She had rejected a psychiatric consultation but was paying to see an alternative medicine therapist.

The findings can be summarized in an individualized form of Table 5.2.7.3 with an emphasis on the perpetuating factors, which can be seen as reversible barriers to recovery.

General management

The five basic steps essential to the care of patients with CFS are listed in Table 5.2.7.5.

The doctor should listen to the patient's story and ask about his or her own understanding of the illness. It is usually also worth seeing the partner or relevant family members. It is important to address misunderstandings about the nature of the illness and especially to make clear that it is not progressive or life threatening. A positive explanation of CFS as a 'dysfunction of the central nervous system' emphasizing reversibility is often helpful. Anxiety and depression can be explained as understandable consequences of illness and treatment given for them. The adverse physiological and psychological effects of prolonged bed rest should be explained, and the patient encouraged to avoid extremes of both inactivity

Table 5.2.7.5 Principles of management of CFS

1. Acknowledge the reality of the patient's symptoms and disability
2. Provide appropriate education about the nature of the syndrome
3. Treat identifiable depression and anxiety disorder
4. Encourage a very gradual return to normal functioning
5. Help the patient overcome occupational and interpersonal obstacles to normal functioning

and exertion. Finally there are often problem in relationships and with employers than the patient may need help to address.

An initial hospital appointment that achieves all the above usually requires at least 45 min. An evidence-based self-help book may be recommended.[22]

(a) Pharmacological treatments

Patients are often reluctant to take antidepressants and careful explanation and follow-up are required. Other pharmacological agents should only be used with care and preferably only as part of randomized controlled trials.

(b) Graded activity and exercise therapy

This should be considered for patients who are physically inactive. However, they need to be slowly graded and tailored to the patients' ability and progress; a simplistic application of fixed exercise regimens, particularly if given without adequate explanation is unlikely to be helpful, and may exacerbate symptoms and damage confidence.

(c) CBT

Whilst CBT is currently the mainstay of management it is not effective in all patients and requires both careful explanation and therapists skilled in its delivery to patients with CFS. It is particularly important that patients do not interpret referral for a behavioural treatment as the doctor implying that their illness is imaginary. It can be explained that the most likely cause of fatigue in CFS is changes in brain and neuroendocrine function and that these can be reversed by these therapies. It is also both useful and true to draw attention to studies showing the effective of CBT in improving symptoms, quality of life and outcome in conditions such as diabetes, cancer, and rheumatoid arthritis. General physical rehabilitation services may be useful for patients with chronic severe disability.

Potential management problems

Several issues may complicate the management of patients with CFS. These include:

(a) Strong illness beliefs

Difficulties are most likely to arise when patient and physician hold differing beliefs about the nature and best management of the illness. This problem can often be overcome by the physician acknowledging the patient's beliefs without either agreeing with them or arguing about them. If the patient's family, friends, or acquaintances suggest or encourage views that the physician regards as unhelpful the problem is more difficult and a meeting with the other parties may be necessary.

(b) Alternative therapies

Patients with CFS often turn to alternative and complementary medicine. Whilst some of these therapies may be easily continued in parallel with rehabilitative management, others may interfere either by the explanation of the illness they offer or by their practical requirements. In such cases it can be helpful to explain the need to pursue one approach at a time in order to learn what helps.

(c) Official reports and financial benefits

Perhaps the greatest difficulty is when patients ask the physician to write reports on their behalf, confirming that they suffer from permanent disability. The physician wants to help the patient and ensure that they get appropriate benefits but also to avoid a self-fulfilling prophecy. This dilemma has no easy solution, but it seems important not to confirm a negative prognosis until potentially effective treatment has been tried.

(d) Poor prognosis patients

For patients who have been identified as having a poor prognosis because of a long history of severely impaired functioning, or poor response to treatment, regular (albeit infrequent) long-term follow-up is, at the least, likely to limit iatrogenic harm from unnecessary investigations and ineffective treatments. Often in such cases the best strategy is simply to tell the patient how well they are managing in difficult circumstances.

Possibilities for prevention

We do not know how to prevent CFS, but its development can probably be modified.

The most important place for such intervention may be the transition from an acute fatigue state to chronic disability. For example, although most of us have been exposed to Epstein–Barr virus infection by the time we reach 30 years of age, few go on to develop CFS. Encouraging modest amounts of activity in the weeks after an acute infection has been shown to be effective in reducing the duration of symptoms. Intervention that maintain activity and prevent a slide into a vicious circle of symptoms, reduced activity, demoralization, disability, and depression might therefore offer an opportunity for prevention. The second area for intervention is in achieving better attitudes to symptoms and distress. Simplistic depictions of illness as either physical or psychological and the corresponding division in medical services are clearly unhelpful. The third area is in employment practice. CFS is often associated with work stress and dissatisfaction and only a minority of employers allow a flexible return to work. Finally, a doctor–patient relationship that both allows the patient to be ill and encourages recovery is probably the most important preventive strategy.

Conclusions and future directions

Chronic fatigue syndrome is best regarded as a descriptive term for a clinical presentation, rather than as a discrete condition. The group of patients it defines is almost certainly aetiologically heterogeneous. While psychiatric diagnosis provides one approach to subclassification of CFS, both the medical and psychiatric current diagnostic systems have significant limitations, and a multidimensional description of the patient's characteristics may be more clinically useful.[23]

The illness defined by the term chronic fatigue syndrome is important because it represents potentially treatable disability and suffering. It is also important because the clinical problems it gives rise to highlight shortcomings in our present approach to illness. Whatever is ultimately discovered about the causes of CFS, the attention it is receiving offers a golden opportunity to reappraise our understanding and classification of human illness and to re-examine our current organization of medical care.

Further information

Wessely, S., Sharpe, M., and Hotopf, M. (1998). *Chronic fatigue and its syndromes*. Oxford University Press, Oxford.

National Institute for Health and Clinical Excellence Guideline number 53. Chronic fatigue syndrome/ myalgic encephalomyelitis (or encephalopathy); diagnosis and management. http://guidance.nice.org.uk/CG53

References

1. Prins, J.B., Van der Meer, J.W., and Bleijenberg, G. (2006). Chronic fatigue syndrome. *Lancet*, **367**, 346–55.
2. Wessely, S., Hotopf, M.H., and Sharpe, M. (1998). *Chronic fatigue and its syndromes*. Oxford University Press, Oxford.
3. Holmes, G.P., Kaplan, J.E., Gantz, N.M., *et al.* (1988). Chronic fatigue syndrome: a working case definition. *Annals of Internal Medicine*, **108**, 387–9.
4. Baker, R. and Shaw, E.J. (2007). Diagnosis and management of chronic fatigue syndrome or myalgic encephalomyelitis (or encephalopathy): summary of NICE guidance. *British Medical Journal*, **335**, 446–8.
5. Sharpe, M. (1998). Doctors' diagnoses and patients' perceptions: lessons from chronic fatigue syndrome. *General Hospital Psychiatry*, **20**, 335–8.
6. Fukuda, K., Straus, S.E., Hickie, I.B., *et al.* (1994). Chronic fatigue syndrome: a comprehensive approach to its definition and management. *Annals of Internal Medicine*, **121**, 953–9.
7. Reeves, W.C., Lloyd, A., Vernon, S.D., *et al.* (2003). Identification of ambiguities in the 1994 chronic fatigue syndrome research case definition and recommendations for resolution. *BMC Health Services Research*, **3**, 25.
8. Wessely, S., Nimnuan, C., and Sharpe, M. (1999). Functional somatic syndromes: one or many? *Lancet*, **354**, 936–9.
9. Skapinakis, P., Lewis, G., and Meltzer, H. (2000). Clarifying the relationship between unexplained chronic fatigue and psychiatric morbidity: results from a community survey in Great Britain. *The American Journal of Psychiatry*, **157**, 1492–8.
10. Katon, W., Buchwald, D.S., Simon, G.E., *et al.* (1991). Psychiatric illness in patients with chronic fatigue and rheumatoid arthritis. *Journal of General Internal Medicine*, **6**, 277–85.
11. Fischler, B., Cluydts, R., De Gucht, Y., *et al.* (1997). Generalized anxiety disorder in chronic fatigue syndrome. *Acta Psychiatrica Scandinavica*, **95**, 405–13.
12. Farmer, A., Jones, I., Hillier, J., *et al.* (1995). Neurasthenia revisited. *British Journal of Psychiatry*, **167**, 496–502.
13. Mayou, R., Kirmayer, L.J., Simon, G., *et al.* (2005). Somatoform disorders: time for a new approach in DSM-V. *American Journal of Psychiatry*, **162**, 847–55.
14. Kato, K., Sullivan, P.F., Evengard, B., *et al.* (2006). Premorbid predictors of chronic fatigue. *Archives of General Psychiatry*, **63**, 1267–72.
15. White, P.D., Thomas, J.M., Amess, J., *et al.* (1995). The existence of a fatigue syndrome after glandular fever. *Psychological Medicine*, **25**, 907–16.
16. Afari, N. and Buchwald, D. (2003). Chronic fatigue syndrome: a review. *American Journal of Psychiatry*, **160**, 221–36.
17. Cho, H.J., Skowera, A., Cleare, A., *et al.* (2006). Chronic fatigue syndrome: an update focusing on phenomenology and pathophysiology. *Current Opinion in Psychiatry*, **19**, 67–73.
18. Surawy, C., Hackmann, A., Hawton, K.E., *et al.* (1995). Chronic fatigue syndrome: a cognitive approach. *Behaviour Research and Therapy*, **33**, 535–44.
19. Cairns, R. and Hotopf, M. (2005). A systematic review describing the prognosis of chronic fatigue syndrome. *Occupational Medicine (London)*, **55**, 20–31.
20. Chambers, D., Bagnall, A.M., Hempel, S., *et al.* (2006). Interventions for the treatment, management and rehabilitation of patients with chronic fatigue syndrome/myalgic encephalomyelitis: an updated systematic review. *Journal of the Royal Society of Medicine*, **99**, 506–20.
21. Sharpe, M., Chalder, T., Palmer, I., *et al.* (1997). Chronic fatigue syndrome. A practical guide to assessment and management. *General Hospital Psychiatry*, **19**, 185–99.
22. Campling, F. and Sharpe, M. (2000). *Chronic fatigue syndrome: the facts*. Oxford University Press, Oxford.
23. Sharpe, M., Mayou, R., and Walker, J. (2006). Bodily symptoms: new approaches to classification. *Journal of Psychosomatic Research*, **60**, 353–6.

5.2.8 Body dysmorphic disorder

Katharine A. Phillips

The dysmorphophobic patient is really miserable; in the middle of his daily routines, talks, while reading, during meals, everywhere and at any time, he is caught by the doubt of deformity . . . . Enrico Morselli, 1891[1]

Introduction

Body dysmorphic disorder (BDD), also known as dysmorphophobia, is a relatively common, severe, and sometimes difficult-to-treat condition that has been described for more than a century.[1–3] BDD consists of a distressing or impairing preoccupation with an imagined or slight defect in one's physical appearance. BDD is classified as a separate disorder in DSM-IV and a type of hypochondriasis in ICD-10. This disorder can cause severe distress and notably impaired functioning. In addition, risk behaviours—suicidality, violence, problematic substance use, and compulsive tanning—appear common in BDD. Despite its severity, BDD is underrecognized in clinical settings.

Clinical features

Demographic characteristics

BDD occurs in all age groups.[3–5] It appears about equally common in females and males or may be somewhat more common in females.[3] Most individuals with BDD have never been married, and a high proportion is unemployed, often because of their psychopathology.[3–7]

Bodily preoccupations

People with BDD are preoccupied with the idea that some aspect of their appearance is ugly, unattractive, deformed, flawed, or defective in some way.[1–6, 8–10] Concerns usually focus on the face or head but can involve any body area.[3–6, 8–10] Skin (e.g. acne, scars, lines, or pale skin), hair (e.g. thinning or excessive body or facial hair), and nose (e.g. size or shape) concerns are most common. Most patients are preoccupied with several body areas. The preoccupation usually focuses on specific areas but may involve overall appearance.

BDD preoccupations are distressing, time consuming (occurring for an average of 3–8 h a day), and usually difficult to resist or control.[3] They are often associated with low self-esteem, shame, rejection sensitivity, and high levels of neuroticism, introversion,

depressed mood, anxiety, anger-hostility, and perceived stress.[3] Patients often believe that they are unacceptable—e.g. worthless, inadequate, unlovable, and an object of ridicule and rejection.[3,9]

Insight/delusionality

Insight is usually poor or absent; 27–39 per cent of patients are currently delusional (completely convinced that their belief is accurate and undistorted).[3,11,12] Most do not recognize that their belief is due to a mental illness or has a psychological/psychiatric cause.[3,12] In addition, a majority have ideas or delusions of reference, believing that others take special notice of the supposed appearance defects—for example, stare at them or mock the person because of how they look.[3,12] Referential thinking can fuel feelings of anger and rejection as well as social isolation.

Compulsive and safety behaviours

Nearly all patients perform BDD-related compulsive or safety behaviours (Table 5.2.8.1), which are time consuming (occurring for hours a day) and difficult to resist or control.[3–6,10] The behaviours usually aim to examine, improve, hide, or obtain reassurance about the perceived defects. These behaviours typically do not alleviate distress and may even worsen it.

Compulsive skin picking, which 27–45 per cent of BDD patients do to try and improve their appearance, can cause considerable skin damage.[3,5] Emergency surgery is sometimes required—for example when sharp implements used for picking rupture major blood vessels. Compulsive tanning to darken 'pale' skin or minimize perceived acne, scarring, or 'marks' can cause skin damage and may increase cancer risk.[3]

Psychosocial functioning and quality of life

Functioning and quality of life are usually very poor.[3,4,7] Some people, with effort, function adequately despite their distress, although usually below their potential. Those with severe BDD may be profoundly impaired by their symptoms—for example, housebound for years, unemployed and socially isolated, and chronically suicidal.

Social impairment is nearly universal.[3,7] People with BDD feel embarrassed and ashamed of their 'ugliness', are anxious around others as a result, and fear being rejected because of how they look. Thus, they may have few or no friends; avoid dating, physical intimacy, and other social interactions; or get divorced. Impairment in academic or occupational functioning is common, due to the time consuming and distracting nature of BDD symptoms and a desire to avoid interactions with others.[3,7] In a broadly ascertained BDD sample (n = 200), 36 per cent of individuals were not currently working and 32 per cent were not able to be in school or do school work because of psychopathology (BDD was the primary diagnosis for most).[7] In two BDD series, more than a quarter of individuals had been completely housebound for at least 1 week because of BDD symptoms, and more than 40 per cent had been psychiatrically hospitalized.[5,13] Mental health related quality of life is markedly poorer than for the general population and even poorer than for patients with diabetes, a recent myocardial infarction, or clinical depression.[7]

Table 5.2.8.1 Common compulsive and safety behaviors in body dysmorphic disorder

- Comparing one's appearance 'defects' with the same body areas of other people
- Checking the perceived defects directly, in mirrors, or in other reflecting surfaces
- Excessive grooming—for example, applying make-up, shaving, hair cutting or removal, hair styling
- Camouflaging—for example with a hat, clothes, sunglasses, hair, body position, hand, or make-up
- Seeking reassurance from others or trying to convince others of the 'defect's' ugliness
- Skin picking
- Excessive exercising or weightlifting
- Tanning
- Frequent clothes changing
- Touching the perceived defect
- Dieting
- Body measuring
- Compulsive shopping for beauty products, remedies, or clothes
- Seeking surgical, dermatologic, and other cosmetic treatment for the perceived deformity

Suicidality

Suicidal ideation and attempts appear very common. Reported lifetime rates of suicidal ideation and suicide attempts are 78–81 and 24–28 per cent, respectively.[9,13,14] Among adolescent inpatients, those with BDD have significantly greater suicidality than those without BDD. The rate of completed suicide, while preliminary, appears markedly high. In a prospective study, the annual suicide rate was 0.35 per cent, which is approximately 45 times higher than for the US population (adjusted for age, gender, and geographic region) and higher than for most other mental disorders.[15] A study of dermatology patients who committed suicide found that most had acne or BDD.[16] Indeed, individuals with BDD have many suicide risk factors.[3,14]

Aggression and violence

In several BDD studies, 36–38 per cent of patients reported lifetime aggression/violence due specifically to BDD symptoms.[3,10] Such behaviour may be fuelled by anger about looking 'deformed', an inability to fix the perceived defect, and misperceptions of being rejected, ridiculed, or mocked because of the appearance 'defects'. Individuals with BDD tend to misinterpret self-referent facial expressions as contemptuous and angry,[17] misinterpret ambiguous social (and other) situations as threatening,[18] and have high levels of anger/hostility.[3] Surgeons and dermatologists may be victims of violence—even murder—fuelled by dissatisfaction with the outcome of cosmetic procedures.[3] In a survey of 265 plastic surgeons, 12 per cent reported that a BDD patient had physically threatened them.[19]

Comorbidity

Major depressive disorder is the most frequently comorbid disorder, occurring in about 75 per cent of individuals with BDD.[5,13] Social phobia, OCD, and substance use disorders are also common.[5,8,13] Of note, one study found that 49 per cent of 200 BDD subjects had a lifetime substance use disorder, 70 per cent of whom reported that BDD contributed to their substance use.[5] Muscle dysmorphia,

a preoccupation with the idea that one's body is insufficiently lean or muscular, may lead to anabolic steroid abuse.[3]

Gender

Men and women appear to have largely similar clinical features.[10, 13, 20] However, in two United States studies (n = 200 and n = 188[13, 20]) males were more likely to be single, have a substance use disorder, and be preoccupied with thinning hair and small body build. Females were more likely to be preoccupied with weight, hips, and excessive body hair, and were more likely to pick their skin and use their hands or make-up for camouflage. One of these two studies, and a study from Italy (n = 58[10]), found that females were more likely to be preoccupied with their breasts/chest and legs, check mirrors, and use camouflaging. In all three studies, a concern with genitals was more common in males, and a comorbid eating disorder more common in females.

BDD in children and adolescents

BDD usually begins during early adolescence and can occur in childhood.[4, 5, 10] While data are limited, BDD's clinical features in youth appear largely similar to those in adults. Of note, children and adolescents appear to have lifetime rates of functional impairment similar to those in adults, despite having had fewer years over which to have developed these problems.[3] In the one study that directly compared adolescents to adults, adolescents were more likely to have delusional BDD beliefs and had a significantly higher lifetime suicide attempt rate (44 per cent versus 24 per cent), underscoring the importance of recognizing BDD in this age group.

Cross-cultural aspects of BDD

Case reports and series from around the world suggest that BDD's clinical features are generally similar across cultures but that cultural factors may produce nuances and accents on a basically invariant, or universal, expression of BDD.[3] It is unclear whether koro (a belief that one's penis is shrinking, which occurs primarily in Southeast Asia) is a form of BDD.

Classification and relationship to other disorders

A clinically important classification controversy is whether delusional and non-delusional BDD are the same or different disorders. Whereas BDD is classified as a somatoform disorder, its delusional variant is classified as a psychotic disorder—a type of delusional disorder. DSM-IV, however, allows delusional patients to be diagnosed with both BDD and delusional disorder, reflecting data suggesting that its delusional and non-delusional variants may constitute the same disorder, which spans a spectrum of insight. Indeed, delusional and non-delusional BDD appear more similar than different, although delusional BDD appears more severe.[11,12] Of clinical significance, delusional and non-delusional patients both appear to respond to serotonin-reuptake inhibitors (SRIs) as monotherapy but not to antipsychotic monotherapy.[21]

Another important question is whether BDD is related to OCD, social phobia, major depressive disorder, or eating disorders.[3] BDD is widely conceptualized as an OCD-spectrum disorder. Supporting this conceptualization, OCD often co-occurs with BDD, and BDD appears more common in first-degree relatives of OCD probands than control probands.[3] Data from a variety of domains suggest that BDD and OCD have many similarities.[3] However, BDD and OCD have some differences and do not appear to be identical disorders.[3] Although BDD's relationship to other disorders has received less investigation, preliminary data suggest that BDD may also be related to major depressive disorder. However, BDD does not appear to simply be a symptom of depression.[3] Although ICD-10 classifies BDD as a type of hypochondriasis, no studies have examined their relationship.

Diagnosis

BDD can be diagnosed using questions at the top of Table 5.2.8.2, which follow DSM-IV's diagnostic criteria. Clinicians should adequately probe for examples of clinically significant distress and impairment in social, occupational, and other aspects of functioning. BDD is diagnosed if the person is excessively preoccupied with a non-existent or slight physical flaw (for example, thinks about it for at least an hour a day), and the concern causes clinically significant distress or impairment in functioning. The appearance concerns should not be better accounted for by an eating disorder. However, BDD and eating disorders may co-occur, in which case both disorders should be diagnosed.

The bottom of Table 5.2.8.2 includes questions that are ***not*** recommended for screening for or diagnosing BDD. The word 'imagined' is problematic, because most patients have poor or absent insight and do not think their appearance problem is imagined. Terms such as 'deformed' or 'disfigured' are too extreme for some patients to endorse. Asking if there is something wrong with one's body is too broad, as patients may interpret this to refer to bodily functioning.

ICD-10's criteria require that patients persistently refuse to accept the advice and reassurance of several different doctors that they do not have an abnormality. However, many people with BDD do not disclose their appearance concerns to doctors or even seek medical care because they are housebound, ashamed, believe they cannot be helped, lack medical insurance, or do not have access to

Table 5.2.8.2 Questions to diagnose BDD

Recommended questions for diagnosing BDD:

1 Are you very worried about your appearance in any way? ***OR:*** Are you unhappy with how you look? ***If yes:*** Can you tell me more about your concern?

2 Does this concern preoccupy you? Do you think about it a lot and wish you could worry about it less? How much time would you estimate you spend each day thinking about how you look, if you were to add up all the time you spend?

3 What effect does this preoccupation with your appearance have on your life? Has it caused you a lot of distress (for example, anxiety or depression)? Has it significantly interfered with your social life, relationships, school work, job, or any other activities? Has it affected your family or friends?

*Questions that are **not** recommended when screening for or diagnosing BDD:*

1 It is recommended that patients ***not*** be asked if they are concerned with an 'imagined' defect in their appearance, whether they think they are 'deformed' or 'disfigured', or whether there is something 'wrong with (their) body'

2 To diagnose BDD, it should not be required that patients refuse to accept the advice and reassurance of doctors that they do not have an abnormality

health care for other reasons. Using this diagnostic criterion will underdiagnose BDD. BDD is also underidentified in many mental health databases because its ICD-9 diagnostic code identifies it as hypochondriasis.

BDD usually goes undiagnosed in clinical settings.[3,4,22] Sufferers often conceal their symptoms due to embarrassment and shame.[3,4,22] They may volunteer only depression, anxiety, or discomfort in social situations. The compulsive and safety behaviours in Table 5.2.8.1 may be clues to BDD's presence. BDD may be misdiagnosed as another disorder[3]: social phobia or agoraphobia (due to secondary social anxiety and isolation), panic disorder (because panic attacks may occur after looking in the mirror or when feeling scrutinized by others), trichotillomania (when hair is cut or plucked to improve perceived flaws, such as uneven eyebrows), or OCD (due to obsessional preoccupations and compulsive behaviours). Delusional patients are sometimes misdiagnosed with schizophrenia, psychotic depression, or psychotic disorder NOS. To diagnose BDD, patients must usually be asked directly about BDD symptoms.

Epidemiology

BDD has been reported in 0.7–1.7 per cent of community samples.[3] In the largest study, a nationwide survey in Germany (n = 2552), BDD was present in 1.9 per cent of women and 1.4 per cent of men.[23] In smaller non-clinical student samples, BDD's prevalence has ranged from 2.3–13 per cent.[3] A prevalence of 9–12 per cent has been reported in dermatology settings, 3–15 per cent in cosmetic surgery or plastic surgery settings, 8–37 per cent in patients with OCD, 11–13 per cent in social phobia, 26 per cent in trichotillomania, and 14–42 per cent in atypical major depression.[3] In a study of 122 general psychiatric inpatients, 13 per cent had BDD's, which was higher than for schizophrenia, OCD, PTSD, and eating disorders.[22]

Pathogenesis

BDD's pathogenesis is likely to be complex and multifactorial.[3] Aetiologic factors likely involve a complex interplay of genetic and environmental risk factors.[3] Preliminary data indicate an association of the $GABA_A$-γ2 gene with BDD. Environmental risk factors may include perceived childhood neglect and/or abuse, teasing, and low parental warmth. A role is also likely for socio-cultural and evolutionary pressures (e.g. symmetrical features may signal reproductive health).

Neuropsychological studies indicate a tendency to focus on isolated details of visual and verbal stimuli rather than more global, configurational attributes[24]—consistent with clinical observations that patients selectively attend to specific aspects of their appearance or minor flaws. Cognitive processing studies indicate that BDD patients tend to misinterpret ambiguous social (and other) situations as threatening and misinterpret self-referent facial expressions as contemptuous and angry.[17,18] These interpretive biases may combine with rejection sensitivity, perfectionism, and a focus on aesthetics to contribute to BDD's development.[3] High neuroticism and low extroversion may also play a role.[3] Many potential risk factors (e.g. neuroticism) are not specific to BDD, but the overall combination of risk factors may be. BDD's neurocircuitry is unknown but likely involves a complex interplay of dysfunction in several neural systems,[3,25] including circuitry involved in OCD (orbitofrontal cortex, anterior cingulate cortex, caudate, thalamus). BDD's shared features with other anxiety disorders and depression points to possible dysfunction in amygdala, prefrontal cortex, and anterior cingulate cortex. Brain regions involved in body image and facial emotion perception (e.g. right parietal cortex, amygdala, occipitotemporal cortex [e.g. fusiform face and extrastriate body areas]) may also be involved.

Course and prognosis

Prospective and retrospective studies indicate that BDD is usually chronic.[5,13,26] More severe BDD, a longer duration of BDD, and the presence of a personality disorder predict a lower probability of remission. However, when BDD is accurately identified and its treatment optimized, the prognosis appears much more favourable.

Treatment

BDD's treatment is described in more detail elsewhere, including in a guideline from the United Kingdom's National Institute of Clinical Excellence (NICE).[3,21,27,28] Serotonin-reuptake inhibitors (SRIs, or SSRIs) and cognitive–behavioural therapy (CBT) are currently recommended as the first-line treatments.[3,21,27,29] Treatment studies are limited, and more research is needed. However, available data consistently indicate that a majority of patients improve with these treatments.

Evaluation of treatments

Surgical, dermatologic, and other cosmetic treatment

A majority of BDD patients seek often-costly cosmetic treatment.[3,8,9,13,30] Dermatologists and surgeons are most often consulted, but any type of physician may be seen. It appears that most BDD patients are dissatisfied with such treatment.[3,9,13,30] Occasionally, dissatisfied patients sue, or are violent towards, the physician. Some patients perform their own surgery[3,6]—e.g. cutting open one's nose with a razor blade and trying to replace nose cartilage with chicken cartilage in the desired shape.

Pharmacotherapy and other somatic treatments

All SRI studies to date indicate that SRIs are often efficacious for BDD.[3,21] These studies include a placebo-controlled fluoxetine study (n = 67), a controlled and blinded cross-over study comparing the SRI clomipramine to the non-SRI antidepressant desipramine (n = 29), and open-label trials of fluvoxamine, citalopram, and escitalopram (n = 15–30). In these studies, 53 per cent to 73 per cent of patients responded to the SRI. The cross-over trial found greater efficacy for clomipramine than desipramine, suggesting that SRIs may be more efficacious than non-SRI antidepressants for BDD. This important finding is consistent with clinical series and retrospective data suggesting that SRIs are more efficacious than a broad range of non-SRI medications for BDD.[3,21]

Response to an SRI usually develops gradually and may require up to 12–14 weeks of treatment (while reaching a relatively high dose) to be evident.[3,21] Although dose-finding studies are lacking, relatively high SRI doses (higher than typically used for depression)

appear to often be needed.[3,21] Response to medication usually includes a decrease in appearance preoccupations, distress, and compulsive/safety behaviours, as well as improved functioning. Suicidality, depressive symptoms, anxiety, and anger-hostility often improve.[3] Of note, delusional patients often improve with SRI monotherapy, whereas limited data suggest that antipsychotic monotherapy is usually ineffective for delusional BDD.[3,21]

Patients who do not improve with one SRI may improve with another SRI.[3, 21] Regarding SRI augmentation, a small double-blind randomized controlled trial found that pimozide was not more efficacious than placebo as a fluoxetine augmenter.[3,21] Clinical series suggest that augmentation of an SSRI with buspirone, lithium, or clomipramine may be helpful.[3,21] (SSRIs may increase clomipramine blood levels, however, which may cause toxicity; thus, if this approach is tried, clomipramine should be started at a very low dose with monitoring of levels.) Clinical observations suggest that SSRI augmentation with venlafaxine, atypical neuroleptics, or bupropion may be helpful for some patients.[3,21]

Monotherapy with agents other than SRIs has not been well studied.[3, 21] A small open-label trial (n = 11) suggests that venlafaxine may be efficacious.[3] For severe and treatment-refractory cases an MAO inhibitor may be worth trying (but should never be combined with an SRI). Available case series and reports, while very limited, suggest that ECT is generally ineffective for BDD and secondary depressive symptoms.[3,21]

Clinical experience suggests that many patients relapse after SRI discontinuation and that long-term treatment may be needed, with efficacy usually sustained over time.[3,21] For patients who appear at high risk of suicide or violence, lifelong treatment with an effective SRI is recommended, as suicides have been known to occur after SRI discontinuation.

Cognitive behavioural therapy (CBT)

Preliminary data suggest that CBT is often efficacious for BDD.[3,27,29] CBT typically includes:[3,27–29]

1 *Cognitive restructuring* to identify cognitive errors and develop more accurate and helpful BDD-related thoughts and beliefs
2 *Behavioural experiments* to test the accuracy of BDD beliefs
3 *Exposure* to avoided situations (e.g. leaving the house, attending social gatherings)
4 *Response prevention* to decrease or stop compulsive behaviours (e.g. stopping excessive mirror checking, limiting grooming time)

In two randomized studies (n = 54 and 19), patients improved more with CBT than on a waiting list.[3,29] Several case series (n = 5–13) found that CBT was efficacious for BDD.[3,29] The number of sessions in these reports ranged from twelve 60-min sessions to sixty 90-min sessions.

Data on exposure and response prevention alone—without a cognitive element—is limited to a retrospective study and small case series with up to 10 subjects.[3,29] These reports note favourable outcomes, although in the author's experience cognitive approaches are a helpful, even necessary, component of treatment for most patients. It can be particularly helpful to work on core beliefs, which typically involve feelings of inadequacy and being unwanted by others.[3, 9] Clinical experience additionally suggests that the following are beneficial components of CBT:[3, 28]

1 *Mirror retraining*, which involves learning to see one's entire body in a non-judgemental and 'holistic' way (rather than focusing on disliked areas), while refraining from excessive mirror checking
2 *Habit reversal* for skin picking and repetitive hair pulling or plucking
3 *Mindfulness skills*
4 *Activity scheduling* and *scheduling pleasant activities* for more severely ill, depressed, and inactive patients
5 *Motivational interviewing*, which may be needed to engage and keep patients in treatment

Future studies (e.g. dismantling studies) are needed to determine which specific components of CBT are necessary and effective.

Other types of psychotherapy have not been well studied and are not currently recommended as first-line treatments.[3] Nonetheless, clinical experience suggests that insight-oriented or supportive therapy—in addition to an SRI and/or CBT—may help some patients cope with their illness or with co-occurring problems or disorders.[3]

Management

Management approaches are described in more detail elsewhere.[3, 21, 27, 28]

1 First try to engage the patient and establish an alliance so they are willing to try treatment. This can be difficult to accomplish, as many patients are delusional, prefer cosmetic treatment, are rejection sensitive, and do not want other people (including a clinician) to see them.
2 Empathize with the patient's suffering.
3 Take patients' appearance concerns seriously, neither dismissing their concerns about how they look nor agreeing that there is something wrong with their appearance. Trying to convince patients (especially delusional patients) that their beliefs are irrational or that they look normal is usually not helpful.
4 Instead, focus on the potential for psychiatric treatment to diminish their distress and preoccupation and improve their functioning and quality of life.
5 Provide psychoeducation about BDD and recommend reading.
6 For patients who wish to pursue cosmetic treatments, explain that such treatment appears ineffective for BDD.
7 Provide education about recommended treatments. It can be helpful to explain, for example, that SRIs are usually well tolerated, are not habit forming, appear to normalize the brain (and do not cause brain damage), and often diminish suicidal thinking in people with BDD. CBT is a practical 'here-and-now' treatment in which patients actively collaborate with the therapist and learn helpful skills by attending sessions and doing homework.

Treatment should be initiated with an SRI and/or CBT. SRIs are also the first-line medication for delusional BDD. All severely ill patients, especially those who are highly suicidal, should, in the author's opinion, receive an SRI. Patients with severe comorbid depression also warrant SRI treatment. Other comorbidity may warrant additional medication.

Before concluding that an SRI is ineffective, it should be tried for 12–16 weeks, reaching the highest dose recommended by the manufacturer or tolerated by the patient (if necessary) for at least 2–3 of those 12–16 weeks. If tolerated, higher doses than those recommended by the manufacturer can be cautiously tried to obtain or optimize a response (excluding clomipramine). If this is not effective, an augmentation strategy or switching to another SRI is indicated. For patients who refuse or are ambivalent about treatment motivational techniques should be tried. Patients who are not improving with CBT may need more frequent sessions, longer sessions, or a change in the current CBT focus. At least 4–6 months of weekly or more frequent CBT sessions, plus daily homework, is generally recommended. More severely ill and delusional patients may require much longer or more intensive treatment. Maintenance/booster sessions following CBT may reduce relapse risk. It may be helpful to add CBT to an SRI, or vice versa, if either treatment alone is insufficient.

For certain patients, adjunctive individual supportive therapy or family therapy may be helpful.[3] Families can be an invaluable support and facilitate treatment.[3] Mental health professionals may need to interface with dermatologists, plastic surgeons, and other physicians from whom patients have requested or are receiving cosmetic treatment.

Prevention

Because little is known about BDD's pathogenesis and risk factors, the disorder cannot currently be prevented. However, BDD should be treated in its early stages, before it causes substantial morbidity or interferes with a child's, adolescent's, or young adult's development.

Conclusions

Although this disorder has received far less investigation than many other serious mental illnesses, BDD research is rapidly advancing. It is important that clinicians screen patients for this often-secret disorder and be aware that it typically goes unrecognized in clinical settings, causing significant morbidity. While much more research is needed, available treatments are often very helpful for this distressing and often-disabling disorder.

Further information

Further information about BDD and its treatment is provided in references 3, 27, and 28 below and at www.bodyimageprogram.com.

References

1. Morselli, E. (1891). Sulla dysmorphophobia e sulla tafefobia. *Bolletinno della R accademia di Genova*, **6**, 110–9.
2. Phillips, K.A. (1991). Body dysmorphic disorder: the distress of imagined ugliness. *The American Journal of Psychiatry*, **148**, 1138–49.
3. Phillips, K.A. (1996, 2005). *The broken mirror: understanding and treating body dysmorphic disorder*. Oxford University Press, New York. Revised and Expanded Edition, 2005.
4. Phillips, K.A., McElroy, S.L., Keck, P.E. Jr., *et al.* (1993). Body dysmorphic disorder: 30 cases of imagined ugliness. *The American Journal of Psychiatry*, **150**, 302–8.
5. Phillips, K.A., Menard, W., Fay, C., *et al.* (2005). Demographic characteristics, phenomenology, comorbidity, and family history in 200 individuals with body dysmorphic disorder. *Psychosomatics*, **46**, 317–32.
6. Fontanelle, L.F., Telles, L.L., Nazar, B.P., *et al.* (2006). A sociodemographic, phenomenological, and long-term follow-up study of patients with body dysmorphic disorder in Brazil. *International Journal of Psychiatry in Medicine*, **36**, 243–59.
7. Phillips, K.A., Menard, W., Fay, C., *et al.* (2005). Psychosocial functioning and quality of life in body dysmorphic disorder. *Comprehensive Psychiatry*, **46**, 254–60.
8. Hollander, E., Cohen, L.J., and Simeon, D. (1993). Body dysmorphic disorder. *Psychiatric Annals*, **23**, 359–64.
9. Veale, D., Gourney, K., Dryden, W., *et al.* (1996). Body dysmorphic disorder: a survey of fifty cases. *The British Journal of Psychiatry*, **169**, 196–201.
10. Perugi, G., Akiskal, H.S., Giannotti, D., *et al.* (1997). Gender-related differences in body dysmorphic disorder (dysmorphophobia). *The Journal of Nervous and Mental Disease*, **185**, 578–82.
11. Phillips, K.A., Menard, W., Pagano, M., *et al.* (2006). Delusional versus nondelusional body dysmorphic disorder: clinical features and course of illness. *Journal of Psychiatric Research*, **40**, 95–104.
12. Phillips, K.A. (2004). Psychosis in body dysmorphic disorder. *Journal of Psychiatric Research*, **38**, 63–72.
13. Phillips, K.A. and Diaz, S. (1997). Gender differences in body dysmorphic disorder. *The Journal of Nervous and Mental Disease*, **185**, 570–7.
14. Phillips, K.A., Coles, M., Menard, W., *et al.* (2005). Suicidal ideation and suicide attempts in body dysmorphic disorder. *The Journal of Clinical Psychiatry*, **66**, 717–25.
15. Phillips, K.A. and Menard, W. (2006). Suicidality in body dysmorphic disorder: a prospective study. *The American Journal of Psychiatry*, **163**, 1280–2.
16. Cotterill, J.A. and Cunliffe, W.J. (1997). Suicide in dermatological patients. *The British Journal of Dermatology*, **137**, 246–50.
17. Buhlmann, U., Etcoff, N.L., and Wilhelm, S. (2006). Emotion recognition bias for contempt and anger in body dysmorphic disorder. *Journal of Psychiatric Research*, **40**, 105–11.
18. Buhlmann, U., Wilhelm, S., McNally, R.J., *et al.* (2002). Interpretive biases for ambiguous information in body dysmorphic disorder. *CNS Spectrums*, **7**, 435–6, 441–3.
19. Sarwer, D.B. (2002). Awareness and identification of body dysmorphic disorder by aesthetic surgeons: results of a survey of American Society for Aesthetic Plastic Surgery members. *Aesthetic Surgery Journal*, **22**, 531–5.
20. Phillips, K.A., Menard, W., and Fay, C. (2006). Gender similarities and differences in 200 individuals with body dysmorphic disorder. *Comprehensive Psychiatry*, **47**, 77–87.
21. Phillips, K.A. and Hollander, E. (in press). Treating body dysmorphic disorder with medication: Evidence, misconceptions, and a suggested approach. *Body Image: An International Journal of Research.*
22. Grant, J.E., Kim, S.W., and Crow, S.J. (2001). Prevalence and clinical features of body dysmorphic disorder in adolescent and adult psychiatric inpatients. *Journal of Clinical Psychiatry*, **62**, 517–22.
23. Rief, W., Buhlmann, U., Wilhelm, S., *et al.* (2006). The prevalence of body dysmorphic disorder: a population-based survey. *Psychological Medicine*, **36**, 877–85.
24. Deckersbach, T., Savage, C.R., Phillips, K.A., *et al.* (2000). Characteristics of memory dysfunction in body dysmorphic disorder. *Journal of the International Neuropsychological Society*, **6**, 673–81.
25. Saxena, S. and Feusner, J.D. (2006). Toward a neurobiology of body dysmorphic disorder. *Primary Psychiatry*, **13**, 41–8.
26. Phillips, K.A., Pagano, M.E., Menard, W., *et al.* (2006). A 12-month follow-up study of the course of body disorder. *The American Journal of Psychiatry*, **163**, 907–12.
27. National Collaborating Centre for Mental Health. (2006). Core interventions in the treatment of obsessive compulsive disorder and body dysmorphic disorder (a guideline from the National Institute

for Health and Clinical Excellence, National Health Service). British Psychiatric Society and Royal College of Psychiatrists, London. http://www.nice.org.uk/page.aspx?o=289817
28. Wilhelm, S., Phillips, K.A., and Steketee, G (in press). *Cognitive-behavioral therapy for body dysmorphic disorder: a modular treatment manual*. Guilford Publications Inc., New York.
29. Neziroglu, F. and Khemlani-Patel, S. (2002). A review of cognitive and behavioral treatment for body dysmorphic disorder. *CNS Spectrums*, **7**, 464–71.
30. Crerand, C.E., Franklin, M.E., and Sarwer, D.B. (2006). Body dysmorphic disorder and cosmetic surgery. *Plastic and Reconstructive Surgery*, **118**, 167e–80e.

5.2.9 Factitious disorder and malingering

Christopher Bass and David Gill

Factitious disorder

Introduction

Patients with factitious disorder feign or simulate illness, are considered not to be aware of the motives that drive them to carry out this behaviour, and keep their simulation or induction of illness secret. In official psychiatric nomenclature, factitious disorder has replaced the eponym Munchausen syndrome, introduced by Asher[1] to describe patients with chronic factitious behaviour. Asher borrowed the term from Raspe's 1785 fictional German cavalry officer, Baron Karl von Munchausen, who always lied, albeit harmlessly, about his extraordinary military exploits.

The criteria for factitious disorder in DSM-IV[2] are (a) the intentional production or feigning of physical or psychological signs or symptoms; (b) motivation to assume the sick role; and (c) lack of external incentives for the behaviour (e.g. economic gain, avoidance of legal responsibility, or improved physical well-being, as in malingering) and lack of a better classification for the disorders.

In the last 10 years there has been increased interest in deception in medical practice, with specific focus on pathological lying and the diagnostic dilemmas in this field: specifically, how to differentiate between hysteria, factitious disorders, and malingering. Some of these topics will be discussed in the next section.

This chapter concentrates on factitious physical complaints; fabricated psychological symptoms are considered under malingering.

Diagnostic problems

The DSM-IV criteria have recently come under attack. Turner[3] has argued that criterion B (motivation to assume the sick role) has no empirical content and fulfils no diagnostic function. He also argues that criterion A, the intentional production of physical or psychological signs or symptoms, emphasizes symptoms and cannot accommodate pseudologia fantastica (PF), voluntary false confessions, and impersonations. He concludes that the two criteria need reformulating in terms of lies and self-harm, respectively. Bass and Halligan[4] have also suggested that because the conceptual justification for factitious disorders is 'empirically unsubstantiated' and the motivation for diagnostic purposes (conscious versus unconscious; voluntary versus involuntary) essentially unknowable, it seems reasonable to question the clinical status and legitimacy of factitious disorder. More recently there has been a resurgence of interest in **pathological lying**, because this is often easier to identify than, for example, the degree of 'voluntariness' or 'motivation' to attain the sick role (however that is defined).

Pathological lying (pseudologia fantastica): a key component of factitious disorder

It is possible to identify pathological lying if the clinician has sufficient information at his disposal (most often the medical notes). If the patient reports, for example, that they are being treated for leukaemia, and when there is evidence that contradicts this, then this suggests dissimulation. On some occasions the patient will admit to lying, but this is rare.

Because pathological lying is often a key component in factitious disorders, evidence for it should be actively sought by the clinician. But what distinguishes the pathological liar from the person who just lies a lot? Dike *et al.*[5] suggest that the diagnosis is made when lying is persistent, pervasive, disproportionate, and not motivated primarily by reward or other external factors. They also suggest, however, that a key characteristic of pathological lying may be its compulsive nature, with pathological liars 'unable to control their lying'. Psychiatric conditions that have been traditionally associated with deception in one form or another include malingering, confabulation, Ganser's syndrome, factitious disorder, borderline personality disorder and antisocial personality disorder. Lying may also occur in histrionic and narcissistic personality disorders. It is important to note however that pathological lying can occur in the absence of a psychiatric disorder, and that there may be different types of pathological lying, e.g. the benefit fraudster and the stereotypical wandering Munchausen patient describe different subgroups. Furthermore, it has been reported that up to 40 per cent of cases of pseudologia fantastica have a history of central nervous system abnormalities, which suggests that brain dysfunction in these patients requires closer study.[6]

In recent years, functional neuroimaging techniques (especially functional magnetic resonance imaging) have been used to study deception. Attempted deception is associated with activation of executive brain regions (particularly prefrontal and anterior cingulate cortices), while truthful responding has not been shown to be associated with any areas of increased activation (relative to deception).[7] Furthermore, Yang *et al.*[8] reported that pathological liars showed a 22–26 per cent increase in prefrontal white matter and a 36–42 per cent reduction in prefrontal grey/white ratios compared with both antisocial controls and normal controls. These findings suggest that increased prefrontal white matter developmentally provides a person with the cognitive capacity to lie, although Spence[9] has urged caution in the interpretation of these results.

Clinical features of factitious disorder

Clinical features are diverse, and attempts to subtype patients have not always been helpful. The majority of patients with factitious disorders are non-wandering, socially conformist young women (often nurses) with relatively stable social networks.[10–12] These patients

are likely to enact their deceptions in general hospitals, especially accident and emergency departments, and the liaison psychiatrist should be alert to these clinical problems, which can be referred from a variety of different medical and surgical specialties.

Factitious disorders typically begin before the age of 30 years;[13,14] there are often prodromal behaviours in childhood and adolescence (see below). These individuals often report an unexpectedly large number of childhood illnesses and operations, and many have some association with the health care field.[10] High rates of substance abuse, mood disorder, and borderline personality disorder have been reported.[10,12] Approximately four-fifths of factitious disorder patients are women, and 20–70 per cent work in medically related occupations.[12]

Clinicians should be alert to the presentation of more exotic forms of factitious presentation. For example, some women present themselves to family cancer or genetic-counselling clinics and provide a false family history of breast cancer to their medical attendants.[15] Another recent example is 'electronic' factitious disorder,[16] used to describe patients who falsify their electronic medical records to create a factitious report (e.g. of cancer). Another important group is encountered in pregnancy, and this will clearly have important implications for child protection.[17]

In medico-legal practice factitious disorders have been described in patients with a diagnosis of reflex sympathetic dystrophy (RSD), specially involving the forearm,[18] and others have reported that the abnormal movements commonly associated with RSD (CPRS Type I) are consistently of somatoform or malingered origin.[19] Cases have been described where patients involved in litigation have died of factors directly related to factitious physical disorder.[20]

It is being increasingly recognized that these disorders can occur in childhood and adolescence, and child psychiatrists need to be alert to factitious presentations, especially in departments of infectious diseases.[21] Unlike adult patients, many of these children admit to their deceptions when confronted, and some have positive outcomes at follow-up. The descriptions of some of these children as bland, depressed, and fascinated with health care are remarkably similar however to adults with factitious disorders.[22]

Classification

Four main subtypes are distinguished in DSM-IV.[2]

1 Factitious disorder with predominantly psychological signs and symptoms. This is more difficult to diagnose than factitious disorder with physical complaints, because there is no way of excluding a 'true' psychiatric disorder by physical examination or laboratory investigation: see below under malingering.

2 Factitious disorders with predominantly physical signs and symptoms. Almost every illness has been produced factitiously. However, four subgroups describe most cases[23]:

   (a) self-induced infections

   (b) simulated illnesses, for example adding blood to urine

   (c) interference with pre-existing lesions or wounds

   (d) surreptitious self-medication, for example self-injection of insulin

   These categories are not mutually exclusive or jointly exhaustive.

3 Factitious disorders with combined psychological and physical symptoms

4 Factitious disorders not otherwise specified. This includes factitious disorder by proxy (see below and Chapter 9.3.3).

Diagnosis

Clinicians should become suspicious that a patient may be fabricating symptoms if the following features are noted:

- The course of the illness is atypical and does not follow the natural history of the presumed disease, e.g. a wound infection does not respond to appropriate antibiotics (self-induced skin lesions often fall into this category, when 'atypical' organisms in the wound may alert the physician).
- Physical evidence of a factitious cause may be discovered during the course of treatment, e.g. a concealed catheter, a ligature applied to a limb to induce oedema.
- The patient may eagerly agree to or request invasive medical procedures or surgery.
- There is a history of numerous previous admissions with poor outcome or failure to respond to surgery (these patients may overlap with the chronic somatoform patient with 'surgery prone behaviour').[24]
- Many physicians have been consulted and have been unable to find a relevant cause for the symptoms.

Additional clues include the patient being socially isolated on the ward and having few visitors, or the patient being prescribed (or obtaining) opiate medication, often pethidine, when this drug is not indicated. When these findings occur in someone who has either worked in or is related to someone who has worked in the health field, the caregivers should have a high index of suspicion for a factitious disorder. Obtaining collateral information from family members, prior physicians, and hospitals is crucial.

Differential diagnosis

Factitious disorder must be distinguished from authentic medical conditions. It is not uncommon in clinical practice however to find patients with **both** factitious disorder and coexisting physical illness. For example, patients with brittle diabetes are usually young females who deliberately interfere with their treatment, causing unstable diabetic control.[25] A syndrome of severely unstable asthma ('brittle asthma') which also affects young females has also been described;[26] this can occur (especially in A and E departments) with paradoxical adduction of the vocal cords during inspiration.[27] Such patients can neglect to take medication at appropriate times and then ignore adequate management of the potentially dangerous consequences. This may lead to repeated admissions to hospital with medical emergencies such as diabetic ketoacidosis, status asthmaticus, or even pseudo status (simulated status epilepticus).

Factitious disorders are differentiated from somatoform disorders, where physical symptoms, although not caused by physical disease, are deemed not to be intentionally produced. Patients with factitious disorder, although they may state that they are not aware of the motives that drive them, voluntarily produce their physical or psychological symptoms. The disorders may overlap, however, and non-wandering female patients with factitious physical disorders may have more in common with those females who have somatoform disorders than with men with factitious disorder or with malingerers. Fink[28] found that 20 per cent of patients with

persistent somatization (i.e. patients with more than six admissions to the general hospital with medically unexplained symptoms) also had a factitious illness. One of the authors (Christopher Bass) has also found coexisting chronic somatoform and factitious disorders in the female perpetrators of factitious or induced illness.[29]

A more difficult distinction is between factitious disorder and malingering. Malingerers, described below, have clear-cut goals, often personal profit, and lack a history of hazardous, unnecessary invasive procedures. In our opinion the boundaries between the two disorders are more porous than the glossaries would have us believe, and we have stated above that the differentiation between conversion disorders, factitious disorders and malingering is extremely difficult in clinical practice. Case reviews have demonstrated how behaviour may shift from somatization to factitious to malingering when patients are followed longitudinally.[30] In our opinion the clinical status and diagnostic legitimacy of factitious disorder as a selective medical disorder is questionable, because it fails to take account of a morally questionable but volitional-based choice to deceive others by feigning illness. Considered as an act of wilful deception, illness deception can be meaningfully conceptualized within a socio-legal or moral model of human nature that recognizes the capacity for choice and the potential for pursuing benefits associated with the sick role. This model, which recognizes the human capacity to exercise free will, is shown diagrammatically in Fig. 5.2.9.1.

Epidemiology

Factitious disorders are relatively uncommon but probably underdiagnosed. Prevalence depends on clinical setting and the investigators' index of suspicion. Factitious disorder (probably underreported) is probably more common than full-blown Munchausen syndrome (probably overreported). In a recent survey of physicians from Germany, frequency estimates of factious disorder among their patients averaged 1.3 per cent, with dermatologists and neurologists giving the highest estimations.[31]

Of 1288 liaison psychiatric consecutive referrals, seen in a North American general hospital, 0.8 per cent had factitious disorder.[13] Similar figures have been reported by Dutch investigators.[14]

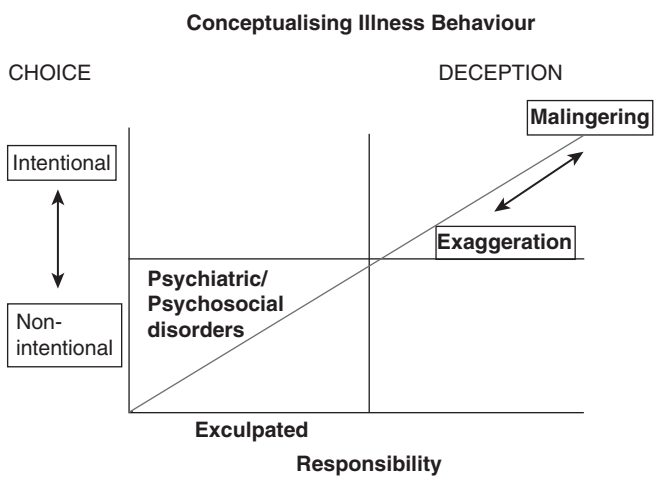


Fig. 5.2.9.1 Model of illness deception incorporating patient choice, free will and intentionality. (Reproduced from Bass, C. and Halligan, P. Illness related deception: social or psychiatric problem? Journal of the Royal Society of Medicine, **100**, 81–4. Copyright 2007, The Royal Society of Medicine Press.)

Prevalence rates for factitious disorder with psychological symptoms in patients under the age of 65 in a psychiatric hospital are approximately 0.4 per cent.[32]

Aetiology

There is little aetiological information, as large studies are lacking and the self-reported histories of many patients are fallacious. However, a number of themes are apparent[13]:

Developmental

- Parental abuse, neglect, or abandonment: Many factitious disorder patients experienced significant deprivation in childhood that left them with unfulfilled cravings for attention and care. These patients then seek to gratify these dependency needs by creating illness to obtain the 'attention and care' of the medical system
- early experiences of chronic illness or hospitalization

Medically related

- a significant relationship with a physician in the past
- experiences of medical mismanagement leading to a grudge against doctors
- paramedical employment

Physical

- Organic brain disorder: There is increasing evidence that neurobiological factors have a role in some patients, and it has been recommended that screening for evidence of brain dysfunction be carried out in these patients.[33] It is thought however that it is the pseudologia fantastica, and not the factitious disorder per se, that is associated with brain dysfunction (see previous section on Pathological Lying), though such a distinction may be difficult to support in clinical practice.

The authors have found it useful to conceptualize the problem using a cognitive behavioural formulation suggested by Fehnel and Brewer.[34] This allows the assessors to examine for relevant developmental factors and recent life events (especially losses or threatened losses and separations; see Fig. 5.2.9.2).

Course and prognosis

Factitious disorder may be limited to one or more brief episodes, but is usually chronic. For example, of 10 patients identified in a general hospital setting and followed up, at least one was known to have died as a result of factitious behaviour 4 months after the index admission.[13] Only one of the remaining nine patients accepted psychiatric treatment after discharge from hospital. Other authors have, however, reported a less gloomy outcome.[10] Outcome may be determined by how patients are managed, once their deceptions become manifest. Regrettably, the unmasking of the disorder is often the end of the story. Psychological support following hospital discharge may be associated with improved outcome.[10] Non-wanderers with more stable social networks may have a better prognosis than wanderers.[14] Engaging a patient with factitious disorder in long-term psychological treatment occurs so rarely that it often becomes the subject of a case report. One such report, of a 20-year follow-up of a patient with factious disorder with a favourable outcome is, in the opinion of the authors, the exception to the rule.[35]

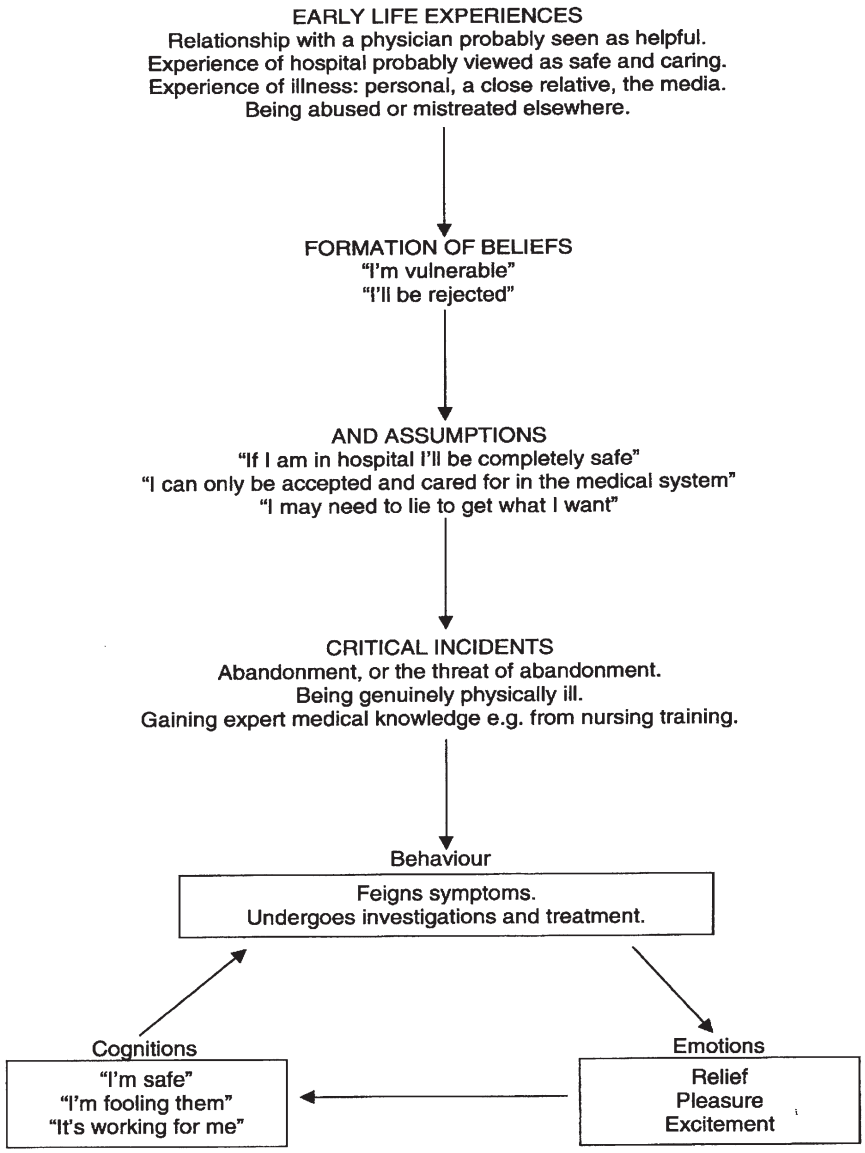


Fig. 5.2.9.2 A cognitive behavioural conceptualization of factitious disorder. Reproduced from Kinsella, P. Factitious disorder: a cognitive behavioural perspective, *Behavioural Cognitive Psychotherapy*; **29**: 195–202, copyright 2001, with permission from Cambridge University Press

Treatment

There are no systematic or controlled treatment studies on patients with factitious disorders. This is hardly surprising, as the patient's primary motive is deception, and the doctor's is to understand or unmask these motives, usually leading to rapid discharge from hospital after the deception has been exposed.

Management

Once the diagnosis is established, the doctor–patient relationship may have become irreparably damaged: negative emotions in the doctor may need to be dealt with before any consideration can be given to 'engaging' the patient in any therapeutic endeavour. Ethical and legal issues may also intrude (see below) and affect management. Although psychotropic medications have been used, the main treatment is psychological, using either confrontational or non-confrontational strategies.

Before treatment takes place however, it is important to establish the diagnosis, which is nearly always initially made by a non-psychiatrist, e.g. A and E physician, infectious disease specialist. A meeting should take place between the physician/surgeon and the psychiatrist and a strategy worked out before any confrontation or other approach is embarked on. These preliminary procedures are important and preparation before the joint interview is crucial (Table 5.2.9.1).

(a) Confrontational approaches

This process is easier if the physician has tangible evidence of fabrication, for example catheters, or medication used in the patient's deception. It is also desirable to have the psychiatrist present when

Table 5.2.9.1 Supportive confrontation: preparation and process (for non-psychiatrists)

- Collect firm evidence of fabrication, e.g. catheter, syringe, medication
- Discuss with psychiatrist (or hospital legal team if no psychiatrist available)
- Arrange meeting to marshall the facts; discuss strategy. Discuss with GP
- CONFRONTATION with patient should be non-judgemental, non-punitive, with . . .
- Proposal of ongoing support/follow-up
- If patient is a health care worker, the doctor should discuss with his/her medical defence organization
- Discuss the outcome of the confrontation with the patient's GP
- Document full record of the meeting and proposed outcome in patient's notes

the physician confronts the patient. The approach during confrontation and thereafter should be non-punitive and supportive, stressing continuity of care, and that the patient is a sick person who needs help. This approach was adopted in perhaps the largest published series of patients with factitious disorder treated systematically.[10] Thirty-three patients were 'confronted' with objects found in their room or with clinical data showing that their conditions were factitious. Only 12 (36 per cent) patients acknowledged the truth; the remaining 21 continued to deny that they played any role in creating their disorders. No confronted patient developed serious psychological disturbance or became suicidal, or discharged themselves against medical advice. Four of the most chronic cases became asymptomatic. Most, however, greeted the idea with either overt hostility or passivity and covert negativism.

More recently Krahn *et al.*[12] replicated these findings, but found that only one in six of their patients acknowledged their factious behaviour. Many patients will experience confrontation as humiliating and seek care from a different hospital. Others will refuse to see the psychiatrist with the treating clinician to discuss the deception, and discharge themselves against clinical advice.

(b) Non-confrontational strategies

These approaches, advocated by Eisendrath and Feder,[36] are less concerned with the origin of the illness and more with shaping future behaviours. Face-saving is a key element, and it is important for the patient to subsequently explain their 'recoveries' without admitting that their original problems were psychiatric.

One strategy is the therapeutic 'double-bind'. In this approach the patient is presented with two choices: prove that his or her disorder is not factitious by responding to a relatively minor and benign medical intervention, or prove that the disorder is factitious by failing to respond. For example, a woman was offered the double-bind for a wound that had failed to heal in 4 years despite numerous surgical closures. Following this strategy the plastic surgeon told her that her wound should respond to a skin grafting procedure. If it did not, it would mean that her disorder was factitious in origin. The graft took place, and there was no recurrence of infection at 2-year follow-up.[36] This approach has also been used with some success in the rehabilitation of three patients with factitious motor disorders.[37] The strategy was successful in providing patients with a face-saving legitimization of both their illnesses and recoveries.

Another face-saving approach uses 'inexact interpretations', i.e. suggesting a relationship between certain events or stressors, for example being abandoned, and emergence of factitious symptoms. It involves presenting a brief formulation of the problem to the patient, stopping short of overtly identifying the factitious origin. By avoiding confrontation the doctor makes it safe for the patient to relinquish the symptom with a feeling of control. Regrettably, none of these non-confrontational techniques have been evaluated in a systematic fashion.

(c) Systemic interventions

Patients with factitious disorders can create havoc on medical and surgical wards. They often elicit negative and hostile emotions in general hospital staff, especially after the deception has been exposed. The psychiatrist can help staff members to vent and reduce the anger they experience when a factitious diagnosis is confirmed, and also help the staff to understand the likely mechanisms underlying the factitious behaviour. These issues are often best addressed at a multi-disciplinary staff meeting. The major task of this group, which should include a member of the hospital medico-legal department as well as the patient's family doctor, is to develop practical treatment guidelines and to discuss the complex legal and ethical issues raised with factitious physical disorders. Some of these issues are discussed in the next section.

Ethical and legal issues

Patients with factitious disorders create unique ethical and medico-legal issues, some of which will be described below.

(a) Confidentiality

If no meaningful doctor–patient relationship exists or can be established, it has been argued that the physician is not bound by ethical codes, and that drastic solutions such as keeping 'blacklists' and the use of a central register can therefore be justified.[38] Objections to these approaches include breach of doctor–patient confidentiality and possible denial of treatment for genuine illness. Anyway, the use of aliases and poor record-keeping reduces the effectiveness of blacklists. Furthermore, physicians who disclose information without patient consent may have to justify the decision to their licensing body. In the United States, there is a consensus that disclosure should only occur where there is a specific risk to the patient and/or another party. In such situations, a multi-disciplinary staff meeting can help to develop treatment policy, and share responsibility for difficult decisions (see above).

(b) Invasion of privacy

The medical literature contains many descriptions of how the diagnosis of factitious disorder was established following a search of the patient's room or belongings. Some physicians, however, consider that such behaviour infringes patients' rights, and that no search should be undertaken without the patient's knowledge and consent. One way of avoiding this dilemma is to make it clear to the patient that factitious disorder is among the differential diagnoses, and then request permission for a room search. If needles or syringes are discovered during the course of treating the patient, the ethical issue of invasion of privacy does not arise.

(c) Involuntary hospitalization or treatment

Because the patient with factitious disorder may engage in behaviour that leads to permanent maiming or even death, it has been argued that in such cases a compulsory order may be used to protect the patient from himself or herself. This will provide time for

not only a more in-depth psychiatric assessment but also the development of a more trusting relationship with a therapist. This is a contentious subject, but some case reports do indicate that extended involuntary hospitalization may result in therapeutic progress.[39]

Induced factitious illness (Munchausen syndrome by proxy)

Meadow[40] first described this disorder as 'the deliberate production or feigning of physical or psychological symptoms or signs in another person who is under that individual's care', which has recently been renamed fabricated or induced illness or FII.[29] The perpetrator is usually the mother and the victim her child: the syndrome is considered to be a form of child abuse. The parent's aim is to have the child considered seriously ill: this may involve providing false histories, poisoning, or persuading doctors to carry out invasive and potentially dangerous procedures.

Factitious disorder and FII can be interrelated. Psychiatrists (and general practitioners) should be aware of the implications of a diagnosis of factitious disorder for any children of the index patient. For example, 75 per cent of mothers of these children have a history of a factitious or somatoform disorder, and most meet criteria for personality disorder.[41] With the birth of a child some mothers with pre-existing factitious disorder abandon dissimulation themselves, only to extend it to the next generation through factitious disorder by proxy. Because factitious disorder and FII can co-occur, the finding of one should always trigger judicious efforts to establish, or hopefully disconfirm, the other.

Long-term outcome in factitious disorder by proxy is poor, so active early intervention is recommended, with child protection agencies working closely with paediatric and psychiatric services (see also Chapter 9.3.3).

Factitious disorder in health care workers

If factitious disorder is diagnosed in a health care worker, the investigating psychiatrist should consider whether their continuing clinical work would pose risks to either patients (often children) with whom that person comes into contact (i.e. of factitious disorder by proxy), and/or the health of the factitious disorder patient himself or herself. These issues have been thrown into sharp focus by some highly publicized cases in both the United States and the United Kingdom.[42,43] Any employee who manufactures crises, for example multiple cardiac arrests and resuscitations, is obviously of great concern. The British commission that investigated the 1993 case made several stringent recommendations, for example 'We recommend that no candidate for nursing in whom there is evidence of major personality disorder should be employed in the profession'.[44] Nevertheless, psychiatrists should seek medico-legal advice before communicating concerns about such patients to any third party, including the employing hospital.

Malingering and exaggeration

Introduction

Malingering is the deliberate simulation or exaggeration of physical or psychiatric symptoms for obvious and understandable gain (e.g. financial compensation, disabled status and benefits, avoidance of criminal prosecution or conscription).

Treating doctors have been understandably reluctant to diagnose malingering lest it adversely affect the patient. However, recent research indicates that it may not be uncommon, especially where financial rewards attach to disability status, such as benefits for sickness or compensation for injury. We here describe some types of malingering seen in psychiatry, and discuss newer specialist psychological tools, especially *symptom validity testing*, which are emerging as useful additions to the overall assessment.

Definition

Malingering is not coded as a mental disorder in either ICD-10 or DSM-IV, although in the latter it is denoted as an 'additional condition that may be the focus of clinical attention'.

DSM-IV[2] suggests that malingering should be 'strongly suspected' when two or more of the following factors apply:

1 medico-legal context

2 antisocial personality disorder

3 discrepancy between complaints and objective findings

4 lack of co-operation with the assessment

This actually sets a very low bar: in effect, DSM advises it should be 'strongly suspected' in any disputed medico-legal case (where points 1 and 3 apply), but does not give guidance as to how this 'suspicion' should be followed up. The ICD-10 definition ('the intentional production or feigning of either physical or psychological symptoms or disabilities, motivated by external stresses or incentives') is broad, but again does not give practical suggestions regarding assessment.

Malingering includes:

1 Pure malingering: complete fabrication of symptoms

2 Partial malingering: exaggerating real symptoms or saying that past symptoms are continuing

3 False attribution: falsely saying that real health problems are due to a compensable accident or other circumstance

Epidemiology

Until recently, there has been little systematic information on prevalence, although it has generally been agreed that exaggeration of real symptoms is more common than outright faking expect them to dissimulate. Exaggeration, dissimulation, or feigning can be considered one of several rational/economic/adaptive options open to patients when seeking health care and/or limited social and welfare resources.

Using the 'Composite Disability Malingering Index' Griffin *et al.*[45] suggested that 19 per cent of disability claimants in the United States malingered to some degree. A recent study of 131 practicing members of the American Board of Clinical Neuropsychology provided estimates of the prevalence of malingering and symptom exaggeration for a variety of different clinical conditions.[46] In this study, estimates of the base rate of malingering/symptom exaggeration were calculated using over 33 000 annual cases seen by a group of clinical neuropsychologists. The reported base rates (when statistically adjusted to remove for the influence of referral source) were 29 per cent for personal injury, 30 per cent in the case of disability or workers compensation, 19 per cent in criminal cases, and 8 per cent in medical or psychiatric cases. The same rates

broken down by diagnosis revealed 39 per cent in the case of mild head injury, 35 per cent in fibromyalgia and chronic fatigue, 31 per cent in chronic pain, 15 per cent for depressive disorders, and 11 per cent in the case of dissociative disorders. In a separate review of 1363 compensation-seeking cases, Larrabee[47] found similar figures for mild head injury of 40 per cent. The use of symptom validity testing has confirmed these high rates of exaggeration (see under 'Psychological tests'). For example, about 42 per cent in the Canadian series of Richman *et al.*[48] and about 60 per cent in the United Kingdom in a series studied by one of the authors (Gill *et al.* submitted for publication).

Clinical features

(a) Malingered neurosis: post-traumatic stress disorder

The archetypal disorder after trauma is post-traumatic stress disorder. Malingering of other disorders seen after trauma, notably depression is less frequent, confirming malingerers' preference for dramatic positive symptoms such as nightmares and flashbacks.

In 1983 Sparr and Pankratz[49] described five men who claimed to have been 'traumatized' in the Vietnam War; three claimed to have been PoWs. It turned out that none had been PoWs, four had never been in Vietnam and two had never been in the services. The patients were seeking the generous benefits, which the United States accords to ex-service (Veterans' Administration) personnel. The key point in this paper was that the authors *supplemented their clinical assessment by seeking external data*, in this case service records; relying on clinical assessment alone would have led to wrong diagnosis. Seeking corroboration is vital in the assessment of possible malingering.

Rosen, in another classic paper, documented malingered PTSD in the case of the *Aleutian Enterprise*, a fish-processing ship, which sank in the Bering Sea in 1990.[50] Of the 31 on board, 9 were lost, 2 went back to sea, and the remaining 20 sued. Nineteen (86 per cent) of these 22 survivors consulted psychiatrists or psychologists with the key features of PTSD. But this is much higher than the expected rates: most individuals exposed to a traumatic event do not develop PTSD. Even if they do, it resolves over a few weeks or months in many cases, whereas here the claimants' symptoms did not show any tendency to resolve. Furthermore, they all had almost all of the classic features of the condition, in other words, they did not display the case to case variability, which would be expected in real individuals. Rosen documented that the patients had in fact 'shared symptoms' and had been 'coached' by their attorneys, some of whom had advanced the claimants money so they would not have to settle.

Most cases of suspected malingered post-traumatic stress disorder involve less dramatic civilian trauma, most commonly road accidents. Again, exaggeration of genuine symptoms is much more common than outright fabrication. For example, a person involved in a car accident, with apparently genuine phobic travel anxiety, may report that they have nightmares and 'flashbacks', but their description of these experiences lacks vividness on close enquiry. Holistic assessment is vital; if a minor accident is to be accepted as causing a severe psychiatric condition such as PTSD, there must be some evidence of pre-existing vulnerability. Otherwise, the apparent result will be disproportionate to the cause, and the possible influence of external incentives on symptom presentation will need to be considered.

(b) Malingered psychosis

This disorder can occur in various circumstances, for example in homeless persons wishing to obtain shelter in hospital, in previously psychotic inpatients whose discharge is imminent, in illegal migrants seeking to avoid deportation or in criminal defendants trying to avoid standing trial or to influence sentencing.

A (perhaps somewhat academic) distinction can be made between malingered and factitious psychosis, in that malingerers are conscious of their motivation, and their goal is not merely confined to gaining patient status. However, the following description covers both. It is 'positive' symptoms, which are usually mimicked (e.g. hallucinations). They are often dramatic or bizarre. Patients are keen to describe them at interviews, unlike, for example, most patients with schizophrenia. Symptoms are obvious during assessments, less so when the patient is unobserved. More subtle features of genuine psychosis, such as thought disorder and negative symptoms, are absent. Florid hallucinations may be unaccompanied by delusions, which would be unusual in genuine psychosis.

For example, Jaffe and Sharma[51] described nine defendants on serious criminal charges that developed uncommon symptoms such as coprophagia and 'seeing little green men'. Eight were judged to be malingering, and fit to plead, based on evidence such as the association between visual hallucinations and organic brain syndromes, of which there was no evidence on investigation. Malingering in forensic settings also includes feigned memory deficits, when isolated amnesia for an alleged crime would place malingering in the differential diagnosis. Forensic psychiatry is dealt with elsewhere in this book.

Ganser syndrome does not appear in current classifications. It was originally described in prisoners, as comprising confusion and so-called 'approximate answers' (or Vorbereiden: *Question*: for example: how many legs has a horse? *Answer*: three). If true confusion (diminished level of consciousness) is present, the diagnosis is the cause of this. Approximate answers may be seen in several conditions, including mental retardation, organic brain disorders, and malingering; again, the diagnosis will be the underlying condition. The term Ganser syndrome has now largely and rightly been dropped.

(c) Malingered cognitive deficit

Study of cognitive deficits following brain injury has recently led to advances in understanding of malingering, through the development of special neuropsychological tests, especially *effort testing*, to gauge the effort the patient brings to cognitive testing. These tests seem likely to have broader application than just brain injury assessment.

Discussion of the question of malingering post head injury has until recent years been rich in opinion, though comparatively light on facts. Miller's view,[52] long influential, was that many patients malingered their memory and other cognitive symptoms and that symptoms were in inverse proportion to injury severity and were only resolved with receipt of compensation. He used the term 'accident neurosis'. Mendelson, by contrast,[53] found that disability continued after settlement in many patients, and inferred from this that disability was generally not malingered.

The question is obviously not capable of being resolved scientifically without data. But such data is now becoming available. Recent findings however have supported Miller's original observations that embellishment rises as injury severity decreases in a

compensable context.[54] Moving forward from mere debate, the American Academy of Neuropsychologists recently published a consensus statement which concluded that 'Symptom exaggeration or fabrication occurs in a sizeable minority of neuropsychological examinees, with greater prevalence in forensic contexts', and that the use of effort testing is mandatory in neuropsychological assessments.[55]

In clinical assessment, immediate recall (e.g. digit span) is important, as even organic amnesic patients (e.g. Korsakoff's syndrome) perform normally; poor performance suggests that poor motivation or malingering should be among the differential diagnoses. Claims of being unable to remember personal information (e.g. name and birthday, also preserved in organic amnesia), yet having been able to come to the assessment independently, are highly suggestive of malingering, but seen only in gross cases. However, these 'bedside' tests are only a guide: psychiatric assessment of brain injury patients is not complete without quantitative assessment of cognitive function, including effort testing (see below).

(d) Malingered physical disease

This usually presents either as a referral to a liaison psychiatrist, or in a medico-legal context. A frequent example is a patient with post-injury back or neck pain who is involved in litigation or seeking disability payments. Often, some form of accident has undoubtedly occurred, so the potential for initial physical injury is not in doubt; but the length and severity of symptoms, disability, and distress may seem out of proportion, and raise the possibility of malingering.

In a seminal paper, Richman[48] administered effort testing to 106 people claiming injury or sickness benefits. Forty-five (42 per cent) failed. On one easy subtest, those who failed the effort test overall had a similar score on average to patients with dementia tested previously, even though none of them had a clinical diagnosis of dementia. Schmand et al.[56] found that 61 per cent of litigants after whiplash neck injury had evidence of underperformance on memory testing compared with 29 per cent of outpatient controls. The underperforming litigants scored as low as controls with definite evidence of closed head injury.

Classification

Malingering should be distinguished from factitious disorder, and from other syndromes such as hypochondriasis, other somatoform disorders, and conversion/dissociation disorder. These distinctions may involve difficult judgements such as how 'intentional' is the production of a symptom, or how 'genuine' it is. As an alternative, it has been suggested that such patients lie on a continuum between those in whom the production of symptoms is assumed to be wholly unconscious (conversion/dissociation disorder) and those in whom it is wholly conscious (malingering, factitious disorder).

However, use of the concept of the unconscious has to be very cautious when there are external incentives. It stretches credulity to think that a claimant would be conscious of there having been an accident so as to pursue litigation, and conscious that the outcome could include financial compensation, but somehow unconscious that presentation of symptoms could form a desired link between the two.

It is possible that the emergence of effort testing may cast new light on the area of unexplained physical symptoms. For example, the concept of somatoform disorders assumes that the symptoms are not consciously produced. However, if large-scale studies reveal that a substantial proportion of somatoform patients turn out to fail effort tests, or in other words to display evidence of conscious symptom exaggeration, then the concept of somatoform disorders may need to be re-examined.

Diagnosis and differential diagnosis

Doctors' training and culture rightly encourage the treating physician towards a generally trusting relationship with his patients. However, assumptions that patients' accounts are generally trustworthy and that malingering is rare are not appropriate if the main responsibility of the doctor lies elsewhere, for example, to the Court, if he is preparing an expert report.

Identifying ungenuine cases requires an enquiring approach, and the methodical use of all sources of information:

- awareness of the possibility of exaggeration or faking of symptoms
- neutral attitude
- open questions initially; use closed questions with caution
- unlikely questions (see below)
- mental state—changes appropriately as sensitive topics discussed?
- informants (but they may also have vested interests)
- observation—overt (e.g. in a ward) or covert (e.g. video surveillance)
- medical records (and legal documents if applicable)
- look for consistency of accounts
- standard psychometric tests—consistency of results across measures
- specialist instruments: symptom validity tests

The medical notes should be read, ideally before interview, especially the general practice records, and any discrepancies noted for specific enquiry. Legal assessments, case papers and previous reports should be studied.

In the clinical interview, a neutral attitude is essential; a confrontational approach, even if malingering is strongly suspected, may cause further exaggeration of symptoms. Open questions should be used at first. Closed questions should be avoided (e.g. 'Do you ever get nightmares where you seem to re-experience the accident?'). The careful use of unlikely questions can be useful. For example, in suspected malingered post-traumatic stress disorder the answer 'Yes' to the question 'Have you had any problems with colour vision since the accident?' would be suggestive, but would still need clarification with open questions. 'I am now colour-blind' might suggest malingering, but 'Red makes me nervous—it was a red car which crashed into me' might not.

Questions must be appropriate to the case, ideally prepared beforehand and introduced tactfully to prevent the patient feeling that the interviewer is attempting to 'catch him out'. The interviewer should look for the clinical characteristics of the particular malingered disorder in question.

Surveillance by video or other means may be used by lawyers or insurers, although is seldom initiated by clinicians, for ethical reasons (except perhaps in suspected factitious disorder by proxy—see Chapter 9.3.3).

Psychological tests

Since the last edition of this book, consensus has developed that symptom validity testing is essential in neuropsychology, that is, in the field primarily concerned with memory problems and brain injury. The tests in practical use are in fact specialized memory tests, and it is likely that their use will be extended from head injury cases to other cases in which memory complaints occur (such complaints are in fact very common in patients with pain or distress for any reason). We are about to discuss what currently appear to be the leading symptom validity tests in UK neuropsychology, the Word Memory Test[57] and the ToMM,[58] but first we mention some other tests by way of historical background.

Inconsistent patterns of response on standard instruments (e.g. high scores on one measure of depression and low on another) might be suggestive. However, this should not be overinterpreted, because there is substantial test–retest variation on many tests, and there is also variation between subscales on the same occasion. Certain subscales of the Minnesota Multiphasic Personality Index have been proposed as measures of tendency to malinger. Specialist instruments have been developed, for example the Structured Inventory of Malingered Symptoms,[59] which identifies features associated with malingering such as endorsement of rare symptoms, which occur only infrequently in clinical populations. However, consensus on such tests remains some way off and they are not in general use in the United Kingdom.

Tests for malingered memory deficits present memory tasks, which appear difficult but are in fact easy. For example, 50/50 psychiatric and 10/16 mentally retarded inpatients were able to recall nine of the 15 items on the Rey 15-item.[60] Forced choice testing (e.g. Portland Digit Recognition Test) is another approach. A sequence of digits is presented. Subjects must identify it among two further sequences, one identical and the other different. By chance alone, they must score around 50 per cent on a large number of items, and so below chance scores (below 50 per cent) strongly suggest malingering.[61] However, this is very gross, and will miss many cases where the degree of exaggeration is less. Again, these tests are not in general use in the United Kingdom.

In the United Kingdom, the Word Memory Test and the ToMM have become established as leading symptom validity tests in neuropsychology. They are regarded as having the best research support and are administered in standard formats.

TOMM stands for *Test Of Malingered Memory*.[58] It is a pictorial test, not computerized, and is mainly used by psychologists. Word Memory Test (WMT) (Green: wordmemorytest.com) is available either on paper or on computer. It is mainly used by psychologists as part of a 'battery' of neuropsychological tests. An abbreviated form for physicians, the Medical Symptom Validity Test, is available.[57] The key point is that patients are given a test of memory, which looks difficult, but is in fact known to be easy from previous administration to control subjects.

Someone making a **good effort:**

- scores well on tests which are in fact easy (even though they may look hard)
- scores lower on more difficult tests

Someone making an **inconsistent or poor effort**:

- may score low on tests which look hard (though they are in fact easy)
- may not score lower on more difficult tests

If the patient's scores are low, where even say primary school children score almost perfectly, it would suggest that he might not have been making a full effort on the test. If at the end of the test he said that he had made a full effort, this would be evidence that his self-report of effort brought to cognitive testing is not accurate. This would be consistent with the proposition that he might have been exaggerating at least the memory aspects of his complaints, and by extension, possibly other symptom areas also.

Aetiology

Malingering is not a mental disorder, so complex general theories about causation are unlikely to be helpful. It has been suggested for example that[62] 'adaptation' is the simplest model for malingering: ' . . . malingering is more likely to occur when the evaluation is perceived as adversarial, when the personal stakes are very high, and when no alternatives appear to be viable'. However, this really just restates the problem in different terms. Nor do the 'response styles' described by psychologists seem to offer fundamental insights. It is more helpful to consider each case in a common sense way, bearing in mind the presence or absence of external incentives, and the results of the holistic assessment process outlined above, in combination with the results of symptom validity testing.

Course and prognosis

Malingerers are so heterogeneous that it is impossible to state prognosis in general. There are few adequate follow-up studies.

In forensic and inpatient settings, malingering is usually episodic, associated with particular circumstances such as impending discharge, trial, or change in conditions of imprisonment (e.g. transfer to a single cell or hospital wing). The behaviour often stops when the circumstances no longer remain, although they may recur. Similar situations occur in persons facing conscription, or in migrants at risk of repatriation.

The prognosis of malingering *in personal injury litigants* is unknown. Studies indicating poor outcome of many compensated litigants cannot be extrapolated, as the number of malingerers in these samples is unknown. Clinical experience suggests that patients with long-standing disability, even if partly or wholly non-organic, often fail to recover fully in any event.

Treatment

There is very little evidence on management, which will largely be dictated by whether the clinician has clinical responsibility, or whether he or she has been asked to give an opinion to a third party. Even if the diagnosis of malingering is clear, it may be appropriate to inform the referrer, rather than the patient directly, as the patient may become very angry. If there is a psychiatric disorder present when, as it were, the dust has settled, then this should be treated in standard fashion.

Clinical experience is that patients for whom the exercise of attempted malingering seems worthwhile do often have substantial pre-existing psychosocial problems, including lack of skills or employment; there is a tendency towards a regionality of such claims, with areas of deprivation ('rust-belt') overrepresented.

Efforts at retraining and vocational rehabilitation may be more likely to be of assistance in the long-term than specialist psychiatric care.

Possibilities for prevention

These seem mainly confined to malingering after injury. First, systems of litigation should be expedited. Prolonged cases certainly make for exaggeration of symptoms and disability. Second, some patients with chronic disability, irrespective of cause, do respond to rehabilitative treatment (e.g. programmes assisting in return to work or in the management of chronic pain), even at a late stage. If such programmes were easily and generally available at an earlier stage in the evolution of the disorder, symptoms and subsequent disability could certainly be ameliorated or even prevented in some patients.

Ethical legal and personal issues

As previously indicated, malingering poses a number of particular challenges to the doctor himself and to the doctor–patient relationship, including amongst others the following.

There is no doubt that seeing substantial numbers of likely ungenuine patients has the potential to affect the practitioner himself. Appropriate 'supervision', as in psychotherapy, may be appropriate.

If the doctor has a treatment responsibility to the patient, he naturally tends to give the patient the benefit of doubt, for example in respect of disability payments. But he will be wise to be cautious: if, for example, he signs a form saying that the patient cannot undertake certain activities of daily living, but without direct observation of that, he could potentially be considered to be an accessory if the claim was subsequently found to be fraudulent.

The use of the word 'malingering' itself in reports must usually be avoided, as it is tantamount to an accusation of a criminal offence such as deception. The medical duty is to present the evidence to the Court, which can then decide the question itself.

Finally, it must not be forgotten that the Data Protection Act gives patients the right to see personal information about themselves such as medical reports. Any suggestion of ungenuine presentation in reports must therefore be well-founded on evidence and properly argued.

Further information

Malleson, A. (2002). *Whiplash and other useful illnesses*. Magill-Queens University Press.

Halligan, P., Bass, C., and Oakley, D. (2003). Willful deception as illness behaviour. In: *Malingering and illness deception* (eds. P. Halligan, C. Bass, and D. Oakley). Oxford University Press.

Vrij, A. (2001). *Detecting lies and deceit*. Wiley.

References

1. Asher, R. (1951). Munchausen's syndrome. *Lancet*, **1**, 339–41.
2. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
3. Turner, M. (2006). Factitious disorders. Reformulating the DSM-IV criteria. *Psychosomatics*, **47**, 23–32.
4. Bass, C. and Halligan, P. (2007). Illness related deception: social or psychiatric problem? *Journal of the Royal Society of Medicine*, **100**, 81–4.
5. Dike, C., Baranoski, M., and Griffith, E. (2005). Pathological lying revisited. *The Journal of the American Academy of Psychiatry and the Law*, **33**, 342–9.
6. King, B. and Ford, C. (1988). Pseudologia fantastica. *Acta Psychiatrica Scandinavica*, **77**, 1–6.
7. Spence, S., Hunter, M., Farrow, T., *et al.* (2004). A cognitive neurobiological account of deception: evidence from functional neuroimaging. *Philosophical Transactions of the Royal Society of London*, **359**, 1755–62.
8. Yang, Y., Raine, A., Lencz, T., *et al.* (2005). Prefrontal white matter in pathological liars. *The British Journal of Psychiatry*, **187**, 320–5.
9. Spence, S. (2005). Letter in *The British Journal of Psychiatry*, **187**, 326–7.
10. Reich, P. and Gottfried, L.A. (1983). Factitious disorders in a teaching hospital. *Annals of Internal Medicine*, **99**, 240–7.
11. Eisendrath, S.J. (1996). Current overview of factitious physical disorders. In *The spectrum of factitious disorders* (eds. M. Feldman and S. Eisendrath), pp. 21–36. American Psychiatric Press, Washington, DC.
12. Krahn, L., Honghzhe, L., and O'Connor, K. (2003). Patients who strive to be ill: factitious disorder with physical symptoms. *The American Journal of Psychiatry*, **160**, 1163–8.
13. Sutherland, A.J. and Rodin, G.M. (1990). Factitious disorders in a general hospital setting: clinical features and review of the literature. *Psychosomatics*, **31**, 392–9.
14. Hengeveld, M.W. (1992). Factitious disorders: what can the psychiatrist do? In *Practical problems in clinical psychiatry* (eds. K. Hawton and P. Cowen), pp. 118–29. Oxford University Press, Oxford.
15. Kerr, B., Foulkes, W., Cade, D., *et al.* (1998). False family history of breast cancer in the family cancer clinic. *European Journal of Surgical Oncology*, **24**, 275–9.
16. Hadeed, V., Trump, D.L., and Mies, C. (1998). Electronic cancer Munchausen syndrome. *Annals of Internal Medicine*, **129**, 73.
17. Feldman, M and Hamilton, J. (2006). Serial factitious disorder and Munchausen by proxy in pregnancy. *International Journal of Clinical Practice*, **60**, 1675–8.
18. Taskaynatan, M.A., Balaban, M., Karlidere, T., *et al.* (2005) Factitious disorders encountered in patients with the diagnosis of reflex sympathetic dystrophy. *Clinical Rheumatology*, **24**, 521–6.
19. Verdugo, R. and Ochoa, J. (2000). Abnormal movements in complex regional pain syndrome: assessment of their nature. *Muscle & Nerve*, **23**, 198–205.
20. Eisendrath, S. and McNeil, D. (2004). Factitious physical disorders, litigation, and mortality. *Psychosomatics*, **45**, 350–3.
21. Peebles, R., Sabella, C., Franco, K., *et al.* (2005). Factitious disorder and malingering in adolescent girls: case series and literature review. *Clinical Pediatrics*, **44**, 237–43.
22. Libow, J. (2000). Child and adolescent illness falsification. *Pediatrics*, **105**, 336–41.
23. Parker, P. (1996). Factitious psychological disorders. In *The spectrum of factitious disorders* (eds. M. Feldman and S. Eisendrath), pp. 37–49. American Psychiatric Press, Washington, DC.
24. DeVaul, R.A. and Faillance, L.A. (1978). Persistent pain and illness insistence. A medical profile of proneness to surgery. *American Journal of Surgery*, **135**, 828–33.
25. Kent, L., Gill, G., and Williams, G. (1994). Mortality and outcome of patients with brittle diabetes and recurrent keto-acidosis. *Lancet*, **344**, 778–81.
26. Barnes, P. and Cheung, K. (1998). Difficult asthma. *British Medical Journal*, **299**, 695–8.
27. Renz, V., Hern, J., Tostevin, T., *et al.* (2000). Functional laryngeal dyskinesia: an important cause of stridor. *The Journal of Laryngology and Otology*, **114**, 790–2.
28. Fink, P. (1992). The use of hospitalisations by persistent somatizing patients. *Psychological Medicine*, **22**, 173–80.
29. Bass, C. and Adshead, G. (2007). Fabrication of illness in children: the psychopathology of abuse. *Advances in Psychiatric treatment*, in press.
30. Eisendrath, S and McNeil, D. (2002). Factitious disorders in civil litigation: twenty cases illustrating the spectrum of abnormal illness-affirming behaviour. *The Journal of the American Academy of Psychiatry and the Law*, **30**, 391–9.

31. Fliege, H., Grimm, A., Eckhardt–Henn, A., *et al.* (2007). Frequency of ICD-10 factitious disorder: survey of senior hospital consultants and physicians in private practice. *Psychosomatics*, **48**, 60–4.
32. Bhugra, D. (1988). Psychiatric Munchausen's syndrome. Literature review with case reports. *Acta Psychiatrica Scandinavica*, **77**, 497–503.
33. Diefenbacher, A. and Heim, S. (1997). Neuropsychiatric aspects in Munchausen syndrome. *General Hospital Psychiatry*, **19**, 281–5.
34. Fehnel, C. and Brewer, E. (2006). Munchausen's syndrome with 20 year follow up. *The American Journal of Psychiatry*, **163**, 547.
35. Kinsella, P. (2001). Factitious disorder: a cognitive behavioural perspective. *Behavioural and Cognitive Psychotherapy*, **29**, 195–202.
36. Eisendrath, S.J. and Feder, A. (1996). Management of factitious disorders. In *The spectrum of factitious disorders* (eds. M. Feldman and S. Eisendrath). American Psychiatric Press, Washington, DC.
37. Teasell, R.W. and Shapiro, A.P. (1994). Strategic–behavioral intervention in the treatment of chronic non-organic motor disorders. *American Journal of Physical Medicine Rehabilitation*, **73**, 44–50.
38. Powell, R. and Boast, N. (1993). The million dollar man: resource implications for chronic Munchausen's syndrome. *The British Journal of Psychiatry*, **162**, 253–6.
39. O'Shea, B., McGennis, A., Cahill, M., *et al.* (1984). Munchausen's syndrome. *British Journal of Hospital Medicine*, **35**, 269–74.
40. Meadow, R. (1977). Munchausen syndrome by proxy: the hinterland of child abuse. *Lancet*, **2**, 343–5.
41. Bools, C., Neale, B., and Meadow, R. (1994). Munchausen syndrome by proxy: a study of psychopathology. *Child Abuse & Neglect*, **18**, 773–84.
42. Elkind, P. (1989). *The death shift: the true story of nurse Genene Jones and the Texas baby murders*. Viking Penguin, New York.
43. Davies, N. (1993). *Murder on ward four*. Chatto and Windus, London.
44. *The Allitt enquiry*. (1994). HMSO, London.
45. Griffin, G. (1996). Assessing dissimulation among social security disability income claimants. *Journal of Consulting and Clinical Psychology*, **64**, 1425–30.
46. Mittenberg, W., Patton, C., Vanyock, E., *et al.* (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, **24**, 1094–102.
47. Larrabee, G. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, **17**, 410–25.
48. Richman, J., Green, P., Gervais, R., *et al.* Objective tests of symptom exaggeration in independent medical examinations. *Journal of Occupational and Environmental Medicine*, **48**, 303–311.
49. Sparr, L.D. and Pankratz, L. (1983). Factitious post traumatic stress disorder. *The American Journal of Psychiatry*, **140**, 1016–9.
50. Rosen, G. (1995). The Aleutian Enterprise sinking and post-traumatic stress disorder: misdiagnosis in clinical and forensic practice. *Professional Psychology: Research and Practice*, **26**, 82–7.
51. Jaffe, M.E. and Sharma, K.K. (1998). Malingering uncommon psychiatric symptoms among defendants charged under California's 'three strikes and you're out' law. *Journal of Forensic Science*, **43**, 549–55.
52. Miller, H. (1961). Accident neurosis. *British Medical Journal*, **1**, 919–25.
53. Mendelson, G. (1995). Compensation neurosis revisited: outcome studies of the effects of litigation. *Journal of Psychosomatic Research*, **39**, 695–706.
54. Grieffenstein, M. and Baker, J. (2005). Miller was (mostly) right: head injury severity inversely related to simulation. *Legal and Criminological Psychology*, **10**, 1–16.
55. Bush S., Ruff, R.M., Tröster, A.I., *et al.* (2005). Symptom validity assessment: Practice issues and medical necessity, NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, **20**, 419–26, NAN position paper.
56. Schmand, B., Lindeboom, J., Schagen, S., *et al.* (1998). Cognitive complaints in patients after whiplash injury: the impact of malingering. *Journal of Neurology, Neurosurgery and Psychiatry*, **64**, 339–43.
57. Word Memory Test (WMT) (Green: wordmemorytest.com)
58. Tombaugh, T.N. (1996). The test of memory malingering. Multi-Health Systems, Toronto, Canada.
59. Smith, G.P. and Burger, G.K. (1997). Detection of malingering: validation of the SIMS. *Journal of the American Academy of Science and the Law*, **25**, 183–9.
60. Goldberg, J.O. and Miller, H.R. (1986). Performance of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints. *Journal of Clinical Psychology*, **42**, 792–5.
61. Binder, L.M. and Willis, S.C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment*, **3**, 175–81.
62. Rogers, R. (1997). *Clinical assessment of malingering and deception*. Guilford Press, New York.

5.2.10 Neurasthenia

Felice Lieh Mak

Introduction

The term neurasthenia has had a variegated history, and although retained as a diagnostic entity in the ICD-10 it does not appear in the DSM-IV. In cultures where neurasthenia still enjoys popular professional and lay acceptance it has a variety of usages:

- a nosological entity
- an idiom for expressing distress
- a culturally sanctioned illness behaviour
- an explanatory model for a constellation of somatic symptoms
- an euphemism for avoiding the stigma of mental disorder.

Therefore, in diagnosing, understanding, and managing neurasthenia the clinician has to be aware of the context in which the term is used.

Concept and diagnostic entity

The concepts of nervous weakness and asthenia (debility, lack of strength) have existed throughout the history of medicine. Hippocrates described the illness of the Scythians as a general asthenia linked to damage to the genitalia caused by horseback riding. In France, Bouchut (1764) described a syndrome similar to the latter-day neurasthenia, which he called 'neuropathie'. Cullen (1772) conceived muscles and nerves as a unitary nervous force and all diseases as movements against the nature of that nervous force. He coined the word neuroses for this process and postulated that diseases were due to the various alternations of excitement and atony in the nervous system. A few years later, his pupil Brown (1780) elaborated on the hypothesis by dividing diseases into sthenic diseases, which were due to excessive excitement, and asthenic diseases, which were due to deficient excitement. These views on the polarity of the nervous system as a cause of mental illness set the scene for neurasthenia to become a disease entity.

By the beginning of nineteenth century the term neurasthenia was already in use. In 1869, Van Deusen in Holland published a monograph on neurasthenia. This was quickly followed by the publication of a paper, which Beard[1] had presented to the New York Medical Journal Association. Beard based his description of the disorder on a series of 30 cases. In reorganizing the subjective nature of the complaints and the unique clustering of symptoms in each patient, Beard had difficulties in attempting to limit the number of symptoms that constituted the syndrome; he started with 50 symptoms and expanded it to 75 in later publications.

Eventually it became clear that the expanding kaleidoscope of symptoms should be managed in a way that made some sense. Beard approached this problem by organizing the symptoms into subtypes of neurasthenia: cerebrasthenia (cerebral exhaustion) characterized by symptoms that were directly or indirectly connected with the head; myelasthenia (spinal exhaustion) was defined by symptoms related to the involvement of the spinal cord; digestive asthenia was characterized by dyspepsia, constipation, and flatulence. As time went on more subtypes were added by other investigators and specific treatment approaches were developed.

Despite the over inclusiveness of the term, Beard maintained that neurasthenia belonged to one family with a common pathology, prognosis, history, and treatment. As more cases were reported, he felt able to claim that neurasthenia was predominantly an American illness.[2] He attributed the increase in prevalence to the pressures of modern civilization.

Notwithstanding its vagueness, or perhaps because of its vagueness, neurasthenia gained popular acceptance not only by the medical profession but also by the general public. Although by the turn of the century it had become practically a household word, its popularity did not preclude dissent. Most of the criticisms focused on the disorder's over inclusiveness and lack of precision; for instance, Brill called it 'the newest garbage can' in medicine.

The first two decades of the twentieth century witnessed an increasing number of discoveries of more specific causes of disease. This period also saw greater attention being paid to the taxonomy of neuroses. These forces combined to bring about the decline of neurasthenia as a diagnostic entity.

In 1895, Freud published two seminal papers in which he drew up the blueprint for reconfiguring the various neurotic disturbances that were grouped together under the term neurasthenia. In the paper entitled 'On the grounds for detaching a particular syndrome from neurasthenia under the description of 'anxiety neurosis'[3] he questioned the validity of continuing to allow neurasthenia to cover all the symptoms described by Beard. He saw the need to classify different categories of neuroses based on the following:

- collection of symptoms that were more closely related to one another
- common aetiology
- common psychical mechanism.

In the paper 'Obsessions and phobias: their psychical mechanism and their aetiology',[4] Freud removed obsessions and phobias from neurasthenia. As a result of these two papers, neurasthenia ceased to be an amorphous concept and was differentiated into the following categories:

- neurasthenia proper
- anxiety neuroses
- obsessions
- phobias
- pseudoneurasthenias due to cachexia, arteriosclerosis, early stages of the general paralysis of the insane, and psychoses.

Intermittent and periodic types of neurasthenia were to be included under melancholia.

The first list of symptoms Freud proposed for neurasthenia proper included headache, spinal irritation, dyspepsia with flatulence, and constipation. Later, he added sexual weakness and fatigue.

The possibility of including some neurasthenic symptoms under melancholia was mentioned but not expanded on by Freud. This task was taken up by Kraepelin.[5] He distinguished three major types of depression: manic–depressive disorder, involutional melancholia, and a milder form of neurasthenic depression. He asserted that all these types of depression were due to an underlying disordered brain function.

Having been so denuded, the use of the term neurasthenia as a diagnostic entity by the medical professions had declined in the United States by the time of the First World War. The first edition of the DSM-I published in 1952 gave no formal recognition to neurasthenia. Instead, it was replaced by the category of 'Psychophysiological nervous system reaction', the predominant symptom of which was general fatigue. In an effort to make DSM-II congruent with ICD-8, neurasthenia reappeared in American psychiatry as neurasthenic neurosis.

In DSM-III neurasthenia disappeared as an entity and appeared only in the index where readers were asked to refer to 'Dysthymic disorder'. However, unlike the DSM classification, neurasthenia consistently remained a subtype of neurosis throughout the many versions of the ICD. ICD-9 defined neurasthenia as follows.

> A neurotic disorder characterized by fatigue, irritability, headache, depression, insomnia, difficulty in concentration, and lack of capacity for enjoyment (anhedonia). It may follow or accompany an infection or exhaustion or arise from continued emotional stress.

The following categories were included:

- fatigue neurosis
- nervous disability
- psychogenic asthenia
- general fatigue.

Spread to other countries

One of the most fascinating aspects of the history of neurasthenia is its ready acceptance by countries other than the United States where it was originally conceived as a peculiarly American phenomenon. The diagnostic entity took firmer root in some countries than in others. In many countries the concept was indigenized and took on local cultural colour.

The reasons for its spread can be summarized as follows:

- The all-embracing nature of the entity provided a foothold for almost everyone involved.
- The concept provided a blend of scientific theory, thus lending legitimacy to a cluster of symptoms, which are mostly subjective.
- It is considered to be a disease resulting from overwork, which affects the upper social class.

Asia and Australia

In all probability neurasthenia was introduced into China in the 1920s by American psychiatrists and returning Chinese doctors who were trained in the United States. Up to the end of the Second World War, Chinese physicians accepted and used the diagnostic concept of neurosis and neurasthenia from the United States. With the firm establishment of communism in 1949, Pavlovian theory was adopted as the sole model on which Chinese psychiatrists practice, teach, and research.[6] In China, as in the former USSR, neuroses were divided into neurasthenia, psychasthenia, and hysteria. The cause of neurasthenia, as indeed of neuroses, followed the Pavlovian theory of overstrain in the excitation and inhibition processes and mobility of the higher nervous system.

The concept of neurasthenia or *shenjing shuairuo* (nerve weakness), as translated by the Chinese, was not an entirely alien idea. The symptoms associated with neurasthenia (fatigue, loss of memory, poor attention span, headache, tension, insomnia, and all varieties of vague pains) are similar to those in patients suffering from a deficiency in *qi* (vital essence), that is weakness of the kidney, spleen, or heart in traditional Chinese medicine. In addition, the theory of nerve weakness and depletion of nervous energy as causes of neurasthenia fits in with the traditional Chinese medicine concept of organ weakness and *yin–yang* deficiency. Thus in no time at all neurasthenia was incorporated into the body of the practice of traditional Chinese medicine and the vocabulary of the lay public.

In the 1950s, the number of patients suffering from neurasthenia increased enormously. Medical or neurology clinics reported that 80 to 90 per cent of their outpatients were suffering from neurasthenia. It was particularly rampant among the 'brain or mind workers'. The Chinese government regarded it as a serious public health problem, so much so that in its First Five Year Plan (1958–1962) a large-scale national campaign was initiated to eradicate neurasthenia. Research on neurasthenia carried out during this period focused on the role of stress as the external factor, and on heredity and personality as endogenous factors. Treatment included intensive group re-education, herbal medicine, and tranquillizers. Lin[7] postulated that the marked increase in neurasthenia was due to the presence of a deep-seated tension in the revolutionary development of China during the 1950s. Neurasthenia became the vehicle to express political, social, and physical stresses.

About a decade after China's 'open-door policy', an epidemiological survey was conducted in 12 districts in China. The instrument used was the Present State Examination. The results showed that neurasthenia affected 12.59 per cent of persons aged from 15 to 59 years, accounting for 56.7 per cent of all neurotic disorders.[8] In 1982, Kleinman[9] conducted a study of 100 patients diagnosed with neurasthenia in the Psychiatric Outpatient Clinic of the Hunan Medical College. He found that 89 patients satisfied the DSM-III diagnostic criteria for 'Major depressive disorder', 70 per cent of whom responded substantially to antidepressant medication. Despite their improvement, few experienced decreased help-seeking behaviour. This led him to conclude that neurasthenia should be regarded as a special form of somatization related to culturally sanctioned idioms of distress.

In Taiwan, neurasthenia attracted little interest among western-trained doctors. However, it became enormously popular among traditional Chinese doctors, and consequently neurasthenia established itself as a major disease in the minds of the Taiwan public during the 1940s and 1950s.[10]

The mostly British-trained doctors in Hong Kong largely ignored neurasthenia as a diagnostic entity. As in Taiwan, neurasthenia became the domain of traditional Chinese doctors.[11]

In the late nineteenth century psychiatry in Japan was essentially German in orientation. Psychiatrists applied the diagnosis of neurasthenia to patients who presented with weakness, headaches, mental distraction, fatigue, and reduced psychic productivity.[12] The diagnostic entity became a popular term until Morita[13] supplanted it with the term *shinkeishitsu* (nervous or nervous disposition). He described this disorder as basically a psychological reaction to anxiety in predisposed personalities—the personality type being characterized by introversion, perfectionism, hypochondria, hypersensitivity, and self-consciousness. He developed a specific treatment aimed at breaking up the vicious cycle of sensitivity and anxiety, the initial phase of which consisted of isolated bed rest followed by a second phase of work therapy.

Doctors in Malaysia, Singapore, India, Pakistan, Burma, and Sri Lanka are mostly trained in the British tradition. After the First World War neurasthenia lost its popularity in Britain. Standard British textbooks regarded the disorder as rare and outmoded. As a result psychiatrists in these countries tended not to diagnose neurasthenia. However, neurasthenia is used in the Chinese communities where traditional Chinese medicine maintains a stronghold. In India and countries where Ayurvedic medicine is practised, neurasthenia was not added on to the more traditional ways of explaining fatigue, pain, dizziness, and headaches. Instead, concepts such as *dhàtu* loss (loss of semen) and *vàta roga* (wind disease) remained the preferred explanation.

In Australia, Paterson[14] reported that over a 15-year period from 1950 to 1965 neurasthenia was one of the 10 major illness categories reported by a large Sydney-based industry. He claimed that, since he had a fairly representative sample, the 10 categories of illness could very well apply to the rest of Australia.

Europe

From 1880 to 1920 neurasthenia was one of the diseases most frequently discussed. From an 'American nervousness' it rapidly evolved into a western European bourgeois illness. Practically every academic neurologist and psychiatrist wrote a major piece on neurasthenia (see Drinka[15]).

In England, neurasthenia was described in Osler's *The Principles and Practice of Medicine* published in 1900.[16] During the First World War it was a common diagnosis used for invaliding out many soldiers. In order to cope with its diagnosis, treatment, and disposal, the Army instituted a short course of training for medical officers who graduated with the title of 'neurasthenic expert' (see Sims[17]).

Russian psychiatry is largely based on Pavlovian psychophysiological theories. The Pavlovian classification of the principle of neuroses was adopted by all countries that came under the influence of the former USSR. Opinions on the subdivisions of neurasthenia were divided in Russia. One school of thought based its classification on the course of the illness, and the other was based on aetiology.[18] Neurasthenia as a cause of inefficiency and low productivity in the workplace was a recurrent theme in both Russia and Eastern Europe.

Current usage

In ICD-10,[19] neurasthenia is classified as a neurotic disorder in which two main, but overlapping, types of neurasthenia are described:

- the predominant symptom is increased fatigue after mental effort
- predominant feelings of bodily or physical weakness and exhaustion after only minimal efforts.

For a definite diagnosis ICD-10 requires the following:

(a) either persistent and distressing complaints of increased fatigue after mental effort, or persistent and distressing complaints of bodily weakness and exhaustion after minimal effort;

(b) at least two of the following:

- feelings of muscular aches and pains
- dizziness
- tension headaches
- sleep disturbances
- inability to relax
- irritability
- dyspepsia;

(c) any autonomic or depressive symptoms present are not sufficiently persistent and severe to fulfil the criteria for any of the more specific disorders in this classification.

The following are excluded:

- asthenia not otherwise specified
- burn-out
- malaise and fatigue
- postviral fatigue syndrome
- psychasthenia.

DSM-IV does not include neurasthenia as a nosological entity. Instead, it is replaced by 'Undifferentiated somatoform disorder'.

In the third edition of the *Chinese Classification of Mental Disorders*[20]neurasthenia is classified under 'Neurotic disorder'. The criteria for diagnosis have been made more stringent, requiring three symptoms out of five non-hierarchical groups of symptoms, which include weakness, emotionality, excitement, nervous pain, and sleep disturbance. The duration of the symptoms should be at least 3 months. Other psychiatric disorders have to be excluded. Because of the different connotations of fatigue and weakness in Chinese culture, fatigue is not included in the list of symptoms.[21]

Differential diagnosis

Fatigue is a ubiquitous symptom. It can occur in many psychiatric illnesses and in a wide range of physical illnesses. In cultures where the term neurasthenia is loosely used, many of the cases would probably meet the ICD-10 or DSM-IV diagnostic criteria for depressive disorder or anxiety disorder. Physical illness is a common cause of fatigue. In this respect, a detailed history and judicious investigation will be necessary.

Epidemiology

Merikangas and Angst[22] studied a cohort of young adults from a community sample in Zurich, Switzerland, and reported the prevalence of neurasthenia, defined according to the ICD-10 criteria, as 1 per cent across 10 years. The sex ratio across the 10 years of follow-up revealed an equal prevalence among males and females during the initial stages of the study, but females exhibited an l.6-fold greater rate than males during the later stages.

The World Health Organization (WHO) international study[23] of patients with psychological problems seen in primary-care settings reported a prevalence of 1.7 per cent of pure neurasthenia. The prevalence rate increased to 5.4 per cent when the syndrome was diagnosed comorbid with depression or anxiety. The prevalence rate in each centre is shown in Table 5.2.10.1.

The differences in the prevalence rate can be due to many factors, including perception of what health services can treat and the existence of alternative sources of care.

The results of an epidemiological study conducted in 1998 in seven areas in China showed a prevalence rate of 2 per cent.[24]

In a national survey by the Australian Bureau of Statistics Hickie *et al.*[25] reported that 1.5 per cent of the general population met the ICD-10 criteria for neurasthenia in the past year.

All the studies were consistent in demonstrating that the syndrome tended to affect patients below the age of 45 and the absence of significant gender differences.

Aetiology

Although theories abound, the predisposing, precipitating, and perpetuating causes of the syndrome remain unclear.

Table 5.2.10.1 Prevalence of neurasthenic syndrome among patients contacting general health care facilities

Centre	Overall prevalence (%)	Males (%)	Females (%)
Ankara	4.1	1.0	5.6
Athens	4.6	3.3	5.2
Bangalore	2.7	1.7	3.7
Berlin	7.4	4.0	9.7
Mainz	7.7	7.4	8.0
Groningen	10.5	7.1	12.8
Ibadan	1.1	3.4	0.2
Manchester	9.7	6.1	11.3
Nagasaki	3.4	3.8	3.0
Paris	9.3	5.0	14.2
Rio de Janiero	4.5	2.3	5.3
Santiago	10.5	6.4	12.1
Seattle	2.1	2.3	2.0
Shanghai	2.0	1.5	2.2
Verona	2.1	1.8	2.3

Course and prognosis

The 10-year follow-up study conducted by Merikangas and Angst[22] revealed that approximately 50 per cent of patients continued to exhibit symptoms. The WHO study[23] reported that patients with a diagnosis of neurasthenia had, on average, been disabled for 8 to 7 days during the month preceding the examination. Hickie *et al.*[25] confirmed the chronicity of the condition. They reported that 80 per cent of people who met the ICD-10 criteria in the past 12 months were also current cases.

Comorbidity

In the Australian study Hickie *et al.*[25] showed that there was more comorbidity with major depression, panic disorder, and generalized anxiety disorder than could be expected by chance after adjustment for the prevalence of the comorbid disorder and the average level of comorbidity of that disorder.

Treatment

Although there are reports that antidepressants can be effective these are not supported by published data on randomized double-blind controlled trials. Indeed, no such trials have been carried out on any form of psychiatric treatment for neurasthenia. In the absence of such data the clinician will have to rely on the adage of 'When not able to do any good, avoid doing any harm'. Thus aggressive treatment and investigations should be avoided. Patients are best managed in a supportive relationship with due regard given to their psychological and psychosocial needs.

General and non-specific strategies may also be used. These can include regular graded increase in exercise, promotion of sleep hygiene, cognitive techniques to break the cycle of symptoms leading to decreased activities, and improving social support.

Clinicians working in an environment where people with mental illnesses are stigmatized might find it easier to accede to social demands. Clinicians work at two levels in these situations. At one level he or she will have made a diagnosis of a psychiatric disorder and have prescribed the appropriate treatment. At another level the practitioner will be using neurasthenia as a euphemism for mental illness. Therefore, until stigmatization can be reduced or abolished, this unenviable state of affairs will continue.

Complementary medicine

The extracts of leaves from the *Ginkgo biloba* tree contain ginkgoflavone glycosides and trepenoids. *Ginkgo* is widely used as a cognitive enhancer. The roots of the *Panax ginseng* contain several triterpine glycosides, which are believed to have physical performance enhancing properties. This root is used in traditional Chinese medicine to treat a large variety of diseases including neurasthenia. Wesnes *et al.*[26] evaluated the effects of a *Ginkgo biloba/ginseng* combination on 64 healthy volunteers who fulfilled the ICD-10 criteria for neurasthenia. This was a 90-day, double-blind, placebo-controlled, parallel group study. They reported that the combined dose of 120 mg *Ginkgo* extract and 300 mg of *ginseng* extract was significantly better than placebo in reducing the symptoms of neurasthenia by day 90. Adverse effects were nausea and abdominal pain.

The mushroom, *Ganoderma lucidum*, known as lingzhi in China has been widely used to treat cancer, diabetes, and neurasthenia. It is the only known source of ganoderic acid, which has a molecular structure similar to steroid hormones and is a source of biologically active polysaccharide. *Ganoderma* is one of the most highly ranked herbal medicines by Asian people. Tang *et al.*[27] conducted a randomized, double-blind, placebo-controlled parallel study to investigate its efficacy and safety in the treatment of Chinese patients who fulfilled the ICD-10 diagnostic criteria for neurasthenia. Their findings indicated that *Ganoderma* was significantly superior to placebo with respect to the clinical improvement of symptoms in neurasthenic. Adverse effects were mild consisting of nausea, dry mouth, and vomiting.

Complementary medicine is increasingly being used either as an alternative or in addition to conventional psychotropic medications. Refer to Chapter 6.2.9 for further information.

Future directions

As to whether neurasthenia will be replaced by chronic fatigue syndrome (Chapter 5.2.7) or subsumed under somatoform disorders (Chapter 5.2.1), and thus be relegated to a footnote in the history of medicine, or will enjoy resurgence with a new set diagnostic criteria will be determined by future research and clinical data.

Further information

Starcevic, V. (1999). Neurasthenia: cross-cultural and conceptual issues in relation to chronic fatigue syndrome. *General Hospital Psychiatry*, **21**, 249–55.

Tang, W., Gao, Y., Chen, G., *et al.* (2005). A randomized, double-blind and placebo-controlled study of a *Ganoderma lucidum* polysaccharide extract in neurasthenia. *Journal of Medicinal Food*, **8**, 53–8.

Web pages: www.kosmix.com/health Type in neurasthenia to search

References

1. Beard, G.M. (1869). Neurasthenia or nervous exhaustion. *Boston Medical and Surgical Journal*, **3**, 217–20.
2. Beard, G. (1881). *American nervousness: its causes and consequences.* Putnam, New York.
3. Freud, S. (1895). On the grounds for detaching a particular syndrome from neurasthenia under the description of 'anxiety neurosis'. In *Standard edition of the complete psychological works of Sigmund Freud*, Vol. 1 (ed. J. Strachey), p. 139. Hogarth Press, London.
4. Freud, S. (1895). Obsessions and phobias: their psychical mechanism and their etiology. In *Standard edition of the complete psychological works of Sigmund Freud*, Vol. 3 (ed. J. Strachey), pp. 74–82. Hogarth Press, London.
5. Kraepelin, E. (1902). *Clinical psychiatry* (trans. A.R. Defendorf), pp. 96–104. Macmillan, New York.
6. Lin, T.S. (1989). Neurasthenia revisited: its place in modern psychiatry. *Culture, Medicine and Psychiatry*, **13**, 105–29.
7. Lin, T.S. (1985). The shaping of Chinese psychiatry in the context of politics and public health. In *Mental health planning for one billion people* (eds. T.S. Lin and L. Eisenberg), pp. 13–14. University of British Columbia Press, Vancouver.
8. Li, C.P. and Zhang, W.H. (1985). Neurasthenia. In *Psychiatry* (ed. Y.C. Shen), pp. 407–12. People's Health Press, Beijing (in Chinese).
9. Kleinman, E.A. (1982). Neurasthenia and depression: a study of somatization and culture in China. *Culture, Medicine and Psychiatry*, **6**, 117–90.

10. Rin, H. and Huang, M.G. (1989). Neurasthenia as a nosological dilemma. *Culture, Medicine and Psychiatry*, **13**, 215–26.
11. Cheung, F. (1989). The indigenization of neurasthenia in Hong Kong. *Culture, Medicine and Psychiatry*, **13**, 227–41.
12. Suzuki, T. (1989). The concept of neurasthenia and its treatment in Japan. *Culture, Medicine and Psychiatry*, **13**, 203–13.
13. Morita, M., Kondo, A. and LeVine, P. (1998). Morita therapy and the true nature of anxiety-based disorders (Shinkeishitsu) translated by Kondo, A. pp. 55–62. SUNY Press, New York..
14. Paterson, G.O. (1969). The economics of illness: an employee sickness study. *The Medical Journal of Australia*, **2**, 249–52.
15. Drinka, G.F. (1984). *The birth of neurosis: myth, malady and the Victorians*. Simon and Schuster, New York.
16. Osler, W. (1900). *The principles and practice of medicine* (7th edn), pp. 166–7. Appleton, New York.
17. Sims, M. (1968). *Guide to psychiatry*, p. 444. Livingstone, Edinburgh.
18. Chatel, J.C. and Peele, R. (1970). A centennial review of neurasthenia. *The American Journal of Psychiatry*, **126**, 1404–13.
19. World Health Organization. (1992). *The ICD–10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. WHO, Geneva.
20. Chinese Psychiatric Society. (2001). *The Chinese classification of mental disorders* (*CCMD-3*). Shandong Publishing House of Science and Technology, Shandong (in English and Chinese).
21. Lee, S. (1996). Cultures in psychiatric nosology: the CCMD–2–R and the International classification of mental disorders. *Culture, Medicine and Psychiatry*, **20**, 421–73.
22. Merikangas, K. and Angst, J. (1994). Neurasthenia in a longitudinal cohort study of young adults. *Psychological Medicine*, **24**, 1013–24.
23. Üstün, T.B. and Sartorius, N. (1995). *Mental illness in general health care: an international study*. Wiley, Chichester.
24. Zhang, W.X., Shen, Y.C., and Li, S.R. (1998). Epidemiological investigations on mental disorders in 7 areas of China. *Chinese Journal of Psychiatry*, **31**, 69–71.
25. Hickie, I., Davenport, I., Issakidis, C., *et al.* (2002). Neurasthenia: prevalence, disability and health care characteristics in the Australian community. *The British Journal of Psychiatry*, **181**, 56–61.
26. Wesnes, K.A., Faleni, R.A., Hefting, N.R., *et al.* (1997). The subjective, cognitive and physical effects of a *Ginkgo biloba*/*Panax ginseng* combination in healthy volunteers with neurasthenic complaints. *Psychopharmacology Bulletin*, **33**, 677–83.
27. Tang, W., Gao, Y., Chen, G., *et al.* (2005). A randomized, double-blind and placebo-controlled study of a *Ganoderma lucidum* polysaccharide extract in neurasthenia. *Journal of Medicinal Food*, **8**, 53–8.

5.3

Medical and surgical conditions and treatments associated with psychiatric disorders

5.3.1 Adjustment to illness and handicap

Allan House

Introduction

Not everybody who develops a serious physical illness will have psychiatric problems as a consequence. To understand why, it is useful to have a model of the normal process of adjustment to stress; psychiatric disorder can then be seen as arising when that process, often called coping, is either maladaptive or is adaptive but only partially successful. This chapter will start with an outline of one theory of stress and coping as it applies to physical illness, followed by a review of disorders of adjustment to illness. A distinction will be drawn between recent-onset illness, which provokes an acute response, and long-standing illness, where the challenge is more often to adjust to chronic disability.

Adjustment to illness and handicap

A number of diseases are reviewed in later chapters, and therefore this chapter will deal with general principles. For more details on particular diseases, the reader should consult specialist textbooks of psychiatry or health psychology.

Illness as a stress

Stress is a word that is used in different ways. Sometimes it refers to an environmental stimulus—a threat or demand from the outside world. This definition lies behind various measures, such as the Social Readjustment Rating Scale[1] or the Bedford College Life Events and Difficulties Schedule[2] which characterize life experiences and produce standardized measures of their severity. According to this view, experiences have properties—as losses, or challenges, or dilemmas—that can be identified by knowing something of the social circumstances of the subject of those experiences but without knowing about the meaning given to them by the person experiencing them.

Another meaning of stress is that it is a bodily state, so that events are only regarded as stressful if they produce changes in the individual. The best-known example of this usage comes from physiology.[3] Stress as a psychological state is also a common lay meaning; when people describe themselves as 'stressed' they are usually referring to a state of tension or autonomic arousal.

Yet another way to understand stress, which is useful in considering physical illness, is that it arises out of an interaction between environmental demands and the resources available to deal with them. This view is articulated in the transactional model of Lazarus and Folkman.[4] According to the theory, when faced with a new experience individuals assesses its likely impact (the primary appraisal) and assess their resources (the secondary appraisal). Stress arises when this double appraisal identifies a mismatch between demands and resources that cannot be narrowed by coping manoeuvres.

(a) Illness as a demand or threat (the primary appraisal)

There are a number of characteristics of an experience that increase the chances of it being appraised as threatening. These include immediacy, ambiguity, uncontrollability, or undesirability. The probability that many people will share an interpretation of a particular episode explains the similarity of people's responses to certain illnesses. The possibility of individual, even idiosyncratic, interpretations can explain sharp differences between people with apparently the same disorder.

A useful way to construe individual appraisals of illness is outlined by Leventhal *et al.*[5] in their theory of internal illness representations. The common elements of the illness representation can be identified from a simple self-report questionnaire[6].

Illness beliefs cannot be assumed solely on the basis of the illness from which a person is suffering, or from his or her social context.[7] Individuals may hold unpredictable beliefs—that an illness is inherited from a family member, or that it is a punishment for a misdemeanour, or that it may be curable by adopting an unusual diet. For some, the representation of illness overlaps with the representation of self, so that sufferers see themselves as living their illness rather than suffering from it.[8] (see Box 5.3.1.1)

The characteristics of a particular disease are not the only component of the illness that can make it threatening. Illness occurs in a social and interpersonal context, and while the responses of other people may be helpful, they may in some cases contribute to the demands of the situation. For example, a partner may withdraw or become depressed, or family members may become intrusive or overcontrolling. Being ill confers a special status, the so-called sick role, but it is a status acquired at a cost in the loss of independence and certain rights. While disability may arise largely from the impairments caused by a disease, much handicap is socially determined.

(b) Resources for responding to illness (the secondary appraisal)

The focus of secondary appraisal is twofold: the person's personal resources, and the resources external to them, mainly in the immediate social network.

Personal resources may be defined in a number of ways, for example cognitive attributes, personal characteristics, or personality traits.

The other resource for the individual is social support. There are many approaches to understanding support, but a useful one[9] is to regard it as having four components:

1 emotional support, conveying a sense of being cared about or loved

2 esteem support, conveying a sense of being valued or respected

3 instrumental support, conveying practical help

4 informational support, conveying knowledge relevant to tackling the problem

Box 5.3.1.1 Components of the illness representation

- identity (label and associated symptoms)
- causal ideas
- consequences (severity and likely impact)
- time-line (natural history)
- curability or controllability

The family's reaction to illness has an important impact on the type of support available. If they are rejecting, intolerant of dependence, or unsympathetic to the needs of the patient—for example, to change their diet, or stop smoking, or take more (or less) exercise—then they may offer too little support. On the other hand, they may be overprotective, refusing to allow the patient a reasonable degree of autonomy and discouraging active coping. Sometimes, members of a family will hold different views about the nature of an illness, leading to conflict, which is not always revealed to doctors. More often, they share views. If such views are inaccurate (so-called family myths) and yet strongly held, then they can be a powerful barrier to the patient accepting medical advice. It is a common observation that patients with chronic illness who are depressed often have a carer who is depressed, and this tendency to share (often dysfunctional) beliefs and coping styles is one reason for that.

(c) Coping with illness

Coping refers to efforts to reduce the gap between demands and resources. Coping is described according to its aims, the techniques used to achieve those aims, and according to the overall coping style adopted.

The ***aims*** of coping are either problem focused, designed to modify the demands of the situation, or emotion focused, designed to modify how one feels about a situation.[4] Emotion focused coping generally works well but only transiently. It is best reserved for brief stresses, such as unpleasant medical procedures, or for situations in which nothing can realistically be done to modify the stress.

The ***techniques*** for coping serve to mobilize available resources. Vocabularies differ for describing them. Cognitive coping techniques include information seeking, downplaying, or adopting a defiant or overoptimistic attitude. In psychodynamic terms, the two most commonly used techniques are probably denial and regression. In common usage, the techniques referred to by these vocabularies overlap. Behavioural coping may involve changing ones lifestyle, such as exercising more or excessive drinking of alcohol. Social coping is a particular form of behavioural coping, and may involve increasing contacts or accepting help from professional agencies. In chronic illness, successful coping may be accompanied by a slower process of reappraisal—in which the patient comes to a different understanding of the illness, from that apparent at initial diagnosis—through for example *benefit-finding* and *downward comparison* (with others who have worse disability, pain, or whatever).

Coping ***styles*** are more general approaches to coping. Two contrasted styles are active/engaged (sometimes called 'approach') coping and passive/disengaged (sometimes called 'avoid') coping.[10] While it is appealing to characterize people as having a particular coping style, and while it is possible to think of typical examples from personal experience, in fact most people do not have a sufficiently unchangeable repertoire of coping techniques to merit the label of a style.

Adjustment disorders

(a) Definition and classification

The emphasis in ICD-10[11] is on emotional disturbance as the characteristic feature of adjustment disorders—some disturbance of behaviour is acknowledged, particularly in adolescence.

However, it is common to encounter cognitive or behavioural changes that interfere with social functioning and quality of life, and yet which are not attributable to the consequences of mood disorder. DSM-IV[12] acknowledges this possibility more directly, including a category of 'Adjustment disorder, unspecified', which covers 'maladaptive reactions (e.g. physical complaints, social withdrawal, or work or academic inhibition)'.

Examples of cognitive problems are extreme helplessness, denial of the existence of illness, or of the handicap associated with it. Behavioural problems may include marked social withdrawal or lack of self-care, or irrational non-adherence to treatment. Emotional problems are typically thought of as anxiety or depression, but irritability is also common.

(b) Diagnosis and differential diagnosis

The diagnostic features of adjustment disorders are relatively non-specific, comprising mood symptoms and behaviour disturbances, which do not meet the criteria for a diagnosis of another disorder, and yet which are sufficient to amount to a mental disorder. The two main diagnostic questions are as follows.

- Does the patient have a diagnosable mental disorder?
- If there is a mental disorder, should it be given another more specific label than 'adjustment disorder'?

What distinguishes normal adjustment from a disorder? The first criterion is whether the symptoms are persisting beyond the time when they might be attributable to the stressor. This judgement is relatively straightforward when the stressor is a single event. However, if illness is more persistent or intermittent—such as cancer followed by intensive treatment, or multiple sclerosis—then it is less easy to judge.

The second criterion is whether the response is causing avoidable social dysfunction. For example, in many cultures illness is followed by a period of convalescence, during which activity is reduced and a return to full social responsibilities is deferred. This may be a healthy avoidance of activity, if it allows full recovery from illness, but prolonged avoidance of activity may lead to secondary physical problems as well as social isolation and loss of role.

When adjustment disorders are associated with chronic illness and handicap, the duration criterion cannot apply. An individual may present symptoms because his or her response is outside the culturally acceptable range; for example, he or she may be too demanding or uncooperative, or too passive and dependent. It is unwise to regard a presentation as disordered simply on these grounds. The best indicator is whether the individual is achieving the highest level of function and the lowest level of distress of which they are capable under the circumstances. This means that each person must be diagnosed according to his or her own context, and that a standardized set of criteria cannot be applied.

The differentiation of adjustment disorders from other psychiatric disorders is more straightforward, and depends on the presence or absence of key symptoms. The main conditions found in association with physical illness are depressive disorders, anxiety disorders, and occasionally post-traumatic stress disorder.

(c) Epidemiology

Little is known about the epidemiology of adjustment disorders other than those involving mood disturbance, because of the absence of standardized diagnostic criteria.

Psychiatric symptoms are distributed in the general population, with a positive skew to the distribution. In the physically ill, the same pattern of distribution is seen, but the curve is shifted to the right. The increase in psychiatric symptoms is contributed to by a general increase in all the common symptoms. The usual way to identify cases is to select those who cross an accepted threshold for symptom levels—as determined, for example, by one of the standardized self-report questionnaires—and then to apply diagnostic criteria. Adopting this approach, rates of diagnosable mood disorder among the physically ill are about double what they are in the general population. That is, 30 to 50 per cent (depending on the population studied and the diagnostic criteria employed) of the physically ill have a mental disorder. Approximately two-thirds of these cases are adjustment disorders, the rest meeting criteria for another disorder (usually depressive).

The elderly report lower rates of psychiatric disturbance. This may be a cultural effect, with the elderly disposed to report fewer symptoms of distress as a result of stoicism learned through experience of adversity earlier in life. Alternatively, the elderly may genuinely respond differently to physical illness.

Mood symptoms and adjustment disorders are commoner in response to acute illness than they are in chronic illness.

(d) Aetiology

There are several reasons why coping might fail.

First, demands may be overwhelming. The news that one has a terminal illness takes time to assimilate—to understand all its meanings, grasp all the threats and losses involved. While that process of appraisal is going on, it is difficult to marshal resources and use them effectively. This explains, in part, why mood disorder is more commonly associated with acute than chronic illness.

Second, resources may be inadequate or missing. One problem associated with physical illness is that it may impair personal resources as a primary effect of the disease process—most importantly when the illness has effects on the central nervous system by virtue of the direct involvement of the brain or through the neurological effects of systemic disturbance.

Third, coping responses may be ineffective. There are few rules about what makes effective coping. In general, a broad and flexible repertoire is desirable, with a strong element of active problem-focused techniques. However not all illnesses, nor all aspects of a particular illness, are likely to be amenable to problem-focused coping. Probably the most effective coping is matched to the situation. That is, the coping matches the demands, so that heavy reliance is not placed on problem-focused coping when little in the situation can change, nor excessive use made of emotion-focused coping when active involvement in illness management is needed.

A common problem of failure to match coping to the situation is found in patients with chronic illness, who are responding to their circumstances as if they none the less have an acute illness. In acute illness, problem-focused coping often involves seeking reversal or even cure of the illness process, while emotion-focused coping involves dealing with the anxiety of uncertainty, or grieving if the prognosis is clearly poor. On the other hand, in chronic illness, problem-focused coping involves symptom management and maximizing function, while emotion-focused coping requires a degree of acceptance.

It is not easy to predict who will develop an adjustment disorder. Certainly the risk is not strongly linked to physical diagnosis, or

within a particular diagnostic group to physical disability. The most robust finding is that a previous history of psychiatric problems increases the risk of psychiatric problems associated with physical illness.

(e) Course and prognosis

By definition, adjustment disorders arise shortly after diagnosis. In practice, there is variation; some people respond immediately and develop symptoms within days, while others develop symptoms weeks or even months after diagnosis. The losses associated with illness may only become apparent when a person leaves hospital and faces functional impairment at home. Carers and others in the social network respond differently to acute and chronic illness, and it may take time for that to become clear. The greater the delay from the onset of illness to the emergence of symptoms, the harder it is to make a diagnosis of adjustment disorder. In clinical practice, it is reasonable to set an upper limit of a year.

Most adjustment disorders provoked by a newly onset illness, resolve within weeks. Slower recovery takes place over 12 or 18 months. If recovery has not occurred by then, the patient has usually developed another mental disorder, such as a depressive disorder. Accurate data are few, but probably no more than 10 per cent of patients develop a prolonged adjustment disorder.

The psychiatric symptoms of adjustment disorder impair quality of life, so much so that all standardized quality-of-life measures include mood symptoms in their profile. Psychiatric morbidity associated with physical illness is also a risk factor for self-harm and for completed suicide. Adjustment disorders are likely to have an effect on the outcomes of treatment for physical disease.[13] Health service costs are greater for patients with physical illness and psychiatric co-morbidity; lengths of stay are longer for hospital inpatients; the functional outcomes of rehabilitation may be poorer, and there is some evidence that there may also be an increased mortality. The mechanism for these effects may be broadly behavioural or physiological. Examples of the former are increased rates of smoking, lack of exercise, and poor adherence to treatment regimes among people with mental disorders. Examples of the latter include activation of the hypothalamic–pituitary–adrenal axis or increased cytokine production associated with chronic emotional disorder.

(f) Treatment

(i) Drug treatments

Antidepressants can be effective in the presence of physical illness.[14] There is no good evidence to support claims for a great superiority in efficacy of serotonin-reuptake inhibitors. Although their long-term tolerability may be greater than older drugs in patients with physical disease, they are not without toxicity—for example they have been associated with increased falls and gastrointestinal haemorrhage. Tricyclic antidepressants have advantages in treating patients with insomnia or chronic pain. Cost differences are substantial.

(ii) Psychological treatments

A number of brief psychological therapies have been shown to be effective in treating depression; namely cognitive-behaviour therapy, problem-solving therapy, interpersonal therapy, and brief dynamic therapy. Such therapies may also be effective in treating adjustment disorder in the physically ill,[15,16] although they may have disappointingly weak effects upon physical outcomes. Therapy may need to be modified to allow for fatigue or concentration problems, and sessions need to be arranged flexibly to accommodate hospital appointments and other treatment needs.

(g) Management

(i) Identifying cases

A major difficulty in delivering treatment to people with adjustment disorders is the difficulty in identifying cases. There are several self-report questionnaires, which may be used to screen for patients with mood disorder. In certain settings such questionnaires can be delivered routinely to all patients, for example by means of computers with touch screen technology, and they are useful for alerting staff to the presence of mood symptoms—but their positive predictive value is too low to allow for accurate use in identifying those who need referral to specialist services. Their use is also difficult to integrate into routine clinical practice, and response rates outside research studies are usually low.[17] Their use is best restricted to specialist services where the clinical staff are clear about what response they will make to a high score, since this is where there is some evidence for the benefits of case-finding. There are no useful standardized instruments for the detection of other problems with adjustment.

Instead, clinicians should be encouraged to consider the possibility of psychiatric disorder when there is a gap between impairment and handicap so that the patient is doing worse in rehabilitation than the severity of their disease would suggest they should be, when there are multiple complaints that are difficult to explain, or when multiple drug treatments are being administered without conspicuous benefit. The clinical interview is the mainstay of diagnosis.

There are a number of common reasons for failing to recognize adjustment disorders. First, the questions simply are not asked, or attempts by the patient to introduce the topic of psychological problems are blocked or sidestepped. Second, questions may be asked, but in circumstances where it is difficult for the patient to answer honestly—when there is no privacy, or the person asking is obviously too busy to listen to any but a conventional answer. Third, expressions of distress may be normalized, and thus dismissed: 'Of course it's natural you will feel like that' means to the patient 'So please don't mention it again'.

(ii) Broadening the repertoire of psychological responses

No single intervention is going to be effective for all patients with psychological problems arising from difficulties with adjustment to illness. Realistic management therefore involves offering what has been called a 'menu of interventions'.[18] A currently favoured model is a so-called ***stepped care*** model in which intervention is organized in a hierarchy according to intensity of treatment and the expertize needed to deliver it e.g. see Box 5.3.1.2. While unfocused 'support' is of limited value (because it does not encourage active secondary appraisal and experiments with different coping strategies) there are now many brief and flexible psychological therapies available, some of which may be deliverable by staff in the primary care or physical healthcare service (see Box 5.3.1.3). In ***collaborative care*** such first line psychological treatments, along with medication management and simple social care, are delivered by non-mental health staff and monitored by a case manager, with a mental health professional providing supervision and back-up consultation. Successful trials have been conducted in (for example) heat disease[19] and diabetes.[20]

Box 5.3.1.2 Stepped care in the treatment of depression (NICE clinical guideline 2004)

	Who is responsible for care?	What is the focus?	What do they do?
Step 5	Inpatient care, crisis teams	Risk to life, severe neglect	Medication, combined treatments, ECT
Step 4	Mental health specialists including crisis teams	Treatment-resistant, recurrent, atypical and psychotic depression, and those at significant risk	Medication, complex psychological interventions, combined treatments
Step 3	Primary care team, primary care mental health worker	Moderate or severe depression	Medication, psychological interventions, social support
Step 2	Primary care team, primary care mental health worker	Mild depression	Watchful waiting, guided self-help, computerized CBT, exercise brief psychological interventions
Step 1	GP, practice nurse	Recognition	Assessment

National Institute for Health and Clinical Excellence (NICE) (2005). CG22 Anxiety: quick reference guide. London: NICE. Available from http://www.nice.org.uk/nicemedia/pdf/CGO22quickrefguideamended.pdf. Reproduced with permission.

Box 5.3.1.3 Brief psychological treatment of use in the medically ill

- **Motivational interviewing** is an approach developed to encourage people to attempt change in addictive behaviours. It may be useful in engaging people in demanding treatments, or in improving adherence to treatment regimes.
- **Graded activity** has been used to treat negative symptoms in mental illness like schizophrenia or depressive disorder. It is effective in improving function in chronic fatigue syndrome, and is worth using in other conditions where inactivity and passivity is out of proportion to physical disability.
- **Anger management** is a modification of cognitive-behaviour therapy, which may be useful where irritability or aggressive behaviour is complicating adjustment.
- **Interpersonal therapy**[21] was initially developed for the treatment of depression, but it has obvious applications in the field of physical illness. In the terminology of interpersonal therapy, illness represents a role transition, and the focus in therapy is therefore on negotiating that transition with key others in the patient's life.
- **Family therapy and couples therapy** are rarely considered (or available) for adults with physical illness, and yet many of the external resources needed for coping are in the family.

For more severe or persistent problems, referral to specialist services is appropriate—ideally liaison services that operate in the primary or secondary care settings where patients receive their main care. Psychiatric treatment of the physically ill, especially in hospital, requires a number of modifications to routine clinical practice, which are sometimes overlooked.

First, an extra effort has to be made to meet the family and carers. They may be reluctant to attend if there is hostility in the family, or if missed time from work is creating financial pressures, but failure to interview others makes it near impossible to come to a full and accurate formulation of the problem.

Second, personal contact with the referrer is highly desirable. The 'real' question may not be that posed in the referral, and can only be identified by probing. Advice is much more likely to be followed if it is delivered face to face, and followed up with a later visit to check on compliance! This direct contact with non-psychiatric colleagues is one of the defining characteristics of liaison psychiatry, and its importance cannot be overemphasized.

Third, it must be recognized that the course of psychiatric treatment needs to be modified. Appointments will be missed, or interrupted, by the demands of physical treatment. And psychological issues may well not be resolved by a single clinical encounter; a relapse of illness may provoke a further episode with new features, and patients often have to return repeatedly to work through themes in therapy, as they are re-challenged with new physical problems.

(h) Prevention

There are two broad approaches to prevention, namely education and support.

Education and the provision of information and advice about the illness and its management is desirable as an informed patient is more likely to be an effective partner in treatment, and because it is popular with patients. Disappointingly, however, it is not an effective means of preventing psychiatric problems. This is probably because, while it facilitates primary appraisal, it does nothing to facilitate secondary appraisal or the use of effective coping strategies.

Provision of support is also popular. It takes a number of forms, including self-help groups, volunteer visiting, and professional support workers—usually with knowledge of a particular disease such as AIDS or a stroke. Again, there is little evidence that it prevents psychiatric problems. Perhaps this is because it usually provides emotional support; which in itself may be worthwhile but which is insufficient if not combined with a more problem-focused approach.

In conclusion, there are no clear indications that we can prevent the development of adjustment disorders. The mainstay of current management is therefore to identify existing cases and to offer specialist care to those who are most symptomatic or handicapped, and to those who are not improving spontaneously.

Adjustment to terminal illness: care of the dying

Adjustment to terminal illness has much in common with adjustment to other severe illness, and is not specifically the province of psychiatrists. For a detailed discussion of the care of the dying, the reader is referred to a more specialized text.[22] Here we will discuss two issues that are commonly presented to the psychiatrist in this setting: the diagnosis of depression and other adjustment disorders, and the issue of suicide.

Diagnosis of depression and adjustment disorders in the terminally ill

As with physical illness, somatic complaints are common in the dying and thus, individually, lose their diagnostic or predictive usefulness in the major depressive syndrome. Even psychologically, a degree of hopelessness may be appropriate. Anxiety is a common symptom in the dying, but it is not necessarily pathological. Like depression, it may result from physical disability, uncontrolled pain, or pre-existing anxiety disorders. In these circumstances a more detailed examination of the attitudes of the patient is necessary. Pervasive global hopelessness, feelings that life has had no meaning, strong feelings of guilt or punishment, and suicidal thoughts are pointers towards depressive illness in the terminally ill.

Assessing suicidal thoughts in the terminally ill

Several studies have shown that the prevalence of suicidal thoughts among patients with terminal cancer is less than 10 per cent.[23] However, this contradicts the clinical impression that most patients admit to either suicidal thoughts or thoughts of assisted suicide as an escape from the imaged consequences of losing control. In some patients, having a belief in a 'way out' can be positive in offering a sense of control.

Completed suicide is an important complication in patients with a terminal illness. General predictors of suicide apply, with the addition of severity of functional impairment, isolation, and delirium. The two most important factors to watch out for are uncontrolled pain and depression. These two factors greatly increase suicide risk but are nevertheless treatable in the terminal illness setting.[24]

People who express a desire to die are nearly always ambivalent. The expression of suicidal ideas should never be accepted as rational without a searching enquiry for evidence of subtle external pressures, fear of terminal symptoms or of being a burden on others, and treatable depression.

Pharmacological management of psychiatric disorders in the dying

Alleviation of distress rather than cure is the guiding principle of the management of the terminally ill. Caution is needed in the selection and prescription of psychotropic drugs.

Anxiolytic medication is usually well tolerated and the concern over dependence and tolerance is less of an issue. Terminal metabolite benzodiazepines, such as lorazepam and oxazapam, can provide symptomatic relief for a variety of conditions even in patients with hepatic impairment. Occasionally, opiates are used in this role where first-line treatments are unsuccessful.

Antidepressant use in the dying may be more problematic due to the adverse effects of sedation, seizures, hypotension, and constipation and urinary retention. For this reason, the choice of drug needs to be individually tailored. Dosage has to be carefully adjusted, beginning at low doses and increasing gradually. The use of psychostimulants (dextroamphetamine, methylphenidate) is worth considering. These drugs are used infrequently in general psychiatry because of the risks of dependence, but in the terminally ill they may have advantageous 'energizing' properties, including increased energy, improved concentration, increased appetite, and possibly, a faster onset of action.

Further information

Sharpe, L. and Curran, L. (2006). Understanding the process of adjustment to illness. *Social Science & Medicine*, **62**, 1153–66.

Fletcher, J., Bower, P., Gask, L., *et al.* (2006). *Primary care services for depression: a guide to best practice care services improvement partnership*. National Institute for mental Health in England, Department of Health, London.

References

1. Holmes, T.H. and Rahe, R.H. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research*, **11**, 213–18.
2. Brown, G. and Harris, T. (1978). *Social origins of depression*. Tavistock Publications, London.
3. Selye, H. (1956). *The stress of life*. McGraw–Hill, New York.
4. Lazarus, R.S. and Folkman, S. (1984). *Stress, appraisal and coping*. Springer, New York.
5. Leventhal, H., Nerenz, D., and Steele, D. (1984). Illness representations and coping with health threats. In *Handbook of psychology and health* (4th edn) (eds. A. Baum, S.E. Taylor, and J.E. Singer), pp. 219–52. Erlbaum, Hillsdale, NJ.
6. Weinman, J., Petrie, K., Moss–Morris, R., *et al.* (1996). The illness perception questionnaire: a new method for assessing the cognitive representations of illness. *Psychology and Health*, **11**, 431–45.
7. Petrie, K. and Weinman, J. (eds.) (1997). *Perceptions of health and illness*. Harwood, Amsterdam.
8. Kleinman, A. (1988). *The illness narratives: suffering, healing and the human condition*. Basic Books, New York.
9. Cohen, S. and Wills, T.A. (1985). Stress, social support and the buffering hypothesis. *Psychological Bulletin*, **98**, 310–57.
10. Carver, C.S., Weintraub, J.K., and Scheier, M.F. (1989). Assessing coping strategies: a theoretically based approach. *Journal of Personality and Social Psychology*, **56**, 267–83.
11. World Health Organization. (1992). *International statistical classification of diseases and related health problems,* 10th revision. WHO, Geneva.
12. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
13. Saravay, S.M. and Lavin, M. (1994). Psychiatric comorbidity and length of stay in the general hospital–a critical review of outcome studies. *Psychosomatics*, **35**, 233–52.
14. Gill, D. and Hatcher, S. (2000). Antidepressants for depression in medical illness. *Cochrane Library*, **4**.
15. MacHale, S. (2002). Managing depression in physical illness. *Advances in Psychiatric Treatment*, **8**, 297—305.
16. Winkley, K., Landau, S., Eisler, I., *et al.* (2006). Psychological interventions to improve glycaemic control in patients with Type I diabetes: systematic review and meta-analysis of randomized controlled trials. *British Medical Journal*, **333**, 65–8.
17. House, A. (1988). Mood disorders in the physically ill: problems of definition and measurement. *Journal of Psychosomatic Research*, **32**, 345–53.
18. Goodheart, C.D. and Lansing, M.H. (1997). *Treating people with chronic disease: a psychological guide*. American Psychological Association, Washington, DC.
19. Schrader, G., Cheok, F., Holdacre, *et al.* (2003). Effect of psychiatry liaison with general practitioners on depression severity in recently hospitalized cardiac patients: a randomized controlled trial. *JAMA: The Journal of the American Medical Association*, **288**, 2836–45.
20. Williams, J.W., Katon, W., Lin, E.H.B., *et al.* (2004). The effectiveness of depression care management on diabetes-related outcomes in older patients. *Annals of Internal Medicine*, **140**(12), 1015—24.
21. Klerman, G.L. and Weissman, M.M. (1993). *New applications of interpersonal psychotherapy*. American Psychiatric Press, Washington, DC.

22. Wiener, I., Breitbart, W., and Holland, J. (1996). Psychiatric issues in the care of dying patients. In *Textbook of consultation–liaison psychiatry* (eds. J.R. Rundell and M.G. Wise), pp. 804–31. American Psychiatric Press, Washington, DC.
23. Brown, J.H., Henteleff, P., Baratat, S., *et al.* (1986). Is it normal for terminally ill patients to desire death? *The American Journal of Psychiatry*, **143**, 208–11.
24. Bolund, C. (1985). Suicide and cancer II: medical and care factors in suicide by cancer patients in Sweden. *Journal of Psychosocial Oncology*, **3**, 17–30.

5.3.2 Psychiatric aspects of neurological disease

Maria A. Ron

Psychiatric abnormalities are an integral part of neurological disease and their study can improve our understanding of the neural basis of psychiatric illness. This chapter deals with common neurological diseases where psychiatric symptoms are prominent.

Stroke

Stroke is defined as the sudden loss of blood supply to an area of the brain resulting in permanent tissue damage and is the commonest neurological disorder. The incidence of stroke for those aged between 35 and 65 is between 90 and 330 per 100 000. It is commoner in men and the incidence increases with advancing age. Ischaemic stroke is commoner than haemorrhagic stroke and accounts for 80 to 85 per cent of all cases.

Depression

Its prevalence is around 30 per cent in the first few weeks after a stroke—two-thirds of the patients fit the criteria for major depression and the rest for minor depression. Survivors remain at an elevated risk for depression for many years.

(a) Clinical features

- Diurnal variation of mood, weight loss, anergia, insomnia, and loss of libido are prominent in the early stages.
- Anhedonia, suicidal ideation, loss of self-esteem, and feelings of guilt become evident later.
- Irritability and aggressive behaviour are common, especially in those with cognitive impairment.[1]
- The onset of depression may occur acutely in the early post-stroke period or be delayed for 6 months or more.

(b) Factors associated with post-stroke depression

Early after stroke, anterior left-sided lesions involving the cortex and subcortical regions, especially the basal ganglia, and right posterior lesions are more frequently associated with depression.[2] In time these associations become less marked.[3] Cognitive impairment is closely associated with depression early after stroke,[4] and is present in 70 per cent of those with major and 43 per cent of those with minor depression. Old age, a past or family history of depression, and negative life events in the preceding 6 months substantially increase the incidence of post-stroke depression.

Disruption of fronto-subcortical circuits, directly or as a distant effect of stroke, plays a central role in the causation of depression. Decreased metabolism in orbitofrontal, anterior cingulate, and inferior temporal regions has been reported using **PET**.[5] Serotonergic mechanisms have been implicated, with a reduction of 5-hydroxytryptamine-2 receptor binding in the temporal cortex, especially in left hemisphere stroke.

(c) Course and prognosis

The average duration of major depression is around 1 year, with spontaneous remission in many patients. Symptoms of minor depression may persist for 2 years or more. The presence of depression and other psychiatric diagnosis in the first 3 years after a stroke substantially increases the risk of death after controlling for cardiovascular and other risk factors.

Anxiety

- About a quarter of patients fulfil criteria for generalized anxiety disorder during the acute post-stroke phase.
- Rates are lower in community studies (5 per cent)[6] and 1 or 2 years after stroke (4 to 18 per cent). Half of those with anxiety disorder also satisfy criteria for depression.
- Right-sided subcortical lesions may be more common in anxiety disorder, while left-sided pathology, usually involving cortical regions, is more likely when anxiety and depression coexist.[2] Serotonergic abnormalities are also likely to be relevant.

(a) Course and prognosis

Anxiety in the acute post-stroke phase is associated with high mortality rates, comparable to those in depressed patients.

Emotionalism (abnormal crying or laughing)

Emotionalism is usually mood-congruent and is triggered by sad or emotional events. Most patients have some degree of voluntary control.

Emotionalism occurs in a quarter of patients during the first year post-stroke, with a peak in the first month and decreasing gradually thereafter.[7] It is associated with the severity of depression and with left-anterior lesions disrupting serotonergic pathways.

Management

The first step in treating the psychiatric manifestations of stroke is for these to be recognized, and patients need to be routinely assessed for the presence of psychiatric symptoms.

Treatment

Double-blind placebo-controlled trials using nortriptyline, trazodone, and selective serotonin reuptake inhibitors (**SSRI**s) have shown these drugs to be effective.[4] Improvement in depression also results in lasting improvement in cognition and physical activity. Treatment within 3 months of stroke may be followed by the best outcome.

The treatment of anxiety symptoms has been less well documented. Short-acting benzodiazepines, buspirone, and SSRIs are the main pharmacological approaches. The usefulness of psychotherapeutic or behavioural interventions has not been established

and may depend on the severity of symptoms and cognitive impairment.

Emotionalism responds well to treatment with SSRIs and tricyclic antidepressants, even in those without associated depression.

Parkinson's disease

The neurological and cognitive features of Parkinson's disease are dealt with in Chapter 4.1.6 and only commonly encountered psychiatric symptoms will be considered here.

Depression

Its overall prevalence is approximately 40 per cent.[8] Depressive symptoms are more common early in the disease (50 per cent) and in those with onset before the age of 55. For many people, adaptation to the disease results in a return to normal mood. Depression becomes more frequent again in the advanced stages of the disease, particularly in those with rapidly progressive disability.

Major depression is commoner in those with akinetic Parkinson's disease (38 per cent) than in those with classical forms of the disease (15 per cent), but dysthymia is equally common in both. Depression, severe in some patients, has also been reported in the first few weeks after bilateral subthalamic stimulation, a successful treatment for the motor symptoms of Parkinson's disease,[9] but mood changes were less commonly observed a year after surgery.

(a) Clinical features

- Anxiety, agitation, and depressed mood are prominent.
- Depressive symptoms are not closely associated with the severity of motor signs, but may be more severe during 'off' periods and are commoner in those with cognitive impairment.
- Early awakening, motor retardation, and apathy in the absence of mood abnormalities may not be indicative of depression.
- Major depression is associated with functional deterioration at 1-year follow-up.[10]
- Three-quarters of patients with depression also fulfil criteria for anxiety disorder, but only 10 per cent have symptoms of anxiety in isolation.[11]

(b) Mechanisms underlying depression

Studies using PET have suggested that depression and anxiety in Parkinson's disease are associated with a specific loss of dopamine and noradrenaline innervation in the limbic system.[12] Postmortem studies have also described loss of dopaminergic neurones in the ventral tegmental area.[13] Degeneration of these neurones leads to dysfunction of the orbitofrontal cortex with secondary effects on the serotonergic cell bodies of the dorsal raphe.[5]

(c) Management

Optimal control of neurological symptoms may lead to improvement in depression and should be a management aim. Repeated assessment may be needed to differentiate features of the disease, such as apathy, from the symptoms of depression.

(d) Treatment

There are few studies describing the treatment of depression in Parkinson's disease. When antidepressants are clinically indicated because of the severity or persistence of symptoms, the antidepressant profile of side-effects should be considered. Antidepressants with strong anticholinergic effects, such as amitriptyline, may increase cognitive impairment and SSRIs may be preferable. Electroconvulsive treatment is also effective for Parkinson's disease patients with depression and may also transiently improve motor symptoms.

Psychotic symptoms

(a) Clinical features

- Up to 40 per cent of patients on long-term treatment experience visual hallucinations.
- Long duration of illness, age, cognitive impairment, and depression are associated with visual hallucinations.
- Visual hallucinations in clear consciousness are usually fully formed images of people or animals, non-threatening, fleeting, and stereotyped. They are recurrent and tend to occur at night, more commonly in the 'on' periods.
- Sleep disturbances (fragmented sleep, alteration of sleep rhythms, and vivid dreams) often precede daytime hallucinations and may be part of a continuum.
- Delusions are less frequent than hallucinations but are more stressful and difficult to manage. They are usually paranoid, with conspiracy and infidelity themes.
- Severe psychosis is associated with institutional placement, progressive dementia, and increased risk of death (over a quarter of patients within 2 years).

(b) Mechanisms

Psychotic symptoms may represent intrusion of REM sleep imagery into wakefulness. They may be more frequent in those receiving anticholinergics and dopamine agonists, but there is no clear association with dosage or duration of treatment. Stimulation of hypersensitive dopaminergic receptors in the nigrostriatal system by dopaminergic drugs may explain psychosis early in the disease, but it is unlikely to explain late psychosis. The therapeutic efficacy of atypical neuroleptics suggests a role for mesolimbic dopaminergic and serotonergic pathways. Cholinergic deficiency may also be relevant, in patients with dementia and atrophy of the nucleus basalis.

(c) Management

In patients with clouded consciousness, infection, cerebrovascular accidents, and other relevant pathologies need to be excluded. Revision of dopaminergic medication should come next, with reduction of polypharmacy and dose tapering. Anticholinergics, selegiline, amantadine, and dopamine agonists may need to be discontinued, as they are more likely to trigger psychosis.

(d) Treatment

In most patients antipsychotic drugs are needed, as dopaminergic medication is needed to preserve acceptable motor function. Atypical neuroleptics such as clozapine, with its low D2 receptor affinity and few extrapyramidal side-effects, are preferable to typical neuroleptics. Clozapine is effective at doses of less than 100 mg per day. Initial recommended doses of 6.25 to 12.5 mg daily should be gradually increased until the symptoms are controlled. The danger of agranulocytosis and the need to monitor the blood picture

are the main drawbacks, and other atypical neuroleptics (particularly quietiapine) may be preferable. Cholinesterase inhibitors may also be helpful in treating psychotic symptoms in patients with cognitive impairment.

Dopamine dysregulation syndrome

The syndrome is characterized by the compulsive use of dopaminergic medication beyond that needed to control motor symptoms. It is commoner in males with young onset and may affect 4 per cent of patients. Drug-hoarding and drug-seeking behaviour, impaired social functioning, aggression, and reluctance to reduce medication despite severe dyskinesias are common features. Hypomania and frank psychosis may follow.[14]

(a) Management

Reduction of medication resolves symptoms, but a withdrawal state characterized by dysphoria and irritability follows. Treatment is difficult and often unsuccessful, and primary prevention is preferable.

Tourette syndrome

The syndrome is characterized by motor and phonic tics of fluctuating severity.

- Motor tics appear early, between the ages of 3 and 8 years. Simple tics, e.g. eye blinking, are followed by complex stereotypies (i.e. touching, licking). Phonictics (sniffing, throat clearing) appear later.
- Severity of tics peaks by the age of 20 and may lessen thereafter, but total recovery is rare. Severe cases may start in adulthood.
- Tics, preceded by premonitory urges, occur many times a day and are exacerbated by anxiety, boredom, fatigue, and excitement and lessened by alcohol, relaxation, and sleep. Patients may be able to suppress tics for long periods at the expense of a build-up of tension.
- Coprolalia (utterance of obscenities), copropraxia (obscene gestures), echolalia and echopraxia (repetition of words and actions), and self-harm are also common.

Tourette syndrome is part of a spectrum that includes transient childhood tic disorders. Secondary tic disorders following trauma, encephalitis, rheumatic fever, and metabolic and toxic encephalopathies, and those present in inherited degenerative conditions such as Huntington's disease and neuroacanthocytosis, need to be considered in the differential diagnosis.

Epidemiology

Tourette syndrome is more frequent in males (4:1). Its prevalence in male adolescents is around 4 per 10 000, but it may be higher (49 per 10 000) in children with behavioural disturbances[15] and is as high as 1 to 3 per cent in school children if a broad definition of chronic motor and phonic tics is used.

Aetiology

Major gene effects with an autosomal mode of inheritance seem likely. Monozygotic twins have a concordance rate between 50 and 70 per cent for the syndrome compared with 10 to 20 per cent for dizygotic twins, and a third of relatives may have features of the syndrome. Several candidate genes have been assessed including dopamine receptors, the dopamine transporter, and various noradrenergic and serotonergic genes, and although isolated changes in a single locus are unlikely to lead to the syndrome, these alleles could have a significant cumulative effect. Gestational and perinatal risk factors may also play a role. Multiple streptococcal infections in the months preceding symptom onset are commoner in Tourette patients than in controls, suggesting that the paediatric autoimmune neuropsychiatric disorder associated with streptococcal infection (PANDAS) may be the cause of Tourette syndrome in vulnerable patients.

The efficacy of dopamine antagonists and SSRIs in the treatment of tics and obsessive–compulsive symptoms has implicated dopaminergic and serotonergic pathways and the regions where dopaminergic and serotonergic neurons interact (i.e. striatum, substantia nigra, and prefrontal cortex). The beneficial effects of basal ganglia deep brain stimulation also point to basal ganglia dysfunction. PET studies have suggested decreased metabolism and blood flow in basal ganglia–thalamo-cortical projection systems.[16] Post-mortem studies have not shown consistent abnormalities in D1/D2 receptors, and *in vivo* receptor-binding studies have been conflicting, as have structural brain-imaging studies.

Psychiatric comorbidity

(a) Clinical features

- About half of the patients meet criteria for other psychiatric disorders,[17] but it is uncertain whether they should be considered part of the phenotype.
- Depression and anxiety occur in about 25 per cent of patients.
- Personality disorder is present in two-thirds of patients. Borderline, obsessive–compulsive, and paranoid types are the commonest types.
- Attention-deficit hyperactivity disorder may be more common in males and may have educational and behavioural implications. Wide variations in prevalence have been reported (8 to 80 per cent).
- Obsessive–compulsive disorder may be more common in females, reaching its peak in late adolescence. Concern for symmetry, violent and sexual thoughts, forced touching, fear of harming self and others, and a need to do things 'just right' are common features.
- Intellectual ability tends to be normal, but poor performance in complex attentional tasks is associated with attention-deficit hyperactivity disorder.

(b) Management

Explanation and reassurance are often enough in mild cases; for the rest, drug treatment or cognitive-behaviour therapy may be indicated.

(c) Treatment

Neuroleptics are useful in treating tics. Haloperidol, pimozide, and sulpiride are commonly prescribed and atypical neuroleptics (e.g. risperidone, zipresidone, and olanzapine) are also useful, although few controlled trials are available. Clonidine, a $\alpha 2$ adrenergic agonist, is useful, with less severe side-effects. Behavioural interventions (e.g. habit reversal training and bio-feedback techniques) may also suppress tics. Deep brain stimulation of the

centromedian-parafascicular complex of the thalamus or the internal segment of the globus pallidus is reported to ameliorate tics and self-harm.

Variable success has been achieved with behavioural techniques, SSRIs (fluoxetine), and risperidone in the treatment of obsessive–compulsive disorder. The treatment of attention-deficit hyperactivity disorder is more controversial, but benefits may follow the use of clonidine and stimulants such as methylphenidate, pemoline, and dextroamphetamine without increasing tic severity. (See also Chapter 9.2.4.)

Multiple sclerosis

Multiple sclerosis is a common neurological disease, with a prevalence of 50 to 60 per 100 000; it is more common in women. It usually starts between the ages of 20 to 40 and is characterized by multiple demyelinating lesions with a predilection for the optic nerves, cerebellum, brainstem, and spinal cord. In most cases the disease initially follows a relapsing–remitting course, entering a secondary, progressive phase after some years. For a few patients the disease is progressive from the outset.

Purely psychiatric presentations of multiple sclerosis other than dementia are rare, but psychiatric symptoms are common in the course of the illness.

Depression

- Depressive symptoms occur in about 50 per cent of patients in cross-sectional studies[18] and their lifetime prevalence is also around 50 per cent.
- The rates of suicide are more than twice those of the general population, and young males and those socially isolated or with drinking problems are at a greater risk.
- Low mood, negative thoughts, anhedonia, and suicidal ideation are common features of depression.
- Fatigue and poor concentration may be features of multiple sclerosis and have less diagnostic value.

Euphoria

- It is only present in about 10 per cent of patients and is characterized by mild, continuous elation.
- It is best considered as an organic type of personality change.

Emotional lability

- It is as frequent as euphoria.
- Excessive crying is more frequent than laughter.
- It tends to be more severe in those with significant depression.[19]

Psychosis

- It is uncommon, but brief affective or schizophrenia-like psychoses may occur in patients with well-established multiple sclerosis, sometimes coinciding with a relapse.[20]
- Persecutory delusions and lack of insight are common.
- In most patients these are single episodes lasting 4 to 6 weeks that respond well to symptomatic treatment.

(a) Mechanisms of psychotic symptoms

Severity of brain disease, as measured by magnetic resonance imaging (**MRI**), and duration of illness are not closely correlated with **depression**, but they are a risk factor. The personal and social limitations imposed by the disease are an important risk factor for depression.[18] A **genetic predisposition** has been reported in multiple sclerosis patients with bipolar illness.

Euphoria and **emotional lability** tend to occur in patients with advanced disease and cognitive impairment and are more closely related to MRI indices of brain damage. MRI lesions tend to cluster around the temporal lobes in patients with **psychotic symptoms**.

(b) Management of psychotic symptoms

All patients should be assessed for the presence of depression. Fatigue and poor concentration have limited diagnostic value, as they may be features of multiple sclerosis. Although disease-modifying treatments do not increase the overall risk of depression, they may do so in the first 6 months of treatment in those with a previous history of depression.[21] Regular psychiatric assessment in the early stages of treatment is, therefore, important.

(c) Treatment of psychotic symptoms

Few studies have assessed the effect of antidepressants in patients with multiple sclerosis, but SSRIs and other antidepressants appear to be effective. Their side-effects need to be carefully considered for their potential to aggravate or improve neurological symptoms. Cognitive-behaviour therapy aimed at improving coping strategies is also useful. **Emotional lability** responds well to small doses of SSRIs or tricyclic antidepressants but tends to recur when these drugs are discontinued.

Psychotic episodes may require the use of neuroleptics for brief periods, but the long-term use of these drugs is rarely required.

Cognitive impairment

Cognitive impairment is present in about 40 per cent of multiple sclerosis patients[22] and contributes significantly to the overall disability. The pattern of impairment is characterized by the following:

- Attention deficits and slowing of information processing speed are often the first manifestations and may be present early in the disease.[23]
- Memory disturbances, with greater impairment of recall over recognition, appear later.
- Executive function deficits, with poor working memory, abstract reasoning, and use of strategy, are common.
- Language skills and visuospatial functions tend to be preserved.

Cognitive impairment is greater in those with progressive, severe disease, although cases presented as dementia have also been described.[24] Depression worsens cognitive impairment by slowing down information processing and interfering with learning and working memory. Cognitive impairment correlates with MRI markers of disease severity, in particular brain atrophy.

(a) Management of cognitive impairment

Disease-modifying treatments may slow down cognitive impairment, but evidence is insufficient. The same applies to cholinesterase inhibitors. Cognitive rehabilitation has so far been disappointing.

Space-occupying lesions

Brain tumours

Their clinical manifestations are determined by location and by the effects of raised intracranial pressure. Psychiatric symptoms occur in 50 per cent of patients[25] and are of three main types:

- Confusional states and/or progressive cognitive deterioration occur in a third of patients. Disorientation with clouding of consciousness, euphoria, apathy, and loss of insight are prominent in those with confusional states. Progressive memory impairment, loss of initiative, and bradyphrenia occur in patients with a more protracted course and may coexist with signs of raised intracranial pressure.
- Behavioural and mood disturbances occur in 20 per cent of patients. Irritability, euphoria, depression, and, less frequently, psychosis are part of the picture.
- Paroxysmal disturbances such as poorly formed visual hallucinations and automatisms, indicating temporal lobe involvement, are less common.

Fast-growing tumours are more likely to cause psychiatric symptoms (60 per cent in patients with gliomas and 42 per cent in those with meningiomas). Frontal lobe tumours may present with psychiatric symptoms in the absence of other neurological abnormalities.[26] Medial and orbitofrontal tumours lead to emotional symptoms while disinhibition and irritability or marked apathy occurs when the anterior cingulate is involved. Tumours involving the dorsolateral prefrontal regions are more likely to produce abnormalities of executive function (planning, goal-directed behaviour, ability to monitor effective performance). Disturbances of micturition are specifically associated with frontal tumours.

(a) Management

Brain tumours should be suspected in patients with the above syndromes or when psychiatric symptoms are accompanied by neurological abnormalities or appear *de novo* late in life. Imaging usually confirms the diagnosis.

Neurofibromatosis

There are two types of neurofibromatosis, both inherited as autosomal dominant disorders.

Neurofibromatosis 1 is the commonest (incidence 1/3000) and is characterized by cutaneous manifestations (café-au-lait pigmentation) and neurofibromas, which are benign nerve sheath tumours. Gliomas and hamartomas, especially in the eye and optic pathways, and bone dysplasia are also features. Severe learning disability occurs in 4 per cent of patients, but milder cognitive impairment is commoner (80 per cent of cases). Attentional difficulties are common and a third of patients fulfil criteria for attention-deficit hyperactivity disorder. Perceptual and executive functions are worse than memory functions.[27, 28] Epilepsy is twice as common as in the general population. Ventricular enlargement and T_2 MRI hyperintensities in the basal ganglia, internal capsule, thalamus, brainstem, and cerebellum not closely related to the severity of the learning disability are often present.

Neurofibromatosis 2 is much less common (incidence 1/40 000) and is characterized by bilateral vestibular schwannomas, which may also occur in other peripheral nerves. Meningiomas and ependymomas also occur. Hearing loss, vestibular disturbances, and cataracts are the commonest clinical presentations. Cognition is normal.

The loci for neurofibromatosis 1 and 2 have been located to 17q 11.2 and 22q 12.2 respectively. The rate of spontaneous mutations is high and both genes may have tumour-suppressant roles.

Management is aimed at dealing with attention-deficit hyperactivity disorder, learning disability, and epilepsy.

Tuberous sclerosis

Tuberous sclerosis is a rare autosomal dominant disorder with a prevalence of 1/27 000, variable expressivity, and a high spontaneous mutation rate. There is genetic heterogeneity, with loci described in chromosomes 9 and 16. It is characterized by the presence of skin lesions (adenoma sebaceum), calcified subependymal nodules (tubers), and cortical dysplasias. Hamartomas and other neoplasms of the brain, heart, kidney, and liver are part of the picture.

- **Epilepsy** occurs in 60 to 80 per cent of patients, and infantile spasms, a type of epilepsy, with onset in the first 6 months of life, are particularly common.
- Moderate to severe **learning disability** is present in over 50 per cent of patients and is associated with the presence of infantile spasms and poorly controlled epilepsy.[29]
- **Autism** is 200 times more common in tuberous sclerosis than in the general population and tuberous sclerosis occurs in 1 per cent of autistic patients.[30]
- The number of cortical tubers detected by MRI is a marker of disease severity and is related to the degree of learning disability.[31] The presence of tubers in the temporal lobes has been reported to be associated with autism in patients with tuberous sclerosis.

Management is aimed at control of epilepsy and learning disability.

Further information

Lishman, W.A. (1997). *Organic psychiatry: psychological consequences of cerebral disorder* (3rd edn). Blackwell, London.

Mitchell, A.J. (2003). *Neuropsychiatry and behavioural neurology explained*. WB Saunders, Edinburgh.

Rickards, H. (2005). Depression in neurological disorders: Parkinson's disease, multiple sclerosis, and stroke. *Journal of Neurology, Neurosurgery, and Psychiatry*, **76**(Suppl. 1), 48–52.

Burn, D.J. and Tröster, A.I. (2004). Neuropsychiatric complications of medical and surgical therapies for Parkinson's disease. *Journal of Geriatric Psychiatry and Neurology*, **17**, 172–80.

Siegert, R.J. and Abernethy, D.A. (2005). Depression in multiple sclerosis: a review. *Journal of Neurology, Neurosurgery and Psychiatry*, **76**, 469–75.

Amato, M.P., Zipoli, V., and Portaccio, E. (2006). Multiple sclerosis-related cognitive changes: a review of cross-sectional and longitudinal studies. *Journal of the Neurological Sciences*, **245**, 41–6.

Albin, R.L. and Mink, J.W. (2006). Recent advances in Tourette syndrome research. *Trends in Neurosciences*, **29**, 175–82.

Prather, P. and de Vries, P.J. (2004). Behavioral and cognitive aspects of tuberous sclerosis complex. *Journal of Child Neurology*, **19**, 666–74.

References

1. Chan, K.-L., Campayo, A., Moser, D.J., *et al.* (2006). Aggressive behavior in patients with stroke: association with psychopathology and results of antidepressant treatment on aggression. *Archives of Physical Medicine and Rehabilitation*, **87**, 793–8.

2. Robinson, R.G. (1998). Relationship of depression to lesion location. In *The clinical neuropsychiatry of stroke: cognitive, behavioural and emotional disorders following vascular brain injury*, pp. 94–124. Cambridge University Press.
3. Sharpe, M., Hawton, K., Seagroatt, V., *et al.* (1994). Depressive disorders in long-term survivors of stroke: associations with demographic and social factors, functional status, and brain lesion volume. *The British Journal of Psychiatry*, **164**, 380–6.
4. Narushima, K., Chan, K.-L., Kosier, J.T., *et al.* (2003). The effect of early versus late antidepressant treatment on physical impairment associated with poststroke depression: is there a time-related therapeutic window? *The Journal of Nervous and Mental Disease*, **191**, 645–52.
5. Mayberg, H.S. and Solomon, D.H. (1995). Depression in Parkinson's disease: a biochemical and organic viewpoint. *Advances in Neurology*, **65**, 49–60.
6. House, A., Dennis, M., Mogridge, L., *et al.* (1991). Mood disorders in the year after stroke. *British Journal of Psychiatry*, **158**, 83–92.
7. House, A., Dennis, M., Molyneau, A., *et al.* (1989). Emotionalism after stroke. *British Medical Journal*, **198**, 991–4.
8. Brown, R. and Jahanshahi, M. (1995). Depression and Parkinson's disease: a psychosocial viewpoint. *Advances in Neurology*, **65**, 61–84.
9. Berney, A., Vingerhoets, F., Perrin, A., *et al.* (2002). Effect on mood of subthalamic DBS for Parkinson's disease: a consecutive series of 24 patients. *Neurology*, **59**, 1427–9.
10. Starkstein, S.E., Mayberg, H.S., Leiguarda, R., *et al.* (1992). A prospective longitudinal study of depression in patients with Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, **55**, 377–82.
11. Menza, M.A., Robertson–Hoffman, D.E., and Bonapace, A.S. (1993). Parkinson's disease and anxiety: comorbidity with depression. *Biological Psychiatry*, **34**, 465–70.
12. Remy, P., Doder, M., Lees, A., *et al.* (2005). Depression in Parkinson's disease: loss of dopamine and noradrenaline innervation in the limbic system. *Brain: a Journal of Neurology*, **128**, 1314–22.
13. Paulus, W. and Jellinger, K. (1991). The neuropathologic basis of different clinical subgroups of Parkinson's disease. *Journal of Neuropathology and Experimental Neurology*, **50**, 743–55.
14. Giovannoni, G., O'Sullivan, J.D., Turner, K., *et al.* (2000). Hedonistic homeostatic dysregulation in patients with Parkinson's disease on dopamine replacement therapies. *Journal of Neurology, Neurosurgery, and Psychiatry*, **68**, 423–8.
15. Robertson, M.M. and Stern, J.S. (1997). The Gilles de la Tourette syndrome. *Critical Reviews in Neurobiology*, **11**, 1–19.
16. Tanner, C.M. and Goldman, S.M. (1997). Epidemiology of Tourette syndrome. *Neurologic Clinics*, **15**, 395–402.
17. Eidelberg, D., Moeller, J.R., Antonini, A., *et al.* (1997). The metabolic anatomy of Tourette's syndrome. *Neurology*, **48**, 927–34.
18. Ron, M.A. and Logsdail, S.J. (1989). Psychiatric morbidity in multiple sclerosis. A clinical and MRI study. *Psychological Medicine*, **19**, 887–95.
19. Dark, F.L., McGrath, J.J., and Ron, M.A. (1996). Pathological laughing and crying. *The Australian and New Zealand Journal of Psychiatry*, **30**, 472–9.
20. Feinstein, A., du Boulay, G., and Ron, M.A. (1992). Psychotic illness in multiple sclerosis. A clinical and magnetic resonance imaging study. *British Journal of Psychiatry*, **161**, 680–5.
21. Goldman Consensus Group. (2005). The Goldman consensus statement on depression in multiple sclerosis. *Multiple Sclerosis*, **11**, 328–37.
22. Rao, S.M., Leo, G.J., Ellington, L., *et al.* (1991). Cognitive dysfunction in multiple sclerosis. II: impact on employment and social functioning. *Neurology*, **41**, 692–6.
23. Feinstein, A., Youl, B., and Ron, M. (1992). Acute optic neuritis: a cognitive and magnetic resonance imaging study. *Brain: a Journal of Neurology*, **115**, 1403–15.
24. Hotopf, M.H., Pollock, S., and Lishman, W.A. (1994). An unusual presentation of multiple sclerosis. Case report. *Psychological Medicine*, **24**, 525–8.
25. Hecaen, H. and Ajuriaguerra, J. (1956). *Troubles mentaux au cours des tumeurs intracraniennes*. Masson, Paris.
26. Ron, M.A. (1989). Psychiatric manifestations of frontal lobe tumours. *British Journal of Psychiatry*, **155**, 735–8.
27. Hyman, S.L., Shores, A., and North, K.N. (2003). The nature and frequency of cognitive deficits in children with neurofibromatosis type 1. *Neurology*, **65**, 1037–44.
28. Yohay, K. (2006). Neurofibromatosis types 1 and 2. *The Neurologist*, **12**, 86–93.
29. Jozwiak, S., Goodman, M., and Lamm, S.H. (1998). Poor mental development in patients with tuberous sclerosis complex: clinical risk factors. *Archives of Neurology*, **55**, 379–84.
30. Fombonne, E., Du Mazaubrun, C., Cans, C., *et al.* (1997). Autism and associated medical disorders in a French epidemiological survey. *Journal of the American Academy of Child and Adolescent Psychiatry*, **36**, 1561–9.
31. Joinson, C., O'Callaghan, F.J., Osborne, J.P., *et al.* (2003). Learning disability and epilepsy in an epidemiological sample of individuals with tuberous sclerosis complex. *Psychological Medicine*, **33**, 335–44.

5.3.3 Epilepsy

Brian Toone

Introduction

An epileptic seizure has been defined as 'a clinical manifestation presumed to result from an abnormal and excessive discharge of a set of neurones in the brain'.[1] A diagnosis of epilepsy applies with the recurrence of two or more discrete and unprovoked seizures (febrile and neonatal seizures are excluded from this definition).

Epilepsy is one of the more common neurological disorders. It carries with it a greater psychiatric morbidity than is to be found in other neurological disorders of comparable severity. Many of its manifestations resemble and may be confused with psychiatric phenomenology. It is often associated with learning difficulties; it may be a manifestation of acquired brain damage or disease; seizures may occur in the course of substance abuse or be caused by psychiatric treatment. For these and for many other reasons psychiatrists should be familiar with epilepsy, its manifold aetiologies, presentations, and treatment.

Classification: epileptic seizures and epilepsy

A comprehensive taxonomy should embrace classifications of both seizure semiology (i.e. the manifestations of abnormal discharge activity) and of epilepsy syndromes. The position of each seizure type in a seizure classification system is determined by its clinical manifestations, by electroencephalographic changes during the seizure, and by the interictal electroencephalographic abnormalities. A classification of epilepsy syndromes takes into account seizure subtype, and also anatomical substrate, aetiology, age of onset, and other characteristics. Seizure classification is dependent upon entities that are immediately ascertainable; epilepsy syndrome classification depends upon entities (e.g. neuroanatomical substrate) that are more speculative. The International League Against Epilepsy

(ILAE) has chosen to give the former priority, while recognizing the importance of the latter.

A familiarity with terminology, with the more common seizure subtypes, and the more commonly encountered epileptic syndromes, will assist in an understanding of the psychiatric disorders that occur in patients with epilepsy. The aura is a simple partial seizure, i.e. a seizure of focal onset in which consciousness is retained. It may progress to a complex partial seizure in which consciousness is disturbed, or into a generalized tonic, clonic, or tonic–clonic seizure. It may subside without further development. It rarely lasts for more than a few seconds, although patients often find time estimation difficult. It is to be distinguished from the epileptic prodrome, a period characterized by dysphoria, impaired memory and concentration, and minor motor manifestations, which precede the seizure and may last for hours or even days. An automatism may be defined as 'a state of clouding of consciousness which occurs during or after a seizure, and during which the individual retains the control of posture and muscle tone and performs simple or complex movements without being aware of what is happening'.[2] The initial phase, consisting of staring or simple chewing movements, may progress to more complex, stereotyped, and repetitive movements such as fumbling or picking. Automatisms rarely last more than a few minutes and are often very brief. They usually arise from temporal lobe discharges, but may be associated with orbital and mesial frontal lesions. The ictus refers to the period of manifest seizure activity. If this persists for 30 min or more it is described as status epilepticus and constitutes a medical emergency (Table 5.3.3.1).

Seizure types

Partial seizures

Partial seizures have a focal onset and may or may not generalize. They may be simple or complex, depending upon whether or not consciousness remains undisturbed. For this purpose, consciousness is defined as the ability to remain aware or to respond.

(a) Simple partial seizures

The content of the simple partial seizure depends upon the site of the focus. One that arises from motor territory may present as a Jacksonian 'march' or as a versive turning of the head. Speech arrest may be present. A focus in the primary sensory cortex may give rise to poorly formed sensations. The more elaborate sensations and the psychic symptoms that arise, respectively, from the association cortex and from the mesial frontotemporal structures are more likely to progress to complex partial seizures.

(b) Complex partial seizures

These may begin as a simple partial seizure or 'aura', or consciousness may be impaired from the beginning. Characteristic auras include epigastric sensations rising into the thorax and olfactory and gustatory hallucinations, elaborated auditory and visual hallucinations, complex changes in perception (e.g. micropsia, depersonalization), and psychic phenomena such as *déjà vu*. As discharge activity spreads, automatic behaviour may supervene or a secondary generalized seizure may ensue. Complex partial seizures commonly arise from temporal, particularly mesial temporal, structures. Hence the obsolete term 'temporal lobe epilepsy'. They may also arise from the orbital and mesial frontal cortices. Complex partial status, previously known as temporal lobe status, is uncommon. It may present as an organic confusional state and may be mistaken for a florid psychosis. Electroencephalography will usually confirm the diagnosis.

Generalized seizures

Both hemispheres are initially and simultaneously involved, the ictal electroencephalographic pattern and motor manifestations are bilateral and consciousness may be impaired from the onset.

Tonic–clonic seizures

This is the common 'major' seizure formerly referred to as 'grand mal'. A brief tonic phase leads into clonic activity, the entire seizure lasting about 2 min. Generalized seizures may be primary, arising from both hemispheres simultaneously, or they may be due to secondary generalization from a focal onset.

Absence seizures

An absence attack is characterized by abrupt cessation of ongoing activity, a vacant stare, and a period of unresponsiveness lasting from a few seconds to half a minute. The absence may be accompanied by brief clonic movements, especially of the eyelids, a reduction in tone causing the body to slump, or automatic movements. Absence seizures occur more commonly during childhood. The absence must be distinguished from the complex partial seizure, which it may resemble. Absence status is not uncommon in childhood and may be mistaken for inattention.

Myoclonic seizures

These are brief shock-like contractions of groups of muscles.

Table 5.3.3.1 Classification of seizures

1 Localization (partial, focal) seizures
(a) Simple partial seizures
(b) Complex partial seizures
(c) Partial seizures evolving to secondary generalized seizures
2 Generalized seizures
(a) Absence seizures
(b) Myoclonic jerks
(c) Clonic seizures
(d) Tonic seizures
(e) Tonic–clonic seizures
(f) Atonic seizures

The epilepsy syndromes

The ILAE, in order to embrace a wider range of clinical features than is possible in a classification based on seizure types, introduced a classification of seizures and epilepsy syndromes. In this classification the broad division lies between the localization-related or partial epilepsies, the great majority of which, in adults, is made up of symptomatic epilepsies (i.e. epilepsy arising from a known or suspected cause) and the generalized epilepsies. The characteristics of the partial epilepsies are determined by the function of the cortical site from which the seizure emanates. The more common generalized epilepsies are age-related. Though admirable in concept, the classification has proved unwieldly and

the terminology clumsy: as such it seems unlikely to displace the seizure-type classification.

Epidemiology

The cumulative incidence (i.e. lifetime risk) is 3.4 per cent for males, 2.8 per cent for females, but the prevalence is only 7 per 1000. This is because prevalence represents the balance between newly diagnosed cases and permanent remission or death, and epilepsy has a good prognosis with 76 per cent of newly diagnosed cases entering long-term remission.[3] Approximately half can be classified as partial seizures and 40 per cent as generalized. Prognosis varies according to the epilepsy subtype, with partial epilepsy having a poorer outcome. Consequently, people who attend a hospital clinic are unrepresentative, in that they are more likely to have treatment-refractory partial seizures along with the other adverse prognostic factors such as mental handicap, neurological dysfunction, or psychiatric disorders.

Aetiology

Aetiology varies according to the age of onset. In childhood- and adolescence-inherited disorders of metabolism, ante- and perinatal complications, infection, migrational errors, and the consequences of febrile convulsions predominate, in middle life trauma and tumour are most common, and in advanced years cerebrovascular disease and degenerative disorders are predominant.

Only one-quarter to one-third of cases of epilepsy are due to known causes. Many others fall into recognizable syndromes about which much is understood. The partial epilepsies are, by definition, due to focal areas of damage and dysfunction usually involving the cortex. However, although the site may be suggested by the seizure semiology, comprehensive investigation may fail to identify any abnormality. Even when it does so, the radiological appearance may lack aetiological specificity. Some generalized seizures may be identified as primary generalized epilepsy syndromes; for example, juvenile myoclonic epilepsy. These are of uncertain aetiology, though genetic factors are considered important; onset is in childhood or adolescence and the prognosis favourable.

In psychiatric practice seizures may arise iatrogenically; they are usually due to pharmacotherapy, less commonly to electroconvulsive therapy. They may result from the overhasty withdrawal of benzodiazepines or to the use of antidepressant or antipsychotic drugs, most of which are epileptogenic. Such seizures are thought to be provoked and do not form grounds for a diagnosis of epilepsy. Adjustment of drug dosage is usually all that is required. Provoked seizures may also occur during alcohol intoxication ('rum fits') or withdrawal; a genetic predisposition may play a part.

Diagnosis

The diagnosis of epilepsy depends first and foremost on historical information; the patient's own account of the seizure and the observations of a reliable informant are of tantamount importance. A family history of epilepsy should be sought; age of onset should be determined when possible. A history of birth complications, febrile fits, early head injury, or cerebral infection is of particular importance in seizures starting in childhood, adolescence, or early adult life. In middle life symptoms suggestive of developing intracranial pathology and in later life cerebrovascular and degenerative disorders should be sought. The clinician should be aware of specific circumstances and situations that may provoke seizures: alcohol or substance abuse, prescribed drugs that have epileptogenic properties, and intermittent photic stimulation. Physical examination will detect not only gross congenital abnormalities such as tuberosclerosis, but also more subtle features, for example facial or skull asymmetries. The differential diagnosis varies according to age group, but will include vasovagal attacks and pseudoseizures, particularly in the young, vertigo and transient ischaemic attacks in the elderly, and cardiogenic syncope, hypoglycaemic episodes, and migraine at any age.

The role of physical investigation is to confirm the diagnosis of epilepsy when this is in doubt and to identify the cause; it may also help to determine the type of epilepsy and, in the partial epilepsies, the site of seizure onset. Magnetic resonance scanning is now widely available and all new cases of adult-onset epilepsy and patients of any age with partial epilepsy should have the benefit of this technology. A sleep electroencephalograph is still mandatory and may be invaluable not only in differential diagnosis, but in determining the epilepsy subtype and seizure localization. Video-telemetry is invaluable when there is continuing diagnostic uncertainty, and in those cases in which precise localization is necessary for presurgical assessment. Functional neuroimaging may also aid localization.

In psychiatric practice epileptic seizures must be distinguished from psychogenic pseudoseizures, panic attacks, and aggressive episodes. The pseudoseizure may take different forms: the patient may fall or slump to the ground and remain still as in a syncopal attack; or there may be jerking or thrashing of limbs resembling a major tonic–clonic seizure. The absence of tongue-biting, incontinence, or significant injury is often said to distinguish the pseudoseizure from the epileptic seizure, but such guides are frequently unreliable. A history of other conversion disorders or illness behaviour may be obtained, but pseudoseizures also occur in individuals with epilepsy. A detailed description of the attack from the patient and from a witness will be of the greatest assistance in diagnosis, but when in serious doubt video-telemetry may offer a definitive answer. Other useful investigations include serum prolactin levels, which rise postictally to reach a peak between 20 and 30 min after a major epileptic seizure. Episodes of panic are sometimes mistaken for epileptic seizures. Extreme anxiety, especially when accompanied by hyperventilation, may lead to a subjective diminution in awareness, altered perception, and other features suggestive of complex partial seizures. The context in which the attack occurs, a description of initial autonomic arousal, and other symptoms of anxiety should lead to the correct diagnosis. Sudden outbursts of aggressive behaviour, particularly when out of character and context, often give rise to suspicions of epilepsy. Aggression, especially directed aggression, is extremely unusual as a feature of the epileptic seizure, though it may occasionally be seen during the phase of postictal confusion or postictal psychosis.

The psychiatric consequences of epilepsy

The prevalence of psychiatric morbidity among persons with epilepsy is greater than in the general population, but the increase in prevalence will vary according to the type of epilepsy, the presence and extent of brain damage, and the presence of cognitive and physical disability. The more reliable studies are drawn from

community samples. Of children between the ages of 5 and 14 years, 29 per cent showed some psychiatric disorder compared with 6.8 per cent of the general population. The figure rose to 58 per cent when brain damage was present.[4] Pond and Bidwell[5] surveyed 14 general practices and reported a prevalence of 29 per cent in adults, increasing to over 50 per cent in patients with temporal lobe epilepsy. Neurotic disability accounted for the great majority of cases. Psychiatric morbidity is over-represented in clinic attenders and in patients with partial epilepsy.[6]

Personality disorder and social development

Notions of an epileptic personality arising out of a hereditary 'taint' persisted well into the present century. The person with epilepsy was said to be explosively aggressive, rigid, egocentric, and irritable. These beliefs were formed by observations of often oversedated inmates of epileptic institutions. The concept of a specific epileptic personality has now largely been abandoned, though it is acknowledged that some features associated with, but not specific to, epilepsy may exercise a powerful influence on personality development. Many of these are consequences of brain damage rather than epilepsy as such. Thus learning difficulties, leading to limited educational opportunity, adult unemployment, and socioeconomic disadvantage may be significant personality determinants. But even in the epileptic individual without brain damage the sedative actions of anticonvulsant medication, the continuing stigma of seizure activity, and the social and occupational constraints are not without their effects on the developing personality.

A particular link between temporal epilepsy and abnormalities of personality has long been debated[7] and certain exaggerated traits (e.g. hypergraphia) are consistently reported,[8] but many of the personality difficulties may be explained by the refractory nature of temporal lobe seizures and the need for increased medication.

Psychoses

(a) Chronic interictal psychoses

Throughout the first half of this century the relationship between epilepsy and schizophrenia was debated at length, usually in terms of whether the presence of one condition encouraged or discouraged the development of the other—the affinity and antagonism hypotheses, respectively. In recent years, particularly following the publication of Slater's seminal studies,[9] informed opinion has moved firmly behind the first view. Epidemiological studies based on national registers[10,11] find a higher prevalence of chronic psychosis in epileptic subjects than in the general population. A neurology outpatient clinic study[12] reported schizophrenia to be nine times more common in epilepsy than in a migraine control group. The onset, cause, and clinical characteristics are, to a very large extent, indistinguishable from those of more usual forms of schizophrenia, although negative symptoms occur less frequently, thought disorder is rarely encountered and the outcome may be more benign. Psychosis usually develops 11–15 years following the onset of epilepsy. The aetiology remains uncertain. Cases in which the epilepsy takes the form of complex partial seizures arising from the mesial temporal or frontal lobes are over-represented; there may be a slight left-sided predominance. A family history of schizophrenia was thought to be unusual, but some recent studies have cast doubt on this assumption. Neuropathological examination, more readily available with the increasing practice of epilepsy surgery, has proved less informative than had been hoped, but subjects undergoing temporal lobectomy for resection of small neurodevelopmental lesions, e.g. ganglioglioma, dysembryoplastic neuroepithelioma (DNET)[13] appear at greater risk of developing a post-operative chronic interictal psychosis. The results of structural neuroimaging studies have been inconclusive with reports of both reduction and increase in amygdala size. The risk of bipolar illness or affective psychosis does not appear to be increased in epilepsy.

(b) Postictal psychosis

The other common form of epileptic psychosis develops following an exacerbation of seizure activity. Because of this close temporal relationship the instrumental role of epilepsy in the aetiology of the psychosis is not open to question. The salient characteristics have been described.[14] The psychosis usually occurs following a cluster of complex partial seizures usually followed by secondary generalization. Characteristically, the subject appears to make a complete recovery, but 1 to 2 days later, the so-called lucid interval, becomes floridly psychotic. Affective, schizophrenic, and confusional elements may be present. An electroencephalogram recorded at the time shows increased focal discharge activity, though less than would be seen in partial status. Spontaneous recovery is to be expected, usually within a week of onset. The first episode is usually delayed until early adult life, a decade and a half after the first seizure. Half of those affected will have further similar episodes. Fifteen to 20 per cent will progress to develop a chronic interictal psychosis. Bilateral EEG discharges are seen more commonly in those patients who develop postictal psychosis. Functional neuroimaging at the time of the psychosis demonstrates relative mesial temporal hyperperfusion suggestive of an active process, either continuous seizure discharge or seizure inhibition.

Rarely, the introduction of certain anticonvulsant drugs may seem to precipitate psychotic episodes. This may be associated with rapid seizure control and give rise to the term 'forced normalization'.[15] In recent years vigabatrin has seemed to be the drug most responsible.

Sexual function

A diminution in sexual interest, a decrease in activity, and impaired performance are the most common aspects of sexual dysfunction in epilepsy. Men have been studied more thoroughly than women. In patients receiving antiepilepsy drugs libido may be diminished and erectile potency impaired. Levels of free testosterone, the biologically active hormone, may be diminished. Sperm concentrations may be reduced, morphological abnormalities may occur more commonly and mobility may be reduced. Menstrual irregularities are increased in women with epilepsy and are related to seizure frequency, polytherapy with antiepilepsy drugs, and the use of sodium valproate. Infertility, ovulation, and the polycystic ovarian syndrome occur more commonly. Hyposexuality may be more pronounced in patients with partial epilepsy, but this may simply reflect the refractory nature of partial epilepsy and the greater amount of drugs prescribed.[16,17]

Epilepsy and crime

There is an association between epilepsy and criminal activity. Male epileptics are three times more likely to receive a criminal conviction,[18] in England and Wales between 0.7 and 0.8 per cent of the

prison population suffers from epilepsy, a figure considerably higher than in the general population,[19] but the pattern of offence does not differ. The reasons for this are unclear. Low intelligence and low socioeconomic class are common to both epilepsy and prison populations; the role of brain damage as distinct from epilepsy has not been fully evaluated. Crimes of violence in the context of disturbed ictal or postictal behaviour do occur, but are extremely rare.[20]

Neurotic illness

Neurotic illness, more especially anxiety and depression, largely account for the increased psychiatric morbidity that is to be found in patients with epilepsy. These disorders have few distinctive characteristics. They may be explained by the adverse social, educational, and economic disadvantages that confront people with epilepsy. A phobic anxiety akin to agoraphobia may be seen in some individuals who fear the onset of a seizure in public places. Obsessive–compulsive disorder does not seem to occur more commonly in epilepsy.

Epilepsy and suicide

Suicide is increased fivefold among patients with epilepsy, but is considerably higher among those with temporal lobe epilepsy.[21] Among patients presenting with self-harm, epileptic subjects are over-represented from five- to sevenfold.[22]

Treatment

Seizure control is most effectively achieved through the appropriate use of anticonvulsant drugs. The use of behavioural techniques to inhibit seizure activity holds promise, but is still in its infancy. Surgical treatment is increasingly available, but should only be considered for those patients who have failed repeatedly to respond to drug therapy and who have resectable lesions. Drug treatment should aim to achieve seizure control through the use of a single anticonvulsant drug, thus minimizing unwanted side-effects. This should be possible in the great majority of patients. If a first-choice drug fails, a second-choice drug should be substituted. Polytherapy may be necessary, notably in the management of partial seizures, but should be avoided wherever possible. Most first-line drugs, except phenytoin, are described in Chapter 6.2.6. Phenytoin, although relatively toxic, especially to the cerebellum, is an effective anticonvulsant and still widely prescribed. Serum monitoring is particularly important. More recently introduced drugs include topiramate and levetiracetam. The benzodiazepines, clobazam and clonazepam, may be used in adjunctive therapy. Lamotrigine and levetiracetam appear to be as efficacious as the longer established drugs and have less side-effects. There is little place for either phenobarbitone, which is unduly sedative and may cause depression and behavioural disturbance especially in children and adolescents, or vigabatrin, which may cause visual field constrictions. Vigabatrin, and to a lesser extent topiramate may cause psychotic episodes. The need for continuing anticonvulsant therapy should be reviewed by a specialist neurologist once the patient has been free of seizures for 2 years. For a more detailed account of the management of epilepsy and of status epilepticus the reader is referred to Shorvon *et al.*[23]

The treatment and outcome for psychogenic seizures have received increasing attention. The importance of early diagnosis is emphasized.[24] Psychological treatment, particularly cognitive–behavioural therapy, can be effective. Reduction or cessation of symptoms can be achieved in at least half of the cases within a 6–12 month period.

Further information

Asbury, A.K., McKhann, G.M., *et al.* (eds.) (1992). *Diseases of the nervous system* (2nd edn). W.B. Saunders, Philadelphia, PA.

Hopkins, A., Shorvon, S., and Cascino, G. (eds.) (1995). *Epilepsy* (2nd edn). Chapman & Hall Medical, London.

Lishman, W.A. (1998). *Organic psychiatry* (3rd edn). Blackwell Science, Oxford.

Trimble, M.R. (1991). *The psychoses of epilepsy*. Raven Press, New York.

References

1. Hopkins, A. and Shorvon, S. (1995). Definitions and epidemiology of epilepsy. In *Epilepsy* (2nd edn) (eds. A. Hopkins, S. Shorvon, and G. Cascino), pp. 1–25. Chapman & Hall Medical, London.
2. Fenton, G.W. (1981). Psychiatric disorder of epilepsy: classification and phenomenology. In *Epilepsy and psychiatry* (eds. E.H. Reynolds and M.R. Trimble), pp. 12–26. Churchill Livingstone, Edinburgh.
3. Annegers, J.F., Hauser, W.A., and Elveback, L.R. (1979). Remission of seizures and relapse in patients with epilepsy. *Epilepsia*, **20**, 729–37.
4. Graham, P. and Rutter, M. (1968). Organic brain dysfunction and child psychiatric disorder. *British Medical Journal*, **3**, 695–700.
5. Pond, D.A. and Bidwell, B.H. (1960). A survey of epilepsy in fourteen general practices. II. Social and psychological aspects. *Epilepsia*, **1**, 285–99.
6. Edeh, J., Toone, B.K., and Corney, R.H. (1990). Epilepsy, psychiatric morbidity, and social dysfunction in general practice. *Neuropsychiatry, Neuropsychology and Behavioural Neurology*, **3**, 180–92.
7. Bear, D.M. and Fedio, P. (1977). Quantitative analysis of inter-ictal behaviour in temporal lobe epilepsy. *Archives of Neurology*, **34**, 454–67.
8. Waxman, S.G. and Geschwind, N. (1974). Hypergraphia in temporal lobe epilepsy. *Neurology*, **24**, 629–36.
9. Slater, E., Beard, A.W., and Glithero, E. (1963). The schizophrenia-like psychoses of epilepsy. *British Journal of Psychiatry*, **109**, 95–150.
10. Bredkjaer, S.R., Mortensen, P.B., and Parnas, J. (1998). Epilepsy and non-organic non-affective psychosis. National epidemiological study. *British Journal of Psychiatry*, **172**, 235–9.
11. Stefansson, S.B., Olafsson, E., and Hauser, W.A. (1998). Psychiatric morbidity in epilepsy: a case control study of adults receiving disability benefit. *Journal of Neurology, Neurosurgery and Psychiatry*, **64**, 238–41.
12. Mendez, M.F., Grau, R., Doss, R.C., *et al.* (1993). Schizophrenia in epilepsy: seizure and psychosis variables. *Neurology*, **43**, 1073–7.
13. Andermann, L.F., Savard, G., Meencke, H.J., *et al.* (1999). Psychosis after resection of ganglioglioma or DNET: evidence for an association. *Epilepsia*, **40**, 83–7.
14. Logsdail, S.J. and Toone, B.K. (1988). Post-ictal psychoses. *British Journal of Psychiatry*, **152**, 246–52.
15. Landolt, H. (1953). Some clinical EEG correlations in epileptic psychoses (twilight states). *EEG and Clinical Neurophysiology*, **5**, 121.
16. Toone, B.K. (1995). Epilepsy and sexual life. In *Epilepsy* (2nd edn) (eds. A. Hopkins, S. Shorvon, and G. Cascino), pp. 555–65. Chapman & Hall Medical, London.
17. Hertzog, A.G. and Fowler, K.M. (2005). Sexual hormones and epilepsy: threat and opportunities. *Current Opinion in Neurology*, **18**, 167–172.
18. Gudmundsson, G. (1966). Epilepsy in Iceland. A clinical and epidemiological investigation. *Acta Neurologica Scandinavica*, **25**, 7–124.
19. Gunn, J. and Fenton, G.W. (1971). Epilepsy, automatism and crime. *Lancet*, **i**, 1173–6.

20. Delgado-Escueta, A.V., Mattson, R.H., King, L., *et al.* (1981). The nature of aggression during epileptic seizures. *New England Journal of Medicine*, **305**, 7110–16.
21. Barraclough, B. (1981). Suicide and epilepsy. In *Epilepsy and psychiatry* (eds. E.H. Reynolds and M.R. Trimble), pp. 72–6. Churchill Livingstone, Edinburgh.
22. Hawton, K., Fagg, J., and Marsack, P. (1980). Association between epilepsy and attempted suicide. *Journal of Neurology, Neurosurgery and Psychiatry*, **43**, 168–70.
23. Shorvon, S., Dreifuss, F., Fish, D., *et al.* (eds.) (1996). *The treatment of epilepsy*. Blackwell Science, Oxford.
24. Reuber, M., Fernandéz, G. Bauer, J., *et al.* (2002). Diagnostic delay in psychogenic non-epileptic seizures. *Neurology*, **58**, 493–5.

5.3.4 Medical conditions associated with psychiatric disorder

James R. Rundell

Introduction

Seven out of 10 office visits to a primary care practitioner are related to a chronic illness.[1] There are high levels of association of many of these chronic conditions with psychiatric disorders.[1] Comorbid medical and psychiatric conditions increase use of medical resources and costs, as well as amplify functional impairment.[2] For example, depression is associated with an approximately 50 per cent increase in medical costs of chronic medical illness, even after controlling for severity of physical illness.[2, 3] Dementia is associated with hospital costs up to 75 per cent higher than for non-demented patients.[4]

As important as a comprehensive knowledge of psychiatric diagnosis and psychosocial formulation is to a consulting psychiatrist, it is also vital to understand the pathophysiology and clinical characteristics of the medical and surgical conditions that frequently coexist with psychiatric disorders. It is also important to know the behavioural and psychiatric side effects of medications and substances. Lacking this data permits only a partial and inadequate approach to diagnosis and treatment.

This section describes general medical disorders associated with psychiatric syndromes. The pathophysiology and clinical characteristics of the medical disorder are described first, followed by psychiatric syndromes often seen with that diagnosis.

Cardiovascular disorders

Ventricular dysrhythmias

Sudden cardiac death is responsible for 300 000 deaths annually in the United States.[5,6] Sympathetic nervous system activity increases the likelihood of ventricular dysrhythmias[6] especially when there is prior ischaemic damage. Sympathetic nervous system stimulation, which increases heart rate, can trigger ectopic sites in the myocardium, which override normal conductive pathways, producing potentially fatal dysrhythmias. Either the peripheral sympathetic nervous system or the central nervous system can generate stimuli leading to this phenomenon. Therefore, anxiety and stress may increase the risk of dysrhythmia.[7] Among individuals with pre-existing heart disease or dysrhythmias, activities which precipitate adrenergic discharge may produce ventricular dysrhythmias—for example, public speaking, road rage, and recall of emotionally charged events.[8] In one series of patients, psychological stressors were more reliable triggers of dysrhythmias than physical manoeuvres such as carotid sinus massage, hyperventilation, and the Valsalva manoeuvre.[8] Simple and inexpensive non-pharmacological techniques such as relaxation training, hypnosis, and medication have been shown to improve ventricular dysrhythmias.[8, 9]

Depression has also been associated with lower threshold for ventricular dysrhythmias.[10] Patients with depression exhibit dysregulation of the sympathoadrenal system—hypothalamic corticotropin releasing factor-containing neurons appear to stimulate several autonomic centres involved in regulating sympathetic activity.[11] Smith *et al.*[8] found that deaths within 18 months of a myocardial infarction were concentrated among depressed patients with 10 or more premature ventricular contractions per hour. In this group of patients, 83 per cent of mortality was due to 'arrhythmic deaths'.

Hypertension

More than 60 million Americans have hypertension. The prevalence among whites is about 15 per cent, but is over 25 per cent in the African–American population. 95 per cent of people with hypertension have primary, or idiopathic, hypertension. The remaining 5 per cent have secondary hypertension, due to conditions or substances such as renal disease, steroids, or oral contraceptives. Managing hypertension reduces the morbidity and mortality of the condition. Constitutional and stress-related factors contribute to hypertension. Patients with hypertension, in general, have a more prolonged vasoconstrictive response to psychological stress than patients with normotension,[12] which may result in both short-term and long-term blood pressure elevation due to this interplay of environmental and constitutional factors.[13]

Myocardial infarction

Myocardial ischaemia often leads to myocardial infarction. Acute myocardial infarction can develop at rest or with normal activity. Deaths associated with acute myocardial infarction occur during the first few hours after the onset of symptoms, and are the result of ventricular fibrillation. It is important that patients know the warning signs and seek care promptly when symptoms develop. Unfortunately, as many as 20 per cent of myocardial infarctions are unrecognized. Denial of acute myocardial infarction symptoms and warning signs by individuals, particularly men, are a frequent source of mortality and morbidity. The roles of gender-specific differences in terms of establishing predictors for clinical outcomes is understudied.[14]

The most common precipitant of myocardial ischaemia among patients with pre-existing coronary artery disease is stress.[15] Stress-induced ischaemia is more common than ischaemia induced by physical stressors. Recovery from a myocardial infarction is also highly dependent on psychosocial factors. Ruberman *et al.*[16] demonstrated that postmyocardial infarction patients, who are socially isolated and have high stress levels, have at least four times

the risk of death, compared to their counterparts who have lower levels of stress and isolation. Particularly, lack of a close confidant predicts negative outcomes after myocardial infarction, including further cardiac events.[17] In addition, emotional distress after myocardial infarction is associated with poorer outcomes in terms of quality of life and psychological adjustment.[18]

Depression may occur in 31 per cent of patients admitted for acute myocardial infarction.[19] Presence of major depressive disorder in a patient with cardiac disease has a significant association with morbidity and mortality. Carney *et al.*[20] found that major depressive disorder was the best single predictor of myocardial infarction, angioplasty, and death during the 12 months following cardiac catheterization. Patients with a history of myocardial infarction and major depressive disorder are up to three to five times more likely to die within 6 months of discharge than non-depressed patients following infarction.[19] As to how depression increases risk include hypothalamic–pituitary–adrenocortical and/or sympathoadrenal hyperactivity, diminished heart rate responsivity, ventricular instability, myocardial ischaemia due to stress, and alterations in platelet receptors and reactivity.[21, 22] The data are limited, antidepressant treatment, stress management, and relaxation training in patients with coronary artery disease or myocardial infarction and major depression probably reduces mortality.[23]

Type A and type D personality

Assessment of the patient's personality and behavioural style is important because type A behavioural patterns increase the risk of a myocardial infarction.[24] The type A behaviour pattern includes ambitiousness, aggressiveness, competitiveness, impatience, muscle tenseness, alertness, rapid and emphatic vocal style, irritation, cynicism, hostility, and an increased potential for anger. Very frequently, such individuals are also hard-working 'workaholics' who deny physical or emotional vulnerability. Their self-esteem is often dependent on constant achievement. Unstable cardiac function poses an immediate and ongoing threat to them, and challenges their need to be in control of their environment and bodies. Many clinicians believe that modifying a type A behaviour pattern is an integral part of preventing future myocardial infarctions.[24] Group or individual psychotherapy that reduces type A behaviour and other behavioural risks has been shown to lower the incidence of recurrent infarction and cardiac death in patients with a previous myocardial infarction.[25]

There has been recent attention to the type D personality construct.[26, 27] D behaviour is characterized by inhibition of negative emotions and avoiding social contacts with others. D personality patients may be at increased risk for cardiovascular morbidity and mortality.[26] Cortisol may be a mediating factor for this increased risk. D personality predicts cardiac events after controlling for concurrent stress and anxiety.[27] Studies are needed to validate this personality construct, further define associations with cardiac outcome, and develop treatment approaches for patients with this personality style.

Respiratory disorders

Asthma

Asthma affects between 3 and 5 per cent of the population of the United States. The three hallmarks of the disease are airway inflammation, airway hyperresponsiveness, and a partially reversible airway obstruction. It is one of the classic 'psychosomatic diseases'. Emotional arousal causes changes in airway tone. The severity of an asthma attack is highly correlated with presence of major depressive disorder, panic attacks, general anxiety, and level of fear among children, adolescents, and adults.[28] Asthma patients with psychiatric disorders have worse asthma control, more frequent exacerbations, and worse quality of life than asthma patients without psychiatric disorders.[29] Education, relaxation, biofeedback, and family therapy have each shown efficacy in the management of asthma.[30] Important in the management of asthma is education about the adverse effects of antiasthma medications, which include jitteriness, palpitations, and insomnia. These side effects may require treatment with behavioural and/or psychopharmacological therapies.

Chronic obstructive pulmonary disease

Patients with chronic obstructive pulmonary disease (COPD) have slowly progressive airway obstruction. The course of the disease is punctuated by exacerbations due to pulmonary infection, heart failure, and poor compliance with prescribed therapy. Generally affects middle-aged and older patients. They present with dyspnoea, exercise intolerance, cough, and sputum production. Physical examination reveals lung overinflation, prominent use of accessory muscles to augment respiration, diminished breath sounds, and diffuse wheezing. As with asthma, pharmacological treatments for COPD can cause psychiatric symptoms, especially higher doses of steroid medications. Patients with COPD must stop smoking; pulmonary function declines faster in smokers who develop COPD than non-smokers who develop COPD.

The chronic hypoxia caused by COPD compromises cognition and mood, which, in turn, can produce delirium, mood lability, mood disorders, and restriction in daily activities. Depression is present in 20–60 per cent of COPD patients.[31] Depression adversely affects treatment adherence and may increase risk for poor outcomes. There is considerable evidence that supplemental oxygen improves cognitive function and quality of life.[30] Unfortunately, mood improvement with supplemental oxygen has not been conclusively demonstrated.

Panic attacks are reported in up to 38 per cent of patients with COPD.[32] Benzodiazepines, which are highly effective for controlling panic attacks, have limited usefulness in patients with COPD because they can suppress respiratory function and if used chronically result in tolerance and dependence. Carbon dioxide likely plays a role in promoting panic attacks; carbon dioxide levels increase with COPD disease progression. Antidepressants are useful in patients with COPD who develop panic attacks. Low-dose neuroleptic medications (e.g. 0.5–1.0 mg risperidone orally two to three times daily) are also sometimes used for severe fear and panic, especially in intensive care unit settings (e.g. when weaning the patient from a respirator). Neuroleptics do not directly suppress respiration, though caution must be exercised so that the sedation induced by neuroleptics—potentially combined with other sedating agents—does not reduce respiratory effort beyond that required to maintain adequate oxygenation. Function must also be monitored to ensure that neuroleptic use does not affect cardiac conduction or cause dysrhythmias.

Pulmonary embolism

Patients with psychiatric disorders, including bipolar disorder, anxiety disorder, and schizophrenia, are at increased risk for pulmonary embolism.[33] Embolism may account for a portion of the

excess risk of death among people with schizophrenia, even after controlling for blood pressure, cholesterol, body mass index, smoking, exercise, alcohol intake, and education level.[34] Most thromboemboli originate in the deep veins of the thigh. The diagnosis of pulmonary embolus is often missed because the clinical findings are non-specific. They include dyspnoea, pleuritic chest pain, haemoptysis, tachypnoea, and wheezing or crackles on pulmonary examination. Number of factors predispose to pulmonary thromboemboli: cancer, stroke, myocardial infarction, congestive heart failure, sepsis, pregnancy, lower extremity fractures, major surgical procedures, polycythaemia vera, and paroxysmal nocturnal haemoglobinuria. Pulmonary emboli are treated with heparin and warfarin. Fibrinolytic drugs and acute embolectomy are used in certain situations. The differential diagnosis of sudden anxiety or a panic attack includes pulmonary embolus.

Sleep apnoea

Apnoea is defined as the complete cessation of respiratory airflow for 10 or more seconds.[35] Apnoea can occur during any sleep stage, but is particularly likely to occur during the period of rapid eye movement sleep. It is important to remember that normal people have apnoeic episodes during sleep. When apnoeic events are frequent and prolonged, they lead to chronically disrupted sleep and excessive daytime somnolence. This defines the condition known as sleep apnoea. Sleep apnoea can be central, obstructive, or a mixture of the two. Central sleep apnoea is caused by an abnormal central drive to the respiratory muscles. Congestive heart failure is the most common cause, followed by neurological disorders involving the brainstem and respiratory centres. Obstructive sleep apnoea is more common; obesity is a major risk factor, but is not always present. Aside from disrupted sleep and daytime somnolence, associated symptoms include an inability to concentrate, depressed mood, irritability, and personality changes. The sleeping partner often sleeps in another room because of the individual's very loud snoring, snorting, gasping, and restlessness. Treatment with continuous positive airway pressure is often effective. Patients should avoid sedatives and alcohol. If obese, they should lose weight.

Gastrointestinal disorders

Oesophageal dysmotility

Oesophageal dysmotility can be demonstrated in 30 per cent of patients with non-cardiac chest pain;[36] a significant number of non-cardiac chest pain patients lack any evidence of oesophageal reflux and have reduced perception thresholds for pain. Cases of oesophageal dysmotility often lead to psychiatric consultation. Situational stress has not been conclusively linked to oesophageal dysmotility, but major psychiatric illness has.[37] The majority of patients with oesophageal motility disorders have an Axis I psychiatric illness, especially major depressive disorder (52 per cent), generalized anxiety disorder (36 per cent), somatization disorder (20 per cent), and substance-related disorders (20 per cent).[38] Smooth muscle relaxants, such as calcium-channel blockers, are superior to psychiatric treatments in improving physiological measures (such as oesophageal motility testing), antidepressants, and behavioural therapies produce more impressive changes in patients' subjective oesophageal complaints and level of psychological well-being.[39]

Irritable bowel syndrome

Irritable bowel syndrome (IBS) ranks second only to the common cold as a cause of absenteeism from work,[40, 41] affecting between 8 and 17 per cent of the general population in the United States.[40] Symptoms include abdominal pain (relieved by defecation), and various forms of disturbed defecation such as altered stool frequency, altered stool form, altered stool passage, passage of mucus, and bloating. Symptoms must be continuous or recur within 3 months to meet the criteria for a diagnosis of irritable bowel syndrome. The severity of this syndrome frequently correlates with periods of emotional stress; the sympathetic nervous system inhibits gastric motility.

The enteric nervous system contains approximately 100 million neurons, close to the same number found in the spinal cord,[41] and more than those distributed to any other organ or physiological system. It therefore, makes sense that the gastrointestinal tract is uniquely sensitive to the neurophysiological aspects of the stress response. With IBS who seek medical care exhibit high rates of psychiatric disorders. The most frequently occurring are panic disorder (26 per cent), generalized anxiety disorder (26 per cent), social phobia (26 per cent), and major depressive disorder (23 per cent).[42] Patients with irritable bowel syndrome who are depressed and complain of diarrhoea may benefit from tricyclic antidepressant treatment, at least partially because of their anticholinergic effects. Anxious patients may also benefit from, and well-tolerate buspirone. At least one in eight IBS patients are offered an antidepressant,[43] though data suggest that antidepressants are more consistent in improving global measures than specific gastrointestinal symptoms. A group of patients with treatment-refractory irritable bowel syndrome—nearly half had no psychiatric disorder—more than 90 per cent benefited from low-dose antidepressant or antianxiety medications: 92 per cent of patients improved, and 56 per cent experienced complete remission of irritable bowel symptoms.

Inflammatory bowel disease

Inflammatory bowel disease is the collective term for patients who have ulcerative colitis or Crohn's disease. The aetiology of inflammatory bowel disease is unknown, but it may involve immunological, infectious, or environmental factors.[44] The primary manifestations of acute ulcerative colitis are rectal bleeding, diarrhoea, urgency, fever, weight loss, and, sometimes, abdominal pain. Crohn's disease presents with malaise, fever, abdominal pain, and frequently rectal bleeding. Surgical treatment (colectomy) cures ulcerative colitis but not Crohn's disease. However, surgery is usually a last resort in ulcerative colitis.

Despite the strong beliefs of early psychosomatic theorists, there is no objective evidence that psychiatric disorders cause inflammatory bowel disease. However, patients with this disease and who have psychiatric disorders are more likely to have unexplained physical symptoms in other organ systems, more disability than patients with similar disease severity and no psychiatric disorder, and prior histories of physical and sexual abuse.[45] Exacerbations of inflammatory bowel disease symptoms are positively associated with major life events and major stressors.[41,46] Stress-induced alterations in gastrointestinal inflammation may be mediated through changes in hypothalamic–pituitary–adrenal axis function and alterations in bacterial-mucosal interactions, and via mucosal mast cells and mediators such as corticotrophin releasing factor.[47] Treatment focuses on the identification and treatment of

psychiatric disorders, if found, and on stress management and quality of life issues. Walker *et al.*[44] treated inflammatory bowel disease patients who had major depression with an antidepressant and found marked improvement in depression and ability to function. Relaxation, stress management,[45, 48] and hypnotherapy were found to reduce abdominal pain and diarrhoea.

Gastroesophageal reflux and peptic ulcer disease

Acid reflux and peptic ulcer disease are common causes of non-cardiac chest pain.[49] Ulcer disease occurs when the balance between stomach acid and mucosal defence factors is disrupted. Gastric acid, *Helicobacter pylori*, and non-steroidal antiinflammatory drugs are the most important risk factors in the development of peptic ulcers.[50] The majority of patients with peptic ulcer disease are present with epigastric pain that begins 1 to 3 h after eating. Treatment is aimed at reducing gastric acid (e.g. using cimetidine and ranitidine), improving mucosal defences (with, for instance, sucralfate), and/or eradicating *H. pylori* (antibiotics).

On examination, at least half of the patients initially suspected of having peptic ulcer disease do not have evidence of an ulcer.[51] Among patients with non-ulcer dyspepsia, psychiatric comorbidity is high. Magni reported that 87 per cent of patients with non-ulcer dyspepsia have one or more anxiety disorders compared with 25 per cent of those with dyspepsia where there is endoscopic evidence of ulcer.[52] Ang *et al.* reported that a majority of patients who have typical symptoms of gastroesophageal reflux do not have erosions on examination; those patients with non-erosive reflux disease have a higher prevalence of psychiatric morbidity.[53]

Metabolic disorders

Obesity

Obesity is becoming an epidemic throughout the developed world. Existing standard treatments in university settings, only 20 per cent of obese patients lose around 9 kg (about 20 lbs) at 2-year follow-up and only 5 per cent of patients lose about 18 kg (40 lbs).[54] The majority of people who lose weight on a diet gain it all back. Weight loss and weight maintenance after loss is associated with more initial weight loss, reaching a self-determined goal weight, having a physically active lifestyle, a regular meal rhythm including breakfast and healthier eating, control of overeating, and self-monitoring of behaviours.[55] Associated with weight regain after significant weight loss include a history of weight cycling, disinhibited eating, binge eating, more hunger, eating in response to negative emotions and stress, and more passive reactions to problems.

There is no ideal treatment for weight loss. Weight-loss programmes vary considerably in terms of risk, cost, and efficacy. For most patients with mild to moderate obesity, a multidimensional approach is best, combining diet, exercise, behaviour modification, and social support. Motivated patients with morbid obesity (more than 100 per cent overdesired body weight) may be considered for very low calorie diets, with the emphasis on long-term diet, behavioural change, exercise, and social support. Increasingly, surgical approaches are being used; patient selection procedures have been developed to address motivations, psychological resilience, dietary education, potentially complicating psychiatric or substance-related factors, and ensuring patients are aware these procedures are not without risk—there are many structural and metabolic complications, including death. It is important to treat comorbid psychiatric illnesses. It is associated with excessive intake of carbohydrate-rich foods and with resistance to engaging in physical activity.[56] With mood disorders and schizophrenia have a high prevalence of risk factors for cardiovascular disease, diabetes, and obesity, which are on the order of 1.5 to 2.5 times higher than in the general population.[56] On the other hand, it is also important to be mindful of metabolic effects of psychopharmacological agents, especially second generation antidepressants. The latter are associated with weight gain, dyslipidemia, and abnormal glucose homeostasis, especially with olanzapine.[56]

Wilson's disease

Wilson's disease, or hepatolenticular degeneration, is an autosomal recessive disorder affecting between one and three persons per 100 000 of the population. The abnormality in Wilson's disease is defective hepatic excretion of copper. The consequence is copper deposition and injury to many organs, particularly the liver and the brain, including diffuse white matter lesions seen on MRI in many patients.[57] The genetic defect occurs on chromosome 13, and the gene product is probably a transmembrane copper transporter.[58] Because copper accumulation is slow, signs and symptoms do not appear before the age of 6 years. Most patients present with manifestations of organ damage between the ages of 8 and 20. Prolonged extrahepatic release of copper not bound to ceruloplasmin causes basal ganglion destruction, and sometimes cerebral cortex destruction. Prominent neuropsychiatric symptoms include irritability, aggression, disinhibition, and recklessness. Depressive features are also common. The severity of psychiatric symptoms correlates with the severity of neurological symptoms, especially dystonic and bulbar manifestations.[58]

Disorders of lipid metabolism

Intervention studies have shown that cholesterol reduction using diet, drugs, or surgery reduces the risk of developing or worsening coronary disease. In general, a 1 per cent reduction in low-density lipoprotein-cholesterol has been associated with roughly a 2 per cent reduction in disease end-points.[59] General agreement exists that eating less saturated fat and cholesterol, and adopting a diet and exercise habits to reduce obesity will benefit the health of most people. Exercise has a much greater effect in reducing triglyceride levels than in reducing low-density lipoprotein-cholesterol concentrations. Triglyceride levels are reduced after even a single exercise session. The efficacy of regular aerobic exercise in mild to moderate hypertriglyceridaemia has been repeatedly demonstrated.[60]

Hepatic encephalopathy

The pathogenesis of hepatic encephalopathy is related to widespread hepatic necrosis, commonly due to an acute viral infection, such as hepatitis B, or exposure to hepatotoxins. Common hepatotoxins that lead to liver failure include acetaminophen, isoniazid, halothane, valproic acid, mushroom toxin, and carbon tetrachloride. Hepatic encephalopathy that accompanies acute fulminant liver failure is frequently associated with cerebral oedema, which might be reversible and a treatable factor. Oedema is the leading cause of death in acute hepatic failure. It may respond to the administration of mannitol and measures to control agitation.[61] For patients with acute hepatic failure who have significant hepatic encephalopathy, liver transplantation increases survival from 20 to 80 per cent, making rapid and accurate diagnosis vital. Survival of

liver transplant patients with neuropsychiatric involvement is significantly lower if there is liver disease alone.[61] There is also an increased incidence of liver disease among patients with primary psychiatric disorders, including substance use disorders.[62] B virus carriers are almost three times more likely to have psychiatric disorders than comparison subjects.[63]

Endocrine disorders

Diabetes mellitus

Type I diabetes (insulin-dependent diabetes mellitus) occurs when the pancreas' ability to secrete insulin is clinically impaired. Hyperglycaemic symptoms emerge when 80 to 90 per cent of islet cells fail to produce insulin. Around 90 per cent of diabetics have type II diabetes (non-insulin-dependent diabetes mellitus). Type II diabetes is characterized by peripheral resistance to the action of insulin and decreased insulin secretion, in spite of the presence of elevated serum glucose levels. Patients with type II diabetes can often avoid or postpone the need for insulin treatment with appropriate diet and exercise. Both type I and II diabetes are associated with a genetic predisposition.[64]

The most frequent psychiatric disorders in patients with diabetes are anxiety and depressive disorders. Among general populations of diabetics, anxiety disorders occurred in up to 45 per cent and depressive disorders in up to 33 per cent.[65] Rosenthal *et al.*[66] in a 3-year prospective study of hospitalizations and mortality in older patients with diabetes, found that the combined presence of retinopathy and a high depression score on the Geriatric Depression scale had the strongest relationship with mortality. Patients with diabetes are twice as likely to experience depression as those without diabetes; this holds true for both type I and II diabetes.[66] Patients with schizophrenia are at increased risk for developing type II diabetes[67] is growing interest in the possibility that there are shared inherited risk factors for the two disorders,[68] though the evidence is weak and largely circumstantial. In addition, second generation antipsychotic medications, especially olanzapine, are associated with type II diabetes mellitus and abnormal glucose metabolism.

Diabetic patients who have psychiatric disorders can have less disease morbidity when their psychiatric disorders are appropriately treated, highlighting the importance of monitoring diabetic patients for psychiatric disorders and monitoring psychiatric patients for excess weight and diabetes.[69] Independent of the level of physical illness present in type I or II diabetes, the presence of anxiety and/or depression is important in determining the quality of a patient's life.[70] Treatment adherence problems complicate care, particularly in children and adolescents with type I diabetes. A great deal of patience, family support, and education is necessary to minimize passive and active non-compliance.

Hypothyroidism

Hypothyroidism is usually the result of primary failure or ablation of the thyroid gland, hypothalamic dysfunction, pituitary dysfunction, autoimmune thyroiditis, or lithium therapy. Clinical manifestations of hypothyroidism include fatigue, cold intolerance, lethargy, weakness, weight gain, constipation, menstrual irregularities, hair loss, slow reaction time, oedema, delayed reflexes, and bradycardia. Hypothyroidism occurs in as many as 10 per cent of patients taking lithium; lithium-induced hypothyroidism is more likely to occur in women.[70]

The association between clinical hypothyroidism and depression is well known. Gold *et al.*[71] found that 5 per cent of a series of 250 patients with major depressive syndromes had at least subclinical hypothyroidism. In many patients with hypothyroidism, the depression responds to thyroid hormone replacement alone,[72] but the response may take a long time. When that is the case, antidepressants are indicated and efficacious.[71]

Hyperthyroidism

The most frequent clinical manifestations of hyperthyroidism are nervousness, diaphoresis, hypersensitivity to heat, palpitations, fatigue, weight loss, tachycardia, dyspnoea, and weakness. The most common causes include Graves' disease, toxic adenoma, and toxic multinodular goitre. Less common causes include Hashimoto's thyroiditis, postpartum hyperthyroidism, and factitious hyperthyroid state.

As with hypothyroidism, depressive and anxiety syndromes are the most common psychiatric conditions seen among patients with hyperthyroid states; there is a three-fold increased risk for development of mood disorder following hospitalization with hyperthyroidism.[73,74] When patients have depressive or anxiety syndromes in the context of hyperthyroidism, and have no past histories of psychiatric disorders, the psychiatric symptoms resolve more than 90 per cent of the time when the hyperthyroidism resolves. This obviates the need for other psychiatric interventions unless antithyroid medication, radioactive iodine, or thyroid surgery has not been successful.[74] Anxiety symptoms will disappear in direct relation to the reduction of thyroid hormone levels. Depressive symptoms are not quite so linearly related and may resolve at a slower pace as thyroid hormone level normalize.

Hypoparathyroidism

Parathyroid hormone mobilizes calcium from bone, induces renal reabsorption of calcium, increases renal clearance of inorganic phosphate, and promotes intestinal reabsorption of calcium. Hypoparathyroidism would be expected to result in hypocalcaemia, which can cause delirium. Hypoparathyroidism can result from autoimmune destruction of the parathyroid glands, removal of the parathyroids, disruption of the glands' blood supply, tumour, or neck irradiation. Medical and neuropsychiatric symptoms and signs are related to the level of serum calcium and the rate at which hypocalcaemia develops. Faster the hypocalcaemia develops, the more likely delirium and other neuropsychiatric symptoms are to occur. The most frequent symptoms are caused by neuromuscular irritability and include paraesthesias, carpal pedal spasm, laryngospasm, blepharospasm, and bronchospasm.[75] Cardiovascular manifestations include prolonged Q–T interval, heart block, and congestive heart failure. The most common neuropsychiatric symptoms and signs are seizures, EEG abnormalities, increased intracranial pressure, disorientation, confusion, and extrapyramidal symptoms. The mainstays of treatment are calcium and vitamin D. Neuropsychiatric syndromes should resolve with the normalization of serum calcium.

Hyperparathyroidism

Typically, hypercalcaemia is discovered by routine laboratory testing in patients without obvious illness. Primary hyperparathyroidism is the most common cause of hypercalcaemia among adult patients; among hospitalized patients, malignancy is the most

common cause.[75] Reversible hyperparathyroidism and hypercalcaemia are also associated with lithium therapy.[76] In primary hyperparathyroidism, parathyroid hormone is secreted inappropriately, despite an elevation in the ionized calcium level. Signs and symptoms of hyperparathyroidism include nausea, vomiting, anorexia, constipation, proximal muscle weakness, polyuria, polydipsia, impaired renal function, hypertension, short Q–T interval, bradycardia, and a number of neuropsychiatric symptoms. The latter include lethargy, drowsiness, impaired concentration ability, and confusion. In severe cases, there may be stupor or coma, psychosis, and cognitive impairment are common in patients who have serum calcium levels above 16 mg/dl. Depressive symptoms, but not cognitive symptoms, tend to resolve with treatment.[77] Cognitive symptoms may improve, but residual symptoms usually remain.

Cushing's syndrome

Hypersecretion of cortisol by the adrenal gland can result in Cushing's syndrome. Cushing's syndrome can also be due to exogenous ACTH or glucocorticoid administration, or endogenous hyperproduction of these hormones. Because a physiological release of cortisol occurs during periods of stress or duress, it is common to see elevations of serum cortisol during the courses of many psychiatric disorders, including major depressive disorder, alcoholism, anorexia nervosa, panic disorder, and psychoactive substance-withdrawal syndromes. The more common clinical signs and symptoms of Cushing's syndrome, whether endogenous or exogenous, include fat redistribution, menstrual irregularities, dysphoria, thin skin, moon facies, increased appetite, sleep disturbances, hypertension, hypercholesterolaemia, hypertriglyceridaemia, poor concentration, impaired memory, euphoria, glucose intolerance, striae, and hirsutism.

At least half of all patients with Cushing's syndrome will experience depressive or manic symptoms[78]; the symptoms will be moderate to severe in half of these patients. Many will also experience psychotic symptoms. Symptoms are dose-related when due to exogenous steroids. Depression or mania due to Cushing's syndrome will eventually remit when the hypercortisolaemia is corrected, the return to euthymia is usually gradual. When depression or mania is slow to remit, treatment with antidepressants or mood stabilizers is warranted.

Addison's disease

Addison's disease is the result of an autoimmune process that destroys the adrenal glands; it is the most common cause of primary adrenal insufficiency in the industrialized world, accounting for about 65 per cent of cases.[70] Both glucocorticoid and mineralocorticoid secretion are diminished in this condition. Clinical manifestations of adrenal insufficiency include weight loss, fatigue, vomiting, diarrhoea, anorexia, and salt-craving. Patients with Addison's disease will require lifelong replacement of both glucocorticoids and mineralocorticoids. Patients may be misdiagnosed with major depressive disorder, personality disorder, dementia, or somatoform disorders. It is not uncommon for the diagnosis to be delayed for many months; Addison's disease is a disorder to be continually mindful about in a patient with treatment-resistant depression.

Hyperprolactinaemia

Prolactin is synthesized in the pituitary gland; its secretion is increased during pregnancy, enhancing breast development. Prolactin secretion is inhibited by glucocorticoids and thyroid hormone, it is predominantly under the inhibitory control of dopamine.[70] Dopamine is, in fact, prolactin inhibiting factor. This is why dopamine blocking medications such as neuroleptics can cause hyperprolactinaemia. Symptoms of hyperprolactinaemia include breast development and lactation. These bothersome side effects of neuroleptics, particularly in men, can be modulated by changing to a more favourable neuroleptic medication, such as a second generation antipsychotic.

Hypopituitarism

Hypopituitarism occurs when multiple pituitary hormones exhibit decreased secretion. It can be due to either gland destruction or inadequate stimulation by factors that regulate pituitary functioning. Common causes of hypopituitarism are pituitary adenomas, hypothalamic tumours, metastatic carcinoma (especially breast and bronchus), and cerebral trauma and haemorrhage.[70] Other causes include vascular disorders, immunological conditions, and a variety of congenital anomalies. Clinical manifestations depend on which hormones are deficient; signs and symptoms are those of the individual deficiency states. Other signs that may be present when there is hypopituitarism include headache, visual loss, and radiographically discovered sella enlargement. Treatment of patients with hypopituitarism involves hormone replacement therapy and surgery when accessible lesions are present.[70]

Autoimmune disorders

Systemic lupus erythematosus

Systemic lupus erythematosus is characterized by the production of autoantibodies that injure tissue in several organ systems, most frequently the skin, central nervous system, kidney, and lungs. In addition, non-specific symptoms occur in a large majority of patients, including fatigue, fever, weight loss, arthralgias, and myalgias.[79] The most common presenting symptoms are fever, arthralgias, butterfly rash, photosensitivity, Raynaud's phenomenon, and mucous ulcers. The laboratory diagnostic hallmark of systemic lupus erythematosus is the production of high-titre autoantibodies directed against a variety of cell nucleus components (antinuclear antibodies). Systemic lupus erythematosus has no known cure, so treatment is based on symptom relief, suppression of inflammation, and preventing future pathology.

The most common psychiatric presentations of active systemic lupus erythematosus are psychosis, delirium, seizures, and cognitive dysfunction.[80] Autoantibodies to neuronal membranes, which interfere with the ability of neurones to respond to stimuli, may account for most neuropsychiatric deficits and symptoms,[81] though CNS vasculitis during the course of this disease may also play a role in some patients. Symptoms in lupus patients are likely to be related to unique sets of autoantigens,[82] and are related to antibodies against N-methyl-D-aspartate (NMDA) receptors.[83] Antibodies may be partially responsible for depression, short-term memory problems, and new learning difficulties in lupus patients. Significantly associated with declining cognitive function are consistently positive antiphospholipid antibodies, consistent steroid use, diabetes, and higher depression scores.[83] Mood syndromes are probably the most common psychiatric presentation of patients with systemic lupus erythematosus,[84] but mood change is not

always due to involvement of the central nervous system. Patients are frequently treated with steroids, which raise the possibility of steroid-induced neuropsychiatric syndromes. However, in many cases, addition of steroids to treatment may improve psychiatric syndromes.

Renal disorders

Acute renal failure

Acute renal failure is an abrupt decrease in renal function sufficient to result in azotaemia—retention of nitrogenous waste in the body.[85] Acute renal failure can result from a decrease of renal blood flow (prerenal azotaemia), intrinsic renal disease (renal azotaemia), or obstruction of urine flow (postrenal azotaemia). Prerenal azotaemia can be caused by renal arterial occlusion or a decrease in the effective blood volume (e.g. haemorrhage, congestive heart failure, diarrhoea). Intrinsic renal azotaemia is most commonly caused by acute tubular necrosis due to an acute ischaemic or nephrotoxic insult. Azotaemia is due to obstruction of the urine collecting system; this may occur when there is bladder outlet obstruction or ureteral obstruction.

Medical complications of acute renal failure include hyperkalaemia, hyperuricaemia, arrhythmias, anaemia, coagulopathies, vomiting, nausea, and urinary tract infections. Metabolic perturbations can lead to delirium. Neuropsychiatric manifestations include somnolence, asterixis (flapping tremor), neuromuscular irritability, and seizures. Mental status abnormalities in acute (but not chronic) renal failure begin to occur for most adults when the serum creatinine level acutely rises to about 4.0 mg/dl. In oliguric renal failure, serum blood urea nitrogen levels can be expected to rise by about 10 to 20 mg/dl per day. Serum creatinine levels can be expected to rise by about 1 mg/dl per day. Neuropsychiatric complications of acute renal failure are best treated by correcting the underlying cause of the renal failure. Dialysis may be used to manage acute manifestations. While awaiting reversal of neuropsychiatric manifestations, symptomatic management with antiseizure medications and neuroleptics may be necessary.

Chronic renal failure and end stage renal disease

Chronic renal failure is a progressive and irreversible loss of renal function.[86] The most common aetiologies of renal insufficiency ultimately leading to end stage renal disease are diabetes, hypertension, and glomerulonephritis. Loss of up to 75 per cent of glomerular filtration rate does not usually result in pronounced clinical symptoms, as the remaining glomeruli adapt with hyperfiltration. Serum creatinine is a sensitive indicator of early, subclinical, chronic renal failure. For example, the doubling of serum creatinine from 0.7 to 1.4 mg/dl signifies a loss of approximately 50 per cent of glomerular filtration rate, emphasizing the importance of early detection and prevention.

Patients with chronic renal failure usually become symptomatic when glomerular filtration rate is less than 10 ml/min. Uraemia affects every organ system, including the central nervous system. Neuropsychiatric manifestations of chronic renal failure include irritability, insomnia, lethargy, anorexia, seizures, and restless legs syndrome.[86] In contrast to acute renal failure—where neuropsychiatric signs and symptoms may appear with a creatinine level as low as 4 mg/dl—in chronic renal failure, patients may have a normal mental status examination with a serum creatinine level as high as 9 to 10 mg/dl. Symptomatic treatments with low-dose neuroleptics, antiseizure medications, or benzodiazepines are sometimes necessary in chronic renal failure.

Kimmel *et al.*[86] studied the prevalence of hospitalizations for psychiatric illness in patients with end stage renal disease and compared that rate with four other chronic medical conditions (diabetes, ischaemic heart disease, cerebrovascular disease, and peptic ulcer disease). Hospitalizations for mental disorders were 1.5 to 3.0 times higher in these patients than in patients with the four other chronic diseases. Dementias and depression were the most common reasons for hospitalization. With end stage renal disease are almost twice as likely to die by suicides as the general populations, after controlling for other potential demographic and clinical contributors to suicide risk.[87]

Sexual dysfunction is very common in patients with end stage renal disease. Abrams *et al.*[88] found that 75 per cent of his sample of men with this disease reported a decrease in frequency of sexual intercourse of at least 50 per cent. Disruptions in sexual function, which may be physiological (e.g. vascular complications of diabetes, fatigue following dialysis treatments) or psychological or both, account for at least a portion of the dysphoria experienced by patients with end stage renal disease.

The definitive treatments for most patients with chronic renal failure are transplantation or haemodialysis. In general, transplantation is encouraged because of a better quality of life and a greater chance for rehabilitation and symptom resolution. Researchers in three separate prospective studies found that patients who received renal transplants experienced better physical and psychological outcomes than patients who remained on dialysis.[89] Neuropsychiatric signs and symptoms resolve much more completely with transplantation than with haemodialysis. Psychiatric aspects of organ transplantation are discussed later. The psychiatric aspects of haemodialysis are discussed next.

Haemodialysis

The average patient on haemodialysis requires 3.5 h of dialysis three times per week to achieve adequate creatinine clearance.[86] Haemodialysis has enabled the survival of countless thousands of patients with chronic renal failure and provides a temporary management tool for patients on transplantation waiting lists. However, it is not a benign procedure, and has a number of potential neuropsychiatric complications. Patients on haemodialysis are at high risk for developing volume overload, pulmonary oedema, hyperkalaemia, hyperphosphataemia, and metabolic bone disease if compliance with restricted diet and fluid intake is not optimal. Patient adherence to these diet and fluid-intake protocols are used as one of the criteria for making decisions about appropriateness for transplantation. Psychiatric reasons for non-adherence should be addressed and are usually reversible, with the exception of personality disorders. These include mood disorders, phobias, panic disorder, substance-related disorders, adjustment disorder, and cognitive disorders.

Haematological disorders

Anaemia due to vitamin deficiency

Both folic acid and cobalamin (vitamin B12) are necessary for the production of DNA; in their absence, the nucleus of the cell cannot

undergo normal mitosis. The main cause of folic acid deficiency is dietary insufficiency.(90) This commonly occurs in severe alcoholics. The main cause of cobalamin deficiency is malabsorption. The major clinical manifestations of folate or vitamin B12 deficiency include fatigue, pallor, and for cobalamin deficiency, neuropsychiatric manifestations. The latter include loss of proprioception in the lower extremities, loss of vibratory perception, anosmial, forgetfulness, and even dementia. Diagnosis is based on measurement of serum levels of vitamin B12 and folate. Treatment is replacement of folate (1 mg/day or improved diet) and vitamin B12 (administered parenterally).

In cobalamin deficiency, neuropsychiatric findings can occur even when megaloblasts and anaemia are absent.(91) Patients with cobalamin levels between 100 and 200 pg/ml, and especially those with levels less than 100 pg/ml, may have cobalamin-reversible neuropsychiatric deficits.(138) Even dementia due to chronic cobalamin deficiency may be partially reversible when diagnosed and aggressively treated. It is also associated with a higher risk of depression, especially in older adults; anaemia severity and depression severity scores are highly correlated.(92) Treatment of the anaemia may result in resolution of depressive symptoms; in some cases depressive symptoms are slower to resolve and may require extended treatment with antidepressant medication, particularly for patients predisposed to mood disorders.

Iron-deficiency anaemia

When erythrocyte-cytoplasm production is abnormally low due to the reduced production or availability of one of the three components of haemoglobin (iron, globin, or haem), the ratio of cytoplasm to the contents of the rest of the cell declines. This results in a microcytic anaemia. In most cases, microcytic anaemia is due to iron-deficient haemopoiesis.(92) Deficiency is usually due to an iron-poor diet or defective iron utilization by the body. Clinical manifestations of iron-deficiency anaemia are related to the severity of the iron deficiency and include severe fatigue, pallor, changes in nail curvature ('spoon nails'), and, at times, pica and cheilosis at the corners of the mouth. A diminished haematocrit and mean corpuscular volume, and low serum iron are the cornerstones of the diagnosis. The severe fatigue associated with iron deficiency can be misdiagnosed as major depressive disorder; treatment with an antidepressant will not help unless there is also clinical major depression apart from the iron deficiency. Iron replenishment to correct the underlying deficiency is the specific. With iron replenishment, the haemoglobin should correct to normal, and symptoms should resolve, within 4 to 6 weeks.

Conclusions

There is increasing recognition that patients with psychiatric signs and symptoms frequently have associated medical disorders. Interactions between the disorders can be quite complex and may involve neuropsychiatric manifestations of medical illness, medical effects of psychiatric treatments, psychiatric effects of medical treatment, increased medical illness related to factors inherent in the psychiatric condition, and maladaptive personality styles or disorders that affect outcomes of medical and psychiatric illness. Psychiatric disorders in patients with other medical illnesses increase mortality, morbidity, health care costs, and decrease adherence to medical therapies. Appropriate and cost-effective management must integrate knowledge from psychiatry and primary care to assure maximal therapeutic gains for the patient. Those involved in consultation to primary care providers must understand both the basic pathophysiology and clinical characteristics of the medical disorders and their treatments, as well as possess the psychiatric knowledge necessary to ensure that the patient with medical illness is adequately managed.

Further information

Robinson, M.J. and Owen, J.A. (2005). Psychopharmacology. In *Textbook of psychosomatic medicine* (ed. J.L. Levenson), pp. 871–3. American Psychiatric Publishing, Inc., Washington, DC.

Saha, S., Chant, D., and McGrath, J. (2007). A systematic review of mortality in schizophrenia. *Archives of General Psychiatry*, **64**, 1123–31.

Schmitz, N., Thefeld, W., and Kruse, J. (2006). Mental disorders and hypertension: factors associated with awareness and treatment of hypertension in the general population of Germany. *Psychosomatic Medicine*, **68**, 246–52.

References

1. Chapman, D.P., Perry, G.S., and Strine, T.W. (2005). The vital link between chronic disease and depressive disorder. *Preventing Chronic Disease*, **2**, 1–12.
2. Barsky, A.J., Orav, J., and Bates, D.W. (2005). Somatization increases medical utilization and costs independent of psychiatric and medical comorbidity. *Archives of General Psychiatry*, **62**, 903–10.
3. Brod, J. (1970). Circulatory changes underlying blood pressure elevation during acute emotional stress (mental arithmetic) in normotensive and hypertensive subjects. *Clinical Scientist*, **18**, 269–78.
4. Torian, L., Davidson, E., Fulop, G., *et al.* (1992). The effect of dementia on acute care in a geriatric medical unit. *International Psychogeriatrics*, **4**, 231–9.
5. Thomas, S.A., Friedmann, E., and Kelley, F.J. (2001). Living with an implantable cardioverter-defibrillator: a review of the current literature related to psychosocial factors. *The Journal of Cardiovascular Nursing*, **12**, 156–63.
6. Lown, B., DeSilva, R.A., and Reich, P. (1980). Psychophysiologic factors in sudden cardiac death. *The American Journal of Psychiatry*, **137**, 1325–35.
7. Tennant, C. (1999). Life stress, social support and coronary heart disease. *The Australian and New Zealand Journal of Psychiatry*, **33**, 636.
8. Frasure-Smith, N., Lesperance, F., and Talajic, M. (1995). Depression and 18-month prognosis after myocardial infarction. *Circulation*, **91**, 999–1005.
9. Benson, H., Alexander, S., and Feldman, C.L. (1975). Decreased premature ventricular contractions through the use of relaxation response in patients with stable ischaemic heart disease. *Lancet*, **2**, 380–2.
10. Grippo, A.J., Santos, C.M., Johnson, R.F., *et al.* (2004). Increased susceptibility to ventricular arrhythmias in a rodent model of experimental depression. *American Journal of Physiology—Heart and Circulatory Physiology*, **286**, H619–26.
11. Musselman, D.L., Evans, D.L., and Nemeroff, C.B. (1998). The relationship of depression to cardiovascular disease: epidemiology, biology, and treatment. *Archives of General Psychiatry*, **55**, 580–92.
12. Brod, J. (1970). Circulatory changes underlying blood pressure elevation during acute emotional stress (mental arithmetic) in normotensive and hypertensive subjects. *Clinical Scientist*, **18**, 269–78.
13. Harburg, E., Erfurt, J.C., and Hauenstein, L.S. (1973). Socio-ecological stress, suppressed hostility, skin color, and black-white male blood pressure: Detroit. *Psychosomatic Medicine*, **35**, 276–96.
14. Bankier, B. and Littman, A.B. (2002). Psychiatric disorders and coronary heart disease in women—a still neglected topic: review of the literature from 1971 to 2000. *Psychotherapy and Psychosomatics*, **71**, 133–40.

15. Rozanski, A., Bairey, C.N., and Krantz, D.S. (1988). Mental stress and the induction of silent myocardial ischemia in patients with coronary artery disease. *The New England Journal of Medicine*, **318**, 1005–11.
16. Ruberman, W., Weinblatt, A.B., Goldberg, J.D., *et al.* (1984). Psychosocial influences on mortality after myocardial infarction. *The New England Journal of Medicine*, **311**, 552–9.
17. Dickens, C.M., McGowan, L., Percival, C., *et al.* (2004). Lack of a close confidant, but not depression, predicts further cardiac events after myocardial infarction. *Heart*, **90**, 518–22.
18. Mayou, R.A., Gill, D., Thompson, D.R., *et al.* (2000). Depression and anxiety as predictors of outcome after myocardial infarction. *Psychosomatic Medicine*, **62**, 212–19.
19. Lesperance, F., Frasure-Smith, N., and Talajic, M. (1996). Major depression before and after myocardial infarction: its nature and consequences. *Psychosomatic Medicine*, **58**, 99–110.
20. Carney, R.M., Rich, M.W., and Freedland, K.E. (1988). Major depressive disorder predicts cardiac events in patients with coronary artery disease. *Psychosomatic Medicine*, **50**, 627–33.
21. Plotsky, P.M., Owens, M.J., and Nemeroff, C.B. (1998). Psychoneuroendocrinology of depression the Hypothalamic-Pituitary-Adrenal Axis. *The Psychiatric Clinics of North America*, **21**, 293–307.
22. Musselman, D.L., Evans, D.L., and Nemeroff, C.B. (1998). The relationship of depression to cardiovascular disease. *Archives of General Psychiatry*, **55**, 580–92.
23. Langosch, W., Seer, P., Brodner, G., *et al.* (1982). Behavior therapy with coronary heart disease patients: results of a comparative study. *Journal of Psychosomatic Research*, **26**, 475–84.
24. Preckel, D., von Kanel, R., Kudielka, B.M., *et al.* (2005). Over commitment to work is associated with vital exhaustion. *International Archives of Occupational and Environmental Health*, **78**, 117–22.
25. Friedman, M. and Thoresen, C.E. (1986). Alteration of type A behavior and its effect on cardiac recurrences in post-myocardial infarction patients: summary results of the recurrent coronary prevention project. *American Heart Journal*, **112**, 653–65.
26. Sher, L. (2005). Type D personality: the heart, stress, and cortisol. *Quarterly Journal of Medicine*, **98**, 323–9.
27. Denollet, J., Pedersen, S.S., Vrints, C.J., *et al.* (2006). Usefulness of type D personality in predicting five-year cardiac events above and beyond concurrent symptoms of stress in patients with coronary heart disease. *The American Journal of Cardiology*, **97**, 970–3.
28. Katon, W.J., Richardson, L., Lozano, P., *et al.* (2004). The relationship of asthma and anxiety disorders. *Psychosomatic Medicine*, **66**, 349–55.
29. Ten Brinke, A., Sterk, P.J., Masclee, A.A.M., *et al.* (2005). Risk factors of frequent exacerbations in difficult-to-treat asthma. *The European Respiratory Journal*, **26**, 812–18.
30. Greenberg, D.B., Halperin, P., Kradin, R.L., *et al.* (2000). Internal medicine and medical subspecialties. In *American psychiatric publishing textbook of consultation–liaison psychiatry* (2nd edn) (eds. M. Wise and J.R. Rundell), pp. 548–608. American Psychiatric Publishing, Washington, DC.
31. Norwood, R. (2006). Prevalence and impacts of depression in chronic obstructive pulmonary disease. *Current Opinion in Pulmonary Medicine*, **12**, 113–17.
32. Porzelius, J., Vest, M., and Nochomovitz, M. (1992). Respiratory function, cognitions, and panic in chronic obstructive pulmonary patients. *Behavioural Research and Therapy*, **30**, 75–7.
33. Strudsholm, U., Johannessen, L., Foldager, L., *et al.* (2005). Increased risk for pulmonary embolism in patients with bipolar disorder. *Bipolar Disorders*, **7**, 77–81.
34. Joukamaa, M., Heliovaara, M., Knekt, P., *et al.* (2006). Schizophrenia, neuroleptic medication, and mortality. *The British Journal of Psychiatry*, **188**, 122–7.
35. Clouse, R.E. and Lustman, P.J. (1989). Value of recent psychological symptoms in identifying patients with esophageal contraction abnormalities. *Psychosomatic Medicine*, **51**, 570–6.
36. Van Handel, D. and Fass, R. (2005). The pathophysiology of non-cardiac chest pain. *Journal of Gastroenterology and Hepatology*, **20**, S6–11.
37. Clouse, R.E. and Lustman, P.J. (1983). Psychiatric illness and contraction abnormalities of the esophagus. *The New England Journal of Medicine*, **309**, 1337–42.
38. Castell, D.O. and Richter, J.E. (1987). Edrophonium testing for esophageal pain: concurrence and discord. *Digestive Disease Scientist*, **32**, 897–9.
39. Cybulska, E.M. (1997). Globus hystericus: a somatic symptom of depression? *Psychosomatic Medicine*, **59**, 67–9.
40. Drossman, D.A., Sandler, R.S., and McKee, D.C. (1982). Bowel patterns among subjects not seeking health care. *Gastroenterology*, **83**, 529–34.
41. Bhatia, V. and Tandon, R.K. (2005). Stress and the gastrointestinal tract. *Journal of Gastroenterology and Hepatology*, **20**, 332–44.
42. Clouse, R.E., Richter, J.E., Heading, R.C., *et al.* Functional esophageal disorders. *Gut*, **45**, 31–6.
43. Clouse, R.E. and Lustman, P.J. (2005). Use of psychopharmacological agents for functional gastrointestinal disorders. *Gut*, **54**, 1332–41.
44. Walker, E.A., Gelfand, M.D., Gelfand, A.N., *et al.* (1996). The relationship of current psychiatric disorders to functional disability and distress in patients with inflammatory bowel disease. *General Hospital Psychiatry*, **18**, 220–9.
45. Bennett, P. and Wilkinson, S. (1985). A comparison of psychological and medical treatment of the irritable bowel syndrome. *The British Journal of Clinical Psychology*, **24**, 215–16.
46. Maunder, R. (2005). Evidence that stress contributes to inflammatory bowel disease: evaluation, synthesis, and future directions. *Inflammatory Bowel Diseases*, **11**, 600–8.
47. Mawdsley, J.E. and Rampton, D.S. (2005). Psychological stressing IBD: new insights into pathogenic and therapeutic implications. *Gut*, **54**, 1481–91.
48. Whorell, P.J., Prior, A., and Faragher, E.B. (1984). Controlled trial of hypnotherapy in the treatment of severe refractory irritable bowel syndrome. *Lancet*, **2**, 1232–4.
49. Cremonini, F., Wise, J., Moayyedi, P., *et al.* (2005). Diagnostic and therapeutic use of proton pump inhibitors in non-cardiac chest pain: a meta-analysis. *The American Journal of Gastroenterology*, **100**, 1226–9.
50. Littman, A.B. and Ketterer, M.W. (2000). Behavioural medicine. In *American psychiatric publishing textbook of consultation–liaison psychiatry* (2nd edn) (eds. M. Wise and J.R. Rundell), pp. 1080–95. American Psychiatric Publishing, Washington, DC.
51. Magni, G. (1987). On the relationship between chronic pain and depression when there is no organic lesion. *Pain*, **31**, 1–21.
52. Whitehead, W.E. (1992). Behavioural medicine approaches to gastrointestinal disorders. *Journal of Consulting and Clinical Psychology*, **60**, 605–12.
53. Ang, T.L., Fock, K.M., Ng, T.M., *et al.* (2005). A comparison of the clinical, demographic and psychiatric profiles among patients with erosive and non-erosive reflux disease in a multi-ethnic Asian country. *World Journal of Gastroenterology*, **11**, 3558–61.
54. Elfhag, K. and Rossner, S. (2005). Who succeeds in maintaining weight loss? A conceptual review of factors associated with weight loss maintenance and weight regain. *Obesity Reviews*, **6**, 67–86.
55. Morriss, R. and Mohammed, F.A. (2005). Metabolism, lifestyle and bipolar disorder. *Journal of Psychopharmacology*, **19**, 94–101.
56. Casey, D.E. (2005). Metabolic issues and cardiovascular disease in patients with psychiatric disorders. *The American Journal of Medicine*, **118**(Suppl. 2), 15S–22S.
57. Prashanth, L.K., Sinha, T.S., Ravishankar, S., *et al.* (2005). Prognostic factors in patients presenting with severe neurological forms of Wilson's disease. *Quarterly Journal of Medicine*, **98**, 557–63.
58. Dening, D.C. and Berrios, G.E. (1989). Wilson's disease: psychiatric symptoms in 195 cases. *Archives of General Psychiatry*, **46**, 1126–34.
59. National Cholesterol Education Program. (1993). Summary of the second report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation and treatment of high blood

cholesterol in adults. (Adult Treatment Panel 11). *The Journal of the American Medical Association*, **264**, 3015–23.
60. Freedman, D.S., Byers, T., Barrett, D.H., *et al.* (1995). Plasma lipid levels and psychological characteristic in men. *American Journal of Epidemiology*, **141**, 507–17.
61. Medici, V., Mirante, V.G., Fassati, L.R., *et al.* (2005). Liver transplantation for Wilson's disease: the burden of neurological and psychiatric disorders. *Liver Transplantation*, **11**, 1056–63.
62. Crone, C.C., Gabriel, G.M., and DiMartini, A. (2006). An overview of psychiatric issues in liver disease for the consultation-liaison psychiatrist. *Psychosomatics*, **47**, 188–205.
63. Atesci, F.C., Cetin, B.C., Oguzhanoglu, N.K., *et al.* (2005). Psychiatric disorders and functioning in Hepatitis B virus carriers. *Psychosomatics*, **46**, 142–7.
64. Lustman, P.J., Griffith, L.S., Clouse, R.E., *et al.* (1986). Psychiatric illness in diabetes mellitus: relationship to symptoms and glucose control. *The Journal of Nervous and Mental Disease*, **174**, 736–42.
65. Kathol, R. (2000). In *American psychiatric publishing textbook of consultation–liaison psychiatry* (2nd edn) (eds. M. Wise and J.R. Rundell), pp. 579–84. American Psychiatric Publishing, Washington, DC.
66. Rosenthal, M.J., Morley, J.E., Fajardo, M., *et al.* (1998). Hospitalization and mortality of diabetes in older adults. *Diabetes Care*, **21**, 231–5.
67. Chafetz, L., White, M.C., Collins-Bride, G., *et al.* (2005). The poor general health of the severely mentally ill: impact of schizophrenic diagnosis. *Community Mental Health Journal*, **41**, 169–84.
68. Gough, S.C.L. and O'Donovan, M.C. (2005). Clustering of metabolic comorbidity in schizophrenia: a genetic contribution? *Journal of Psychopharmacology*, **19**, 47–55.
69. McIntyre, R.S., Mancini, D.A., Pearce, M.M., *et al.* (2005). Mood and psychotic disorders and type 2 diabetes: a metabolic trial. *Canadian Journal of Diabetes*, **29**, 122–32.
70. Goldman, M.B. (2002). Neuropsychiatric features of endocrine disorders. In *The American psychiatric publishing textbook of neuropsychiatry* (4th edn) (eds. S. Yudofsky and R. Hales), pp. 519–40. American Psychiatric Publishing, Washington, DC.
71. Gold, M.S., Pottash, A.L.C., and Extein, I. (1981). Hypothyroidism and depression. *The Journal of the American Medical Association*, **245**, 1919–22.
72. Rouchell, A.M., Pounds, R., and Tierney, J.G. (2000). Depression. In *American psychiatric publishing textbook of consultation–liaison psychiatry* (2nd edn) (eds. M. Wise and J.R. Rundell), pp. 310–45. American Psychiatric Publishing, Washington, DC.
73. Williams, R.H. (1946). Thiouracil treatment of thyrotoxicosis. *The Journal of Clinical Endocrinology and Metabolism*, **6**, 1–22.
74. Thomsen, A.F., Kvist, T.K., Andersen, P.K., *et al.* (2005). Increased risk of affective disorder following hospitalization with hyperthyroidism—a register based study. *European Journal of Endocrinology*, **152**, 535–43.
75. Brown, G.G., Preisman, R.C., and Kleerekoper, M. (1987). Neurobehavioural symptoms in mild primary hyperparathyroidism: related to hypercalcemia but not improved by parathyroidectomy. *Henry Ford Hospital Medical Journal*, **35**, 211–15.
76. Khandwala, H.M. and Van Uum, S. (2006). Reversible hypercalcemia and hyperparathyroidism associated with lithium therapy: case report and review of the literature. *Endocrine Practice*, **12**, 54–8.
77. Allerheiligen, D.A., Schoeber, J., Houston, R.E., *et al.* (1998). Hyperparathyroidism. *American Family Physician*, **57**, 1795–802.
78. Thompson, J.M., Gallagher, P., Hughes, J.H., *et al.* (2005). Neurocognitive impairment in euthymic patients with bipolar affective disorder. *The British Journal of Psychiatry*, **186**, 32–40.
79. Yagniak, P.M. and Cohen, M.M. (1988). Systemic lupus erythematosus nervous system involvement. In *Diagnosis and management of rheumatic diseases* (2nd edn) (ed. W. Katz), pp. 220–3. J.B. Lippincott, Philadelphia, PA.
80. Greenberg, D.B. (2000). Systemic lupus erythematosus. In *American psychiatric publishing textbook of consultation–liaison psychiatry* (2nd edn) (eds. M. Wise and J.R. Rundell), p. 585. American Psychiatric Publishing, Washington, DC.
81. Baker, M. (1973). Psychopathology in SLE: psychiatric observations. *Seminars in Arthritis and Rheumatology*, **3**, 95–110.
82. Margutti, P., Sorice, M., Conti, F., *et al.* (2005). Screening of an endothelial cDNA library identifies the C-terminal region of Nedd5 as a novel autoantigen in systemic lupus erythematosus with psychiatric manifestations. *Arthritis Research & Therapy*, **7**, R896–903.
83. Omdal, R., Brokstad, K., Waterloo, K., *et al.* (2005). Neuropsychiatric disturbances in systemic lupus erythematosus are associated with antibodies against NMDA receptors. *European Journal of Neurology*, **12**, 392–400.
84. McLaurin, E.Y., Holliday, S.L., Williams, P., *et al.* (2005). Predictors of cognitive dysfunction in patients with systemic lupus erythematosus. *Neurology*, **64**, 297–303.
85. Cohen, L.M. (2000). Renal disease. In *American psychiatric publishing textbook of consultation–liaison psychiatry* (2nd edn) (eds. M. Wise and J.R. Rundell), pp. 573–8. American Psychiatric Publishing, Washington, DC.
86. Kimmel, P.L., Thamer, M., Richard, C.M., *et al.* (1998). Psychiatric illness in patients with end-stage renal disease. *The American Journal of Medicine*, **105**, 214–21.
87. Kurella, M., Kimmel, P.L., Young, B.S., *et al.* (2005). Suicide in the United States end-stage renal disease program. *Journal of the American Society of Nephrology*, **16**, 774–81.
88. Abrams, H.S., Hester, L.R., and Sheridan, W.F. (1975). Sexual functioning in patients with chronic renal failure. *The Journal of Nervous and Mental Diseases*, **160**, 220–6.
89. Russell, J.D., Beecroft, M.L., Ludlow, D., *et al.* (1992). The quality of life in renal transplantation—a prospective study. *Transplantation*, **54**, 656–60.
90. Stabler, S.P. and Allen, R.H. (1990). Clinical spectrum and diagnosis of cobalamin deficiency. *Blood*, **76**, 871–81.
91. Greenberg, D.B. (2000). Cobalamin deficiency. In *American psychiatric publishing textbook of consultation–liaison psychiatry* (2nd edn) (eds. M. Wise and J.R. Rundell), p. 586. American Psychiatric Publishing, Washington, DC.
92. Onder, G., Penninx, B.W., Cesari, M., *et al.* (2005). Anemia is correlated with depression in older adults: results from the InCHIANTI study. *Gerontology*, **60**, 1168–72.
93. Badminton, M.N. and Elder, G.H. (2005). *Journal of Inherited Metabolic Disease*, **28**, 277–86.

5.3.5 Psychiatric aspects of infections

José Luis Ayuso-Mateos

Neuropsychiatric disturbances stemming from infectious diseases are widespread in both the industrialized world and developing countries. Such neuropsychiatric syndromes are not necessarily the result of infectious processes directly involving the central nervous system, they may also be complications of systemic infections. There are many microbial, viral, and parasitic agents, as well as other types of infectious substances, which can affect the central nervous system, leading to the appearance of neurological and psychiatric symptoms that may cause suffering to the patient, and even be disabling.

When considering the psychiatric manifestations of infectious illness, it is important to consider clinical manifestations derived

from a possible systemic infection, which can be less obvious than a direct involvement of the central nervous system. Acute organic reactions may accompany many systemic infections, especially at the extremes of life. A clear example is the delirium that frequently occurs with pneumonia in the elderly. In these clinical syndromes, several factors could be responsible for the alterations in cerebral metabolism. The mere fact of having a fever could be involved. Cerebral anoxia often appears to be responsible, or the influence of toxins derived from the infecting micro-organism. More complex metabolic disturbances or the accumulation of toxic intermediate products can also be implicated.

Likewise, infections that course as chronic or subacute illnesses are frequently accompanied by the onset of depressive syndromes. One of the factors implied in clinical depression that occurs within the context of systemic infectious illnesses (e.g. tuberculosis and infectious mononucleosis), is a sense of physical vulnerability, possibly heightened by a loss of strength and negative changes in the patient's appearance. Patients are often afraid of losing their earning capacity or even their jobs, as well as other social and occupational problems associated with the illness.

Another very important factor, above all with the human immunodeficiency virus (**HIV**) and other sexually transmitted disease (**STD**), is the social stigma that these patients may suffer.[1] Sexually transmitted disease infection implies sexual activity that historically carries connotations of illicit, casual, sexual encounters, and acquiring an STD is frequently associated with embarrassment and social stigma.

In addition to the disease itself, the medications commonly used to treat infectious illnesses can have side-effects that alter patients' behaviour, as well as their cognitive and affective functioning (Table 5.3.5.1).

In this chapter we consider infections of clinical interest in the practice of psychiatry. These conditions will be dealt with briefly, and textbooks of general medicine should be consulted for further details. Prion diseases and chronic fatigue syndromes, which are also related to the subject of the present chapter, are discussed in Chapters 4.1.4 and 5.2.7, respectively.

Table 5.3.5.1 Neuropsychiatric adverse effects of drugs frequently used in the treatment of infectious diseases

Drug	Adverse effect
Aciclovir	Headache, somnolence, tremor, confusion, lethargy, seizures, agitation, major depression with psychotic symptoms
Amphotericin B	Delirium
Chloramphenicol	Memory impairment, confusion, depersonalization, hallucinations
Cycloserine	Depression, anxiety, confusion, hallucinations, paranoia, agoraphobia
Didanosine (ddI)	Headache, asthenia, polyneuropathy
Efavirenz	Dizziness, headache, insomnia, inappropriate behaviour, depression, concentration impairment, agitation, abnormal dreaming, and somnolence
Foscarnet	Asthenia
Gentamicin	Confusion, hallucinations
Interferon	Depression, anxiety, irritability, delirium
Isoniazid (INH)	Headaches, vertigo, hyper-reflexia, neuritis, convulsions, ataxia, toxic, encephalopathy, confusion, psychosis, antidepressant effect
Ketoconazole	Somnolence, delirium
Para-aminosalicylate (PAS)	Toxic psychosis
Penicillin G (procaine)	Hallucinations, seizures, agitation, confusion
Rifampicin (rifampin)	Myopathy, headache if hypersensitivity
Streptomycin	Toxic effects on cranial nerve VIII (vestibular), vertigo, nystagmus, ataxia, neuromuscular junction blockade
Sulphonamide	Anxiety, depression, insomnia, hallucinations
Trimethoprim-sulphamethoxazole	Vertigo and confusion
Zalcitabine (ddC)	Polyneuropathy
Zidovudine (AZT)	Headache, myalgia, insomnia, asthenia, somnolence, anxiety, depression, mania, restlessness

HIV infection

Patients infected with HIV are at an increased risk for a variety of mental disorders. Those encountered most frequently in psychiatric practice are discussed below. HIV dementia is discussed in Chapter 4.1.9.

Nature of neuropsychiatric disorders in HIV-infected patients

Neuropsychiatric disorders are common in HIV-infected patients, and they can be either primary or secondary. **Primary** complications are those that can be attributed directly to the infection of the central nervous system by the virus, or to immunopathological events precipitated by HIV infection. Primary HIV-related brain disorders include HIV-related dementia and minor cognitive disorder.[2] Immune suppression can lead to a variety of secondary complications affecting the brain, including opportunistic infections (e.g. cerebral toxoplasmosis and progressive multifocal leucoencephalopathy) and tumours (e.g. cerebral lymphoma). **Secondary** complications in the form of acute and subacute syndromes (e.g. delirium) often occur as a result of cerebrovascular complications and toxic states induced by various therapeutic agents.

HIV-associated acute stress reaction

This transitory syndrome appears in some individuals after they are notified of their seropositivity. It is equally frequent among those who, after a period as an asymptomatic carrier, are informed that the infection has progressed towards full-blown AIDS. The appearance of these symptoms is closely linked in time to the stressful circumstance, and generally remits in hours or days.

The symptoms are highly varied. Some patients suffer from intrusive thoughts or brooding related to their uncertainties regarding health, the future, the risk of contagion to others (especially loved ones), and the idea of death. The vegetative symptoms of panic attacks are also usually present. In more severe cases, the patient may also present social isolation, verbal expressions of rage or feelings of desperation, and other forms of altered behaviour.

Depression

(a) Clinical features

Depression is one of the most common psychiatric disorders found among HIV-infected individuals. Symptomatic stages of HIV infection are associated with an increased prevalence of depressive symptoms and a syndromal diagnosis of major depression.[3]

There are several factors behind the increased morbidity for affective disorders found in this population. First of all, the patient's discovery of the infection has a dramatic **psychological impact**, as does the disease's relentless progression. Second, the **neurotropism of the virus** itself produces neuropathological changes in deep grey structures whose dysfunction is known to cause mood disturbances and changes in the neurotransmission systems, which may contribute to the development of depression. Finally, the groups that in Western countries are at the highest risk for HIV infection (intravenous drug users and male homosexuals/bisexuals) are also known to be at a **high risk for depressive syndromes**, independently of having the virus. The risk factors for depression appear to be similar to those for HIV-seronegative patients and include, besides advanced HIV infection: loss of social support; personal and family history of depression; drug use; and lack of confidants.

When severe physical disease is present the diagnosis of major depression can be difficult to make, because the disease itself may be the real source of many depressive symptoms, for example insomnia, loss of appetite and weight, fatigue, lack of energy, retardation, and concentration difficulties. To avoid misdiagnosing depression, it is important to focus on the more psychological, as opposed to somatic, symptoms associated with low mood. These include **persistent low mood**, **loss of enjoyment** of usually pleasurable activities, **suicidal thoughts** and marked **feelings of hopelessness**, **guilt**, and **self-reproach**. Suicidal ideation may not be expressed directly, but may be expressed more passively, for example poor adherence to medical treatment. Assessment of depressed mood also requires evaluation of the probable contributing factors.

(b) Management

(i) Pharmacological treatment

Antidepressants are the treatment of choice in major depression, as well as in less severe depressive syndromes that are unresponsive to psychological and social intervention. Tricyclic antidepressants have been shown to be effective in treating depressed HIV-positive patients.[4] AIDS patients can respond to lower dosages of tricyclics (25–100 mg), but they may also suffer severe anticholinergic effects at reduced dosages. Therefore, the choice of an antidepressant for these patients should be guided by its side-effect profile.

Several studies have been published showing therapeutic response to selective serotonin reuptake inhibitors in seropositive patients with major depression.[5] Many clinicians prefer the newer drugs in the medically ill, not only because of their higher acceptance among patients, but also because of their greater overdose safety margin.

(ii) Psychotherapy

Psychosocial interventions derived from a wide variety of theoretical orientations are effective in treating depression among individuals infected with HIV. There is good evidence for the value of psychological intervention in the management of HIV patients. Both interpersonal psychotherapy[6] and cognitive–behavioural group therapy[7] may be particularly beneficial for HIV patients with depressive symptoms.

Psychosis

Psychotic disorders sometimes occur in people with HIV infection. While their prevalence is not high, such a development can lead to complicated diagnostic and management problems. The fact that psychosis can be related to HIV infection does not imply that a new disease entity or diagnostic category has been identified. When seropositive individuals present with psychotic symptoms, efforts should be made to clarify the clinical features and to establish their aetiology, which could well be unrelated to HIV. While in some cases the psychotic symptoms may be the result of subtle or gross brain pathology associated with HIV infection, in others it may be iatrogenic or secondary to substance misuse. Psychiatric patients per se may be considered a group at risk for contracting HIV infection.[8]

Neuroleptics are the treatment of choice for controlling psychotic symptoms. The risk of developing antipsychotic-induced extrapyramidal symptoms is higher in psychotic patients with AIDS than in psychotic patients without AIDS. AIDS patients may have an increased risk of developing tardive dyskinesia, neuroleptic malignant syndrome, and severe dystonic reactions.[9] The presence of organic cerebral deterioration, in particular HIV-associated dementia, is a risk factor for the development of neuroleptic malignant syndrome. In general, when using neuroleptics in this population, the best course is to start off with low doses, and increase the dosage slowly and progressively. The new antipsychotic risperidone has been associated with fewer extrapyramidal side-effects and used successfully in this group of patients.[10]

Mania

HIV seems to increase the risk of manic episodes, and mania is a frequent reason for psychiatric hospitalization among people with the virus.[11] In some cases illicit drug use or iatrogenic causes are implicated, for example the chance association of HIV infection and bipolar affective disorders, but generally no obvious aetiological factors can be identified. Mania has been found to be a side-effect of medication frequently used for HIV/AIDS, including **didanosine (ddI)**, **ganciclovir**, **procarbazine**, **estavudine** (d4T), steroids, and **zidovudine** (**AZT**). Most cases of new-onset mania occur in advanced HIV disease and they are often associated with the presence of substantial cognitive impairment. New-onset mania in severe symptomatic disease is predictive of reduced survival.

Standard pharmacotherapy with neuroleptics and lithium are effective, but the usefulness of these drugs may be restricted by the development of severe adverse effects in immunosuppressed HIV-infected patients. Most psychiatrists choose atypical neuroleptics for HIV. However, these agents are not without risk for extrapyramidal side-effects in HIV patients, including the risk for metabolic inhibition of some agents by protease inhibitors. Potent antiretroviral therapy has been documented to protect against the development of HIV-associated mania.[12]

Delirium

Delirium is one of the organic mental disorders observed most frequently in hospitalized HIV-infected patients. The exact prevalence of delirium or acute organic brain syndrome in HIV is unknown.

Table 5.3.5.2 Aetiology of delirium in HIV-infected patients

Infections	Encephalitis due to HIV, syphilis, toxoplasmosis, cryptococcosis, coccidioidomycosis, progressive multifocal leucoencephalopathy, herpesvirus
Abstinence	Alcohol, opiates
Metabolic	Depletion of volume, hydro-electrolytic alterations, transfusions
Hypoxia	Pneumonia with respiratory, compromise
Deficiencies	B-complex vitamins
Cerebral vascular event	
Medication	Anticholinergics, central nervous system depressors
Intracranial mass	Haematoma, neoplasias
Toxic	Drugs of abuse

Patients with advanced systemic disease and dementia are at a high risk for delirium, the cause of which is often multifactorial. The precipitant organic factors involved are listed in Table 5.3.5.2.

A conservative attitude has been recommended for the management of these conditions, with the use of low oral or intramuscular doses of neuroleptics, and correction of the organic disorders responsible for the development of disturbances in the level of consciousness.[13] However, other authors have postulated that patients suffering from delirium and agitation should be given high doses of neuroleptics—alone or in combination with lorazepam—in cases where quick control of the symptoms is vital.[14] The efficacy of pharmacological interventions in patients with delirium is heightened if treatment is begun as soon as the first symptoms appear.

Other central nervous infections in HIV-related illness

In the advanced phases of AIDS, opportunistic infections are highly varied, as are the neoplasias that can develop in immunodepressed individuals, which affect the central nervous system. The more frequents are:

- **Progressive multifocal leucoencephalopathy.** This is a grave neurological complication, linked to papovavirus infection. Dementia can develop rapidly, with focal neurological alterations such as blindness, ataxia, and hemiparesia. Death follows very quickly thereafter, and there is no known treatment. Computerized brain images taken from these patients show a characteristic involvement of the white matter.
- **Cerebral toxoplasmosis.** It is linked to the reactivation of a latent cerebral infection by *Toxoplasma gondii,* an opportunistic intracellular protozoan. The clinical presentation can vary greatly, but it is characterized by the rapid development of a marked alteration in the mental state. The focal involvement can produce headache and lateralized neurological effects. The lesions tend to be located in basal ganglia. Diagnosis is based on structural neuroimaging tests, and treatment is with pyrimethamine and sulphadiazine.
- **Cryptococcal meningitis (torulosis).** This form of meningitis, caused by infestation with the yeast-like fungus *Cryptococcus neoformans,* is characterized by headache, meningism (although it sometimes courses without this symptom), photophobia, nausea, fever, and delirium. The diagnosis is made after a lumbar puncture, and analysis of the culture and antibodies.

Syphilis

A century ago, patients with general paresis due to cerebral syphilitic infections constituted a high proportion of mental hospital admissions, and accounted for an appreciable part of the chronic population of such institutions. With the identification in the early twentieth century of the causative agent, *Treponema pallidum*, and the development of effective methods of treating syphilis, this condition has become relatively rare. Historically, the study of syphilis of the central nervous system has been of great interest to psychiatrists due to the light it sheds on the nature of the relationship between cerebral and mental disease. It was one of the first mental disorders for which a specific organic aetiology was demonstrated, and the first to respond to a medical treatment.

Syphilis remains a major problem in certain areas of the world. Because during its early stages it is a genital ulcerative disease, syphilis facilitates the transmission of HIV and may be particularly important in contributing to HIV transmission in those regions where the rates of both infections are high.

Clinical features

Syphilis is a complex STD with an extremely variable clinical course. Neurosyphilis presents 5 years or more after the initial infection. It affects 10 per cent of non-treated cases, and can take several clinical forms.[15]

- **Asymptomatic neurosyphilis**. Infected subjects have abnormalities in the cerebrospinal fluid (pleocytosis, elevated protein, and reactive VDRL score), but no symptoms or signs of central nervous system disorder. It can evolve into a symptomatic form or remit on its own.
- **Meningovascular syphilis.** Appears within 1 to 5 years of primary infection, although it can occur as early as 6 months and as late as 12 years.[16] In the clinical picture, the patient may develop stroke syndromes of subacute onset with a preceding encephalic picture, including psychiatric disturbances such as lability or personality changes. The patient may complain of headache, lethargy, and malaise, and may experience difficulty in concentration and exhibit faulty judgement. Emotional instability and irritability are common. Mental deterioration may progress to dementia, which can be accompanied by delusional symptomatology and episodes of excitation.[16]
- **General paresis.** This form of parenchymal neurosyphilis is also known as *dementia paralytica* or *general paralysis of the insane.* It usually first appears some 20 years after the initial infection. Its initial symptoms are memory disturbance, dysarthria, and hyper-reflexia, which may be accompanied by personality changes and irritability—in many cases the latter are the presenting abnormalities. The symptoms progress to dementia with abnormal motor function and psychotic symptoms. These organic psychoses were frequent in the pre-penicillin era and are known for their florid clinical picture—prominent euphoric mood, expansive demeanour, and delusions of power, wealth, or social position. Other cases may resemble depressive psychosis with somatic delusions.
- **Tabes dorsalis.** This condition is a degeneration of the ascending fibres from the dorsal root ganglia, resulting in atrophy of the dorsal roots and demyelinization in the posterior columns of the

cord. It can develop from 3 to 20 years after the initial infection. Main symptoms are the loss of position reflexes, ataxia, vibration sense, incontinence, and lacinating pains involving many areas of the body.

Diagnosis and management

The clinical picture of neurosyphilis is so variable that routine serological testing upon admission to psychiatric inpatient units has been recommended. In clinical practice, a rise in atypical syndromes with minor symptomatology has been attributed to partial suppression of the infection during its early stages by antibiotics taken for other reasons.[16] The diagnosis of neurosyphilis should be considered on the basis of the patient's symptoms and clinical signs, and confirmed by serology and analysis of the cerebrospinal fluid.

Penicillin is the drug of choice for *T. pallidum* infections, and in the treatment of neurosyphilis, but the dosage must achieve treponemicidal levels within the cerebrospinal fluid. In untreated cases, death usually occurs within 4 to 5 years. If treatment is given early, the condition usually remits; in already established cases, the progression of the disease can be halted. Antipsychotics are indicated for the symptomatic management of the excitement and psychotic symptoms that these patients may present. Clinicians should have in mind that all patients who have syphilis should be tested for HIV.

Other sexually transmitted diseases

A diverse range of psychological symptomatology is associated with STDs, in which maladaptive or pathological responses to infection (or fear of infection) may occur. Most of the studies that evaluated the psychological effects of having a STD have been carried out in patients attending genitourinary clinics, and focused on genital herpes, a common, recurrent, and painful infection. The response to diagnosis of a STD can include depression, anxiety, anger, social withdrawal, feelings of loneliness, and sexual dysfunction.[17,18] Also, high rates of hypochondriasis and veneroneurosis (a strong but unfounded conviction of having a venereal disease) are found in STD clinics and are frequently associated with psychiatric morbidity. Psychological interventions can effectively reduce the distress associated with STDs, contribute to the control of the infection, increase compliance with medication regimens, and reduce somatic symptoms misattributed to a STD.

Tuberculosis

Clinical picture

In spite of pharmacological advances, tuberculosis continues to be a serious public health problem in many parts of the world, especially due to the increased tuberculosis rate in HIV-infected patients and the appearance of multidrug-resistant tuberculosis.

Tuberculosis patients may present with vegetative signs suggestive of depression, especially the elderly and those in the symptomatic stages of HIV infection. The initial depressive symptoms are weight loss, lethargy, lack of interest, and mental confusion. They may also develop sleep disturbances due to night sweats and nocturnal coughing.

Tuberculosis patients quite frequently present with neuropsychiatric symptoms, which can be related to very different circumstances. First, clinicians should bear in mind that tuberculosis infection often develops in patients who previously had a severe psychiatric pathology, such as alcoholism or intravenous drug abuse, or in the chronic mentally ill. Also, the most common psychiatric symptoms, such as emotional lability and depression, could be related, among other factors, to the feeling of invalidity that accompanies the illness, and its social stigma. In addition, the preventive treatment of those in contact with the patient can trigger feelings of guilt. In some cases, tuberculosis leads to chronic respiratory disease, which is also associated with depressive symptoms, suicidal ideation, and cognitive impairment, particularly in debilitated patients.

Neuropsychiatric disorders in patients with tuberculosis can also be related to cerebral infections: tuberculous meningitis, potentially a very serious complication, develops in 5 per cent of all cases. Since these patients frequently present with mental symptoms that can figure prominently from the outset and even precede overt signs of meningeal infection, the correct diagnosis should be established urgently, in order to institute specific therapy as soon as possible.

Diagnosis and management

Sputum tests, cultures, and chest radiographs, as well as a tuberculin skin test, are standard diagnostic tools. In cases of suspected tuberculous meningitis, a spinal puncture is necessary to determine whether the patient has lymphocytosis and a moderate increase in protein. The diagnosis is confirmed when tubercle bacilli can be identified or cultured from the fluid; however, since irreversible brain damage may result from waiting for cultural confirmation, it is often necessary to begin therapy on the basis of a presumptive clinical diagnosis. The brain scan may show hydrocephalus, focal infarcts, and exudate in basal brain cisterns.[19]

It is important in psychiatric settings that suspected tuberculosis patients receive a proper diagnostic evaluation, not only for the sake of their own health but also for that of other patients who may be exposed to the infection in the unit. It may be necessary to transfer psychiatric inpatients with tuberculosis to a ward where isolation can be assured. The Centres for Disease Control recommend routine tuberculosis screening for patients in HIV risk groups, and for residents of mental health facilities.[20]

The most commonly used drugs for the treatment of tuberculosis are isoniazid, rifampicin, pyrazinamide, ethambutol, and streptomycin. Compliance is vital to achieve effective treatment.

Lyme disease

First described in the United States in 1975, this infection has also been reported in Europe, Australia, and other parts of the world. It is caused by the spirochaete *Borrelia burgdorferi*, which is carried and transmitted by the deer tick. The somatic symptomatology features a characteristic skin lesion, an expanding erythematous annular lesion which usually first appears 3 to 32 days after the initial transmission and may last for several weeks. In 15 per cent of the patients, the disease progresses to a secondary phase marked by neurological symptoms, for example meningoencephalitis, radiculitis, central and peripheral neuropathy, and myelitis.

The neuropsychiatric symptomatology consists in difficulties involving memory, orientation, and calculation. Even years after the first infection, patients can present with violent and impulsive

behaviour, labile affect, and depression. Cases of psychotic or catatonic syndromes and chronic dementia have been described. Many patients with Lyme disease, who suffer from neurological symptoms, present with signs of encephalopathy with alterations in their sleep, affect, and memory. The diagnosis can be established from a serological analysis.

Encephalitis

Encephalitis may be caused primarily by a viral disease affecting the brain or can be a complication of bacterial meningitis, septicaemia, or brain abscesses. It can occur after influenza, herpes simplex, measles, rubella infections, and also after vaccination. In the acute stage, the patient may present with headache, vomiting, and seizures. Patients may develop a confusional syndrome. In rare cases, the encephalitis may present with predominantly psychiatric symptoms. That is the case of herpesvirus encephalitis, which due to its damage to temporal lobes can cause a serious amnestic syndrome.

In clinical practice, the psychiatrist is more likely to see the complications that appear after the acute episode in the form of anxiety and depressive syndromes, personality change, and dementia. In the early years of life, encephalitis may be followed by behavioural disorders.[16]

Infectious mononucleosis

The Epstein–Barr virus causes 90 per cent of all cases of infectious mononucleosis. It can appear at any age, but the illness tends to manifest itself clinically in adolescents and young adults.

The most important symptoms are fever, general malaise, diffuse lymphadenopathies, and laryngitis. However, complications can lead to encephalitis and paralysis of the cranial nerves. In the case of encephalitic compromise, delirium can result. Depressive syndromes have also been observed after an acute infectious episode, accompanied by fatigue.

Brucellosis

This infection is produced by micro-organisms of the genus *Brucella*, and is transmitted by exposure to or ingestion of contaminated animal products, especially unpasteurized milk products, or contact with infected animal tissues. Onset can be insidious, since it mimics other more common illnesses, with low fever, fatigue, and sweating, but 10 to 20 per cent of cases present with splenomegaly. The psychiatric manifestations of the disease can include depressive or anxious syndromes. Diagnosis can be confirmed by blood or lymph cultures or bone marrow biopsy, although the majority of diagnoses are made serologically.

Further information

Fernandez, F. and Ruiz, P. (eds.) (2006). *Psychiatric aspects of HIV/AIDS*. Lippincott Williams & Wilkins, Philadelphia, PA.

APA AIDS resource center: www.psych.org/AIDS

CDC resource center: www.cdc.gov/std

References

1. Forstein, M. (1988). Understanding the psychological impact of AIDS: the other epidemic. *New England Journal of Public Policy*, **4**, 159–73.
2. Catalan, J., Burgess, A., and Klimes, I. (1995). *Psychological medicine of HIV infection*. Oxford Medical Publications, Oxford.
3. Ciesla, J.A. and Roberts, J.E. (2001). Meta-analysis of the relationship between HIV infection and risk for depressive disorders. *The American Journal of Psychiatry*, **158**, 725–30.
4. Markowitz, J.C., Rabkin, J.G., and Perry, S.W. (1994). Treating depression in HIV-positive patients. *AIDS*, **8**, 403–12.
5. Rabkin, J.G., Wagner, G.J., and Rabkin, R. (1999). Fluoxetine treatment for depression in patients with HIV and AIDS: a randomized, placebo-controlled trial. *The American Journal of Psychiatry*, **156**, 101–7.
6. Markowitz, J., Kocsis, J.H., Fishman, B., *et al.* (1998). Treatment of depressive symptoms in human immunodeficiency virus-positive patients. *Archives of General Psychiatry*, **55**, 452–7.
7. Blanch, J., Rousaud, A., Hautzinger, M., *et al.* (2002). Assessment of the efficacy of a cognitive-behavioral group psychotherapy programme for HIV infected patients referred to a consultation-liaison psychiatry department. *Psychotherapy and Psychosomatics*, **71**, 77–84.
8. Ayuso-Mateos, J.L., Montañes, F., Lastra, I., *et al.* (1997). HIV infection in psychiatric inpatients: an unlinked anonymous study. *British Journal of Psychiatry*, **170**, 181–5.
9. Ayuso, J.L. (1994). Use of psychotropic drugs in patients with HIV infection. *Drugs*, **4**, 599–610.
10. Singh, A.N., Golledge, H., and Catalan, J. (1997). Treatment of HIV related psychotic disorders with risperidone: a series of 21 cases. *Journal of Psychosomatic Research*, **42**, 489–93.
11. Kierburz, K., Zettelmaier, A., Ketonen, L., *et al.* (1991). Manic syndrome in AIDS. *The American Journal of Psychiatry*, **148**, 1068–70.
12. Mijch, A.M., Judd, F.K., Lyketsos, C.G., *et al.* (1999). Secondary mania in patients with HIV infection: are antiretrovirals protective? *Journal of Neuropsychiatry and Clinical Neurosciences*, **11**, 475–80.
13. Breitbart, W., Marotta, R., Platt, M.M., *et al.* (1996). A double-blind trial of haloperidol, chlorpromazine and lorazepam in the treatment of delirium in hospitalised AIDS patients. *The American Journal of Psychiatry*, **153**, 231–7.
14. Fernandez, F., Levy, J.K., and Mansell, P.W. (1989). Management of delirium in terminally ill AIDS patients. *International Journal Psychiatry in Medicine*, **19**, 165–72.
15. Simon, R.P. (1985). Neurosyphilis. *Archives of Neurology*, **42**, 606–13.
16. Lishman, W.A. (1998). *Organic psychiatry* (3rd edn). Blackwell Science, Oxford.
17. Catalan, J., Bradley, M., Gallwey, J., *et al.* (1981). Sexual dysfunction and psychiatric morbidity in patients attending a clinic for sexually transmitted diseases. *British Journal of Psychiatry*, **138**, 292–6.
18. Hedge, B. (1997). Sexually transmitted diseases. In *Cambridge handbook of psychology, health and medicine* (eds. A. Baum, S. Newman, J. Weinman, R. West, and C. McManus), pp. 584–5. Cambridge University Press, Cambridge.
19. Rovira, M., Romero, F., and Torrent, O. (1980). Study of tuberculosis meningitis by CT. *Neuroradiology*, **19**, 137–41.
20. Dooley, S.W., Castro, K.G., Hutton, M.D., *et al.* (1990). Guidelines for preventing the transmission of tuberculosis in health-care settings, with special focus on HIV-related issues. *Morbidity and Mortality Weekly Report*, **39** (RR-17), 1–29.

5.3.6 Psychiatric aspects of surgery (including transplantation)

S. A. Hales, S. E. Abbey, and G. M. Rodin

Attention to psychiatric disturbances and to emotional distress is important in the surgical setting, from the time of the initial diagnostic assessment, to the perioperative period and the phase of subsequent recovery and rehabilitation. Psychiatric illness and psychological factors, which are not taken into account prior to surgery, may contribute to inaccurate diagnoses, unrealistic assessment of the surgical risk, unnecessary surgery, and complications that could have been avoided or minimized. This chapter will address these factors and provide an approach to the consideration of psychiatric factors and interventions in this setting.

Preoperative assessment and intervention

The assessment of all patients being considered for surgery should include a brief evaluation of their current emotional state, cognitive functioning, personal circumstances, present or past history of psychiatric illness, and personality and coping style, as these factors may affect their adjustment to surgery. Psychiatric consultation may be indicated for a number of specific reasons discussed below.

Psychological contributors to the patient's physical symptoms

The most common psychological factor that complicates the surgical assessment is a low pain threshold and a tendency to somatize, i.e. to experience and communicate emotional distress in physical terms. When emotional factors amplify somatic symptoms, it is more difficult to distinguish organic from functional disorders on the basis of the clinical history. At the extreme end of the continuum of somatization is the dramatic presentation of physical symptoms, which may mimic a surgical condition, in the absence of organic disease. This syndrome may fulfil criteria for a somatoform disorder, such as a somatization disorder, conversion disorder, or somatoform pain disorder.[1] In such cases, careful attention to the objective indications for surgery is required. Individuals with a body dysmorphic disorder, a syndrome of perceived or imagined ugliness, may present with repeated requests for cosmetic surgery.[2] However, cosmetic surgery is unlikely to relieve the body dissatisfaction of such patients, whose condition has much in common with obsessive–compulsive disorder. Dissatisfaction with the results of surgery is common in patients with body dysmorphic disorder, and litigiousness and threats towards the treating surgeon may occur in a small proportion of cases. In general, the failure of clinicians to consider the contribution of somatization to the clinical presentation may lead to unnecessary or inappropriate surgery. When this occurs, postsurgical complications may interfere with the subsequent evaluation of persistent physical symptoms. Somatoform disorders are discussed further in Chapter 5.2.2.

Although physicians commonly consider that emotional factors may amplify physical symptoms, the possibility that disease has been intentionally simulated or fabricated is usually not entertained. Factitious disorder refers to a syndrome in which such behaviour is enacted for no apparent reason, other than to assume the patient role. This is in contrast with malingering, in which there may be reports of physical symptoms motivated by the desire for some specific secondary gain, which may be financial or compensation-related. While relatively rare, factitious disorder poses particular challenges, in both detection and treatment, to psychiatric and surgical teams. Individuals with this disorder may produce or simulate disease in various ways, such as by self-inflicting wounds that require surgical intervention or by surreptitiously contaminating themselves to produce infection. Patients with this condition may also communicate plausible symptoms of a surgical condition. Such patients are at risk to receive unnecessary surgery and should be regarded as suffering from a serious and potentially life-threatening condition. The majority of such patients are unwilling to accept psychological or psychiatric assistance, but an ongoing, supportive relationship with a medical caregiver may diminish this symptom pattern. Factitious disorder is discussed further in Chapter 5.2.9.

Capacity to consent to surgery

Providing information and obtaining informed consent are routine and essential aspects of preoperative care provided by the surgical team. Informed consent to a surgical procedure requires disclosure of pertinent information by the treating physician, and understanding of the information, decisional capacity, and voluntary choice on the part of the patient.[3] The decisional capacity of patients depends on their ability not only to understand information relevant to the decision, but also to apply it to their own situation and to express a consistent voluntary choice.[4] Information required by patients to make an informed surgical decision includes the rationale, risks, and benefits of the surgery, the potential alternative treatments, and the risk of not proceeding with surgery. In most jurisdictions, the emergency treatment of incapable persons is permitted, when substitute consent is not available, unless the clinician has reason to believe that the person would refuse such treatment if he or she were capable.[3] When it is not an emergency, substitute consent must be obtained on behalf of individuals who are incapable of providing informed consent. The legal requirements for substitute consent vary in different jurisdictions.

The capacity to provide informed consent may be impaired by cognitive dysfunction, by psychiatric illness, or by contextual factors, such as the clarity and relevance of the information disclosed or the manner of disclosure. If screening by the treating surgeon indicates that the patient may be incapable, a psychiatric consultation may be requested to evaluate the patient's decisional capacity.[5] Patients with cognitive impairment or a major psychiatric disorder, such as schizophrenia, are not necessarily incapable of making treatment decisions, unless these conditions affect their understanding and appreciation of information relevant to the decision. Numerous tools have been developed to assess decisional capacity but there is no current gold standard.[6]

Obtaining informed consent for surgery is not only a legal and ethical requirement, but also a crucial dimension of the surgeon–patient relationship. In some cases, treatment refusal reflects a breakdown in the relationship between the surgeon and the patient more than it does an informed decision of the patient to reject a

recommendation for surgery. When this occurs, attention to the physician–patient relationship, and the provision of additional information, may help to relieve the impasse so that an informed decision can be made. In other cases, treatment of a major psychiatric illness, such as a psychotic episode in a patient with schizophrenia, is necessary to restore the patient's capacity to provide consent.

Assessment of the response to surgery

Surgical patients face numerous stressors, including the fear of pain, disfigurement, and the loss of control, as well as the possibility of major medical complications and death. The response to these stressors may be affected by the nature of the illness and the surgical procedure, its personal meaning, the prior history of trauma, the support which is anticipated and perceived from medical caregivers and significant others, and the prior experience of the individual with medical or surgical procedures. The age and life stage of the individual, the risk associated with the procedure, and the prognosis of the underlying or associated medical conditions may also affect the psychological response in the perioperative period. Apprehension and mistrust are more common in those who have previously suffered from the adverse effects of missed or delayed diagnosis or treatment. Attitudinal factors, including positive expectations and the desire to participate actively in the recovery process, may also affect clinical outcomes. The desire to maintain a sense of control may be adaptive during the preparation and rehabilitation phases but may be associated with greater distress immediately following surgery, when there is an inescapable and predominant requirement to depend on others.[7] Those with more attachment anxiety, i.e. concern about the availability of support from others, may benefit from predictability and reliability in relation to caregivers, whereas those who tend to be more self-sufficient may benefit most from strategies which promote self-reliance and self-care.[8, 9]

There has been particular interest in psychosocial issues in the setting of transplantation surgery. This occurred, in part, because of the desire of transplant programmes to select optimal candidates for organ transplants, which were experimental and/or in scarce supply. However, the psychiatric and psychosocial selection criteria for transplant surgery have become less stringent, as the transplantation of particular organs has become more routine.[10] At present, psychosocial evaluation of transplant candidates by a multidisciplinary team allows for the identification, treatment, and monitoring of factors that may affect compliance, morbidity, and psychosocial outcomes. Organ transplants from living donors, such as for bone marrow, kidney, and liver transplantation, are unlike most other surgical interventions in that they necessitate surgery for individuals without pre-existing disease. Psychosocial evaluation of such donors includes consideration of the process of decision-making and informed consent, the adaptive capacities of the individual, the degree of social support, and the relationship of the donor to the recipient.[11] Although there has been concern about the psychological consequences of such surgery, the available evidence suggests that organ donation is usually well tolerated and experienced in positive terms by the donor, particularly when the surgical and medical outcomes are favourable.[12]

There is now increasing evidence that the systematic preoperative education of patients and their family caregivers in a therapeutic context may enhance adjustment to surgical procedures.[13] Postoperative education may also improve subsequent rehabilitation following surgical procedures. Such approaches are consistent with modern Western trends towards consumerism and patient empowerment, in which greater emphasis is placed on assisting patients to assume more responsibility for their medical course and treatment outcome. This approach has also been necessitated by the trend towards earlier hospital discharge of surgical patients into the community where much more self-care is required.

Anxiety

Preoperative anxiety is common and may be particularly problematic in patients awaiting procedures such as transplantation, which usually occur in the course of a life-threatening condition, and are associated with long and unpredictable waiting periods for surgery. Anxiety has been reported to be more common in younger patients, in females, and in those who are unmarried or who have less perceived social support.[14] Research suggests that preoperative anxiety may complicate postoperative recovery through behavioural and physiological mechanisms.[15] Symptoms of anxiety can usually be managed with education and reassurance, but when they are persistent and problematic, interventions such as progressive relaxation and guided imagery may be helpful both to reduce symptoms of anxiety and to enhance feelings of self-control.[16] Some patients benefit from a benzodiazepine to reduce preoperative anxiety, but those with antecedent anxiety disorders may require more intensive intervention, as outlined in Chapter 4.6.1. Prior to elective procedures, patients with specific blood or needle phobias may benefit from systematic desensitization. Those with comorbid panic disorder may require a higher dose of anxiolytic medication in the preoperative period. The surgical team should be aware of such treatment so that the medication can be restarted promptly after surgery, to avoid symptoms of withdrawal and anxiety. If oral medications cannot be reinstituted for a prolonged period of time after surgery, intramuscular lorazepam or intravenous lorazepam or diazepam may be used.

Mood disorders

It is important to detect and treat mood disorders prior to surgery because they are associated with increased surgical morbidity and mortality, and with reduced treatment compliance in the postoperative period.[17] Anaesthetists must be aware of any drugs taken to treat and prevent bipolar and depressive disorders, because some can significantly prolong muscle paralysis secondary to neuromuscular blockade. Furthermore, attention should be paid to serum lithium levels and signs of lithium toxicity since they may be affected by the patient's fluid and volume status. Conventional heterocyclic antidepressants and selective serotonin reuptake inhibitors can be continued until the time of surgery, and then restarted postoperatively, when oral medications can be tolerated. Selective serotonin reuptake inhibitors are known to affect platelet serotonin levels and platelet aggregation and to be occasionally associated with prolonged bleeding times, increased perioperative blood loss, and an increased subsequent need for transfusion.[18] Patients receiving monoamine oxidase inhibitors (MAOIs) are usually advised to discontinue this medication for 1 to 2 weeks prior to their surgery, although this recommendation must be weighed against the risk of withdrawal symptoms and of precipitating

a current depression. The medical charts of patients receiving MAOIs should be clearly labelled to advise that all drugs administered should be screened for their interactions with these drugs and that pethidine (meperidine) and dextromethorphan in particular should not be prescribed due to risk of a serotonin syndrome, which is associated with gastrointestinal, neurological, cardiovascular, and psychiatric symptoms. In addition, hospital charts of patients taking MAOIs should be clearly marked to indicate that they must avoid foods containing tyramine. Patients with bipolar disorder should be monitored for mood alterations since they may be at risk of developing hypomania or mania when steroid medications are used following transplantation surgery.

Psychotic disorders

These disorders, most commonly related to schizophrenia or bipolar disorder, pose challenges which vary depending upon the requirements of the surgery and the patient's mental status. Patients with schizophrenia may be at increased perioperative risk for hypotension, hypothermia, confusion, infection, and for ileus, in those who undergo abdominal surgery. These complications may occur due both to pathology of the endocrine, immune, and cardiovascular systems associated with schizophrenia and to the effects of antipsychotic medications.[19] Those who are being treated with a low-potency antipsychotic may be switched to a high-potency agent to decrease the risk of hypotension, particularly with cardiovascular surgery in the postoperative period. For their own comfort and for that of others, individuals who are actively psychotic may require special arrangements, such as a single room, close or constant observation, and, when feasible, greater family involvement. Such patients require closer monitoring for many reasons, including the increased likelihood that they may misinterpret common ward events as threatening and because they may be at increased risk for exacerbation of their underlying condition and for the occurrence of delirium.

Cognitive disorders

The capacity of patients to understand information and to provide a coherent account of their symptoms is fundamental to the process of diagnosis, informed consent, and assessment of the indications for surgery and the risk of specific postoperative complications. Cognitive impairment prior to surgery may complicate these processes and may be associated with an increased risk of delirium or dementia in the postoperative period. In such cases, neuropsychological testing may be indicated prior to elective surgery to establish a baseline, to assist in the evaluation of decisional capacity, and, to aid in the prediction of postoperative delirium or worsening of dementia.

Personality disorders

Patients with personality disorders are more likely to have greater difficulty than others adapting to the multiple and unpredictable stresses associated with surgery. Those with impulsivity may have difficulty adhering to the preoperative and postoperative regimen and those who are suspicious and mistrustful may be more limited in their ability to form effective treatment relationships and to make treatment decisions, in which an enormous degree of trust is required. Those with a borderline personality disorder may be highly sensitive to feelings of personal injury or neglect and may tend to idealize some caregivers and to denigrate others. These responses may create problematic divisions amongst the treatment team and may adversely affect the care of the patient. A psychiatric consultant may be of help to provide patient support and to educate staff about the underlying psychiatric disorder in such individuals, who may otherwise be viewed negatively by staff who regard their behaviour as simply wilful or manipulative.

Substance abuse disorders

The preoperative assessment should include enquiry about substance use, in order to adjust current medication appropriately and to prevent the occurrence of postoperative withdrawal syndromes. Consultation psychiatrists may play a role in continuing medical education about the value of routine, non-judgemental screening for substance use in medical patients, and about the importance of recording these details in the medical record.

Assessment of psychotropic medications

The pharmacological effects of psychotropic medications must be taken into account in the perioperative period. Important factors that affect the risks and benefits associated with psychotropic medication use perioperatively include:

1 End-organ sensitivity to side effects based on medical comorbidity and organ dysfunction;

2 direct effects of psychotropic medications and their potential interactions with anaesthetic and analgesic agents likely to be prescribed;

3 route of access available (oral, suppository, subcutaneous, intramuscular, intravenous);

4 risk of withdrawal symptoms and recurrence or relapse of a psychiatric disorder if psychotropic medications are to be discontinued.[20]

Information on drug–drug interactions is constantly being updated and current information can be obtained from internet sites such as http://www.drugdigest.org/DD/Home or http://search.medscape.com/drug-reference-search.

Postoperative complications and interventions

Agitation and delirium

Agitation is a common postoperative problem, the frequency of which depends upon the characteristics of the disease, the nature of the surgery and its complications, and the pre-existing vulnerability of the patient. A concerted effort should be made to identify the source of the agitation, which may be a worsening of the medical condition, inadequately controlled pain, or delirium. Delirium, which is described in more detail in Chapter 4.1.1, is a common complication, which develops in more than one-third of cases following surgery.[21] Higher rates of delirium are found following longer procedures, due to intraoperative hypoxemia, following cardiac surgeries, due to hypoperfusion and microemboli formation, following orthopaedic procedures, due to fat emboli, and following cataract surgery, due to the impact of vision loss and of ophthalmic drugs with anticholinergic side effects.[22]

Measures to prevent and ameliorate delirium include the identification and treatment of predisposing risk factors and reversible causes, symptomatic treatment, and environmental interventions to reduce distress and agitation. The latter may include measures to prevent sensory deprivation and disorientation, to monitor safety, and to educate patients and family members about the condition. When distress or agitation associated with delirium threaten the safety and care of the patient, pharmacologic interventions may be necessary. There have been trials of cholinesterase inhibitors[23] and of atypical antipsychotics in those patients who can take oral medications, but haloperidol remains the first-choice medication for management of delirium-associated agitation. It is preferred because it has fewer active metabolites and fewer anticholinergic and sedative effects than other antipsychotic medication and because it can be administered intravenously. This route of administration is usually safe, although arrhythmias with its intravenous use have been reported in patients with histories of alcohol abuse or with cardiomyopathy.[24] Benzodiazepines should be used to treat withdrawal from alcohol and sedative-hypnotics using standard protocols.

Delirium may be highly distressing for both patients and family caregivers.[25] Some patients subsequently retain disturbing memories of their experience during a delirium. Such traumatic recall may also occur, in rare instances, when there has been inadequate anaesthesia. In these cases, patients and their family caregivers typically appreciate a discussion of their concerns and the opportunity to review these events with the surgeon and the anaesthetist. A small number of patients with these disturbing experiences develop symptoms of post-traumatic stress disorder and may benefit from a brief course of psychotherapy or pharmacotherapy to alleviate their symptoms.

Ventilator weaning

Some patients experience difficulty being weaned from the ventilator due to anxiety, depression, delirium, or other psychological factors related to their disease or to the ICU environment. Behavioural approaches to facilitate weaning may include relaxation techniques that do not depend on observation or manipulation of breathing, guided imagery, and/or biofeedback.[26] Weaning problems due to anxiety typically respond to benzodiazepines or to haloperidol, administered in a single dose prior to weaning. Apathetic or depressed patients who have difficulty being weaned may benefit from a psychostimulant.[27] When delirium is the cause of weaning problems, its underlying cause should be identified and treated.

Pain management

Effective pain management is a fundamental aspect of postoperative treatment and may reduce distress, agitation, sleep disturbance, anxiety, mood symptoms, and behavioural disorders. Suboptimal pain management may occur due to inadequate assessment of this symptom, insufficient knowledge of the pharmacokinetic and pharmacodynamic properties of analgesic medication, and unfounded concerns about 'addiction'. Further, some patients refuse to accept adequate analgesia because of misconceptions or personal beliefs regarding the importance of stoicism, vigilance, or personal control. Consultant psychiatrists may help patients to address their concerns about analgesic medication and analgesic adjuvants and may act also as advocates or intermediaries for patients to ensure adequate analgesia.

Sleep disturbance

Sleep disturbances, including reduced total sleep time, fragmentation of sleep, frequent arousals and awakenings, and reduced slow-wave sleep, are common in the immediate postoperative period. These disturbances may be caused by multiple factors, including the noise, temperature and light of the hospital environment, neuroimmunological and other changes associated with the surgical insult, and anaesthetic and analgesic medications.[28] Daytime sleeping, related to prolonged bed rest, lack of intellectual and social stimulation, and reduced circadian cues, may disrupt normal sleep chronobiology and increase difficulty with night-time sleep.

Treatment of sleep disturbances should be directed to the identified cause. In those patients in whom anxiety or the intrusiveness of the hospital environment is the main cause of the sleep disturbance, benzodiazepines may be temporarily used. When there are concerns about substance abuse, newer non-benzodiazepine hypnotics may be preferable, due to the lower risk of tolerance associated with their use.[28] Caution should be exercised with patients for whom benzodiazepine-induced nocturnal respiratory compromise may be problematic or with elderly patients who are susceptible to cognitive compromise or to falls.

Cognitive impairment

Cognitive impairment is a common short-term and long-term complication of major surgery, particularly in those with more advanced age. Gradual recovery of cognitive functions occurs in most patients within 3 months after surgery, although it may take as long as 6 to 12 months.[29] Delirium that is slow to resolve may be an early sign of an associated dementing illness or of a cerebral insult that has occurred during or after surgery. Cognitive impairment may be permanent when there has been irreversible brain damage due to neurosurgery or perioperative complications, such as hypoxia or stroke. Neuropsychological testing may be helpful to document or track changes in cognitive functioning. Education of primary caregivers about the risk and manifestations of cognitive impairment is essential, and institution of home supports and respite care for patients and families facing this complication may be necessary. Some patients with cognitive impairment may also benefit from neurorehabilitation interventions.

Adjustment issues

Longer-term problems in adjustment may occur, particularly following disfiguring surgeries, such as facial surgery, amputations, ostomies, or following procedures such as organ transplantation, which impose complicated postoperative regimens. Sexual difficulties that result from procedures that compromise the neural input or functional integrity of genital structures, or that negatively affect body image or feelings of attractiveness, may be disturbing to patients and spouses who may benefit from specific enquiry and assistance. Patients and their families should be informed about the possibility of adjustment problems with such surgeries and should be given information about available resources. A minority of patients develop clinically significant mood or anxiety disorders or significant compliance problems during the

rehabilitation phase which may necessitate psychiatric consultation and intervention.

Prolonged dysfunction

Prolonged and disproportionate pain and disability occur in a subset of patients. In some of these cases, the surgery was undertaken to relieve refractory symptoms that subsequently proved to be functional or medically unexplained. This may occur with hysterectomies performed to relieve pelvic pain and in surgery to relieve chronic back pain. In other cases, persistent disproportionate symptoms may be perpetuated by secondary gain of a financial or social nature, or by opiate dependence. Undiagnosed and untreated mood, anxiety, and somatoform disorders may also contribute to persistent symptoms. The consultation psychiatrist may be called upon to identify these factors and to help distinguish them from undetected medical/surgical pathology or from the effects of inadequate pain regimens.

Further information

Levenson, J.L. (ed.) (2005). *American psychiatry publishing textbook of psychosomatic medicine*. American Psychiatric Press, Washington, DC.

Lloyd, G. and Guthrie, E. (eds.) (2007). *Handbook of liaison psychiatry*. Cambridge University Press, Cambridge.

References

1. Abbey, S.E. (2002). Somatization and somatoform disorders. In: *The American psychiatric publishing textbook of consultation-liaison psychiatry* (2nd edn) (ed. M.G. Wise and J.R. Rundell), pp. 361–92. American Psychiatric Press, Washington, DC.
2. Phillips, K.A., McElroy, S.L., Keck, P.E. Jr., *et al.* (1993). Body dysmorphic disorder: 30 cases of imagined ugliness. *The American Journal of Psychiatry*, **150**, 302–8.
3. Etchells, E., Sharpe, G., Walsh, P., *et al.* (1996). Bioethics for clinicians: 1. consent. *CMAJ: Canadian Medical Association Journal*, **155**, 177–80.
4. Grisso, L.B. and Appelbaum, P.S. (1998). *Assessing competence to consent to treatment: a guide for physicians and other health professionals.* Oxford University Press, New York.
5. Etchells, E., Sharpe, G., Elliott, C., *et al.* (1996). Bioethics for clinicians: 3. Capacity. *CMAJ: Canadian Medical Association Journal*, **155**, 657–61.
6. Dunn, L.B., Nowrangi, M.A., Palmer, B.W., *et al.* (2006). Assessing decisional capacity for clinical research or treatment: a review of instruments. *The American Journal of Psychiatry*, **163**, 1323–34.
7. Rosenberger, P.H., Jokl, P., and Ickovics, J. (2006). Psychosocial factors and surgical outcomes: an evidence-based literature review. *The Journal of the American Academy of Orthopaedic Surgeons*, **14**, 397–405.
8. Hunter, J.J. and Maunder, R.G. (2002). Using attachment theory to understand illness behavior. *General Hospital Psychiatry*, **23**, 177–82.
9. Tan, A., Zimmermann, C., and Rodin, G. (2005). Interpersonal processes in palliative care: an attachment perspective on the patient–clinician relationship. *Palliative Medicine*, **19**, 143–50.
10. Olbrisch, M.E., Benedict, S.M., Ashe, K., *et al.* (2002). Psychological assessment and care of organ transplant patients. *Journal of Consulting and Clinical Psychology*, **3**, 771–83.
11. Olbrisch, M.E., Benedict, S.M., Haller, D.L., *et al.* (2001). Psychosocial assessment of living organ donors: clinical and ethical considerations. *Progress in Transplantation*, **11**, 40–9.
12. Shrestha, R. (2003). Psychosocial assessment of adult living donors. *Liver Transplantation*, **9**, S8–11.
13. Walker, J. (2002). Emotional and psychological preoperative preparation in adults. *British Journal of Nursing*, **11**, 567–75.
14. Karanci, A.N. and Dirik, G. (2003). Predictors of pre- and postoperative anxiety in emergency surgery patients. *Journal of Psychosomatic Research*, **55**, 363–9.
15. Kiecolt-Glaser, J.K., Page, G.G., Marucha, P.T., *et al.* (1998). Psychological influences on surgical recovery: perspectives from psychoneuroimmunology. *The American Psychologist*, **53**, 1209–18.
16. Horne, D. (1994). Preparing patients for invasive medical and surgical procedures. 2: using psychological interventions with adults and children. *Behavioral Medicine*, **20**, 15–21.
17. Rodin, G., Craven, J., and Littlefield, C. (1991). *Depression in the medically ill: an integrated approach*. Brunner–Mazel, New York.
18. Movig, K.L., Janssen, M.W.H.E., de Waal Malefijt, J., *et al.* (2003). Relationship of serotonergic antidepressants and need for blood transfusion in orthopedic surgical patients. *Archives of Internal Medicine*, **163**, 2354–8.
19. Kudoh, A. (2005). Perioperative management for chronic schizophrenic patients. *Anesthesia and Analgesia*, **101**,1867–72.
20. Huyse, F.J., Tuow, D.J., van Schijndel, R.S., *et al.* (2006). Psychotropic drugs and the perioperative period: a proposal for a guideline in elective surgery. *Psychosomatics*, **47**, 8–22.
21. Dyer, C.B., Ashton, C.M., and Teasdale, T.A. (1995). Postoperative delirium. A review of 80 primary data-collection studies. *Archives of Internal Medicine*, **155**, 461–5.
22. Winawer, N. (2001). Postoperative delirium. *The Medical Clinics of North America*, **85**, 1229–39.
23. Overshott, R., Burns, A., and Karim, S. (2005). Cholinesterase inhibitors for delirium. (Protocol) Cochrane Database of Syst Rev [serial online] [cited 2005 Apr 20] Issue 2. Art. No.: CD005317. DOI: 10.1002/14651858.CD005317. Available from: URL:http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD005317/frame.html
24. Schwartz, T. and Masand, P. (2002). The role of atypical antipsychotics in the treatment of delirium. *Psychosomatics*, **43**, 171–4.
25. Breitbart, W., Gibson, C., and Tremblay, A. (2001). The delirium experience: delirium recall and delirium-related distress in hospitalized patients with cancer, their spouses/caregivers, and their nurses. *Psychosomatics*, **43**, 183–94.
26. Hannich, H.J., Hartmann, U., Lehmann, C., *et al.* (2004). Biofeedback as a supportive method in weaning long-term ventilated critically ill patients. *Medical Hypotheses*, **63**, 21–5.
27. Johnson, C.J., Auger, W.R., Fedullo, P.F., *et al.* (1995). Methylphenidate in the 'hard to wean' patient. *Journal of Psychosomatic Research*, **39**, 63–8.
28. Morin, A.K., Jarvis, C.I., and Lynch, A.M. (2007). Therapeutic options for sleep-maintenance and sleep-onset insomnia. *Pharmacotherapy*, **27**, 89–110.
29. Dijkstra, J.B. and Jolles, J. (2002). Postoperative cognitive dysfunction versus complaints: a discrepancy in long-term findings. *Neuropsychology Review*, **12**, 1–14.

5.3.7 Psychiatric aspects of cancer

Jimmie C. Holland and Jessica Stiles

Introduction

Psycho-oncology addresses the two major psychiatric and psychological dimensions of cancer: first, the responses of patients and their families at all stages of disease and the psychological stresses on health professionals delivering their care. The patient and physician relationship, dependent on effective communication, impacts

the care of all patients, at every visit, at all sites and stages of cancer, and during all treatments. The second dimension addresses the psychological, behavioural, and social factors that influence cancer risk, detection, and survival.

Many cancer centres and hospitals now have multi-disciplinary psychosocial teams consisting of clinicians and clinical investigators from psychology, psychiatry, social work, nursing, and clergy. These teams provide consultation for patients and their caregivers, psychosocial education for oncology staff, and collaboration in studies in which quality of life is important. In addition, active research in brain, immune, and endocrine links is occurring, particularly in the mechanism of cytokines in producing 'sickness behaviour' that may provide a biological basis for common symptoms of fatigue, depression, anxiety, weakness, and cognitive chances in cancer patients.[1, 2]

Despite the fact that many cancer centres and oncology divisions now have a psycho-oncology or psychosocial unit, only a few centres have programmes that include both research and training.

This chapter describes the common psychiatric disorders and psychosocial challenges experienced by cancer patients and the range of interventions available.

Psychiatric disorders

A key challenge for the oncologist is the differentiation of expected, tolerable, transient distress associated with cancer, such as fear, worry, and sadness, from excessive, disabling, persistent distress requiring therapeutic intervention. Most psychiatric disturbances in patients with cancer relate to their illness or treatment side effects.[3] One-third of patients will experience distress that requires evaluation and treatment.[3–6] The percentage is greater among younger patients, those with sites of cancer with poorer prognosis, for example, brain, pancreas, lung, and those who are hospitalized with greater level of illness causing confusional states and greater anxiety and depression.[7,8]

Anxiety

Anxiety is the most common form of distress experienced by patients in the oncology setting (Table 5.3.7.1). It occurs with abnormal metabolic states: hypoxia, pulmonary embolus, sepsis, delirium, bleeding, cardiac arrhythmia, and hypoglycemia. Hormone-secreting neoplasms that produce psychiatric symptoms consistent with mood or anxiety disorder are pheochromocytoma, thyroid tumour, carcinoid, parathyroid adenoma, adrenocorticotropic hormone-producing tumour, insulinoma, and paraneoplastic syndrome, an immunologic non-metastatic central nervous system complications of several tumours (particularly, lung and ovary) that may present with mood or cognitive changes.

Numerous medications produce symptoms of anxiety: corticosteroids, neuroleptics, bronchodilators, thyroxine, and psychostimulants. The antiemetics, including metoclopramide and prochlorperazine, which are widely used for chemotherapy-related nausea and vomiting, produce restlessness, akathisias, and dystonias. Benzodiazepines promptly reduce the restless movements, anxiety, and agitation. Withdrawal states from alcohol, benzodiazepines, sedative-hypnotics, and opioids produce anxiety as prominent symptoms.

Some patients undergoing cyclic chemotherapy receiving highly emetogenic regimens develop anticipatory anxiety, nausea, and vomiting days to hours in advance of receiving the next cycle of treatment.[9–11] More effective antiemetic regimens have significantly reduced the frequency and severity of this problem. However, behavioural interventions paired with antianxiety medications continue to assist in providing relief from this distress.

Table 5.3.7.1 Causes of anxiety in patients with cancer

Situational
Diagnosis of cancer, prognosis discussion
Crisis, illness/treatment
Conflicts with family or staff
Anticipating a frightening procedure
Awaiting results of tests
Fears of recurrence after completing treatment
Disease-related
Poorly controlled pain
Abnormal metabolic states
Hormone secreting tumors
Paraneoplastic syndromes (remote CNS effects)
Treatment-related
Frightening or painful procedures (MRI, scans, wound debridement)
Anxiety-producing drugs (antiemetic neuroleptics, bronchodilators)
Withdrawal states (opioids, benzodiazepines, alcohol)
Conditioned (anticipatory) anxiety, nausea, and vomiting with cyclic chemotherapy
Exacerbation of preexisting anxiety disorder
Phobias (needles, claustrophobia)
Panic or generalized anxiety disorder
Posttraumatic stress disorder (Holocaust survivors, Vietnam veterans, recall of the death of a relative with cancer)
Obsessive compulsive disorder

Patients who have pre-existing phobias, panic attacks, generalized anxiety disorder, or obsessive–compulsive disorder are at risk of experiencing symptom exacerbations during treatment (Table 5.3.7.1).[12] Phobias of needles, blood, hospitals, magnetic resonance imaging machines, or radiation simulators complicate a patient's ability to tolerate hospital procedures or adhere to recommended treatments. Panic attacks superimposed on physical symptoms of dyspnea and tachycardia may be partially alarming to patients.[3,13] Patients with previous traumatic experiences may suffer a recurrence of intrusive re-experiences of painful memories, maladaptive avoidant behaviour or withdrawal, and hypervigilance.[14,15]

Cancer patients with OCD may have increased difficulty during treatment. Intrusive fears may lead to indecisiveness regarding treatment options and reluctance to accept interventions with known therapeutic efficacy. Excessively time-consuming rituals may interfere with a patient's adherence to medical appointments. Inflexibility of thought, hostility, overwhelming distress, and occasionally poor insight contribute to the challenge of engaging these patients and assisting them in accepting interventions.

Management. Anticipatory anxiety prior to medical interventions responds to empathic validation of the fear, adequate preparation to set realistic expectations for the encounter, and rehearsal of the dreaded event.

Significant disabling anxiety symptoms are frequently treated pharmacologically with benzodiazepines, selective serotonin-reuptake inhibitors (SSRIs), mirtazapine, venlafaxine, buspirone, antihistamines, beta-blockers, or neuroleptics. Table 5.3.7.2 outlines the benzodiazepines commonly used and their initial and

Table 5.3.7.2 Common anxiolytic agents

Drug	Brand name	Starting dose/day	Theraputic dose/day
SSRIs			
Sertraline	Zoloft	25–50 mg AM	50–150 mg
Fluoxetine	Prozac	10–20 mg AM	20–60 mg
Paroxetine	Paxil	10–20 mg	20–60 mg
Citalopram	Celexa	10–20 mg	20–60 mg
Escitalopram	Lexapro	5–10 mg	10–30 mg
Benzodiazepines			
Alprazolam, XR	Xanax	0.25–0.5 mg	0.5–2.0 mg
Clonazepam, wafers	Klonopin	0.25–0.5 mg	0.5–2.0 mg
*Lorazepam	Ativan	0.25–0.5 mg	0.5–2.0 mg
*Diazepam	Valium	2 mg	5–20 mg
Hypnotics			
Temazepam	Restoril	15 mg	15–45 mg
Zolpidem	Ambien	5 mg	5–20 mg
Zaleplon	Sonata	5 mg	5–20 mg
Eszopiclone	Lunesta	2 mg	2–3 mg

* Also IV, IM

therapeutic doses. A shorter half-life enhances control during the upward titration process and decreases the risk of accumulation and intoxications.

Mood disorders

Depression in cancer patients requires early recognition and therapeutic intervention. Depression is more challenging to diagnose in patients with cancer because illness produces many neurovegetative symptoms: sleep disturbances, appetite reduction and weight loss, psychomotor retardation, fatigue, apathy, and poor concentration.[16] Focusing the assessment on the psychological symptoms of dysphoria, anhedonia, hopelessness, worthlessness, excessive guilt, and suicidal ideation helps distinguish depression in the context of medical illness.[17]

Table 5.3.7.3 outlines the medically related risk factors for developing depression: increasing levels of debilitation, advanced disease, and concurrent presence of other chronic illnesses or disabilities. Medications frequently encountered in the oncology setting that contribute to depressive symptoms are corticosteroids (dexamethasone and prednisone), chemotherapeutics (interferon, interleukin-2, vincristine, procarbazine, l-asparaginase), and supportive care medications.[17] Depression may relate to organ failure or nutritional, endocrine, and neurologic complications of cancer. Depression is a common symptom of pancreatic cancer, which led to speculation about a tumour-induced mood disturbance mediated by alteration of brain serotonergic function through the effect of proinflammatory cytokines.[18–21]

Management. Psychotropic medications are effective in reducing depressive symptoms present in cancer patients.[22] Table 5.3.7.4 lists the most frequently used antidepressant medications in patients with cancer and their initial and maintenance doses. The antidepressants commonly used today are SSRIs, mirtazapine, venlafaxine, or buproprion. Tricyclic antidepressants and duloxetine are beneficial for patients with depressive symptoms and neuropathic pain. Psychostimulants treat depressive symptoms, and counter fatigue related to advanced illness and the somnolence

Table 5.3.7.3 Medical-related risk factors for depression in patients with cancer

Poorly controlled pain
Other chronic disease/disability; advanced stage
Medications
Corticosteroids
Prednisone, dexamethasone
Inteferon and Interleukin-2
Chemotherapeutic agents
Vincristine, vinblastine, procarbazine, L-asparaginase
Other medications
Cimetidine
Indomethacin
Levodopa
Methyldopa
Pentazocine
Phenmetrazine
Phenobarbital
Propranolol
Rauwolfia alkaloids
Tamoxifen
Antibiotics
(Amphotericin B)
Other medical conditions
Metabolic (anemia; hypercalcemia)
Nutritional (B_{12} or folate)
Endocrine (hyper-hypothyroidism; adrenal insufficiency)
Neurologic (paraneoplastic syndrome)
Sites of cancer
Pancreatic, small cell lung, breast cancer, lymphoma (producing remote CNS effects)

associated with opioids. For depressed cancer patients not expected to survive weeks to months, psychostimulants provide more rapid relief from distressing depressive symptoms. Initiating antidepressants at low doses for elderly and debilitated patients and titration upward as tolerated provides similar benefits, but over a longer period of time.

Suicide and cancer

The incidence of suicide is increased in patients with cancer compared with the general population, but it is not as high as is often assumed. Suicide is more likely to occur in advanced disease as depression, hopelessness, and the presence of poorly controlled symptoms (especially pain) escalate.

Evaluation of suicidal thoughts should take into account disease stage and prognosis. Almost all patients who receive a diagnosis of cancer, even if the prognosis is optimistic, consider or contemplate suicide in the event of developing unbearable or intolerable distress. Some patients maintain supplies of medications for this purpose. This practice allows the patient to maintain a perception of control over progressive disease and feared intolerable pain and inevitable distress. Maintaining this option sometimes allows patients to tolerate difficult treatments.

Morbid preoccupation with suicide or ruminative plans to commit suicide in cancer patients for whom the disease is in remission or in whom a good prognosis exists require careful evaluation.[23] Patients with a poor prognosis, advanced disease, and poorly controlled symptoms often have thoughts of suicide that are more

Table 5.3.7.4 Commonly used antidepressants in cancer

Drug	Brand name	Starting daily dosage PO (mg)	Therapeutic daily dosage PO (mg)
Selective serotonin-reuptake inhibitors			
Sertraline	(Zoloft)	25–50 mg AM	50–150 mg
Fluoxetine	(Prozac)	10–20 mg AM	20–60 mg
Paroxetine	(Paxil)	10–20 mg AM	20–60 mg
Citalopram	(Celexa)	10–20 mg AM	20–60 mg
Escitalopram	(Lexapro)	5–10 mg	10–20 mg
Tricyclics (neuropathic pain management primarily)			
Nortriptyline	(Pamelor)	25–50 mg	50–200 mg
Amitriptyline	(Elavil)	25–50 mg	50–200 mg
Desipramine	(Norpramin)	25–50 mg	50–200 mg
Other agents			
Venlafaxine	(Effexor)	18.75–37.5 mg	75–225 mg
Trazodone	(Desyrel)	50–100 mg	100–200 mg
Bupropion (XL, SR)	(Wellbutrin, Zyban)	50–75 mg	150–400 mg
Mirtazapine	(Remeron)	15 mg HS	15–45 mg
Psychostimulants			
Methylphenidate	(Ritalin)	5–10 mg (8AM & Noon)	10–30 mg
Modafinil	(Provigil)	50–100 mg (8AM & Noon)	100–400 mg
Dextroamphetamine	(Dexedrine)	5–10 mg (8AM &Noon)	10–20 mg

Lithium and mood stabilizers only for bipolar disorder;

MAOIs not recommended.

likely to be viewed as rational by physicians.[24] These patients may request assistance from a physician in obtaining a prescription for medications to use to commit suicide. A treatable major depressive episode may precipitate their suicidal ideation, so it is particularly important to evaluate for the presence of hopelessness, which is a better predictor of suicidal risk than depression itself.[23]

Management. Attentiveness to uncontrolled physical symptoms, especially pain, is crucial. Adequate pain control may have a dual effect of hastening death, while ameliorating suffering. Most physicians feel comfortable providing comfort and relieving distress. Increasing numbers of physicians do not consider this practice to be assisted suicide but as best medical care geared to maximal comfort.[25]

Poorly controlled pain in patients with organ failure and metabolic encephalopathy may result in poor judgement and impulse control leading to unpredictable suicide attempts.[26] These patients benefit from a 24 h companion, nurse, or family member who understands the patient's compromised state and treatment for delirium.

Delirium

Delirium is a global cerebral dysfunction characterized by a fluctuating level of arousal and cognitive disturbances. Symptoms include disorientation and confusion, inattention and poor concentration, perceptual disturbances, disordered thought process, psychomotor agitation or retardation, and an altered sleep–wake cycle. Delirium is distinguished from dementia in part by its reversibility. However, in advanced cancer, as organ failure progresses and results in refractory metabolic derangements, delirium may be irreversible. The primary goal is ensuring the safety of the patient and caregivers. Protecting others from aggressive or combative behaviour is essential. Family members should be told that the cause of the behaviour is brain dysfunction, not a mental aberration, and a given guidance in understanding the patient's states.

In patients with cancer, especially those in advanced stages, an abrupt shift in mood or behaviour is most often related to a change in neurologic, vascular, or metabolic status; a psychological basis is far less likely. In fact, up to three-fourths of terminally ill patients may develop a delirium before death. Common causes of delirium in cancer are outlined in Table 5.3.7.5.

Management begins with attention to the patient's safety. It is important to have constant 1:1 observation, preferably by a person who can correct the patient's misinterpretations of reality. Providing

Table 5.3.7.5 Common causes of delirium in cancer

Causes	Examples
Metabolic encephalopathy because of vital organ failure	Liver, kidney, lung (hypoxia), thyroid, adrenal
Electrolyte imbalance	Sodium, potassium, calcium, glucose
Treatment side effects	Narcotic/analgesics Anticholinergics Phenothiazines Antihistamines Chemotherapuetic agents Steroids Radiation therapy to brain
Infection	Septicemia
Hematologic abnormalities	Microcytic and macrocytic anemias, coagulopathies
Nutritional	General malnutrition, thiamine, folic acid, vitamin B_{12}
Paraneoplastic syndromes	Remote effects of tumors
Metastatic or primary brain tumor	Glioblastoma multiforme, primary CNS lymphoma

Table 5.3.7.6 Behavioral symptoms of delirium in patients with cancer

State	Symptom
Early, mild	Alteration of sleep-wake cycle, transient periods of disorientation Unexplained anxiety and sense of dread Increased irritability, anger, temper outbursts Withdrawal, refusal to talk to staff or relatives New onset of forgetfulness
Late, severe with behavioral changes	Refusal to cooperate with reasonable requests; pulling out tubes and lines Angry, swearing, shouting, abusive Demanding to go home, pacing corridor Illusions (misidentifies staff, visual and sensory clues) Delusions (misinterprets events, usually paranoid, fears of being harmed) Hallucinations (visual and auditory)

consistent caregivers, structured interactions with others, and frequent reorientation may limit the distress experienced by the patient. Elderly patients are more vulnerable to developing delirium, and those with cognitive impairment or dementia are at even higher risk. It is preferable to avoid physical restraints. However, containment of severe agitation may temporarily require restraints to prevent removal of endotracheal tubes, intravenous access, and loss of indwelling catheters, and also to avert falls. Behavioural symptoms of delirium in patients with cancer are outlined in Table 5.3.7.6.

Medications commonly used in managing delirium are summarized in Table 5.3.7.7. Identification and corrections of the underlying aetiology of the delirium is not always possible and in these circumstances, neuroleptics providing relief for the patient from distressing symptoms should be the primary intervention. While haloperidol and lorazepam administered in conjunction provide additional sedative effects for patients with sever agitation,[27] patients with hypoactive subtype of delirium will benefit from the administration of neuroleptics for symptomatic relief.[26,28] In the terminal stage of cancer, delirium may be irreversible and refractory to neuroleptics. Additional sedation with alternative agents may be required to provide comfort and safety for the patient and family.

Table 5.3.7.7 Medications for managing delirium in cancer patients

Drug	Brand name	Approximate daily dosage
Neuroleptics		
Haloperidol	Haldol	0.5–5 mg every 2–12 h, PO, IV, SC, IM
Chlorpromazine	Thorazine	12.5–50 mg every 4–12 h, PO, IV, IM
Risperidone	Risperdal	1–3 mg every 12 h, PO
Olanzapine	Zyprexa	2.5–5 mg every 6–8 h, PO
Quetiapine	Seroquel	12.5–50 mg every 12 h, PO
Benzodiazepines		
Lorazepam	Ativan	0.5–2.0 mg every 1–4 h, PO, IV, IM
Midazolam*		30–100 mg every 24 h, IV, SC
Anesthetics		
Propofol*		10–50 mg every h, IV

PO, orally; IV, intravenously; SC, subcutaneously; IM, intramuscularly.

* Usually IV continuous infusion in intensive care setting.

Further information

www.apos-society.org

Free online education program for multidisciplinary training in Psycho-Oncology. Website contains fifteen web-cast lectures in the five following tracks: Introduction to oncology, program administration, symptom detection and management (eight web-casts), interventions (four web-casts), and population-specific issues.

Direct link: http://www.apos-society.org/professionals/meetings-ed/webcasts/webcasts-multidisciplinary.aspx

www.ipos-society.org

Free online lectures: Multilingual core curriculum in psycho-oncology. Five web-cast lectures translated into English, French, German, Hungarian, Italian, and Spanish.

Direct link: http://www.ipos-society.org/professionals/meetings-ed/core-curriculum/core-curriculum-pres.htm

Holland, J.C. (2002). History of psycho-oncology: overcoming attitudinal and conceptual barriers. *Psychosomatic Medicine*, **64**, 206–21.

Holland, J.C., (ed.) (2006). *Quick reference for oncology clinicians: the psychiatric and psychological dimensions of symptom management*. IPOS Press, Charlottesville.

Psycho-oncology: Journal of the American Psychosocial Oncology Society, Behavioral and Ethical Aspects of Cancer. Published monthly in hard copy and online by John Wiley & Sons, Ltd.

References

1. Cleeland, C.S., Bennett, G.J., and Dantzen, R. (2003). Are the symptoms of cancer and cancer treatment due to a shared biologic mechanism? *Cancer*, **97**, 2919–25.
2. Musselman, D.L., Miller, A.H., Parter, M.R., *et al.* (2001). Higher than normal interleukin-6 concentrations in cancer patients with depression: preliminary findings. *The American Journal of Psychiatry*, **158**, 1252–7.
3. Strain, J. (1998). Adjustment disorders. In *Psycho-oncology* (ed. J.C. Holland), pp. 509–17. Oxford University Press, New York.
4. Hewitt, M., Herdman, R., and Holland, J. (eds.) (2004). *Meeting psychosocial needs of women with breast cancer*. Institute of Medicine, National Academies Press, Washington, DC.
5. Barg, F.K., Cooley, M., Pasacreta, J., *et al.* (1994). Development of a self-administered psychosocial cancer screening tool. *Cancer Practice*, **2**, 288–96.
6. Farber, D.M., Wienerman, B.H., and Kuypers, J.A. (1984). Psychosocial distress in oncology outpatients. *Journal of Psychosocial Oncology*, **2**, 109.
7. Zabora, J. (2001). The prevalence of psychological distress by cancer site. *Psycho-oncology*, **10**, 19–28.
8. Bukberg, J., Penman, D., and Holland, J.C. (1984). Depression in hospitalized cancer patients *Psychosomatic Medicine*, **46**, 199.
9. Jacobsen, P.B., Bovbjerg, D.H., Schwartz, M., *et al.* (1995). Conditioned emotional distress in women receiving chemotherapy for breast cancer. *Journal of Consulting and Clinical Psychology*, **63**, 108–14.
10. Andrykowski, M.A. and Redd, W.H. (1987). Longitudinal analysis of the development of anticipatory nausea. *Journal of Consulting and Clinical Psychology*, **55**, 36.
11. Redd, W.H., Jacobsen, P.B., Die-Trill, M., *et al.* (1987). Cognitive/attentional distraction in the control of conditioned nausea in pediatric oncology patients receiving chemotherapy. *Journal of Consulting and Clinical Psychology*, **55**, 391.
12. Payne, D. and Massie, M. (2000). Anxiety in palliative care. In *Handbook of psychiatry in palliative medicine* (eds. H. Chochinov and W. Breitbart). pp. 63–74. Oxford University Press, New York.
13. Noyes, R. Jr., Holt, C., and Massie, M. (1998). Anxiety disorders. In *Psycho-oncology* (ed. J.C. Holland), pp. 548–63. Oxford University, New York.

14. Rowland, J.H. (1989). Intrapersonal resources: developmental stage of adaptation: adult model. In *Handbook of psycho-oncology: psychological care of the patient with cancer* (eds. J.C. Holland and J.H. Rowland), pp. 25–43. Oxford University Press, New York.
15. Peretz, T., Baider, L., Ever-Hadani, P., *et al.* (1994). Psychological distress in female cancer patients with holocaust experience. *General Hospital Psychiatry*, **16**, 413–18.
16. Roth, A.J. and Holland, J.C. (1994). Treatment of depression in cancer patients. *Primary Care in Cancer*, **14**, 23–9.
17. Coups, E., Winell, J., and Holland, J. (2005). Depression in the context of cancer. In *Biology of depression: from novel insights to therapeutic strategies*, Vol. 1 (eds. J. Lucinio and W. Ma-LeWong), pp. 365–85. Wiley, Weinheim, Germany.
18. Ebrahimi, B., Tucker, S.L., Li, D., *et al.* (2004). Cytokines in pancreatic carcinoma. *Cancer*, **101**, 2727–36.
19. McDaniel, J.S., Musselman, D.L., and Porter, M.R. (1995). Depression in cancer: diagnosis, biology and treatment. *The Journal of General Psychology*, **52**, 89–99.
20. Lerner, D.M., Stoudemire, A., and Rosenstein, D.L. (2001). Cytokine-induced neuropsychiatric toxicity. In *Cytokine therapeutics in infectious diseases* (ed. S.M. Holland), pp. 323–32. Lippincott Williams & Wilkins, Philadelphia, PA.
21. Mussleman, D.L., Lawson, D.H., Gumnick, J.F., *et al.* (2001). Paroxetine for the prevention of depression induced by high-dose interferon alpha. *The New England Journal of Medicine*, **344**, 961–6.
22. Wilson, K., Chochinov, H., de Faye, B., *et al.* (2000). Diagnosis and management of depression in palliative care. In *Handbook of psychiatry in palliative medicine* (eds. H. Chochinov and W. Breitbart), pp. 25–51. Oxford University Press, New York.
23. Rosenfeld, B., Krevo, S., Breitbart, W., *et al.* (2000). Suicide, assisted suicide, and euthanasia in the terminally ill. In *Handbook of psychiatry in palliative medicine* (eds. H. Chochinov and W. Breitbart), pp. 51–63. Oxford University Press, New York.
24. Conwell, Y. and Caine, E.D. (1991). Rational suicide and the right to die: reality and myth. *The New England Journal of Medicine*, **325**, 1100.
25. Chochinov, H.M. (2002). Dignity-conserving care- a new model for palliative care: helping the patient feel valued. *The Journal of the American Medical Association*, **287**, 2253–60.
26. Breitbart, W. and Cohen, K. (1998). Delirium. In *Psycho-oncology* (ed. J.C. Holland), pp. 564–75. Oxford University Press, New York.
27. Breitbart, W., Marotta, R., Platt, M., *et al.* (1996). A double-blind trial of haloperidol, chlorpromazine and lorazepam in treatment of delirium in hospitalized AIDS patients. *The American Journal of Psychiatry*, **153**, 231–7.
28. Breitbart, W. (2002). Spirituality and meaning- centered group psychotherapy interventions in advanced cancer. *Support Care in Cancer*, **10**, 272–80.

5.3.8 Psychiatric aspects of accidents, burns, and other physical trauma

Ulrik Fredrik Malt

Epidemiology of accidents and injury

The one-year prevalence of accidents is about 15–20 per cent with highest prevalence in the younger age groups. About 80 per cent of accidents cause personal injury, and 1/3 to 1/2 of these injuries result in medical attention. About 10 per cent of medically attended injured victims require hospitalization.[1] In the UK (population about 60 million) 31 845 people were killed or seriously injured in 2006 due to road accidents and there were 2 58 404 road casualties.

Accident occurrence and psychiatric disorders

On a group basis, lower social classes, subjects with less education and lower intelligence tend to sustain more accidents and injuries (and have higher morbidity and mortality in general). The ratio of males to females for both fatal and non-fatal accidents is about 2:1 in subjects below 60 years of age. Individual variables associated with increased liability of being involved in an accident include antisocial tendencies, aggressiveness, impulsiveness, thrill and adventure-seeking behaviour. Conscious or unconscious intention is not an important explanation of the overall prevalence of accidents or injuries in the society.

Patients with significant psychological problems (psychopathology including substance abuse) sustain more severe injuries than healthy subjects and the prevalence of psychiatric disorders is increased among hospitalized injured adults compared to surgical patients admitted for other reasons. At least 15–20 per cent of persons brought to hospital emergency rooms due to accidental injury have clinical significant blood concentrations of alcohol. Furthermore, patients with schizophrenia, affective illness and post-traumatic stress disorder have more accidental *deaths* (and suicides) compared to the general population.

Physical injuries

Most non-fatal injuries treated in hospitals are minor head concussion and lacerations, strains/sprains, contusions/abrasions and fractures to body parts such as limbs. More severe injuries are mostly related to high energy accidents (e.g. motor vehicle accidents) and often involve both the head and limbs. Injuries to the inner organs are less frequent, but mostly more severe. The anatomical based Abbreviated Injury Scale (AIS) and Injury Severity Score (ISS) are the most widely used classification system of physical injury. Other classification systems based on physiological impact of trauma (e.g. Revised Trauma Score, Glasgow Coma Score) and combinations of anatomical injury and physiological impact (e.g. Trauma and Injury Severity Score) exist as well. See: http://www.trauma.org/archive/scores/ais.html;

Physical injury as psychological trauma

Accidental injury implies several important sources of threat, loss or conflict which may cause psychological distress or psychiatric disorders. The most important accident related variables associated with subsequent psychological problems include,

- Severity of the accident (e.g. real degree of threat to life of one self and others)
- Degree of helplessness
- Duration of the stressor
- Presence and type of actual physical injury
- Exposure to dead and mutilated bodies.

Nevertheless, pre-accident adjustment, personality and the personal meaning of the accident or injury are the strongest predictors

of both acute psychological responses and long-term psychiatric outcome. This observation holds even in the presence of a severe injury,[2] although the type of injury *per se* may influence the short- and long-term outcome. The relative contribution of 'objective' accident related compared to 'subjective' appraisal related variables in shaping the acute response varies. A rule of thumb is that the less severe the accident, the more important are variables not directly related to the accident *per se* (i.e. the personal meaning of the accident and its consequences for the individual.[3] Important individual variables include,[4–6]

- Pre-injury mental health and adjustment problems
- Personality traits (e.g. neuroticism, quality of attachment)
- Trauma history

The accident *per se* may represent a blow to the person's feeling of invulnerability (narcissistic loss). In some, the accident situation may provoke conflictual feelings (e.g. self-blame, survivor guilt) or shame (e.g. own actions or fantasies prior to the situation). Injury to the body may threaten self-esteem and body image; or represent a loss of function. In some cases, the injury may even serve as a primary gain in a psychodynamic sense. The immediate responses will also be influenced by psychological issues like fear of losing control, or the effect of that phenomenon if it occurs. Conflicts related to secondary gains may also influence the clinical response observed by others.

Clinical features and assessment of trauma at the accident scene

The ABC rule of assessment (Airway, Blood pressure, Circulation) should always be the first step in any medical assessment of acute injury followed by physical examination of the thorax, abdomen, head and finally the extremities. However, except for head injuries associated with impaired cognitive function and injuries that significantly interfere with ventilation or cardiovascular function (e.g. agitation due to hypoxia or apathy due to cardiovascular hypotension), the injury it self plays a minor role for the immediate *psychological* responses to trauma.

Early and marked psychophysiological arousal symptoms like (in decreasing frequency) heartbeat, tremor, dry mouth, restlessness, shaking/trembling, weakness in legs, and sweating are common responses to an acute accident. However, the majority of accident victims appear reasonably calm[5] although many have some degree of inner turmoil that may impair the ability to receive, retrieve, and handle information. If behavioural disturbances are seen during the first seconds to minutes, they mimic phylogenetic responses known from all mammals exposed to acute and severe stressful events: flight, freeze, or fight.

(a) Flight response (anxiety, panic)

The patient appears frightened, may scream or cry. Clear cut panic (e.g. overt confusion, bewildered or aimless behaviour or running away), is rather infrequent even during disasters (<1 per cent). Although lowering of blood pressure is not part of the clinical features of panic, panic is often included in the concept of 'shock' used by lay people and media.

Physiological response to physical injury may be misunderstood as flight response. Patients with injury to the thorax hyperventilate and may appear anxious and scared. Cyanosis is not a sign of emotional distress in adults, and hyperventilation should always be considered as sign of respiratory problems needing urgent medical attention (e.g. pneumothorax). Patients with head injury may be confused and bewildered, but they seldom display the open anxiety seen in patients who panic.

Panic with severe behavioural disturbances may threaten the safety of the subject and provoke anxiety in bystanders and other victims who may themselves be afraid to lose control. Thus, whenever possible, patients with strong anxiety or panic should be offered immediate psychological support. Establishing physical (e.g. hand around the shoulder) and verbal contact is important to reduce panic and provide a sense of security and control. Verbal contact may also reveal the subject's real or imagined fears and provide the subject with an alternative way to express their inner turmoil and despair and thus pave the way for more optimal coping and subsequent behavioural control. The subject should be removed from the accident scene, but not left alone. These subjects need to move around and should not be forcefully immobilized. A helper may walk with the patient until he calms down. The exception to this rule is rare instances where the subject's behaviour is completely out of control representing an immediate threat to the physical safety of self or others.

Reuniting family members may reduce anxiety and worries.

Hyperventilation is treated as usual (breathing into a bag to increase the CO_2 -level) combined with physical and verbal contact as described above. It is crucial that somatic causes (e.g. pneumothorax, intoxication) have been ruled out.

(b) Freeze response (apathy)

Freeze responses include halted surprise or in more extreme cases emotional numbness (apathy). Apathy causing lack of appropriate lifesaving activities occur rather infrequently among random samples of accidentally injured adults (less than 10 per cent). In less than 1 per cent, significant parasympathetic (vagus) responses with lowering of blood pressure occurs ('emotional shock'). These patients appear pale and silent. The look of their eyes gives an impression of detached distance, if they were looking onto their own personal world somewhere far away from the actual accident scene. Rarely, an atypical freeze response characterized by blank denial of having sustained an injury when one, in fact, exists may be seen. These subjects may continue to behave as if nothing had happened and not take appropriate precautions at the accident scene.

Several physical injuries may mimic freeze-response. Patients with internal bleedings (e.g. liver, spleen) may appear pale and silent as if in emotional shock ('freeze response'). The pulse is weak and fast (tachycardia), however, in contrast to the vagus tonus induced bradycardia of the freeze response.

If there is a risk of further injury associated with remaining at the accident scene, patients with freeze responses must be removed to a safe place. They should not be left alone, but covered with a jacket or a blanket over their shoulders and attended to in a calm and gentle way, encouraging them to express some of their thoughts and emotions. If the freeze response is severe and prolonged, the patient should be brought to an emergency room for renewed and extended medical evaluation and basic psychological care. Cases of complete denial of having sustained an injury despite evidence for the opposite, should clinically be handled as a freeze response.

(c) Dissociative symptoms

Dissociative symptoms occur in about 15 per cent during the 1st second to minutes after an accident and may be associated with

flight or freeze responses. Brief symptoms of derealization are most common, even in relatively minor accidents (e.g. 'unreal', like a 'dream' or 'slow movie'). Symptoms of depersonalization (e.g. 'I watched my body burn from a distance') are less common and usually signal a more severe psychological response. Brief symptoms of dissociation do not predict later psychiatric problems,[5,7,8] but marked and prolonged dissociative symptoms still present weeks after the accident.[7]

(d) Fight response (aggression)

Fight responses include irritability, anger and more rarely, open aggression. This response is most often seen among bystanders or helpers who feel threatened by the exposure of dead and mutilated bodies. They may quarrel with the rescue team, and sometimes even interfere with the work of police or helpers. Open aggression is rare among victims themselves with the exception of intoxicated victims with severe personality disorders and a few who have sustained severe head injuries (e.g. subdural hematoma, frontal brain contusion).

Irritability and aggressive comments should not be taken personally by the helpers, but interpreted a symptom of helplessness. In most cases, this response is psychological, but impaired behavioural control due to drug or alcohol may be contributing factors. The patient should be treated as being extremely anxious and under high emotional distress. Reuniting with family or significant others if possible may be helpful. Physical activity may reduce aggression. If suitable, simple tasks which require physical movements may be therapeutic ('Can you give me a hand with . . . . .'), but subjects under stress should never be involved in important rescue tasks due to their impaired judgement ability and tendency to act irrationally.

(e) Acute stress reaction

Marked or severe flight, freeze or fight reponses are included in the ICD-10 (F43.0) definition of acute stress reaction (ASR). ASR is defined as immediate onset of marked psychological symptoms (within 1 hour) following exposure to an exceptional mental or physical stressor. The symptoms must begin to decrease after 8 hours if the stressor is transient (e.g. accident). If exposure to the stress continues (e.g. combat zone, hostage situation) the symptoms must begin to diminish after 48 hours. In contrast, the DSM-IV concept 'acute stress disorder' (ASD) describes development of symptoms not earlier than 2 days after the trauma but within one month after exposure.

Psychotropic drugs are seldom needed to treat acute psychological responses at the accident scene if proper medical care including emotional contact from skilled, empathic helpers is offered. Violence towards victims having lost behavioural control may increase the anxiety among other victims and bystanders, and in fact, increase the risk for more behavioural disturbance within the group, and should thus be avoided.

(f) Acute pain

Some injured persons do not report pain complaints during the 1st second to minutes after even severe physical injury, and some may even continue to perform tasks as usual. This response occurs particularly in situations with continous threat to others or own life (e.g. wounded soldiers). This is part of a brief dissociative response which may be life saving and does not reflect psychopathology. However, a few accident victims respond differently. They may report the most painful physical sensations ever experienced. In the absence of severe physical injury, this response most often reflects catastrophic cognitions associated with severe anxiety[5] and should be treated accordingly. Most injured patients report some degree of pain as minutes pass, however.

Severe pain should be treated at the accident scene and will contribute to psychological and physical recovery from the injury.[9] Anxiety and fear may lead to increased pain complaints, so may imagined (!) severe injuries. For those reasons, it is important not only to examine the presence of actual injury, but also explicitly ask the victim if he or she *believes* or *fears* having sustained serious or life-threatening injuries not detected by the medical personnel. If yes, factual information combined with additional proper medical examination if needed, should be provided to reduce the subject's fears and worries. Faced with true life-threatening injuries, the helper should admit facts if asked, but nevertheless provide some hope and cautioned optimism. It is often hard to evaluate true prognosis at the accident scene and advanced trauma surgery may save the life of many severely injured subjects who would have died a few decades ago.

Responses seen in the emergency room

In urban areas, most subjects will be brought to emergency rooms within less than an hour. At that time most victims have started the process of working through the accident, the injury and its implications. This process is reflected in a characteristic cluster of emotions, cognitions and physiological symptoms observed in humans exposed to all types of stressful situations.[10–12]

- Intrusion includes images of the accident popping into the victims mind, and thinking about the accident even when the person do not want to do so. The main load of intrusive symptoms are related to the severity of the accident and the personal meaning. Intrusive symptoms are common both in post-traumatic depression and anxiety.[5]
- Avoidance includes trying not to talk about the accident or avoiding any cognitive or behavioural activities which reminds the person about the accident. Such symptoms and signs are strongly related to accident-independent variables such as personality traits (e.g. coping style) and more often associated with anxiety than depression.[5]
- Hyperarousal includes startle response, strongly increased heart rate, shivering and trembling, irritability, difficulty in concentrating, hypervigilance and disturbed sleep. With the exception of difficulty to sleep, clinically significant hyperarousal is rather infrequent in randomly selected accidentally injured subjects (less than 10 per cent). However, severe hyperarousal signifies a strong physiological and emotional response and is increased among injured compared to non-injured accident victims and is in some studies associated with later post-traumatic distress problems.[13]

The three most common types of behavioural problems seen in the emergency room are,

- Uncontrolled crying or screaming
- Strong anxiety which may include excessive pain complaints
- Aggression and dyssocial behaviour

Crying and anxiety are associated with high levels of intrusion and avoidance, and may be part of ASR. Systematic and carefully

conducted medical examinations accompanied by supporting questions about the patients emotions, thoughts, and fantasies are the most effective way to put the patient at ease. Sedating drugs are seldom needed if the necessary psychological support is provided. Separation from family members or significant others may increase anxiety and despair, and family reunion may be helpful. If symptoms of high arousal persist, prazosin, a central nervous system (CNS) active alpha-1 adrenoreceptor antagonist or a beta-blocker (e.g. 40–60 mg propranolol) or alfa—may be given to attenuate extreme adrenergic tonus.[14]

Aggressive behaviour occurs in about 5 per cent of injured persons brought to hospital, mostly among intoxicated subjects. The presence of head injury must be ruled out. Most cases can be brought under control with the help of significant others and firm, but calm attitude, addressing the fear or helplessness. In a few cases, acute administration of benzodiazepines or a sedating neuroleptic may be necessary. If the patient is intoxicated or suffer from respiration difficulties, neuroleptics may be the safest option. In cases of armed patients, the necessary precautions must be taken.

Psychotic forms of ASR are seldom seen in injured adults and even patients with schizophrenia or other psychotic disorder prior to the accident appear remarkably calm and collected upon arrival in the hospital. If psychosis is present at arrival in the emergency room, influence of psychoactive substances, severe injury (e.g. brain injury, respiratory failure) or a concurrent psychotic disorder must be ruled out.

(a) Whiplash injury

Rear end collision may cause a whiplash like movement of the neck. Biomechanical studies suggest overstretch of cervical facet-joint capsules as a possible source of pain. Neck pain, stiffness or tenderness may occur minutes to hours after the accident. A medical examination including an X-ray of the cervical columna seldom reveals pathological findings (Quebec classification grade I). In more severe cases, distortion and minor bleeding in capsules, ligaments, tendons or muscles (grade II) may lead to additional musculoskeletal signs such as decreased range of motion and point tenderness. In severe injuries, neurological findings (impaired myostatic reflexes, pareses, loss of sensibility, grade III) or even fractures (grade IV) may be present. In patients with whiplash related injury grade I or II, acute psychological distress and associated neck pain is the most important predictor of long-term outcome.

In the emergency room, treatment should aim at providing the subject with adequate information about the good prognosis. Pain after whiplash-injury usually lasts for four-to-six weeks (!), but gradually disappears. In cases of pain without somatic findings, pain killers or antiflogistic medication have uncertain effect and should not be prescribed for more than a week. Sick leave should be avoided or be as short as possible. Mobilization and early return to work is recommended. Overtreatment by physicians or physiotherapists (e.g. application of stiff collar despite no findings of injury to the cervical columna) may lead to permanent illness behaviour and pain-fixation.[15] The optimal physical treatment of whiplash injury is still unsettled,[16] but premorbid pain and psychiatric disorders represent a risk for development of chronic disabling symptoms and should be treated.

(b) Significant others' needs

Relatives or survivors may want to see dead significant others brought to hospital, and touch them. This process helps the relatives to work through the traumatic event and should be encouraged. If the dead body is grotesquely disfigured, the most horrifying parts should be covered prior to exposure. In any case, a physician or a skilled nurse should accompany the relatives during exposure. Small childrens' emotional response to dead bodies mirror the adults' response. Accordingly, reducing the anxiety and fear of the adults is the best way to help children cope with dead ones. Correspondingly, in cases of severe anxiety in accompanying small children, addressing the helplessness and anxiety of the parents is important. If dead bodies are stored in hospital chapels, care must be taken to cover the presence of religious symbols incongruent to the religious status of the dead one and his family (e.g. Christian crosses should be covered in case of a Jew or a Muslim). The reader is referred to chapter 4.16 for more information on culture specific responses to stress and trauma.

In disaster situations, the need for information varies among relatives, depending on whether their loved ones are missing, injured, or dead (survivor status). Those who have lost loved ones often want to talk to rescuers or get information with regard to any hint about the emotional status of the dead one at the time of death. Accordingly, in situations with several hundred relatives come to the hospital, information is provided in separate groups according to the significant other survivor status. The logistics of such procedures should be outlined in the hospital's disaster plan.

Psychiatric treatment during hospital stay

Most studies indicate that risk factors, emotional, and behavioural responses correspond to that of medically ill patients and identifying those who are at increased risk can follow the same guidelines as for medicine in general.[17] Some patients may complain about physical symptoms suggesting undetected injury. Such complaints may in fact be true. If not addressed and attended to, psychological distress presented by means of somatic complaints or symptoms is the rule.

Clinical syndromes requiring psychiatric attention during hospital stay are listed in Table 5.3.8.1. Complete denial of severity of injury or avoidant coping is maladaptive and should be counteracted.[18] Relatives or significant others should be contacted. They may convey unrealistic fears—or hopes (e.g. 'you will be able to walk'—attitudes in patients with permanent paralysis of legs)—which strongly influence the behaviour and emotional well-being of the patient. They may also provide information which may be helpful to understand current behaviour (e.g. previous dysfunction, 'silent' delirium undetected by staff).

(a) Anxiety and acute stress reaction

Worrying and compulsive thoughts about the accident or the injury (intrusion) is seen both in anxiety and depression. Extreme anxiety may infrequently lead to cardiovascular complications (e.g. pulmonary embolia) in subjects with cardiovascular risk factors (e.g. elderly subjects often smokers with hypertension and arterosclerosis).

Sleep problems may be present or related to physical pain and treatment procedures. The aetiology of nightmares following traumatic injury is complex.[19] They mostly emerge a couple of days after the accident and disappear gradually. Persisting nightmares for more than two weeks without any signs of mastery in the dream content, suggest development of post-traumatic stress disorder and should thus be treated.

Table 5.3.8.1 The most frequently seen psychiatric syndromes during hospital stay following accidental injury

Type of syndrome or clinical problem	Clinical symptoms and signs	Comment
Delirium	Confused, strange behaviour; episodic disorientation; irritability; episodic fearful look	May occur without obvious signs of agitation if sedated; Relatives may detect it and be upset. Diagnosis: 'Draw a clock test' helpful
Abstinence from drug or alcohol	increased pulse; sweating; tremor; insomnia; agitation; anxiety; nausea; abdominal pain; dysphoric.	May be interpreted falsely as accident-provoked anxiety
Antisocial personality, histrionic or borderline personality disorder	Aggressive behaviour; poor compliance with treatment; abusive language; high demand for analgesics	Undetected brain dysfunction must be ruled out; Relatives may provide important pre-injury information
Hypomanic or manic responses	Elated mood; emotions do not correspond to the situation; uncritical behaviour	Undetected brain dysfunction; bipolar disorder or hypomanic response as a defence against survivor guilt
Anxiety	Tense, anxious, restless, worrying, increased startle reflexes; insomnia; dissociative symptoms may occur.	Prolonged or delayed stress response or disorder; imagined or real threat from accident or injury; physical complication (e.g. hypoxia, delirium); abstinent or side effects of drug. Obsessive-compulsive traits and high inner tension with fear of losing emotional control. If unexplained, consult relatives for psychological clues.
Depression	Withdrawn, loss of appetite; inability to feel; sad; worrying; passive; lassitude; anxiety symptoms frequent	Grief; survivor guilt; psychological response to disfigurement or loss (real or imagined) of function or self-esteem; reactivating of previous painful memories
Medically unexplained physiological events including delayed healing of wounds	Rare phenomena; typical senior physicians statement 'I've never seen something like it before'	Secondary gain by extended hospital stay (e.g. alternative prison); factitious disorder; extreme stress (psychophysiological activation)
Excessive pain	Complaints of pain; poor sleep and appetite; poor performance; do not reveal emotions.	Undetected physical complication; insufficient pain treatment; anxiety or depression response in past with obsessive-compulsive traits; withdrawal syndrome; drug abuser.
Partial or complete denial of actual injury	As-if-nothing-has-happened behaviour; refuse treatment; request early dismissal from hospital	Undetected brain dysfunction; psychotic disorder. If male and partial denial, consider obsessive-compulsive traits and high inner tension with fear of passivity

Psychological interventions should be based on clear indication and be brief, distress focused and time limited. Symptoms of intrusion including nightmares may be treated by simple psychological techniques. If one specific traumatic event which can be delineated (e.g. visual image of a traumatic moment), psychological video replay techniques (VRT) may be useful. The subject is taught how to relax. Subsequently, the subject reviews the pre-accident and accident situation on an imagined (i.e. mental) video screen. When the anxiety rises to unacceptable levels, the subject is asked to push the (imagined) stop button and press fast replay until a pre-accident situation where the subjects is at ease is reached. When calm, the procedure is repeated, until the subject can view the whole accident without strong anxiety.

If the anxiety level is high or the traumatic event is more complex, more comprehensive interventions are needed, e.g. Eye Movement Desensitization and Reprocessing (EMDR) or Trauma-Focused Cognitive Behavioural Therapy (TF-CBT).[20] If strong anxiety is not brought under control by means of psychotherapy or other behavioural techniques (e.g. applied relaxation), psychoactive drugs can be added. In injured patients with childhood trauma or other traumatic events in the past, a selective serotonin reuptake inhibitor (SSRI) probably should be the first choice. Sleep problems are best dealt with by environmental adjustment whenever possible, or by optimal pain control if appropiate. Sedative drugs are secondary option. Mianserin or mirtazapine combine sedative and antidepressant effects and may be alternative to benzodiazepines.

Acute stress disorder (ASD) may be conceptualized as an acute form of PTSD and may predict chronic PTSD.[21] The main treatment is psychotherapy (i.e. EMDR, TF-CBT). In PTSD, prazosin reduces nightmares and sleep disturbance in placebo-controlled studies, and may thus be an option also during the first days to weeks after trauma. Some studies have reported propranolol given within days following a traumatic event to be useful for mitigating PTSD symptoms or perhaps even preventing the development of PTSD. The mechanism is thought to be explained by reduced consolidation of emotional memory. Small doses of glucocorticoids may reduce traumatic memories in ASD as well,[22] but larger controlled studies are needed to verify this finding. Benzodiazepines may also reduce acute distress, but may not reduce the risk of 6-month psychiatric anxiety problems. In conclusion, all psychopharmacological treatments must be provided together with psychological interventions addressing the key psychological sources of distress and worry.

(b) Depression

Depressed mood during the first days to weeks following an accident are mostly due to guilt, shame, or grief due to real or

imagined losses. The key to understanding the response is the meaning of the accident or the injury for the patient. Guilt, shame or rumination over real or imagined losses is associated with long-term problems[23] and may require specific therapy.[24] Premorbid causes of depression (e.g. bereavement, mood disorder at the time of the accident), must be kept in mind. In patients with immobilizing injuries staying in hospital for an extended period, some degree of depressive symptoms is the rule. If the symptomatology is severe and persistent, antidepressants are indicated.

Persisting depression has been found to be highly predictive of a long-term psychiatric consequence, and moderate to severe depression predicts less likeliness of returning to preaccident functional level. Thus, depression should always be taken seriously. If marked depression presist for more then 2–3 weeks, an antidepressant should be given. Due to lower incidence of side effects, the newer low-toxic antidepressants are preferred. Delirinm and abstinence are treated as usual. Detailed presentations of the psychopharmacology of the injured or medically ill are available.[34–36]

(c) Pain

Both the injury itself (e.g. burn injuries, injuries to the pelvis, penetrating traumas) or medical treatment (e.g. physiotherapy to prevent contractures, ICU) may be associated with psychological distress and pain.[25,26] Pre-accident psychopathology increases the prevalence and severity of pain complaints. Pain or fear of bringing about pain leads to diminished movement, which can engender contractures, muscle atrophy, and bed ulcers. Traumatic amputation may be associated with phantom pain and exacerbation of pain in response to imagined movements has been reported in subjects with spinal cord injury.[27] Poorly treated pain can be demoralizing to patients and provoke psychological regression, giving-up responses and long-term psychiatric problems[28] including increased risk for suicide.[29] Thus active pain control is crucial and may reduce the prevalence of long-term suffering.[30] Comprehensive reviews of the psychological care of burned subjects are available.[31]

Concerns that trauma patients with injury related pain will become addicted if treated properly with analgesics is neither supported by clinical experience nor by empirical data.[32] However, co-morbid psychiatric disoders must be taken into account when treating injury-associated pain. Patients with a history of substance abuse may have greater tolerance to analgesics and will have to be titrated to higher doses.[33]

(d) Confusion and psychoses

In civilian life situations, confusion or psychotic responses appearing for the first time days after being admitted to hospital are almost exclusively due to a central nervous system dysfunction. Risk factors are severe injuries (ISS >15; third degree burn injury) and major head injuries (e.g. contusion). Impaired cognitive functions or drug or alcohol abuse prior to accident and age >50 increase the risk of organic mental dysfunction.

In a few cases, psychotic-like confusional and agitated responses occurring during the hospital stay may be due to abrupt disruption of intake of psychotropic medication taken for long periods prior to the accident. If in doubt about the aetiology of a psychotic response seen during the first days to weeks (including severe manic episodes), the psychiatrist should consider the response to be of organic origin and explore the pathophysiological processes as done in cases of delirium. Asking the patient to draw a clock, may be a simple and effective way of detecting organic dysfunction.[37]

(e) Alcohol and drug abuse

A significant number of accident victims brought to hospital have an alcohol or drug problem, and symptoms of abstinence may be misinterpreted as psychological anxiety.[38] In cases of grossly deviant behaviour, the presence of co-morbid severe personality disorder should be considered. Alternative explanations include delirium or side effects to drugs (e.g. steroids).

(f) Psychological needs of rescue personnel and staff

Debriefing is a psychological treatment intended to reduce the psychological morbidity that may arise after exposure to accident or injury. Debriefing involves promoting emotional processing/ catharsis or ventilation by systematically encouraging recollection/ ventilation/ reworking of the traumatic event. There is no evidence that this method reduces the incidence of post-accident problems in civilian life.[20,39] Accordingly, psychiatric intervention in the emergency room should be limited to those who display acute and severe psychiatric disorders.

Rescue personnel and medical staff may be psychologically affected by sudden exposure to grotesquely mutilated bodies.[40] The same individual vulnerability found in injured subjects apply. Group debriefing has been recommended if the rescue operation was extremely difficult; there were many dead or there was explicit harsh critique of the rescue operation from the media. Debriefing offered and conducted by a respected senior member of the rescue team is probably more appropriate than debriefing offered by psychologists or psychiatrists. However, empirical data regarding efficacy of emotionally focused group debriefing is scarce.

Long-term behavioural and psychiatric consequences of physical trauma

Physical injury may cause permanent physical change including neurological dysfunction,[41] impaired physical function, changes in perceived somatic health including pain,[30] decreased capacity to work (in children: play), decreased social contact and decreased leisure pleasure.[42] In an unselected population of hospitalized accidentally injured adults, about half will report some complaints three years later. Among those with most severe injuries (ISS >15), only 1/3 will have made full recovery after three years, and about half will report at least moderate disabilities.[43]

The prevalence of non-organic mental disorders among hospitalized adults is about 20 per cent after six months and 10 per cent after two years.[29, 44–49] Depressive symptoms and disorder are most frequently seen followed by specific accident-related phobia and PTSD. Subsyndromal PTSD-cases must be added to these numbers. PTSD is associated with several physical health problems including cardiovascular diseases, respiratory diseases, chronic pain conditions, gastrointestinal illnesses, and cancer.[12] The prevalence of alcohol and drug abuse is increased as well.

The prevalence of long-term psychiatric is increased in injuries associated with visual disfigurement, loss of body parts or physical function (e.g. spinal cord injury), neck injuries and injuries to the pelvis and genital areas.[42] Chronic pain following accidental injury is often associated with concommittant mental disorders, in particular mood disorders or PTSD.[50] Man-made accidental injury

(e.g. assault, combat, rape, terrorism) cause more long-term mental problems than other types of accidental injury (e.g. natural disasters). Studies comparing outcomes in *men versus women* have been mixed. Current evidence suggests that women are at higher risk for anxiety and depression, and men are more at risk for substance abuse and antisocial behaviour.

Following extreme psychological and physical trauma (e.g. torture, concentration camp survivors, hostage situations), permanent change in the person's pattern of perceiving, relating to, and thinking about the environment and the self may occur (ICD-10 F62.0: Enduring personality change after catastrophic experience). The changes should not fully be explained by the presence of PTSD. This diagnostic category does not exist in the DSM-IV.

Assessment of long-term psychiatric consequences of traumatic injury

The following key-points need to be explored when evaluating long-term effects of traumatic injury

- Social and cognitive resources (including social support)
- History of mental disorder, social dysfunction or trauma in the past
- Overlooked physical injury (increased risk if high energy accident or severe injury, e.g. undetected frontal brain damage or other neurological injury)
- Deviant behaviour or accident-related psychiatric disorders following the accident (including ASR or ASD)
- Painful treatment procedures
- Accident-independent traumatic life-events during the post-injury period
- Current psychiatric disorders

Patients, relatives and physicians may evaluate long-term problems differently[51] and psychiatric co-morbidity is prevalent. Thus the clinical assessment should be supplemented by a systematic screening for the most common psychiatric disorders (e.g. MINI neuropsychiatric interview) and cognitive, behavioural and quality-of-life issues (e.g. Impact of Event scale, General Health Questionnaire). Questionnaires specifically designed to address physical, emotional and social outcome of accidental injury are available.[42] A proper, complaint-focused medical examination is often necessary as well.[52] The psychiatrist may improve the quality of the medical examination by providing the examining physician with specific diagnostic questions based on information of the patient's trauma history and symptom complaints.

Treatment of long-term problems

Psychiatric disorders occurring in the aftermath of injury are treated according to general treatment guidelines of mental disorders with some modifications. EMDR and TF-CBT are the best validated psychotherapeutic interventions for trauma related PTSD.[20] If the psychological themes are related to conflicts, family issues or secondary events, short-term psychotherapy as outlined by Horowitz and his group[10] may be conducted. Body-focused treatments may be helpful in some subjects with chronic pain problems after trauma.[53] Randomized controlled treatment trials of accidentally injured adults with post-injury psychosomatic and psychiatric problems are few, however. Comprehensive treatment of patients may provide better results than intervention performed by one single professional only.[54]

Antidepressants should be given in cases of mood disorders or PTSD not responding to psychotherapeutic interventions alone. SSRIs and related drugs are first choice. Drugs acting on nor-adrenalin reuptake alone (e.g. atomoxetine, reboxetin) may increase anxiety and should be avoided. Psychopharmacological treatment of somatoform pain disorders should target both serotonin and nor-adrenalin (e.g. amitryptyline, chlomipramine, duloxetine, venlafaxine). In cases of chronic PTSD with high level of intrusive symptoms, prazosin or propranolol may be added. Betablockers may be valuable as a supplement to anxiety provoking exposure therapy. Benzodiazepines may reduce PTSD-related anxiety, but differences in modulation of skin conductance compared to patients with panic disorder support clinical experience that drug treatment should be supplemented with psychological interventions in order to achieve optimal results. Guidelines for psychopharmacological treatment in patients with co-morbid physical disorders exist (e.g.[36]).

Compensation claims and litigation

Most accidentally injured subjects do not exaggerate their loss,[55] and in non-litigant situations malingering is an unlikely explanation in most cases of chronic disturbances after accidents. Neither is economical settlement followed by significant change in clinical situation in most cases. However, in litigation situations, the patient's problem report may sometimes be exaggerated or even invalid. Studies of personal injury plaintiffs indicate that a significant number report pre-injury functioning superior to that of controls, and malingering has been estimated to 20–30 per cent.

Studies consistently show that delayed-onset PTSD in the absence of any prior symptoms is rare, whereas delayed onsets that represent exacerbations or reactivations of prior symptoms may occur.[56] Untrained subjects are able to endorse symptoms on checklists to meet criteria for diagnoses of major depression, PTSD and GAD, and PTSD self-report measures cannot be used for diagnosis.[57] Furthermore, intrusive symptoms are not PTSD-specific and may be significant in depression as well.[5] This fact is often neglected which explains why some expert testimonies misinterpret depression as being PTSD.

The physician should always try to get patient-independent information from reliable sources (e.g. medical records, general practitioners) also related to pre-injury function[58] before concluding about long-term problems due to physical injury. The possibility that clinically significant brain injury or non-injury related illnesses or psychiatric disorders occurring after the injury, have been overlooked during the medical evaluation part of litigation and compensation cases must be kept in mind.

There is no evidence that physical injury provoke *de novo* bipolar disorder or disorganized schizophrenia, even among severely maltreated subjects.[59] However, permanent injury to frontal and temporal lobes of the brain may provoke manic episodes, paranoid psychoses with schizophrenic-like symptomatology and chronic depression. Both in clinical and court settings, such brain dysfunctions may be overlooked due to lack of classical neurological signs.

Expert testimony should be based on the best available evidence and standards of care, which requires that experts stay current in their field of expertise, and revise old opinions as new information is published. Personal experience alone is rarely sufficient. The psychological difficulties and challenges faced by an expert witness is discussed elsewhere.[60]

Further information

Web links

Physical injury scoring systems

This web link provides access to description and on-line calculation of physical injury trauma scores and links to other resources about physical trauma.

http://www.trauma.org/archive/scores/ais.html

Psychiatric treatment

Several databases providing information about treatment in medicine, including psychiatry and psychosomatic medicine are available. These databases include systematic reviews metaanalysis, clinical trials, and more, including both psychological and biological interventions.

The Cochrane Library

http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME

NICE (National Institute for Health and Clinical Excellence):

http://www.nice.org.uk/).

The limitation of such databases is infrequent updates. In areas with limited research, the conclusions reported may be outdated even shortly after they are published. Thus these databases cannot replace continuous updates from databases of original research like:

PubMed http://www.ncbi.nlm.nih.gov/sites/entrez

PsychInfo.http://www.apa.org/psycinfo/

or National center for post-traumatic stress disorder database with information about treatment of PTSD and traumatic stress for Mental Health Care Providers. http://www.ncptsd.va.gov/ncmain/index.jsp

Other useful web links

Practice guidelines for psychiatric consultation in the general medical settings provided by the

Academy of Psychosomatic Medicine: http://www.apm.org/prac-gui/psy 39-s8.shtml

The European association for consultation-liaison psychiatry and psychosomatics publishes power-point presentations about different aspects of psychiatry in the medical ill or injured: http://www.eaclpp.org/

References

1. Zatzick, D.F., Rivara, F.P., Nathens, A.B., *et al.* (2007). A nationwide US study of post-traumatic stress after hospitalization for physical injury. *Psychological Medicine*, **37**(10), 1469–80.
2. Zatzick, D.F., Grossman, D.C., Russo, J., *et al.* (2006). Predicting posttraumatic stress symptoms longitudinally in a representative sample of hospitalized injured adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry* **45**(10), 1188–95.
3. O'Donnell, M.L., Elliott, P., Wolfgang, B.J., *et al.* (2007). Posttraumatic appraisals in the development and persistence of posttraumatic stress symptoms. *Journal of Traumatic Stress*, **20**(2), 173–82.
4. Krupnick, J.L. and Horowitz, M.J. (1981). Stress response syndromes. Recurrent themes. *Archives of General Psychiatry*, **38**(4), 428–35.
5. Schnyder, U. and Malt, U.F. (1998). Acute stress response patterns to accidental injuries. *Journal of Psychosomatic Research*, **45**(5), 419–24.
6. Creamer, M., McFarlane, A.C., Burgess, P., *et al.* (2005). Psychopathology following trauma: the role of subjective experience. *Journal of affective disorders*, **86**(2–3), 175–82.
7. Murray, J., Ehlers, A., Mayou, R.A., *et al.* (2002). Dissociation and post-traumatic stress disorder: two prospective studies of road traffic accident survivors. *The British Journal of Psychiatry*, **180**, 363–8.
8. Wittmann, L., Moergeli, H., Schnyder, U., *et al.* (2006). Low predictive power of peritraumatic dissociation for PTSD symptoms in accident survivors. *Journal of Traumatic Stress*, **19**(5), 639–51.
9. Saxe, G., Stoddard, F., Courtney, D., *et al.* (2001). Relationship between acute morphine and the course of PTSD in children with burns. *Journal of the American Academy of Child and Adolescent Psychiatry*, **40**(8),915–21.
10. Horowitz, M.J. (1997). *Stress response syndromes: PTSD, grief and adjustment disorders* (3rd edn.). Jason Aronson, Northvale, New Jersey.
11. Skari, H., Malt, U.F., Bjornland, K., *et al.* (2006). Prenatal diagnosis of congenital malformations and parental psychological distress–a prospective longitudinal cohort study. *Prenatal Diagnosis*, **26**(11), 1001–9.
12. Sareen, J., Cox, B.J., Stein, M.B., *et al.* (2007). Physical and mental comorbidity, disability, and suicidal behavior associated with posttraumatic stress disorder in a large community sample. *Psychosomatic Medicine* , **69**(3), 242–8.
13. O'Donnell, M.L., Creamer, M., Elliott, P., *et al.* (2007). Tonic and phasic heart rate as predictors of posttraumatic stress disorder. *Psychosomatic Medicine*, **69**(3), 256–61.
14. Schelling, G. (2007). Post-traumatic stress disorder in somatic disease: lessons from critically ill patients. *Prog Brain Res*, **167**, 229–37.
15. Malleson, A. (2002). *Whiplash and other useful illnesses*. McGill-Queens University Press, Montreal.
16. Verhagen, A.P., Scholten-Peeters, G.G., van, W.S., *et al.* (2007). Conservative treatments for whiplash. *Cochrane database of systematic reviews*, (2):CD003338.
17. Huyse, F.J., and Stiefel, F.C. (2006). Integrated care for the complex medically ill. pp. 1-767. Philadelphia, Saunders. *Medical Clinics of North America*.
18. Dougall, A.L., Ursano, R.J., Posluszny, D.M., *et al.* (2001). Predictors of posttraumatic stress among victims of motor vehicle accidents. *Psychosomatic Medicine*, **63**(3), 402–11.
19. Phelps, A.J., Forbes, D., Creamer, M., *et al.* (2007). Understanding posttraumatic nightmares: An empirical and conceptual review. *Clinical psychology review*.
20. Bisson, J.I., Ehlers, A., Matthews, R., *et al.* (2007). Psychological treatments for chronic post-traumatic stress disorder. Systematic review and meta-analysis. *British Journal of Psychiatry*, **190**, 97–104.
21. Meiser-Stedman, R., Yule, W., Smith, P., *et al.* (2005). Acute stress disorder and posttraumatic stress disorder in children and adolescents involved in assaults or motor vehicle accidents. *American Journal of Psychiatry*, **162**(7), 1381–3.
22. de Quervain, D.J.F. (2008). Glucocorticoid-induced reduction of traumatic memories: implications for the treatment of PTSD. In *Progress in Brain Research* (eds. E.R. de Kloet, M.S. Oitzl, E. Vermetten), pp. 239–47. Elsevier.
23. Kleim, B., Ehlers, A., Glucksman, E., *et al.* (2007). Early predictors of chronicpost-traumatic stress disorder in assault survivors. *Psychological Medicine* , **37**(10), 1457–67.
24. Speckens, A.E., Ehlers, A., Hackmann, A., *et al.* (2007). Intrusive memories and rumination in patients with post-traumatic stress disorder: a phenomenological comparison. *Memory,* **15**(3), 249–57.
25. Richter, J.C., Waydhas, C., Pajonk, F.G., *et al.* (2006). Incidence of posttraumatic stress disorder after prolonged surgical intensive care unit treatment. *Psychosomatics*, **47**(3), 223–30.
26. Berben, S.A., Meijs, T.H., van Dongen, R.T., *et al.* (2007). Pain prevalence and pain relief in trauma patients in the Accident & Emergency department. *Injury*.

27. Gustin, S.M., Wrigley, P.J., Gandevia, S.C., *et al.* (2008). Movement imagery increases pain in people with neuropathic pain following complete thoracic spinal cord injury. *Pain.* **137** (2), 237–24.
28. Norman, S.B., Stein, M.B., Dimsdale, J.E., *et al.* (2007). Pain in the aftermath of trauma is a risk factor for post-traumatic stress disorder. *Psychological Medicine* , **10**, 1–10.
29. Edwards, R.R., Magyar-Russell, G., Thombs, B., *et al.* (2007). Acute pain at discharge from hospitalization is a prospective predictor of long-term suicidal ideation after burn injury. *Archives of Physical Medicine and Rehabilitation,* **88**(12 Suppl 2), S36–S42.
30. Castillo, R.C., Mackenzie, E.J., Wegener, S.T., *et al.* (2006). Prevalence of chronic pain seven years following limb threatening lower extremity trauma. *Pain*, **124**(3), 321–9.
31. Van Loey, N.E., and Van Son, M.J. (2003). Psychopathology and psychological problems in patients with burn scars: epidemiology and management. *American Journal of Clinical Dermatology*, **4**(4), 245–72.
32. Abdi, S., and Zhou, Y. (2002). Management of pain after burn injury. *Current Opinions in Anaesthesiology*, **15**(5), 563–7.
33. Alford, D.P., Compton, P., Samet, J.H., *et al.* (2006). Acute pain management for patients receiving maintenance methadone or buprenorphine therapy. *Annals of Internal Medicine*, **144**(2), 127–34.
34. Simon, A., and Gorman, J. (2004). Psychopharmacological possibilities in the acute disaster setting. *The Psychiatric clinics of North America,* **27**(3), 425–58.
35. Fleminger, S., Greenwood, R.J., Oliver, D.L., *et al.* (2006). Pharmacological management for agitation and aggression in people with acquired brain injury. *Cochrane Database of Systematic Reviews*, **4**: CD003299.
36. Malt, U.F., Llody, G.G. (2007). Psychopharmacological treatment in liaison psychiatry. In (eds. G.G. Lloyd, E. Guthrie), pp. 763–94. Cambridge University Press, Cambridge.
37. Royall, D.R., Cordes, J.A., Polk, M., *et al.* (2006). CLOX: an executive clock drawing task. *Journal of Neurology, Neurosurgery, and Psychiatry*, **64**(5), 588–94.
38. Moss, M., and Burnham, E.L. (2006). Alcohol abuse in the critically ill patient. *Lancet,* **368**(9554), 2231–42.
39. Rose, S., Bisson, J., Wessely, S., *et al.* (2003). A systematic review of single-session psychological interventions ('debriefing') following trauma. *Psychotherapy and Psychosomatics,* **72**(4), 176–84.
40. Benedek, D.M., Fullerton, C., Ursano, R.J., *et al.* (2007). First responders: mental health consequences of natural and human-made disasters for public health and public safety workers. *Annual review of public health,* **28**, 55–68.
41. Gurvits, T.V., Gilbertson, M.W., Lasko, N.B., *et al.* (1997). Neurological status of combat veterans and adult survivors of sexual abuse PTSD. *Annals of the New York Academy of Sciences* , **821**, 468–71.
42. Malt, U.F. (1994). Traumatic effects of accidents. In: *Individual and community responses to trauma and disaster*. (eds. R.J. Ursano, B.G. McCaughey, and C.S. Fullerton), pp. 103–35. Cambridge University Press, Cambridge.
43. Schnyder, U., Moergeli, H., Klaghofer, R., *et al.* (2001). Incidence and prediction of posttraumatic stress disorder symptoms in severely injured accident victims. *American Journal of Psychiatry*, **158**(4), 594–9.
44. Bryant, B., Mayou, R., Wiggs, L., *et al.* (2004). Psychological consequences of road traffi c accidents for children and their mothers. *Psychological Medicine*, **34**(2), 335–46.
45. Creamer, M., O'Donnell, M.L., Pattison, P., *et al.* (2004). The relationship between acute stress disorder and posttraumatic stress disorder in severely injured trauma survivors. *Behaviour Research and Therapy*, **42**(3), 315–28.
46. Fann, J.R., Burington, B., Leonetti, A., *et al.* (2004). Psychiatric illness following traumatic brain injury in an adult health maintenance organization population. *Archives of General Psychiatry*, **61**(1), 53–61.
47. Grieger, T.A., Cozza, S.J., Ursano, R.J., *et al.* (2006). Posttraumatic stress disorder and depression in battle-injured soldiers. *American Journal of Psychiatry* , **163**(10), 1777–83.
48. Glynn, S.M., Shetty, V., Elliot-Brown, K., *et al.* (2007). Chronic posttraumatic stress disorder after facial injury: a 1-year prospective cohort study. *Journal of Trauma* , **62**(2), 410–8.
49. Hoge, C.W., Terhakopian, A., Castro, C.A., *et al.* (2007). Association of posttraumatic stress disorder with somatic symptoms, health care visits, and absenteeism among Iraq war veterans. *American Journal of Psychiatry*, **164**(1), 150–3.
50. Geisser, M.E., Roth, R.S., Bachman, J.E., *et al.* (1996). The relationship between symptoms of post-traumatic stress disorder and pain, affective disturbance and disability among patients with accident and non-accident related pain. *Pain,* **66**(2–3), 207–14.
51. Biddle, D., Elliott, P., Creamer, M., *et al.* (2002). Self-reported problems: a comparison between PTSD-diagnosed veterans, their spouses, and clinicians. *Behaviour Research and Therapy*, **40**(7), 853-65.
52. Frenisy, M.C., Benony, H., Chahraoui, K., *et al.* (2006). Brain injured patients versus multiple trauma patients: some neurobehavioral and psychopathological aspects. *Journal of Trauma*, **60**(5), 1018–26.
53. Haugstad, G.K., Haugstad, T.S., Kirste, U.M., *et al.* (2006). Mensendieck somatocognitive therapy as treatment approach to chronic pelvic pain: results of a randomized controlled intervention study. *American Journal of Obstetrics and Gynecology*, **194**(5), 1303–10.
54. Zatzick, D., Roy-Byrne, P., Russo, J., *et al.* (2004). A randomized effectiveness trial of stepped collaborative care for acutely injured trauma survivors. *Archives of General Psychiatry*, **61**(5), 498–506.
55. Bryant, B., Mayou, R., Lloyd-Bostock, S., *et al.* (1997). Compensation claims following road accidents: a six-year follow-up study. *Medicine, Science, and the Law,* **37**(4), 326–36.
56. Andrews, B., Brewin, C.R., Philpott, R., *et al.* (2007). Delayed-onset posttraumatic stress disorder: a systematic review of the evidence. *American Journal of Psychiatry* , **164**(9), 1319–26.
57. Sumpter, R.E., and McMillan, T.M. (2005). Misdiagnosis of post-traumatic stress disorder following severe traumatic brain injury. *British Journal of Psychiatry*, **186**, 423–6.
58. Duckworth, M.P., and Iezzi, T. (2005). Chronic pain and posttraumatic stress symptoms in litigating motor vehicle accident victims. *Clinical Journal of Pain*, **21**(3), 251–61.
59. Eitinger, L., and Strøm, A. (1973). Mortality and morbidity after excessive stress. Humanities press, New York.
60. Gutheil, T.G., and Simon, R.I. (2005).Narcissistic dimensions of expert witness practice. *Journal of the American Academy of Psychiatry and the Law*, **33**(1), 55–8.

5.4

Obstetric and gynaecological conditions associated with psychiatric disorder

Ian Brockington

Introduction

This chapter covers the psychiatry of menstruation, various manifestations of the desire for children (such as surrogate pregnancy and pseudocyesis), pregnancy and mental health, the psychopathology of parturition, infant loss, postpartum psychiatric disorders, the mother–infant relationship and infanticide.

The psychiatry of menstruation

It has long been realized that menstruation and mental illness are linked. As early as 1827 menstrual mood disorder was used as a defence in filicide.[1] In the 1850s, Brière de Boismont[2] and Schlager[3] carried out the first surveys showing that 20–30 per cent of women suffered a mood disorder before or during the menses—usually irritability or depression, occasionally euphoria. There are descriptions of a wide variety of deviant behaviours, including nymphomania, food cravings, binge drinking, pathological lying, shoplifting, and fire-setting, as well as suicide, violence, homicide, and morbid jealousy. There are other nervous diseases associated with menstruation, including epilepsy, migraine, and hypersomnia.

Recently, there has been much research into the biological basis and treatment of 'premenstrual tension' (or its synonyms). A number of daily rating schedules have been published, but self-devised rating scales, tailored to an individual patient's symptoms, can be used, provided that they are carefully completed every day. Scientific studies are bedevilled by difficulties in defining the disorders.[4] It is not known whether this is one syndrome or many. Irritability is striking, but otherwise the symptoms are common to many other disorders.

Although little is known about the aetiology, progress has been made in treatment. There may be a response to serotonin-reuptake-inhibiting antidepressants (e.g. fluoxetine, chlomipramine). In so far as a luteal-phase defect may be a factor, ovulation-promoting drugs such as clomiphene can be tried. The synthetic steroid danazol, and the gonadorelin agonists (which suppress menstruation), are draconian treatments for severe cases. All interventions should be prescribed in the context of a long-term study using daily ratings.

Rarely, menstruation is linked to a psychosis with acute onset, brief duration, and full recovery. Premenstrual, catamenial, paramenstrual, mid-cycle, and 'epochal' variants have been described.[5] Menstrual psychosis is rare, but perhaps not excessively so. There is a clustering of episodes around puberty and after childbirth, although only a small proportion of menstrual cycles are involved. There are sufficient case reports from Japan, India, and Islamic countries to suggest a worldwide disorder. This is not a specific entity, and most typical examples manifest non-menstrual bipolar disorder at another stage of life. Clinically, it resembles puerperal psychosis. The close relationship between these two psychoses is emphasized by women who develop puerperal and menstrual psychosis at different times.[6] A Japanese investigation showed an association with anovulatory cycles.[7] Pregnancy has a beneficial effect, and there are claims of successful treatment with oral contraceptives, progesterone, clomiphene, danazol, and gonadorelins. The basis for intervention is a long-term study, with a good baseline and exact timing of events in relation to the menstrual cycle.

Infertility

Motherhood is among the strongest and most universal of motivations. For many infertile women, childlessness is the most upsetting experience of their lives, and the yearning for children dominates everything. Infertility is stigmatizing, especially in some cultures. Infertile couples often suffer from self-reproach over sexual indiscretions, abortions, or contraception. They envy fertile couples, and contacts with other people's children, family celebrations, and friends' pregnancies are problematic. The security of the marriage may be threatened by the fear that the spouse will desert to a fertile partner; nevertheless, the marriages themselves are often happy.

Infertility differs from other stresses in its duration. The psychological reaction unfolds over years. When treatment begins, there is a cycle of optimism and hope, with a build-up of tension towards the end of the cycle, followed by disappointment and despair. Sexual functioning comes under strain during the investigation, and the discovery of azoospermia is especially stressful. There is some evidence that stress affects conception, though more prospective studies are needed.

Assisted reproduction

Artificial insemination (using the husband's or partner's semen) has been available from the late eighteenth century, and donor insemination since 1884. Its psychological effects on marriage seem minimal; husbands or partners rarely react with jealousy to the

baby, any more than to an adopted child. The proof that the experience is acceptable is that it is often repeated. One of the principles is privacy, ensuring that donor and couple never meet and remain ignorant of each other's identity. It is felt that violating anonymity might compromise the marriage, since donor and mother are too deeply involved in procreation to regard their relationship with detachment; but times may be changing. The interests of the children have to be considered; donor insemination obscures the genetic lineage, and the child cannot benefit from advances in genetics.

In vitro fertilization (IVF) was first performed in 1978, and was achieved with a donated oocyte in 1984; it is now widely used—in Holland, 1/60 babies are born by IVF. The procedure is harrowing, and counselling is mandatory. There is an increase in multiple births, which are more stressful. But the quality of parenting may be superior to that of families with naturally conceived children.

Surrogate motherhood

This has two meanings:

- A woman is inseminated (artificially or naturally) with the husband's or partner's semen, and surrenders the child to the genetic father and adoptive mother. The surrogate provides oocyte and womb, and is a substitute spouse.
- The wife donates a fertilized oocyte to the surrogate gestational mother. This method, involving *in vitro* fertilization and embryo transfer, is the only way a woman without a uterus can have a child that is genetically her own.

A considerable number of women apply to become surrogate mothers, for motives of financial gain, altruism, pleasure in being pregnant, or atonement.[8] A child can now have 3 mothers—genetic, gestational, and rearing.

Surrogate pregnancy has stirred up an ethical debate. Apart from religious objections, there is concern about the physical and psychological consequences for the gestational mother, and there are endless opportunities for custody disputes and other legal complications. It has been found that the gestational mother does not bond strongly to the foetus, and most surrogate and commissioning mothers do not suffer from psychological problems.

Pseudocyesis

When a woman believes herself to be pregnant and develops symptoms and signs of pregnancy, this is called pseudocyesis. In a classic monograph, Bivin and Klinger[9] collected 444 cases from the literature. Many sufferers were parous, including women with as many as 10 children, and as many as six episodes of pseudocyesis.

The differential diagnosis includes delusions of pregnancy, in which there are no somatic changes. This is a common delusion and can also occur in men. There is also pregnancy simulated for social, mercenary, or legal purposes (e.g. to escape the death penalty).

The clinical features include:

- a firm belief in the pregnancy, usually lasting until the onset of a false labour at 9 months, after which the disorder usually resolves
- amenorrhoea
- morning sickness and/or pica
- enlargement of the breasts and nipples, and even a discharge of colostrum
- abdominal enlargement, caused by muscular contraction, tympanites, fat, or pathological lesions, but without effacement of the navel
- an illusion of foetal movements
- enlargement of the uterus to the size of a 6-week pregnancy.

Modern diagnostic tests have greatly reduced the frequency. The diagnosis should be made on ultrasound examination. Where radiology or ultrasound are unavailable, an examination under anaesthetic is recommended—in the presence of a family member to avoid accusations of abortion.

The psychological basis is usually an intense desire for children, especially in older childless women. In some cases, however, a guilty fear of pregnancy has been the background; this has occasionally led to dangerous attempts at abortion by non-pregnant women. Pseudocyesis is a demonstration of the influence of psyche over soma, mediated by hormonal secretion. It occurs in dogs, cattle, and rodents. Persistence of the corpus luteum would explain breast changes, moderate uterine enlargement, and secretory endometrium; but it is not the only basis: hormonal measurements have been made in at least 30 patients, some of whom had chronic anovulatory states, hyperprolactinaemia, or androgen excess.

These women require psychotherapy. Simply revealing the diagnosis is unsatisfactory because the patient may consult another doctor with the same symptoms, or develop a recurrence. The underlying conflicts must be explored, helping the patient to accept that she is not pregnant.

Sterilization

Women can be prevented from bearing children by various operations on the uterus and Fallopian tubes, indications for which are contraceptive, medical, eugenic, or psychiatric. Sterilization is the most effective and widespread contraceptive method. A large number of studies have looked at its effect on mental health, but many had methodological weaknesses. Ekblad, however, published two thorough studies in 1950s—a general study of 225 women, of whom 99 per cent were interviewed 5 to 6 years later, and a unique study of 60 sterilized women with no living children.[10]

There have been two modern prospective studies. Cooper and colleagues in Oxford[11] interviewed 201 women 4 weeks before non-puerperal tubal sterilization for contraceptive reasons; 190 were re-interviewed 6 months later, and 193 at 18 months after sterilization: the number with psychiatric illness fell from 21 before the operation to 9 at 6 months, and rose to 18 at 18 months. Not surprisingly, the presence of psychiatric disorder before the operation was a predictor of its continued presence; only two who were in good psychological health before the operation developed psychiatric illness 6 months later. A WHO collaborative study, involving five countries (India, Colombia, Nigeria, Philippines, and England), compared 926 sterilized women with 924 who used other methods of contraception: those who chose sterilization had more preoperative psychiatric disorder. The results from the Nottingham field centre[12] were published separately, and found that 9/138 sterilized women had psychiatric disorder before the operation; after surgery there were only three new cases at 6 weeks, and four more at 6 months, less than the control group.

A small minority of sterilized women are troubled by frigidity or severe regret. The most concrete evidence is a request for reversal, several studies of which have been published. Regret is more common in the following groups of women:

- Younger women or those with fewer children: Ekblad[10] found that none of his 60 childless women required hospitalization for depression, but 16 were seriously distressed and 29 expressed a longing for children of their own.
- Those in whom sterilization was the condition for a termination—a barbaric and punitive practice that used to be the rule in some countries.
- Those sterilized at a time of crisis—after parturition, or during a psychiatric illness—when it is difficult to make a balanced judgement.
- Those under external pressure.
- Those with learning difficulties: the issue of sterilization, which has from time to time been practised in various countries, is becoming more important. With a policy of community care and a greater tolerance of sexual activity, there is an increased risk of pregnancy in women with severe learning difficulties, with the spectre of inherited disorders and problems in mothering. Yet these women greatly desire children, and do not have the same resources to compensate for their lack.
- Those sterilized for medical reasons such as inherited disorders, for which medical advances have later provided alternative solutions (e.g. amniocentesis).
- Those who seek sterilization in a context of marital disharmony: after the marriage has failed, the wife may remarry and change her mind about further children.
- Those with religious scruples.

Hysterectomy

This is one of the commonest operations, and is performed in about 10 per cent of women. There have been claims that it leads to 'post-hysterectomy depression'. But this idea has been thoroughly and systematically refuted. Several prospective investigations have shown that mental health improves after hysterectomy. Three comparable Oxford studies, conducted between 1975 and 1990, have addressed this problem: all showed that psychiatric morbidity fell below its preoperative level, or remained low.[13] The ranks of women with 'post-hysterectomy' depression are swollen by those seeking a surgical remedy for psychosomatic complaints.

In younger women, infertility can be a source of discontent. It would not be surprising if the loss of the womb affected feminine identity and libido; but this is probably also a myth. Prospective studies from Oxford,[13] St Louis,[14] and Aberdeen showed an increase in the frequency of intercourse, and of enjoyment. Concomitant oöphorectomy does not adversely affect psychiatric well-being.[15]

The psychiatry of pregnancy

Pregnancy adjustment

The psychopathology of pregnancy needs to be understood in terms of the adjustment all women must make when they conceive. Pregnancy is not only a biological event, but also an adaptive process.[16] A pregnant woman must carry the baby safely through to delivery, and adjust to the sacrifices that motherhood demands. She must ensure the acceptance of the child by the family, develop an attachment to the baby within, and prepare for the birth. She must adjust to the alteration in her physical appearance, and develop a somewhat different relationship with the child's father.

Many pregnancies are unplanned and not initially welcomed. Many women react to conception with grief and anger. A random sample of English mothers showed that 44 per cent of pregnancies were unintentional, including 17 per cent that ended by legal abortion. In married women aged 25 to 29 years with one child, 80 to 84 per cent of pregnancies were planned, compared with 26 per cent in the unmarried.[17] The planning of pregnancy and its acceptance are two different things. The fact of planning does not guarantee acceptance; 6 to 12 per cent of those who plan their pregnancies subsequently regret them. Most unplanned pregnancies are immediately accepted; even if the initial response is negative, gradual acceptance usually follows. In a small proportion of cases, rejection continues to the end of the pregnancy.

Pregnancy has a profound effect on the relationship with the child's father. At every stage this relationship is of the highest importance. A pregnant woman needs increased attention and care and is sensitive to perceived rejection. Pregnancy alters other relationships as well—with the wider family and friends. Many women become closer to their families-of-origin and in-laws.

The change in appearance and shape is sometimes distressing. Some take pride and pleasure in these changes, enjoy the extra attention, and feel an enhanced sense of womanliness. Others are concerned about their loss of figure and facial bloom, weight gain, and stretch marks. Dysmorphophobia, with ideas of reference and social avoidance can ensue.

Pregnancy may be accompanied by medical disorders, and in all there is an interaction between physical and psychological factors. Pica is common, especially geophagia (eating earth or clay), which can lead to iron deficiency anaemia, bowel obstruction, and roundworm infection; other forms of pica can lead to lead poisoning or hypokalaemia. Rarely, hyperemesis can cause Wernicke's encephalopathy, and delirium can complicate chorea gravidarum.[18]

Denial of pregnancy

In women who do not realize they are pregnant, one must distinguish between three different phenomena: unnoticed pregnancy, deliberate concealment, and dissociative denial. A German survey of 29 000 births found 62 women who failed to recognize pregnancy until the 20th week (1/475 births); 12 were not diagnosed until they were in labour with a viable infant (1/2455).[19] A Welsh study obtained similar figures.[20]

The late discovery of an unwelcome pregnancy carries a small risk of suicide. The mother is also at risk of all those complications of delivery that, with modern antenatal care, have become rare. For the child there are increased hazards, including prematurity and neonaticide.

Prenatal attachment

The mother 'bonds' or 'affiliates' to the unborn child in a way analogous to the formation of the mother–infant relationship after birth. Prepartum bonding is catalysed by quickening and probably by ultrasound examination. The mother begins to have fantasies

about the baby and talks affectionately to it. She may engage the husband or partner and other children in 'playing' with the baby. At the same time she prepares for the birth and motherhood ('nesting behaviour').

There is a pathology of the affiliative stage. In some mothers there is minimal attachment even at term. The foetus is viewed as an intrusion, whose movements annoy the mother and disturb her sleep. A poor mother–foetus relationship is one of the predictors of impaired mother–infant bonding. When the mother's attitude to the pregnancy is obstinately rejecting, therapists can direct her attention to the relationship with the child within. Stroking the abdomen and identifying foetal body parts, or telling stories about the baby's future life, have been suggested.

Foetal abuse

When a mother deeply resents her pregnancy, she may try to harm the foetus. This occurs, with determined intent, in self-induced abortion. It may also occur as a manifestation of rage against the baby;[21] a pregnant woman may pound on her abdomen, even to the point of causing bruising.

It is not only the mother who may 'batter' the foetus. Domestic violence is common and may increase during pregnancy, when kicks and blows are directed at the abdomen, rather than the face. The main factors are sexual frustration, substance abuse, jealousy, the mother's irritability, and unreadiness for fatherhood.

The foetus is cushioned from external violence by the amniotic fluid, but can still be damaged by severe abdominal or pelvic injuries. Domestic violence can lead to miscarriage, foetal death, and premature birth. Infants can be damaged by penetrating wounds, and there are over 100 instances of gunshot wounds to the gravid uterus—the result of murderous assaults, attempts to induce a late abortion, or suicide attempts.

Mental illness during pregnancy

(a) Anxiety

For many mothers, pregnancy is a time of considerable anxiety. The first trimester may involve an anguished decision whether to continue or terminate the pregnancy. Those who have previously suffered from prolonged infertility, multiple miscarriages or foetal loss are especially prone to prepartum anxiety. In the third trimester anxiety is centred on three main themes: fears of parturition (tocophobia), of foetal abnormality, and of failure to cope with motherhood.

These anxieties will usually be managed by ventilation and support, but anxiolytic medication can be used cautiously. Of the anxiolytic agents, phenothiazines are relatively safe. Benzodiazepines are contraindicated in the last stages of pregnancy because of foetal intoxication ('the floppy infant syndrome'). Propranolol is best avoided, because of reports of intrauterine growth retardation, and neonatal cardiac and respiratory symptoms.

(b) Depression

Although prepartum depression has not aroused the same interest as postpartum depression, it is no less common. Depression is common in all women in the reproductive age group, and pregnancy is not protective. Depression can be recurrent, and there is an association with puerperal mania.

The frequency of suicide is a vexed question. There are problems about the accuracy of the data since not all suicides are reported to the coroner, not all have necropsies, and not all necropsies include an examination of the uterus. In addition, both suicide and pregnancy are often concealed. One must therefore treat with scepticism those enquiries which do not scrutinize the primary records. Nevertheless, there is evidence that the suicide rate has declined throughout this century; in the first quarter, about 13 per cent of women who committed suicide were pregnant—a rather high figure, suggesting that pregnancy was a risk factor at a time when illegitimate pregnancy was stigmatized. This was confirmed by the thorough mid-twentieth century study of Weir.[22] More recent studies show rates below those in the general population.

Severe prepartum depression is sometimes left untreated, because of fears about the effect of drugs on the foetus. These fears have been exaggerated. No antidepressive drug is known to have teratogenic effects. Most have no effect on the foetus, though fluoxetine may reduce uterine blood flow and paroxetine may cause neonatal pulmonary hypertension. There are reports of toxic effects or withdrawal symptoms in neonates, so that medication is more to be avoided during the last trimester. Electroconvulsive therapy is safe, provided that the mother is competently oxygenated during anaesthesia; pregnant women should be screened for rare syndromes of pseudocholinesterase deficiency before receiving this treatment.

(c) Alcoholism

Pregnancy has a beneficial effect on alcohol addiction, but, if heavy abuse continues, there are severe effects on the foetus. The main effect is retardation of intrauterine growth[23]; although ethanol shortens gestation, the low birth weight is not explained by prematurity, rather the infants are small for gestational age. The infant becomes addicted and may suffer neonatal withdrawal symptoms. Ethanol is also teratogenic, causing 'the foetal alcohol syndrome' (or 'spectrum disorder'), first described in France in 1968.[24] The features include facial dysmorphism due to maxillary hypoplasia, and brain damage, resulting in long-term cognitive impairment and behavioural disorders (see also Chapters 9.2.7 and 10.4). In the detection of these severe complications, systematic prenatal screening for alcohol abuse is useful.

(d) Other addictions

Cannabis is commonly abused by pregnant women; it affects foetal growth, and may lead to long-term neurobehavioural and cognitive deficits. **Lysergic acid diethylamide** may have teratogenic or mutagenic effects. **Phencyclidine** addiction leads to withdrawal symptoms.

Narcotic addicts, like alcoholics, have multiple emotional and social problems, and many do not seek antenatal care. The infants may be affected by maternal malnutrition and infections such as venereal disease, hepatitis, endocarditis, and AIDS. Narcotics are not teratogenic, but a high proportion of the infants are of low birth weight, partly explained by prematurity, and partly by intrauterine growth retardation. A withdrawal syndrome develops in most babies. The perinatal mortality rate and frequency of sudden infant death, are increased. There is an increased incidence of microcephaly, and there may be impaired mental development, although other factors in the maternal life style may account for this. Methadone maintenance reduces the effect on birth weight; but it may depress respiration in the newborn, and lead to a more severe and prolonged withdrawal syndrome, with a greater frequency of seizures. Buprenorphine may be a more suitable maintenance therapy, with milder withdrawal effects. If it is decided to withdraw heroin, this should be done in the second trimester,

replacing it by methadone. Naloxone, which can be given by implant, has been used, although there are concerns about foetal abstinence syndromes. After birth, the infants should be kept in hospital for at least 14 days. Respiratory depression can be treated by naloxone, and seizures and withdrawal symptoms by sedatives such as diazepam, or by tincture of opium.

Cocaine may be teratogenic, causing genitourinary and cardiac abnormalities, but the evidence is conflicting. Its main effects are cardiovascular: it causes uterine vasoconstriction, and this can lead to placental abruption. The infants may suffer cerebral infarction. There is intrauterine growth reduction and an increased incidence of microcephaly. Premature labour is common. There is a withdrawal syndrome, but this is less severe than with narcotics. There is some evidence of an increased risk of sudden infant death. Long-term effects on language development and behaviour are controversial, and may be due to confounding factors such as maternal depression, other drugs, and the environment.

All these mothers should receive close psychiatric supervision and social casework. Hair and meconium analysis improves the diagnosis of opiate and cocaine abuse in mothers who present unexpectedly in labour.

(e) Eating disorders

There are psychological and somatic reasons for an antagonism between pregnancy and anorexia nervosa; nonetheless, most anorexic women recover, and menstruate when their weight reaches about 80 per cent of the standard weight. Ovulation can be induced by clomiphene or menopausal gonadotrophin in those who fail to menstruate. There are numerous case reports and several long-term studies showing that many women with a history of anorexia nervosa give birth to children in the normal way. The overall effect on fertility has been quantified by a 12-year Danish study; the average number of children (0.6) was about one-third the usual figure.[25] The desire for children is shown by the frequency of infertility treatment, planned pregnancy, and breast feeding.

A minority become pregnant while in the throes of the disease. Anorexic amenorrhoea may delay the diagnosis. Pregnancy usually has a beneficial effect; but if the mother continues to restrict her diet, the foetus may suffer from malnutrition. Occasionally it has been necessary to rescue the infant by elective Caesarean section. There is a tendency to relapse in the puerperium. When mothers are actively anorexic, there is often conflict at mealtimes; occasionally children may become involved in their mother's asceticism, and suffer stunted growth.

Bulimia nervosa is often improved by pregnancy. The pressure of the enlarging uterus on the stomach makes bingeing more difficult. About half relapse after delivery.[26] Pregnancy is not much affected by bulimia, but low birth weight has been reported. Bulimic mothers sometimes show deviant mothering, ignoring or excluding their children while overeating or vomiting, or restricting food supplies.

(f) Obstetric factitious disorder

Self-induced illness behaviour can extend into the obstetric domain.[27] Women may induce bleeding to simulate threatened miscarriage, placenta praevia, or postpartum haemorrhage. They may stimulate rupture of the membranes to precipitate an early delivery. Others have been caught manipulating instruments, for example an external tachodynamometer. Two patients even attempted to simulate hydatidiform mole, by adding human chorionic gonadotrophin to blood samples.

(g) Psychosis

Numerous asylum surveys have testified to the lower frequency of psychosis during pregnancy than after delivery. This was confirmed by Kendell and colleagues, in their linkage of Edinburgh obstetric and psychiatric case registers[28]: in a study of 54 087 births, they found rates of 2.1 per month before conception and 2.0 per month during pregnancy, much lower than after childbirth (51 in the first month).

Pregnancy probably has no effect on chronic delusional states, but it does have a beneficial effect on menstrual, bipolar, and possibly cycloid (acute polymorphic) psychoses.[29] Nonetheless, acute manic and cycloid episodes occur during pregnancy, and some seem remarkably similar to puerperal psychosis. They would be regarded as sporadic or random, except that they have been observed in women with a history of puerperal psychosis (at least 13 in the literature).[30] There is an association with multiparity, with the postpartum episode occurring first.

Neuroleptic agents appear to be safe during pregnancy. Phenothiazines and butyrophenones are not teratogenic. The main (but infrequent) hazard is sedation and extrapyramidal symptoms in the newborn. Lithium is relatively dangerous; at least 12 cases of the rare Ebstein's anomaly have been reported. As delivery approaches, reduced renal clearance can result in toxicity with normal doses; eight cases of alarming blood levels (up to 5 mmol/l) have been reported, with coma and convulsions in the mother. Even at normal blood levels, babies exposed to lithium have suffered lethargy, hypotonicity, and other effects. Carbamazepine has been associated with rather high rates of congenital abnormality, and sodium valproate is particularly dangerous, with major abnormalities especially spina bifida, and a foetal valproate syndrome.

(h) Obstetric liaison services

In view of the complexity of the psychological response to pregnancy, and the frequency of anxiety, depression, and other psychiatric disorders, there should be good liaison between obstetric and psychiatric services. In addition to the need to diagnose and treat prepartum psychiatric disorders, the high level of supervision in the antenatal clinics offers an opportunity for preventive psychiatry, by screening for vulnerable women, including those with unwanted pregnancies, severe social problems, or a history of psychosis, addictions, or depression.

The psychopathology of parturition

Childbirth can be one of the severest of human ordeals, and in spite of its brevity, is a time of risk for psychopathology.[18] In advanced countries, all these complications are rare, but may still be common where obstetrics is primitive, or pregnancy denied. Acts of desperation, such as auto-Caesarean section or suicide, and rage attacks, endangering the foetus, are fully described in the older literature. Delirium is well documented; in most cases, it lasts a few hours, starting shortly before delivery and disappearing after the birth, with amnesia for the event; but it can continue into the puerperium, or start immediately after the birth. Engelhard[31] gave the best estimate of its frequency: in a 10-year survey, there were five cases of transitory confusional states in 19 910 births. The existence of this phenomenon aggravates the jurisprudential problem of neonaticide, because, in an unattended delivery, it is impossible

to know whether or not the mother was temporarily confused. Unexplained stupor or coma has also been described during and immediately after delivery.

Infant loss

The child may be lost for a variety of reasons:

- termination of pregnancy at the behest of the mother
- miscarriage, ectopic pregnancy, and late termination of a wanted child for medical reasons
- foetal death *in utero*, stillbirth, neonatal death, and sudden infant death ('cot death', SIDS)
- relinquishment to adoption.

(a) Termination of pregnancy

The indications for abortion include the following:

- medical—to preserve the health and life of the mother
- humanitarian—when pregnancy has resulted from rape or incest
- eugenic—where there is a risk of congenital abnormality
- psychiatric
- social—because pregnancy is untimely and disruptive
- on demand—in the belief that women should be free to decide when to have children.

There has been a debate on the validity of the psychiatric indications; this turns on the psychiatric consequences of a refusal to terminate. Suicide threats are common, but are rarely carried out; nevertheless there can be no doubt that unwanted pregnancy is a factor in completed suicide. A history of puerperal psychosis is not an indication, because it is equally likely to follow abortion; but there are other, arguably more serious, puerperal complications such as mother–infant relationship disorders, which are more common and severe after unwanted pregnancy. These can be avoided by adoption, but the psychological effects of relinquishment are not negligible.

The psychological effects of termination have been thoroughly explored. Most who voluntarily abort suffer no adverse effects, either in the short or long-term. There is often relief, even euphoria, and a reduction in anxiety, depression, anger, guilt, and shame. A minority experience regret and self-reproach over the 'murder' of the baby. Some feel like criminals and worry about punishment, a nemesis of sterility or future congenital malformations. A few develop clinical depression. A Finnish study showed that the suicide rate was increased from $11/10^5$ to $35/10^5$.[32]

There is a literature on 'postabortion psychosis', but both parts of the term have multiple meanings; 'abortion' refers to miscarriage, termination, criminal abortion, and even stillbirth after short gestation, and 'psychosis' includes delirium, Wernicke–Korsakow syndrome, melancholia, and psychogenic paranoid disorders. Manic or cycloid episodes, similar to puerperal psychosis, occur after abortion: apart from epidemiological evidence[33] and individual cases, the association of postabortion and postpartum psychosis in the same woman has been reported on at least 14 occasions.[30] Some episodes occurred after miscarriage, but several followed termination, in most cases performed to prevent a recurrence of puerperal psychosis.

To minimize the psychological risk, prudent decision taking is of the essence, and counselling has a valuable role. The most difficult part of the experience is the loneliness and isolation. Many do not inform their parents, and, when they do, face censure and unwelcome pressure. The attitude of the child's father is crucial. Attempts should be made to involve him in all aspects of the experience; unfortunately his reaction is often unhelpful. It is axiomatic that a woman should make her own decision—one of the most difficult she will ever take. It often has to be taken hastily, in an atmosphere of conflict and turmoil. The best outcomes are found when a woman makes her decision in a context of respect and support from partner, parents, friends, or counsellor.

(b) Miscarriage

This is a common event, perhaps 40 per cent of all conceptions, but only 10 per cent occur after pregnancy is recognized by amenorrhoea or other signs. An ectopic pregnancy is gynaecologically more serious, but has the same psychological effects. The emotional consequences of miscarriage are not trivial, and can be compared to perinatal death—less severe, because there has been little time for attachment to the newly conceived, but still the loss of a greatly desired child. The event itself, with foetal tissue passed suddenly and painfully, may be disturbing. Some of the psychological symptoms may resemble post-traumatic stress disorder, with intrusive re-experiencing ('flashbacks') and nightmares. There is a sense of failure, guilt, and anger. The incidence of depression is four times the rate found in the general population. There may be depressive episodes at the time of the expected delivery, anniversary reactions, and an increased risk of postpartum emotional disorder after a later normal delivery.

Helping a mother who has suffered a miscarriage is a variant of grief therapy, in which her intense distress is shared, and sadness, guilt, and anger ventilated.

Late termination for medical reasons, although a deliberate intervention, is psychologically similar to miscarriage and to foetal death *in utero*. Some wish to continue the pregnancy in the full knowledge that the baby will be abnormal. Depression is common, and grief long-lasting. All these women require counselling, before and after the termination.

(c) Foetal death *in utero*, stillbirth, neonatal death, and sudden infant death

Reactions to these events are generally more severe than to miscarriage, and each has its special characteristics. When the baby dies in late pregnancy, the mother carries a corpse within her, and must undergo a futile labour. If it dies during labour, the loss is sudden and shock pronounced, with a strong sense of unreality. When the child dies in the first week, the parents have to endure great anxiety, with dwindling hope; they may be involved in the decision to switch off the respirator, and witness the child dying. The later death of an infant, when the maternal emotional response is fully developed, especially sudden infant death, is at the very top of the catalogue of calamities; there is no warning or preparation, and the death is followed by a forensic investigation.

(d) Grief after infant loss

This is similar to other grieving, but has its own special character. There is shock, followed by emotional numbness and emptiness, then long-lasting and agonizing sadness. Grief hallucinations (of foetal movements, the baby's face, the infant crying or playing in the cot) may be experienced. There is guilt, anger, and recrimination. There are various crises, especially the disposal of toys, baby

clothes, and nursery furniture, as well as meeting friends and relatives; some, floundering in their embarrassment, are evasive and unable to comfort or sympathize ('wall of silence'). Especially after SIDS, there may be shame, stigma, and even ostracism or malicious speculation. Envy of successful mothers is a problem; there may even be a temptation to steal babies. Surviving children may be confused by their parents' grief, upset by family turmoil, and deprived of attention and care; they are also grieving and preoccupied with their own search for the meaning of death.[34]

When helping the parents,[35] the principles are as follows:

- **Honesty and openness in communication**. The admission of errors is delicate, but the parents' guilt should not be reinforced by the obstetric team's refusal to accept responsibility. Recrimination, litigation, or querulant reactions are common. Staff should accept this as normal, and try not to be defensive. After a stillbirth, most mothers prefer to be segregated, and discharged early. One or more interviews with the consultant obstetrician are indicated. It is essential that the mother is visited by a member of the primary care team. A lactating mother may need bromocriptine, or to donate milk to a milk bank. Hypnotics may help mothers troubled by insomnia. The doctor should be alert for secondary depression.
- **All parents want to know why the baby died**. The necropsy can help, but parents should be warned that often no explanation is found. Necropsies in SIDS are specialized; the pathologist can play a vital psychological role, and should be available for discussion.
- **Mementoes** should be kept, including a photograph. The dignity of naming and a burial ceremony is helpful. The value of seeing and holding the dead baby has been challenged.[36]
- **The bereaved mother needs to share her distress**. A sensitive and sympathetic person, with the time and interest to listen, can help her grieve and accept her loss. This support will often come from the husband or partner, family, or friends. If not, professionals, especially chaplains or nurses, should step in. Self-help groups and voluntary agencies are invaluable for some mothers.
- **The next pregnancy**. No doctrinaire advice can be given about the timing of the next pregnancy. Increased anxiety during pregnancy and the puerperium can be expected.
- **The grieving sibling**. The routine and rhythm of family life should be disturbed as little as possible. The parents should not be afraid to show their emotions—it is best to acknowledge their sadness, and how much they will miss the baby. They should try to give a factual account of what happened, avoiding euphemisms. It is important to reassure the children: they are not responsible and will not lose the love of the parents; neither they nor their parents are in imminent danger of death. The child can be helped to grieve by looking at pictures of the dead sibling, attending the funeral, and visiting the graveyard. (see Chapter 9.3.7 for further information about bereavement in childhood.)

(e) Relinquishment

Adoption used to be the main way to satisfy the longing to rear children and to handle accidental pregnancy, but there has been a great social change in Europe and North America. Since 1950, the number of children born to single mothers has climbed steeply, and is still climbing, but, despite this increase, the number of adoptions is falling steadily. This is not due to spectacular improvements in the infertility treatment, reducing the demand, nor to the relaxation in the abortion laws, reducing the supply, but a new tolerance of single motherhood. In partial compensation for the scarcity of relinquished babies, the practice of adopting foreign-born children has arisen.

Although adoption is on the wane, attention has been focused on the psychological effects of relinquishment.[37] For some relinquishing mothers, giving up the child is a painful, loving act of selfless courage. In others it is the enforced loss of a living child, with a charade of informed consent. Relinquishment is among the most stressful of events. Instead of understanding and support, there is often loneliness and ostracism. Time is no healer; the child continues to exist and can be seen again, and there is often a fantasy of reunion or restitution. As time goes on, there is a new component; the adult child may seek its biological mother, and there is the hope that this event, which she cannot influence, may happen. There has been a growth in the number of organizations to help relinquishing parents find their offspring. Many countries are grappling with the problems of legislating for reunions.

To avoid these severe and prolonged psychological effects, a relinquishing mother needs counselling during the pregnancy. The aim is to emerge from the experience with self-respect and dignity. After delivery the mother should be encouraged to see the infant and photographs should be filed. Follow-up counselling should be continued for at least 6 months. The mother may wish to join a society for relinquishing parents. There is also the relationship with the adopting family to consider. Adoptive parents should accept any gift or token of the natural mother's love. Information on the outcome of the child should be available. Some birth-mothers wish to provide up-to-date information, so that the child knows they are now respected citizens. A recent innovation is the practice of 'open adoption', in which both sets of parents meet. There is even 'continuing open adoption', which means that they remain in contact over the course of the child's development.

The psychiatry of the postpartum period

The normal puerperium

For many or most mothers, giving birth is a supreme moment, and euphoria or elation is common. Some may be too excited to sleep. These feelings of peace, fulfilment, and accomplishment help to sustain mothers during the weeks of strain that follow. Prolonged euphoric reactions, lasting a week or more, are probably mild puerperal mania, and are often followed by depression.

Newly delivered mothers have to face a number of challenges, including the following.

- **Physical exhaustion**. This can be coupled with the painful sequelae of pelvic trauma.
- **Breast feeding**. Although this has many advantages, it is often difficult to establish.
- **Insomnia**. Sleep deprivation, especially during the first month, is a cause of irritability, and should be borne in mind when mothers present 'at the end of their tether'.
- **Recovery of normal figure and attractiveness**. This may be threatened by weight gain and stretch marks. Mothers may occasionally develop a state similar to dysmorphophobia.[38]

- **Loss of libido**. Episiotomy and vaginal trauma often cause dyspareunia; fatigue may depress sexual activity. Nevertheless sexual relations are usually resumed within 1 to 3 months, though reduced in frequency, and with a delayed return of orgasm. For this and other reasons (e.g. jealousy) the marriage may come under strain.
- **Social privation**. The loss of employment, income, and leisure, as well as confinement to the house, are all contributory factors.

With this background of rapid biological, social, and emotional transition, it is not surprising that a wide variety of psychiatric disorders occur; indeed the psychiatric complications of childbirth are more numerous and complex than in any other human situation.

The **maternity 'blues'** is so common as to be almost normal. Usually between the third and fifth days, many mothers experience a sudden, fleeting, and unexpected period of sensitivity and uncharacteristic weeping. In the great majority this passes off within a few hours, or a day or two. There is some evidence for an association between this brief dysphoric reaction and postpartum depression.

Reactions to severe labours

(a) Post-traumatic stress disorder

After excessively painful labours, some women suffer nightmares, and repetitive daytime intrusion of images and memories, similar to those that occur after the harrowing experiences of war and natural disaster. Since the original description,[39] over 40 papers have been published on this subject, including 10 quantitative studies showing rates of up to 5.9 per cent of deliveries. Many of these women avoid further pregnancy (secondary tocophobia), and those who become pregnant again may experience a return of symptoms, especially in the last trimester.

This disorder can be treated by counselling and by specific psychological therapies. Tocophobia is an indication for elective Caesarean section.

(b) Querulant reactions

Another reaction to a severe labour experience is pathological complaining (*Querulantenwahn*). These women complain bitterly about perceived mismanagement, and their angry rumination may continue for weeks or months, interfering with infant care. Some confine themselves to vengeful fantasies and verbal or written criticism, but others proceed to litigation. Careful assessment is needed to distinguish these reactions from reasonable complaining.

This disorder can be treated by a psychotherapeutic approach, which distracts the mother from her grievances and reinforces productive child-centred activity.

Postpartum anxiety disorders

Recent research has shown that postpartum anxiety disorders are just as common as postpartum depression.[38,40] A review of eight studies of 'panic disorder' showed that 44 per cent of anxious women had an exacerbation, and 10 per cent a new onset, in the puerperium.[41] It is important to identify the focus, as well as the form, of anxiety, because there are several themes that indicate specific psychological therapies. Benzodiazepines should be used with caution in lactating mothers. They are well absorbed from the gut, and more slowly metabolized in the neonatal liver, and occasionally cause lethargy and weight loss in breast-fed infants.

(a) Puerperal panic, and phobic avoidance of the infant

Some mothers, especially *primiparae* in isolated 'nuclear' families, are overwhelmed by the responsibility of caring for the newborn.[42] The panic and agitation seen in extreme examples is an exaggeration of the anxiety that many women experience when they first confront this awesome task. If no help is available, a mother can develop a phobic avoidance of the infant,[43] and risks losing her mothering role.

These disorders can often be handled by the wider family, without invoking professional help. The mother needs sedation, especially at night. During waking hours, she should remain with the baby, but must be supported at all times. Treatment is by desensitization. Gradually the mother takes over, at her own pace, undertaking the easiest tasks first, and involved in all decisions. In severe cases, conjoint admission may be the only way to rescue the situation. With correct diagnosis and management the prognosis is excellent.

(b) Anxieties about infant health and survival

The care of an infant involves ceaseless vigilance. In women prone to anxiety and excessive worrying, or in those who have suffered years of infertility or recurrent miscarriage, motherhood can lead to excessive solicitude about banal tasks that put the baby at risk (e.g. bathing) and sensitivity to the slightest indication of illness. In some, the anxiety is focused on the possibility of sudden infant death.[44] These mothers lie awake listening to the baby's breathing; sleep with their hand on the infant's chest, check the infant many times each night, or even wake the baby to ensure that he or she is still alive. This results in excruciating tension, insomnia, and exhaustion.

These mothers require anxiety management. Day-hospital attendance, with relaxation therapy and group support is ideal. A mother with 'fear of cot death syndrome' may be helped by explanations about the rarity of SIDS, and the infant's resistance to asphyxia, as well as devices to monitor the infant's breathing. The vicious cycle of insomnia and hypervigilance can be interrupted periodically by involving relatives or friends, so that she can sleep under sedation. Ventilation, and the support of mothers who have recovered from similar problems, is helpful. However, these are only palliatives, because the underlying cause is an event, which, albeit uncommon, remains possible during a period of several months.

Puerperal obsessional disorders

There is evidence that the puerperium is one of the main precipitants of obsessive–compulsive disorders.[45,46] In addition to obsessional rituals, the disorder may present with thoughts, images, or impulses of child harm. These impulses to attack the child must be distinguished from the pathological anger that precedes child abuse. The mother is gentle and devoted. She experiences extravagant infanticidal images, such as stabbing, decapitation, or strangulation. She fears being left alone with her infant, and may take extraordinary precautions.[47] The obsessional content may be of child sexual abuse, for instance masturbating or castrating their sons.

Ventilation, explanation, and psychotropic medication are part of the treatment, but are rarely sufficient. It is important to discourage avoidance of the child, and encourage cuddling and play, thus strengthening positive maternal feelings. Cognitive–behavioural treatment can help her to achieve mastery over

irrational impulses. (For the treatment of obsessive–compulsive disorder, see Chapter 6.3.2.1.)

Depression

Puerperal melancholia was one of the first postpartum psychiatric disorders to be identified. During the asylum era, only the most severe cases were admitted, and the occurrence was underestimated when compared with 'puerperal mania'. When, in the 1950s, attention turned to milder disorders, postpartum depression was found to be common in the general population. The pioneering work of the Gordons in New Jersey[48] was soon widely confirmed. In the last 10 years there has been a flood of papers from all over the world. Surveys have shown rates of at least 10–20 per cent, or even higher in the 'Third-World'.[49] 'Postnatal depression' has become a household word. It is an important lay concept, which has legitimized maternal depression in the minds of the public, providing a valid explanation for role failure, diminishing stigma, enabling mothers to accept that they are ill, and to come forward for treatment. It is a slogan that can be wielded in the political struggle to obtain better services for mothers of young families. There is a need for such concepts, which have social influence.

However, one must examine the scientific value of this concept with scepticism. Depression after childbirth is clinically similar to any other depression, and the association of depression with the puerperium is not striking. Whatever the prevalence in surveys, only about 5 per cent consult their general practitioners. The epidemiological evidence is weak. The suicide rate in the first postpartum year is below the female rate.[32] Depression is common in women during the reproductive years, whether they are infertile, pregnant, puerperal, menopausal, or involved in child rearing. The term 'postnatal depression' has the danger of introducing into the minds of the unwary the mirage of a homogeneous disorder with a single cause. Rather, it is a rubric for a heterogeneous group of disorders. Many mothers with anxiety, obsessional, or post-traumatic disorders, or with a disturbed infant relationship, are depressed, but the setting, causes, and treatment are different. Not surprisingly, research into its causes has found that they are the same as those that cause depression at all ages—heredity, a history of previous or prepartum depression, 'neuroticism', adverse events or social conditions, difficult relationships, and social isolation. It has been suggested that the burden of child rearing, rather than child bearing, is a factor, but this has been challenged by a Swedish twin study[50] and a Norwegian suicide study[51] showing that parous women have a lower risk of depression than nulliparous women. In mothers with recurrent puerperal depression one would expect to find specific factors. Adjustment to motherhood has received much less emphasis than it deserves; unwanted pregnancy has been found to be a predictor of antepartum and postpartum depression.

Whatever its frequency, the effects of depression on family life, and the emotional climate in which children are reared, is of great concern. A growing child needs emotional support, attention, approbation, and stimulation. The mother is the child's primary environment, and her mood dominates his or her world. Even very young infants are disturbed by deviant social behaviour in the mother. Although deficits are not universal,[52] her depression can lead to inattention through anergia or brooding, reduced quantity, quality, and variety of interaction, and loss of the reinforcement of the mother's gaiety and tenderness. Her anger may be misdirected at the children. Frequent irritability, impatience, and criticism induce social withdrawal, anxiety, and reciprocal anger. There may be educational deficits. These effects depend on the degree and duration of maternal depression, and the extent to which it involves interactions with the child. (See Chapter 9.3.6 for further information on the effects of maternal depression on child development.) In extreme cases, maternal depression can lead to the tragedy of combined suicide and filicide.

Treatment begins with effective diagnosis. Many more mothers are depressed than ever make their way to the surgery. The reasons for the failure to seek help are not fully understood: some recover early, some do not realize they are ill, and some are ashamed of confessing their symptoms, suffering in silence because of ignorance, stigma, and fears of losing their baby. Screening procedures help the primary care services to identify cases; an example is the Edinburgh Postnatal Depression Scale,[53] which has high sensitivity and specificity. Patients identified by screening, or self-referral, require a full psychiatric examination, in order to identify vulnerability factors and the specific components of postpartum disorders. This initial interview is best held at home, because clinic attendance is an obstacle for mothers fettered with the care of young children, and because domiciliary assessment has a quality that cannot be achieved in the office. The interview should explore the symptoms and course of the illness, study its context in the mother's life history, personality, and circumstances, review the events of this pregnancy, explore the mother's relationships with her spouse, baby, other children, and family of origin, and establish the available supports.

Treatment is focused on depression and any underlying vulnerability. It will always involve psychotherapy, if only in the form of a single interview; it will usually include medication or other specific treatments; a few require electroconvulsive therapy. Working with the baby's father, potentially the main supporter, is important; fathers can come under strain, either because their wife's intimacy with the baby disturbs conjugal dynamics, or because her depression has a domino effect on him. Home visits by community nurses are an ideal method of delivering continuing care and psychotherapy. An extensive literature has accumulated demonstrating the efficacy of psychological treatments by double-blind randomized controlled trials. As for drug treatment, there is no evidence that any drug is superior to others. There are at least 50 reviews of drug treatment in lactating mothers. The suckling infant has little body fat, less plasma protein-binding, an immature liver and kidney and an undeveloped blood-brain barrier. But the risks are minor. Only a minute dose is delivered to the infant: Epperson and colleagues demonstrated that serotonin re-uptake blocking agents do not affect serotonin levels in breast fed infants.[54] Occasionally, babies have been over sedated. It is not recommended that antidepressant agents be withheld, or that breast-feeding be stopped; but it is wise to use these drugs cautiously, and it may be helpful to take the drug after breast feeding (see also Chapter 6.2.3).

Prevention is important in mothers with a history of severe or prolonged postpartum depression. They often present during the next pregnancy, requesting advice or prophylaxis. If they are already symptomatic, or have obvious risk factors such as marital friction or social isolation, they need support from community psychiatric nurses, voluntary agencies, or other groups. If they are well, it is only necessary to establish contact, so that a recurrence is diagnosed and treated promptly. Prophylactic antidepressant medication can be considered.

Mother–infant relationship disorders

Just as the emerging relationship with the foetus is important during pregnancy, so also the growth of the mother–infant relationship is the key psychological process in the puerperium. 'Bonding' is a popular lay term; some professionals prefer 'attachment', but one must not confuse this with infant–mother attachment. The mother–infant relationship consists essentially of ideas and emotions aroused by the infant, which find their expression in affectionate and protective behaviour. Its immense power is revealed in self-sacrifice, and the pains of separation. Its inner presence is betrayed by external signs—touching and fondling, kissing, cuddling and comforting, prolonged gazing and smiling, baby talk and cooing, recognizing signals, tolerating demands, and resisting separation; but it is hard to select a single activity that lies at the core. Particular behaviours wax and wane, but the relationship endures, even when the child is absent, even when it is gone for ever. This emotional response enables the mother to maintain the never ending vigilance, and endure the exhausting toil of the nurture of the newborn.

There is no 'critical period' in the development of the maternal response. Close proximity from the start ('rooming-in') gives confidence in mothering skills, and breast-feeding may help. The infant plays an important part. At an early stage, it can discriminate speech, and reacts preferentially to the human face and voice. Eye-to-eye contact mediates the interaction, and gazing becomes an absorbing activity on both sides. The baby's smile is another catalyst. Videotape studies have shown the infant contributing to a dialogue with its caregiver. Sometimes the maternal response is immediate, primed by affiliation to the foetus, but sometimes there is a worrying delay. For the first 3 to 4 weeks many mothers feel bruised, tired, and insecure, and their babies seem strange and distanced. As the baby begins to respond socially, a normal relationship develops rapidly.

The term 'mother–infant relationship (or bonding) disorders' covers a spectrum of clinical states, which has two main dimensions:

- An absent or negative emotional response. In severe cases, the mother regrets the pregnancy and expresses dislike or hatred of her baby. She may try to persuade her own mother, or another relative to take over, and may demand that the infant be fostered or adopted. The most poignant manifestation is a secret wish that the baby 'disappear'—be stolen, or die.
- Pathological anger. The infant's demands anger the mother and provoke aggressive impulses, which may lead to shouting, cursing, screaming, or assaults.

There is at present little data on the frequency of these disorders in the general community. At the level of 'threatened rejection'—where the mother has an aversion to her child and seeks temporary escape from child care (the threshold for active intervention)— the frequency of these disorders in the general community is probably about 1 per cent. It is much higher in mothers who seek help for 'postnatal depression'—about 10 per cent at the level of established rejection, and another 15 per cent for threatened rejection.[55]

These disorders are usually accompanied by depression, but there are many reasons to reject the euphemism 'postpartum depression with impaired mother–infant interaction'.

- A disturbed relationship is different from a mood disorder.
- When depression is associated with phobias, obsessions, or deviant behaviour, these co-morbid phenomena are still considered worthy of study and treatment in their own right.
- Impaired interaction, although it can be recorded and measured, is not the essence of the phenomenon, but merely its behavioural manifestation; it has other causes, especially infant-centred anxiety.
- Aversion to the infant is not confined to depressed mothers;[56,57]
- When they coexist, their severity and course often differ.
- Only a minority of depressed mothers have this problem. It is important to select them for treatment and not to stigmatize the others.
- The risks—including child abuse and neglect—are higher. It is these disorders, rather than uncomplicated depression, that have serious and long-term effects on the child's development.[58] These mothers are a high-risk group that can contribute to the vital task of preventing child abuse and neglect.
- The treatment is different and specific (see below). Assessment and treatment of this relationship forms an important part of the work of mother–infant mental health teams.
- Management must be aware of the need for training and service provision.
- The aetiology is different, with more emphasis on unwanted pregnancy and challenging infant behaviour.

The diagnosis can be facilitated by self-rating questionnaires, but the main clinical resource is an interview probing the mother's emotional response and behaviour. In severe cases and in research, the 'gold standard' is direct observation, preferably over a substantial length of time.

The treatment proceeds in stages:

- Where there is a delay in the maternal emotional response, explanation and reassurance are usually sufficient.
- When hostility, rejection, and anger are prominent, the primary decision is whether to attempt treatment or not. The mother must be given freedom of choice; it is dangerous for her to feel trapped in unwelcome motherhood. At the same time, the father has his rights. The option of relinquishing the infant must be openly acknowledged, and fully discussed with both parents.
- If it is decided to embark on treatment (as in most cases), depression should be treated with psychotherapy, drugs, or (occasionally) electroconvulsive therapy.
- The specific element of therapy is working on the dyadic relationship. This relationship, like others, grows through shared pleasure. The baby alone has the power to awaken its mother's feelings, so the aim is to create circumstances in which mother and child can enjoy each other. It is a mistake to separate the mother and baby completely, which merely compounds the problem by adding an element of avoidance. If there is any hint of abuse or aggressive impulses, the mother must never be left alone with her infant. She must be relieved of irksome burdens of infant care. When mother and baby are calm, she is encouraged and helped to interact with him—to cuddle, talk, play, and bring out his smile and laughter. Participant play therapy and baby massage may assist.

Treatment can take place in various settings. Home treatment can be successful, provided there is enough support to relieve the mother of night care and stressful duties: the maternal grandmother,

an understanding husband or partner or a family group can sometimes achieve this. Day-hospital treatment provides individual support and group discussion, as well as specific therapies. In the most severe and refractory cases, the proper setting is an inpatient mother-and-baby unit, where an experienced team of psychiatric and nursery nurses, available 24 h a day and 7 days a week, can provide full support. Even in the most severe cases, one can feel optimistic about a successful outcome. (For further information about child abuse see Chapter 9.3.3.)

Postpartum psychoses

These fall into two main groups—organic psychoses and bipolar disorders; a third group—reactive or psychogenic psychosis—is most convincingly seen in adoptive mothers and fathers. There are many causes of delirium after childbirth.[18] Organic psychoses are hardly ever seen in Europe or North America, but may still be important in Africa,[59] India, South-East Asia, and Latin America, where the majority of children are born. Historically the most important causes have been infection and eclampsia psychosis, but cerebral venous thrombosis is common in India.[60]

The form of psychosis still seen in Europe and North America was described by Osiander,[61] and illuminated by the case studies of Esquirol.[62] The long-standing controversy about its nosology has been resolved in favour of a relationship with the bipolar group; but there is also a connection with acute polymorphic (cycloid) psychosis, which may also belong to the bipolar rubric. These are biological brain disorders, with high heritability and an inborn tendency to develop episodes throughout life. The problem of causation can be broken down into three subsidiary questions: the nature of the diathesis, the determinants of clinical polarity (mania, depression, or cycloid), and the trigger that provokes the episode. The first two questions belong to the wider study of bipolar disorder. The third is specific to puerperal psychosis. The clinical facts suggest not one, but several triggers related to the female reproductive process—abortion, pregnancy itself (especially the last trimester), the early puerperium (especially the first 10 days), postpartum menstruation, menstruation in general (see menstrual psychosis above), and weaning. These triggers can be added to the list of other biological events that trigger bipolar episodes, including surgery, adrenocortical steroid treatment, and seasonal climatic changes. Instances can be given of the combination of all these triggers in the life history of individual women, and there may be a shared pathway in these diverse precipitants. The incidence is somewhat less than 1 in 1000 pregnancies.[28,63] The Edinburgh study showed no link with twin pregnancies, breast feeding, single parenthood, or stillbirth. The miscarriage rate has been low in two studies. This psychosis has high heritability—not just for bipolar disorder, but also the puerperal trigger.[64]

Postpartum bipolar psychoses are acute, rapidly reaching a climax of severity. The onset is usually between 2 and 14 days after delivery. Mania is severe, often with 'schizoaffective' symptoms or extreme excitement. Almost every psychotic symptom may be seen—the whole gamut of delusions, verbal hallucinations, disorders of the will and self, and catatonic features. There is often an apparent confusion or perplexity. Since the advent of electroconvulsive therapy and neuroleptic medication, the duration has fallen to a few weeks. A minority of patients show a tendency to relapse in rhythm with menstruation. Puerperal recurrences occur after 20 to 25 per cent of subsequent pregnancies. Non-puerperal recurrences are also common.

There are no specific treatments. The first resource is sedation by neuroleptic agents, but these should be used with caution because of the risk of severe extrapyramidal side effects, which include neuroleptic malignant syndrome. It is usual to stop breast-feeding, although this may not be necessary, because the infant receives only a minute dose of the neuroleptic and adverse effects have not been noted; Clozapine may, however, accumulate in breast milk. Lithium has been used increasingly since the link with manic-depressive psychosis was recognized. There is also evidence for its prophylactic value in women at high-risk. However, it may have adverse effects on breast-fed infants. Electroconvulsive therapy is highly effective in all varieties of puerperal psychosis, including puerperal mania. The location of treatment is an important issue. Since hospitalization can be disruptive to the family, this disorder should for preference be treated at home, where the patient can maintain her role as wife, homemaker, and mother, and her relationship with the newborn; but its severity and the lack of community resources make this a counsel of perfection. If hospital admission is necessary, there are great advantages in conjoint mother and baby admission.

Services for mothers with mental illness

An outline of the services required for this area of psychiatry is slowly emerging. Its aims are prevention in those who are vulnerable, early and accurate diagnosis, and rapid effective intervention, with minimal disruption of family life. These aims require the following:

- **A multi-disciplinary specialist team** This team, a key resource whatever the cultural background, should consist of psychiatrists, nursing staff of various kinds, psychologists and social workers; it can serve a population of several million inhabitants, handle severe and intractable illness, train all staff, develop services, and conduct research.
- **A community service** Domiciliary assessment and home treatment are appropriate for mothers.
- **Day care** A day hospital can provide a full range of interventions, including groups, play therapy, motherhood classes, anxiety management, and occupational therapy, with minimal family disruption. The presence of mothers with similar disorders is an additional support. The children are cared for in a crèche.
- **Inpatient facilities** Conjoint admission of mother and infant is superior to the admission of the mother alone.[65] Wards dedicated to conjoint admission also have advantages over admissions to general psychiatric wards, although they are more expensive.
- **An obstetric liaison service** Apart from treating prepartum mental illness, this provides an opportunity for preventive psychiatry, by detecting vulnerability during pregnancy.
- **Links with other agencies providing services for mothers** The social services have a key role. Their family centres fulfil a similar function to mother-and-baby day hospitals. They can relieve the burden on the mother, and safeguard the child, by providing emergency foster care. Other agencies include the National Society for the Prevention of Cruelty to Children (in the United Kingdom), midwifery services, primary care teams, and child psychiatry services.

- **A network of voluntary organizations** These are independent organizations, but can have close cordial ties with the professional service. There is no one better suited to support a depressed mother than another mother who has suffered a similar problem and is now well—she knows the stratagems or words of comfort that were helpful, and is living proof of the hope of recovery. For each disorder, a panel of recovered mothers is an important resource.
- **Medico-legal expertise** Expert advice is often required in cases of child abuse or infanticide, and where a mother with mental illness is seeking custody of, or access to, her children.

The psychiatry of parental children-killing

The term 'infanticide' covers the killing of infants and children by their mother or father in a wide variety of circumstances, broadly divided into 'neonaticide' (killing the newborn) and 'filicide' (the later murder of a child).

Neonaticide

Killing neonates (especially female infants) has been customary in certain societies as an official policy or 'grass-roots' custom for controlling population growth. This is completely different from criminal neonaticide, in which a mother, who has concealed her pregnancy and given birth in secret, kills the infant immediately after parturition. This was a major public health problem in Europe during the nineteenth century. Its frequency has dwindled as a result of contraception, a relaxation of the abortion laws, and changed attitudes to single motherhood; but it still occurs.

The mental state of mothers who kill the newborn can be deduced from the methods used. Suffocation is by far the most common,[66] and this testifies to the mother's panic, faced by a crying baby. In a minority, brutal head injuries, stabbing, or decapitation testify to rage and hatred.

In Europe, starting in Russia in 1647, the public has gradually taken a humane view of this felony. By 1881, all European states, with the exception of England and Wales, made a distinction between infanticide and other forms of murder, and assigned a more lenient penalty. England and Wales at last came into line with the Infanticide Act 1922. In some American states no distinction is made between this and other forms of murder.

There has been much debate whether the defence of insanity can be invoked. Most of these babies die when the mother is in the grip of an emotional crisis—seized by fear or fury. This is not generally acceptable in law as evidence of insanity, which is defined as a defect of reason. However, impairment of consciousness undoubtedly occurs during labour (see above); it is rare in hospital practice, but may be more common in clandestine deliveries, and is hard to exclude. If the defence is burdened with the proof of insanity, there can be no valid evidence in unwitnessed deliveries; but there is the possibility of a miscarriage of justice—that a mother, who killed her baby when her consciousness was clouded, is wrongly condemned.

Filicide

The majority of murdered children are killed by their parents, and the majority of female murderers kill their own child. A survey in Queensland gave an estimate of the frequency: of 49 infanticides between 1969 and 1978, there were 11 neonaticides and 38 filicides; this is about 3 in 100 000 per year of children under 5 years of age.[67] In Sweden, between 1971 and 1980, there were 79 cases involving 96 children—an annual rate of 2 in 100 000 children under the age of 5 years.[68] There are a variety of causes.

- **Depression** This is the most common cause. Studies of convicted mothers underestimate the frequency of depressive filicide, because many complete suicide. Melancholic filicide is committed in the belief that the child's best interests are being served (delusional mercy-killing). Mothers surviving depressive filicide usually make no attempt to conceal the crime; they confess and seek punishment. Mothers may kill more than one child, but family murder seems more common in men.
- **Child abuse** This is the other relatively common cause. Death results from ill-tempered assaults or overzealous punishment, without homicidal intent. Fathers are often involved.
- **Psychosis** In non-affective psychosis, filicide may occur if delusions involve the child, or as a result of command hallucinations.
- **Trance states** A few filicides have occurred during epileptic automatism or somnambulism.

Not all parental child killing occurs as a complication of mental illness. Unwanted infants are occasionally murdered in cold blood. Euthanasia of an incurably ill and suffering child can also occur.

Further information

The only modern texts that cover most of this subject are *Motherhood and Mental Health* (reference 30), and *Psychological Aspects of Women's Health Care* (2002), editors D. E. Stewart & N. L. Stotland, Washington, American Psychiatric Press.

References

1. Hitzig, J.E. (1827). Mord in einem durch Eintreten des Monatsflusses herbeigeführten unfreien Zustande. *Zeitschrift für Criminal-rechts-pflege*, **6**, 237–331.
2. Brière de Boismont, A. (1851). Recherches bibliographiques et cliniques sur la folie puerpérale, précédées d'un aperçu sur les rapports de la menstruation et de l'aliénation mentale. *Annales Médico–psychologiques*, **3**, 574–610.
3. Schlager, L. (1858). Die Bedeutung des Menstrualprocesses und seiner Anomalieen für die Entwicklung und den Verlauf der psychischen Störungen. *Zeitschrift für Psychiatrie*, **15**, 459–98.
4. Bancroft, J. (1993). The premenstrual syndrome—a reappraisal of the concept and the evidence. *Psychological Medicine*, (Suppl. 24), 1–47.
5. von Krafft-Ebing, R. (1902). *Psychosis Menstrualis. Eine klinisch–forensische Studie*. Enke, Stuttgart.
6. Brockington, I.F. (2005). Menstrual psychosis. *World Psychiatry*, **4**, 9–17.
7. Kitayama, I., Yamaguchi, T., Harada, M., *et al.* (1984). Periodic psychoses and hypothalamo-pituitary function. *Mie Medical Journal*, **34**, 127–38.
8. Parker, P.J. (1983). Motivation of surrogate mothers: initial findings. *The American Journal of Psychiatry*, **140**, 117–18.
9. Bivin, G.D. and Klinger, M.P. (1937). *Pseudocyesis*. Principia Press, Bloomington.
10. Ekblad, M. (1963). Social-psychiatric prognosis after sterilisation of women without children. *Acta Psychiatrica Scandinavica*, **39**, 481–514.
11. Cooper, P., Gath, D., Rose, N., *et al.* (1982). Psychological sequelae to elective sterilisation: a prospective study. *British Medical Journal*, **284**, 461–4.
12. Bledin, K.D., Cooper, J.E., MacKenzie, S., *et al.* (1984). Psychological sequelae of female sterilisation: short-term outcome in a prospective controlled study. *Psychological Medicine*, **14**, 379–90.

13. Gath, D., Rose, N., Bond, A., *et al*. (1995). Hysterectomy and psychiatric disorder: are the levels of psychiatric morbidity falling? *Psychological Medicine*, **25**, 277–83.
14. Martin, R.L., Roberts, W.V., and Clayton, P.J. (1980). Psychiatric status after hysterectomy. A one-year prospective follow-up. *The Journal of the American Medical Association*, **244**, 350–3.
15. Aziz, A., Bergquist, C., Nordholm, L., *et al.* (2005). Prophylactic oöphorectomy at elective hysterectomy. Effects on psychological well-being at 1-year follow-up and its correlations to sexuality. *Maturitas*, **16**, 349–57.
16. Cohen, R.L. (1988). *Psychiatric consultation in childbirth settings*. Plenum, New York.
17. Cartwright, A. (1988). Unintended pregnancies that lead to babies. *Social Science and Medicine*, **27**, 249–54.
18. Brockington, I.F. (2006). *Eileithyia's mischief: the organic psychoses of pregnancy, parturition and the puerperium*. Eyry Press, Bredenbury.
19. Wessel, J. and Buscher, U. (2002). Denial of pregnancy: a population based study. *British Medical Journal*, **324**, 458.
20. Nirmal, D., Thijs, I., Bethel, J., *et al.* (2006). The incidence and outcome of concealed pregnancies among hospital deliveries: an 11-year population-based study in South Glamorgan. *The Journal of Obstetrics and Gynaecology Research*, **26**, 118–21.
21. Condon, J.T. (1987). The battered foetus syndrome. *The Journal of Nervous and Mental Disease*, **175**, 722–5.
22. Weir, J.G. (1984). Suicide during pregnancy in London 1943–1962. In *Suicide in pregnancy* (eds. G.J. Kleiner and W.M. Greston), pp. 59–62. Wright, Boston.
23. Ulleland, C.N. (1972). The offspring of alcoholic mothers. *Annals of the New York Academy of Sciences*, **197**, 167–9.
24. Lemoine, P., Harousseau, H., Borteyru, J.P., *et al.* (1968). Les enfants de parents alcooliques: anomalies observées. *Ouest-Médical*, **25**, 476–82.
25. Brinch, M., Isager, T., and Tolstrup, K. (1988). Anorexia nervosa and motherhood: reproduction pattern and mothering behaviour of 50 women. *Acta Psychiatrica Scandinavica*, **77**, 611–17.
26. Morgan, J.F., Lacey, J.H., and Sedgwick, P.M. (1999). Impact of pregnancy on bulimia nervosa. *The British Journal of Psychiatry*, **174**, 135–40.
27. Jureidini, J. (1993). Obstetric factitious disorder and munchausen syndrome by proxy. *The Journal of Nervous and Mental Disease*, **181**, 135–7.
28. Kendell, R.E., Chalmers, J.C., and Platz, C. (1987). Epidemiology of puerperal psychoses. *The British Journal of Psychiatry*, **150**, 662–73.
29. Grof, P., Robbins, W., Alda, M., *et al*. (2000). Protective effect of pregnancy in women with lithium-responsive bipolar disorder. *Journal of Affective Disorders*, **61**, 31–9.
30. Brockington, I.F. (1996). *Motherhood and mental health*, pp. 92, 111–12. Oxford University Press, Oxford.
31. Engelhard, J.L.B. (1912). Über Generationspsychosen und der Einflus des Gestationsperiode auf schon bestehende psychische und neurologische Krankheiten. *Zeitschrift für Geburtshülfe und Gynäkologie*, **70**, 727–812.
32. Gissler, M., Hemminki, E., and Lönnqvist, J. (1996). Suicides after pregnancy in Finland, 1987–1994: register linkage study. *British Medical Journal*, **313**, 1431–4.
33. David, H.P. (1985). Post-abortion and post-partum psychiatric hospitalization. *CIBA Foundation Symposium*, **115**, 150–64.
34. Nagy, M. (1948). The child's theories concerning death. *The Journal of Genetic Psychology*, **73**, 3–27.
35. Smialek, Z. (1978). Observations on immediate reactions of families to sudden infant death. *Pediatrics*, **62**, 160–5.
36. Hughes, P., Turton, P., Hopper, E., *et al.* (2002). Assessment of guidelines for good practice in psychosocial care of mothers after stillbirth: a cohort study. *Lancet*, **360**, 114–48.
37. Condon, J. T. (1986). Psychological disability in women who relinquish a baby for adoption. *The Medical Journal of Australia*, **144**, 117–19.
38. Brockington, I.F., Macdonald, E., and Wainscott, G. (2006). Anxiety, obsessions and morbid preoccupations in pregnancy and the puerperium. *Archives of Women's Mental Health*, **9**, 253–64.
39. Bydlowski, M. and Raoul-Duval, A. (1978). Un avatar psychique méconnu de la puerpéralité: la nevrose traumatique post-obstétricale. *Perspectives Psychiatriques*, **4**, 321–8.
40. Wenzel, A., Haugen, E.N., Jackson, L.C., *et al.* (2005). Anxiety symptoms and disorders at eight weeks postpartum. *Journal of Anxiety Disorders*, **19**, 295–311.
41. Hertzberg, T. and Wahlbeck, K. (1999). The impact of pregnancy and puerperium on panic disorder: a review. *Journal of Psychosomatic Obstetrics and Gynaecology*, **20**, 59–64.
42. De Armond, M. (1954). A type of post partum anxiety reaction. *Diseases of the Nervous System*, **15**, 26–9.
43. Sved-Williams, A.E. (1992). Phobic reactions of mothers to their own babies. *The Australian and New Zealand Journal of Psychiatry*, **26**, 631–8.
44. Weightman, J., Dalal, B.M., and Brockington, I.F. (1999). Pathological fear of cot death. *Psychopathology*, **167**, 246–9.
45. Maina, G., Vaschetto, P., Ziero, S., *et al.* (2001). Il postpartum come fattore di rischio specifico per l'esordio del disturbo ossessivo-compulsivo: studio clinico controllato. *Epidemiologia e Psychiatria Sociale*, **10**, 90–5.
46. Labad, J., Menchon, J.M., Alonso, P., *et al.* (2005). Female reproductive cycle and obsessive-compulsive disorder. *The Journal of Clinical Psychiatry*, **66**, 428–35.
47. Jennings, K.D., Ross, S., Popper, S., *et al.* (1999). Thoughts of harming infants in depressed and non-depressed mothers. *Journal of Affective Disorders*, **54**, 21–8.
48. Gordon, R.E. and Gordon, K.K. (1959). Social factors in the prediction and treatment of emotional disorders of pregnancy. *American Journal of Obstetrics and Gynecology*, **77**, 1074–83.
49. Affonso, D.D., De, A.K., Horowitz, J.A., *et al.* (2000). An international study exploring levels of postpartum depressive symptomatology. *Journal of Psychosomatic Research*, **49**, 207–16.
50. Malmquist, A. and Kaij, L. (1971). Motherhood and childlessness in monozygotic twins. Part II. The influence of motherhood on health. *The British Journal of Psychiatry*, **118**, 22–8.
51. Høyer, G. and Lund, E. (1993). Suicide among women related to number of children in marriage. *Archives of General Psychiatry*, **50**, 134–7.
52. Pound, A., Puckering, C., Cox, A., *et al.* (1988). The impact of maternal depression on young children. *British Journal of Psychotherapy*, **4**, 240–52.
53. Cox, J.L., Holden, J.M., and Sagovsky, R. (1987). Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. *The British Journal of Psychiatry*, **150**, 782–6.
54. Epperson, C.N., Czarkowski, K.A., Ward-O'Brien, D., *et al.* (2001). Maternal sertraline treatment and serotonin transport in breast-feeding mother-infant pairs. *The American Journal of Psychiatry*, **158**, 1631–7.
55. Brockington, I.F., Aucamp, H.M., and Fraser, C. (2006). Severe disorders of the mother-infant relationship: definitions and frequency. *Archives of Women's Mental Health*, **9**, 243–52.
56. Righetti-Veltema, M., Conne-Perréard, E., Bousquet, A., *et al.* (2002). Postpartum depression and mother-infant relationship at 3 months. *Journal of Affective Disorders*, **70**, 291–306.
57. Bernazzani, O., Marks, M.N., Bifulco, A., *et al.* (2005). Assessing psychosocial risk in pregnant/postpartum women using the contextual assessment of maternity experience (CAME). *Social Psychiatry and Psychiatric Epidemiology*, **40**, 497–508.
58. Murray, L., Hipwell, A., and Hooper, R. (1996). The cognitive development of 5-year-old children of postnatally depressed mothers. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **37**, 927–35.

59. Ndosi, N.K. and Mtawali, M.L. (2002). The nature of puerperal psychosis at Muhimbili national hospital: its co-morbidity, and associated main obstetric and social factors. *African Journal of Reproductive Health*, **6**, 41–9.
60. Srinavasan, K. (1983). Cerebral venous and arterial thrombosis in pregnancy and the puerperium: a study of 135 patients. *Angiology*, **34**, 731–46.
61. Osiander, F.B. (1797). *Neue Denkwürdigkeiten für Aerzte und Geburtshelfer*, pp. 52–128. Rosenbusch, Göttingen.
62. Esquirol, J.E.D. (1818). Observations sur l'aliénation mentale à la suite de couches. *Journal Général de Médecine, de Chirurgie et de Pharmacie Françaises et Étrangères* (Series 2), **1**, 148–64.
63. Terp, I.M. and Mortensen, P.B. (1998). Post-partum psychoses: clinical diagnoses and relative risk of admission after parturition. *The British Journal of Psychiatry*, **172**, 521–6.
64. Jones, I. and Craddock, N. (2001). Familiarity of the puerperal trigger in bipolar disorder: results of a family study. *The American Journal of Psychiatry*, **158**, 913–17.
65. Main, T.F. (1958). Mothers with children in a psychiatric hospital. *Lancet*, **ii**, 845–7.
66. Tardieu, A. (1868). *Étude Médico–Légale sur l'Infanticide*. Baillière, Paris.
67. Wilkey, I., Pearn, J., Petrie, G., *et al*. (1982). Neonaticide, infanticide and child homicide. *Medicine, Science and the Law*, **22**, 31–4.
68. Somander, L.K.H. and Rammer, L.M. (1991). Intra- and extrafamilial child homicide in Sweden 1971–1980. *Child Abuse & Neglect*, **15**, 45–55.

5.5

Management of psychiatric disorders in medically ill patients, including emergencies

Pier Maria Furlan and Luca Ostacoli

The coexistence of psychiatric disorders in patients with medical illnesses may influence both the diagnosis and the course of the illness by their effects on pathophysiological, diagnostic, and therapeutic processes. There may also be effects on patients' collaboration with treatment and on their relationships with health care staff. Several factors change the management of, medical illnesses and psychiatric disorders, and their inter-relation

- increased life-expectancy and increasing survival of people with-severe illness alter the risk of other medical and psychiatric disorders;
- social changes affecting family structure can affect care giving. Other social factors include changes in the role of women (work, delayed maternity); increased immigration with consequent cultural diversity including different concepts of medical and psychiatric disorders (see Chapter 1.3.2);
- increased use of medication in medical and in psychiatric treatment, and changes in the organization of health care and social assistance from hospital-based to community-based.

This chapter describes how to recognize, treat and manage psychiatric disorders in medical illnesses.

The frequency of psychiatric disorders among the medically ill

The prevalence of psychiatric disorders in medical illnesses ranges from 16–60 per cent depending on the research methodology (self-reports or interviews; inclusion or exclusion of somatic items), the setting (out-patient or hospitalized), and the sample. In general, the frequency of psychiatric disorders in patients with heart disease[1] (coronary disease, heart failure), gastrointestinal diseases (irritable bowel syndrome), lung diseases (asthma, chronic bronchitis), and diabetes is 15–20 per cent. In patients with cancer and chronic pain it is 30–40 per cent; and in neurological diseases (Parkinson's, multiple sclerosis, epilepsy) and dialysis it is 50 per cent. Ten to 20 per cent of patients have sub-threshold symptoms that nevertheless influence psychosocial functioning. The prevalence of psychiatric disorders among family members of people with chronic disabling conditions is only slightly lower. The most frequent are organic mental disorders (5–44 per cent), followed by substance abuse (10–25 per cent), anxiety disorders (10–30 per cent), mood disorders (9–13 per cent), personality disorders (6–9 per cent), somatoform disorders (5–9 per cent), mania and psychosis (1 per cent). Recognition by medical doctors is below 50 per cent and the referral rate to liaison services is approximately 1–3 per cent.

The frequency of medical illnesses among psychiatric patients

The most severe psychiatric disorders are frequently associated with social isolation, difficult relations with health-care providers, poor adherence to treatment, unhealthy lifestyle[2] (nutrition, smoking, hygiene), side-effects of medication and substance dependence. The presence of psychiatric symptoms can also lead to failure to recognize physical symptoms. And yet some medical illnesses are more frequent in people with schizophrenia than in the general population.[3] These conditions are cardiovascular risk (9.4 per cent in men, 7 per cent in women), diabetes (13 per cent), hypertension (27 per cent) and chronic conditions in general (41 per cent).[4]

Table 5.5.1 Prevalence of medical illnesses in patients with mood disorders

Disease	%
Hypertension	18.1–34.8
Stroke	1.7–1.9
Headache	19.3
Chronic pulmonary disease	10.6–12.9
Hypothyroidism	9.6
Obesity	4.6
Alcohol abuse	12.2–24.7
Nicotine	9.1–12.6
Illicit drug abuse	9.7

The diagnosis of psychiatric disorder in medically-ill patients

Anxiety, fear, demoralization, a sense of loss, decreased pleasure, and thoughts of death are frequent in advanced debilitating physical disease even when there is no coexisting anxiety or depressive disorder.

Physical disease and its treatment may cause somatic symptoms similar to those of psychiatric disorders. And the 'aetiological' criterion of the DSM-IV-TR that requires exclusion of a physical cause is often difficult to apply in advanced medical illness, as are the criteria for depression. Endicott[5] proposed replacing the four somatic items for depression (fatigue, insomnia, weight-loss, and difficulty in concentrating) with four psychological symptoms: depressed appearance, social withdrawal, brooding, non-reactive mood. However, this proposal risks excluding somatic symptoms which are a core manifestation of more severe forms of depressive disorder. In doubtful cases, an inclusive approach to somatic symptoms is preferable, and the risk of severe psychiatric disorders should not be underestimated.

Self-abasement and guilt are less frequent in medically ill patients. In assessing guilt, ethnic and cultural factors must be taken into account, for example, feelings of guilt are uncommon in depressed Arabs, whereas somatization is common.

Some syndromes in medically ill patients do not correspond to standard diagnostic categories but nevertheless influence functioning and the course of disease. The Diagnostic Criteria for Psychosomatic Research[6] mention illness denial, thanatophobia, demoralization, and alexithymia. In medical illness, psychiatric disorders may manifest with somatic symptoms (see Chapter 5.2.3).

Table 5.5.2 gives some broad indications for the differential diagnosis of psychiatric disorders in the presence of medical illness.[7]

To be classed as a psychological reaction, the psychiatric disorder must develop at the same time as the onset of the medical illness or the treatment. In some conditions such as, pancreatic cancer, multiple sclerosis, the onset of psychiatric disorder may precede the recognition of the medical illness (e.g. Multidimensional evaluation of the care requirements is essential and codified approaches exist).[8]

Atypical symptoms occur in psychiatric disorders due to medical conditions. Drium is often complex with auditory hallucinations prevailing, whereas tactile, olfactory, and gustatory hallucinations are rare.

Causes of psychiatric illness among medical patients

These are both psychological (see Chapter 5.6). and medical (see Chapter 5.3. 4.

Course and prognosis

If properly treated, psychiatric disorders in medically ill patients have the same prognosis as those occurring without medical illness, except in some very advanced and debilitating cases, and in these, the few reported studies give contrasting results. Psychiatric disorders may significantly influence the outcome of the medical condition. Depression is associated with an increased risk for subsequent development of ischaemic heart disease, Parkinson's disease, Alzheimer's disease (and other dementias) and medical diseases in general. It is an independent predictor of severe complications in diabetes and of mortality in ischaemic heart disease, heart failure,[9] stroke, dementia, cancer and HIV. Anxiety may exacerbate angina, arrhythmia, asthma, movement disorders, hypertension and irritable bowel syndrome and is associated with increased health-seeking behaviour and prescription of inappropriate drugs.

Delirium is reversible in 70–80 per cent of cases, but in terminally ill patients may be progressive and intractable, and is associated with increased short-term mortality.[10] Mania and psychosis may worsen the medical outcome due to behavioural alterations, poor adherence and increased drug adverse effects.

Table 5.5.2 Differential diagnosis among psychiatric disorders (PD) in medical illnesses (MI)

	Comorbidity (MI – PD)	Latent (PD)	Psychological Reaction	Psychoorganic PD
Onset with mi	-	+	+	+/-
Medical aetiology	-	-/+	-	+
Life events	-	+/-	+	-
Personal/family medical history	+	+	-	-
Cognitive disorders	-	-	-	+
Altered awareness	-	-	-	+
Fluctuation in severity of psychic symptoms	-	-	+	+
Atypical psychic symptoms	-	-	-	+
Self-abasement	+	+/-	-	-
Family history for psychic symptoms	+	+/-	-	-
Empathy of doctor	-/+	+/-	+	-
Response to psychiatric treatment	+/-	+	+	-

Treatment

In specific populations of patients such as those with diabetes, asthma, myocardial infarction, irritable bowel syndrome, cancer, Parkinson's disease, multiple sclerosis, and rheumatic diseases, psychosocial interventions and a variety of psychological treatments have a positive effect on psychiatric disorders, and the quality of life and relationships.[11] In medical conditions requiring active patient participation, psychological treatments improve adherence to the therapeutic programmes. In diabetes they reduce glycosylated haemoglobin, in Parkinson's they produce cognitive and motor improvement. Such treatments can reduce physical symptoms including pain, nausea, dyspnoea and disability. For other indices such as mortality in heart attack, longevity in cancer, severity of hypertension and peptic ulcer, and inflammatory activity in rheumatoid arthritis, there are psychological benefits and improvements in quality of life but poor effects on medical outcome. Indeed, in the more severe psychiatric disorders, combined treatment with psychopharmacological drugs is more effective. Studies of cost-effectiveness and length of hospitalization have reported conflicting results.[12]

Management

The consultation process

(a) Forming an alliance with the medical team

The psychiatrist should initially aim to work with medical staff on their clinical rounds, and interview patients who have no psychiatric disorder to learn what it means to have the medical condition and to undergo its treatment. The psychiatrist should be present at informal discussions, such as those during coffee breaks,and share the clinical team's emotional experiences.

(b) Interview with medical team

The aim is to evaluate the clinical situation, review the medical records; identify the most significant reason for referral and why the consultation has been made at this time; identify the team's approach to the patient, clarify whether the patient has been informed of the consultation and in what way. If the referral seems inappropriate for the patient, does this reflect a problem within the medical team such as burn out?

(c) Interview with the patient

The interview should move dynamically between the 'objective' position (clinical data, psychosocial information) and the 'subjective' position. Feelings evoked in the psychiatrist frequently mirror perceptions of the patient and the medical team. To recognize them helps empathetic relations and emotional containment. A protocol for the interview is shown below. [13]

(d) Reporting back to the medical team

Reporting should be clear and concise with both a verbal and a written report. The focus should be on the reason for referral. Risk factors and points of strength should be identified. A verbal report of an 'image' such as an episode fom the patient's experience, a memory or even a dream can sometimes aid empathetic understanding by the medical team. Practical advice on management should be provided.

(e) Psychopharmacological treatment

The psychopharmacological treatment of medically ill patients may be difficult for several reasons including the stigma associated with psychiatric disorders, weariness with the many medical treatments already undergone, increased sensitivity to side-effects due to pharmacokinetic changes produced by interactions with medical drugs and any underlying liver or renal disease.[14] However when adherence is adequate, the therapeutic response is similar to that of patients without medical illnesses. Nevertheless, the prescription of medication must not replace receptiveness to the patients problems and emotional support, which are frequently the most effective intervention.

It is often useful to offer the patient a drug not for symptoms such as depression when these arise in a discouraging physical situation because this may be seen as disparaging—but for other

Goal	Approach
Overcome stigma	The medical team: present the psychiatrist to the patient as a team member and motivate consultation The psychiatrist: explores any ambivalence or negative feelings empathetically (it may not be clear to the patient why a psychiatrist is discussing his/her disorder)
Open questions: background and information	Why is the patient hospitalized? What does he/she think of the illness?
Principal emotions and fears	What does the patient feel? What is he/she most afraid of?
Consider constructive ways of coping; and discuss dysfunction	What and who most help the patient to overcome difficulties; who is important, what other resources are available
Completing the information	Medical history and current quality of life
Develop a shared understanding of the situation in which medical and psychological aspects are linked and not viewed as alternatives.	"Normalize" emotional disorders as reactions to illness that may amplify symptoms and influence course. Describe physical mechanisms of symptom production such as muscular contraction, vasodilation/constriction, hyperventilation, asthenia, inactivity, immune defences. Use clear, descriptive language with imagery: e.g. tension is like a tight shoe.
Defining goals and reducing unrealistic expectations	Discuss the main problems Aim to improve problems, not solve them. Set realistic goals
Propose a treatment plan	What interventions, "first steps" Practical advice about day-to-day matters such as interpersonal relations, how to reduce tension

realistic goals. The specific contribution that the drug can make to achieving these goals should be explained and a clear description provided of its somatic effects. The psychiatrist should also:

- Aim to simplify treatment as far as possible. The distinction between 'psychiatric' and 'medical' drugs is arbitrary since many drugs have both physical and psychological effects.
- Rationalize treatment using the fewest possible drugs. First choice should be drugs that act on more than one symptom, psychological or medical (e.g. reduce agitation and nausea, insomnia and hyporexia, depression and headache, pain and anxiety).
- Consider side-effects and interactions before deciding treatment including their time of onset. Thus, SSRI should be prescribed at least one week apart from any treatment with significant gastrointestinal side-effects. Investigate complementary drug use: e.g. 20–40 per cent of oncology patients take herbal remedies, many of which have significant pharmacological interactions (see Chapter 6.2.9).
- Prefer drugs that were effective for the patient in the past; starting with low doses and increasing them gradually.
- Sometimes the best action is to discontinue a drug.

Psychosocial treatments

Differences from traditional psychotherapy

The major aims are to control distress, maintain self-respect, maintain significant relations and doctor-patient communication, work through information, and develop adaptive coping mechanisms.

Timing: At times of crisis (e.g. immediately after the communication of a serious diagnosis), the need for control is paramount. At the onset of complex diseases, short cycles of interviews may prevent subsequent psychiatric disorder.

Flexibility: This is needed due to variations over time in the conditions of treatment (in hospital, or as a day patient or outpatient), the symptoms, and the motivation of the patient, who may alternate between the need for emotional sharing and moments of self-withdrawal. Existential uncertainty is reflected in relations with the psychiatrist and each interview is an entity in itself.

Eclecticism: The complexity of the situation often requires integration of different approaches at different times: expressive, cognitive, body-mediated or psycho educational. The intervention may focus on the patient, on family members or on the medical team. Flexibility and eclecticism must originate from the integration of the patient's needs and the psychiatrist's empathic and comprehensive evaluation of these needs.

Regression: Illness, hospitalization, fear, and the 'invalid role' may make a patient, who would otherwise refuse it, become receptive to the psychiatrist's intervention. During medical illness, emotional defences are more fluid, the relationship with the psychiatrist may form more rapidly and short interventions can be effective. After discharge the psychiatrist must be ready to change approach.

Existential context: A severe disease often casts doubt on the meaning of existence. Patients may fall into despair or ask themselves what really matters. They may want to reorganize their lives around new priorities. In these circumstances, interviews should focus on the present, abandoning the 'past-future' approach, and define the psychiatric disorder as an accentuation of natural human emotions. Fear should be dealt with directly, including fear of dying, and the patient's relatives, including their doctors, should be helped to provide support since fear is greater when experienced alone.

Physical contact: Faced with severe disease, some patients 'speak' only after having been touched. Sometimes bodily contact, for example through simple massage, may keep open communication with family members and staff, even without the use of words. A glance, a voice, the sense of touch, or other bodily sensations may form an intense dialogue between patient and therapist and be the most effective way to understand how aware the patient is of his/her condition.

Treatment of urgent situations

Prevention

- Medical treatment should address the overall quality of life, including the gap between expectations and reality. Many medical interviews target practical matters but place less emphasis on working through expectations. A doctor-patient relationship capable of offering relief while gradually reducing the gap between hope and reality is fundamental.
- The patient's chief supports are family members and caregivers. Improving their psychological skills through simple psychoeducational programmes may be more cost-effective than generally increasing psychiatric consultations.
- Psychosocial and biological risk factors should be identified early, including both current factors and those in the medical history.

Acute anxiety

Treatment of acute anxiety should maximize the beneficial effects of the doctor-patient relationship as well as providing specific psychosocial and pharmacological interventions. At times of crisis, one-on-one companionship is useful, sometimes with simple relaxation or massage that family members can provide. When a serious medical diagnosis is communicated, it is helpful to listen, and for medical staff to be receptive to emotional reactions in the subsequent 24 hours; solitude should be alleviated, if necessary with the help of voluntary workers. Benzodiazepines should be prescribed only when really necessary and for short periods to reduce the risk of tolerance and addiction. Benzodiazepines may reduce respiratory function further in patients who retain CO_2, thereby worsening their asthenia. If prolonged pharmacological therapy is necessary, sedative antidepressants may be an alternative to benzodiazepines.

Panic attacks: Somatic symptoms of a panic attack may be confused with an excacerbation of medical conditions such as chest pain, irritable bowel syndrome, and asthma. Treatment with a selective serotonin reuptake inhibitor is usually effective but half of these patients require long-term treatment since relapse is common after discontinuation. Psychological treatment[15] (see Chapter 4.7.3) can be effective.

Post-traumatic stress disorder: Stress factors in medical contexts may cause and prolong this disorder. Such factors include acute medical events, intensive care, post-confusional reactions, communication of a serious diagnosis, and presence at unexpected deaths. Pharmacological therapy has limited results but psychological treatments may be effective (see Chapter 4.6.2), Risk prevention is important.

Depression

Treatment should be based on prevention, and psycho education, psychosocial and pharmacological interventions.

Antidepressants: The choice of drug should be based on the side-effect profile and interactions (see Chapter 6.2.3). The somatic effect of treating depression can be important. Cognitive and motor functions can improve in stroke patients, glycaemic control can improve in diabetics, as can chronic pain, and dyspnoea in lung-disease. SSRIs may produce gastrointestinal side effects, bleeding due to platelet dysfunction, so that monitoring is essential in patients on anticoagulants. Venlafaxine in high doses may increase blood pressure. Mirtazapine has no sexual side-effects, has anti-nausea activity and stimulates the appetite, but may cause weight gain and sedation. TCAs, being anticholinergic, affecting heart conduction and the peripheral autonomic nervous system, are contraindicated in heart disease, cognitive impairment, orthostatic hypotension, hypertrophic prostate, glaucoma and epilepsy. Secondary amines, such as nortriptyline and desipramine, are preferable.

Methylphenidate may be useful in particular medical situations, such as palliative care or advanced cancer, to alleviate fatigue, increase appetite, and reduce opiate-induced sedation. It may elevate mood and has rapid onset. Agitation and insomnia are the main problems.

Personality disorders

Personality disorders, particularly antisocial and borderline disorders, may create emergencies due either to aggressive behaviour towards self or others or to conflict within the team. Screening for substance abuse and depression is necessary since these increase impulsiveness. Patients can become aggressive when their demands are not met, e.g. for increased painkillers, or interviews with their doctors. Staff require support to manage conflicts and frustration, and to maintain a caring approach.[16]

Staff must be empathetic but at the same time act as a team. Limits and rules should be established at admission treatment to prevent escalation of aggression. Interviews must be in conditions of safety; if necessary with other staff or even police present. The agreed plan must be respected by all staff, avoiding the extremes of excessive permissiveness and excessive rigidity, which may be induced by the patient's disorder. Patients with personality disorders often use truth-based observations pathologically to exploit the therapist's and institution's weak points. Recognizing what is true in their criticisms may improve patient management.

Delirium and dementia (see Chapters 4.1.1 and 4.1.13)

Mania

A manic crisis affects the care of the medical condition dramatically both through excessive activity, weight-loss, reduction of sleep, and through non-compliance with treatment and conflict with staff. The medical history should be evaluated for disease-related stress factors which may trigger crises. Manic symptoms recognized early together with possible causes of insomnia such as pain, nocturia, dyspnoea, or environmental factors including noise, unsuitable temperature, frequent entrance of night-staff for other patients. Where possible, precipitating factors should be removed. When this is impossible (e.g. a requirement for high-dose cortisone), treatment is as that of mania in other situations (see Chapter 4.5.8) taking medication side-effects into account (see Chapter 6.2.4). Antipsychotic drugs are generally faster acting at lower doses than in primary mania. More potent antipsychotics are preferable since they have lower anticholinergic and alpha-blocking effects. It is useful to focus on symptoms the patient finds disturbing rather than ego-syntonic symptoms. Restraint or hospitalization in a psychiatric ward is necessary when hyperactivity could compromise physical safety or the underlying condition.

Psychosis

Close collaboration between the psychiatrist and the medical team is necessary, with the most appropriate environment for treatment, medical or psychiatric unit, being decided case-by-case. Medical investigations and interviews should be simplified and clear basic information provided. Continuity of care among medical professionals is important. The psychological effect of medical treatment should be assessed. When there are hallucinations, medical staff should be educated about their nature and supported, so that they do not criticize any content expressed by the patient, but are able to empathize with the patient's distress. Management may require constant observation and in some cases restraint and involuntary treatment. Medically ill patients are especially sensitive to the side-effects of antipsychotic drugs but generally respond to lower doses. The choice of drug and dose are based on the side-effects and the underlying disease (see Chapter 6.2.2)). Cardiac pathology and QTc enlargement require ECG monitoring and ziprasidone, haloperidol, chlorpromazine, and thioridazine are contraindicated. Caution is also necessary in alcoholics, and patients with hypokalemia and hypomagnesemia.[17]

Extra-pyramidal effects of high doses of typical neuroleptics may cause laryngospasm and affect the diaphragm, worsening any respiratory insufficiency. For chronic treatment, newer drugs are preferable. Caution is required in hepatic and renal diseases and where there is a risk of epileptic fits. Fluid intake should be monitored carefully because dehydration increases the risk of neuroleptic malignant syndrome. In secondary psychosis, the family should be informed and supported and, after the psychotic episode has been resolved, the patient must be helped to work through the disorientating experience.

Aggression and restraint

Aggression should be assessed, thoroughly defining the sequence of events that preceded it and the role of medical and psychiatric factors, stressful events and environmental factors. Major predictors related to the patient include a history of violence, difficulty in communication, and the psychological condition. Predictors related to the environment include: crowding, and a medical team unreceptive to the patient's discomfort. The best strategy is prevention through providing information in a clear and empathetic way, recognizing risk situations and early signs of agitation, and using de-escalation techniques if necessary with pharmacological support (see also Chapter 11.14).

Environmental measures to limit psychomotor agitation include: making the patient as comfortable as possible, reducing background noise, removing potentially harmful objects, and identifying situations of which the patients is intolerant. Patients can be

distracted by walking or occupational activities, avoiding a forced stay in bed and making the timetable flexible. If possible, family members can be asked to collaboration.

Where permitted under national law, restraint may become necessary during unmanageable agitation and aggression in the course of an acute psychiatric disorder, in confusional states, especially in the elderly and aftersurgery, with metabolic diseases or during abstinence from alcohol and drugs (seen especially during the first few days of hospitalization).

If restraint is permissible, staff training is necessary because improper use may cause injury to patients or staff. Restraint should be used only when strictly necessary and for the shortest possible time, with constant monitoring of its continuing necessity and the patient's condition.

In medically ill patients restraint may increase the risk of thromboembolism, reduce respiratory function with increased risk of infection, increase cognitive deficit, pressure ulcers, urinary and intestinal incontinence. When restraint is removed there is an increased risk of falls.

Deliberate self-harm and suicidality (see also Chapter 4.15.4)

In advanced medical illness, suicidal ideation is frequent as an expression of the wish for release from suffering, to regain control over a condition that is perceived as unstoppable, and is wearing the patient down. The actual risk of suicide is slightly above that of the general population. Specific risk factors are: chronic disease, uncontrolled physical symptoms such as pain or dyspnoea, disfiguring surgery, unrecognized psychiatric disorder (delirium, depression, personality disorder), substance abuse, and lack of social support. The most frequent form of self-harm is drug overdose, particularly of analgesics and psychopharmacological drugs. Most lethal suicidal acts are of the impulsive type, in particular jumping. [18] The doctor-patient relationship is often the best help against loss of hope. Safety of the environment is crucial since medical contexts facilitate access to potentially lethal objects and to the possibility of jumping. In cancer, the risk is higher in the first year, decreasing with time after diagnosis. In the end stage of illness e.g. in renal failure, treatment withdrawal may have to be considered, with the: clinical and ethical problems of distinguishing a rational decision by the patient from one affected by a psychiatric disorder or another potentially modifiable factor.

Patients refusing treatment (see also Chapter 1.5.1)

Background culture should be taken into account as some 'overall rejections' are actually related to this rather than to factors in the individual. An alliance with family members is necessary. Factors in the doctor-patient relationship should be evaluated, together with the information that the patient has received. In some cases refusal relates to fear, anger or despair. If it seems that the patient would feel belittled by a psychiatric consultation it is better to support medical staff who are receptive to the patient's emotional sate. This is done patients often then accept psychiatric support more readily. In complex cases, the history of the patient's attitude to treatment should be evaluated on admission to treatment. Aa history of interrupted treatment, changes of doctor or diffidence towards caregivers, may indicate a need for greater efforts to develop good doctor-patient relationships. When obtaining items of medical history from family members, it is important to identify whom the patient trusts most. In relations with the medical team, the most appropriate figure for dialogue should be identified. In some cases, informal conversations, for example in the evening, are more useful than formal medical interviews.

Further information

Levenson, J.L., (2005). *Textbook of psychosomatic medicine*. American Psychiatric Publishing, Washington, DC.

Wise, M.G. and Rundell, J.R., (2005). *Clinical manual of Psychosomatic medicine: A Guide to Consultation–Liaison Psychiatry*. American Psychiatric Press, Washington, DC.

Guthrie, E. and Creed, F. (1996). *Seminars in Liaison Psychiatry*. College seminars series. Royal College of Psychiatrists, London.

Web sites

The European Association for Consultation Liaison Psychiatry and Psychosomatics. http://www.eaclpp.org

International Organization for Consultation- Liaison Psychiatry. http://www. med.monash.edu.au/psychmed/ioclp.

The International College of Psychosomatic Medicine. http://www.icpm.org

References

1. Rutledge, T., Reis, V.A., Linke, S.E., *et al.* (2006). Depression in heart failure a meta-analytic review of prevalence, intervention effects, and associations with clinical outcomes. *Journal of American College of Cardiology,* **48**(8), 1527–37.
2. Michael, T., Compton, M.T., Daumit, G.L., *et al.* (2006). Cigarette Smoking and Overweight/Obesity Among Individuals with Serious Mental Illnesses: A Preventive Perspective. *Harvard Review of Psychiatry,* **14**(4), 212–22.
3. Carney, C.P. and Jones, L., (2006). Medical Comorbidity in women and men with bipolar disorders: a population – based controlled study. *Psychosomatic Medicine,* **68**, 684–91.
4. Goff, D.C., Sullivan, L.M., McEvoy, J.P., *et al.* (2005). A comparison of ten-year cardiac risk estimates in schizophrenia patients from the CATIE study and matched controls. *Schizophrenia Research,* **80**, 45– 53.
5. Endicott, J., (1984). Measurement of depression in patients with cancer. *Cancer,* **53**(Suppl 10), 2243–9.
6. Fava, G.A., Freyberger, H.J., Bech, P., *et al.* (1995). Diagnostic criteria for use in psychosomatic research. *Psychotherapy Psychosomatic,* **63**(1), 1–8.
7. Lipowski, Z.J., (1967) Review of consultation psychiatry and psychosomatic medicine, II: clinical aspects. *Psychosomatic Medicine,* **29**, 201–24.
8. Huyse, F.J., Lyons, J.S., Stiefel, F.C., *et al*, (1999). INTERMED: a method to assess health service needs. I. Development and reliability. *General Hospital Psychiatry,* **21**, 39–48.
9. Barth, J., Schumacher, M. and Lingen, C.H. (2004). Depression as a Risk Factor for Mortality in Patients With Coronary Heart Disease: A Meta-analysis, *Psychosomatic Medicine,* **66,** 802–13.
10. Breitbart, W., Gibson, C. and Tremblay, A., (2002). The delirium experience: delirium recall and delirium-related distress in hospitalized patients with cancer, their spouses/caregivers, and their nurses. *Psychosomatis,* **43**, 183–94.
11. Guthrie, E. and Creed, F. (1996). Treatment methods and their effectiveness. In *Seminars in Liaison Psychiatry* (eds. E., Guthrie, and F., Creed), pp. 238-73. College seminars series. Royal College of Psychiatrists, London.

12. Andreoli, P.B., Citero, Vde A. and Mari Jde, J., (2003). A systematic review of studies of the cost-effectiveness of mental health consultation-liaison interventions in general hospitals. *Psychosomatics,* **44**(6), 499–507.
13. Stuart, M.R. and Lieberman, J.A., (2002). *The fifteen minute hour. Practical therapeutic interventions in Primary care*. Saunders, Elsevier (USA). 14 Levenson JL, (2005). Psychopharmacology. In: Levenson JL, ed. *Textbook of psychosomatic medicine*, American Psychiatric Publishing, Washington, DC.
15. Shapiro, F. (2001). *Eye Movement Desensitization and Reprocessing: Basic Principles, Protocols and Procedures*. Guilford Press, New York.
16. Hay, J.L. and PassiK, S.D. (2000). The cancer patient with borderline personality disorder: suggestions for symptom-focused management in the medical setting. *Psychooncology,* **9**(2), 91–100.
17. Gasper, J.J. and Tsai, C. (2006). Community Behavioral Health Sciences, Community Health Network of San Francisco, San Francisco General Hospital, Revised: *Guidelines for the use of atypical antipsychotics in adults*.
18. Druss, B. and Pincus, H. (2000). Suicidal ideation and suicide attempts in general medical illnesses, *Arch Intern Med,* **160**(10), 1522–6.

5.6

Health psychology

John Weinman and Keith J. Petrie

Introduction

Health psychology is concerned with understanding human behaviour in the context of health, illness, and health care. It is the study of the psychological factors, which determine how people stay healthy, why they become ill, and how they respond to illness and health care.

Health psychology has emerged as a separate discipline in the past 30 years and there are many reasons for its rapid development. An important background factor is the major change in the nature of health problems in industrialized societies during the twentieth century. Chronic illnesses such as heart disease and cancer have become the leading causes of death, and behavioural factors such as smoking, diet, and stress are now recognized as playing a major role in the aetiology and progression of these diseases.[1] The provision of health care has grown enormously and there is an increased awareness of good communication as a central ingredient of medical care and of the importance of such factors as patient satisfaction and quality of life as key outcomes in evaluating the efficacy of medical interventions.

Although health psychology has developed over a similar time period to general hospital/liaison psychiatry and shares some common areas of interest, there are some clear differences between these two fields. Liaison psychiatry has a primary focus on hospital patients, particularly those experiencing psychological difficulties in the face of a physical health problem. In contrast, health psychology has a much broader focus on both healthy and ill populations and on the psychological processes that influence their level of health or their degree of adaptation to disease. Whereas health psychology has been mainly concerned with developing explanations based on theory, for health-related[2] and illness-related behaviour,[3] liaison psychiatry has concentrated on the diagnosis and treatment of either unexplained symptoms or psychiatric disorders occurring in people with medical conditions (see the other chapters in Part 5 of this volume).

In this chapter we provide an overview of the main themes and areas in health psychology. Four broad areas of behaviour will be reviewed, namely behavioural factors influencing health, symptom and illness behaviour, health care behaviour, and treatment behaviour. Inevitably such an overview is selective and the interested reader should seek out a more comprehensive introductory text[4,5] or more in-depth accounts of specific areas.[2,3]

Behavioural factors influencing health

A wide range of behavioural factors can influence health. In the following section there is a focus on stress, personality, and the main theories that have been developed to explain the variation in health-related behaviours.

Stress and health

The term 'stress' is usually used to describe situations, in which individuals are faced with demands that exceed their immediate ability to cope. Stressful situations are typically those that are novel, unpredictable, and uncontrollable as well as those involving change or negative events such as a loss. These situations can give rise to adverse psychological and physiological changes which, in turn, may result in disease.[6]

Stress may have indirect effects on health by increasing levels of risk behaviour (e.g. smoking, alcohol consumption), or may have direct effects on specific physiological mechanisms (e.g. increase in blood pressure) as well as affecting the individual's resistance to disease through suppression of the immune system, or by exacerbating or triggering a disease process in an already vulnerable individual.

A range of behavioural and emotional responses are shown by individuals as they attempt to cope with stressful situations and these are accompanied by autonomic, neuroendocrine, and immunological changes. During stressful episodes, releasing factors from the brain cause the pituitary to release ACTH which gives rise to the release of corticosteroids from the cortex of the adrenal glands. In addition to producing a number of well-known changes associated with the mobilization of both short- and longer-term physical resources (e.g. release of adrenaline (epinephrine) or noradrenaline (norepinephrine), release of glucose, activation of endorphins/encephalins, etc.), these steroids can also have effects on the immune system.[7]

The effects of stress on immunity have sparked the development of the new multi-disciplinary field of psychoneuroimmunology which focuses on the links between psychological, endocrine, and immunological processes (see Chapter 2.3.10). A large amount of work in this area has concentrated on the links between stress and immune function, but less work has focused on impaired immunity and the later development of disease. Acute stressors, such as examinations, or more chronic stressors, such as caring for

a dependent elderly relative, have been shown to lead to deleterious immunological changes. Work has also associated stress with a greater susceptibility to viral infection[8] as well as longer healing times for experimental puncture wounds[9] and wounds from surgical operations.[10] A recent meta-analysis of studies of stress and immunity shows substantial evidence for a relationship between stress and impaired immune system effectiveness, particularly for chronic uncontrollable sources of stress.[7]

Personality and health

Although there is no consistent empirical support for the older idea that different diseases are linked with specific personality types, there is evidence from different, more credible sources that personality factors can influence health and play a role in determining illness in other ways.[11]

Probably the best known work in this area concerns the link between the so-called 'type A' personality and coronary heart disease. The *type A personality* was originally characterized by competitiveness, time urgency, hostility, and related behavioural factors, which were associated with a significantly increased risk of coronary heart disease (CHD). However, it is now thought that only certain components (e.g. anger and hostility) of the original type A formulation are 'pathogenic'.[12]

Type A individuals show a greater physiological reactivity (e.g. in blood pressure and heart rate) to environmental demands and may even generate more demands by their style of behaviour. The more frequent elevations in blood pressure and higher levels of hormonal change, characteristic of this behavioural style, may eventually cause adverse physical changes to the heart and blood vessels. Also, type A individuals are more likely to engage in unhealthy behaviours since they drink more alcohol than type B individuals and, if they smoke, they inhale their cigarette smoke for a longer time.

Type A behaviour is probably the most extensively investigated personality factor in current health psychology research, and there have been interventions developed to change the behaviour pattern, with positive health outcomes.[13] More recently the concept of the *type D personality* has been described as another major psychological risk factor for CHD. Type D refers to the tendency to experience negative emotional states and to inhibit the expression of these emotions in social settings. Type D patients with CHD have been found to have a significantly higher risk of further cardiac morbidity in the short- and longer-term.[14]

More generally, patterns of positive or negative emotional responses, associated with personality, can influence various aspects of health.[11,15] Individuals who are high in negative affect (i.e. experience more negative emotions, particularly anxiety) do not seem to be more prone to disease, but they are more likely to notice bodily changes and symptoms and consequently seek medical help more frequently (see Wiebe and Smith[15] for a more detailed account of negative affect and the links between personality and health).

Another aspect of personality which has been shown to be health protective is optimism, which describes a tendency towards positive expectations in life and which enables individuals to cope better with stressors and engage in healthier lifestyles. There is emerging evidence that optimistic individuals not only cope more effectively with illness and other life crises but also show better health outcomes than those with lower levels of optimism.[16]

Lifestyle and health

The effects on health of behaviours such as smoking and high alcohol use are well documented.[1] There is overwhelming evidence that smokers not only are much more likely to die from lung cancer and other cancers but also have much higher rates of cardiovascular disease and chronic respiratory disorders, particularly emphysema and chronic bronchitis. Moreover, the disease risk is dose related in that higher levels of smoking are more strongly associated with all these diseases. With sustained high levels of alcohol use a different but equally unpleasant spectrum of health problems can be seen. Drinking is a major cause of accidents particularly motoring accidents and can cause liver damage as well as having detrimental effects on brain functioning.[1]

For health psychologists, the key questions about health-risk behaviours concern their origin, their maintenance, and their prevention or treatment. There are diverse determinants of these behaviours since they may start as ways of coping with stress, in response to peer pressure, for pleasure or for a number of other reasons. Similarly, they will be maintained by a variety of psychological, social, and biological factors.

There are many other risky behaviours that cannot be discussed in detail in an overview; these include drug abuse, poor diet, and accidents, and the health effects of all these are also well documented.[1] Although health psychology has an important role to play in describing, explaining, and intervening in all risk behaviours, these problems should not be conceptualized exclusively in individual behavioural terms since they often reflect adverse social circumstances or particular cultural contexts.[17]

The same caveats about the influence of social and cultural factors must also be applied to the understanding of health-protective or health-enhancing behaviours. Prospective cohort studies have confirmed that various daily behaviours (e.g. patterns of eating, sleeping, and exercise) can have significant long-term effects on health.[18] For example, there is now a growing body of evidence to indicate that regular exercise has a beneficial effect on both physical and psychological health.[19] Exercise can reduce the incidence of physical health problems in elderly people and facilitate recovery from heart attack. However, there can be significant problems in ensuring that exercise and other health-promoting activities are adhered to. Interventions need to be planned carefully, because it has been shown that it is usually very difficult to make and maintain changes in health-related behaviour. Information provision is rarely sufficient to promote behaviour change since it is also necessary to elicit and modify beliefs (see below) as well as influencing social networks in order to ensure success.

Beliefs and health-related behaviour

Even though health psychologists acknowledge the importance of situational, dispositional, and socio-cultural factors as determinants of health-related behaviour, most current research has a primary focus on the role of beliefs in explaining variance in health-related behaviour. The most widely used explanatory approaches have been described generically as 'social cognition models' (see Conner and Norman[2] for an excellent overview of these models). These models are based on the premise that, when a person is faced with having to make a decision about a particular health behaviour (e.g. attend for a screening test; wear a seat belt, etc.), their decision-making and behaviour can be best understood in terms of their perceptions or beliefs about the health issue and the behaviour in

question. The best known models here are the Health Belief Model, Theory of Reasoned Action/Theory of Planned Behaviour, and Protection-Motivation Theory. Broadly these models locate the strength of certain beliefs or evaluations of the health threat (e.g. 'is it serious? Is it likely to affect me?') and/or the associated health behaviour ('Is it an acceptable or worthwhile thing for me to do?') as the key determinants of an individual's motivation or intention to carry out the behaviour. More recent models incorporate other beliefs, such as self-efficacy, which reflect the individual's belief about their ability to implement or carry out the health-related behaviour.

For habitual and addictive health-related behaviours (e.g. dietary behaviour; substance abuse) there have also been attempts to develop stage-based models, such as the Precaution Adoption model and the Transtheoretical model[2] as ways of describing the stages which people may go through in evaluating the health issue through to thinking about, planning, and maintaining behaviour change. Although these stage models provide a framework for identifying the patient's state of readiness for a health behaviour change intervention as well as an immediate target for an intervention, the evidence for them is weak and there are now a number of serious critiques of their validity and applicability.[20]

Symptoms and illness behaviour

The psychology of physical symptoms

Understanding how symptoms are perceived is critical to explaining variation in illness behaviour. Psychological factors play an important role in the appraisal of symptoms. There is considerable evidence that bodily symptoms and functions are not perceived with a high degree of accuracy and individuals vary widely in what symptoms are noticed and whether medical help is sought for symptoms.[21]

The probability that individuals will attend to somatic information will depend on the competition for attention from other sources of available stimuli. When the environment is lacking in stimulation individuals tend to pay more attention to bodily symptoms. Conversely, when an individual's attention is drawn to the external environment, bodily symptoms are less likely to be noticed. This finding has wide day-to-day applications ranging from why people cough in the boring parts of movies and lectures to explaining demographic differences in symptoms reports, such as increased symptom reporting among the socially isolated and the unemployed. It also has clinical applications in chronic pain and other chronic medical conditions where patients' isolation may exacerbate the condition by increasing preoccupation with symptoms.

Cognitive schemas can also strongly influence the reporting of physical symptoms by guiding the way individuals pay attention to their body. Schemas determine the organization of incoming information and guide health directed behaviour. There is a strong tendency for individuals to search for information that is consistent with existing schemas and disregard information that does not fit. Individuals also attach more importance to symptoms consistent with a current cognitive schema than other symptoms. Schemas may develop through personal experience with the condition or by having come across the illness through family, friends, or in the media. Illness schemas can vary from vague ideas about the types of symptoms that represent an illness to more elaborate and detailed conceptions of individual illnesses. Medical students' disease, where students studying a particular illness notice they also have the symptoms of the condition, and episodes of mass psychogenic illness are more dramatic demonstrations of this phenomenon, but the process is seen on a more subtle level with response to placebos (see below). Here, following treatment, a new cognitive schema may shift attention towards symptoms that indicate recovery rather than those of the illness.

Patient delay

There is growing research to suggest that patients' interpretation of their symptoms can influence help-seeking behaviour.[22] One medical condition where delay can have serious consequences is myocardial infarction, as early arrival at hospital is strongly associated with improved chances of survival. There is a large variation in how long patients delay before seeking help, and a strong predictor of early arrival at hospital is the belief that the symptoms are a heart attack.[23] Heart attacks are generally seen as sudden and dramatic events that involve severe chest pain and collapse. In the case of myocardial infarction patients, the mismatch between these expectations and the symptoms experienced gives rise to patient delay.

Research investigating the stages of patient delay for medical conditions has generally found three main stages prior to entering treatment, with each stage influenced by a different set of factors and decisional processes. The first interval is generally referred to as appraisal delay, which is the time period from when the individual first detects symptoms to when an illness is inferred. The main influences on this period are factors related to interpretation of symptoms. The second interval is called illness delay—the period from the time the individual decides he or she is ill until the decision is made to seek medical help. The final period called utilization delay is the time until the individual enters hospital or has contact with medical personnel. This first period of appraisal delay has been generally found to cause the largest contribution to overall delay.[24]

High health service users

A large percentage of medical consultations are made for non-medical complaints. This is particularly so for primary health care services. A number of studies have found that a small percentage of individuals without significant medical illness use a disproportionately large amount of medical services and at considerable cost.[25] These individuals have been variously labelled as somatizers, hypochondriacs, the worried well, patients with medically unexplained symptoms, and multiple attenders.

Research on high health service users suggests they are higher in trait anxiety. This is consistent with research showing a strong relationship between the somatic complaints and high levels of psychological distress or neuroticism. Individuals high in anxiety tend to be more introspective, watchful for any unusual symptoms, and develop more negative interpretations of symptoms they experience.[26] Symptoms of anxiety, such as tachycardia, can also be misinterpreted as signs of a physical illness by some patients.

Some individuals also seem to have a tendency to make catastrophic interpretations about physical symptoms and this may influence frequency of presentation to medical services and recovery from illness. Catastrophizing in pain patients has been associated with disability, and drop-out from pain-management programmes.[27] It has also been associated with higher levels of fatigue and disability in chronic fatigue syndrome patients.

Catastrophizing is also seen in 'cardiac invalidism'. Here patients adopt an extremely passive, dependent, and helpless role in the belief that any form of overly vigorous activity will bring on another myocardial infarction. A hypersensitivity to bodily symptoms means that normal sensations may be misconstrued to indicate overexertion or an impending fatal myocardial infarction. This pattern often results in a cycle of inactivity and loss of physical condition, which in turn can support these beliefs when patients exert themselves. Many patients who develop highly negative illness beliefs overuse medical services mainly for reassurance about symptoms.

The issue of reassurance in medical consultations is relevant here. One of the common patient expectations in primary care consultations is to have a better understanding of current symptoms. For many patients, being told there is no serious medical problem underlying their symptoms is effective in reducing concern about their condition, but for a significant number there remains worry about their health status. Continued anxiety in this group often results in further needless consultations and investigations. Evidence suggests that patients' existing beliefs about their condition are predictive of reassurance failure and that for reassurance to be effective, patients' concerns need to be elicited and appropriate information provided to explain either the patient's symptoms or why serious pathology has been ruled out.[28]

Recent work on improving reassurance following medical testing has suggested providing information to patients about the meaning of normal test results before testing, may weaken patients' preconceived beliefs about their condition and provide a context to help understand the test result. In this study, providing patients with information about normal test results prior to testing, improved their reassurance, reduced their symptoms, and lessened their use of unnecessary medication.[29]

If patients' ideas and beliefs about their symptoms are not addressed when symptoms persist or recur it is likely that health worry will also be reactivated as the patient still lacks a satisfactory cognitive model or explanation that enables them to interpret their symptoms as benign. A practical consequence of these findings is the need for clinicians to elicit the patient's own attributions and concerns about their symptoms and to use these as the basis for dealing with misconceptions and providing the patient with a more benign explanation of their symptoms.[29]

Cognitive models of illness

Research suggests patients cluster their ideas about an illness around five coherent themes or components, which health psychologists have called illness perceptions.[30] These provide a framework for patients to make sense of their symptoms, assess health risk, and direct action in the recovery phase. The major cognitive components are as follows.

- Identity: the label of the illness and the symptoms the patient views as being part of the disease.
- Cause: personal ideas about aetiology, which may include simple single causes or more complex multiple causal models.
- Time-line: the patient's belief about the likely time course of the illness (e.g. acute, chronic, or episodic).
- Consequences: expected impact of the illness on the patient's life.
- Cure/control: the patient's beliefs about the extent to which the illness is amenable to cure or control either through personal actions or by treatment.

These components show logical interrelationships. For example, a strong belief that the illness can be cured or controlled is typically associated with short perceived illness duration and relatively minor consequences.

The theoretical framework for this research is derived from the self-regulatory model developed by Leventhal *et al.*[30] This model views illness perceptions as critical in guiding the patient's coping efforts to deal with symptoms, illness, and threats to health. It consists of four components: the cognitive representation of the illness, the emotional response to the illness and treatment, the coping directed by the illness representation, and the individual's appraisal of the coping outcome.

Patient cognitive models of their illness are, by their nature, private. Patients' are often reluctant to discuss their beliefs about their illness in medical consultations because they fear being seen as ignorant or misinformed. Until recently, assessment of illness perceptions has been by open-ended interviews designed to encourage patients to elaborate their own ideas on the illness. However, questionnaires have been developed to measure illness perceptions in a variety of illnesses[31,32] as well as specific beliefs about medication.[33]

The illness perception approach has recently been applied to a large number of health conditions (see Hagger and Orbell[34] for a meta-analysis). Current research in this area is building on these findings to develop cognitive–behavioural interventions designed to modify dysfunctional illness perceptions and provide better recovery. A good example of this is a study showing that the early elicitation and modification of dysfunctional illness beliefs can improve recovery and return to function in patients with a recent myocardial infarction.[35]

Health care behaviour

In this section we examine the role of psychological processes in the delivery of health care by focusing on two broad areas: doctor–patient communication and health care in hospital.

Doctor–patient communication

There is now considerable evidence not only of patient dissatisfaction with medical communication but also of widespread non-compliance with subsequent treatment recommendations. Early research revealed that patient dissatisfaction was often associated with receiving insufficient information, poor understanding of the medical advice, and subsequent reluctance or inability to follow recommended treatment or advice. Another source of patients' dissatisfaction is the perception that the doctor lacks interest and empathy, and is unwilling to involve them in decision-making during the consultation. Thus, an overview of research in this area[36] revealed that patient satisfaction was higher following consultations in which the doctor engaged in more social conversation, positive verbal and non-verbal behaviour, and partnership building.

A range of frameworks have been developed for describing the process of the consultation. Similarly various methods have been devised for analysing the interactional processes which occur during the consultation[36] and Roter *et al.*[37] have used these

analyses to propose five distinct patterns of communication in doctors:

- narrowly biomedical, characterized by closed-ended medical questions and biomedical talk
- expanded biomedical, similar to the narrowly biomedical but with moderate levels of psychosocial discussion
- biopsychosocial, reflecting a balance of psychosocial and biomedical topics
- psychosocial, characterized by psychosocial exchange
- consumerist, characterized by patient questions and information giving by the doctor.

The highest levels of patient satisfaction were found with those who had seen doctors using the psychosocial communication pattern, whereas the lowest satisfaction scores were recorded in those who had experienced either of the two biomedical patterns.

An alternative and broader distinction has been made between consultations which are described as patient centred and those which are doctor centred, reflecting the extent to which the doctor or patient determines what is discussed. Doctor-centred consultations are ones in which closed questions are used more often and the direction is determined by the doctor, typically with a primary focus on medical problems. In contrast, patient-centred encounters involve more open-ended questions with greater scope for patients to raise their own concerns and agendas.

Patient satisfaction and understanding of their illness following the medical consultation can play a major role in influencing adherence with treatment or advice as well as other outcomes including health and well-being. A number of studies have demonstrated beneficial effects on patients' health and well-being arising from positive experiences in medical consultations.[38] These have focused on psychological states such as anxiety as well as changes in specific physical variables such as blood pressure and blood glucose control. Some of the most impressive findings here have been found in the patient-intervention studies, which are described below.

One important spin-off from the findings in this area has been the development of communication skills training packages for medical undergraduates and for experienced clinicians, particularly for improving skills in difficult areas of communication such as giving 'bad news'. There have also been a number of specific interventions aimed at patients. Generally, these have involved interventions for patients prior to a consultation in order to increase their level of participation, particularly to ensure that their own concerns are dealt with and that information provided by the doctor is clearly understood.

Health care in hospital

Patients experience many stressors in hospital and these arise from a range of factors including enforced lifestyle changes and the demands involved in developing good relations with hospital staff.[39] Other hospital stressors include worries about aspects of communication with staff, as well as concerns about investigations and treatment. Even such factors as the layout and colour of the ward, and the view from the patient's bed have been found to affect recovery. Not surprisingly, studies that have compared home-treated and hospitalized patients with the same condition have shown less psychosocial distress in those remaining at home.

In addition to these general psychological impacts of hospitalization, there may be specific problems or demands which occur either as a result of the particular health problem or the type of treatment which the patient has to undergo. An example of the way in which patients' health problems may influence their experience of hospital care can be seen in some of the studies of patients with HIV/AIDS who may experience negative or blaming attitudes from staff or other patients. For example, with AIDS patients being treated in either special care units or integrated in more general hospital settings, the latter group reported higher levels of stress associated with feelings of abandonment, and impersonal or discriminatory treatment. Where staff perceive patients as instrumental in having brought about their own condition through their own behaviour or neglect, they may be less committed, motivated, and sympathetic towards them.

A number of studies have been made of the psychological effects of specific treatment settings such as intensive care units (ICU) and haemodialysis units. Studies of patients in intensive care reveal high levels of psychological distress both during and for some time after their stay.[40] A range of factors seem to be involved, including being intubated and not being able to communicate. Even physical aspects of the ICU can have significant effects. Thus comparisons of patients in intensive care units with and without windows found that those in the windowless units were less well oriented during their stay and had a less accurate recall of their length of stay afterwards. In addition to these general problems associated with the intensive care units, other studies have assessed the degree of stress experienced by staff and visitors. For patients' relatives there is evidence that they find the time spent by the patient on life support in the intensive care unit particularly worrying. During this time they experience considerable fear and uncertainty but this can be improved by seeking information and the use of other resources.

In contrast with the acute psychological restrictions and demands of intensive care, some patients such as those on renal dialysis are subject to much more chronic restrictions. Dialysis can have major effects on an individual's psychological and social functioning, particularly giving rise to vocational impairment, reduced sexual activity, and mood changes.[41] In addition to the physical limitations and demands of dialysis, patients are also faced with the need to adhere to strict recommendations regarding diet and fluid consumption, as well as complex medication regimens. A number of aspects of dialysis can give rise to psychological distress, including the constant threat of death, dependence on the dialysis machine and medical staff. The stringent dietary and liquid restrictions are also important factors in patients' feelings of helplessness and lack of control. The ways in which patients cope can have important influences on their well-being and outcome. For example, problem-focused types of coping have been shown to be associated with better adherence to fluid intake restrictions, when these coping strategies were used in response to stressors arising from a relatively controllable aspect of dialysis. For those stressors, which patients perceived as less controllable, emotion-focused coping strategies provide better levels of adherence.

Many medical procedures in hospital can give rise to considerable discomfort and anxiety. These include certain treatments such as surgery, and specific investigative procedures such as barium radiography, endoscopy, and cardiac catheterization, which may not only be uncomfortable and sometimes physically distressing but which also carry the threat of uncovering a serious medical

condition.[42] Consequently a number of psychological interventions have been developed to prepare patients for surgery or other stressful procedures in the hospital setting. In broad terms they can help by providing the patient with information to reduce the uncertainty of the event, or with specific behavioural or cognitive skills to help with some of the discomfort or pain.[43]

These interventions have been found to improve a range of post-surgical outcomes, including anxiety, pain and use of pain medication, length of stay in hospital, and various indicators of recovery. All the interventions have been found to be successful in improving at least one aspect of outcome, and the majority of them have a positive impact on many of the outcomes. A meta-analysis by Johnston and Vogele[43] revealed that the largest recovery effects were obtained for pain, negative affect, and physiological indices of recovery but there was considerable variation in the magnitude of these effects. Smaller but more consistent advantages of psychological preparation were found on pain medication and length of hospital stay. The interventions, which had the most widespread overall effects on all the outcomes, were found to be procedural information provision and behavioural instructions. In addition to these specific psychological preparations, there is now evidence that the pre- and post-surgical social setting can have a significant effect on recovery. Studies have also revealed clear beneficial effects of sharing a room with someone who was recovering from surgery. Patients who had post-surgical room-mates, who had undergone the same type of surgery, have been shown to be less anxious prior to surgery, engaged in more post-surgical physical activity, and were discharged home sooner.

Treatment behaviour

Patients respond to their treatment in a range of ways and these can have very significant effects on clinical outcomes. Two major areas of patient behavioural variation are seen in the extent to which patients adhere to their prescribed treatment and in the non-specific or placebo effects of the treatment on clinical outcome. An overview of research on these two areas is now presented.

Adherence

The extent to which the patient adheres to the advice or treatment offered in health care consultations has been widely studied. Most medical consultations result in the prescription of treatment or advice, and the use of medicines is a key aspect to the self-management of most chronic illnesses. However, many patients fail to do this and low rates of adherence to recommended treatment are seen as problematic in chronic physical and psychiatric illnesses.[44]

The incidence of reported medication non-adherence varies greatly from 4 to 92 per cent across studies, converging at 30 to 50 per cent in chronic illness. In primary prevention studies, it has been found that many participants drop out of lifestyle change programmes, designed to improve diet or reduce health-risk behaviours. Even patients who have experienced major health problems, such as heart attacks, may show low levels of uptake of rehabilitation programmes as well as considerable variation in the adoption of recommended lifestyle change. In the area of mental health, there is also evidence of significant rates of non-adherence to various recommendations from health care providers.

Non-adherence behaviours may be categorized as either intentional or unintentional. Intentional non-adherence arises when the patient makes a strategic decision not to take the treatment as instructed. An example of this type of behaviour has been found among hypertensive patients who believed that they could judge when their blood pressure was high by the presence of symptoms such as stress or headache and thus took antihypertensive medication only when these symptoms were experienced. From a self-regulatory perspective, the level of treatment adherence may be indicative of a strategic coping response, which is entirely consistent with the patient's view of their problem. Thus, patients who believe that their problem will not last for long have been found to be less likely than those with a more chronic time-line representation to adhere to their medication over a long period of time.

Non-adherence may be unintentional when the patient's intentions to follow treatment recommendations are thwarted by barriers such as forgetting, and inability to follow treatment instructions because of a lack of understanding or physical problems such as poor eyesight or impaired manual dexterity. Thus, if the quality of communication is poor and patients receive information, which is difficult to understand or recall, as has been outlined above, then this makes it less likely that treatment will be adhered to.

The determinants of non-adherence

One very obvious explanation for non-adherence arises from poor understanding and recall of information presented in the medical consultation. Many patients lack basic knowledge about their medication but there is no simple relationship between this and their adherence. Reviews of adherence research fail to demonstrate a consistent positive association between knowledge and adherence. Moreover, interventions that enhance knowledge do not necessarily improve adherence. Patient satisfaction can act as a mediator between information provision, recall, and adherence since patient surveys reveal that many patients wanted more information than they were given. Dissatisfaction with attributes of the practitioner or the amount of information and explanation provided may act as a barrier to adherence by making the patient less motivated towards treatment.

The emphasis of adherence research over the last decade or so has moved away from attempts to identify stable trait factors which characterize the non-adherent patient to achieving a greater understanding of how and why patients decide to take some treatments and not others. Much of this research is informed by psychological theories, which conceptualize behaviour as the product of cognition which occurs within a social framework.

The application of the social cognition models, described earlier in this chapter, indicates that medication non-adherence may arise from a rational decision on the part of the patient and identifies some of the cognitions which are salient to these decisions. The types of beliefs and attitudes specified by such theories as the Health Belief Model, the Theory of Planned Behaviour and the Self-Regulatory Model (SRM) have all been used to explain aspects of treatment adherence. The SRM also acknowledges the importance of symptom perception in influencing illness representations and adherence as a coping behaviour. Confirmatory evidence for this is provided by findings from studies of patients with hypertension and with diabetes, both of whom commonly use perceived symptoms to indicate their blood pressure and glucose levels respectively, and to guide self-treatment. However, patients' beliefs about their symptoms and estimations of their own blood pressure

and glucose levels are often erroneous, and this can result in poor control of symptoms and illness.

More recent research has begun to focus on the role of people's beliefs about medicines and the ways in which these could influence adherence.[45] This research has revealed two broad factors describing people's beliefs about their prescribed medicines: their perceived necessity for maintaining health (specific-necessity) and concerns based on beliefs about the potential for dependence or harmful long-term effects and that medication taking is disruptive (specific-concerns). Two factors were also found to describe people's beliefs about medicines in general. The first relates to the intrinsic properties of medicines and the extent to which they are harmful addictive substances (general-harm) and the second factor comprises concerns that medicines are overused by doctors (general-overuse).

People's views about the specific medication regimen prescribed for them were found to be much more strongly related to adherence reports than are more general views about medicines as a whole. Moreover, interplay was found between concerns and necessity beliefs, which suggests that people engage in a risk–benefit analysis and consequently attempt to moderate the perceived potential for harm by taking less. Patients with stronger concerns based on beliefs about the potential for long-term effects and dependence reported lower adherence rates, whilst those with stronger beliefs in the necessity of their medication reported greater adherence to medication regimen.[45] This work points to the importance of accessing patients' beliefs as a prerequisite of any intervention designed to increase medication adherence. In particular, it would seem important to identify specific concerns about treatment and to allay these in ways which make sense to the patient.

The placebo response

The term 'placebo' is used to describe a treatment that gains a response due to its therapeutic intent rather than the specific ingredients of the treatment itself. Placebo responses have been shown for a wide variety of medical treatments including surgery, psychotherapy, medication, therapeutic ultrasound, injections, and aerosol sprays. Placebos have also been demonstrated to have effects in countless medical conditions and also on a number of physiological functions such as blood pressure, heart rate, gastric motility, lung function, and postoperative swelling. Adverse effects from placebos or so-called 'nocebo' effects have also been noted in the literature.

Characteristics of the treatment itself and the setting it is administered in can have a strong influence on the magnitude of the placebo response. In general, treatments that involve more serious rituals and sophisticated equipment such as surgery have stronger placebo effects. Likewise, other treatments imparting a powerful impression to the patient such as foul- or strong-tasting medicine, injections, and precise instructions also enhance the placebo response. The colour of medication has been shown to have some effect depending on the condition, and known brand names seem to have an edge in placebo response over unknown drug companies.

A similar theme runs through the clinician characteristics that increase the placebo response. Clinicians and clinics seen as having high status and having high levels of credibility have an improved placebo response. At the same time, the doctor–patient interaction is also important. If the doctor shows high levels of concern and empathy for the patient then the response increased. High confidence shown by the doctor in the treatment administered to the patient along with a clear indication of the expected response of the treatment is also likely to improve the likelihood of placebo response.

In contrast, isolating characteristics of patients who are placebo responders has yielded inconclusive results. Much of the evidence for the role of demographic, intellectual, or personality characteristics of patients likely to respond to placebos is mixed and inconsistent. Studies have found individuals who responded to placebos in one setting to be unresponsive in another. Likewise, conditioning studies have shown individuals who have been unresponsive can later respond. These findings point to the fact that individual characteristics probably play a less significant role than situational factors and the doctor–patient interaction in influencing the placebo response.

Treatment response has been divided into specific and non-specific components, with the non-specific component encompassing factors such as clinician attention, expectation, reputation, treatment setting, etc. Determining the magnitude of the specific and non-specific components of medical treatment is a difficult and probably impossible task. In an attempt to determine how powerful non-specific effects are under ideal circumstances, Roberts *et al.*[46] chose to look at the effect of a number of medical treatments later shown to be ineffective but where clinician faith in the treatment was initially positive. Pooling the data from diverse treatments such as gastric freezing for duodenal ulcer and glomectomy for asthma, this study found 40 per cent of patients had an excellent response to the treatments, 30 per cent good, and 30 per cent poor. This suggests that under ideal circumstances where clinician and patient expectations are high and the treatment is administered in a credible way, non-specific factors can by themselves exert a powerful effect.

The role of compliance with placebos also appears to be important. In a review of five placebo-controlled studies measuring both compliance with medication and outcome, Epstein[47] found that subjects who were more compliant did better on outcome measures, regardless of whether they were on placebo or active treatment. Outcomes included prevention of relapse in schizophrenia, reduction of fever or infection in cancer patients, alcohol abstinence, reduction of weight, and prevention of mortality in patients with heart disease. In a later study, Horwitz *et al.*[48] also found the risk of death was substantially less in patients who took more than 75 per cent of their medication regardless of whether the medication was placebo or β-blocker. This suggests the act of compliance may have some effects of other health-promoting behaviours or cause cognitive or emotional changes that may influence health in the long-term.

There have been a number of theories proposed to explain the placebo response but no one theory yet provides an adequate integrated theoretical framework. The reduction of anxiety following treatment and consequent effect on symptoms has been proposed as one mechanism by which the placebo effect may operate, but changes in anxiety states have not reliably been associated with placebo responses. The role of the medical situation and its accoutrements being associated through classical conditioning with symptom relief is likely to play some role. There is, however, little direct research on this proposed mechanism, although

classical conditioning of drug responses has been shown in certain situations.

Two other theories proposed have been the role of cognitive dissonance and patient expectations. The cognitive dissonance argument proposes that the placebo effect may be due to the pressure on individuals to show consistency in their views and actions. Therefore, for some individuals, having treatment is inconsistent with not showing any change in symptoms and this may encourage the person to reduce this inconsistency. The role of patient expectations and placebo effects is an area that has not received a great deal of systematic research. It is suggested that patient expectations may cause changes in cognitive schemas that influence the types and nature of symptoms that patients pay attention to following treatment. New developments in research on illness perception and beliefs about treatment, outlined above, as well as in the fields of neurobiology and neuroimaging[49] hold considerable promise for increasing our understanding of the nature of the placebo effect and its determinants.

Conclusions

This selective overview of health psychology has demonstrated the range of psychological processes in health, illness, and health care. At the present time it is primarily a disciplinary area of psychology with an emphasis on research into health and illness behaviour. However, many interventions have been developed for healthy individuals, patients, and health care staff. This practitioner aspect of health psychology is now being accompanied by specific professional developments, and formal postgraduate training in health psychology is now available in many countries.[50]

Health psychology has established itself rapidly but it is still very much an emerging discipline. Greater insights are needed into the ways in which psychological processes can influence health and illness, and more comprehensive models are required for explaining all aspects of health and illness behaviour. In the long term this will result in the increasing use of psychological interventions for preventing and managing health problems and for the effective delivery of health care.

Further information

Weinman, J., Johnston, M., and Molloy, G. (eds.) (2007). *Health psychology, 4 volume set*. Sage, London.

http://www.health-psych.org/

Ayers, S., Baum, A. McManus, I.C. (eds.), *et al.* (2007). *Cambridge handbook of psychology, health and medicine* (2nd edn). Cambridge University Press, Cambridge.

Johnston, M., Weinman, J., and Wright, S. (1995). *Health psychology: an assessment portfolio*. NFER Nelson, Windsor.

References

1. Mokdad, A.H., Marks, J.S., Stroup, D.F., *et al.* (2004). Actual causes of death in the United States, 2000. *JAMA*, **291**, 1238–45.
2. Conner, M. and Norman, P. (eds.) (2005). *Predicting health behaviour*. Open University Press, Buckingham.
3. Petrie, K.J. and Weinman, J. (eds.) (1997). *Perceptions of health and illness: current research and applications*. Harwood Academic, London.
4. Ayers, S., Baum, A., Newman, S. (eds.), *et al.* (2007). *Cambridge handbook of psychology, health and medicine* (2nd edn). Cambridge University Press, Cambridge.
5. Ogden, J. (2007). *Health psychology: a textbook* (4th edn). Open University Press, Buckingham.
6. McEwen, B.S. (1998). Protective and damaging effects of stress mediators. *New England Journal of Medicine*, **338**(3), 171–9.
7. Segerstrom, S.C. and Miller, G.E. (2004). Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. *Psychological Bulletin*, **130**(4), 601–30.
8. Cohen, S., Tyrell, D.A., and Smith, A.P. (1993). Psychological stress and susceptibility to the common cold. *New England Journal of Medicine*, 325, 606–12.
9. Kiecolt-Glaser, J.K., Marucha, P.T., Malarkey, W.B., *et al.* (1995). Slowing of wound healing by psychological stress. *Lancet*, **346**, 1194–6.
10. Broadbent, E., Petrie, K.J., Booth, R., *et al.* (2003). Psychological stress impairs early wound repair following surgery. *Psychosomatic Medicine*, **65**, 865–9.
11. Friedman, H.S. and Booth-Kewley, S. (1987). The "disease-prone personality": a meta-analytic view of the construct. *The American Psychologist*, **42**(6), 539–55.
12. Hecker, M.H.L., Chesney, M.A., Black, G.W., *et al.* (1988). Coronary-prone behaviors in the western collaborative group study. *Psychosomatic Medicine*, **50**, 153–64.
13. Thoresen, C., Friedman, M., Powell, L.H., *et al.* (1985). Altering the Type A behavior pattern in postinfarction patients. *Journal of Cardiopulmonary Rehabilitation*, **5**, 258–66.
14. Denollet, J., Pedersen, S.S., Vrints, C.J., *et al.* (2006). Usefulness of type D personality in predicting five-year cardiac events above and beyond concurrent symptoms of stress in patients with coronary heart disease. *The American Journal of Cardiology*, **97**(7), 970–3.
15. Wiebe, D. and Smith, T.W. (1997). Personality and health. In *Handbook of personality psychology* (eds. R. Hogan and J.A. Johnson), pp. 891–918. Academic Press, San Diego, CA.
16. Carver, C.S., Smith, R.G., Antoni, M.H., *et al.* (2005). Optimistic personality and psychosocial well-being during treatment predict psychosocial well-being among long-term survivors of breast cancer. *Health Psychology*, **24**(5), 508–16.
17. Adler, N.E., Boyce, T., Chesney, M.A., *et al.* (1994). Socioeconomic status and health: the challenge of the gradient. *The American Psychologist*, **49**(1), 15–24.
18. Schoenborn, C.A. (1993). The Alameda Study—25 years later. In *International review of health psychology*, Vol.**2** (eds. S. Maes, H. Leventhal, and M. Johnston), pp. 81–116. John Wiley & Sons Ltd., Chichester.
19. Warburton, D.E.R., Nicol, C., and Bredin, S.D.S. (2006). Health benefits of exercise: the evidence. *CMAJ: Canadian Medical Association Journal*, **174**, 801–9.
20. West, R.J. (2005). Time for a change: putting the transtheoretical (stages of change) model to rest. *Addiction*, **100**(8), 1036–9.
21. Pennebaker, J.W. (1982). *The psychology of physical symptoms*. Springer-Verlag, New York.
22. Cameron, L., Leventhal, E.A., and Leventhal, H. (1995). Seeking medical care in response to symptoms and life stress. *Psychosomatic Medicine*, **57**, 37–47.
23. Horne, R., James, D., Petrie, K.J., *et al.* (2000). Patients' interpretation of symptoms as a cause of delay in reaching hospital during acute myocardial infarction. *Heart*, **83**, 388–93.
24. Andersen, B.L. and Cacioppo, J.T. (1995). Delay in seeking a cancer diagnosis: delay stages and psychophysiological comparison processes. Special issue: social psychology and health. *The British Journal of Social Psychology*, **34**, 33–52.
25. Verhaak, P.F.M., Meijer, S.A., Visser, A.P., *et al.* (2006). Persistent presentation of medically unexplained symptoms in general practice. *Family Practice*, **23**, 414–20.
26. Sensky, T., Macleod, A.K., and Rigby, A.F. (1996). Causal attributions about common somatic sensations among frequent general practice attenders. *Psychological Medicine*, **26**, 641–6.

27. Sullivan, M.J.L., Thorn, B., Haythornthwaite, J.A., *et al.* (2001). Theoretical perspectives on the relation between catastrophising and pain. *The Clinical Journal of Pain*, **17**, 52–64.
28. Donkin L, Ellis, C.J., Powell, R., *et al.* (2006). Illness perceptions predict reassurance following negative exercise testing result. *Psychology and Health*, **21**, 421–30.
29. Petrie, K.J., Müller, J.T., Schirmbeck, F., *et al.* (2007). Effect of providing information about normal test results on patients' reassurance: randomized controlled trial. *British Medical Journal*, **334**, 352–5.
30. Leventhal, H., Nerenz, D.R., and Steele, D.J. (1984). Illness representations and coping with health threats. In *A handbook of psychology and health*, Vol. 4 (eds. A. Baum and J. Singer), pp. 219–52. Erlbaum, Hillsdale, NJ.
31. Weinman, J., Petrie, K.J., Moss-Morris, R., *et al.* (1996). The illness perception questionnaire: a new method for assessing illness perceptions. *Psychology and Health*, **11**, 431–46.
32. Broadbent, E., Petrie, K.J., Main, J., *et al.* (2006). The brief illness perception questionnaire (BIPQ). *Journal of Psychosomatic Research*, **60**, 631–7.
33. Horne, R., Weinman, J., and Hankins, M. (1999). The beliefs about medicines questionnaire: a new method for assessing cognitive representations of medication. *Psychology and Health*, **14**, 1–24.
34. Hagger, M.S. and Orbell, S. (2003). A meta-analytic review of the common-sense model of illness representations. *Psychology and Health*, **18**, 141–84.
35. Petrie, K.J., Cameron, L.D., Ellis, C.J., *et al.* (2002). Changing illness perceptions after myocardial infarction: an early intervention randomized controlled trial. *Psychosomatic Medicine*, **64**, 580–6.
36. Roter, D. and Hall, J.A. (1989). Studies of doctor patient interaction. *Annual Review of Public Health*, **10**, 163–80.
37. Roter, D., Stewart, M., Putnam, S.M., *et al.* (1997). The patient-physician relationship: communication patterns of primary care physicians. *The Journal of the American Medical Association*, **277**, 350–6.
38. Stewart, M.A. (1995). Effective physician–patient communication and health outcomes: a review. *Canadian Medical Association Journal*, **152**, 1423–33.
39. Koenig, H.G., George, L.K., Stangl, D., *et al.* (1995). Hospital stressors experienced by elderly medical in-patients: developing a hospital stress index. *International Journal of Psychiatry in Medicine*, **25**, 103–22.
40. Andrews, P., Azoulay, E., Antonelli, M., *et al.* (2005). Year in review in intensive care medicine, 2004. III. Outcome, ICU organization, scoring, quality of life, ethics, psychological problems and communication in the ICU, immunity and hemodynamics during sepsis, pediatric and neonatal critical care, experimental studies. *Intensive Care Medicine*, **31**, 356–72.
41. Kimmel, P. (2000). Psychosocial factors in end-stage renal disease patients treated with hemodialysis: correlates and outcomes. *American Journal of Kidney Diseases*, **35**, S132–40.
42. Weinman, J. and Johnston, M. (1988). Stressful medical procedures: an analysis of the effects of psychological interventions and of the stressfulness of the procedure. In *Topics in health psychology* (eds. S. Maes, P. Defares, I.G. Sarason, and C.D. Speilberger), pp. 205–17. Wiley, Chichester.
43. Johnston, M. and Vogele, C. (1993). Benefits of psychological preparation for surgery: a meta-analysis. *Annals of Behavioral Medicine*, **15**, 245–56.
44. Osterberg, L. and Blaschke, T. (2005). Adherence to medication. *The New England Journal of Medicine*, **353**, 487–97.
45. Horne, R. and Weinman, J. (1999). Patients' beliefs about prescribed medicines and their role in adherence to treatment in chronic physical illness. *Journal of Psychosomatic Research*, **47**, 555–67.
46. Roberts, A.H., Kewman, D.G., Mercier, L., *et al.* (1993). The power of non-specific effects in healing: implications of psychosocial and biological treatments. *Clinical Psychology Review*, **13**, 375–91.
47. Epstein, L.H. (1984). The direct effects of compliance on health outcome. *Health Psychology*, **3**, 385–93.
48. Horwitz, R.I., Viscoli, C.M., Berkman, L., *et al.* (1990). Treatment adherence and risk of death after a myocardial infarction. *Lancet*, **336**, 542–5.
49. Benedetti, F., Mayberg, H.S., Wager, T.D., *et al.* (2005). Neurobiological mechanisms of the placebo effect. *The Journal of Neuroscience*, **24**(45), 10390–402.
50. Jansen, M. and Weinman, J. (eds.) (1991). *The international development of health psychology*. Harwood Academic, London.

5.7

The organization of psychiatric services for general hospital departments

Frits J. Huyse, Roger G. Kathol, Wolfgang Söllner, and Lawson Wulsin

Introduction

The organization of psychiatric services for general hospital departments might change in far-reaching ways in the coming decades. Whereas the focus was primarily on reactive services for inpatients on medical and surgical wards, the future should focus on more proactive integrated service delivery for the complex medically ill. The essential difference from other psychiatric services is that the population served is taken care of by medical specialists in the general health setting. Consequently services are delivered in the context of the medical-psychiatric interface. Consult requests are always formulated in this perspective: the patient is treated for a medical illness or physical complaints and there are signs of an interfering psychiatric disorder.[1] Nowadays these patients are referred to as the 'complex medically ill'.[1] Therefore triage and treatment integrated in the medical context is the area of expertise of consultation-liaison (CL) psychiatrists.

The development of this area of psychiatry has been hampered by dysfunctional splits in health care, such as between general and mental health care, both on the level of its organization as well as its reimbursement.[2, 3] Recent reports, such as the report of the joint working group of the United Kingdom Royal College of Physicians and the Royal College of Psychiatrists, which describe the psychological needs of the medically and surgically ill, provide guidance to counteract these dysfunctional splits.[4] As the delivery of care-trajectories for comorbid patients becomes more and more an issue on the health care agenda, CL psychiatrists should seize this opportunity and become advocates for integrated service delivery for the complex medically ill.

[1] Whereas in mental health comorbidity refers to making more than one criteria based psychiatric diagnosis, in the CL literature the term 'comorbidity' is generally used to describe the combination of physical diseases and psychiatric disorders.

Current levels of service delivery

Around 1990 the extent of inpatient CL psychiatric service delivery was evaluated, based on the records of a representative national sample of hospitals (United States)[5] and based on a prospective multicentred study (Europe).[6] Both studies reported an average consult rate of 1 per cent, ranging up to 4–5 per cent in some university settings. This rate is much lower than the prevalence of psychiatric disorders in medical populations.[6] Taking this underutilization into account, the most striking finding was still the large variation in departments served and types of patients seen. The European Consultation-Liaison Workgroup's (ECLW) Collaborative Study made clear that CL psychiatric service delivery is primarily an emergency service. Most referrals were late, as reflected by an average time of 11 days after admission before patients were referred. In addition one-third was emergency referrals: 'See the patient the same day'.[6, 7] Exceptions were the German psychosomatic services driven by their primary interest in patients with unexplained physical complaints and problems of coping with somatic illness using a more integrated liaison approach. These services showed higher consultation rates (between 2 and 4 per cent), provided more follow-up visits, and communication with aftercare providers.[6, 7–9]

It is now evident that mental disorders and physical diseases cluster in vulnerable patients. The prevalence of mental disorders in the general hospital population is on average twice as high compared to that of the general population. However, when focusing on specific populations such as cardiac, diabetes, or transplantation, the rates of major depression may reach up to 30 per cent[10, 11] (see other chapters of this section). Patients in the general hospital setting are primarily treated for their physical diseases. However, the multiple interactions between the comorbid medical and psychiatric disorders make them complex. This justifies an integrated approach and requires individualized multimodal and multidisciplinary care.[12,13] These complex patients are the target population for CL psychiatrists. They are in need of integrated services.

Table 5.7.1 Types of service delivery

1 Emergency services
• Attempted suicide
• Acute behavioural disturbances and their prevention
  • Deliria
  • Withdrawal

2 Regular consults for patients with possible interfering psychiatric complications, such as anorexia, factitious disorder, anxiety- or depressive disorders, adjustment disorders, somatization and organic mental disorders.
3 Integrated services • Participation in multidisciplinary clinics, such as pain, memory, or transplant • Participation in multi-disciplinary rounds on 'liaison-wards' or of disease management programmes, such as for patients with Parkinson disease, diabetes, cancer, or chronic heart failure • Screening for depression or complexity in at risk populations, including the development of related care trajectories • Clinical services for highly complex patients with both medical and psychiatric acuity, such as the medical psychiatric unit

Types of service delivery

Here several models of service delivery are described (Table 5.7.1). The models have an increasing level of sophistication determined by their level of integration and the related procedural collaborative activities. Service delivery requires by definition, negotiations with health plans for their reimbursement. This is especially true for the integrated models of service delivery.[7,13,14]

Consultations

Consultations are the classical mechanism for doctors to involve other medical specialists in the treatment of patients with additional medical problems. Patients are referred if the treating physician recognizes psychiatric comorbidity or a psychological problem and if he or she thinks that psychiatric evaluation and/or intervention may be helpful. The problem linked with this type of service delivery is that physicians often do not recognize psychiatric disturbance in medical patients.[15] In some cases, this problem is avoided by organizing a 'contract type' of consultation where every patient with a defined clinical problem is referred, for instance patients with attempted suicide.

Liaison[2]

Whereas in the consultation function psychiatrists wait for the referral, the liaison function is proactive. A preventive approach is implemented through weekly multidisciplinary rounds. In orderto establish such a role the consultant and a departmental head formulate a liaison arrangement for the provision of psychiatric services for a certain population, clinic or ward. An important additional aspect of the liaison model is its educational focus. Though every consult offers an educational opportunity, in the liaison function the consultant is better equipped to enhance the skills of the teams through weekly attendance of clinical rounds. Currently, the liaison model is restricted to tertiary care hospitals with more extensive CL psychiatric services. In the European collaborative study only 5 per cent of the consults came from a liaison arrangement.[6,8]

Psychiatric-medical, medical-psychiatric units, or psychosomatic units

The 'Psych-Med unit' is an integrated clinical service for high complex patients with unstable medical disease, such as diabetes, and psychiatric disorders. Due to their mutual interactions such patients require not sequential but integrated assessment and treatment, including both intensive medical and psychiatric nursing. Depending on the required acuity levels of physical and psychiatric nursing, different types of Psych-Med units can be described.[14] Dual-trained or combined staffs are selected to provide these levels of integrated care. Organizational prototypes of this function are the US-initiated psychiatric-medical clinics and the Germfan psychosomatic wards, which focus on adjustment disorders in medical patients and complex somatization.[14] During the last years, the efforts to improve integration between inpatient and outpatient care for complex patients led to the implementation of multidisciplinary, and specialized integrated treatment programmes for specific patient groups in day hospitals (e.g. for chronic pain patients and geriatric patients).

Screening

As the selection criteria for patients in the liaison function are not operationalized, referrals are intuitively generated on the basis of clinical expertise.[15] Nowadays instruments are available to support both clinical work and research. The liaison function can be seen as a precursor of screening. It will gradually merge into more structured preventive functions, defined by the needs of a target population and guided by screening.[16] Currently, two lines of screening are in its development. First of all there is a model with a primary focus on psychopathology and primarily depression, using the patient health questionnaire (PHQ) as an indicator of psychiatric comorbidity. It is used in (elderly) patients with physical disorders, such diabetes, or physical complaints of unknown origin.[11,17] Until now this is mainly used in an outpatient setting. The other approach taken is screening for 'complexity'. A European group has taken the approach to operationalize complexity and to develop a screener and an assessment tool to detect and analyse the complex medically ill.[12,16] The Complexity Prediction Instrument (COMPRI)—the screener—is to be applied at admission on an internal medicine ward to detect patients at risk for negative outcomes of care. At the same time a comparable instrument has been developed for the elderly population to detect patients who are frail and are or have an increased risk of becoming complex.[18] Other indicators of complexity such as administrative and clinical are discussed elsewhere.[12] The INTERMED-method has been developed for complexity assessment and the design of related integrated interventions.[19] It starts with a structured interview evaluating 16 health risk variables (Table 5.7.2). The fitting of these 16 risks with 4 prognostic variables in a biopsychosocial schema and the uniform-scoring system providing different levels of action visualized in different colours, supports decision-making and facilitates interdisciplinary communication. The integrated multidisciplinary interventions designed might require case-management.

[2] Here the term 'liaison' is only used to describe the specific 'liaison function' in addition to the basic consultation function.

Table 5.7.2 Health risks evaluated for complexity assessment with INTERMED-method

Chronicity	Is patient known with physical illness/disease
Diagnostic dilemma	Were physical symptoms clarified
Severity of symptoms	Physical functioning
Diagnostic challenge	Complexity of current medical problem
Restrictions in coping	Interferences of coping with medical problems
Psychiatric dysfunctioning	Psychiatric history
Resistance to treatment	Capacity to collaborate with treatment
Psychiatric symptoms	Severity of symptoms
Restrictions in integration	Social integration reflected by work and leisure
Social dysfunctioning	Quality of relations
Residential instability	Stability of housing
Restrictions of network	Availability of help
Intensity of treatment	Utilization
Treatment experience	Trust in health professionals
Organization of care	Participating health professionals
Appropriateness of referral	Capacity to deliver appropriate care

Until now this is an area of health care in which mental health care is not formally integrated. It is to be expected that screening for psychopathology or complexity in the chronically ill will be become important future tools to initiate integrated care and allow CL psychiatric teams to actively contribute to the care of complex patients.

The organization of a consultation-liaison psychiatric service

It is unrealistic to assume that in the future the needs of general hospital patients with psychiatric morbidity can be met simply by increasing staff. To see all patients with psychiatric comorbidity would require many times the present staffing levels. Consequently, CL psychiatrists should plan their services carefully together with their medical colleagues. The following points should be considered.

The population to be seen

Every consultant working in a general hospital, in primary care, or in a nursing-home setting needs to define what patients have to be seen and what services are to be delivered. In a general hospital, emergency services will be required for patients who attend for psychological reasons, including attempted suicide, and for patients with substance abuse withdrawal and acute deliria (Table 5.7.1). In addition consults should be done for patients with unexplained physical problems and other complex illness behaviour. The consult capacity beyond these two consult categories should be used for the development of more preventive consults integrated in existing forms of multidisciplinary service delivery. Selection of areas of interest depends on several factors, such as service delivery or research priorities of the hospital (for instance transplant or oncology) or an own research agenda. In primary care the target population in addition to the chronically ill will be patients with somatization problems including affective and anxiety disorders (see Chapter 5.2.3).

Psychiatric assessment

As in other settings, the formal psychiatric assessment is a crucial part of the services delivered. Specific to the setting is the differential diagnosis with physical disorders, the role and the meaning of physical deregulations, effects of pharmaca, the effects of the psychiatric disturbance on compliance with the treatment of the physical disease, the consequences of the assessment for the integrated prognosis, and the subsequent long-term integrated management of the patient. As the outcome of psychiatric disorders is clearly related to the interfering problems, which contribute to the complexity of patients, inclusion of the assessment of potential risks for such problems should be considered.[19]

Disciplines and staffing

The size and composition of the CL team needs to be defined depending on the size and type of the hospital and of the target group of patients (Table 5.7.1) as well on the financial possibilities and other available services. For the basic function, the assessment and treatment of patients seen for attempted suicide, one is referred to Chapter 4.15.4. In a European study (ECLW Collaborative Study) it became clear that there was a variation in team composition from monodisciplinary (medical model) to multidisciplinary (mental health model) depending on the size of the service as well as the country.[8] In addition to individual psychotherapeutic treatment, mutual adjustment with and instructions of other caretakers is a key aspect of CL work. Consequently, nowadays CL psychiatry cannot provide optimal care without team members focusing on psychological treatment and the organization of case-management required for long-term individualized care-trajectories. Good evidence is becoming available that psychological interventions (cognitive–behavioural, problem solving, and interpersonal psychotherapies) are effective in patients with physical illness and depression as well as unexplained physical complaints;[20,21] (see also Chapter 5.2.3). The effectiveness of interventions of CL psychiatric nurses depends on their roles. For the effectiveness of case-management in these patients is less evidence according to a recent systematic review.[22] Turning it another way around, as CL psychiatric nurses will often work in the chronic medically ill, a recent review has provided an overview to assess the effective elements of chronic disease management (Table 5.7.3).[23] To be able to contribute to integrated care programmes, such as for diabetes or for haemodialysis, tertiary care hospitals should have teams with, on average, one full-time equivalent of psychiatric staff per 300 beds and a secretary, in addition to psychiatric residents, nurses, and psychologists.[24] Both psychiatric and complexity screening functions need to be supported by manpower to translate the findings into clinical action, such as the design and implementation of a long-term individualized care trajectory and prevent decompensation in those who are vulnerable.

Relationship between medical staff, hospital board, and regional mental health facilities

For the development of more integrated services beyond emergency consultations good working relationships are required. The psychiatrist should be a formal member of the medical staff. Negotiations on the size and focus of CL psychiatric service delivery

Table 5.7.3 Evidence for effective chronic care management

There is evidence to support the following initiatives
- Broad chronic care management models
- Integrated community and hospital care
- Greater reliance on primary care
- Identifying people at greatest risk of complications and hospitalizations
- Involving people with long-term conditions in decision-making
- Providing accessible structured information for people with long-term conditions and their families
- Self-management education
- Self-monitoring and referral systems
- Electronic monitoring and telemonitoring
- Using nurse-led strategies, where appropriate

There is less evidence to support the following initiatives
- Case-management
- Evidence-based care pathways
- Shared learning among health professionals

There is limited information about
- New models of commissioning services
- Appropriate data collection and monitoring
- Linking health services with voluntary and community services

(From D. Singh (2006), *Transforming chronic care*, NHS Surrey and Suffolk UK. © Crown copyright.)

should be organized with representatives of the general hospital board, the regional mental health provider, and the health plan. They should decide on functions, budgets, and facilities. The lack of medical facilities in mental health institutions is a good reason to include the need to develop a psychiatric medical unit in the general hospital to serve the more serious medically ill from mental health institutions. Wards with both a psychiatric and medical function can solve the problems created by the artificial division between general and mental health care.

Audit

Both for financial purposes as well as for strategic planning practice audit is required. It does not make much sense to have an extensive audit system such as used in studies, unless this is used for projects. Otherwise an audit system integrated in the hospital mainframe seems the most appropriate, including the basic patient documentation, the reason for referral, the referring department, their diagnoses, and treatment.[25]

Training

In the 'western' world CL psychiatry is becoming more and more an area of special interest, which is reflected in a subspecialty-status in several countries, such as the United States, the United Kingdom, and Australia. Training should include specific medicopsychiatric aspects of the work, including psychopharmacology in the medically ill. Guidelines have been formulated by several associations and have been published.[26, 27]

Further information

Associations for consultation-liaison psychiatry exist, which organizational format differs by country. Since the first decade of the twenty-first century there is an increasing international exchange between leading organizations as well as among leaders in the field. Leading associations are

- The Academy of Psychosomatic Medicine in the USA: www.apm.org. This organization has an international membership and focus and
- The European Association of Consultation and Liaison Psychiatry and Psychosomatics: www.eaclpp.org

References

1. Huyse, F.J. and Stiefel, F.C. (eds.) (July 2006). *Integrated care for the complex medically ill*. Medical Clinics of North America, Elsevier.
2. Institute of Medicine. (2001). *Crossing the quality chasm: a new health system for the 21st century. Committee on quality of health care in America*. National Academy Press, Washington, DC.
3. Institute of Medicine. (2005). *Improving the quality of health care for mental and substance-use conditions: quality chasm series*. National Academy Press, Washington, DC.
4. Royal College of Physicians and Royal College of Psychiatrists. (1995). *The psychological care of medical patients. Recognition of need and service provision. Council report CR35*. Royal College of Physicians and Royal College of Psychiatrists, London.
5. Wallen, J., Pincus, H.A., Goldman, H.H., *et al.* (1987). Psychiatric consultations in short-term general hospitals. *Archives of General Psychiatry*, **44**, 163–8.
6. Huyse, F.J., Herzog, T., Lobo, A., *et al.* (2001). Consultation-liaison psychiatric service delivery: results from a European study. *General Hospital Psychiatry*, **23**(3), 124–32.
7. Huyse, F.J., Herzog, T., Lobo, A., *et al.* (2000). European consultation-liaison psychiatric services: the ECLW collaborative study. *Acta Psychiatrica Scandinavica*, **101**, 360—6.
8. Herzog, T., Creed, F., and Huyse, F.J., *et al.* (1994). Psychosomatic medicine in the general hospital. In *Psychiatry in Europe: directions and developments* (eds. C. Katona, S. Montgomery, and T. Sensky), pp. 143–51. Gaskell, London.
9. de Cruppe, W., Hennch, C., and Buchholz, C., *et al.* (2005). Communication between psychosomatic C-L consultants and general practitioners in a German health care system. *General Hospital Psychiatry*, **27**(1), 63–72.
10. Kathol, R., Saravay, S.M., Lobo, A., *et al.* (2006). Epidemiologic trends and costs of fragmentation. *The Medical Clinics of North America*, **90**, 549–72.
11. Egede, L.E. (2006). Disease-focussed or integrated treatment: diabetes and depression. *The Medical Clinics of North America*, **90**, 627–46.
12. De Jonge, P., Huyse, F.J., and Stiefel, F.C. (2006). Case and care complexity in the medically ill. *The Medical Clinics of North America*, **90**, 679–92.
13. Smith, G.C. and Clarke, D. (2006). Assessing the effectiveness of integrated interventions: terminology and approach. *The Medical Clinics of North America*, **90**, 533–48.
14. Wulsin, L.R., Söllner, W., and Pincus, H.A. (2006). Models of integrated care. *The Medical Clinics of North America*, **90**, 647–77.
15. Söllner, W., DeVries, A., and Steixner, E., *et al.* (2001). How successful are oncologists in identifying patient distress, perceived social support, and need for psychosocial counselling? *British Journal of Cancer*, **84**, 179–85.
16. Huyse, F.J., Stiefel, F.C., and de Jonge, P. (2006). Identifiers, or "red flags" of complexity and need for integrated care. *The Medical Clinics of North America*, **90**, 703–12.
17. Kroenke, K. and Rosmalen, J.G.M. (2006). Symptoms, syndromes, and the value of psychiatric diagnostics in patients who have functional somatic disorders. *The Medical Clinics of North America*, **90**, 603–26.
18. Slaets, J.P.J. (2006). Vulnerability in the elderly: frailty. *The Medical Clinics of North America*, **90**, 593–601.

19. Stiefel, F.C., Huyse, F.J., Söllner, W., *et al.* (2006). Operationalizing integrated care on a clinical level: the INTERMED project. *The Medical Clinics of North America*, **90**, 713–58.
20. Katon, W.J., Von Korff, M., Lin, E.H., *et al.* (2004). The pathways study: a randomized trial of collaborative care in patients with diabetes and depression. *Archives of General Psychiatry*, **61**(10), 1042–9.
21. Unutzer, J., Katon, W., and Callahan, C.M., (2002). Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *The Journal of the American Medical Association*, **288**(22), 2836–45.
22. Latour, C.H., van der Windt, D.A., de Jonge, P., *et al.* (2007). Nurse-led case management for ambulatory complex patients in general health care: a systematic review. *Journal of Psychosomatic Research*, **62**(3), 385–95.
23. Singh, D. (2006). *Transforming chronic care*. NHS Surrey and Suffolk, UK.
24. Herzog, T., Stein, B., Söllner, W., *et al.* (2002). *Practice guidelines for consultation-liaison psychosomatics*. Schattauer, Stuttgart.
25. Söllner, W., Stein, B., and Hendrischke, A., *et al.* (2005). A documentation form for the consultation-liaison services: development of the CL-BaDo. *Zeitschrift fur Psychosomatische Medizin und Psychotherapie*, **51**, 310–22.
26. Gitlin, D.F., Schindler, B.A., Stern, T.A., *et al.* (1996). Recommended guidelines for CL psychiatric training in psychiatry residency programs: a report from the academy of psychosomatic medicine task force on psychiatric training in CL psychiatry. *Psychosomatics*, **37**, 3–11.
27. Söllner, W., Creed, F. and the EACLPP Workgroup on training in CL. (2007). European guidelines for training in CL psychiatry and psychosomatics. *Journal of Psychosomatic Research*, **62**(4), 501–09.

SECTION 6

Treatment Methods in Psychiatry

6.1

The evaluation of treatments

Contents

6.1.1 The evaluation of physical treatments

Clive E. Adams

The strengths and weakness of the single trial

Strengths

New treatments, or variations of older therapies, rarely represent a revolutionary departure from what has gone before. As progress is usually made in modest steps, evaluation in prospective randomized trials is needed. These studies, comparing a new treatment with a relevant control, may be able to highlight and quantify relatively subtle but important differences in outcome.

Randomization controls for selection biases. If undertaken carefully, it should ensure that both known and unknown confounding variables, such as age, sex, and additional medications, are evenly distributed between groups. Any differences in outcome should then be due to the treatment, or the intention to give the treatment (see below). In 1991, the World Health Organization stated that the randomized controlled trial, if ethical and feasible, is the most objective means of evaluating mental health interventions.[1]

Certainly, large well-conducted trials, with participants, interventions, and outcomes recognizable to those working in health services, are potent guides to clinical practice. Nevertheless, even when such trials exist, it is important to view them alongside all other comparable evidence. Should the large study affirm the findings of smaller trials the clinician can proceed with confidence. If there is a discrepancy then debate will be generated, which should clarify important issues relating to the participants, interventions, or outcomes measured or to the methods by which the trial was conducted.[2]

Power

As numbers within a study increase so does the precision of results, enabling important but subtle differences to be detected, if they do indeed exist. Should a new treatment be considerably better than its predecessors few people would have to be randomized in order to demonstrate clearly the advantage of the innovative approach. As the advantage expected of new treatments is usually modest, reasonably large studies are often needed.

The power calculation is an important prerequisite for any randomized trial. For example, if clinical observation suggests that a new treatment can help 20 per cent more people avoid admission than the standard care, this can form the basis for a power calculation for a trial. Using a simple formula[3] the trialist can work out how many people would have to be randomized in order to have a known probability of highlighting such a difference, should one really exist. In this case, about 150 people would have to be allocated to each arm of the trial to be reasonably confident of detecting a true 20 per cent difference ($\alpha = 0.05$, $\beta = 0.8$). Most mental health trials are far too small to show up anything but very gross differences between treatments. For example, the average number of participants in schizophrenia trials is about 100 with only a slow increase over time.[4]

A single small trial should not greatly influence the clinician, but the combined results of several studies may begin to have the power to inform practice.

Biases

Randomization attempts to control for the biases that would influence treatment allocation (selection biases). Blinding at outcome attempts to control for biases that would result from participants or raters knowing which treatment had been allocated to whom (observer bias). Inadequate randomization leads to an overestimate of effect in the region of 31 to 40 per cent, and poor blinding at outcome to that of about 17 per cent.[5] Further overestimate results from the use of unpublished or modified scales, commonly seen in mental health studies, and a financial or academic investment in the therapy by the trialists.[6,7]

There are many threats to the validity of a single trial. Viewing all relevant studies, each of which was subject to different degrees of bias, should give a more balanced picture. Of course, the reader of a review should be vigilant for the systematic bias, across all trials, that may consistently sway results one particular direction.

Generalizability

Even if a study is adequately powered and undertaken with due regard for bias, a single trial may be difficult to apply to everyday practice.

(a) Participants

Most studies involve unusual participants. Frequently those eligible for trials have to give informed consent, their problems are well defined and do not involve multiple pathologies, and they are expected to tolerate the demands of a study.

(b) Interventions

Applying the results of a single study is made even more difficult because study interventions are often impractical. For example, drug trials may use rigid dose regimens impossible to apply to routine care. Psychosocial therapies tested within a trial are often of such high quality that they bear little resemblance to what an overstretched clinical service can provide.

(c) Outcomes

Measurement of outcome may also limit the value of a single trial. In a survey of 2000 schizophrenia trials, 640 different scales were used to record outcomes such as mental state, behaviour, global impression, and adverse effects.[4] Specific subspecialties within psychiatry may have an even greater propensity to create scales for trials.[8] Even within poorly powered trials, these sensitive tools may be able to detect real differences between treatments that may be statistically if not clinically significant. However, few clinicians use such scales in everyday practice and interpreting results becomes a matter of conjecture.

Trials that involve carefully defined groups of participants receiving meticulously controlled treatments and having outcomes measured on sensitive scales are called 'explanatory' studies.[9] Such trials dominate the literature, although calls for more pragmatic or 'real world' methodology are increasing[4,10] and there are now examples of this broader approach.[11,12,13,14] Currently, generalizing from the results of a single trial to day-to-day practice is inadvisable. If, however, several explanatory studies, all undertaken with constrained, but different, methodologies are giving a similar result, the clinician can feel a little more comfortable when acting on their findings.

The rogue result

Even the well-conducted generalizable trial can produce a rogue result. Currently, the acceptable level of chance is one in 20. A statistically significant result, often denoted as $p<0.05$, suggests that the finding, if the experiment was to be replicated, should occur 19 out of 20 times. One time in 20, however, a different result will appear simply because of chance. This can lead to an interesting paradox. A single trial may not provide the best evidence of how to manage people, even in the locality that the study was undertaken. The play of chance may result in an erroneous result and unless that trial is viewed in the company of all relevant evidence, clinicians will be mislead.

Time

It is inadvisable to act on the results of a single trial because of issues of power, biases, generalizability, and the possibility of chance erroneous results. There is, however, also the issue of time. Clinicians may often prefer to read the results of a single review rather than spending time assimilating information from several similar trials. Most practitioners have very little time to keep up with research relevant to their practice. Clinicians admit to having half an hour of reading time per week,[15] and much of that may not be retained.[16] Reading reviews is time efficient.

Reviews

There are two main approaches to the reviewing process—the traditional and the systematic.

The traditional review

The traditional review is often undertaken by a person well respected in the relevant field who uses knowledge and acumen, supplemented by research, to produce a synopsis of the literature. This approach still dominates the current texts, journals, and lecture tours. For example, in 1987 in four major North American medical journals, 86 per cent of review articles depended on qualitative synthesis and contained no 'methods' section whatsoever.[17] In only 6 per cent of the reviews was quantification attempted in order to support opinion and the situation did not improve much across the next decade.[18] Therefore, the clinician is left in a situation where it is difficult not to operate under a double standard. On the one hand, a large relevant trial providing objective evaluation would be desirable, but frequently, a traditional subjective review is all that is available.

Systematic reviews

The form of a systematic review encourages the introduction of basic epidemiological principles and quantification into the process of reviewing. Gene Glass, an educational psychologist, was the first to add the results of similar studies in the hope of quantifying the effects of a treatment.[19] Glass defined 'meta-analysis' as 'the statistical analysis of a large collection of analyses results from individual studies for the purpose of integrating the findings'.[20] Unsurprisingly, in the sensitive area of the psychotherapies, their first and flawed attempts in the new discipline generated controversy.[21] Critics were quick to point out that drawing conclusions from summation of very different types of therapies, undertaken by practitioners of varied experience, was likely to be inadvisable. These pioneers, who even years later are still being criticized for adding 'apples and oranges',[22] are nevertheless owed a great debt by the rest of medicine. After all, it depends on the question being asked. It is fine to mix apples and oranges, if your question is about fruit.[23]

Systematic reviews attempt to minimize bias in the identification, extraction, and summation of relevant data by applying good survey methods to the process of literature reviewing. An analogy may help. In a community survey of the prevalence of mental disorders, a researcher stands on the doorstep of the hospital and suggests that 5 per cent of the population suffers from serious mental illness. By chance, the final estimate may even be correct, but the work could not be seen as methodologically rigorous. The researcher should have written a study protocol, clearly defined an

unbiased sample of individuals to interview, and specified *a priori* the analyses to be undertaken. A systematic review should do this for a survey of a 'population' of relevant literature. Within such a review, the objectives, criteria for selection of relevant studies, search strategy, methods of study selection, data extraction, and assimilation are all made explicit.

The advantages of the systematic approach

As is suggested above, a systematic review may, by adding the results of similar studies together, at least begin to address the issue of the underpowered study, single trials of biased methodology, poor generalizability, and idiosyncratic results. Although often longer than most traditional reviews, the systematic review is still a time-efficient way to appraise research. Additional advantages are both intuitive and practical.

(a) Objectivity

Medicine remains a scientific discipline and the attempt at objective quantification must be an integral part of this approach. However, the systematic review and meta-analysis should never become a source of clinical tyranny. Individual clinicians will always have to use wisdom and judgement in their day-to-day decision-making, but to exclude objective appraisal from this process is foolhardy.

(b) Clinical empowerment

Systematic reviews can provide clear information to clinicians, policy makers, and recipients of care, and so help inform the decision-making process. For example, a systematic review of family therapy suggests that this educational, psychosocial package can help those with schizophrenia avoid or postpone relapse.[24] This finding is very much in line with the suggestions of traditional reviews.[25] However, the systematic review is able to illustrate how seven families have to undergo regular therapy, for up to a year, in order for a single relapse to be postponed. Such data, of course, mean different things to different people. Clinicians may find this an acceptable degree of effort, whereas managers of services, or even families of those with schizophrenia, may not. Although the findings may not decrease controversy, at least debate can be informed.

The quantification of trial data can sometimes provide information quite at odds with the advice of traditional reviewers. The best example comes from outside mental health care. In 1992, Antman *et al.* undertook a meta-analysis of randomized trials evaluating the care of those with acute myocardial infarction.[26] As the reviewers added trial data, they found that by 1973 enough studies existed to show clearly that thrombolytic treatment saved lives. Subsequent trials added precision to the result but did not change the finding. Antman and colleagues also showed how traditional reviews continued to fail to mention thrombolytic therapy up to 15 years after the summated trial data could have shown its value.[26] These traditional reviews recommended treatments for myocardial infarction that were positively harmful. Examples have emerged from mental health. Sometimes, traditional reviews make bold claims or recommendations which are not supported by the evidence in quantitative, systematic reviews. For example, some claims made for cognitive therapy for schizophrenia[27] go well beyond objectively summarized evidence.[28,29] Conversely, even when evidence is of high quality and readily accessible, traditional reviewers can be blind. For example, both the strengths and the limitations of new generation antipsychotic drugs have been evident for years[30,31,32] but traditional reviews and texts have encouraged uncritical enthusiasm for their use[33,34] and even guilt for failing to prescribe.[35]

(c) Gaps in research

Often a systematic review will highlight unsuspected gaps in research. The trial-base of much routine practice is not strong, and systematic reviews can help shape questions to be tested in well-planned and conducted trials.[36] Certainly, some research funders are now requiring that a systematic review be undertaken before a randomized trial is funded. This also avoids wasting resources on questions that have already been answered.

The limitations of the systematic approach

(a) Qualitative information

Systematic reviews focusing on the value of treatments given to those with mental health problems usually involves quantitative synthesis of data from randomized trials. Incorporating the great wealth of information from more qualitative approaches in an unbiased way is problematic.

(b) Trial content and quality

Systematic reviews are limited by trial content and quality. For example, it is feasible that, on average, those taking a new drug may have a statistically significant 10-point-greater decrease in a modified Brief Psychiatric Rating Scale[37] score than those taking the comparison treatment. First, this finding is difficult to put into clinically meaningful terms. Second, most scales do not provide 'interval' data. A 10-point change for someone who started with a very high score may not mean the same as the same change for a person entering the study with a lower rating. Third, more problems stem from the modification of the scale. This may well not be published and so validity is questionable. The use of such data is associated with an overestimate of effect.

Undertaking a systematic review of poor-quality trials is an important prerequisite for the design of good studies, but clinical interpretation can be problematic.[38]

(c) Rare outcomes

Randomized trials are not a powerful means of identifying rare but important outcomes. For example, large cohorts of those taking the 'atypical' antipsychotic drug, clozapine, suggest a rate of about 1 per cent for agranulocytosis, a serious adverse effect.[39] However, a systematic review of all relevant randomized trials finds a much lower incidence.[40] As the most vulnerable period for the occurrence of agranulocytosis is from weeks 6 to 18 of treatment,[39] and most studies in the systematic review were of shorter duration, the incidence was underestimated.

Trials have limited power to identify rare outcomes and, although systematic reviews may increase this power, reviews of studies of different methodologies may be needed to quantify these important effects.

(d) Limited statistical methods

The statistics used to summate data within meta-analyses are still evolving. For example, much of the continuous scale data, seen so frequently in mental health trials, is not normally distributed. How robust the commonly used methods of meta-analysis are for these non-parametric data is not clear. In addition, as mental health begins to evaluate preventive interventions then cluster randomization, where communities or institutions are randomized rather

than individuals, will become more common. The statistics for a meta-analysis of these studies is still a matter of debate.[41] Frequently, a systematic review of mental health trials must present, but not summate, relevant data.

The methods of systematic reviews

Setting the question

Clinical questions regarding the effects of treatments have three parts: the participants (who are the people of interest to the questioner?), the interventions (what are the specific treatments that are to be the focus of the review?), and the outcomes (what are the outcomes of interest to the reviewers?) . Although the reviewers may have knowledge of existing trials and their limitations, it is important that the questions set are relevant to the review's readership. If the review is to service clinicians then clinically relevant outcomes must be a priority and not necessarily those anticipated by foreknowledge of the trials. If all studies then provide data on mean change in the Brief Psychiatric Rating Scale and fail to mention the outcome of 'clinically important improvement', the review can highlight this important gap in knowledge.

Developing an answerable question. The next stage for formulating the question is to decide on the type of study that is best suited to answering the question. For questions related to the efficacy of treatments this is usually the randomized trial. At first glance, this may seem straightforward, but it is important to state *a priori* whether studies that implied, but did not state, randomization should be included. No other methodological parameter is so consistently linked with exaggerated estimations of effect than poor description of randomization.[5] Studies that describe themselves as 'prospective, double-blind, evaluative controlled trials' would be excluded from a review if the entry criteria demanded an explicit description of randomization.

Identifying studies

Studies are usually identified by searching bibliographic databases such as EMBASE, MEDLINE, or PsycINFO. Hand-searching relevant journals, conference proceedings, and references is also often undertaken.

A systematic review would be a misnomer if the researchers did not make the means of identification of studies clear and reproducible. The exact source of trials, and the search strategies, must be explicit. It is at this poin`t that the advice of an information specialist is important. The coverage of mental health journals in many bibliographic databases is poor and often limited by region or language[42] so that searching several sources is advisable. For example, the last decade China has become highly productive of mental health trials[43] but few are reported in mainstream bibliographic databases.[44]

In recent years, however, the situation for those wishing to identify all treatment trials relevant to a particular topic has become easier. The Cochrane Controlled Trials Register within the Cochrane Library,[45] is the largest and most comprehensive bibliographic database of published and unpublished randomized trials, and controlled clinical studies, in existence. For citations of trials, this specialist register has eclipsed databases such as EMBASE, Medline, and PsycINFO. Searching the Cochrane Controlled Trials Register also avoids the problem of the numerous 'false' positive citations produced by searches of unspecialized biomedical databases.

Identifying every possible study is important. Potent biases operate in this area. Trials that have statistically significant results are more likely to be published than those reporting equivocal findings,[46] and they are more likely to be published in English.[47] Systematic reviews incorporating only Anglophone published data or trials from one region are likely to produce, at best, imprecise, and at worse, overoptimistic views of efficacy.

Selecting studies

Once a search is completed, relevant studies must then be selected without bias. Reviewers usually work independently and document the outcome of all disputed decisions. Some feel that those selecting the trials should be blind to the study's author, source (usually a journal), and the institution where the trial was undertaken. All have potential to bias the study selection. However, such blinding would often involve prohibitive effort. In any event, a systematic review should make explicit the degree of effort made to avoid selection bias at this crucial stage.

Quality assessment of trials

Once studies are identified and prespecified entry criteria met, a last set of quality criteria may be applied. Scales are available, but essentially they rate selection and observation bias (see above). The description of concealment of allocation is central, as this methodological parameter has consistently been shown to be linked with an estimate of effect. If this is poorly reported, the trial is likely to overestimate the effect of the experimental intervention.[5] For trials that describe allocation with nothing more than 'randomized', this single parameter may not be a sensitive measure of quality. A scale addressing both selection and observation bias by rating the description of randomization, blinding, and reasons for people withdrawing, may be more appropriate.[48]

A systematic review should prestate the level of quality that is acceptable, or, at the very least, how the data from poor-quality studies are to be managed.

Data extraction

Reliable data extraction is important. Just as studies must be selected with due regard for the inclusion of bias, so data must be extracted carefully and reproducibly. Often reviewers ensure maximum reliability by organizing double data extraction by an independent reviewer.

Data management

(a) *A priori* primary analysis

As with any quantitative research, a systematic review will generate the potential for multiple analyses. As one in 20 will be statistically significant by chance, it is important to state *a priori* the primary analyses to be undertaken. Although multiple secondary analysis are often undertaken, these are only hypothesis-generating as data have been multiply tested.

(b) Unacceptable loss to follow-up

In every study, there must be a certain attrition that renders data meaningless. For example, in a trial of tacrine for those with Alzheimer's disease 68 per cent of people taking the experimental compound were withdrawn or lost to follow-up.[49] Drawing conclusions from the data provided by the 32 per cent of 'completers' is problematic as selection bias, originally addressed by

randomization, is likely to be great. Trial attrition may not be immediately apparent from first glance. For example, a meta-analysis of studies comparing the antipsychotic quetiapine with chlorpromazine and haloperidol for schizophrenia shows considerable loss to follow-up at only a few weeks.[50] The last observation of those leaving was carried forward to the results, so that data presented in the trials were on the numbers originally randomized. The trialists made an assumption that data collected just before leaving the study would reflect the situation at the end of the trial. These assumptions by the trialists may or may not be justified[51] but reviewers must also make judgements. It is crucial to make these important decisions explicit and to make them before seeing the data.

The limit at which data become meaningless may differ depending on the question addressed. For example, in the situation of trialling a new oral drug for schizophrenia, clearly a loss of nearly 60 per cent of people at 6 weeks is clinically untenable. The reviewer may judge that the unfortunate clinician may lose up to 30 per cent of people by 6 weeks but that any greater loss would reflect more than misfortune and render data of little use. In different circumstances, such as the acute care of very disturbed people in closed wards, the loss of even 10 per cent of participants could be seen as a threat to the value of the data presented.

(c) Intention-to-treat analysis

Interventions are not randomized in trials — it is the *intention* to give treatments that is randomly allocated. Once people are lost to follow-up, the property of the randomization to distribute known and unknown confounding variables is under threat. The randomization has, in effect, been broken. The real threat of an introduction of selection bias has led to the phrase — once randomized, always analyze.[3]

Once a limit to trial attrition has been set, reviewers must, before seeing trial data, exercise more judgement in what outcome is to be attributed to those who were lost. It is impossible to avoid assumptions, but these should be based on common sense if not evidence. For example, when presenting data for the outcome of 'clinically improved', reviewers could assume, unless contrary information is provided in the trials, that those who left early did not have an important recovery. If good quality sources of information are available, this assumption can become more evidence-based. Perhaps an exemplary trial within a systematic review managed 100 per cent follow-up even on those who left the study early. If this trial found that 90 per cent of those who had not complied with the study protocol were not 'clinically improved', it would provide a rationale for applying this figure to the other trials in the meta-analysis. Unless individual patient data are available, this process is impossible for continuous outcomes and only 'completer' data must be presented.

(d) Continuous data

Data on continuous outcomes are frequently skewed, with the mean not being the centre of the distribution. The statistics for meta-analysis are thought to be able to cope with some skew, but were formulated for parametric normally distributed data. Reviewers may wish to build in simple rules to avoid the potential pitfall of applying parametric tests to very skewed data. For example, in scale data where a mean endpoint score is provided with a standard deviation, when the latter is multiplied by 2 and is then greater than the mean, data could be stated to be too skewed to summate.[52] This rule cannot be applied for scale data reporting change, rather than endpoint, scores.

A wide range of rating scales are available to measure outcomes in mental health trials. These scales vary in quality and many are poorly validated. It is generally accepted that measuring instruments should have the properties of reliability (the extent to which a test effectively measures anything at all) and validity (the extent to which a test measures that which it is supposed to measure). Before publication of an instrument, most scientific journals insist that both reliability and validity be demonstrated to the satisfaction of referees. Reviewers may well decide, as a minimum standard, to exclude data from unpublished rating scales.[38]

(e) Individual patient data

Most mental health meta-analyses are of aggregate data from published reports. Other specialities have set a 'gold standard' for systematic reviews by acquiring, checking, and reanalyzing each person's data from the original trialists.[53] Collecting individual patient data allows reviewers to undertake time-to-event analyses and subgroup analyses, to ensure the quality of the randomization and data through detailed checking and correction of errors by communication with triallists, and finally to update follow-up information through patient record systems (such as mortality registers).

Limited empirical evidence exists for some of the advantages of individual patient data reviews over other types of review. The former does help to control publication bias, to ensure use of the intention-to-treat principle in the analysis, and to obtain a fuller picture of the effects of different treatments over time. Undertaking individual patient data reviews requires considerable additional skills, time, and effort on the part of the reviewers when compared to meta-analyses of published aggregate data.[54]

Statistics

(a) Inappropriate meta-analyses

Systematic reviews may, or may not, contain meta-analyses. Where participants, interventions, or outcomes are clearly too different to summate, reviewers must resist the temptation to use powerful statistics on inappropriate data.[41]

(b) Summary measures

Much has been written about the statistics for meta-analysis,[55] but if their meaning is not conveyed to the user of the review, they are of little value. Summary measures such as odds ratios or relative risk are frequently employed for dichotomous outcomes and weighted and standard mean difference for continuous data. Where continuous data are presented from different scales measuring similar phenomena then standard mean difference is often calculated and called the effect size. This estimate has statistical integrity, but is even more problematic to interpret clinically than weighted mean difference. In each of these summary measures, an individual trial contributes to the final statistic, inversely proportional to the precision of its result.

Currently, in this new discipline, statistics for meta-analysis are powerful and evolving. In recent years, there has been a much better understanding of how best to summate data from cluster randomized trials and crossover studies. Techniques for these types of analyses are now widely accessible. Further statistical flexibility is available through use of generic inverse variance. This statistical

technique facilitates analyses of properly analyzed cross-over trials, cluster randomized trials and non-randomized studies, as well as outcome data that are ordinal, time-to-event or rates. Better methods for managing non-parametric data can confidently be expected in the next few years.[41]

(c) Sensitivity analyses

This is where an analysis is used to determine how sensitive the results of the review are to changes in how it was undertaken. For example, the reviewers may state, again *a priori*, that they wish to compare the size of effect of industry-sponsored trials versus those undertaken independently of the manufacturer of the experimental drug.[7] The sensitivity of the final result to adding and subtracting sets of trials is then tested. Sensitivity analyses can be proposed on many variables, such as severity of illness, age of participant, means of diagnosis, subtype of intervention, and quality of trial. This can easily lead to the problems of multiple testing, although for meta-analyses of published data the quality and extent of trial reporting severely restricts the numbers of sensitivity analyses that are possible.

(d) Heterogeneity

In systematic reviews heterogeneity refers to variability or differences between studies' estimates of effects and is a function of clinical and/or methodological diversity among the studies. Despite rigorous definition and application of inclusion criteria, the trials eventually selected may not be homogeneous enough to summate. Heterogeneity should be considered, sought, and measured, and, if present, investigated. Statistical tests of heterogeneity are used to assess whether the observed variability in study results (effect sizes) is greater than that expected to occur by chance. These tests, however, have low statistical power and careful inspection of results for outlying findings is just as valuable. Recently, partly because it is more intuitive than previous measures, the I^2 has become widely used. This quantifies inconsistency across studies, moves the focus away from testing whether heterogeneity is present to assessing whether the heterogeneity, that many argue is inevitable, will impact on the meta-analysis. A value greater than 50 per cent is often considered to indicate substantial heterogeneity.[41]

Heterogeneity can be caused by various factors, and its presence generates debate about differences in study design (methodological diversity) and differences between studies in key characteristics of the participants, interventions, or outcome measures (clinical diversity). Heterogeneity can be explored by undertaking subgroup analyses or employing meta-regression. In subgroup analyses, the results of one group of studies with a characteristic thought to be causing heterogeneity is compared to those of another group of trials without that characteristic. For example, trials of a new interpersonal therapy for depression may be heterogeneous. The reviewers may have stated, *a priori*, that the new therapy, targeted at young people, may not be effective for those over 30 years of age. Overall, although the trials may have heterogeneous results, if the two subgroups of studies involving people over and under the age of 30 have homogeneous results, it is feasible that the original hypothesis of the reviews regarding age was correct. Meta-regression, is an extension to subgroup analyses that allows the effect of variables, even several simultaneously, to be investigated. Again these variables should be pre-stated and this technique necessitates much high quality data to be meaningful. Even then, the regression is essentially an observational study with all the dangers of unexplained or residual confounding. In mental health reviews, investigation of heterogeneity often generates valuable debate, rather than providing definitive answers.

(e) Reporting bias

There are several ways to assess whether reporting (publication) bias (see above) is operating within a review. The reviews may use a funnel plot technique[56] where the results of a trial are plotted against its size. Large studies with any result, positive or negative, tend to be published. Small positive studies are also usually easily identified, but it quickly becomes apparent if small 'negative' studies have not been found. This could be due to a variety of reasons, but one of which is a selective reporting of positive outcomes.

Sources of systematic reviews

Journals

Systematic reviews are increasingly seen in major journals. These can be identified by simple or comprehensive methodology-specific searches of bibliographic databases such as EMBASE, Medline, or PsycINFO. Examples of these searches can be readily identified[57] but need to be updated periodically to reflect changes in notation or indexing. These search phrases would then be linked to a subject-specific phrase such as 'depression' or 'cognitive therapy', by the word 'and' to limit and focus the number of identified citations.

DARE database

The Database of Abstracts of Reviews of Effectiveness (DARE) is a specialist database containing thousands of citations of systematic reviews of health care. These reviews are assessed according to explicit criteria, and structured abstracts describing methodology and results, and commenting on quality and clinical implications, are also included.

Cochrane reviews

The Cochrane Database of Systematic Reviews is an electronic publication increasingly favoured by clinicians, researchers and those compiling guidelines. It contains the full text and data of reviews undertaken to the most rigorous systematic standards[58] and is updated quarterly. There is often a considerable lag time for traditional journal publication and this can result in the publication of good-quality but misleading systematic reviews. For example, an important systematic review on the effects of family intervention for schizophrenia was published in August 1994.[59] Just 2 weeks later, the same authors re-summated relevant data in an electronic version and the results were much less favourable than had been previously reported.[60] In the considerable period between completion of the paper version and its publication, less favourable trials appeared. Further updates of this review suggest that the trend is continuing. By using electronic publishing, the Cochrane Database of Systematic Reviews allows trends over time to be highlighted and reviews to be maintained.

The Cochrane Library is a collection of databases supplying high-quality evidence to inform all those interested in the evaluation of health care.[45] It is published quarterly on CD-ROM and the Internet, and includes the Cochrane Database of Systematic Reviews, the source of many of the reviews quoted in this

chapter, the DARE abstracts, and The Cochrane Controlled Trials Register.

Summary publications

There are now several periodically updated, systematically compiled, publications designed primarily for busy clinicians. For example, one such publication, Clinical Evidence,[61] now has a comprehensive mental health subset, and is available in short or very short versions, providing clinical 'bottom lines' of relevant systematic reviews or randomized trials.

Further information

The Cochrane Handbook - http://www.cochrane-handbook.org/ contains up-to-date methodological details of systematic reviews and meta-analysis

Clinical Evidence - http://clinicalevidence.bmj.com/ceweb/index.jsp an on-line, regularly updated compendium of systematic evidence summaries

References

1. WHO Scientific Group on Treatment of Psychiatric Disorders (1991). Evaluation of methods for the treatment of mental disorders. *World Health Organization Technical Report Series*, **812**, 1–75.
2. Egger, M. and Davey-Smith, G. (1995). Misleading meta-analysis. *British Medical Journal*, **311**, 753–4.
3. Pocock, S. (1989) *Clinical trials: a practical approach*. Wiley, Chichester.
4. Thornley, B. and Adams, C. (1998). Content and quality of 2000 randomised controlled trials in schizophrenia over 50 years. *British Medical Journal*, 1181–4.
5. Juni, P., Altman, D. and Egger, M. (2001). Systematic reviews in health care: Assessing the quality of controlled clinical trials. *British Medical Journal*, **323**, 42–6.
6. Heres, S., Davis, J., Maino, K., *et al.* (2006). Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of headto- head comparison studies of second-generation antipsychotics. *American Journal of Psychiatry*, **163**, 185–94.
7. Montgomery, J., Byerly, M., Carmody, T., *et al.* (2004). An analysis of the effect of funding source in randomized clinical trials of second generation antipsychotics for the treatment of schizophrenia. *Controlled Clinical Trials*, **25**, 598–612.
8. Cure, S., Chua, W., Duggan, L. *et al.* (2005). Randomised controlled trials relevant to aggressive and violent people, 1955-2000: a survey. *British Journal of Psychiatry*, **186**, 185–9.
9. Roland, M. and Torgerson, D. (1998). What are pragmatic trials? *British Medical Journal,* **316**, 285.
10. Simon, G., Wagner, E. and Vonkorff, M. (1995). Cost-effectiveness comparisons using «real world» randomized trials: the case of new antidepressant drugs. *Journal of Clinical Epidemiology*, **48**, 363–73.
11. Lieberman, J., Stroup, T., McEvoy, J., *et al.* (2005). Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New England Journal of Medicine*, **353**, 1209–23.
12. Lewis, S., Davies, L., Jones, P., *et al.* (2006). Randomised controlled trials of conventional antipsychotic versus new atypical drugs, and new atypical drugs versus clozapine, in people with schizophrenia responding poorly to, or intolerant of, current drug treatment. *Health Technology Assessment*, **10**, iii-iv, ix-xi, 1–165.
13. TREC Collaborative Group (2003). Rapid tranquillisation for agitated patients in emergency psychiatric rooms: a randomised trial of midazolam versus haloperidol plus promethazine. *British Medical Journal*, **327**, 708–13.
14. Alexander, J., Tharyan, P., Adams, C., *et al.* (2004). Rapid tranquillisation of violent or agitated patients in a psychiatric emergency setting. Pragmatic randomised trial of intramuscular lorazepam v. haloperidol plus promethazine. *British Journal of Psychiatry*, **185**, 63–9.
15. Sackett, D., Richardson, W., Rosenberg, W. *et al.* (1997). *Evidence based medicine: how to practice and teach EBM*. Churchill Livingstone, London.
16. Kellett, C., Hart, A., Price, C., *et al.* (1996). Poor recall performance of journal-browsing doctors. *Lancet*, **348**, 479.
17. Mulrow, C. (1987). The medical review article: state of the science. *Annals of Internal Medicine*, **106**, 485–8.
18. McAlister, F., Clark, H., van, W.C., *et al.* (1999). The medical review article revisited: has the science improved? *Annals of Internal Medicine*, **131**, 947–51.
19. Smith, M. and Glass, G. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, **32**, 752–60.
20. Glass, G. (1976). Primary, secondary and meta-analysis of research. *Education Research*, 3–8.
21. Eysenck, H. (1978). An exercise in mega-silliness. *American Psychologist*, **33**, 517.
22. Eysenck, H. (1994). Meta-analysis and its problems. *British Medical Journal*, **309**, 789–92.
23. Higgins, J. and Green, S. (2005). *Cochrane Handbook for Systematic Reviews of Interventions 4.2.5 [updated May 2005]*, Chichester, UK, John Wiley & Sons, Ltd.
24. Pharoah, F., Mari, J., Rathbone, J. *et al.* (2006). Family intervention for schizophrenia. *Cochrane Database of Systematic Reviews*, CD000088.
25. Leff, J. (1994). Working with the families of schizophrenic patients. *British Journal of Psychiatry. Supplement*, 71–6.
26. Antman, E., Lau, J., Kupelnick, B., *et al.* (1992). A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *Journal of American Medical Association*, **268**, 240–8.
27. Kingdon, D. (2006). Psychological and social interventions for schizophrenia. *British Medical Journal*, **333**, 212–3.
28. McKenna, P. (2006). What works in schizophrenia: cognitive behaviour therapy is not effective. *British Medical Journal*, **333**, 353.
29. Jones, C., Cormac, I., Silveira, D.M.N.J. *et al.* (2004). Cognitive behaviour therapy for schizophrenia. *Cochrane Database of Systematic Reviews*, CD000524.
30. Srisurapanont, M., Disayavanish, C. and Taimkaew, K. (2000). Quetiapine for schizophrenia. *Cochrane Database of Systematic Reviews*, CD000967.
31. Duggan, L., Fenton, M., Dardennes, R., *et al.* (2000). Olanzapine for schizophrenia. *Cochrane Database of Systematic Reviews*, CD001359.
32. Kennedy, E., Song, F., Hunter, R., *et al.* (2000). Risperidone versus typical antipsychotic medication for schizophrenia. *Cochrane Database of Systematic Reviews*, CD000440.
33. Adams, C. and Jayaram, M. (2007). Do findings from new trials for schizophrenia fit with existing evidence: not duped ……. just beguiled? *Epidemiologia e Psichiatria Sociale*, **16**, 199–202.
34. Vedantam, S. (Oct 3, 2006). In antipsychotics, newer isn't better – drug find shocks researchers. *Washington Post* A01.
35. Adams, C., Tharyan, P., Coutinho, E. *et al.* (2006). The schizophrenia drug-treatment paradox: pharmacological treatment based on best possible evidence may be hardest to practise in high-income countries. *British Journal of Psychiatry*, **189**, 391–2.
36. Brown, P., Brunnhuber, K., Chalkidou, K., *et al.* (2006). How to formulate research recommendations. *British Medical Journal*, **333**, 804–6.
37. Overall, J. and Gorham, D. (1962). The Brief Psychiatric Rating Scale. *Psychological Reports*, **10**, 799–812.
38. Marshall, M., Lockwood, A., Bradley, C., *et al.* (2000). Unpublished rating scales: a major source of bias in randomised controlled

trials of treatments for schizophrenia. *British Journal of Psychiatry*, **176**, 249–52.
39. Atkin, K., Kendall, F., Gould, D., *et al.* (1996). Neutropenia and agranulocytosis in patients receiving clozapine in the UK and Ireland. *British Journal of Psychiatry*, **169**, 483–8.
40. Wahlbeck, K., Cheine, M. and Essali, M. (2000). Clozapine versus typical neuroleptic medication for schizophrenia. *Cochrane Database of Systematic Reviews*, CD000059.
41. Deeks, J., Higgins, J. and Altman, D. (2005). Analysing and presenting results. In: Higgins, J. and Green, S., (Eds.) *Cochrane Handbook for Systematic Reviews of Interventions 4.2.5 [updated May 2005] Section 8. In: The Cochrane Library*, John Wiley & Sons, Ltd., Chichester, UK.
42. McDonald, S., Taylor, L. and Adams, C. (1999). Searching the right database. A comparison of four databases for psychiatry journals. *Health Libraries Review*, **16**, 151–6.
43. Chakrabarti, A., Adams, C., Rathbone, J., *et al.* (2007). Schizophrenia trials in China: a survey. *Acta Psychiatrica Scandinavicia*, (in press).
44. Xia, J., Wright, J. and Adams, C. (2007). Five large Chinese biomedical bibliographic databases: accessibility and coverage. *Health Infomation Libraries Journal*, (in press).
45. Cochrane Collaboration (2007). *The Cochrane Library*, (3rd edn.). John Wiley & Sons, Ltd., Chichester, UK.
46. Easterbrook, P., Berlin, J., Gopalan, R. *et al.* (1991). Publication bias in clinical research. *Lancet*, **337**, 867–72.
47. Egger, M., Zellweger-Zahner, T., Schneider, M., *et al.* (1997). Language bias in randomised controlled trials published in English and German. *Lancet*, **350**, 326–9.
48. Jadad, A., Moore, R., Carroll, D., *et al.* (1996). Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clinical Trials*, **17**, 1–12.
49. Knapp, M., Knopman, D., Solomon, P., *et al.* (1994). A 30-week randomized controlled trial of high-dose tacrine in patients with Alzheimer's disease. The Tacrine Study Group. *Journal of American Medical Association*, **271**, 985–91.
50. Srisurapanont, M., Maneeton, B. and Maneeton, N. (2004). Quetiapine for schizophrenia. *Cochrane Database of Systematic Reviews*, CD000967.
51. Leucht, S., Engel, R., Bauml, J. (2007). Is the superior efficacy of new generation antipsychotics an artifact of LOCF? *Schizophrenia Bulletin*, **33**, 183–91.
52. Altman, D. and Bland, J. (1996). Detecting skewness from summary information. *British Medical Journal*, **313**, 1200.
53. Early Breast Cancer Trialist's Collaborative Group (1992). Systematic treatment of early breast cancer by hormonal, cytoxic, or immune therapy: 133 randomised trials involving 31 000 recurrences and 24 000 deaths among 75 000 women. *Lancet*, **1-15**, 71–85.
54. Stewart, L. and Clarke, M. (1995). Practical methodology of meta-analyses (overviews) using updated individual patient data. Cochrane Working Group. *Statistics in Medicine*, **14**, 2057–79.
55. Petitti, D. (1994). *Meta-analysis, decision analysis and cost-effectiveness analysis. Methods for quantitative synthesis in medicine*. Oxford University Press, Oxford.
56. Egger, M., Davey-Smith, G., Schneider, M. *et al.* (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, **315**, 629–34.
57. Shojania, K. and Bero, L. (2001). Taking advantage of the explosion of systematic reviews: an efficient MEDLINE search strategy. *Effective Clinical Practice*, **4**, 157–62.
58. Jadad, A., Cook, D., Jones, A., *et al.* (1998). Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *Journal of American Medical Association*, **280**, 278–80.
59. Mari, J. and Streiner, D. (1994). An overview of family interventions and relapse on schizophrenia: meta-analysis of research findings. *Psychological Medicine*, **24**, 565–78.
60. Mari, J. and, S.D. Family intervention for schizophrenia. *The Cochrane Library*.
61. Tovey, D. (2007). *BMJ Clinical Evidence*, BMJ Publishing Group Ltd., London.

6.1.2 The evaluation of psychological treatment

Paul Crits-Christoph and
Mary Beth Connolly Gibbons

Introduction

Psychotherapy continues to be a widely practised treatment for psychiatric disorders and other problems in living. Since publication in 1952 of the well-known article by Hans Eysenck,[1] in which he claimed that there was no evidence that psychotherapy was effective, there has been an accelerating literature concerned with methodologies for evaluating psychotherapy, as well as specific studies demonstrating the efficacy, or lack thereof, of various psychotherapies. In more recent years, pressures from the government agencies and insurance companies that bear much of the cost of mental health treatments have added to the call for accountability regarding psychotherapeutic treatment.

Despite a vast literature of over 1000 outcome studies of the effects of psychotherapy, questions remain about the role of psychotherapy as a treatment for mental disorders. Extensive meta-analytical reviews of the psychotherapy outcome literature provided evidence that, generally speaking, psychotherapy appears to be efficacious.[2] While encouraging, this information was not particularly useful. As with any medical problem or disorder, the relevant public health clinical question is whether a treatment is beneficial for the presenting problem or psychiatric disorder for which help is sought. Along these lines, a number of efforts have been made at summarizing the results of the psychotherapy outcome literature in terms of what works for different disorders or problems.[3, 4] For example, these efforts have arrived at conclusions such as 'cognitive therapy is efficacious in the treatment of major depressive disorder'.

The simplicity and clinical appeal of such conclusions, about which psychotherapy treatments work for which patient problems, belies a host of more complex issues regarding how one evaluates psychotherapy and makes a decision about whether treatment 'works' or not. Other treatments within psychiatry, such as pharmacotherapy, lend themselves to rather straightforward designs (namely placebo-controlled randomized clinical trials) that permit clear inferences about the efficacy of a treatment approach. In contrast, research on psychotherapy as a verbal interchange between two or more participants does not have the luxury of such straightforward pharmacotherapy research designs. Instead, psychotherapy outcome research is characterized by the use of a variety of research designs and methods that, while often not without limitations to strong scientific inferences about treatment efficacy, can provide incremental scientific advance in the understanding of the usefulness of psychotherapeutic treatments. The aim of the current chapter is to provide an overview of approaches to the evaluation

of psychological treatments. We begin with a discussion of specific research designs employed in psychotherapy outcome research, with a discussion of some of the broad issues that currently guide the selection among these different experimental designs. This is followed by a selective review of assessment strategies for outcome evaluation, with discussion of examples of instruments.

Issues in planning research evaluating psychotherapy

A number of other sources provide a detailed discussion of issues involved in planning a study on psychotherapy, as well as explication of various research designs. In particular, our presentation draws heavily from Kazdin,[5] supplemented with writings that illustrate more recent trends in both design and methodology.

There are, of course, a wide range of decisions to be made in designing an evaluation study of psychotherapy. These decisions affect the choice of patients, therapists, control groups, data analytical strategies, etc. Table 6.1.2.1 presents a list of the types of questions that need to be asked in designing or evaluating a study of psychotherapy outcome. A discussion of some of the key methodological issues that cut across many of the questions raised in Table 6.1.2.1 follows.

Internal versus external validity

An initial important decision in planning an evaluation of psychotherapy outcome, or any intervention, is the relative emphasis on internal versus external validity of the inferences from the investigation. Internal validity refers to the extent to which inferences can be attributed to the intervention *per se*, as opposed to other factors. In order to maximize internal validity, the investigator attempts to control as many of the extraneous factors as possible through a variety of procedures including, among others, random assignment, the use of control groups, assessing subjects in the same ways and at the same point in time, and careful selection of a relatively homogeneous subject sample. With as many factors as possible held constant across treatment groups except for the nature of the intervention, an outcome difference detected between the experimental and control group(s) can be attributed to the intervention, rather than other factors.

External validity, in contrast, refers to the extent to which the results of a study can be generalized to other subjects, settings, treatment durations, and treatment providers other than those used within the specific study. In regard to the evaluation of psychotherapy outcomes, external validity is often invoked to raise the question of whether study results pertain to the 'real world' in which psychotherapy is practiced—the diverse set of patients, therapists, and settings occurring in the community that may be quite different from the conditions of an investigation conducted in a research setting.

Clearly, both internal and external validity are important, but it is difficult to maximize each within the context of the same study. Studies of homogeneous patient samples, for example, may have high internal validity, but generalize poorly to the mix of heterogeneous patients seen in clinical practice. The relative merits of studies with high internal versus external validity have been a source of ongoing debate among psychotherapy outcome researchers.

Different research designs are more or less appropriate depending upon the scientific question of interest. For example, the process of developing and testing new treatments generally proceeds stepwise, beginning with individual case reports and then progressing to an 'open-label' (a term derived from pharmacotherapy research) trial involving the application of a single treatment to a relatively small group of patients. Following an open-label trial, a promising treatment would then be tested within the context of a controlled, efficacy trial. In this efficacy trial, the treatment would be tested under ideal circumstances (for instance, by highly trained clinicians). If an effect is found in the controlled efficacy trial, a controlled, effectiveness trial is the next step. In this effectiveness trial, the treatment would now be tested under more 'real world' conditions. This line of research is oriented towards understanding whether the treatment *per se* is responsible for change (efficacy trial) and whether the effect generalizes (effectiveness trial).

Naturalistic studies represent an alternative type of effectiveness trial in which the scientific question is usually not a focus of type of treatment. Instead, such studies might examine the relationship of patient characteristics, therapist factors, or length (dose) of treatment to outcome.

Selection criteria for psychotherapy outcome studies

The choice of selection criteria for a psychotherapy outcome study depends, of course, on the nature of the research question to be asked. From a public health perspective, samples are usually chosen based upon the presence of a discrete disorder or problem that has significance to society. The selection of the target disorder, however, is only the beginning of the selection process. For studies of DSM Axis I non-psychotic disorders, it is typical that other major psychotic disorders such as schizophrenia and bipolar disorder are excluded from the study. However, there is wide variability across research studies in the extent to which other Axis I and Axis II disorders are included in a study or not.

This aspect of selection criteria relates primarily to the internal versus external validity distinction discussed above. Studies that emphasize internal validity will probably exclude many comorbid diagnoses, while studies that maximize external validity will tend to be more inclusive. As the comorbidities among Axis I diagnoses can be high, the impact on the nature of the patient sample selected can be considerable.

Naturalistic studies that focus on psychotherapy *per se*, rather than public health concerns, are oriented towards external validity and typically do not have restrictive selection criteria. For these studies, the question is 'how effective is psychotherapy for the types of patients that end up in psychotherapeutic treatment in the community?' Thus, few, if any, selection criteria are specified.

One particular selection problem that affects any type of psychotherapy outcome study is whether or not patients currently treated with psychotropic medication are included in the evaluation study. Once again, from the point of view of internal validity—attempting to attribute the treatment outcome to the psychotherapy treatment *per se*—patients on concurrent medication treatment are usually excluded. In contrast, external validity concerns would lead to the inclusion of patients on medications, since increasing numbers of patients in the community with anxiety and affective

Table 6.1.2.1 Selected questions to raise in planning a study of psychotherapy

Sample characteristics
1 Who are the subjects and how many of them are there in this study?
2 Why was this sample selected in light of the research goals?
3 How was this sample obtained, recruited, and selected?
4 What are the subject and demographic characteristics of the sample (e.g. sex, age, ethnicity, race, socio-economic status)?
5 What, if any, inclusion and exclusion criteria were invoked (i.e. selection rules to obtain participants)?
6 How many of those subjects eligible or recruited actually were selected and participated in the study?
7 With regard to clinical dysfunction or subject and demographic characteristics, is this a relatively homogeneous or heterogeneous sample?
Design
1 How were subjects assigned to groups or conditions?
2 How many groups were included in the design?
3 How are the groups similar and different in how they are treated in the study?
4 Why are these groups critical for addressing the questions of interest?
Procedures
1 Where was the study conducted (setting)?
2 What measures, materials, equipment, and/or apparatus were used in the study?
3 What is the chronological sequence of events to which subjects were exposed?
4 What intervals elapsed between different aspects of the study (assessment, treatment, follow-up)?
5 What variation in administration of conditions emerged over the course of the study that may introduce variation within and between conditions?
6 What procedural checks were completed to avert potential sources of bias in implementing the manipulation and assessment of dependent measures?
7 What checks were made to ensure that the conditions were carried out as intended?
8 What other information does the reader need to know to understand how subjects were treated and what conditions were provided?
Therapists
1 Who are the therapists, and why are these individuals selected?
2 Can the influence of the therapist be evaluated in the design as a 'factor' (as in a factorial design) or can therapist efforts be evaluated within a condition?
3 Are the therapists adequately trained? By what criteria?
4 Can the quantity and quality of their training and implementation of treatment be measured?
Treatment
1 What characteristics of the clinical problem or cases make this particular treatment a reasonable approach?
2 Does the version of treatment represent the treatment as it is usually carried out?
3 Does the investigation provide a strong test of treatment? On what basis has one decided that this is a strong test?
4 Has treatment been specified in manual form or have explicit guidelines been provided?
5 Has the treatment been carried out as intended? (Integrity is examined during the study but evaluated after it is completed.)
6 Can the degree of adherence of therapists to the treatment manual be codified?
7 What defines a completed case (e.g. completion of so many sessions)?
Assessment
1 If specific processes in the clients or their interpersonal environment are hypothesized to change with treatment, are these to be assessed?
2 If therapy is having the intended effect on these processes, how would performance be evident on the measure? How would groups differ on this measure?
3 Are there additional processes in therapy that are essential or facilitative to this treatment, and are these being assessed?
4 Does the outcome assessment battery include a diverse range of measures to reflect different perspectives, methods, and domains of functioning?
5 What data can be brought to bear regarding pertinent types of reliability and validity for these measures?
6 Are treatment effects evident in measures of daily functioning (e.g. work, social activities)?
7 Are outcomes being assessed at different times after treatment?
Data evaluation
1 What are the primary measures and data upon which the predictions depend?
2 What statistical analyses are to be used and how specifically do these address the original hypotheses and purposes?
3 Are the assumptions of the data analyses met?
4 What is the likely effect size that will be found based on other treatment studies or meta-analyses?
5 Given the likely effect size, how large a sample is needed to provide a strong powerful test of treatment (e.g. power ≥ 0.80)?
6 Are there subdivisions of the sample that will be made to reduce the power of tests of interest to the investigator?
7 What is the likely rate of attrition over the course of treatment, and post-treatment and follow-up assessments?
8 With the anticipated loss of cases, is the test likely to be sufficiently powerful to demonstrate differences between groups if all cases complete treatment?
9 If multiple tests are used, what means are provided to control error rates?
10 Prior to the experimental conditions, were groups similar on variables that might otherwise explain the results (e.g. diagnosis, age)?
11 Are data missing due to incomplete measures (not filled out completely by the subject(s) or loss of subjects)? If so, how are these handled in the data analyses?
12 Will the clinical significance of client improvement be evaluated and if so by what method(s)?
13 Are there ancillary analyses that might further inform the primary analyses or exploratory analyses that might stimulate further work?

disorders are receiving psychotropic medications for their problems. Often, a compromise is struck: patients on medications are eligible for the psychotherapy study as long as they (and their prescribing doctor) agree to maintain a stable dosage of the medication for the duration of the psychotherapy study.

Treatment standardization

Psychotherapy efficacy research, like pharmacotherapy research, requires that the treatment be standardized. Such standardization serves two related purposes. First, from a clinical point of view, it is necessary that the treatment be clearly specified, so that any conclusions about differential treatment efficacy can be translated into clear treatment recommendations. From a research point of view, treatment standardization allows studies to be replicated. In addition, by making the delivery of a treatment more standardized, differences between therapists and the statistical problems that result from the non-independence that 'therapist effects' introduce can be avoided.[6]

Standardization of pharmacological interventions is relatively straightforward—a per-day dosage (or range of dosages) is set in advance. But for psychotherapy, how can something so complex as patient–therapist dialogue be standardized? The central ingredient in standardization of a psychosocial treatment is a treatment manual. A psychotherapy manual describes the treatment in detail, with case examples and instruments for psychotherapists. Some treatment manuals, particularly those coming from a cognitive behavioural perspective, present a highly systematized step-by-step programme which therapists follow over the course of therapy. The relative success of treatment manuals in standardizing psychotherapy has been supported by a meta-analysis,[7] which documented that studies employing treatment manuals had fewer outcome differences between therapists compared with studies that did not employ treatment manuals. Thus, when a treatment manual is used, therapists appear to produce relatively more uniform outcomes. In contrast, when no treatment manual is used, therapists differ considerably in their typical outcomes with patients; suggesting that different therapists are likely to be conducting sessions in discrepant ways, with some therapists producing more favourable outcomes and other therapists producing less favourable outcomes.

Treatment standardization, however, does not simply translate to the use of a treatment manual. A variety of steps are needed to ensure that therapists are delivering the intended treatment (Table 6.1.2.2), including: the careful selection of therapists; training of therapists in the intended modality using a treatment manual; certification of therapists based upon their adherence to the treatment model during training; and continuing adherence and competence monitoring of therapists during a clinical trial.

Concerns have been raised about the 'treatment manual' concept applied to less directive treatments such as psychodynamic therapy. The belief is that session-by-session manuals would remove the essence of good psychotherapy, and good dynamic therapy in particular, by making treatment artificially rigid and taking away the necessary clinical flexibility and creativity. Psychodynamic treatment manuals, however, are perhaps better described as 'guides', which specify the principles of treatment but do not overly constrain the necessary clinical flexibility and creativity. The flexibility of treatment is fully retained through the principle of tailoring the treatment intervention to the specific idiosyncratic issues that are salient for each patient. The actual learning of the practice of treatment is accomplished through supervision in the application of the treatment manual. Because dynamic treatment manuals are less like 'cookbooks', there may be a greater reliance on the supervision process compared with perhaps more straightforward behavioural treatments.

Table 6.1.2.2 Steps involved in the standardization of psychotherapy for outcome research

Selection of therapists
Training of therapists using a treatment manual
Certification of therapists based upon adherence to the treatment model
Continued assessment of therapist adherence and competence during a clinical trial

Research designs

Unlike pharmacological research where a single form of research design (placebo-controlled study) dominates the literature, psychotherapy researchers have employed a host of different research designs to understand the effects of psychotherapy. Some of the more common designs are listed in Table 6.1.2.3 and are explicated in the next section.

Single-case designs

Clinical evaluation of the effects of psychotherapy dates back to Freud's descriptions of individual cases in treatment. However, methods to systematically examine the effects of interventions with individual patients were developed by behaviour therapists.[8] These single-case experiments rely on comparing patient responses to differing experimental conditions over time. Typically, such single-case studies begin with an extended baseline period where patient behaviours or symptoms are recorded without any intervention. Then, different intervention phases are introduced, usually followed by more baseline (no intervention) assessment phases.

While experimental, single-case designs lend themselves well to the investigation of behavioural treatments that include a focus on immediate overt behaviour, such designs have rarely been employed with other verbal psychotherapies that emphasize longer term processes such as patient psychological growth and functioning. The generalizability of findings from single-case research is another limitation to this form of research.

Table 6.1.2.3 Common research designs for the evaluation of psychotherapy

Single-case designs
Randomized controlled trials with non-specific or psychological 'placebo' control
Comparative designs
Dismantling or additive designs
Comparisons with medication and pill placebos
Naturalistic designs

Randomized controlled trials

Random assignment of subjects to treatment and control conditions is generally viewed as the preferred method of evaluating the effects of interventions in psychiatry, and in medicine in general. However, perhaps the single most vexing problem in research of psychotherapy outcome is the design of control groups. In pharmacotherapy research, a pill placebo (with 'double blinding', i.e. the patient and the doctor are unaware of treatment assignment) serves to control for all elements of treatment except for the chemical ingredient of interest. Thus, the overall effects of treatment are the sum of the specific effect of the chemical agent plus the effects of 'extraneous' or non-specific factors such as patient expectancy, hope, and aspects of the doctor-patient relationship. With psychotherapeutic treatments, it is less clear which aspects of treatment are specific and which are non-specific. Furthermore, designing a credible control treatment that contains only the non-specific elements of the treatment package is inherently difficult.

(a) Types of control groups

A variety of types of control conditions have been implemented in psychotherapy efficacy studies. One common form of control group is the 'waiting-list' control. Patients are randomly assigned to either the experimental group or to delayed treatment. While such a control condition appears to control for the passage of time (i.e. patients generally improve over time even without treatment), assignment to a waiting-list control condition is likely to immediately change patient hope and expectations about relief of their problem. Patients in the waiting-list condition will not expect to improve until they actually receive treatment, while patients assigned to the experimental condition will likely be more hopeful about change. Thus, the treatment and control conditions are not balanced with regard to expectancy or other non-specific factors such as regular meetings with a competent caring professional. The potential superiority of the experimental condition over the control condition cannot be attributed to the hypothesized active ingredients in the experimental condition. Nevertheless, such waiting-list controls can serve as a useful initial step in evaluating a treatment.

A scientifically stronger type of control group for psychotherapy evaluation studies is one that involves regular meetings with a psychotherapist, but does not contain certain important hypothesized active elements. Some forms of supportive psychotherapy have been used as a control group in this regard. For example, in evaluating the efficacy of cognitive behavioural therapy for posttraumatic stress reactions, a supportive therapy control condition has been successfully used.[9] With some psychotherapy treatments, however, the supportive elements are an important part of the overall treatment model and are hypothesized to be curative in their own right. In this case, a supportive psychotherapy control condition is not appropriate for evaluating the efficacy of the overall package. In fact, sometimes supportive therapy has been found to be superior to other therapies.[10]

Comparative designs

An alternative to attempting to create an adequate psychological 'placebo' control is to only compare active treatments rather than comparing active treatments to 'placebo' control groups. However, rather than solve the problem of how to evaluate the efficacy of a psychotherapeutic intervention, comparative designs simply change the scientific question from 'Is this treatment efficacious?' to 'Which treatment works best?' Comparative designs are generally only informative if one active treatment proves to be more effective than another. If, as is commonly the case in psychotherapy studies, both active treatments produce equivalent results, the investigator is left not knowing whether both treatments are effective, or whether both treatments are non-effective (beyond the effects of non-specific elements).

Dismantling designs

Dismantling designs are an alternative to psychological placebos, waiting lists, or comparative designs. In a dismantling design, the full clinical treatment package is compared with the full package minus one element in order to establish which elements are necessary and sufficient for change. Variations on this theme include 'additive' or constructive designs that examine whether adding a new element enhances the efficacy of a treatment package.

As an example, dismantling designs have been usefully applied in the evaluation of the behavioural treatment of obsessive-compulsive disorder. The full treatment package involving exposure and response prevention techniques has been compared with each of the individual components of the package.[11] Patients were randomly assigned to either the full exposure and response prevention package, exposure alone, or response prevention alone. Those receiving the full package improved significantly more than did the patients in either of the control treatments. At follow-up assessments, 80 per cent of patients receiving the full package remained improved at follow-up, whereas only 27 per cent of those in the single components groups remained 'improved'.

The advantage of dismantling strategies in psychotherapy research is clear. Causal statements about differences in improvement between treatment conditions can be made, since all factors (including non-specific elements) except one are held constant, and the problems involved in other types of control groups are avoided. These designs, in general, should probably be used more often than they are. Some psychotherapies, however, may not easily lend themselves to such dismantling strategies. Moreover, it may be premature to attempt to dismantle a treatment package when questions about the efficacy of the whole package need to be resolved first.

Comparisons with medication

One type of design for the evaluation of psychotherapy that has increased in importance consists of comparisons to medication treatments. An example of this type of study is a recent investigation comparing cognitive therapy, medication (paroxetine), and pill placebo in the treatment of moderate to severe major depressive disorder.[12] The rationale for this design is that the medication group provides the standard reference condition with established efficacy. The pill-placebo group allows for establishing that the particular sample in the study showed the typical medication-placebo difference that would be expected (in other words, the sample was not unusual), and controls for some of the non-specific effects of the psychotherapy (for example, regular visits to a professional, positive expectancies for change). While the pill placebo is clearly not a perfect control condition for psychotherapy, it serves as a practical function—i.e. if a specific psychotherapy is not better than a pill placebo, should the psychotherapy be pursued as a treatment option?

Attribute by treatment interactions

An emerging emphasis in psychotherapy outcome research is the investigation of attribute by treatment interactions. Partly influenced by the common finding of no differences between psychotherapies, as well as the desire to make more specific clinical recommendations, investigators have hypothesized that certain matches of client characteristics and treatment modalities produce superior outcomes. One initial problem with pursuing potential matches of patient characteristics with the type of treatment is that there are a large number of potential variables (for instance, various combinations of diagnosis, therapist, treatment, patient's problem, setting).

The investigation of patient-treatment matches has intuitive clinical appeal, as most therapists believe that some patients seem to 'fit' a form of treatment better than others. Research on patient-treatment matching, however, is inherently difficult, particularly because large sample sizes are generally needed to adequately test interaction effects. Two of the largest randomized clinical trials ever performed with psychotherapeutic treatments have failed to find much support for specified patient-treatment interactions.[13,14] Thus, it remains to be seen whether aptitude by treatment-interaction designs will provide useful information about psychotherapy outcome.

Naturalistic studies

Naturalistic designs are used to examine issues such as the effectiveness of treatments in real-world setting with the types of patients that typically seek treatment. For example, the effectiveness of cognitive behavioural therapy for panic disorder and agoraphobia has been examined in a naturalistic study.[15] The authors report that outcomes were better for patients receiving naturalistic cognitive behaviour therapy, compared to a non-randomized wait-list control group. However, the overall outcomes for cognitive behavioural therapy were not as robust as previously seen in controlled clinical trials. While this form of naturalistic study has several advantages, including the fact that the data are drawn from actual clinical services in the real world, such data do not add to an understanding of which forms of psychotherapy work and which do not, and they do not provide strong causal statements. Hypotheses generated from naturalistic studies can inform the planning of experiments that can make stronger statements about causality. In addition, naturalistic studies provide useful descriptive information about the service delivery system.

Strategies for assessing psychotherapy outcome

Once a particular experimental design has been decided upon, the next crucial question in evaluating psychotherapy pertains to the selection of instruments for the study. Instruments include those needed to adequately select a patient population, such as measures of psychiatric diagnoses and initial symptomatology. Therapeutic change is often evaluated on a broad range of measures including dimensional measures of symptoms, personality, self-esteem, quality of life, and functioning in a variety of areas (for example, social and occupational functioning). In addition, the impact of treatment on specific theoretical constructs (mediators) can be examined. Discussion of each of these domains follows.

Patient selection

(a) Diagnosis

Because the identification of effective therapies for specific disorders is an important research priority from a public health point of view, an accurate and thorough assessment of these disorders is central to the evaluation process. The most widely used instrument for assessing the major DSM-IV Axis I diagnoses in research settings is the Structured Clinical Interview for DSM-IV.[16] The format of this semi-structured interview, with questions grouped by criteria and by diagnosis, allows an experienced clinical evaluator to assign diagnosis as the interview progresses. It is most commonly used to select subjects according to study inclusion or exclusion criteria or to characterize a study population. It can also be used to document change in diagnostic status post-treatment or over the course of a longitudinal study.

(b) Symptom measures

Whereas categorical measures are critical for the selection of a relatively homogeneous patient sample at intake, the evaluation of patient improvement requires the measurement of symptoms and functioning on a continuum. These continuous measures of the amount, timing, and nature of change are typically the primary outcomes of intervention studies. They include ratings of a single construct representing a core feature of the disorder, scales which cover the range of symptoms present in a general diagnostic category (e.g. rating scales of 'depression' or 'anxiety'), and measures that cut across many diagnoses and are indicative of overall psychopathology or symptomatology. At each of these levels of assessment, one can find both clinician-rated and self-report tools; for practical reasons, the self-report method predominates.

Measures of therapeutic change

(a) Dimensional assessment of core symptoms

In general, the more circumscribed and behavioural the problem, the simpler the assessment method. For some disorders, investigators have found that single-item measures of target symptoms suffice. For example, treatment studies of panic disorder and bulimia rely on self-reporting the number of 'episodes' that occur in a well-defined interval, such as the past week or the past month. Daily diaries are often used to facilitate recall and to enhance accuracy. Obsessive-compulsive symptoms can be assessed by Likert-type ratings of the severity of compulsions and obsessions, and the improvement of specific phobias by ratings of fear and avoidance. Severity ratings can be completed by both the patient and an independent evaluator.

For most disorders, core symptom measures of greater psychometric sophistication are needed to supplement simpler methods. The Penn State Worry Questionnaire,[17] for example, assesses the central feature of generalized anxiety disorder, and the Yale-Brown Obsessive-Compulsive Scale[18] measures the various symptoms that occur with obsessive compulsive disorder. In eating disorder research, the self-report EDI-2[19] and the Eating Disorder Examination[20] are commonly used.

In depression research, the Inventory of Depressive Symptomatology[21] is the result of efforts to develop a continuous measure of the nine DSM symptoms for major depressive episode. Other widely used scales include the interview-based Hamilton Rating Scale for Depression[22] and the self-report Beck Depression

Inventory,[23] although these instruments do not map directly on to DSM criteria. The severity of specific manic symptoms can be assessed with either the Bech-Rafaelsen Mania Scale[24] or the Young Mania Rating Scale.[25]

(b) General measures of anxiety, depression, and other symptoms

The Hamilton Rating Scale for Depression and the Beck Depression Inventory are often administered concurrently to provide clinician and patient perspectives on the level of depression, and are used both as primary outcome measures in studies of depression, and as secondary measures in treatment studies of other disorders. Because of the comorbidity of Axis I disorders in general, and the interrelatedness of anxiety and depression in particular, it is useful to include assessments of both depression and anxiety. General measures of anxiety include the Hamilton Anxiety Rating Scale[26] and the Beck Anxiety Inventory.[27] These general measures are used with a variety of diagnostic groups, but it is important to remember that they have varying relevance to different disorders within a given diagnostic group. For example, a score on the Beck Anxiety Inventory might be a good index of the severity of the generalized anxiety disorder, but be less informative about specific phobias or obsessive-compulsive behaviour.

(i) Psychotic symptoms

In psychological treatment studies of severe depression, mania, and psychotic disorders, it is often necessary to assess the level of psychotic symptomatology. The Brief Psychiatric Rating Scale[28] is recommended for this purpose. It has items covering five symptom clusters: thinking disturbance, anxious depression, withdrawal/retardation, hostile/suspiciousness, and agitation/excitement. As with other clinician-rated scales, it should be administered by an experienced clinician who has received standardized training on this instrument.

(ii) Substance use problems

Because of the frequent comorbidity of substance use problems with other Axis I disorders, it is important to evaluate the level of drug and alcohol use when screening patients for study enrollment or when characterizing a sample. The substance use modules of the DSM-IV Structured Clinical Interview are frequently used for this purpose. The Addiction Severity Index[29] has the advantage of providing more comprehensive and detailed information on problem areas associated with abuse, and yields scores that can be compared across patients, and from the beginning to the end of treatment. Because it is a rather time-consuming interview, it is recommended only in instances when the rates of substance use and related problems are expected to be high and their measurement is a research priority.

(c) Measure of global psychological functioning/ psychopathology

The rationale for using measures that cover a broad range of psychopathology is that they characterize the sample in terms of associated symptoms (i.e. symptoms other than those of primary interest) and provide a global measure of subjective distress. They have also proved to be quite useful in detecting treatment-related changes in evaluations of diverse psychotherapies. A popular self-report measure of general psychopathology is the SCL-90-R (and its abbreviated version, the Brief Symptom Inventory).[30] The SCL-90-R is a 90-item scale; the Brief Symptom Inventory is composed of 53 items. Both yield nine symptom dimensions and three global indices of distress. Because correlations between the two instruments are very high, it is recommended that the latter be used when time is an issue or when this measure is included as part of a larger core battery of assessment.

(d) Measures of self

Many psychotherapies explicitly attempt to improve self-esteem, self-concept, and self-confidence, and therefore it is relevant to examine the extent to which treatment successfully impacts on these domains. One of the oldest and most widely used measures of self-esteem is the Rosenberg Self-Esteem Scale,[31] an easily administered 10-item Likert-type scale yielding a uni-dimensional indicator of global self-esteem. More recent work in this area aims to distinguish other self-related constructs (such as self-concept) from self-esteem, develop more theoretically based multifactorial models, and improve upon the psychometrics of earlier measures. Some resulting scales, such as the Beck Self-Concept Test[32] or the Selves Questionnaire,[33] are appropriate for patients with a fairly wide range of psychiatric diagnoses.

(e) Personality assessment

Personality variables typically appear in evaluations of psychological treatment as either primary outcomes (as in the study of psychotherapy for personality pathology), or as prognostic indicators in studies of other Axis I disorders. They might also be included as part of a larger effort to thoroughly describe the patient sample. There are two ways to approach the evaluation of personality within these contexts: determination of the presence or absence of a DSM personality disorder, and dimensional ratings of personality features.

(i) Categorical: DSM-IV Axis II

Treatments that target specific personality disorders tend to rely on the first approach to assessment, namely the use of interviewer-administered instruments which assess criteria for the 10 specific personality disorders listed in the DSM-IV. The Structured Clinical Interview for DSM-IV Axis II[34] was designed for this purpose. A positive feature of this assessment is its efficiency in eliciting the information required to assign Axis II diagnoses, especially when used in conjunction with the self-report Personality Disorder Questionnaire.

(ii) Dimensional

Dimensional methods of assessing personality arise from either theoretical or factor-analytical models of the essential elements of personality structure. The Five Factor Model[35] proposes that personality, in both patient and non-patient samples, can be measured along five dimensions: neuroticism, extraversion, openness to experience, agreeable, and conscientiousness. Extremes of these traits define personality pathology. This model and several instruments used to generate scores on these factors, including the self-report NEO Personality Inventory and NEO-PI-R[36] and a semistructured interview (the Structured Interview of the Five-Factor Model of Personality[37]), have received considerable empirical support in both clinical and non-clinical samples.

Another dimensional measure of personality is the SWAP-200.[38] This instrument consists of a set of 200 personality-descriptive statements designed to be used by a clinician who knows a given patient well. The clinician arranges the 200 statements into eight categories, from those that are not descriptive to those that are highly descriptive of the patient. These scores are then used to generate both individualized, ideographic case descriptions as well as dimensional scores for the 10 personality disorders included in DSM–IV.

The categorical and dimensional methods are not entirely mutually exclusive and both are valuable in ongoing research into the phenomenology, aetiology, correlates, and treatment of personality disorder.

(f) Measures of functioning and quality of life

In recent years, the definition of mental health has been broadened to encompass its more positive, and somewhat paradoxically, its more negative aspects. Mental health is now viewed as more than the absence of illness, but the illness itself is increasingly seen as chronic and quite disabling. This has prompted investigators to turn their attention to outcomes other than symptom change—to evaluate fully the effectiveness of a psychotherapeutic intervention means to document its broad effect on a number of relevant domains including social/interpersonal functioning, work functioning, and quality of life.

(i) Overall functioning

Incorporated into the DSM system as a separate axis (Axis V), the Global Assessment of Functioning scale[39] (GAF) is the most widely used measure of psychosocial impairment. Clinicians rate on a scale of 1 to 100 the current level of functioning, or if desired, the highest or lowest level of functioning within some designated time period. General well being and functioning are also assessed with scales that have been developed for the routine outcome assessment in clinical practice, rather than for efficacy trials. An example is the OQ-45,[40] a self-report measure consisting of three subscales representing broad content areas: (1) symptom distress, (2) interpersonal relations, and (3) social role (dissatisfaction and distress in tasks related to work, family roles, and leisure life).

(ii) Social and occupational functioning

The assessment of role performance has a fairly long history in psychotherapy research. The Social Adjustment Scale (SAS)[41] was developed to document the level in functioning in six areas: work as worker, 'housewife', or student; social activities; relationship with family; relationship with spouse; parental responsibilities; member of a family unit. There is much documentation of its psychometric properties and clinical utility, and it has been used with a wide range of adult outpatients. A unique feature of the Longitudinal Interval Follow-up Evaluation (LIFE)[42] is that it was designed to collect information on psychosocial functioning (as well as on diagnosis and symptoms) over longer periods of time. In the hands of trained evaluators, it is a structured interview that has been shown to have good reliability.

The concern among both scientists and health-care managers about the cost-effectiveness of specific treatments has generated interest in more comprehensive assessments of work functioning and productivity. Endicott and Nee developed the Endicott Work Productivity Scale,[43] a self-report instrument containing 25 items designed to be sensitive to more subtle differences among patients in work attitudes and behaviour.

Two self-report instruments measuring interpersonal difficulties have been used in studies of psychotherapy: The Inventory of Interpersonal Problems[44] and the Dyadic Adjustment Scale.[45] The Inventory of Interpersonal Problems is a 127-item self-report measure with high internal consistency and test-retest reliability of each of the six subscales: assertive, social, intimate, submissive, responsible, and controlling. The Dyadic Adjustment Scale is a 32-item scale designed to assess the severity of relationship discord in married and unmarried cohabiting couples, with higher scores indicating better adjustment. Responses on the Dyadic Adjustment Scale discriminate between distressed and non-distressed couples and yield a total score based on the factors of dyadic consensus, dyadic cohesion, dyadic satisfaction, and affectional expression.

(iii) Quality of life

The most promising of the quality-of-life measures for mental health outcome research are those based on a broad definition of quality of life—covering role functioning and social-material conditions as well as life satisfaction or well being—and which can be applied across disorders and across treatments. However, more widespread use and testing of quality-of-life instruments is needed to establish their utility and to help resolve a number of important measurement issues, including the value of more 'objective' indices of quality of life and the relationship between symptoms and quality-of-life judgments.

(g) Utilization of treatment services

The impact of a treatment on usage of other health-related services is a factor to be considered in determining the cost-effectiveness of that treatment. One broad instrument designed to assess changes in service usage over the course of treatment is the Treatment Services Review.[46] This is a 5-minute interview that documents the number and types of treatment services received during rehabilitation from substance abuse, but it also can be adapted for use with other clinical populations.

(h) Theory-based measures

Evaluation of the hypothesized important psychological constructs of a particular psychotherapy can serve as outcome measures in their own right or as mediators of change in symptoms and functioning. For example, the cognitive model of depression holds that distorted cognitions about the self and world are responsible for generating and maintaining negative emotions. Measures of depressogenic cognitions are therefore included as outcomes and mediators of symptom change in studies of cognitive therapy for depression. The Hopelessness Scale[47] is a 20-item self-report scale that assesses the hopelessness and pessimism associated with suicidal ideation and intent. The Dysfunctional Attitudes Scale[48] is a 40-item index of general attitudes and beliefs hypothesized by Beck and colleagues to underlie a propensity for depressive thinking, whereas the Automatic Thoughts Questionnaire[49] covers 30 negative thoughts proposed to occur during a symptomatic depressed state.

In regard to psychodynamic psychotherapy, theory-specific mediators include measures of core conflicts[50] and self-understanding.[51]

Conclusions

Of all the treatments in medicine and psychiatry in particular, the evaluation of psychotherapy offers some of the greatest challenges to researchers. A wide variety of research designs and instruments have been employed, with distinct advantages and disadvantages to each. No one study of psychotherapy can answer all the important questions that need to be asked. Given the complexity of research on psychotherapy, knowledge accumulates slowly. Despite the problems, complexities, and slow pace of scientific advance,

psychotherapy outcome research now has the tools and emerging findings to begin to influence the practice of psychotherapy and mental health treatments in general. We expect that in the next decade methodological advances in the evaluation of psychotherapy will lead to a stronger link between research and practice.

Further informaion

Lambert, M. (2004). *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th edn). John Wiley and Sons, New York.

Society for Psychotherapy Research. http://www.psychotherapyresearch.org

Rush, A.J. (2007). *Handbook of psychiatric measures* (2nd edn). American Psychiatric Publishing, Washington, DC.

References

1. Eysenck, H.J. (1952). The effects of psychotherapy: an evaluation. *Journal of Consulting Psychology*, **16**, 319–24.
2. Smith, M.L., Glass, G.V., and Miller, T.I. (1980). *The benefits of psychotherapy*. Johns Hopkins University Press, Baltimore, MD.
3. Castonguay, L.G., and Beutler, L.E. (2006). (eds). *Principles of therapeutic change that work*. Oxford University Press, New York.
4. Nathan, P., and Gorman, J. (in press) (eds.). *Treatments that Work,* Oxford Unversity Press (3rd edn.), New York.
5. Kazdin, A. (1994). Methodology, design, and evaluation in psychotherapy research. In *Handbook of psychotherapy and behavior change* (ed. A.E. Bergin and S.L. Garfield), pp. 19–71. Wiley, New York.
6. Crits-Christoph, P. and Gallop, R. (2006). Therapist effects in the TDCRP and other psychotherapy studies. *Psychotherapy Research,* **16**, 178–81.
7. Crits–Christoph, P., Baramackie, K., Kurcias, J., *et al*. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research*, **1**, 81–91.
8. Kazdin, A. (1982). *Single–case research designs: methods for clinical and applied settings*. Oxford University Press, New York.
9. Blanchard, E.B., Hickling, E.J., Devineni, T., *et al*. (2003). A controlled evaluation of cognitive behavioral therapy for posttraumatic stress in motor vehicle accident survivors. *Behaviour Research and Therapy* **41**, 79–96.
10. McIntosh, V.V., Jordan, J., Carter, FA., *et al*. (2005). Three psychotherapies for anorexia nervosa: a randomized, controlled trial. *American Journal of Psychiatry,* **162**, 741–7.
11. Foa, E., Steketee, G., Grayson, J., *et al.* (1984). Deliberate exposure and blocking of obsessive–compulsive rituals: immediate and long-term effects. *Behavior Therapy*, **15**, 450–72.
12. DeRubeis, R.J., Hollon, S.D., Amsterdam, J.D., *et al. (2005)*. Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry,* **62**, 409–16.
13. Project MATCH Research Group (1997). Matching Alcoholism Treatments to Client Heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol*, **58**, 7–29.
14. Crits–Christoph, P., Sigueland, L., Blaine, J., *et al.* (1999). Psychosocial treatments for cocaine dependence: National Institute on Drug Abuse Collaborative Cocaine Treatment Study. *Archives of General Psychiatry*, **56**, 493–502.
15. Rosenberg, N.K., and Hougaard E. (2005). Cognitive behavioural group treatment of panic disorder and agoraphobia in a psychiatric setting: A naturalistic study of effectiveness. *Nordic Journal of Psychiatry.* **59**, 198–204.
16. First, M.B., Spitzer, R.L., Gibbon, M., *et al.* (1994). *Structured Clinical Interview for Axis–I DSM–IV Disorders. Patient edition (SCID-II, version 2.0)*. Biometrics Research Department, New York State Psychiatric Institute.
17. Meyer, T.J., Miller, M.L., Metzger, R.L., *et al.* (1990). Development and validation of the Penn State Worry Questionnaire. *Behavior Research and Therapy*, **28**, 487–95.
18. Goodman, W.M., Price, L.H., Rasmussen, C.A., *et al.* (1989). The Yale–Brown Obsessive–Compulsive Scale. I: development, use, and reliability. *Archives of General Psychiatry*, **46**, 1006–11.
19. Garner, D.M. (1991). *Eating Disorder Inventory–2 manual.* Psychological Assessment Resources, Odessa, FL.
20. Fairburn, C. G., and Cooper, Z. (1993).The Eating Disorder Examination (12th edn.). In *Binge eating: Nature, assessment, and treatment* (eds. C.G.Fairburn and G.T. Wilson) pp. 317–60. Guilford Press, New York.
21. Rush, A.J., Carmody, T.J., Ibrahim, H.M., *et al.* (2006). Comparison of self-report and clinician ratings on two inventories of depressive symptomatology. *Psychiatric Services*, **57**, 829–37.
22. Hamilton, M. (1960). A rating scale for depression. *Journal of Neurological and Neurosurgical Psychiatry*, **23**, 56–62.
23. Beck, A.T., Steer, R.A., and Brown, G.K. (1996). *BDI-II manual* (2nd edn). Psychological Corporation, San Antonio, TX.
24. Bech P. (2002). The Bech-Rafaelsen Mania Scale in clinical trials of therapies for bipolar disorder: a 20-year review of its use as an outcome measure. *CNS Drugs*, **16**, 47–63.
25. Young, R.C., Biggs, J.T., Ziegler, V.E., *et al.* (1978). A rating scale for mania: reliability, validity and sensitivity. *British Journal of Psychiatry*, **133**, 429–35.
26. Hamilton, M. (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology*, **32**, 50–5.
27. Steer, R.A., Beck, A.T. (1997). In *Evaluating stress: A book of resources* (eds. C.P. Zalaquett and R.J Wood), pp. 23–40. Scarecrow Education, Lanham, MD.
28. Thomas, A., Donnell, A.J., and Young, T.R. (2004). Factor structure and differential validity of the expanded Brief Psychiatric Rating Scale. *Assessment*, **11**, 177–87.
29. McLellan, A.T., Kushner, H., Metzger, D. *et al.* (1992). The fifth edition of the addiction severity index: historical critique and normative data, *Journal of Substance Abuse Treament*, **9,** 199–213.
30. Derogatis, L.R., and Fitzpatrick, M. (2004). The SCL-90-R, the Brief Symptom Inventory (BSI), and the BSI-18 In *The use of psychological testing for treatment planning and outcomes assessment: Volume 3: Instruments for adults* (3rd edn.) (ed. M.E. Maruish). (pp. 1–41). Lawrence Erlbaum Associates Publishers, Mahwah, NJ.
31. Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press.
32. Beck, A.T., Steer, R.A., Brown, G., *et al.* (1990). The Beck Self-Concept Test. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, **2**, 191–7.
33. Higgins, E.T., Bond, R.N., Klein, R., *et al.* (1986). Self-discrepancies and emotional vulnerability: how magnitude, accessibility, and type of discrepancy influence affect. *Journal of Personality and Social Psychology*, **51**, 5–15.
34. First, M.B., Spitzer, R.L., Gibbon, M., *et al.* (1994). *Structured Clinical Interview for DSM-IV Axis-II Personality Disorders (SCID-II, version 2.0)*. Biometrics Research Department, New York State Psychiatric Institute.
35. McCrae, R.R., and Costa, P.T. (2003). *Personality in adulthood: A five-factor theory perspective* (2nd edn.) Guilford Press, New York, NY.
36. McCrae, R.R., and Costa, P. T. (2005). The NEO-PI-3: A More Readable Revised NEO Personality Inventory. *Journal of Personality Assessment*, **84**, 261–70.
37. Trull, T.J., and Widiger, T.A. (2002). The structured interview for the five factor model of personality (SIFFM) In *Big five assessment* (eds. B. de Raad & M. Perugini). (pp. 148–66). Ashland, OH: Hogrefe & Huber Publishers.
38. Westen, D., and Muderrisoglu, S. (2006). Clinical assessment of pathological personality traits. *American Journal of Psychiatry*, **163**, 1285–7.

39. American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, DC.
40. Lambert., M.J., Gregersen A.T., and Burlingame, G.M. (2004). The Outcome Questionnaire-45. In *The use of psychological testing for treatment planning and outcomes assessment: Volume 3: Instruments for adults* (3rd edn) (ed. M.E. Maruish), pp. 191–234. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
41. Weissman, M.M., Klerman, G.L., Paykel, E.S., *et al.* (1974). Treatment effects in the social adjustment of depressed patients. *Archives of General Psychiatry*, **30**, 771–8.
42. Keller, M.B., Lavori, P.W., Friedman, B., *et al.* (1987). The longitudinal interval follow-up evaluation: a comprehensive method for assessing outcome in prospective longitudinal studies. *Archives of General Psychiatry*, **44**, 540–8.
43. Endicott, J. and Nee, J. (1997). Endicott Work Productivity Scale (EWPS): a new measure to assess treatment effects. *Psychopharmacology Bulletin*, **33**, 13–16.
44. Horowitz, L.M., Rosenberg, S.E., Baer, B.A, *et al.* (1988). Inventory of Interpersonal Problems: psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology*, **56**, 885–92.
45. Spanier, G.B. (1976). Measuring dyadic adjustment: new scales for assessing the quality of marriage and similar dyads. *Journal of Marital and Family Therapy*, **38**, 15–38.
46. McLellan, A.T., Alterman, A.I., Cacciola, J., *et al.* (1992). A new measure of substance abuse treatment. Initial studies of the treatment services review. *Journal of Nervous and Mental Disease*, **180**, 101–10.
47. Beck, A.T. and Steer, R.A. (1988). *Manual for Beck Hopelessness Scale*. Psychological Corporation, San Antonio, TX.
48. Weissman, A.N. (1979).The dysfunctional attitudes scale: a validation study. *Dissertation Abstracts International*, **40**, 1389–90.
49. Hollon, S.D. and Kendall, P.C. (1980). Cognitive self-statements in depression: development of an automatic thoughts questionnaire. *Cognitive Therapy and Research*, **4**, 88–100.
50. Luborsky, L. and Crits–Christoph, P. (1998). *Understanding transference: the core conflictual relationship theme method*. American Psychological Association, Washington, DC.
51. Connolly, M.B., Crits–Christoph, P., Shelton, R.C., *et al.* (1999). The reliability and validity of a measure of self–understanding of interpersonal patterns. *Journal of Counseling Psychology*, **46,** 472–82.

6.2

Somatic treatments

Contents

6.2.1 General principles of drug therapy in psychiatry

J. K. Aronson

The successful use of psychotropic drugs demands an understanding of their pharmaceutical, pharmacokinetic, and pharmacodynamic properties.

- *Pharmaceutical properties*: Pharmaceutical formulations can be manipulated to produce different durations of action, for example the use of oily emulsions of antipsychotic drugs in depot formulations.
- *Pharmacokinetic properties*: Pharmacokinetics is the mathematical description of the disposition of drugs in the body by absorption, distribution (to plasma proteins and tissues), and elimination (usually by hepatic metabolism and renal excretion). Differences in drug disposition determine differences in dosage regimens and are important for drug interactions.
- *Pharmacodynamic properties*: Pharmacodynamics is the study of the pharmacological actions of drugs and how actions at the molecular level are translated, via actions at cellular, tissue, and organ levels, into therapeutic or adverse effects. The known pharmacological actions of psychotropic drugs are not necessarily the actions that produce their therapeutic or adverse effects.

Dosage regimens

A drug dosage regimen is a recipe for drug administration, intended to produce the desired therapeutic effect with a minimum of unwanted effects. It is described in terms of the pharmaceutical formulation, the dose, and the frequency and route of administration used. The duration of administration is also important.

In treating any condition, it is best to learn initially to use a few drugs, preferably well-established ones, and to expand one's repertoire with increasing experience.

The choice of drug depends firstly on the indication—obviously an antidepressant will be the drug of choice for a patient with depression, if drug therapy is thought to be required. The choice of

antidepressant will depend on features of the disease and other factors. For example, some antidepressants are more sedative and anxiolytic than others, and can be helpful in patients who are agitated. The avoidance of adverse effects or interactions can also dictate the choice; for example, tricyclic antidepressants should be avoided in men with prostatic hyperplasia and selective serotonin reuptake inhibitors (SSRIs) should not be used in children, because of the increased risk of suicidal ideation.

It is usual to start therapy with published dosage recommendations, generally beginning at the lower end of the recommended dosage range and monitoring for a therapeutic effect. A common error is to give a starting dose of a drug and then to add or substitute another drug if the first does not work. This is usually bad practice. If the desired effect does not occur with the initial dosage, increase the dose gradually until the effect occurs or the upper limit of the recommended range is reached (although adverse effects may limit this process). Only then should another drug be tried. Sometimes a poor response is due to poor adherence to therapy; careful explanation of the condition and the need for therapy helps.

Psychotropic drugs can be given orally or parenterally, and as immediate-release or modified-release formulations. Most drug administration is oral, but parenteral therapy can be useful to guarantee administration (e.g. depot formulations in schizophrenia) and for a quicker onset of action (e.g. in the treatment of acute mania). Modified-release products are used for long-term therapy. They can be given less often and produce a smoother profile of blood concentrations (Fig. 6.2.1.1).[1] The advantage of intramuscular modified-release (depot) formulations of antipsychotic drugs is that drug delivery can be ensured by supervised infrequent administration (say every 2 weeks). Different modified-release formulations of the same compounds have different release characteristics and are not interchangeable; for example, when prescribing a modified-release oral formulation of lithium always give the patient the same formulation and specify the brand name on the prescription.

Combination formulations (e.g. a phenothiazine plus a tricyclic antidepressant in a single tablet) do not allow flexibility of prescribing and should generally be avoided. Important exceptions include combination analgesic formulations (e.g. co-codamol, which contains paracetamol plus codeine) and combinations of levodopa with a dopa decarboxylase inhibitor (benserazide or carbidopa).

Treatment of children

There are no uniform rules for determining dosage regimens in children. Pharmacokinetics and pharmacodynamics are different for some drugs but not others.

- Absorption is not greatly different from absorption in adults.
- The distribution of water-soluble drugs is different, but psychotropic drugs are lipid-soluble.
- Protein binding is reduced in neonates; phenytoin is affected.
- Hepatic oxidative metabolism and glucuronide conjugation are deficient in neonates, and mature at variable rates; this is important for psychotropic drugs.
- Glomerular and renal tubular functions are immature in neonates and take about 6 months to reach adult values.

If a child needs a psychotropic drug, consult the manufacturer's literature and always start with a low dosage.

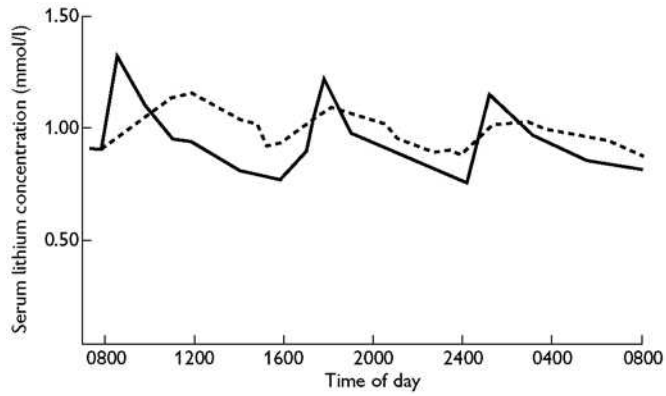


Fig. 6.2.1.1 Administration of lithium in immediate-release and modified-release oral formulations. The immediate-release formulation (solid line) produces rapid peaks of serum concentration and large fluctuations during a dosage interval. In contrast, the modified-release formulation (dotted line) is more slowly absorbed but produces much less fluctuation in serum concentrations. Note also that the apparent half-life of lithium is longer after administration of the modified-release formulation; this is not the true half-life of lithium, but the half-life of its release from the modified-release formulation. (Adapted from A. Amdisen, Variation of serum lithium concentration during the day in relation to treatment control, absorptive side effects and the use of slow-release tablets, *Acta Psychiatrica Scandinavica*, **207**, 55–7, copyright 1969, John Wiley & Sons, Inc.)

Treatment of elderly people

Pharmacokinetic differences in old age are more predictable than in children, but pharmacodynamic changes are variable.

- Absorption is not greatly affected.
- Elderly people have less body fat, and so lipid-soluble drugs may be more highly concentrated in the brain; however, this effect varies unpredictably from drug to drug (e.g. the apparent volume of distribution of diazepam is increased while that of nitrazepam is not).
- Protein binding is reduced in elderly people; phenytoin is affected.
- Hepatic metabolism is reduced in frail but not in fit old people; this effect is proportional to liver size.
- Renal function is impaired with age; use creatinine clearance, measured or estimated (not eGFR), as a guide.
- Inappropriate polypharmacy is common in old people, increasing the risk of drug interactions.

When treating an elderly person with a psychotropic drug always start with a low dosage and increase dosages more slowly.

Pregnancy and breast feeding

Anticonvulsants are teratogenic.[2] For example, sodium valproate has been associated with spina bifida, cardiac malformations, hypospadias, anomalies of the brain and face, coarctation of the aorta, and limb reduction defects.[3]

Few other psychoactive drugs are teratogenic. However, most of them cross the placenta and some can cause withdrawal symptoms in the neonate. The teratogenicity of lithium has been overstated in the past; the main risk is cardiovascular teratogenicity, but although the risk of Ebstein's anomaly is increased, the absolute risk (0.05–0.1 per cent) is still very small;[4] nevertheless, some

advise that it should be avoided or used with caution in the first trimester of pregnancy,[5] and fetal sonography is recommended at 18–20 weeks after first-trimester exposure.[3]

Although most psychoactive drugs are lipid-soluble and therefore enter the breast milk, few do so in high enough amounts to trouble the neonate; if a neonate becomes drowsy while breast feeding, reduce the mother's dosage or stop breast feeding. Lithium appears in the breast milk and can be found in the serum of breast-fed babies in variable concentrations, up to half of those in the mother. Because neonates have immature renal function, some recommend avoiding breast feeding.[6] However, others consider that the benefits of breast feeding to mother and child outweigh the small risk of lithium toxicity.[4] The following advice has been given:[3]

- Educate the mother about the manifestations of toxicity.
- Explain the risks of dehydration.
- Consider partial or total formula supplements during episodes of illness or dehydration.
- Suspend breast feeding if toxicity is suspected.
- Check infant and maternal serum concentrations.

Pharmacokinetics—drug disposition

Most psychotropic drugs are rapidly and well absorbed after oral administration. However, drugs can be removed by various processes before they reach the systemic circulation. The fraction of drug that reaches the systemic circulation is called its systemic availability (or, more commonly, bioavailability).

After oral administration a formulation will generally disintegrate in the stomach and the drug it contains will dissolve in gastric contents. However, drugs are not generally absorbed in the stomach. After gastric emptying they are for the most part absorbed in the jejunum and ileum, and some are absorbed from the colon as well. During transit across the gut wall they may be metabolized by an oxidative isozyme of cytochrome P450, CYP3A4, and can be secreted back into the gut lumen by P glycoprotein. When they enter the portal circulation they may be eliminated by the liver. If hepatic metabolism is extensive, a large amount of drug will be removed during this first passage through the liver. For example, clomethiazole has extensive first-pass metabolism in the liver and its systemic availability is low (about 40 per cent); thus, intravenous doses are considerably lower than oral doses. In severe liver disease, such as cirrhosis, or when there is arteriovenous shunting, this presystemic metabolism is reduced and the systemic availability increases up to 90 per cent; oral doses of clomethiazole should be reduced in liver disease.[7]

In the systemic circulation drugs are bound to plasma proteins and distributed to the tissues. Protein binding is important for drugs that are highly bound (over 90 per cent) and not widely distributed to the body tissues; in those cases protein-binding displacement can result in a large rise in the amount of unbound drug available to the target tissue. This is important for phenytoin, which is 90 per cent bound to plasma albumin and has a low volume of distribution. The binding of phenytoin is reduced when the serum albumin concentration falls (in chronic liver disease, the nephrotic syndrome, protein malnutrition, or the third trimester of pregnancy), when binding to the protein is abnormal (in chronic renal insufficiency), or when another drug (e.g. sodium valproate) causes displacement. Acute displacement causes phenytoin toxicity, but only transiently, because in the case of phenytoin an increase in unbound concentration causes it to be more rapidly eliminated. When measuring plasma phenytoin concentrations in patients in whom protein binding is reduced, the target concentration (and the laboratory will measure total drug, i.e. bound plus unbound) is reduced (see Fig. 6.2.1.2).

In chronic renal insufficiency the protein binding of phenytoin is reduced. This leads to an increase in the unbound plasma (or serum) concentration relative to the total concentration; the target total concentration therefore falls. The shaded area shows the range of plasma phenytoin concentrations that one would generally aim to achieve (the target concentration range) in treating a patient with epilepsy. As renal function deteriorates (indicated here by an increase in serum creatinine concentration), the target range for plasma phenytoin concentration falls from 40–80 μmol/l when renal function is normal to 10–30 μmol/l in severe renal insufficiency.

After absorption and distribution most psychoactive drugs are cleared from the body by hepatic metabolism;[8] impaired liver function, if severe (for the liver has a large capacity), reduces their elimination, and dosages should be reduced.

Lithium is cleared solely by renal elimination and therefore the dosage should be reduced in proportion to the creatinine clearance. Since renal function falls with age, lithium dosages should be lower in older people.[9]

The half-life of a drug is a function of its clearance and its distribution volume: the slower the rate of clearance or the more extensive the distribution the longer the half-life. If a drug is given in a regular maintenance dose, the amount of drug in the body will gradually accumulate; however, as the amount in the body increases, the rate at which it is eliminated also rises, and eventually a plateau (or steady state) is reached when the amount eliminated during a dosage interval equals the dose (the maintenance dose). The time it takes to reach this steady state depends on the half-life

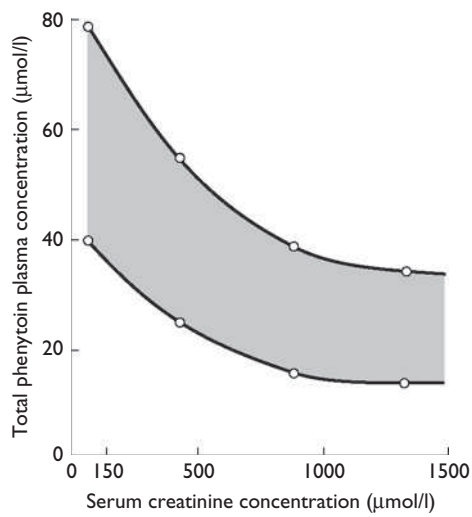


Fig. 6.2.1.2 In chronic renal insufficiency the protein binding of phenytoin is reduced. This leads to an increase in the unbound plasma (or serum) concentration relative to the total concentration; the target total concentration therefore falls. The shaded area shows the range of plasma phenytoin concentrations that one would generally aim to achieve (the target concentration range) in treating a patient with epilepsy. As renal function deteriorates (indicated here by an increase in serum creatinine concentration) the target range for plasma phenytoin concentration falls, from 40–80 μmol/l when renal function is normal to 10–30 μmol/l in severe renal insufficiency (I. Odar-Cederlöf, unpublished data.)

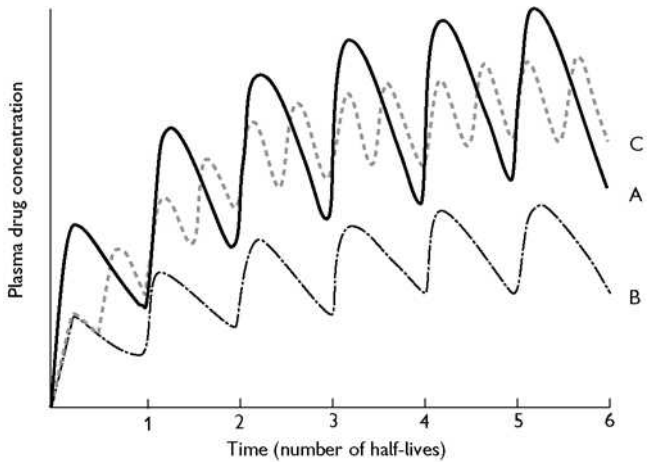


Fig. 6.2.1.3 Curve A—during the regular administration of a maintenance dose of a drug the amount of drug in the body rises after a dose, reaches a peak, and then falls as the drug is distributed to the tissues and eliminated. If another dose is given soon after the first, the plasma concentration will rise by the same amount as before but will fall faster after peaking, since most drugs obey first-order kinetics and the plasma concentration falls exponentially. Thus, when a drug is given repeatedly the mean plasma concentration rises more slowly with each successive dose, until eventually a steady state is reached, when the amount eliminated in a dosage interval is equal to the dose itself. This takes about four half-lives of the drug. Curve B represents the concentrations during administration of half the dose given at the same frequency. The time taken to reach steady state is the same in both cases, but the eventual steady-state concentration in case B is half that in case A, being proportional to the dose. Curve C represents the concentrations during administration of half the dose given twice as often (i.e. the total dose is unchanged). Neither the time taken to reach steady state nor the eventual mean steady-state concentration is affected. However, the fluctuations in plasma concentration during a dosage interval are reduced (cf. Fig. 6.2.1.1). (Adapted from Amdisen, A. Variation of serum lithium concentration during the day in relation to treatment control, absorptive side effects and the use of slow-release tablets, *Acta Psychiatrica Scandinavica*, **207**, 55–57. Copyright 1969, John Wiley & Sons, Inc.)

of the drug; about 94 per cent of the steady-state value will be reached after four half-lives (Fig. 6.2.1.3, curve A). For example, lithium has a half-life of about 24 h; after 4 days of maintenance therapy with the same regular dose a steady state will be reached; this does not depend on the dose or frequency of administration (Fig. 6.2.1.3, curves B and C). If a modified-release formulation is used and the half-life of absorption of the drug from the formulation is longer than the drug's own half-life, the longer (apparent) half-life will determine the time to steady state; for example, the apparent half-life of flupentixol after the administration of flupentixol decanoate is 17 days, compared with 36 h for flupentixol after oral administration. When using depot antipsychotic drugs, which have long half-lives of absorption, steady-state therapy should first be established with an ordinary formulation.

Curve A shows that during the regular administration of a maintenance dose of a drug the amount of drug in the body rises after a dose, reaches a peak, and then falls as the drug is distributed to the tissues and eliminated. If another dose is given soon after the first, the plasma concentration will rise by the same amount as before but will fall faster after peaking, since most drugs obey first-order kinetics and the plasma concentration falls exponentially. Thus, when a drug is given repeatedly the mean plasma concentration rises more slowly with each successive dose, until eventually a steady state is reached, when the amount eliminated in a dosage interval is equal to the dose itself. This takes about four half-lives of the drug. Curve B represents the concentrations during administration of half the dose given at the same frequency. The time taken to reach steady state is the same in both cases, but the eventual steady-state concentration in case B is half that in case A, being proportional to the dose. Curve C represents the concentrations during administration of half the dose given twice as often (i.e. the total dose is unchanged). Neither the time taken to reach steady state nor the eventual mean steady-state concentration is affected. However, the fluctuations in plasma concentration during a dosage interval are reduced (cf. Fig. 6.2.1.1). Kinetic characteristics of some psychotropic drugs are shown in Table 6.2.1.1.

Pharmacological actions of drugs

Psychotropic drugs interfere with neurotransmitter functions in several ways—via actions on neurotransmitter receptors, storage, release, reuptake, and metabolism. Transmembrane neurotransmitter receptors are broadly speaking of two types—ionotropic and metabotropic receptors. Ionotropic receptors (e.g. nicotinic acetylcholine, glycine, GABA, and NMDA, AMPA, and kainate receptors) incorporate ion channels in their structures and mediate rapid responses. Metabotropic receptors (e.g. G protein-coupled receptors such as adrenaline, noradrenaline, cannabinoid, dopamine, opioid, and serotonin receptors other than $5HT_3$ receptors) produce their effects via signal transduction systems, which activate second messengers or ion channels, and produce longer lasting responses.

Agonist action at a receptor

Agonists are substances that act by stimulating the action of a receptor.

Benzodiazepines bind to benzodiazepine receptors in the spinal cord, brainstem, cerebellum, limbic system, and cerebral cortex. These receptors are associated with receptors for the inhibitory neurotransmitter γ-aminobutyric acid (GABA), linked to a chloride channel.[10] The benzodiazepines enhance the action of GABA through its chloride channel, the presumed mechanism whereby they are anxiolytic and hypnotic. Some other hypnotics that have non-benzodiazepine structures also act via benzodiazepine receptors: zopiclone binds to the GABA–benzodiazepine receptor complex, but at a site different from that of benzodiazepines;[11] clomethiazole binds to a binding site distinct from those of benzodiazepines and barbiturates;[12] zolpidem binds to a subtype of binding site called BZ_1, found on GABA neurones in the sensorimotor cortex and extrapyramidal tracts.[10]

The triptans (such as sumatriptan, naratriptan, zolmitriptan), which are used to treat migraine, are agonists at 5-hydroxytryptamine ($5\text{-}HT_{1B/D}$) receptors, causing vasoconstriction. They are therefore contraindicated in patients with cardiovascular disease and in those with hemiplegic or basilar migraine because of the fear of stroke.[13]

Antagonist action at a receptor

Antagonists are substances that have no actions of their own at receptors and act by preventing the action of an agonist, usually an endogenous one.

The antipsychotic (neuroleptic) drugs are all antagonists at receptors for the endogenous neurotransmitter dopamine; this is thought to be the basis of their antipsychotic actions in the mesolimbic system (via D_1 and D_4 receptors) and undoubtedly produces their

Table 6.2.1.1 Pharmacokinetic information about some psychotropic drugs

Drug	Systemic availability (%)	Half-life (h)	Route of hepatic elimination[a]
Benzodiazepines			
Alprazolam	90	12	CYP3A
Chlordiazepoxide[b]	95	15	
Clobazam[b]	90	20	
Diazepam[b]	100	30	CYP2E1/2C19
Flurazepam[b]	30	3	CYP2E1
Lorazepam	90	15	
Nitrazepam	75	24	
Oxazepam	100	9	
Temazepam	95	10	
Antidepressonts			
Lithium	100	24	(Renally excreted)
Tricyclics			
Amitriptyline	50	20	CYP1A2/2D6
Clomipramine	50	20	CYP1A2/2D6/2C19
Desipramine	40	24	CYP2D6
Imipramine[b]	50	14	CYP1A2/2D6
Nortriptyline	60	36	CYP2D6
Tetracyclics			
Maprotiline	95	40	
Mianserin	25	16	CYP2D6
Triazolopyridines			
Trazodone	100	10	
Selective serotonin reuptake inhibitors			
Fluoxetine	95	48	CYP2D6/3A4
Fluvoxamine	90	20	CYP1A2/2C19/3A4
Paroxetine	Variable	24	CYP2D6
Sertraline	Low	24	CYP2D6
Monoamine oxidase inhibitors			
Phenelzine	High	1	Polymorphically acetylated
Tranylcypromine	Moderate	2	
Moclobemide	40	2	CYP2C19
Antipsychotic drugs			
Phenothiazines			
Chlorpromazine[b]	10	12	
Thioridazine	60	10	CYP2D6
Trifluoperazine	Low	14	
Butyrophenones			
Haloperidol	60	20	CYP2D6
Droperidol	75	2	
Thioxanthenes			
Flupentixol	40	36	
Zuclopenthixol	50	20	CYP2D6
Others			
Clozapine	50	12	
Risperidone	75	3/20[c]	CYP2D6

[a] CYP refers to isozymes of the cytochrome P-450 family of enzymes.

[b] All these drugs are partly metabolized to active metabolites, Some of the metabolites have long half-lives (e.g. diazepam is metabolized to desmethyldiazepam). Some benzodiazepines (e.g. clorazepate and prazepam) are completely metabolized to active metabolites with long half-lives.

[c] Extensive and poor metabolizers respectively.

adverse effects in the extrapyramidal tracts (via D_2 receptors). The so-called atypical antipsychotic drugs (including clozapine and risperidone) have little effect on D_2 receptors and less commonly cause extrapyramidal adverse effects.[14]

Flumazenil is a competitive antagonist of benzodiazepines at benzodiazepine receptors and is used to reverse their effects.[15]

Partial agonist action at a receptor

Partial agonists are substances that can be agonists or antagonists at receptors, depending on the endogenous tone of the system upon which they are acting. If the degree of endogenous stimulation of the receptor is low, a partial agonist will tend to act as an agonist; if high, it will end to act as an antagonist.

The anxiolytic buspirone is a partial agonist at 5-HT_{1A} autoreceptors and reduces the firing of 5-hydroxytryptamine (5-HT) neurones by stimulating the auto-receptors.[16]

Actions via second messengers and ion channels

Some drugs act directly on second messenger systems and ion channels, without actions at receptors.

Lithium inhibits enzymes involved in the metabolism of inositol phosphates and may deplete cells of phosphoinositides, which are important as second messengers in neurotransmission.[17] However, other mechanisms have been proposed, including effects on the synthesis, turnover, and functional activity of brain 5-HT and effects on neuronal membrane function by effects on sodium and potassium fluxes via the sodium/potassium pump enzyme.[18]

Altered neurotransmitter storage

Reserpine, which causes depression, inhibits the incorporation of neurotransmitters into presynaptic storage vesicles and thus causes depletion of neurotransmitter stores.

Increased neurotransmitter release

Amphetamines cause increased release of noradrenaline (norepinephrine) and dopamine and have mood-enhancing effects.[19]

Inhibition of neurotransmitter reuptake

Most antidepressants inhibit the reuptake of monoamines into the presynaptic nerve ending after their release.[20] The effects of different antidepressants on monoamine reuptake are listed in Table 6.2.1.2. Reduced reuptake of monoamines occurs immediately, but the full therapeutic effects of antidepressants take some weeks to occur. This is explained by the occurrence of adaptive changes in presynaptic and postsynaptic receptors, including down-regulation of β-adrenoceptors, reduced sensitivity of β- and α_2-adrenoceptors, and increased sensitivity of α_1-adrenoceptors and 5-HT receptors. However, it is not known how these actions are translated into the therapeutic effect. The last of these effects has led to the use of 5-HT autoreceptor antagonists, in the hope of producing a quicker onset of antidepressant action by enhancing 5-HT neurotransmission.[21,22] For example, the partial β-adrenoceptor agonist and 5-HT_{1A} receptor antagonist pindolol hastens the response to SSRIs, although it does not affect the extent of the response.[23] Another strategy involves drugs with several actions, which inhibit more than one reuptake system and are antagonists at neurotransmitter receptors.[24]

Table 6.2.1.2 Effects of antidepressants on monoamine reuptake

Drug	Inhibition of uptake		
	Noradrenaline	5-HT	Dopamine
Tricyclics			
Amitriptyline	+++	++	—
Clomipramine	+++	+++	—
Desipramine	+++	++	—
Imipramine[a]	++[a]	+	—
Nortriptyline	++	+	—
Tetracyclics			
Maprotiline	++	—	—
Mianserin	++	—	—
Phenylethylamines			
Venlafaxine	++	++++	—
Triazolopyridines			
Trazodone	—	+	—
Specific serotonin reuptake inhibitors			
Fluoxetine	—	+++	—
Fluvoxamine	+	++	—
Paroxetine	—	+++	—
Sertraline	+	++++	+

[a] Through its metabolite desipramine.

Altered neurotransmitter metabolism

Inhibitors of monoamine oxidase (MAO) enhance monoamine neurotransmission by irreversibly and non-selectively inhibiting the breakdown of monoamines after their release. Moclobemide[25] is a RIMA, a reversible inhibitor of MAO type A, which metabolizes 5-HT and noradrenaline. Moclobemide is therefore less likely to cause hypertension when taken in combination with amine-containing foods, such as cheese, since this reaction requires inhibition of both MAO type A and MAO type B.

Valproate partly acts by inhibiting GABA transaminase, thus enhancing GABA inhibitory transmission. However, it has other actions: it inhibits GABA reuptake, increases the sensitivity of GABA receptors to GABA, reduces the concentrations of the excitatory neurotransmitter aspartate, and may open potassium channels, thus stabilizing neuronal cell membranes.

Adverse effects of drugs

Unwanted effects of drugs are commonly referred to as toxic effects or side effects. However, these are ambiguous terms, for several reasons:

- Toxic effects occur through exaggeration of the desired pharmacological action of the drug, and therefore occur at doses that are above those usually associated with a therapeutic effect. For example, antipsychotic drugs produce some toxic effects by antagonism at dopamine receptors in the extrapyramidal tracts.
- Toxic effects can also occur through exaggeration of actions other than those that are thought to produce the therapeutic action. Paracetamol toxicity occurs because an active metabolite binds covalently to liver proteins, damaging them.
- Side effects occur either through actions that are unrelated to the desired pharmacological effect or through actions that are related to the desired pharmacological effect but occur in another tissue. Tricyclic antidepressants cause dry mouth, glaucoma, and urinary retention by anticholinergic action. Sildenafil causes colour vision disturbances by inhibiting phosphodiesterase type V in the eye, the action by which it has its therapeutic effect in erectile dysfunction. True side effects are more properly called collateral effects.[26]

It is therefore better to use the terms 'unwanted effects' or 'adverse effects/reactions'. Adverse reactions and adverse effects are identical—the former are seen from the point of view of the patient, the latter from the point of view of the drug.

Adverse drug effects are classified according to the scheme known as DoTS (Dose, Time, and Susceptibility).[26]

Classification according to dose-relatedness

1 *Hypersusceptibility reactions*: Here the dose–response curve for harm is far to the left of the dose–response curve for benefit; hypersusceptibility adverse reactions therefore occur at doses below those that are normally beneficial. Penicillin allergy is an example.

2 *Collateral reactions*: Here the dose–response curve for harm is in a region that is bounded by a curve that is just to the left of the dose–response curve for benefit and one that is just to the right; collateral adverse reactions therefore occur at doses within the range of those that are normally beneficial. They can occur (i) through a pharmacological effect that is distinct from that involved in the beneficial effect (for example an anticholinergic effect of a tricyclic antidepressant) or (ii) through the same pharmacological effect as that associated with the beneficial effect, but in a different tissue (for example colour vision disturbance due to sildenafil).

3 *Toxic reactions*: Here the harm occurs through the same mechanism as benefit (i.e. is on the same dose–response curve) but at doses that are above those that are normally beneficial. An example is serotonin syndrome due to fluoxetine.

Classification according to time-relatedness

Adverse drug reactions can be either time-dependent or time-independent (Table 6.2.1.3).

1 Time-independent reactions can occur at any time during therapy and are generally toxic reactions. They occur when the actual concentration of the drug increases or dose response curve shifts to the left, for whatever reason. An example is digoxin toxicity, which can occur for pharmaceutical reasons (for example administration of the wrong tablets), pharmacokinetic reasons (for example renal insufficiency), or pharmacodynamic reasons (for example hypokalemia). In the first two cases the concentration at the site of action increases and in the last the dose–response curve is shifted to the left.

2 Time-dependent adverse drug reactions are of six types; examples are given in Table 6.2.1.3.
   - Immediate or rapid reactions occur when a drug is given too quickly.
   - First-dose reactions occur only after the first dose of a course.
   - Early reactions occur soon after the first administration; they either wear off with time (early tolerant effects) or persist (early persistent effects).

Table 6.2.1.3 Time-related classification of adverse drug reactions in the DoTS method. (Reproduced from *British Medical Journal*, Aronson, J.K. and Ferner, R.E. (2005), **327**, 1222–5, with permission from BMJ Publishing Group Ltd.)

Type of reaction	Examples	Implications
Time independent		
Due to a change in dose or concentration (pharmaceutical effects)	Toxicity due to increased systemic availability	Beware of changing formulations of some drugs (e.g. modified-release formulations of lithium)
Due to a change in dose or concentration (pharmacokinetic effects)	Lithium toxicity due to renal insufficiency	Forewarn the patient; monitor carefully throughout treatment; alter dosage when pharmacokinetics change (e.g. renal insufficiency); avoid interacting drugs
Occurs without a change in dose (pharmacodynamic effects)	Digitalis toxicity due to hypokalaemia	Forewarn the patient; monitor carefully throughout treatment; avoid precipitating (pharmacodynamic) factors; avoid interacting drugs
Time dependent		
Immediate (due to rapid administration)	Red man syndrome (vancomycin) Hypertension (digitalis) Hypotension (iodipamide)	Administer slowly
First dose [of a course]	Hypotension (Ð1 adrenoceptor antagonists and angiotensin converting enzyme inhibitors) Type I hypersensitivity reactions	Take special precautions for the first dose Careful history taking; if a reaction occurs, avoid re-exposure; counsel the patient
Early tolerant	Adverse reactions that involve tolerance (e.g. nitrate-induced headache)	Monitor during the early stages; give appropriate reassurance; expect adverse effects if strategies to avoid tolerance are adopted
Early persistent	Glucocorticoid-induced diabetes mellitus	Monitor during the early stages and treat appropriately or withhold
Intermediate (risk increases at first, then diminishes)	Venous thromboembolism (classical antipsychotic drugs) Neutropenia (clozapine) Hypersensitivity reactions types II, III, and IV	Monitoring not needed after the high-risk period unless susceptibility changes; withdraw drug if a reaction develops
Late (risk increases with time)	Osteoporosis (glucocorticosteroids) Tardive dyskinesia (dopamine receptor antagonists) Retinopathy (chloroquine) Tissue phospholipid deposition (amiodarone)	Assess baseline function; forewarn the patient; monitor periodically during prolonged treatment
	Withdrawal syndromes: opiates, benzodiazepines, hypertension (clonidine and methyldopa), myocardial infarction (beta-blockers)	Withdraw slowly; forewarn the patient; replace with a longer acting drug if withdrawal is not possible
Delayed	Carcinogenesis (ciclosporin, diethylstilbestrol) Teratogenesis (thalidomide)	Avoid or screen; counsel or forewarn the patient

- Intermediate reactions occur within the first few weeks or months of administration but not thereafter; those who are susceptible will suffer the reaction and those who are not will not (healthy survivors); for example, clozapine causes neutropenia predominantly during the first 24 weeks of therapy—thereafter the risk is small.
- Late reactions occur late in the course of administration, the risk increasing with time; this group includes withdrawal reactions.
- Delayed reactions are seen at some distant time after the initial exposure, even if the drug is withdrawn before the reaction appears.

Classification according to susceptibility factors

The risk of an adverse drug reaction differs among members of an exposed population. For some reactions some individuals are susceptible, others are not—for example, prolonged muscle relaxation due to suxamethonium in people with pseudocholinesterase deficiency. In other cases susceptibility follows a continuous distribution—for example, increasing susceptibility with increasing impairment of renal function. Although reasons for increased susceptibility may be unknown, several types are recognized (Table 6.2.1.4).[27–34] These include:

- genetic variation;
- age;
- sex;
- physiological variation (for example pregnancy, body weight);
- exogenous factors (for example drugs and food);
- diseases (for example renal or hepatic impairment).

More than one susceptibility factor can be present in an individual.

Table 6.2.1.4 Sources of susceptibility to adverse drug reactions

Source of susceptibility*	Examples	Implications
Genetic	Porphyria Suxamethonium sensitivity Malignant hyperthermia CYP isozyme polymorphisms	Screen for abnormalities; avoid specific drugs
Age	Neonates (chloramphenicol[27]) Elderly people (hypnotics[28])	Adjust dosages according to age
Sex	Alcohol intoxication Mefloquine, neuropsychiatric effects[29] Lupus-like syndrome[30]	Use different doses in men and women
Physiology altered	Phenytoin in pregnancy[31]	Alter dosage or avoid
Exogenous factors	Drug interactions Interactions with food (e.g. grapefruit juice with drugs cleared by CYP3A4[32]; see Table 6.1)	Alter dosage or avoid co-administration
Diseases	Renal insufficiency (e.g. lithium[33]) Hepatic cirrhosis (e.g. morphine[34])	Screen for abnormalities; avoid specific drugs; use reduced dosages

*Mnemonic GASPED

Factors that increase the risk of an adverse effect

(a) Pharmaceutical factors

Adverse effects can arise from changes in the pharmaceutical formulation. There is a risk of lithium toxicity or loss of action when one modified-release formulation of lithium is replaced by another.

(b) Pharmacokinetic factors

Changes in the pharmacokinetics of a drug can result in toxic or collateral adverse effects. This most commonly occurs through impaired liver function (Table 6.2.1.1) or renal insufficiency (lithium).

(c) Pharmacodynamic factors

Changes in the sensitivity of a tissue to a drug occur during long-term therapy and can result in adverse effects. Tardive dyskinesia with dopamine receptor antagonists may be related to altered sensitivity of dopamine receptors,[35] although there are problems with this hypothesis, and complex interactions with other neurotransmitters may be involved.[36]

When adaptive changes occur during long-term therapy, sudden withdrawal of the drug can result in rebound reactions. Examples include the typical syndromes that occur after the sudden withdrawal of narcotic analgesics[37,38] or of alcohol (delirium tremens).[39] Sudden withdrawal of barbiturates can cause restlessness, sleeplessness, mental confusion, and convulsions; a similar syndrome, in which anxiety features prominently, can occur after the sudden withdrawal of benzodiazepines.[24,40]

Drug interactions

In a drug interaction one drug alters the effects of another, resulting in increased or decreased effects. Fluvoxamine inhibits the hepatic metabolism of warfarin and increases its anticoagulant effect,[41] whereas carbamazepine reduces the anticoagulant effect of warfarin by increasing its metabolism.[42]

1 Pharmaceutical interactions occur when there is a physicochemical interaction between two compounds in solution. Lists of such incompatibilities are too long to remember. To avoid pharmaceutical interactions, do not combine drugs in an infusion solution and use only sodium chloride 0.9% or glucose 5% in drug infusions.

2 Pharmacokinetic interactions occur when one drug interferes with the disposition of another during absorption, distribution, or elimination. Here are examples that are relevant to psychotropic drugs.

- Sucralfate reduces the absorption of amitriptyline by about 50 per cent;[43] this might be of clinical importance, but drug absorption interactions are not usually important.
- Phenytoin is displaced from protein-binding sites by valproate (which also inhibits its metabolism).[44]
- The metabolism of psychotropic drugs can theoretically be inhibited by drugs that inhibit metabolism via cytochrome P450 in the liver (Table 6.2.1.1). Reports of such interactions appear frequently. Inhibitory drugs include cimetidine, antifungal imidazoles (such as fluconazole), and macrolides (such as erythromycin).
- Non-selective MAO inhibitors, such as tranylcypromine, inhibit the metabolism (by MAO type A) of dietary amines in the gut and metabolism (by MAO type B) of the noradrenaline that they release, resulting in hypertension. Avoid this combination.
- Diuretics inhibit the renal excretion of lithium; alter the dosage of lithium and monitor the serum concentration when starting a diuretic or changing the diuretic dosage.

3 Pharmacodynamic interactions occur when two drugs interact at the same site of action. Alcohol potentiates the actions of all psychotropic drugs; patients taking psychotropic drugs should be warned that even one alcoholic drink can impair their ability to drive or operate machinery. The combination of reuptake inhibitors with non-selective irreversible monoamine oxidase inhibitors can cause the serotonin syndrome,[45] which is

sometimes fatal.[46] The combination of lithium with either SSRIs[47] or the SNRI venlafaxine[48] can also cause the serotonin syndrome. Flumazenil reverses the effects of benzodiazepines—a beneficial pharmacodynamic interaction.

Monitoring drug therapy

If possible, monitor drug therapy by observing the clinical outcome. In psychiatric disorders this is difficult, but it can be done by asking patients or their carers to keep diaries of symptoms.

Next best is to monitor some pharmacological effect of the drug, in the way that one measures the international normalized ratio (INR) in patients taking warfarin (thus measuring the effect of the drug on the blood, being unable to monitor its clinical effect of preventing pulmonary embolism). However, there are no comparable routine tests available for monitoring the pharmacological effects of psychotropic drugs.

Because of these difficulties one falls back on measurements of serum concentrations of some drugs. The assumptions in doing this are that the serum concentration reflects the concentration at the site of action and that there is a concentration–effect relationship. Few psychoactive drugs can be monitored in this way, the principal ones being lithium and phenytoin. Some advocate monitoring treatment with some tricyclic antidepressants,[49] but this use is controversial.

Serum concentration measurement can be used to individualize therapy in the early stages of treatment or when the dosage is being changed, to check adherence to therapy, to help diagnose toxicity, and to monitor the effects of drug interactions.

Carbamazepine

The target plasma carbamazepine concentration range is 17–42 mmol/l, but plasma concentrations do not correlate well with effect, since it has an active metabolite, oxcarbazepine; higher concentrations are associated with an increased risk of toxicity.[50] Carbamazepine induces its own metabolism, and its half-life is therefore shortened during long-term therapy. So, after an initial apparent steady state has been reached 3 or 4 days after starting therapy, a new steady state occurs at a lower concentration a few weeks later, and the dose may need to be increased at that time. Blood samples should be taken immediately before a dose.

Lithium

The target serum lithium concentration is 0.4–1.0 mmol/l. Concentrations above 1 mmol/l are associated with an increased risk of toxicity. Take blood samples at a standard time—12 h after the last dose. In my view, routine monitoring is unnecessary, and the serum lithium concentration should be measured at times when toxicity is most likely, for example in patients with changing renal function or with acute alterations in electrolyte balance.[51,52] Many psychiatrists prefer to measure the serum lithium concentration routinely at, say, 3- or 6-monthly intervals. However, although regular monitoring may emphasize the dangers to the patient and give an opportunity for a consultation, it is no substitute for proper monitoring at the appropriate times.

Valproate

The target serum valproate concentration range is 40–80 μmol/1, although this range is based on its use in epilepsy rather than bipolar affective disorder;[53] higher concentrations are associated with an increased risk of toxicity. The blood sample can be taken at any time after the last dose.

Further information

Aronson, J.K. (ed.) (2006). *Meyler's side effects of drugs. The international encyclopedia of adverse drug reactions and interactions* (15th edn). Elsevier, New York.

Aronson, J.K., Hardman, M., and Reynolds, D.J.M. (1993). *ABC of monitoring drug therapy*. BMJ Publishing, London.

Association of British Pharmaceutical Industries. *ABPI compendium of data sheets and summaries of product characteristics*. Datapharm Publications, London (published annually).

Baxter, K. (ed.) (2006). *Stockley's drug interactions: a source book of interactions, their mechanisms, clinical importance and management* (7th edn). Pharmaceutical Press, London.

Bennett, P.N. (ed.) (1996). *Drugs and human lactation* (2nd edn). Elsevier, Amsterdam.

Denham, M.J. and George, C.F. (eds.) (1990). Drugs in old age. *British Medical Bulletin*, **46**.

Dollery, C. (ed.) (1999). *Therapeutic drugs* (2nd edn). Churchill Livingstone, Edinburgh.

Grahame-Smith, D.G. and Aronson, J.K. (2002). *Oxford textbook of clinical pharmacology* (3rd edn). Oxford University Press, Oxford.

Gilstrap, L.C. and Little, B.B. (eds.) (1998). *Drugs and pregnancy* (2nd edn). Chapman & Hall, New York.

Joint Formulary Committee. *British national formulary*. British Medical Association and The Pharmaceutical Society of Great Britain, London (published every 6 months).

Paediatric Formulary Committee. *British national formulary for children*. British Medical Association and The Pharmaceutical Society of Great Britain, London (published annually).

References

1. Amdisen, A. (1969). Variation of serum lithium concentration during the day in relation to treatment control, absorptive side effects and the use of slow-release tablets. *Acta Psychiatrica Scandinavica*, **207**, 55–7.
2. Malone, F.D. and D'Alton, M.E. (1997). Drugs in pregnancy: anticonvulsants. *Seminars in Perinatology*, **21**, 114–23.
3. Rodriguez-Pinilla, E., Arroyo, I., Fondevilla, J., *et al.* (2000). Prenatal exposure to valproic acid during pregnancy and limb deficiencies: a case-control study. *American Journal of Medical Genetics*, **90**, 376–81.
4. Jefferson, J.W. (2006). Lithium. In *Meyler's side effects of drugs. The international encyclopedia of adverse drug reactions and interactions* (15th edn) (ed. J.K. Aronson), pp. 2073–116. Elsevier, Amsterdam.
5. Anonymous (1999). Lithium. In *Therapeutic drugs* (2nd edn), Vol. 2 (ed. C. Dollery), pp. L71–5. Churchill Livingstone, Edinburgh.
6. Chisholm, C.A. and Kuller, J.A. (1997). A guide to the safety of CNS–active agents during breastfeeding. *Drug Safety*, **17**, 127–42.
7. Pentikainen, P.J., Neuvonen, P.J., and Jostell, K.G. (1980). Pharmacokinetics of chlormethiazole in healthy volunteers and patients with cirrhosis of the liver. *European Journal of Clinical Pharmacology*, **17**, 275–84.
8. Caccia, S. (1998). Metabolism of the newer antidepressants. An overview of the pharmacological and pharmacokinetic implications. *Clinical Pharmacokinetics*, **34**, 281–302.
9. Norman, T.R., Walker, R.G., and Burrows, G.D. (1984). Renal function related changes in lithium kinetics. *Clinical Pharmacokinetics*, **9**, 349–53.

10. Korpi, E.R., Mattila, M.J., Wisden, W., *et al.* (1997). GABA(A) receptor subtypes: clinical efficacy and selectivity of benzodiazepine site ligands. *Annals of Medicine*, **29**, 275–82.
11. Wagner, J., Wagner, M.L., and Hening, W.A. (1998). Beyond benzodiazepines: alternative pharmacologic agents for the treatment of insomnia. *Annals of Pharmacotherapy*, **32**, 680–91.
12. Cross, A.J., Stirling, J.M., Robinson, T.N., *et al.* (1989). The modulation by chlormethiazole of the $GABA_A$-receptor complex in rat brain. *British Journal of Pharmacology*, **98**, 284–90.
13. Evans, R.W. and Lipton R.B. (2001). Topics in migraine management: a survey of headache specialists highlights some controversies. *Neurologic Clinics*, **19**, 1–21.
14. Seeman, P. and Tallerico, T. (1998). Antipsychotic drugs which elicit little or no parkinsonism bind more loosely than dopamine to brain D_2 receptors, yet occupy high levels of these receptors. *Molecular Psychiatry*, **3**, 123–34.
15. Weinbroum, A.A., Flaishon, R., Sorkine, P., *et al.* (1997). A risk benefit assessment of flumazenil in the management of benzodiazepine overdose. *Drug Safety*, **17**, 181–96.
16. Gardner, C.R. (1988). Potential use of drugs modulating 5HT activity in the treatment of anxiety. *General Pharmacology*, **19**, 347–56.
17. Harwood, A.J. (2005). Lithium and bipolar mood disorder: the inositol-depletion hypothesis revisited. *Molecular Psychiatry*, **10**, 117–26.
18. Marmol, F. (2006). Litio: 55 anos de historia en el tratamiento del trastorno bipolar. *Medica Clinica* (*Barcelona*), **127**, 189–95.
19. Fleckenstein, A.E., Volz, T.J., Riddle. E.L., *et al.* (2007). New insights into the mechanism of action of amphetamines. *Annual Reviews of Pharmacology and Toxicology*, **47**, 681–98.
20. Stahl, S.M. (1998). Basic psychopharmacology of antidepressants, part 1: antidepressants have seven distinct mechanisms of action. *The Journal of Clinical Psychiatry*, **59**(Suppl. 4), 5–14.
21. Blier, P. and Bergeron, R. (1997). Early onset of therapeutic action in depression and greater efficacy of antidepressant treatments: are they related? *International Clinical Psychopharmacology*, **12**(Suppl. 3), S21–8.
22. Sussman, N. and Joffe, R.T. (1998). Antidepressant augmentation: conclusions and recommendations. *The Journal of Clinical Psychiatry*, **59**(Suppl. 5), 70–3.
23. Artigas, F., Adell, A., and Celada, P. (2006). Pindolol augmentation of antidepressant response. *Current Drug Targets*, **7**, 139–47.
24. Sambunaris, A., Hesselink, J.K., Pinder, R., *et al.* (1997). Development of new antidepressants. *The Journal of Clinical Psychiatry*, **58**(Suppl. 6), 40–53.
25. Fulton, B. and Benfield, P. (1996). Moclobemide. An update of its pharmacological properties and therapeutic use. *Drugs*, **52**, 450–74 (erratum 869).
26. Aronson, J.K. and Ferner, R.E. (2003). Joining the DoTS. New approach to classifying adverse drug reactions. *British Medical Journal*, **327**, 1222–5.
27. Mulhall, A., de Louvois, J., and Hurley, R. (1983). Chloramphenicol toxicity in neonates: its incidence and prevention. *British Medical Journal*, **287**, 1424–7.
28. Greenblatt, D.J., Divoll, M., Harmatz, J.S., *et al.* (1981). Kinetics and clinical effects of flurazepam in young and elderly noninsomniacs. *Clinical Pharmacology and Therapeutics*, **30**, 475–86.
29. Schwartz, E., Potasman, I., Rotenberg, M., *et al.* (2001). Serious adverse events of mefloquine in relation to blood level and gender. *American Journal of Tropical Medicine and Hygiene*, **65**, 189–92.
30. Batchelor, J.R., Welsh, K.I., Tinoco, R.M., *et al.* (1980). Hydralazine-induced systemic lupus erythematosus: influence of HLA-DR and sex on susceptibility. *Lancet*, **1**, 1107–9.
31. Dickinson, R.G., Hooper, W.D., Wood, B., *et al.* (1989). The effect of pregnancy in humans on the pharmacokinetics of stable isotope labelled phenytoin. *British Journal of Clinical Pharmacology*, **28**, 17–27.
32. Aronson, J.K. (2001). Forbidden fruit. *Nature Medicine*, **7**, 7–8.
33. Clericetti, N. and Beretta-Piccoli, C. (1991). Lithium clearance in patients with chronic renal diseases. *Clinical Nephrology*, **36**, 281–9.
34. Hasselstrom, J., Eriksson, S., Persson, A., *et al.* (1990). The metabolism and bioavailability of morphine in patients with severe liver cirrhosis. *British Journal of Clinical Pharmacology*, **29**, 289–97.
35. Jenner, P. and Marsden, C.D. (1987). Chronic pharmacological manipulation of dopamine receptors in brain. *Neuropharmacology*, **26**, 931–40.
36. Gill, H.S., DeVane, C.L., and Risch, S.C. (1997). Extrapyramidal symptoms associated with cyclic antidepressant treatment: a review of the literature and consolidating hypotheses. *Journal of Clinical Psychopharmacology*, **17**, 377–89.
37. Puntillo, K., Casella, V., and Reid, M. (1997). Opioid and benzodiazepine tolerance and dependence: application of theory to critical care practice. *Heart & Lung*, **26**, 317–24.
38. O'Connor, P.G. and Kosten, T.R. (1998). Rapid and ultrarapid opioid detoxification techniques. *The Journal of the American Medical Association*, **279**, 229–34.
39. Miller, N.S. (1995). Pharmacotherapy in alcoholism. *Journal of Addictive Diseases*, **14**, 23–46.
40. Ashton, H. (1994). The treatment of benzodiazepine dependence. *Addiction*, **89**, 1535–41.
41. Perucca, E., Gatti, G., and Spina, E. (1994). Clinical pharmacokinetics of fluvoxamine. *Clinical Pharmacokinetics*, **27**, 175–90.
42. Cropp, J.S. and Bussey, H.I. (1997). A review of enzyme induction of warfarin metabolism with recommendations for patient management. *Pharmacotherapy*, **17**, 917–28.
43. Ryan, R., Carlson, J., and Farris, F. (1986). Effect of sucralfate on the absorption and disposition of amitriptyline in humans. *Federation Proceedings*, **45**, 205.
44. Suzuki, Y., Nagai, T., Mano, T., *et al.* (1995). Interaction between valproate formulation and phenytoin concentrations. *European Journal of Clinical Pharmacology*, **48**, 61–3.
45. Hilton, S.E., Maradit, H., and Moller, H.J. (1997). Serotonin syndrome and drug combinations: focus on MAOI and RIMA. *European Archives of Psychiatry and Clinical Neuroscience*, **247**, 113–9.
46. Keltner, N. and Harris, C.P. (1994). Serotonin syndrome: a case of fatal SSRI/MAOI interaction. *Perspectives in Psychiatric Care*, **30**, 26–31.
47. Sobanski, T., Bagli, M., Laux, G., *et al.* (1997). Serotonin syndrome after lithium add-on medication to paroxetine. *Pharmacopsychiatry*, **30**, 106–7.
48. Adan-Manes, J., Novalbos, J., Lopez-Rodriguez, R., *et al.* (2006). Lithium and venlafaxine interaction: a case of serotonin syndrome. *Journal of Clinical Pharmacy and Therapy*, **31**, 397–400.
49. Isacsson, G., Bergman, U., Wasserman, D., *et al.* (1996). The use of antidepressants and therapeutic drug monitoring by general practitioners and psychiatrists: findings from a questionnaire survey in two Swedish areas. *Annals of Clinical Psychiatry*, **8**, 153–60.
50. Vasudev, K., Goswami, U., and Kohli, K. (2000). Carbamazepine and valproate monotherapy: feasibility, relative safety and efficacy, and therapeutic drug monitoring in manic disorder. *Psychopharmacology* (*Berlin*), **150**, 15–23.
51. Aronson, J.K. and Reynolds, D.J. (1992). ABC of monitoring drug therapy. Lithium. *British Medical Journal*, **305**, 1273–4.
52. Glasziou, P. and Aronson, J.K. (2008). An introduction to monitoring therapeutic interventions in clinical practice. In *Evidence-based medical monitoring: from principles to practice* (eds. P.P. Glasziou, L. Irwig, and J.K. Aronson), Chap. 1. Wiley-Blackwell Publications, Oxford.
53. Fleming, J. and Chetty, M. (2006). Therapeutic monitoring of valproate in psychiatry: how far have we progressed? *Clinical Neuropharmacology*, **29**, 350–60.

6.2.2 Anxiolytics and hypnotics

Malcolm Lader

Introduction

Anxiety is a commonly experienced emotion that becomes a clinical disorder when it is too severe, too protracted, or too pervasive for the subject to bear. Insomnia is a failure to experience satisfying sleep, together with a feeling of tiredness during the day. Many compounds, the anxiolytics and hypnotics, are used to treat these conditions, but the two groups of drugs overlap.

The classical antianxiety drugs (anxiolytics) are alcohol, the opioids, and the barbiturates. For the past 45 years, the benzodiazepines, such as diazepam and lorazepam, have dominated the field. They are effective anxiolytics in the short term but their long-term efficacy remains in dispute. Their disadvantages include cognitive and psychomotor impairment, paradoxical reactions, tolerance, and dependence, and they are major drugs of abuse.

Other anxiolytics act on the 5-hydroxytryptamine (5-HT; serotonin) systems of the brain and include buspirone and the selective serotonin reuptake inhibitors (SSRIs). Newer compounds are still being introduced that lie outside these groups.

The use of benzodiazepine and benzodiazepine-like hypnotics, by contrast, continues apace. Some switching to the shorter-acting benzodiazepines has occurred, together with the introduction of the 'z-compounds', zopiclone, zolpidem, and zaleplon. These drugs tend to have fewer residual effects the next day than the benzodiazepines, and are claimed to be less likely to induce rebound and dependence than equivalent benzodiazepines. Particular care is needed in prescribing such hypnotics to the elderly.

The rational use of both anxiolytics and hypnotics requires minimal dosage, short durations of use, and simultaneous exploitation of non-pharmacological methods.

Definitions

'Sedative' originally meant a substance that has the property of allaying anxiety. However, it has now come to denote feelings of drowsiness or torpor. This state was originally called 'oversedation', and was often noted with the barbiturates and other older drugs such as chloral. Next, the term 'tranquillizer' was introduced 40 or more years ago in an attempt to distinguish between the older sedatives and the newer drugs, supposedly non-sedative, such as the benzodiazepines. But this distinction is artificial as, apart from safety in overdosage, the benzodiazepines closely resemble the barbiturates in pharmacological and clinical properties. The term 'anxiolytic' is now generally favoured.

Anxiolytic drugs

Anxiety-allaying drugs have been used for thousands of years, dating back to the discovery that, among its psychotropic properties, alcohol could induce sedation. The nineteenth century saw the development of inorganic and, later, organic chemical compounds. Bromides were introduced as sedatives and became widely used despite their poor effectiveness, toxicity, and potential abuse. Organic chemists in the second half of the nineteenth century introduced sedatives such as chloral and paraldehyde.

The first barbiturate was introduced over a 100 years ago. This group of drugs is divisible into the ultrashort-acting (e.g. anaesthetic-induction agents such as thiopentone and methohexital), short-acting (e.g. secobarbital), medium-acting (e.g. butobarbital), and long-acting (e.g. phenobarbital) barbiturates. Most of the rest are of medium duration with half-lives of 16 h or so. The disadvantages of the barbiturates include drowsiness, tolerance to their effects, dangers of overdose, and possible physical and psychological dependence with severe withdrawal syndromes.[1] Meprobamate was introduced as the first of the 'tranquillizers', but its advantages over the barbiturates proved minimal.

The benzodiazepines were first synthesized in the 1930s, but not developed until 2 years later. The prototype, chlordiazepoxide, was evaluated in the clinic, found effective, and soon introduced into medical practice. More than 1000 benzodiazepines and related compounds have been synthesized, including diazepam, the most widely used of all. Anxiolytic and hypnotic, as well as muscle-relaxant and anticonvulsant properties are licensed indications. However, the distinction between anxiolytic and hypnotic uses often seems to owe more to commercial expediency than to scientific rationale; some compounds, such as lorazepam, are marketed for both indications.

The benzodiazepines

The main reason for the original popularity of the benzodiazepines was the perceived safety in overdose compared with the quite marked toxicity of the barbiturates. In turn, concern has mounted concerning the benzodiazepines.[2] These drugs are widely prescribed by many physicians for patients with emotional problems, circulatory disorders, tension headaches, and pains in the chest and back as well as digestive disorders, all with the common symptom of anxiety. This widespread use, even overuse and the induction of dependence even at normal therapeutic dose has led to official injunctions for greater caution in prescribing.

Pharmacokinetics

Two aspects of the pharmacokinetics of the benzodiazepines are relevant to the prescriber—speed of onset of action and the duration of that action. The speed of onset depends on the mode of administration and the penetration time to the brain. Given by mouth, most benzodiazepines are rapidly absorbed and exert a prompt anxiolytic effect, for instance in panic states. Diazepam and lorazepam are prime examples. Although temazepam enters the brain more rapidly than, say, oxazepam, it still takes an appreciable time to induce sleep. The redistribution phase can be pronounced and will then largely determine the duration of effect of single doses of benzodiazepines such as diazepam and flunitrazepam.

The metabolic half-lives of the benzodiazepines also vary greatly. *N*-desmethyldiazepam (nordiazepam) is the major and active metabolite of diazepam and several other benzodiazepines. It has a long half-life, about 60 h, and accumulates over the first month of treatment. Metabolism of these drugs is even slower in the elderly and in patients with liver damage.

Benzodiazepines with a 3-hydroxyl grouping, such as lorazepam, oxazepam, and temazepam, have half-lives averaging 12 h or less. Liver damage has to be severe before the metabolism of these drugs

is affected. Alprazolam is a triazolobenzodiazepine with a half-life of 9 to 16 h, and with hydroxy metabolites of low biological activity. Both chlordiazepoxide and diazepam are absorbed erratically after intramuscular injection. Lorazepam, however, is well absorbed after intramuscular injection.

Basic pharmacology

The benzodiazepines potentiate the widespread inhibitory neurotransmitter γ-amino butyric acid (GABA). Benzodiazepines do not act directly on GABA receptors but have their own receptors. Because of this widespread inhibitory effect, benzodiazepines alter the turnover of neurotransmitters such as norepinephrine and serotonin. The main sites of action of the benzodiazepines are in the spinal cord where muscle-relaxant effects are mediated, the brainstem (perhaps accounting for their anticonvulsant properties), the cerebellum (causing ataxia), and the limbic and cortical areas involved in the organization of emotional experience and behaviour.

Clinical pharmacology

The depressant effects of single therapeutic doses of a benzodiazepine can usually be readily detected. However, lower doses may fail to impair psychological functioning and subjective effects are usually absent. In the clinical context with anxious patients and with repeated higher doses, sustained impairment of functioning is more difficult to demonstrate. Some studies have shown decrements in performance after the first dose, but improvements in functioning, in comparison to predrug levels, may become apparent by the end of a week of repeated usage. This suggests that the well-known impairment of performance produced by pathologically high levels of anxiety is first worsened by the sedative effects. Then as the antianxiety effects build-up, the patient's psychological functions may improve.

A second mechanism concerns tolerance, which reflects several biochemical mechanisms including alteration in benzodiazepine-receptor type. Patients who have a high alcohol intake are tolerant to benzodiazepines.

The benzodiazepines have marked and selective effects on memory by interfering with episodic memory, that is to say the system concerned with remembering personal experiences.[3] This effect seems independent of any sedation or attentional impairment. Alcohol adds to the cognitive impairment induced by the benzodiazepines but does not necessarily potentiate it.

The dependence potential of benzodiazepines is seen in drug-preference studies, but these drugs are much less preferred than say the amphetamines. Differences among benzodiazepines have been documented; for example, oxazepam seems to have less abuse liability than diazepam.

The largest gap in our knowledge of these drugs is on their long-term usage, which has been evaluated in relatively few studies.[4] Thus, it is still largely unclear whether therapeutic effects are maintained in most patients for longer than a few weeks and when dependence supervenes in the minority of patients who encounter problems on protracted usage.

Hypnotic drugs

The main groups of drugs used in the modern treatment of insomnia are the benzodiazepines, and the newer compounds, zopiclone, eszopiclone, zolpidem, and zaleplon. The pharmacology of these benzodiazepines is essentially the same as that of the anxiolytic compounds.

Nitrazepam is a long-acting benzodiazepine with an elimination half-life ranging between 25 and 35 h, but it is longer in the elderly. Because of this, it is likely to produce residual effects and to accumulate. Flunitrazepam is more potent, but is somewhat shorter acting with a half-life of 10 to 20 h. It has a rapid redistribution phase, which can result in a short duration of intense action. It has earned an undeserved reputation as the 'date-rape' drug. Flurazepam is still widely used in the United States. It has a very long-acting metabolite, which can produce psychological impairment on regular dosage, especially in the elderly. Of the intermediate-acting compounds, temazepam has a half-life of 10 to 15 h, without active metabolites. At modest dose (10–15 mg daily), it results in few residual effects and is fairly well tolerated by the elderly. Major problems with abuse have limited its popularity, but it is still widely prescribed worldwide. Lormetazepam is slightly shorter acting, loprazolam has a fairly short half-life, but its absorption may be slow and erratic.

Triazolam is the archetypal short-acting benzodiazepine, with a mean half-life of around 3 to 4 h, and no clinically significant metabolites. Daytime sedation is seen after high doses (0.5 mg daily), but not usually with lower ones. These higher doses have also been associated with an increased incidence of anterograde amnesia and unusual behaviours, including depressive reactions and hostility.

Zopiclone is a cyclopyrrolone derivative believed to bind close to, but not exactly at, the benzodiazepine receptor. It has a half-life of about 5 h in younger subjects and about 8 h in the elderly. Its sedative and hypnotic effects are similar to those of the benzodiazepines, but its side effect profile is generally superior with fewer central nervous system effects such as oversedation, confusion, and memory impairment. Rebound and withdrawal problems also seem to be less.

Eszopiclone is the S-enantiomer of zopiclone, which is a racemic mixture. It is licensed for the long-term treatment of insomnia in the United States, following successful clinical trials.

Zolpidem is an imidazopyridine compound that binds selectively to one subtype of the benzodiazepine receptor. It is rapidly absorbed and has a short elimination half-life of 0.7 to 3.5 h (mean 2.4 h). It decreases sleep-onset latency but has less consistent effects on total sleep time.[5] Residual effects are minimal, as are memory disturbances. Rebound and withdrawal are uncommon but have been documented.

Zaleplon is also a selective compound with a very short half-life averaging only 1 h. It shortens sleep onset without usually prolonging total sleep time. Residual effects are absent, and memory is minimally disturbed.

Clinical effects of anxiolytics

Although the usual licensed indications are generalized anxiety and panic disorder,[6,7] the main practical application of the benzodiazepines is to aid in the symptomatic management of anxiety and stress-related conditions.[8] These indications are often so wide as to be difficult to define in terms of recognized disorders. Instead the symptoms of anxiety, in whatever context, are the main indication.

Thousands of comparative trials among the benzodiazepines have been carried out, but few differences with respect to risk–benefit ratios have been found.

Antianxiety medications are difficult to assess. Anxiety disorders are very varied in their natural history; some resolve over a few weeks, whereas others become chronic for no apparent reason, with subsequent acute-on-chronic exacerbations. The patients with chronic, severe unresponsive illnesses tend to be referred to psychiatric outpatient departments. Uncontrolled observations on family practice patients will give a more encouraging impression of antianxiety drugs than will assessment of the more chronic patients attending psychiatric clinics. Even in the latter type of patient, useful symptomatic relief is often obtained without complete resolution of the illness.

Drugs such as diazepam have a long elimination half-life so that once daily or nightly dosage is sufficient. Nevertheless, many patients prefer to take a divided dosage during the day, often claiming that they can detect further antianxiety activity after each dose and are thereby reassured. For episodic anxiety, shorter-acting compounds such as lorazepam can be used, taken 30 min or so before entering the anxiety-provoking situation. If the panic has already started, lorazepam can still be given and will exert a fairly prompt action. Lorazepam is also invaluable in the emergency management of the acutely anxious and disturbed psychotic patient.

Antipanic actions have been claimed for the benzodiazepines, in particular alprazolam, acting to prevent the episodes rather than aborting them. However, although suppression of the panic attack is often quite effective, relapse, and even rebound may occur when the benzodiazepine is discontinued, even if it is tapered off.[9] Because of this SSRI antidepressants are generally preferred.[7]

The short-acting benzodiazepines are also used as adjuncts to relaxation therapy, preoperative medication, and deep sedation for minor operative procedures such as dentistry. The drugs render the patient calm, conscious, and cooperative, with often total anterograde amnesia for the operation.

Unwanted effects

The commonest unwanted effects of the anxiolytic benzodiazepines are tiredness, drowsiness, and torpor, features of 'oversedation'. The effects are dose and time related, being maximal within the first 2 h after large doses. Drowsiness is most noticeable during the first week of treatment, after which it largely disappears probably due to a true tolerance effect. Patients should be warned of the potential side effects of any prescribed benzodiazepine and the initial dosage should be cautious. Both psychomotor skills and intellectual and cognitive skills are affected. In particular, patients should be advised not to drive during the initial adjustment of dosage. Important decisions should be deferred during this period because judgement may be clouded.

Benzodiazepines have major effects on cognitive function in long-term users. A meta-analysis of 13 research studies revealed impairments across all cognitive categories examined.[4] The drugs differ in their ability to produce memory deficits, with lorazepam being especially powerful.[3] However, most benzodiazepines can cause problems, especially in higher dose and in the elderly.

Psychomotor performance is also affected, with elderly drivers particularly at risk. As with other depressant drugs, potentiation of the effects of alcohol can occur. Patients must be warned not to drink alcohol when taking benzodiazepines, either chronically or intermittently. Patients taking benzodiazepines may develop paradoxical behavioural responses such as uncontrollable weeping, increased aggression and hostility, and acute rage reactions or uncharacteristic criminal behaviour such as shoplifting.[10] This phenomenon is by no means confined to the benzodiazepines; alcohol is a cardinal example of a drug whose use may lead to excessive violence or criminal behaviour. Paradoxical reactions, including the release of anxiety or hostility, are most common during the initial week of treatment, and usually resolve spontaneously or respond to dose adjustment. Reports of the induction of depression by the benzodiazepines in patients with apparent generalized anxiety disorder are probably the result of an initial misdiagnosis and a failure to detect the underlying depression.

Other unwanted effects include respiratory depression, excessive weight gain, skin rash, impairment of sexual function, menstrual irregularities, and rarely, blood dyscrasias. The use of benzodiazepines in pregnancy is generally regarded as reasonably safe. However, benzodiazepines pass readily into the foetus and can produce respiratory depression in the neonate. Finally, benzodiazepines pass into the mother's milk and can over sedate the baby, so breastfeeding should be discouraged if benzodiazepines are prescribed, especially in high dose.

Overdosage

Overdosage with benzodiazepines is common; deaths are not. Although fatal-overdose statistics contain deaths ascribed to benzodiazepines alone,[11] many such attributions are suspect. Only in children and the physically frail, especially those with respiratory illness, are the benzodiazepines on their own hazardous. However, they can markedly potentiate other central nervous system depressant drugs such as alcohol. Typically, the person falls into a deep sleep but can be roused the administration of the benzodiazepine antagonist, flumazenil.

Tolerance and dependence

If tolerance occurred regularly, then escalation of dosage would be the norm. This does occur with the benzodiazepines, but is fairly uncommon. Escalation of dose is often stepwise, with each increment following a temporary deterioration in psychosocial circumstances. Most patients later reduce the dose as the stress resolves, but others continue the higher dose to which they presumably have developed some tolerance.

Tolerance to the clinical effects in patients maintaining moderate doses of benzodiazepines is now generally accepted.[12] Few controlled observations concern the long-term efficacy of antianxiety compounds in chronically anxious patients. If medication is withdrawn, the original symptoms may reappear. This is taken as evidence that therapeutic benefit still continues. However, it may reflect 'rebound' rather than long-term clinical benefit. Undoubtedly, many chronically anxious patients are helped by their treatment with benzodiazepines, but this raises the question as to the frequency of psychological and physical dependence on these drugs. Dependence is easily demonstrable in those patients who have attained high doses. Rebound and withdrawal symptoms after the long-acting benzodiazepine diazepam are not usually apparent until about 5 to 10 days after discontinuation. It is much shorter in patients discontinuing the shorter-acting benzodiazepines (2–4 days). The mildest symptoms and signs are anxiety, tension, apprehension, dizziness, tremulousness, insomnia, and

anorexia. More severe physical dependence is shown by the withdrawal symptoms of nausea and vomiting, severe tremor, muscle weakness, postural hypotension, and tachycardia. Occasionally, hyperthermia, muscle twitches, convulsions, and confusional psychoses may develop.

After normal dose usage, perceptual changes can be particularly troublesome.[13] The proportion of patients taking benzodiazepines chronically who experience withdrawal symptoms on discontinuing medication ranges between 27 and 45 per cent, depending on the criteria used. Sometimes the withdrawal reactions seem very prolonged[14] or depression may supervene.

Management of withdrawal

It is widely accepted that the most appropriate way to manage patients withdrawing from benzodiazepines is to taper the dose gradually, because the severe symptoms of withdrawal, such as epileptic fits and confusional episodes, are more likely to follow abrupt than gradual withdrawal. Views differ as to the rate of withdrawal. Detailed guidelines,[15,16] based on a consensus view in the United Kingdom, recommend minimal intervention first, usually by a general practitioner (Fig. 6.2.2.1). This may comprise a letter to the long-term user, or an interview on a routine visit, with advice to taper the medication. More active intervention involves careful assessment, education, and then the adoption, with the patient's agreement, of a timetable of about 6 to 8 weeks for withdrawal. Some agencies suggest a month of tapering for every year of benzodiazepine use, but this may result in patients becoming preoccupied with their symptoms. One strategy is to try a fairly brisk withdrawal, say over 6 to 8 weeks, and only resort to more gradual tapering if the symptoms become intolerable. Another ploy is first to substitute a long-acting for a short-acting benzodiazepine, say 10 mg of diazepam for 1 mg of lorazepam, and then to taper off the diazepam later.

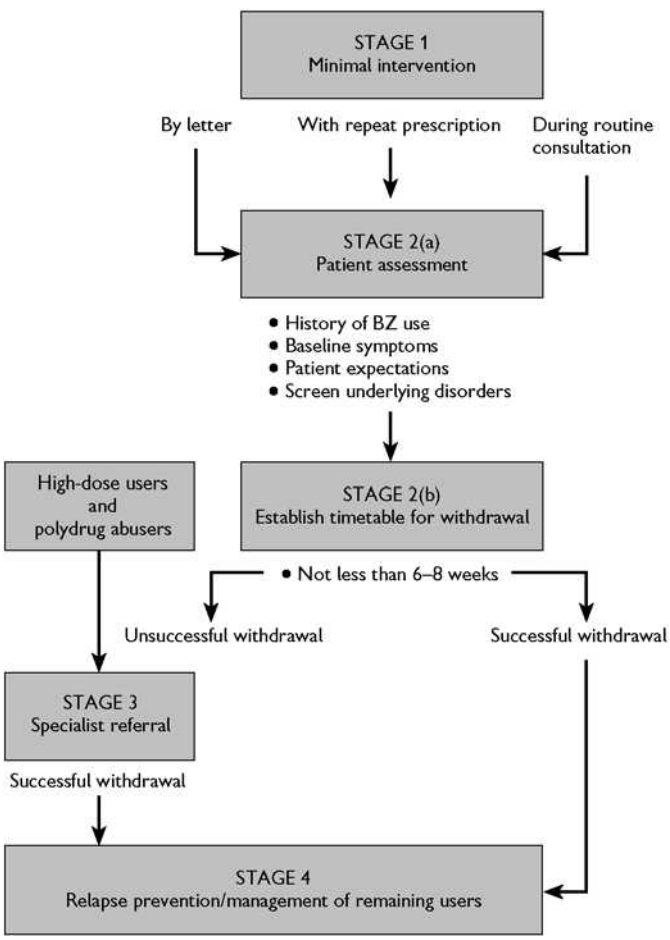


Fig. 6.2.2.1 Guidelines for withdrawal from benzodiazepines (BZ). (Reproduced with permission from J. Russell and M. Lader (eds.), *Guidelines for the prevention and treatment of benzodiazepine dependence*, Mental Health Foundation, London. Copyright 1993, Mental Health Foundation, London.)

Patients must be carefully followed up as a depressive illness is not uncommon and may need vigorous treatment. Such an illness may be reactive to the stress of withdrawal or be a recurrence of an earlier affective episode.

Other drugs have been advocated, but most patients are loath to substitute yet another medication. Depressed patients should have the depression treated before attempting withdrawal. Based on evidence from animal studies, fairly large single doses of flumazenil have been tried with some success.

Psychological support is essential, with the doctor or a practice nurse maintaining close contact with the patient during withdrawal. The physician should show clearly that he understands the problems of withdrawal in order to capture the confidence of the patient. He or she must recognize that patients frequently incubate numerous misconceptions and negative expectations about tranquillizers and withdrawal. These must be elicited, identified, discussed, and corrected.

Cognitive behavioural treatment is currently favoured and is often effective if administered by an experienced professional. Relaxation treatment and training in anxiety management skills in the framework of group therapy can boast of only moderate effectiveness.

Other anxiolytics

Benzodiazepine-receptor partial agonists

The disadvantages of the benzodiazepines include sedation, psychomotor and cognitive impairment, and withdrawal symptoms after long-term use. Increased understanding of benzodiazepine-receptor mechanisms suggested that compounds might be developed which are partial agonists and/or selective to some subtypes of receptor.[17] Such compounds would be less efficacious than full agonists but might have better adverse-effect profiles and less dependence potential, that is superior risk–benefit ratios. Their promise has not been fulfilled, as the risk–benefit ratios of these compounds do not seem superior to those of the full agonists.

5-HT_{1A} partial agonists

These drugs have a complex pharmacology. The first, buspirone, was licensed in many countries some years ago. These drugs suppress activity in presynaptic serotonergic neurones, diminishing serotonin activity, and leading on to down-regulation of 5-HT_2 and perhaps other 5-HT receptors. Buspirone is much less sedative than the benzodiazepines and causes little or no psychomotor or cognitive impairment, nor does it potentiate the effects of alcohol. In formal clinical trials, buspirone was equi-effective and

equipotent to diazepam, but patients taking buspirone improve more slowly.[18] The side effects of buspirone include headache, dizziness, and nausea. Discontinuation is not accompanied by either rebound or withdrawal.

Other 5-HT_{1A} partial agonists have been developed mainly as potential antidepressants and antianxiety agents, but few have been marketed, mainly because of disappointing efficacy.

Pregabalin

This compound is a structural analogue of GABA although it is not active at GABA receptors.[19] It binds with high affinity to an auxiliary α_2-δ subunit protein of voltage-gated calcium channels in the CNS: it acts as a presynaptic modulator of the excessive release of neurotransmitters in hyperexcited neurones. It is predominantly excreted unchanged in the urine. It was initially developed for use as an adjunctive treatment in epilepsy and neuropathic pain. Several clinical trials in GAD have shown it to have efficacy akin to those for benzodiazepines and venlafaxine.[20] It has a rapid onset of action and was effective in preventing relapse over 34 weeks. Tolerance was good during dosage escalation to the usual dose of 150–600 mg/day, mild dizziness and somnolence being the usual adverse effects. No clinically significant withdrawal was seen after tapering. It is approved for the treatment of GAD in Europe, and is a significant introduction.

Antiepileptic drugs

There is a long history of the use of many of these compounds in anxiety disorders,[21] but they are not routine choices.

Antipsychotic drugs

Phenothiazines, such as chlorpromazine and trifluoperazine, and a range of other antipsychotic drugs have been advocated for treating anxiety. The dosage recommended is quite low, typically less than half the initial antipsychotic dose used in psychotic patients. Sometimes, even at this dosage, the antipsychotic drug is not well tolerated by the anxious patient because unwanted autonomic effects, such as dry mouth and dizziness, too closely resemble the symptoms of anxiety. Even more unwelcome are extrapyramidal symptoms such as restlessness (mild akathisia) and parkinsonism, although at the low doses advocated such unwanted effects are uncommon. There may even be a risk of tardive dyskinesia. The chief advantage of this medication is that dependence is virtually unknown, so the main indication for their use is in patients with histories of dependence on other central nervous system depressant drugs such as alcohol or barbiturates.

Antidepressants

Several of these drugs, such as amitriptyline, doxepin, and trazodone, have useful secondary sedative properties. They are widely prescribed for depressed patients with anxiety or agitation. More recently, several SSRI antidepressants such as paroxetine and escitalopram have been evaluated in the treatment of various anxiety disorders. SSRIs are now the treatment of choice in chronic anxiety disorders.[22,23]

MAOIs have been used for many years to treat phobic states, but the well-known range of unwanted effects, including hypotension, limb oedema, and dietary and drug interactions, preclude their routine use.

β-adrenoceptor antagonists

Beta-adrenoceptor antagonists may help patients with anxiety, but usually only those complaining of somatic symptoms. They are still favoured by some primary care doctors.

Antihistamines

The older compounds penetrate the brain readily and are quite sedative. Hydroxyzine has been evaluated in at least two placebo-controlled trials in doses of 50 mg/day: it proved to be significantly better than placebo. The advantages are that paradoxical reactions are rare, cognitive function including memory is largely unaffected, and rebound and withdrawal seem rare.

Clinical effects of hypnotics

Insomnia is a common symptom,[24] especially in the elderly.[25] Nonetheless, many complaints of insomnia are unfounded as the patient has unreal expectations concerning sleep. Elderly people fail to appreciate that it is normal to sleep less and less deeply as they age. Napping during the day also decreases the need for sleep at night. Some people can manage on 5 to 6 h a night indefinitely, and yet worry that this is insufficient. Explanation and reassurance relieve their worries.

In many patients complaining of more severe insomnia, the cause is a physical complaint such as pain, breathlessness, or pruritus. The treatment is for that of the primary complaint. In many other cases, the insomnia is either a symptom of psychiatric distress, anxiety, or depression, or it is iatrogenic, caused by the very drugs prescribed to relieve the insomnia. In the first instance, treatment is directed towards the primary condition; in the second, a careful regimen of drug withdrawal, or substitution and subsequent withdrawal, should be instigated, as discussed earlier for anxiolytic medication. Some drugs, of which caffeine is the most common, induce insomnia. Alcohol may also disrupt sleep, particularly during the latter half of the night.

Despite this, a substantial number of patients cannot be placed into these categories and yet they persistently complain of insomnia (primary insomnia). Careful evaluation of the issues may yet reveal some relationship to stresses, both transient and persistent. It can be established that these patients are responding to unusual or protracted pressures of life: a man worries over possible redundancy, his wife is concerned about their delinquent son, their daughter is lovelorn, and grandmother is anxious over her increasing frailness. Giving drugs may set in train a long-term process culminating in drug-related insomnia without solving the basic problems.

Short-term symptomatic relief is acceptable when the stress is undoubtedly severe but transient. Even so, the hypnotic agent must be chosen carefully. The elimination half-life is the most important consideration. Those with half-lives over 12 h, such as nitrazepam, are only appropriate where an anxiolytic effect is required during the day as well as sleep induction at night. Even here, diazepam 5 to 15 mg, one dose at night, may be preferred. Temazepam with its shorter half-life will encourage sleep onset without leaving the patient with too many residual sedative effects the next day. Unfortunately, it has been extensively abused.

The management of chronic insomnia is much more problematic.[26] The newer compounds zopiclone and zolpidem are also short-acting agents and can help assure a good night's

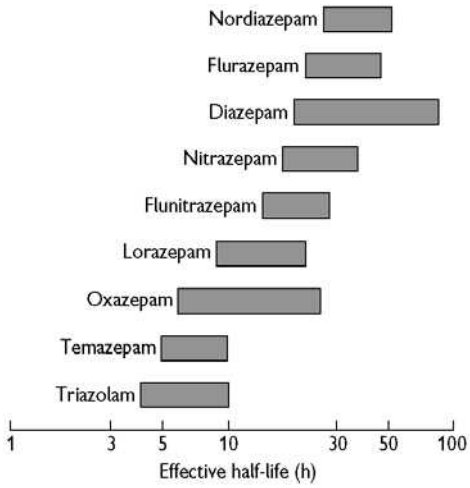


Fig. 6.2.2.2 Table of half-lives.

sleep without much risk of residual sedative effects the next day.[27] This is dependent on the dosage being kept modest, especially in the elderly. Eszopiclone is licensed in the United States for long-term use in chronic insomnia and is already used extensively there.

Zolpidem and zaleplon can be used in a different strategic ways from other longer-acting drugs. Hypnotics are traditionally taken every night before going to bed to induce or maintain sleep. However, the severity of insomnia usually varies from night to night. Consequently, regular usage may be partly, or even largely, unnecessary, and increases the risk of habituation and dependence. Very-short-acting compounds are unlikely to leave residual effects the next day, even taken up to 5 h or so before the expected time of wakening. Consequently, the insomniac can refrain from regular hypnotic usage but, instead, wait up to an hour or so after going to bed to see if natural sleep supervenes before resorting to medication. This changes the regular prophylactic use of hypnotics to 'as needed', and lessens the risk of habituation and dependence. Furthermore, the patient is gratified to feel that he or she has control of the medication instead of vice versa.

Residual effects

Residual effects can be a problem especially when long-acting drugs are used repeatedly. Dosage is important here, since residual effects increase in both magnitude and duration as the dose is increased. It should be remembered that hypnotics are the only class of drugs in which the main therapeutic effect (drowsiness) is identical with the main unwanted effect; the two are merely separated by 8 h in time. A short-acting hypnotic compound will be devoid of residual effects the next day, but the patient may wake early. After taking a longer-acting compound, sleep may be prolonged but hangover effects pronounced.

Idiosyncratic effects

Adverse effects with triazolam alerted prescribers and regulators to possible major adverse effects of short-acting benzodiazepines. The adverse reactions in question include daytime anxiety, amnesic effects, and episodes, and morbid affects such as depression and hostility. In summary, the evidence suggests that these are class effects common to the benzodiazepines, although more likely to occur the shorter the duration of action of the drug and the higher the dose. Alcohol is also capable of producing these effects.

Rebound

Discontinuation of many hypnotics is often followed by worsening of sleep compared with pretreatment levels. In practical terms, insomniac patients find that their sleep is disturbed for a night or two after abrupt discontinuation of what appeared to be effective medication. Some of this rebound is subjective as patients taking sleeping pills tend to overestimate their sleeping time (compared with sleep laboratory recordings); on withdrawal, they underestimate their sleep. The intensity of rebound insomnia is strongly related to dose but less clearly to the duration of use, and marked individual differences exist. The risk of rebound is greater with short-half-life compared with the long-half-life compounds. Tapering off medication lessens the likelihood of rebound. However, despite clinical impressions that rebound insomnia might lead to the resumption of medication, there is little evidence for this.[28]

Dependence

Dependence may supervene on the longer-term use of hypnotics; giving a long-acting benzodiazepine drug only once in 24 h does not protect against such an eventuality. The management of the withdrawal syndrome that may occur is largely the same as with the anxiolytic benzodiazepines.

Abuse

A growing problem with these drugs is abuse—non-medical use, on a regular or sporadic basis, often in a polydrug context. Worldwide, flunitrazepam is the main problem and can be taken orally, by injection, or by sniffing. In the United Kingdom, temazepam is widely abused by injection. The injected drug has a marked sedative and/or disinhibiting effect, resulting in chaotic behaviour, carelessness, and an enhanced risk of the transmission of communicable diseases such as HIV infection and hepatitis.

Other hypnotics

Gaboxadol is a GABA agonist. Preliminary data suggest useful hypnotic properties.

Melatonin preparations

A series of compounds are being developed based on melatonin. This hormone is important in the regulation of sleep and is secreted at night. Some elderly insomniacs seem to be deficient in melatonin. Preparations include Circadin®, licensed in the UK for short-term use in insomnia in over 55s. Ramelteon is a melatonin$_1$ and melatonin$_2$ agonist. It is licensed in the United States for insomnia, and is effective in inducing sleep: it has a favourable safety profile.[29]

Conclusions

In many countries the drug treatment of both anxiety and insomnia still largely revolves around the use of the benzodiazepines. Nevertheless, controversy and disagreement still rage about the risk–benefit ratio of compounds in this area. Short-term use in both indications is well established, with a favourable database as a

rationale for this approach. However, long-term use is still only researched in a limited way. While both the efficacy and safety of long-term use remain unclear, acceptance of current guidelines limiting the use of benzodiazepines seems wise.

The advent of the SSRIs as anxiolytics has driven a wedge between the treatment methods for anxiety and insomnia. Anxiety can be treated just as effectively with an SSRI (and probably, pregabalin) as with a benzodiazepine, and more safely. The treatment of insomnia still relies on the benzodiazepines until the risk–benefit ratio of newer drugs such as the melatonin-related compounds becomes clear.

Nevertheless, in the author's opinion the most important outstanding issue is the relationship between drug and non-drug treatments.[30] The management of anxiety disorders and of insomnia is complex and is hampered by a dearth of information concerning the relative merits of various treatment modalities. Much research is also needed on the optimum strategies for combining all the therapies available to us, and on identifying predictors of response.

Developments in the neuropharmacology of insomnia hold out the promise of new compounds with novel and perhaps more effective modes of action.[31] With respect to anxiety disorders, a major shift of emphasis has followed the demonstration of the efficacy of the SSRIs.[32]

Further information

Taylor, D., Kerwin, R., and Paton, C. (2005). *The Maudsley prescribing guidelines* (8th edn). South London and Maudsley NHS Trust.

Nutt, D. and Ballenger, J. (eds.) (2003). *Anxiety disorders*. Blackwell Science, Oxford.

References

1. Allgulander, C. (1986). History and current status of sedative-hypnotic drug use and abuse. *Acta Psychiatrica Scandinavica*, **73**, 465–78.
2. Lader, M. (1994). Benzodiazepines. A risk–benefit profile. *CNS Drugs*, **1**, 377–87.
3. Curran, H.V. (1991). Benzodiazepines, memory and mood: a review. *Psychopharmacology*, **105**, 1–8.
4. Barker, M.J., Greenwood, K.M., Jackson, M., *et al.* (2004). Cognitive effects of long-term benzodiazepine use. *CNS Drugs*, **18**, 38–45.
5. Langtry, H.D. and Benfield, P. (1990). Zolpidem. A review of its pharmacodynamic and pharmacokinetic properties and therapeutic potential. *Drugs*, **40**, 291–313.
6. Sramek, J.J., Zarotsky, V., and Cutler, N.R. (2002). Generalised anxiety disorder. Treatment options. *Drugs*, **62**, 1635–45.
7. Taylor, C.B. (2006). Panic disorder. *British Medical Journal*, **332**, 951–5.
8. Hoehn–Saric, R. (1998). Generalised anxiety disorders. Guidelines for diagnosis and treatment. *CNS Drugs*, **9**, 85–98.
9. Otto, M.W., Pollack, M.H., and Sachs, G.S. (1993). Discontinuation of benzodiazepine treatment: efficacy of cognitive–behavioral therapy for patients with panic disorder. *The American Journal of Psychiatry*, **150**, 1485–90.
10. Dietch, J.T. and Jennings, R.K. (1988). Aggressive dyscontrol in patients treated with benzodiazepines. *The Journal of Clinical Psychiatry*, **49**, 184–7.
11. Serfaty, M. and Masterton, G. (1993). Fatal poisonings attributed to benzodiazepines in Britain during the 1980s. *The British Journal of Psychiatry*, **163**, 386–93.
12. Michelini, S., Cassano, G.B., Frare, F., *et al.* (1996). Long-term use of benzodiazepines: tolerance, dependence and clinical problems in anxiety and mood disorders. *Pharmacopsychiatry*, **29**, 127–34.
13. Petursson, H. and Lader, M.H. (1981). Withdrawal from long–term benzodiazepine treatment. *British Medical Journal*, **283**, 643–5.
14. Tyrer, P. (1991). The benzodiazepine post-withdrawal syndrome. *Stress Medicine*, **7**, 1–2.
15. Russell, J. and Lader, M. (eds.) (1993). *Guidelines for the prevention and treatment of benzodiazepine dependence*. Mental Health Foundation, London.
16. Voshaar, R.C., Couvee, J.E., van Balkom, A.J., *et al.* (2006). Strategies for discontinuing long-term benzodiazepine use: meta-analysis. *The British Journal of Psychiatry*, **189**, 213–20.
17. Potokar, J. and Nutt, D.J. (1994). Anxiolytic potential of benzodiazepine receptor partial agonists. *CNS Drugs*, **1**, 305–15.
18. Fulton, B. and Brogden, R.N. (1997). Buspirone. An updated review of its clinical pharmacology and therapeutic applications. *CNS Drugs*, **7**, 68–88.
19. Kavoussi, R. (2006). Pregabalin: from molecule to medicine. *European Neuropsychopharmacology*, **16**, S128–33.
20. Frampton, J.E. and Foster, R.H. (2006). Pregabalin in the treatment of generalised anxiety disorder. *CNS Drugs*, **20**, 685–93.
21. Van Ameringen, M., Mancini, C., Pipe B., *et al.* (2004). Antiepileptic drugs in the treatment of anxiety disorders. Role in therapy. *Drugs*, **64**, 2199–220.
22. Stein, D.J. (2006). Evidence-based treatment of anxiety disorders. *International Journal of Psychiatry in Clinical Practice*, **10**(Suppl. 1), 16–21.
23. Baldwin, D.S., Anderson, I.M., Nutt, D.J., *et al.* (2005). Evidence-based guidelines for the pharmacological treatment of anxiety disorders: recommendations from the British Association for Psychopharmacology. *Journal of Psychopharmacology*, **19**, 567–96.
24. Üstün, T.B., Privett, M., Lecrubier, Y., *et al.* (1996). Form, frequency and burden of sleep problems in general health care: a report from the WHO collaborative study on psychological problems in general health care. *European Psychiatry*, **11**(Suppl. 1), 5S–10S.
25. Kamel, N. and Gammack, J.K. (2006). Insomnia in the elderly: cause, approach, and treatment. *The American Journal of Medicine*, **119**, 463–9.
26. Kupfer, D.J. and Reynolds, C.F. (1997). Management of insomnia. *The New England Journal of Medicine*, **336**, 341–6.
27. Nowell, P.D., Mazumdar, S., Buysse, D.J., *et al.* (1997). Benzodiazepines and zolpidem for chronic insomnia. A meta-analysis of treatment efficacy. *The Journal of the American Medical Association*, **278**, 2170–7.
28. Roehrs, T., Merlotti, L., Zorick, F., *et al.* (1992). Rebound insomnia and hypnotic self administration. *Psychopharmacology*, **107**, 480–4.
29. Johnson, M.W., Suess, P.E., and Griffiths, R.R. (2006). Ramelteon: a novel hypnotic lacking abuse liability and sedative adverse effects. *Archives of General Psychiatry*, **63**, 1149–57.
30. Fineberg, N. and Drummond, L.M. (1995). Anxiety disorders. Drug treatment or behavioural cognitive psychotherapy? *CNS Drugs*, **3**, 448–66.
31. Szabadi, E. (2006). Drugs for sleep disorders: mechanisms and therapeutic prospects. *British Journal of Clinical Pharmacology*, **61**, 761–6.
32. Tyrer, P. and Baldwin, D. (2006). Generalised anxiety disorder. *Lancet*, **368**, 2156–66.

6.2.3 Antidepressants

Zubin Bhagwagar and George R. Heninger

Introduction

Major depressive disorder is a serious, recurrent illness which levies a crippling toll on individuals, families, and society in general. The importance of depression as a major public health problem is emphasized by findings from the World Health Organization Global Burden of Disease survey in showing that in 1990 it was the fourth largest cause of burden of disease (i.e. years of life lost due either to premature mortality or to years lived with a disability). It has been estimated that by the year 2020 it is expected to be the second largest cause of burden of disease.[1] Depression is underdiagnosed and frequently under-treated, and depressed individuals have a much higher risk for suicide. The primary treatment for depression involves the use of antidepressant drugs, and it is therefore important that clinicians become familiar with and adept in utilizing this important group of compounds. Although primarily used for the treatment of depression, drugs within this category also have a number of other important uses. A thorough understanding of the pharmacology of antidepressants will aid the clinician in the selective use of these drugs for patients with depression as well as patients with a number of other disorders.

A brief history of antidepressant discovery and theories of action

Table 6.2.3.1 gives a brief chronology of antidepressant drug discoveries and theories of drug action. It is a comment on our understanding of the illness that some of the major advances in the pharmacotherapy of depression have been serendipitous. Prior to 1954, except for the use of electroconvulsive therapy, there were few effective drug treatments for depression. In 1954, the antidepressant era was initiated with the observation that some patients with tuberculosis displayed mood elevations following treatment with the antituberculosis agent iproniazid.(2) Following this initial serendipitous observation, the antidepressant effect of iproniazid was confirmed(3) and its action of inhibiting monoamine oxidase was reported. Iproniazid had significant toxicity and other monoamine oxidase inhibitors (MAOIs) were subsequently introduced. Independent from the work on MAOIs, imipramine, which has a chemical structure similar to the phenothiazines, was assessed as an agent to treat agitation in psychotic patients where it was found to be ineffective. However, it was noticed (again serendipitously) that imipramine produced an improvement of mood in the subset of patients who had symptoms of depression. Kuhn then reported in 1958 that imipramine was an effective antidepressant.(4)

One of the earliest theories of antidepressant drug action was that the antidepressant effect was produced by an increase of serotonin (5-hydroxytryptamine (**5-HT**)) in brain. This was supported by an initial study showing that an MAOI plus tryptophan, the precursor of 5-HT, was a more effective antidepressant treatment than an MAOI alone.[5] Subsequently, the discovery that imipramine and desipramine had effects in inhibiting the reuptake of noradrenaline (norepinephrine) and adrenaline (epinephrine) into the synapse led to the catecholamine theory of depression, which proposed that antidepressant treatments act by increasing the level of catecholamines at brain synapses.[6,7] Ten years later, it was reported that in laboratory animals most antidepressant treatments lead to downregulation of β-adrenergic receptors. This supported the proposal that antidepressants act by reducing β-adrenergic receptor sensitivity.[8] However, the reduction in β-adrenergic receptor sensitivity occurred within hours and antidepressant effect requires 1 to 3 weeks and futher not all effective antidepressant treatments produce reductions in β-adrenergic receptor sensitivity.

In the 1970s and 1980s, a large number of studies on antidepressants were conducted in laboratory animals which demonstrated that they produced a number of changes in monoamine receptor sensitivity.[9] In the late 1980s, a number of neurophysiological

Table 6.2.3.1 History of discovery of antidepressants and pharmacological theories of antidepressant drug action

Year	Discovery or theory	Reference
1954	Discovery that MAOIs have antidepressant effects	2, 3
1958	Discovery that the tricyclic drug imipramine is an effective antidepressant	4
1963	*Serotonin theory of depression:* MAOIs act by increasing serotonin and tryptamine in brain	5
1965	*Catecholamine theory of depression:* ADTs act by increasing cathecolamines in brain	6, 7
1975	β-*Adrenergic receptor theory of depression:* ADTs act by altering the sensitivity of several monoamine receptor subtypes in brain	8
1981	*Monoamine receptor sensitivity theory of depression:* ADTs act by altering the sensitivity of several monoamine receptor subtypes in brain	9
1987	*Serotonergic augmentation theory of depression:* ADTs act by decreasing sensitivity of presynaptic serotonergic autoreceptors and increasing sensitivity of serotonergic postsynaptic receptors to increase overall efficacy in serotonergic transmission	10
1996	*A molecular and cellular theory of depression:* ADTs act by producing a sustained activation of the CAMP system which increases brain levels of neurotrophic factors that reverse the effects of stres in certain brain areas	14
1998	Discovery that a substance P antagonist that does not interact with monoamine systems is as effective an antidepressant as an SSRI (paroxetine)	16
2000/2006	Demonstration of antidepressant properties of ketamine implicating the glutamatergic system in the pathophysiology of depression	19, 20

ADT, antidepressant treatment.

studies provided evidence that the delay in onset of antidepressant effects could be accounted for by a slow decrease in sensitivity at presynaptic serotonergic autoreceptors which has the overall result of increasing serotonergic function after days and weeks of treatment.[10] An elaboration on the receptor sensitivity theory was the discovery that most antidepressants produce alterations in the sensitivity of a specific glycine-sensitive site on the *N*-methyl-d-aspartate (NMDA) receptor.[11] A subsequent study showed that an NMDA antagonist may have antidepressant actions[12]and this line of thought has borne fruit recently in a possible novel mechanism of action for antidepressant treatment. An additional receptor sensitivity change thought to be important in the mechanism of action of antidepressants involved changes in the sensitivity of receptors for glucocorticoids. It was found that antidepressants produce an overall improvement of inhibitory feedback on the hypothalamic-pituitary-adrenal axis[13] and that specific corticotrophin releasing hormone (CRH) antagonists have antidepressant properties.[14]

A more recent theory of antidepressant drug action involves findings that antidepressant treatments affect intracellular pathways and neurotrophins. It was found that many antidepressants, in spite of β-adrenergic receptor downregulation, continue to produce sustained activation of the cAMP system and that this is related to increases of neurotrophic factors in brain.[15] Neurotrophins reverse the effects of stress in some brain areas and this raise the possibility that antidepressants act by increasing neurotrophins which reverse the effects of stress in important brain areas of depressed patients.

Throughout the 1980s, and 1990s a number of compounds that do not fit the standard monoamine theories of depression have been found to be effective clinical antidepressants. One of these drugs, tianeptine, actually increases the uptake of 5-HT into nerve endings, an effect that is opposite to the standard selective serotonin reuptake inhibitors (**SSRI**s).[16] Similarly, while there was intense interest in a report of possible antidepressant efficacy of a substance P receptor antagonist,[17] which does not interact with monoamine systems, clinical trials for this specific compound were disappointing.

Although no single mechanism has been discovered that will account for the antidepressant effects of all effective antidepressant treatments, it is clear that initial effects on monoamine metabolism with subsequent effects of intracellular pathways is important. While clinical wisdom and data suggest that there is a lag of 7–21 days to antidepressant action, recent reports question this notion of delayed onset of efficacy.[18] Recently ketamine, an NMDA antagonist has been shown to have an onset of action much faster than that traditionally seen with conventional antidepressants[19,20] suggesting that the pursuit of novel mechanisms may indeed result in advances in the pharmacotherapy of depression. Using preclinical models, it has been suggested that both nonselective NMDA antagonists as well as NR2B selective antagonists exert their antidepressant effects by regulating the functional interplay between AMPA and NMDA throughput.[21]

Pharmacology and types of compounds available

Antidepressant drugs fall into a wide variety of chemical classes and they have a wide range of neuropharmacological effects. They are grouped in Tables 6.2.3.2, 6.2.3.3, and 6.2.3.4 based on the presumed primary action that leads to an antidepressant effect. Table 6.2.3.2 lists the drugs that inhibit the uptake of the monoamines noradrenaline, 5-HT, and dopamine into nerve endings which in turn is thought to increase the function of the respective monoamine systems in brain. Table 6.2.3.3 lists the drugs that inhibit monoamine oxidase and thereby increase the concentration of many amines in brain. Table 6.2.3.4 lists the drugs with other primary actions that do not primarily involve inhibition of monoamine uptake or monoamine oxidase inhibition.

In Table 6.2.3.2, the first 12 compounds are inhibitors of noradrenaline uptake with a variable potency of inhibiting 5-HT uptake. The drugs with secondary amine structures, desipramine, nortriptyline, protriptyline, amoxapine, and maprotiline are predominantly noradrenaline uptake inhibitors with little effect on 5-HT uptake.[22] It can be seen in Table 6.2.3.2 that clomipramine, in addition to inhibiting noradrenaline uptake, is also a strong 5-HT uptake inhibitor. There are currently three selective serotonin and noradrenaline reuptake inhibitor (SNRI) drugs available; milnacipran (not licensed in the US), venlafaxine and duloxetine. Venlafaxine, inhibits both 5-HT and noradrenaline and 5-HT reuptake,[23] as do milnacipran and duloxetine though in varying proportions. While affinities vary depending on the system studied, milnacipran blocks 5-HT and norepinephrine reuptake with relatively equal affinity, while duloxetine has been suggested to have a slightly greater selectivity for 5-HT and venlafaxine a much greater selectivity for 5-HT.[24] Reboxetine (not licensed in the US) is a highly selective and potent inhibitor of noradrenaline reuptake.[25] It has only a weak effect on the 5-HT reuptake and does not affect the uptake of dopamine.

A key issue in the pharmacology of all antidepressant drugs is the relative specificity of their action. Drugs with a tertiary amine structure tend to produce more antagonism of α_1-adrenergic receptors which can produce hypotension, histamine receptors which can produce sedation, and muscarinic cholinergic receptors which can produce blurred vision, dry mouth, and urinary retention. This leads to more side-effects for these compounds than the drugs with a secondary amine structure. Venlafaxine has relatively less effect on these receptors and thus fewer side-effects[23] (see Table 6.2.3.6).

SSRIs are probably the most widely prescribed antidepressants and represent a class of drugs that selectively inhibit 5-HT reuptake from the synapse. Unlike the tricyclics, they each have different chemical structures. The drugs listed in Table 6.2.3.2 have a relatively specific effect in inhibiting 5-HT uptake,[22] and because of their relatively specific effect on this monoamine system and the lack of antagonism of many other receptors, they have been found to have fewer side-effects. Escitalopram was introduced following the discovery that all of the inhibitory activity of citalopram on 5-HT reuptake resides in the *S*-(+)-enantiomer (*S*-citalopram),[26] with *S*-citalopram being 167 times more potent than *R*-citalopram at inhibiting 5-HT reuptake into rat brain synaptosomes.

MAOIs are listed in Table 6.2.3.3. Two isozymes, monoamine oxidases A and B, are present in many discrete cell populations within the central nervous system, and glial cells also express monoamine oxidases A and B. The main substrates for monoamine oxidase A include adrenaline, noradrenaline, and 5-HT. The breakdown of dopamine in striatal regions of the brain is preferentially by monoamine oxidase B, but it can also be broken down by monoamine oxidase A. Since monoamine oxidase is located on the

Table 6.2.3.2 Pharmacological actions of antidepressants: drugs that inhibit monoamine reuptake at the synapse

Drug	Chemical class	Relative reuptake inhibition		
		Noradrenaline	5-Hydroxytryptamine	Dopamine
Imipramine	Tricyclic	++	+	0
Desipramine[a]	Tricyclic	++++	0	0
Amitriptyline[a]	Tricyclic	++	+	0
Nortriptyline[a]	Tricyclic	+++	0/+	0
Trimipramine	Tricyclic	+	0	0
Clomipramine	Tricyclic	+	+++	0
Protriptyline[a]	Tricyclic	++++	0	0
Doxepin	Tricyclic	++	0/+	0
Amoxapine[a]	Tricyclic	+++	0	+
Maprotiline[a]	Tetracyclic	+++	0	0
Venlafaxine	Bicyclic	+	++	0/+
Milnacipran	SNRI	+++	+++	0/+
Duloxetine	SNRI	+++	++	0/+
Reboxetine	NARI	++++	0	+
Fluoxetine	SSRI	0	+++	0
Sertraline	SSRI	0	++++	+
Fluvoxamine	SSRI	0	+++	0
Paroxetine	SSRI	+	++++	0
Citalopram	SSRI	0	++++	0
Escitalopram	SSRI	0	++++	0

0, None; 0/+, minimal; +, low; ++, moderate; +++, High; ++++, very high.

[a]Secondary amine.

outside of the plasma membrane of the mitochondria in neurones, it is not able to eliminate amines that are stored inside vesicles. MAOI produces an increase in monoamines in the cytoplasm. It is thought that the increase in monoamine content is the primary mechanism of action of MAOIs, and other secondary changes including β-adrenergic receptor downregulation and other receptor changes are secondary to the increased amine levels.[27]

Four of the six drugs listed in Table 6.2.3.3 are irreversible inhibitors. The two reversible inhibitors are essentially inert substrate analogues, and there is usually a correlation between their plasma concentration and the reversible inhibition of monoamine oxidase A. Since isocarboxazid, phenelzine, and tranylcypromine are irreversible inhibitors of monoamine oxidases A and B, there can be serious side-effects when foods that are high in tyramine or other amines are ingested. In addition, these three drugs have strong interactions with other drugs that alter monoamine

Table 6.2.3.3 Pharmacological actions of antidepressants: drugs that inhibit monoamine oxidase

Drug	Chemical class	MAO A	MAO B	Reversible
Isocarboxazid	Hydrazine	Yes	Yes	No
Phenelzine	Hydrazine	Yes	Yes	No
Tranylcypromine	Amphetamine	Yes	Yes	No
Moclobemide	Morpholine	Yes	No	Yes
Brofaromine	Piperidine	Yes	No	Yes
Selegiline	Phenethylamine	No	Yes[a]	No

MAO, Monoamine oxidase.

[a]Selective at lower doses; becomes non-selective at higher doses.

Table 6.2.3.4 Pharmacological actions of antidepressants: drugs that do not act by strong inhibition of monoamine uptake or inhibition of monoamine oxidase

Drug	Chemical class	Possible pharmacological action
Trazodone	Triazolopyridine	Mixed 5-HT agonist/antagonist
Nefazodone	Phenylpiperazine	Mixed 5-HT agonist/antagonist, weak monoamine uptake inhibitor
Bupropion	Unicyclic amino ketone	Weak noradrenaline and dopamine uptake inhibitor
Mianserin	Tetracyclic	Antagonist α_2-adrenergic auto- and heteroreceptors, increased 5-HT and noradrenaline release
Mirtazapine	Tetracyclic	Antagonist α_2-adrenergic auto and heteroreceptors, increased 5-HT and noradrenaline release

metabolism and therefore their use as antidepressants is much more limited than the tricyclics, SSRIs, or other antidepressant compounds. Tranylcypromine, which has a structure similar to amfetamine in addition to being an MAOI, is also thought to have a stimulant-type action of rapid onset. With the reversible MAOIs moclobemide and brofaromine, the recovery of monoamine oxidase back to normal levels after the drug is stopped is much shorter than with the irreversible MAOIs. These drugs increase concentrations of 5-HT, noradrenaline, and adrenaline that are short and parallel the time course of the monoamine oxidase A inhibition. These two drugs are more easily displaced by the pressor amines such as tyramine, and therefore, are thought to be safer than the irreversible inhibitors.

Selegiline, which has recently become available as a transdermal patch,[28] is selective at lower doses for monoamine oxidase B but at higher doses it becomes non-selective.[29] It has been primarily used for the treatment of Parkinson's disease and the doses for treating depression need to be much higher (note: selegiline is not licensed in the UK for depression). Since monoamine oxidase B is not involved in the intestinal tyramine interaction, selegiline interactions with ingested monoamines have been minimal.

In addition to inhibiting monoamine oxidase, these compounds have other effects on monoamine systems that can produce side-effects. However, the major concerns are the interactions with dietary amines and other drugs that influence amine function. The combination of dietary interactions and slow recovery of monoamine oxidase following with the irreversible inhibitors makes these drugs one of the more difficult treatments to administer. They are generally reserved for patients not otherwise responding to the other less toxic antidepressants.

In Table 6.2.3.4, compounds that are effective antidepressants but do not inhibit monoamine oxidase or have strong monoamine uptake inhibition are listed. Trazodone has shown receptor antagonist activity at several 5-HT receptor subtypes although its active metabolite *m*-chlorophenylpiperazine (mCPP) is a potent direct serotonin agonist. It is a weak but relatively selective inhibitor of 5-HT reuptake, is an antagonist at 5-HT1A and 5-HT2 receptors in addition to its active metabolite mCPP being a potent 5-HT agonist.[30] This leads to trazodone being classified as a mixed 5-HT agonist/antagonist. It also has relatively weak 5-HT uptake inhibiting properties but with no effect on noradrenaline or dopamine uptake. Trazodone is virtually devoid of anticholingeric activity and therefore it has few side-effects in this area. However, it does produce considerable sedation and hypotension secondary to antagonism of α_1-adrenergic receptors and histamine receptors.

Nefazodone is an analogue of trazodone that was developed to overcome the orthostatic hypotension and sedation caused by the latter. Like trazodone it is a 5-HT receptor antagonist with weak monoamine uptake inhibition activity.[31] It has less affinity for the α-adrenergic receptors and is inactive on many other receptors. It too is metabolized to *m*-chlorophenylpiperazine which is an active serotonergic agonist. Although the initial effects of nefazodone involve alterations of 5-HT neurotransmission, these effects are complex and depend on the biological test used.

Bupropion resulted from focussed research to find antidepressant compounds that would have fewer side-effects than traditional tricyclics (note: buproprion is not licensed in the UK for depression). Bupropion is a mild inhibitor of noradrenaline uptake, has some effects on inhibiting dopamine uptake but has no effect on 5-HT uptake.[32] These effects are not associated with β-adrenergic receptor downregulation as is seen with many other antidepressants. One of the active metabolites is hydroxybupropion which also has an antidepressant profile in laboratory animals. It is of interest that bupropion is one of the few drugs that reduce REM latency since most other treatments increase it. Although the specific mechanisms of bupropions antidepressant effects are not known, its unique profile has led to its use in the treatment of bipolar disorder[33] as well as its use in the treatment of smoking cessation.[34]

Mianserin and mirtazapine both have potent effects on antagonizing α_2-adrenergic auto- and heteroreceptors.[35] They also antagonize other 5-HT receptors but have minimal effects on monoamine uptake or monoamine oxidase activity. Since α_2 receptors inhibit noradrenaline release, their antagonism leads to an increase in noradrenaline release in many brain areas. In addition, antagonism of α_2-adrenergic heteroreceptors located on serotonergic neurones results in an enhanced 5-HT release. With mirtazapine, since 5-HT_2 and 5-HT_3 receptors are blocked, this could result in selective enhancement of 5-HT_1-receptor-mediated neurotransmission. These drugs have low affinity for muscarinic, cholinergic, and dopamine receptors and this is related to a reduced side-effect profile. The combination of increased noradrenaline release and increased 5-HT release resulting from the α_2-antagonism on auto- and heteroreceptors is hypothesized to be the central mechanism of action.

Pharmacokinetics

Data on the pharmacokinetics of antidepressants are listed in Table 6.2.3.5. The tricyclic antidepressants are by and large well absorbed although time to peak plasma concentration can vary from 1 to 12 h depending on the drug and the individual. In general, these drugs are metabolized in the liver to a variety of metabolites, some of which are active. For instance, desipramine is a metabolite of imipramine and nortriptyline is a metabolite of amitriptyline. Most of these compounds have a long half-life (close to 24 h) that will allow for once-daily dosing. All the compounds are highly bound to plasma protein except for venlafaxine and milnacipran. Although many of the compounds have active metabolites, the exact percentage of each metabolite in patients and in their clinical effects is still largely unknown.

There is considerable individual variation in the metabolism of tricyclics, and a large component of this may be genetic. Up to 7 to 9 per cent of the Caucasian population have been classified as slow metabolizers (slow hydroxylators) which can be measured by the rate of hydroxylation of debrisoquin. The slow hydoxylation has been determined to be caused by a polymorphism in a cytochrome P-450 macrosomal enzyme (CYP2D6). It is of interest that many SSRIs are inhibitors of P-450 isoenzymes which can considerably influence the metabolism of tricyclic antidepressants.[36] In general, the increased renal clearance in children and a decreased renal clearance with age need to be taken into account with dosing.

The SSRIs are rapidly absorbed, although there is variability within the drug half-lives. The metabolism into active metabolites can vary the pharmacodynamic effects considerably. For example, fluoxetine is metabolized to norfluoxetine which has similar activity on 5-HT reuptake as fluoxetine. The elimination half-life of norfluoxetine is longer (4–16 days) than that of fluoxetine (4–6 days). The desmethyl metabolite of sertraline although not nearly as

potent as the parent compound, also has a much longer half life. The desmethyl metabolite of citalopram or escitalopram, although a potent noradrenaline uptake inhibitor, is much lower in concentration than citalopram and it weakly crosses the blood-brain barrier. Fluvoxamine, paroxetine, duloxetine or milnacipran do not have any active metabolites. The relatively long half-lives of some of the SSRIs, particularly fluoxetine, require longer drug-free periods before switching to other classes of compounds especially before starting an MAOI.

The MAOIs are all rapidly absorbed. For the irreversible MAOIs, the elimination half-life and protein binding patterns are not as relevant because of the irreversible effects on monoamine oxidase. The reversible MAOIs have shorter half-lives and require multiple daily dosing.[29] With the irreversible MAOIs, once the drug is stopped, there needs to be time for new synthesis of monoamine oxidase. This requires a minimum of 5 to 7 days and the safest recommendation is to wait 2 weeks before starting other drugs that may interact with the MAOIs.

The five other antidepressants listed in Table 6.2.3.5 are rapidly absorbed but there is some variation in their elimination half-life. In general, the half-lives are short enough that multiple daily dosing is required. They are generally bound to plasma protein at a high level. The metabolites of trazodone and nefazodone have mixed effects on 5-HT receptors which results in a complex overall effect. Trazodone and nefazodone undergo extensive hepatic metabolism and one major metabolite is *m*-chlorophenylpiperazine which stimulates 5-HT receptors. Many metabolites have biological activity with half-lives different to the parent compounds.

Bupropion is metabolized in the liver and its metabolites can be at higher concentration than the parent compound. The relationship between plasma bupropion and clinical response has been poor.

Table 6.2.3.5 Pharmacokinetics of antidepressants

Drug	Absorption time to peak plasma concentration (h)	Elimination half-life (h)	Percentage plasma protein binding	Important metabolite
Monoamine reuptake inhibitors				
Imipramine	1.5–3	11–25	92	Desipramine
Desipramine	3–6	11–31	90	2-OH-desipramine
Amitriptyline	1–5	10–26	94	Nortriptyline
Nortrityline	3–12	18–44	92	10-OH-nortriptylene
Trimipramine	3	9–11	95	None
Clomipramine	2–6	21–31	97	Desmethylclomipramine
Protriptyline	6–12	67–89	93	None
Doxepin	1–4	11–23	80	Desmethyldoxepin
Amoxapine	1–2	8–30	90	8-OH-amoxapine
Maprotiline	4–12	28–58	88	Desmethylmaprotiline
Venlafaxine	2	5	30	O-desmethylvenlafaxine
Milnacipran	0.5–4	8	13	None
Duloxetine	6–10	8–17	95	None
Reboxetine	1.5–2.4	12–14	97	NA
SSRIs				
Fluoxetine	4–8	24–120	94	Norfluoxetine
Sertraline	6–8	27	99	n-Desmethylsertraline
Fluvoxamine	2–8	15–26	77	None
Paroxetine	5–7	24–31	95	None
Citalopram	1–6	33	80	NA (monodesmethylcitalopram)
Escitalopram	3–6	22–32	56	S-desmethylcitalopram
MAOIs				
Isocarboxazid	3–5	NA	NA	NA
Phenelzine	2–4	NA	NA	NA
Tranylcypromine	1.5–3	1.5–3.5	NA	NA
Moclobemide	1–1.5	1.4	NA	Numerous
Brofaromine	1–2	12–15	NA	n-Desmethylbrofaromine
Selegiline	1–3	2–10	NA	n-Desmethylselegiline
Other antidepressants				
Trazodone	1–2	6–11	92	m-Chlor ophenylpiperazine
Nefazodone	1	2–4	99	m-Chlorophenylpiperazine
Bupropion	3	10–21	85	Bupropion threoamino alcohol
Mianserin	2–3	15–22	NA	NA
Mirtazapine	2–3	20–40	85	None

NA, data not available

Side-effects

The history of new drugs becoming available for the treatment of depression reflects the efforts by the pharmaceutical industry to find compounds with reduced side-effects. This is in particularly important in the treatment of patients with medical illness because some of the side-effects can have considerable negative medical consequences. In Table 6.2.3.6, the propensity of the different drugs to produce some of the side-effects caused by antidepressants can be compared. The drugs that have high affinity for the α_1-adrenergic receptors can produce hypotension. Antagonism of histamine receptors has been associated with sedation and there is a long list of anticholinergic effects associated with antagonism of muscarinic cholinergic receptors.

For the tricyclics, it can be seen that drugs with a tertiary amine structure produce increased sedation. There is also an increase in the frequency of side-effects associated with antagonism of muscarinic cholinergic receptors such as dry mouth, constipation, blurred vision, urinary retention, dizziness, tachycardia, memory impairment, and at high and toxic doses, delirium. There is also an increased tendency for these same compounds to produce hypotension and to have unwanted cardiac effects that can lead to serious complications. In addition, the tertiary amines have a tendency to produce more weight gain than the secondary amines. The adverse effects have a particular impact on the tolerance of the patients to taking the medication. Most importantly the anticholinergic and cardiac effects can produce difficult complications in the elderly even leading to delirium when too high a dose is given. Amoxapine can cause extrapyramidal symptoms which are thought to be secondary to blocking dopamine receptors.[37] The most common adverse effects in patients taking reboxetine during clinical trials were insomnia, sweating, constipation, dry mouth, and urinary hesitancy compared with placebo, and the rates of nausea, diarrhea, and somnolence were lower compared with fluoxetine.[38] Nausea, dry mouth, dizziness, headache, somnolence, constipation, and fatigue were reported most frequently with duloxetine.[39]

The propensity to produce orthostatic hypotension is also a serious side-effect, particularly in the elderly. With the increased risk of falls and subsequent fractures in the elderly, this can be a serious health risk. A number of methods such as teaching patients to rise slowly from a supine position, tilting the bed upward, and maintenance of fluid uptake could help prevent this. However, other equally effective newer antidepressants produce much less of many of these side-effects, and they can be more safely used in the elderly.

Many of the drugs that are monoamine uptake inhibitors can cause cardiac conduction delays which may even lead to heart block in patients with pre-existing conditions. Severe overdose of these compounds can produce major and life-threatening cardiac arrhythmias. The secondary amines are generally thought to produce less cardiac effects than the tertiary amines. One of the characteristics of the SSRIs which has led to their widespread use is their low rate of side-effects. The pharmacological specificity of these compounds which bind to the 5-HT transporter, while not binding to the other neurotransmitter receptor types, results in their producing a therapeutic effect without many of the unwanted side-effects. In placebo-controlled trials the incidence of early discontinuation of SSRIs because of adverse events is intermediate between patients treated with placebo and patients treated with tricyclic antidepressants. Some of the symptoms reported with these compounds include agitation, anxiety, headache, sleep disturbance, and tremor. One of the more troublesome side-effects is sexual dysfunction especially anorgasmia. Less frequently, there are changes in appetite with nausea, dry mouth, sweating, and weight change. In general, these effects are less than those observed with the non-SSRI monoamine uptake inhibitors. The interaction of SSRIs with MAOIs to produce the serotonin syndrome is discussed under toxic effects below. Fluoxetine and sertraline induced higher rates of sedation as dosages are increased but in contrast, paroxetine produces a dose-dependent increase in arousal.

In contrast with the fewer side-effects produced by SSRIs and the other antidepressants that are not monoamine inhibitors, the MAOIs tend to produce frequent and often much more serious side-effects. Frequent side-effects include dizziness, headache, insomnia, dry mouth, blurred vision, nausea, constipation, forgetfulness, difficulty with urination, and weakness. There is also sexual dysfunction, including anorgasmia, impotence, delayed ejaculation, and decreased desire. Insomnia has also been reported. The original MAOIs iproniazid and isocarboxazid had a higher frequency of impairing liver function, but this is less with the other drugs. Pyridoxine deficiency has been reported and should be considered in evaluating side-effects. The largest problem with the MAOIs is the interactions with foods and with other drugs. Food interaction is much less of a problem with the reversible MAOIs moclobemide and brofaromine.[37–39]

Trazodone and nefazodone lack the anticholeringic side-effects of many of the tricyclic drugs. This makes them useful compounds in many medical conditions where this effect would be problematic. Trazodone has an acute sedative effect which is useful in the treatment of agitation, anxiety, and insomnia. However, this can be a troublesome side-effect when the patient performs tasks that require full alertness. Trazodone appears to have more propensity to produce orthostatic hypotension than nefazodone, possibility related to the degree of α_1-adrenergic receptor antagonism. Both trazodone and nefazodone, because of their lack of anticholinergic effects, have a low probability of producing difficulties in patients with cardiac illness. There is a slight tendency for weight gain but not nearly as strong as for some of the other antidepressants.[40] A relatively rare but important side-effect with trazodone is priapism. The risk for this side-effect is greatest during the early phase of treatment and the reporting of abnormal erectile function, including inappropriate or prolonged erections, should prompt quick discontinuation of trazodone treatment. Sexual side-effects have also been reported in women.

Bupropion has a very different side-effect profile than the conventional tricyclic antidepressants. It has no anticholinergic effects, is not sedating, and instead of weight gain, it suppresses appetite in some patients. In comparison to the SSRIs and trazodone and nefazodone, it also does not cause sexual dysfunction. There is no orthostatic hypotension, and bupropion does not produce cardiac side-effects. The possible stimulation of dopaminergic systems by bupropion can be related to its activating effects. This may be useful in patients with retardation but may exacerbate patients with agitation and insomnia. Bupropion can make tics in attention-deficit hyperactivity disorder and Tourette's disorder worse.[41] Patients have been described with bupropion-related

Table 6.2.3.6 Side-effects of antidepressants

Drug	Sedation	Anticholinergic effects	Hypotension	Cardiac effects	Weight gain
Monoamine reuptake inhibitors					
Imipramine	+++	++	++	+++	++
Desipramine	+	+	+	++	+
Amitriptyline	+++	++++	+++	+++	+++
Nortriptyline	+	+	+	++	+
Trimipramine	+++	+++	++	+++	++
Clomipramine	++	+++	++	+++	+
Protriptyline	0/+	++	+	+++	+
Doxepin	+++	++	+++	++	++
Amoxapine	+	++	+	++	+
Maprotiline	++	++	++	++	+
Venlafaxine	0/+	0/+	0	+	0
Milnacipran	0/+	+/++	0/+	+	0
Duloxetine	+	++	0/+	0/+	+
Reboxetine	0	+/++	0/+	0/+	0
SSRIs					
Fluoxetine	0/+	0	0	0	0
Sertraline	0	0	0	0	0
Fluvoxamine	0	0	0	0	0
Paroxetine	+	+	0	0	0
Citalopram	+	+	0/+	0/+	0
Escitalopram	+	+	0/+	0/+	0/+
MAOIs					
Isocarboxazid	+	+	+++	0	+
Phenelzine	+	0	+++	0	++
Tranylcypromine	+	0	++	0	0/+
Moclobemide	0	0	0	0	0
Brofaromine	0	0/+	0/+	0	0
Selegiline	0	0	+	0	0
Other antidepressants					
Trazodone	+++	0	++	0/+	+
Nefazodone	+	0	+	0/+	0/+
Bupropion	0	0	0	+	0
Mianserin	+++	0/+	0/+	+	+
Mirtazapine	+++	+	0/+	+	+

0, None; 0/+, 0ccasional; +, law; ++, moderate; +++, veryhigh.

psychosis which includes hallucinations and delusions. Psychotropic drugs also modulate seizure threshold and this needs to be carefully evaluated.[42] For example, a serious side-effect of bupropion that is rare but clinically important is the propensity to induce seizures in doses over 450 mg/day. Thus, bupropion should not be used at a dose higher than this and careful evaluation of history of seizures and other medical conditions or treatments that might lower seizure threshold should be evaluated in each patient.

Mianserin often produces drowsiness during the first weeks of treatment but has much less anticholinergic side-effects than other tricyclic antidepressants. It has less effects on producing hypotension and cardiac effects and there is only a low propensity for weight gain. Mirtazapine also has an increased amount of drowsiness and sedation. These side-effects are usually mild and transient. Mirtazapine has a low propensity to produce orthostatic hypotension or cardiac effects. There is a tendency for increased appetite and weight gain, however, which does not appear to be as severe as with tricyclics such as amitriptyline. Mianserin and mirtazapine have not been shown to produce high rates of sexual dysfunction as has been seen with trazodone and there is little evidence of lowering of the seizure threshold.

The side-effect profiles of the antidepressants are thought to relate to their respective effects on a variety of neurotransmitter systems. Clinicians should be aware of the profile of side-effects for each of the antidepressants they prescribe. The dose and duration of treatment interact with the intensity and type of side-effect and should be considered relative to antidepressant effects when evaluating, switching or stopping treatment. Although all antidepressant treatments can provoke switches into mania in vulnerable patients with bipolar disorder, it would appear that the MAOIs have a somewhat higher propensity to do this than the other compounds.

It is important that the nature of somatic and behavioural symptoms be carefully recorded before the onset of treatment so that the emergence of side-effects can be documented for the individual patient.

Toxic effects

There is ongoing concern recently regarding the issue of antidepressant use and suicide. The field has been grapplling with two inter-related issues: the possible risk of suicidal behaviour attributable to antidepressant treatment versus the potential decrease in suicidal behaviour afforded by antidepressant therapy. In 2004, there was an 18 per cent increase in adolescent suicides over the previous year.[43] This coincided with increased publicity about the relationship between antidepressant treatment and suicide risk in children and adolescents and a subsequent decline in antidepressant prescriptions. The Food and Drug Administration (FDA) has issued a black box warning to warn the public about the increased risk of suicidal thoughts and behaviour ('suicidality') in children and adolescents being treated with antidepressant medications.[44]

Often the most serious toxic effects are the result of overdose. Since depressed patients are at increased risk of suicide there is always the possibility that suicidally depressed patients will overdose on their antidepressants. This is a very serious consideration and should be carefully evaluated when prescribing antidepressants. The symptoms and course of events following acute antidepressant overdose are complex and can be confusing unless a clear history of overdose is obtained. With tricyclic antidepressants, restlessness and excitement are initially seen with possible myoclonus, and dystonia and seizures leading to the development of coma. Seriously compromized patients can have depressed respiration with hypoxia, depressed reflexes, hypertension, and hypothermia. With the antidepressants that have antimuscarinic activity, there can be strong anticholinergic effects with mydriasis, flushed skin, dry membranes, and tachycardia.

Antidepressant overdose can be life-threatening and patients should receive immediate emergency medical evaluation. The local poison control centre should be contacted in any case of suspected antidepressant overdose. Appropriate follow-up can include the use of activated charcoal to absorb the drug as well as other medical supportive measures. Different compounds have different probability of serious complications following an overdose and this is related to the amount ingested. However, drugs are often taken in combination, and it is difficult to know the exact composition and amount of the overdose.

Another serious toxic side-effect is the interaction of MAOIs and foods that are high in tyramine and other monoamines. Tyramine has both direct and indirect sympathomimetic actions, has a pressor action, and is present in a number of foodstuffs. It is normally broken down by the MAO enzymes and in the presence of a MAOI will increase in concentration. Some of the foods that should be restricted in the diet of patients taking MAOIs are listed in Table 6.2.3.7. The reaction usually develops 20 min to 1 h following ingestion of food and is characterized by nausea, apprehension, occasional chills, sweating, restlessness and hypotension with occipital headache, palpitations, and possibly vomiting. Neck stiffness, piloerection, dilated pupils, fever, and motor agitation are seen on examination. In severe forms, the reaction can lead to delirium, hyperpyrexia, cerebral hemorrhage, and death. The interaction of the irreversible MAOIs with certain dietary components leading to the hypertensive reaction is one of the most serious drawbacks to the use of these types of compounds. The reversible MAOIs moclobemide and brofaromine have not been found to interact with tyramine in the same fashion as the irreversible MAOIs. Thus, they have much less liability in terms of producing the hypertensive crisis seen with the irreversible MAOIs.[29] Phentolamine (5 mg) administered intravenously or nifedipine (onset of action 5 minutes), a calcium channel blocker, have been shown to be useful in the treatment of hypertensive reactions.

Serotonin syndrome is most often encountered when a MAOI is combined with an SSRI and there is an excess of 5-HT which overstimulates serotonin receptors.[45] This syndrome can manifest itself with sweating, diarrhoea, abdominal pain, fever, tachycardia, elevated blood pressure, myoclonus, hyper-reflexia, and with irritability and agitation. In its severe form, there can be severe hyperpyrexia, motor irritability, cardiovascular shock, and death. This toxic effect can result from the use of irreversible MAOIs and the addition of high amounts of tryptophan or other drugs that release serotonin in the brain. In addition, a common cause of this syndrome can result from the use of SSRIs and irreversible MAOIs concomitantly. Thus, it is strongly recommended that when MAOIs or SSRIs are utilized in sequence that the switch of treatment from one drug to the other has a minimum of a 14-day washout drug-free period before the second drug is started. In the case of discontinuing drugs with long half-lives such as fluoxetine an even longer period of up to 3 to 5 weeks may be necessary to safely avoid any possibility of producing the serotonin syndrome as a possible reaction to the drug combination.

The use of antidepressants during pregnancy is controversial and as always clinicians need to balance up the risks and benefits of treatments for individuals in this situation. Some recent studies have suggested that the use of SSRI medication in the perinatal period may be associated with adverse events like low birth weight and respiratory distress in the new born.[46, 47] However, this needs to be balanced against the fact that women at risk for depression may be at risk if not treated with antidepressants during pregnancy[48] and both these risks need to be balanced against each other.

Table 6.2.3.7 Dietary restrictions for patients on MAOIs

Aged cheeses	Liver	Raisins
American cheese[a]	Aged meats	Soy Sauce
Cottage cheese[a]	Canned meats	Ripe avocado
Yogurt[a]	Processed meats	Sauerkraut
Sour cream[a]	Meat extract	Licorice
Wine	Fermented foods	Chocolate[a]
Beer	Snails	Coffee[a]
Yeast extract	Anchovies	
Herring	Canned figs	
Sardines	Fava beans	

[a] Not over 50g daily.

The elderly are much more susceptible to toxic effects of antidepressants than younger individuals. In elderly patients, there may be other illnesses and the compensatory biological systems are not as resistant as in younger individuals. Mild toxic effects can be life threatening in the elderly. The newer antidepressants with fewer side-effects are the best drugs to use in the elderly.

There are a number of other toxic events that occur with less regularity. Isocarboxizid and phenelzine, since they are hydrazines, have some propensity to produce liver toxicity. Other much less frequent toxic events have occurred following antidepressants such as idiosyncratic individualized allergic reactions to the drug, suppression of the haematopoietic system, unusual dermatological reactions and hyponatremia.[49] There are reports of death in children receiving desipramine[50] though the cause of these deaths is unknown. Similarly, there is literature documenting increased incidence of gastric bleeding in association with SSRI treatment though there is a confounding effect of concomitant non-steroidal anti-inflammatory medication.[51]

Indications

Table 6.2.3.8 lists conditions where some antidepressant drugs have been found to be effective. Not all drugs are equally effective in each condition and very few clinical trials of the different compounds in each of the conditions have been conducted. Since the efficacy of antidepressant drugs is in part related to the dose administered and/or blood levels, it is difficult to be certain of the relative efficacy of one compound versus another when only single fixed doses are used. The expense and difficulty of multidose designs in comparing two treatments are extremely large and this is the main factor limiting comparisons of different drugs across the conditions listed in Table 6.2.3.8. In addition, the large number of compounds available would make this a very difficult task indeed. Another issue is that many of the drugs are only officially approved by the American Food and Drug Administration for use in depressed patients. Many of the indications listed in Table 6.2.3.8 are 'off-label' use of the medication. Since depression is the most prevalent illness, pharmaceutical companies have developed and brought forward drugs with depression as the primary indication. The expense of clinical trials to gain approval for other indications is high. Thus, for many of the conditions listed in Table 6.2.3.8, there is only fragmentary evidence for efficacy of some antidepressants and almost no data or comparable efficacy across drugs.

There are a number of different diagnostic approaches to depression as listed in Table 6.2.3.8. By and large, all of the drugs listed in Table 6.2.3.5 have been shown to be effective in the treatment of major depression. Most drugs have been studied in outpatient samples of patients with major depression. Their relative efficacy in the treatment of more severe conditions such as melancholia, psychotic depression, or bipolar depression remains limited. In addition, the relative efficacy of the different compounds as treatments for the depressive subtype, such as atypical depression, dysthymia, or secondary depression, has not been fully studied. There have been some reports that the MAOIs may be more effective in atypical depression[52] but not all studies have validated this. When depression in the elderly is under consideration, the side-effect profile for each drug becomes a much more relevant consideration when choosing a specific drug.

Anxiety disorders have considerable comorbidity with depression. Imipramine was initially found to be effective in the treatment of panic disorder and since then SSRIs have also been effective as well as MAOIs.[53] Clomipramine was found to be effective in obsessive-compulsive disorder though more recently SSRIs tend to be favoured as they generally have fewer side-effects. Depending on the studies, both SSRIs and MAOIs have been effective treatments in social phobia as well as some tricyclic drugs. In general anxiety disorder and post-traumatic stress disorder, antidepressants have also shown efficacy but not to the same extent as seen in panic disorder.[54]

It is of interest that some antidepressants have been effective in treating eating disorders. They are effective in bulimia nervosa[55] but not in anorexia nervosa.[56] The dose of fluoxetine to treat

Table 6.2.3.8 Clinical indications for antidepressant treatments

Clinical condition
Depression
Major depression
Melancholia
Psychotic depression
Bipolar depression
Atypical depression
Secondary depression
Dysthymia
Depression in elderly (pseudodementia)
Prevention of depression relapse
Anxiety disorders
Panic disorder
Obsessive–compulsive disorder
Social phobia
Generalized anxiety disorder
Post-traumatic stress disorder
Eating disorders
Bulimia nervosa
Obesity
Nausea with chemotherapy
Sleep disorders
Insomnia
Narcolepsy
Sleep apnea
Pain
Migraine headache
Atypical facial pain
Chronic pain syndromes
Diabetic neuropathy
Other disorders
Substance abuse
Alcoholism
Smoking cessation
Borderline personality disorder
Neurological disorders
Enuresis
Attention-deficit disorder
Premenstrual dysphoric disorder
Peptic ulcer
Urticaria pruritus
Premature ejaculation

bulimia nervosa is higher than the treatment of depression. The increase of weight seen following many antidepressants contrasts with some reports of the usefulness of SSRIs in the treatment of obesity.

In clinical practice, many clinicians have used trazodone as a night-time sedative. In the treatment of depression, sleep is one of the first symptoms to show improvement following initiation of most antidepressant treatments. Various reports of use of antidepressants in the treatment of narcolepsy and sleep apnoea have also been published.

Antidepressants have been effective in various pain syndromes. Since there is a wide range of the medical conditions producing pain, the results have been quite variable. In general the antidepressants have been able to reduce many of the painful symptoms as well as be effective in treating the secondary depression associated with chronic pain. However, they do not demonstrate the clear analgesic effect of drugs such as opioids.

Antidepressants have been reported to be effective in many other disorders including substance abuse, alcoholism, and smoking cessation.[57] In children with enuresis a dose of imipramine as low as 25 mg has been seen to be safe and effective. In both children and adults, imipramine, desipramine, bupropion, and nortriptyline have been effective in the treatment of attention-deficient disorder.[58] Antidepressants have found use in the treatment of premenstrual disorders,[59] and they are also useful in the treatment of several neurological disorders.

In general the indications and uses of a specific antidepressant in part depend on their side-effect profile and on the previously demonstrated efficacy. A major issue in the use of drugs to treat the large number of depressed patients with a comorbid medical condition is the careful choice of drug to minimize possible negative interactions with the medical disease.

Contraindications

The major contraindications in the use of antidepressants arise from the interaction of the pharmacological effects of antidepressant treatment with a comorbid condition of the patient or with diet or drug interactions. As mentioned above, the most serious contraindications arise from the use of irreversible MAOIs in patients taking other drugs or a diet that interacts and potentiates monoamine function resulting in a hypertensive crisis. A major contraindication is the use of MAOI in patients who receive anaesthesia.[59] Patients on MAOIs should carry a card for medical emergencies warning of drug interactions. Drugs that potentiate serotonin can interact with SSRIs to give the serotonin syndrome. The more relative contraindications involve the interaction of the side-effect profile of the antidepressant treatment with either the primary medical disease or with other medications that the patient may be taking. Another relative contraindication is the use of antidepressants during pregnancy and breast feeding though clearly there needs to be a risk benefit analysis of the use of the medication.

Drug interactions

Many patients may be taking other medications and many are prescribed more than one psychotrophic drug at a time. Because of this, it is important that clinicians are aware of drug-drug interactions. Drugs that impair the cytochrome P-450 microsomal enzyme system in the liver can interact with other drugs that are dependent on hepatic metabolism. For example, barbiturates and carbamazepine which induce hepatic enzymes can accelerate tricyclic metabolism and reduce steady state blood levels. Another anticonvulsant increasingly prescribed in the control of affective disorders, valproate, can reduce tricyclic drug clearance. Neuroleptics can elevate tricyclic blood levels which may be related to the impairment of the hydroxylation pathway for tricyclic metabolism. One of the more important drug-drug interactions involves the use of SSRIs and tricyclic drugs. This is related to the competitive inhibition of cytochrome by all of the SSRIs except fluvoxamine. This can result in clear elevations of steady state plasma concentrations. If combinations like this are utilized, the tricyclic doses need to be reduced. The utilization of a drug where plasma concentrations can be monitored (see Table 6.2.3.10) would help in the adjustment of dose if the tricyclic is combined with an SSRI.

One of the more serious drug-drug interactions previously mentioned is the interaction of tricyclic drugs with MAOIs. This can lead to hypertensive reactions and possibly stroke as well as possible induction of the serotonin syndrome.[45] Often antidepressants are combined with phenothiazines. There is some evidence that chlorpromazine can block the metabolism of tricyclics and thus when these two treatments are combined a possible reduction in the tricyclic treatment may be required. Other drugs that have been shown to increase tricyclic levels through blocking their metabolism include cimetidine, methylphenidate, and haloperidol. Tricyclic drugs can reduce the effects of clonidine and guanethidine in reducing blood pressure; an anticonvulsant, phenytoin, may be elevated; and the drug warfarin may be increased following tricyclic drugs. With the SSRIs, since there is a narrow pharmacological effect, interactions with anticholinergic agents or antihistaminics or alcohol are generally less than the tricyclics. The one major interaction is through the cytochrome P-450 family of enzymes which are inhibited by most SSRIs and interact with the metabolism of other drugs.

Some of the more serious drug interactions occur with the MAOIs. In addition to the dietary interactions, the MAOIs can interact with many of the 'over-the-counter' medications such as cough syrups and decongestants. Table 6.2.3.9 lists a number of compounds that have adverse drug interactions with MAOIs. Certainly many of the drugs that are direct or indirect adrenergic and dopaminergic agonists can produce overstimulation of the sympathetic nervous system. This can result in increased blood pressure and possibly adverse effects on the central nervous system. These drugs include all of the sympathomimetics, amphetamines, methylphenidate, and other stimulants. This can also occur with drugs such as other MAOIs and tricyclics or SSRIs that increase monoamine levels. MAOIs may worsen hypoglycaemia and require readjustment of the dosage of hypoglycaemic agents. Major concerns arise when patients on MAOIs need surgery because of the interaction with a number of compounds used in anaesthesia. This is more likely to occur with the use of pethidine. Careful consideration should be given to using a minimal 2-week washout for patients on MAOIs under going elective surgery.

The mixed 5-HT agonist drugs also have important drug-drug interactions. Trazodone can potentiate barbiturate and alcohol and it can increase drowsiness and sedation in patients taking these

Table 6.2.3.9 Adverse drug interactions with MAOIs

Other antidepressants	Other MAOIs
Buspirone	Carbamazepine
Stimulants	L-Dopa
Sympathomimetics	Methyldopa
Dopamine	Guanethidine
Amphetamines	Dextromethorphan
Methylphenidate	Pethidine
Adrenaline	Cocaine
Asthma inhalants	Reserpine
Decongestants	Tryptophan
Appetite suppressants	Fenfluramine

agents. It has also been reported to produce the serotonin syndrome at times when combined with an SSRI and possibly buspirone. Trazodone has altered the kinetics of benzodiazepines including alprazolam and triazolam. Trazodone interacts with cytochrome P-450 enzymes and has been shown to increase plasma levels of digoxin. Unlike trazodone, nefazodone does not appear to potentiate the sedative effects of alcohol.

Bupropion undergoes hepatic metabolism and its levels can be altered by other drugs effecting this metabolic route. There is some dopaminergic activation with bupropion and it has had adverse interactions with MAOIs. Because of the dopaminergic activity, there have been interactions with anti-parkinsonian medication. Because bupropion lowers the seizure threshold at higher dosages it can interact and with other medications that would have similar effects to produce seizures. A combination of bupropion and lithium may increase the likelihood of seizures. There are some reports that carbamazepine may decrease bupropion drug levels.

Mirtazapine is metabolized by cytochrome P-450 enzyme systems. There is the potential for interaction with other drugs via this system. The extent of use of mirtazapine is not as great as the older drugs and the drug-drug interactions are not as extensively reported.

In general, a large number of drug-drug interactions have been reported for the antidepressants. The drug-drug interactions can be quite variable depending on the patient and the dosage and duration of treatments. Thus, the adverse drug-drug interactions are one of the main reasons for the recommendations to use monotherapy rather than more than one drug. The use of drug combinations should, on average, be restricted to patients who have a poor response to a single treatment because of the possibility of adverse drug-drug interactions.

Drug withdrawal

Long-term administration of a drug can produce adaptive changes in many aspects of the human biology. When the drug is abruptly discontinued, 'rebound' drug withdrawal symptoms can be observed. This is most clearly seen with longer-term opiate, benzodiazepine, or barbiturate use. Antidepressants are not addictive and dependence does not develop. With the antidepressants some degree of tolerance to the sedative and autonomic effects tends to develop and on abrupt withdrawal patients can have emerging symptoms consisting of malaise, dizziness, nausea, diarrhoea, chills, insomnia, restlessness, and muscle aches. Symptoms emerging during drug withdrawal have been seen following treatment with tricyclics as well as SSRIs.[60] They have been described on occasion for MAOIs and the other non-monoamine uptake antidepressants. One main factor is the drug elimination half-life. Abrupt discontinuation of drugs with a short elimination half-life will produce more emergent side-effects than drugs with a long half-life. Thus, it has been found that patients taking shorter half-life drugs such as paroxetine and sertraline have more of an emergent symptom increase than patients taking the longer elimination half-life drug fluoxetine. Therefore, the dose of drugs with a shorter elimination half-life should generally be tapered over a 2- to 3-week period when being discontinued rather than being stopped abruptly. Drugs with a longer elimination half-life can be stopped abruptly since the parent drug and metabolite may last for many days. Clinicians must be aware that slow elimination means that parent drug and active metabolites remain in the body for up to several weeks.

With the irreversible MAOI inhibitors, since there is a 1- or 2-week period during with monoamine oxidase must recover following discontinuation of the MAOI, emergent side-effects have not been as regularly observed. In general with the other antidepressants the withdrawal syndromes have not been permanent. Clinicians must make special efforts to discriminate between the return of symptoms and the emergence of new symptoms related to drug withdrawal.

Dosages and administration

In Table 6.2.3.10 some of the suggested optimal plasma concentrations for different antidepressant drugs are listed. For nortriptyline, desipramine, imipramine, and amitriptyline there is some evidence for a minimal plasma and concentration necessary for clinical response. An established therapeutic range is available for nortriptyline. Thus, patients with nortriptyline concentrations between 50 and 150 mg/ml seem to do much better. With the other drugs, it is generally thought that the plasma levels reflect a minimal threshold of plasma concentration for clinical response. Below this level patients are less likely to respond and the upper limit indicates that there is increased possibility of the systemic or cardiac toxicity. Bupropion levels between 50 and 100 ng/ml may possibly be the best range. However, it is important that the dose be kept below 450 mg/day because of the possibility of seizures.

Except for nortriptyline and the use of plasma concentrations to obtain a minimal effective level, it is generally the patient's clinical response that dictates dosage adjustments. One difficulty is that some patients with plasma concentrations outside the therapeutic range do respond and many patients with concentrations within the therapeutic range do not. Thus, the dosage needs to be adjusted depending on the individual patient's response. Clearly plasma concentration monitoring can be helpful in many situations such as evaluating plasma levels when higher than standard doses are used, assessing toxicity, use in elderly patients or patients with comorbid conditions to evaluate possible drug interactions, or where compliance is questioned. Blood for drug levels is usually obtained for plasma levels during elimination phase which is usually in the morning 12 h after the last dose.

Table 6.2.3.10 Dosage and administration of antidepressant drugs. Doses in brackets refer to doses recommended for elderly patients

Drug	Initial dose (mg/day)	Theraupeutic dose range (mg/day)	Recommended optimal plasma concentration (ng/ml)
Monoamine reuptake inhibitors			
Imipramine	25–75 (10)	150–300 (30–50)	200–300[d]
Desipramine	50–75	100–300 (25–150)	125–300[d]
Amitriptyline	75 (30–75)	150–200	120–250[d]
Nortriptyline	25	75–150	50–150
Trimipramine	50–75 (30–75)	150–300 (75–150)	NA
Clomipramine	10	30–250 (30–75)	NA
Protriptyline	20–40 (15)	20–60	70–240
Doxepin	75 (10–50)	30–300 (30–50)	110–250[d]
Amoxapine	100–150 (50–75)	100–600 (150–300)	200–400[e]
Maprotiline	25–75 (25)	25–225 (50–75)	200–300[d]
Venlafaxine	75	150–375	NA
Milnacipran	50	50–100	NA
Duloxetine	60	60	NA
Reboxetine	8	10–12	NA
SSRIs			
Fluoxetine	20	20–60	NA
Sertraline	50	50–200	NA
Fluvoxamine	50–100	100–300	NA
Paroxetine	20	50 (40)	NA
Citalopram	20	20–60 (20–40)	NA
Escitalopram	10 (5)	10–20 (10–20)	NA
MAOIs			
Isocarboxazid	30	10–60 (5–10)	NA
Phenelzine	45	15–90	NA
Tranylcypromine	20	10–30	NA
Moclobemide	300	150–600	NA
Brofaromine	50	50–150	NA
Selegiline	6[a]	6–12[a]	NA
Other antidepressants			
Trazodone	100 (150)	150–600	NA
Nefazodone	200 (100)[b]	200–600[b]	NA
Buproprion	200[c]	300–450[c]	50–100
Mianserin	30–40 (30)	30–90	NA
Mirtazapine	15	45	NA

NA, data not available.

[a] Selegiline not licensed in UK for depression. Doses quoted refer to transdermal patches marketed in the US.

[b] Nefazodone not licensed in UK.

[c] Buproprion not licensed in Uk for depression.

[d] Parent drug plus demethylated metabolite.

[e] Parent drug plus hydroxymetabolite.

These doses should not be used for patient prescriptions; clinicians should consult manufacturer's literature for recommendation of doses and frequency of administration.

Acknowledgement

The authors would like to thank Aybala Saricicek MD for her assistance in the preparation of the chapter.

Further information

Biological Psychiatry (Elsevier)—the journal of the Society of Biological Psychiatry http://www.sobp.org/

Neuropsychopharmacology (Nature Publishing group)—the journal of the American College of Neuropsychopharmacology http://www.acnp.org/

Schatzberg, A.F. and Nemeroff, C. B. (eds). (2004). *American Psychiatric Publishing Textbook of Psychopharmacology* (3rd edn.) American Psychiatric Publishing, Arlington, VA.

References

1. Lopez, A.D., Murray, C.C.J.L. (*1*998). The global burden of disease, 1990-2020. *Nature Medicine,* **4**, 1241–3.
2. Bloch, R.G., Doonief, A.S., Buchberg, A.S., *et al.* (1954). The clinical effect of isoniazid and iproniazid in the treatment of pulmonary tuberculosis. *Annals of Internal Medicine,* **40**, 881–900.
3. Crane, G.E. (1957). Iproniazid (marsilid) phosphate, a therapeutic agent for mental disorders and debilitating diseases. *Psychiatric research reports,* **135**, 142–52.
4. Kuhn, R. (1958). The treatment of depressive states with G 22355 (imipramine hydrochloride). *American Journal of Psychiatry,* **15**, 459–64.
5. Coppen, A., Shaw, D.M., Farrell, J.P. (1963). Potentiation of the antidepressive effect of a monoamine-oxidase inhibitor by tryptophan. *Lancet*, **1**, 79–81.

6. Schildkraut, J.J. (1965). The catecholamine hypothesis of affective disorders: a review of supporting evidence. *American Journal of Psychiatry,* **122**, 509–22.
7. Bunney, W.E., Jr., Davis, J.M. (1965). Norepinephrine in depressive reactions. A review. *Archives of General Psychiatry,* **13**, 483–94.
8. Vetulani, J., Sulser, F. (1975). Action of various antidepressant treatments reduces reactivity of noradrenergic cyclic AMP-generating system in limbic forebrain. *Nature,* **257**, 495–6.
9. Charney, D.S., Menkes, D.B., Heninger, G.R. (1981). Receptor sensitivity and the mechanism of action of antidepressant treatment. Implications for the etiology and therapy of depression. *Archives of General Psychiatry,* **38**, 1160–80.
10. Blier, P., de Montigny, C., Chaput, Y. (1987). Modifications of the serotonin system by antidepressant treatments: implications for the therapeutic response in major depression. *Journal of Clinical Psychopharmacology,* **7**, 24S–35S.
11. Paul, I.A., Nowak, G., Layer, R.T., *et al.* (1994). Adaptation of the N-methyl-D-aspartate receptor complex following chronic antidepressant treatments. *Journal of Pharmacology and Experimental Therapeutics,* **269**, 95–102.
12. Layer, R.T., Popik, P., Olds, T., *et al.* (1995). Antidepressant-like actions of the polyamine site NMDA antagonist, eliprodil (SL-82.0715). *Pharmacology, Biochemistry, and Behaviour*, **52**, 621–7.
13. Holsboer, F., Barden, N. (1996). Antidepressants and hypothalamic-pituitaryadrenocortical regulation. *Endocrine Reviews*, **17**, 187–205.
14. Griebel, G., Perrault, G., Sanger, D.J. (1998). Characterization of the behavioural profile of the non-peptide CRF receptor antagonist CP-154,526 in anxiety models in rodents. Comparison with diazepam and buspirone. *Psychopharmacology (Berl),* **138**, 55–66.
15. Duman, R.S., Heninger, G.R., Nestler, E.J. (1997). A molecular and cellular theory of depression. *Archives of General Psychiatry,* **54**, 597–606.
16. Wilde, M.I., Benfi eld, P. Tianeptine. (1995). A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic effi cacy in depression and coexisting anxiety and depression. *Drugs,* **49**, 411–39.
17. Kramer, M.S., Cutler, N., Feighner, J., *et al.* (1998). Distinct mechanism for antidepressant activity by blockade of central substance P receptors. *Science,* **281**, 1640–5.
18. Taylor, M.J., Freemantle, N., Geddes, J.R., *et al.* (2006). Early Onset of Selective Serotonin Reuptake Inhibitor Antidepressant Action: Systematic Review and Meta-analysis. *Archives of General Psychiatry,* **63**, 1217–23.
19. Berman, R.M., Cappiello, A., Anand, A. *et al.* (2000). Antidepressant effects of ketamine in depressed patients. *Biological Psychiatry,* **47**, 351–4.
20. Zarate, C.A., Jr., Singh, J.B., Carlson, P.J. *et al.* (2006). A Randomized Trial of an N-methyl-D-aspartate Antagonist in Treatment-Resistant Major Depression. *Archives of General Psychiatry,* **63**, 856–64.
21. Maeng, S., Zarate, C.A., Jr., Du, J. *et al.* (207). Cellular Mechanisms Underlying the Antidepressant Effects of Ketamine: Role of alpha-Amino-3-Hydroxy-5-Methylisoxazole-4-Propionic Acid Receptors. *Biological Psychiatry,* **63**, 349–52.
22. Richelson, E. (2003). Interactions of antidepressants with neurotransmitter transporters and receptors and their clinical relevance. *Journal of Clinical Psychiatry,* **64** Suppl **13**, 5–12.
23. Gutierrez, M.A., Stimmel, G.L., Aiso, J.Y. (2003). Venlafaxine: a 2003 update. *Clinical Therapeutics,* **25**, 2138–54.
24. Stahl, S.M., Grady, M.M., Moret, C., *et al.* (2005). SNRIs: their pharmacology, clinical efficacy, and tolerability in comparison with other classes of antidepressants. *CNS Spectrums,* **10**, 732–47.
25. Preskorn, S.H. (2004). Reboxetine: a norepinephrine selective reuptake pump inhibitor. *Journal of Psychiatric Practice,* **10**, 57–63.
26. Hyttel, J., Bogeso, K.P., Perregaard, J., *et al.* (1992). The pharmacological effect of citalopram residues in the (S)-(+)-enantiomer. *Journal of Neural Transmission. General Section,* **88**, 157–60.
27. Kennedy, S.H. (1997). Continuation and maintenance treatments in major depression: the neglected role of monoamine oxidase inhibitors. *Journal of Psychiatry & Neuroscience,* **22**, 127–31.
28. Patkar, A.A., Pae, C.U., Masand, P.S. (2006). Transdermal selegiline: the new generation of monoamine oxidase inhibitors. *CNS Spectrums,* **11**, 363–75.
29. Robinson, D.S. (2002). Monoamine oxidase inhibitors: a new generation. *Psychopharmacology Bulletin,* **36**, 124–38.
30. Haria, M., Fitton, A., McTavish, D. (1994). Trazodone: A review of its pharmacology, therapeutic use in depression and therapeutic potential in other disorders. *Drugs & Aging,* **4**, 331–55.
31. Dunner, D.L., Laird, L.K., Zajecka, J., *et al.* (2002). Six-year perspectives on the safety and tolerability of nefazodone. *Journal of Clinical Psychiatry,* 63 Suppl **1**, 32–41.
32. Ascher, J.A., Cole, J.O., Colin, J.N. *et al.* (1995). Bupropion: a review of its mechanism of antidepressant activity. *Journal of Clinical Psychiatry,* **56**, 395–401.
33. Leverich, G.S., Altshuler, L.L., Frye, M.A. *et al.* (2006). Risk of Switch in Mood Polarity to Hypomania or Mania in Patients With Bipolar Depression During Acute and Continuation Trials of Venlafaxine, Sertraline, and Bupropion as Adjuncts to Mood Stabilizers. *American Journal of Psychiatry,* **163**, 232–9.
34. Hurt, R.D., Sachs, D.P., Glover, E.D. *et al.* (1997). A comparison of sustained-release bupropion and placebo for smoking cessation. *New England Journal of Medicine,* **337**, 1195–202.
35. Szegedi, A., Schwertfeger, N. (2005). Mirtazapine: a review of its clinical efficacy and tolerability. *Expert Opinion on Pharmacotheraphy,* **6**, 631–41.
36. Hemeryck, A., Belpaire, F.M. (2002). Selective serotonin reuptake inhibitors and cytochrome P-450 mediated drug-drug interactions: an update. *Current Drug Metabolism,* **3**,13–37.
37. Kapur, S., Cho, R., Jones, C., *et al.* (1999). Is amoxapine an atypical antipsychotic? Positron-emission tomography investigation of its dopamine2 and serotonin2 occupancy. *Biological Psychiatry,* **45**, 1217–20.
38. Scates, A.C., Doraiswamy, P.M. (2000). Reboxetine: a selective norepinephrine reuptake inhibitor for the treatment of depression. *Annals of Pharmacotheraphy,* **34**, 1302–12.
39. Detke, M.J., Lu, Y., Goldstein, D.J., *et al.* (2002). Duloxetine, 60 mg once daily, for major depressive disorder: a randomized double-blind placebo-controlled trial. *Journal of Clinical Psychiatry,* **63**, 308–15.
40. Sussman, N., Ginsberg, D.L., Bikoff, J. (2001). Effects of nefazodone on body weight: a pooled analysis of selective serotonin reuptake inhibitor- and imipramine-controlled trials. *Journal of Clinical Psychiatry,* **62**, 256–60.
41. Spencer, T., Biederman, J., Steingard, R., *et al.* (1993). Bupropion exacerbates tics in children with attention-deficit hyperactivity disorder and Tourette's syndrome. *Journal of the American Academy of Child Adolescent Psychiatry,* **32**, 211–4.
42. Alper, K., Schwartz, K.A., Kolts, R.L., *et al.* (2007). Seizure Incidence in Psychopharmacological Clinical Trials: An Analysis of Food and Drug Administration (FDA) Summary Basis of Approval Reports. *Biological Psychiatry*, **62**, 345–54.
43. Hamilton, B.E., Minino, A.M., Martin, *et al.* (2007). Annual summary of vital statistics: 2005. *Pediatrics,* **119**, 345–60.
44. Hammad, T.A., Laughren, T., Racoosin, J. (2006). Suicidality in Pediatric Patients Treated With Antidepressant Drugs. *Archives of General Psychiatry,* **63**, 332–9.
45. Boyer, E.W., Shannon, M. (2005). The serotonin syndrome. *New England Journal of Medicine, 352*, 1112–20.
46. Chambers, C.D., Hernandez-Diaz, S., Van Marter, L.J. *et al.* (2006). Selective serotonin-reuptake inhibitors and risk of persistent pulmonary hypertension of the newborn. *New England Journal of Medicine,* **354**, 579–87.

47. Oberlander, T.F., Warburton, W., Misri, S., *et al.* (2006). Neonatal Outcomes After Prenatal Exposure to Selective Serotonin Reuptake Inhibitor Antidepressants and Maternal Depression Using Population-Based Linked Health Data. *Archives of General Psychiatry,* **63**, 898–906.
48. Cohen, L.S., Altshuler, L.L., Harlow, B.L. *et al.* (2006). Relapse of Major Depression During Pregnancy in Women Who Maintain or Discontinue Antidepressant Treatment. *Journal of the American Medical Association,* **295**: 499–507.
49. Madhusoodanan, S., Bogunovic, O.J., Moise, D., *et al.* (2002). Hyponatraemia associated with psychotropic medications. A review of the literature and spontaneous reports. *Adverse Drug Reactions and Toxicological Reviews,* **21**, 17–29.
50. Amitai, Y., Frischer, H. (2006). Excess fatality from desipramine in children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry,* **45**, 54–60.
51. Yuan, Y., Tsoi, K., Hunt, R.H. (2006). Selective serotonin reuptake inhibitors and risk of upper GI bleeding: confusion or confounding? *American Journal of Medicine,* **119**, 719–27.
52. Quitkin, F.M., Stewart, J.W., McGrath, P.J., *et al.* (1993). Columbia atypical depression. A subgroup of depressives with better response to MAOI than to tricyclic antidepressants or placebo. *British Journal Psychiatry Supplement,* 30–4.
53. Otto, M.W., Tuby, K.S., Gould, R.A., *et al.* (2001). An effect-size analysis of the relative efficacy and tolerability of serotonin selective reuptake inhibitors for panic disorder. *American Journal of Psychiatry,* **158**, 1989–92.
54. Ball, S.G., Kuhn, A., Wall, D., *et al.* (2005). Selective serotonin reuptake inhibitor treatment for generalized anxiety disorder: a double-blind, prospective comparison between paroxetine and sertraline. *Journal of Clinical Psychiatry,* **66**, 94–9.
55. Group FBNCS. (1992). Fluoxetine in the treatment of bulimia nervosa. A multicenter, placebo-controlled, double-blind trial. Fluoxetine Bulimia Nervosa Collaborative Study Group. *Archives of General Psychiatry,* **49**, 139–47.
56. Attia, E., Haiman, C., Walsh, B.T., *et al.* (1998). Does fluoxetine augment the inpatient treatment of anorexia nervosa? *American Journal of Psychiatry,* **155**, 548–51.
57. Nunes, E.V., Levin, F.R. (2004). Treatment of depression in patients with alcohol or other drug dependence: a meta-analysis. *Journal of the American Medical Association,* **291**, 1887–96.
58. Culpepper, L. (2006). Primary care treatment of attention-deficit/hyperactivity disorder. *Journal of Clinical Psychiatry,* **67 Suppl 8**, 51–8.
59. Yonkers, K.A. (2004). Management strategies for PMS/PMDD. *Journal of Family Practice,* Suppl: S15–20.
60. Ditto, K.E. (2003). SSRI discontinuation syndrome. Awareness as an approach to prevention. *Postgraduate Medicine*, **114**, 79–84.

6.2.4 Lithium and related mood stabilizers

Robert M. Post

Introduction

Global categorization of the mood stabilizers

Lithium is the paradigmatic mood stabilizer. It is effective in the acute and prophylactic treatment of both mania and, to a lesser magnitude, depression. These characteristics are generally paralleled by the widely accepted anticonvulsant mood stabilizers valproate, carbamazepine (Table 6.2.4.1), and potentially by the less well studied putative mood stabilizers oxcarbazepine, zonisamide, and the dihydropyridine L-type calcium channel blocker nimodipine. In contrast, lamotrigine has a profile of better antidepressant effects acutely and prophylactically than antimanic effects.

Differential responsivity among individual patients

Having grouped lithium, valproate, and carbamazepine together, it is important to note they have subtle differences in their therapeutic profiles and differential clinical predictors of response (Table 6.2.4.1). **Response to one of these agents is not predictive of either a positive or negative response to the others.**[1,2] Thus, clinicians are left with only rough estimates and guesses about which drug may be preferentially effective in which patients. Only sequential clinical trials of agents either alone or in combination can verify responsivity in an individual patient.[3] **Individual response trumps FDA-approval**.

Requirements for method of longitudinal assessment

Given this clinical conundrum, it is advisable that patients, family members, clinicians, or others carefully rate patients on a longitudinal scale in order to most carefully assess responses and side effects. These are available from the Depression Bipolar Support Alliance (DBSA), the STEP-BD NIMH Network, or www.bipolarnetworknews.org and are highly recommended.

Increasing need for complex combination treatment

The importance of careful longitudinal documentation of symptoms and side effects is highlighted by the increasing use of multiple drugs in combination.[2] This is often required because patients may delay treatment-seeking until after many episodes, and very different patterns and frequencies of depressions, manias, mixed states, as well as multiple comorbidities may be present. Treating patients to the new accepted **goal of remission** of their mood and other anxillary symptoms usually requires use of several medications. If each component of the regimen is kept below an individual's side-effects threshold, judicious use of multiple agents can reduce rather than increase the overall side-effect burden.

Patients and family education is a must

There is increasing evidence of reliable abnormalities of biochemistry, function, and anatomy in the brains of patients with bipolar disorder, and some of these are directly related to either duration of illness or number of episodes.[4,5] Therefore, as treatment resistance to most therapeutic agents is related to number of prior episodes, and brain abnormalities may also increase as well, it behooves the patient to begin and sustain acute and **long-term treatment as early as possible**.

Early age of onset and treatment delay are related to an adverse outcome in adulthood

Despite the above academic, personal, and public health recommendations, bipolar disorder often takes ten years or more to diagnose and, hence, treat properly. In fact, a **younger age** of onset is highly related to presence of a **longer delay** from illness onset to **first treatment, and** as well, to **a poorer outcome** assessed both retrospectively and prospectively.[6]

Table 6.2.4.1 Neuroprotective effects of lithium in cultured cells and animal models of diseases

I. Therapeutic Spectrum	Lithium	Carbamazepine	Valproate	Lamotrigine	Nimodipine
A Acute Mania	++++	++++	++++	0	(++)
B. Mania Prophylaxis	++++	+++	++++	+	(++)
C. Acute Depression	++	++	++	+++	(+)
D. Depression Prophylaxis	++++	++++	+++	++++	(++)
II. Correlates of Response					
A. Family History + = positive – = negative	+ Mania –Depression	– Bipolar Disorder	ND	+ Anxiety Disorder	ND
B. Bipolar Type (BP)	I	II	I,II	I,II	I,II,NOS
Manic Type: Pattern:	Euphoric Intermittent	Dysphoric Continuous	Dysphoric Non-accelerating	Continuous	Euphoric Ultradian
C. Comorbidities					
Anxiety Disorder	None	++	+++	+++	(+)
Substance Use	None	++	+++	+/–	(+)
D. PTSD Utility	0	++	++	++	0
E. Other Unique Targets	M-D-I vs. D-M-I Antisuicidal AntimedicalMortality	Alcohol Withdrawal Tigeminal Neuralgia	Alcohol Abstinence Migraine Prophylaxis		(Alzheimer's Dementia) (Migraine Prophylaxis) Subarachnoid Hemorrhage
F. Baseline PET Activity	?	Hyper- metabolism	?	Hypo-metabolism	Hypo-metabolism
G. CSF SRIF (Somatostatin)					
Predicts Response:	?	No Prediction	?	?	Low SRIF
Effect of Drug:	?	Decreases SRIF			Increases SRIF
III. Neurotropic Effects					
A. Increase BDNF	Yes	Yes	Yes	?	?
B. Increase Neurogenesis	Yes	?	Yes	?	?
C. Neuroprotective	Yes	Yes	Yes	Yes	(Yes)

Legend: ++++ = very marked; +++ = marked; ++ = moderate; + = mild or some effect; +/– = equivocal; () = ambiguous data; 0 = no effect; – = worse.

BDNF: A role in vulnerability, onset, progression, and treatment

New data indicate that the brain growth factor BDNF (brain-derived neurotrophic factor) which is initially important to synaptogenesis and neural development, and later neuroplasticity and long-term memory in the adult is involved in all phases of bipolar disorder and its treatment.[7]

It appears to be: 1) both a genetic (the val-66-val allele of BDNF) and environmental (low BDNF from childhood adversity) *risk factor*; 2) *episode-related* (serum BDNF decreasing with each episode of depression or mania in proportion to symptom severity; 3) related to some *substance abuse* comorbidity (BDNF increases in the VTA with defeat stress and cocaine self-administration); and 4) related to *treatment*. **Lithium, valproate, and carbamazepine increase BDNF and quetiapine and ziprasidone block the decreases in hippocampal BDNF that occur with stress (as do antidepressants).**

More episodes convey more problems

A greater number of prior episodes is related to increased likelihood of: 1) a rapid cycling course; 2) more severe depressive symptoms; 3) more disability; 4) more cognitive dysfunction; and 5) even the incidence of late life dementia.[4,8,9,10]

Early effective treatment may protect the brain

Taken together, the new data suggest a new view not only of bipolar disorder, but its treatment. Adequate effective **treatment may** not only (a) prevent affective episodes (with their accompanying risk of morbidity, dysfunction, and even death by suicide or the increased medical mortality associated with depression), but may also (b) **reverse or prevent some of the biological abnormalities associated with the illness from progressing.**

Thus, patients should be given timely information pertinent to their stage of illness and recovery that emphasizes not only the risk of treatments, but also their potential, figuratively and literally, life-saving benefits. Long-term treatment and education and targeted psychotherapies are critical to a good outcome.

Therapeutic strategy

We next highlight several attributes of each mood stabilizer, but recognize that the choice of each agent itself is based on inadequate information from the literature, and sequencing of treatments and their combinations is currently more an art than an evidence-based science. We look forward to these informational and clinical trial deficits being reduced in the near future and the development of single nucleotide polymorphism (SNP) and other neurobiological predictors of individual clinical response to individual drugs.

In the meantime, patients and clinicians must struggle with treatment choice based on: 1) the most appropriate targetting of the predominant symptom picture with the most likely effective agent (Table 6.2.4.1 and 6.2.4.2) the best side-effects profile for that patient (Table 6.2.4.2 and 6.2.4.3) using combinations of drugs with different therapeutic targets and mechanisms of action

Table 6.2.4.2 Global putative mechanisms of action

	Li	CBZ	VPA	LTG	NIMOD
Antiglutamineric:	+	+	+	+	?
Via:					
Glutamate Uptake	+				
Sodium Blockade		+	(+)	+	
↑ Brain GABA	+	+/–	++	—	0
↑ GABA-B R in hippocampus (with chronic administration)	+	+	+		
↓ Calcium Influx	+	+	+	+	++
Via:					
Weak NMDA Receptor Inhibition	+	+	+	+	0
Inhibition of Calcium-Channel Type	—	(L)	T	(N,P)	L
↓ DA Turnover	+	+	+		
Second Messenger System					
↓ c-AMP, G proteins	++	++	—	—	—
PI Turnover	↓	(↑)	N.C.	?	+/–
PKC Inhibition	++		++		
↓ ras, MEK, Erk Pathway	++		++		
↓ Inositol Transport	+	+	+	?	?
↑ BDNF	+	+	+		
↑ Bcl-2	+		+		
Histone Deacetylase Inhibition	—	+	++	?	?

Table 6.2.4.3 Global assessment of relative side-effects

	Li	CBZ	VPA	LTG	NIMOD
Weight Gain	++	+	++	0	0
Tremor	++	+/–	++	+/–	0
GI Upset	++	+	++	+/–	0
Memory Disturbance	+	+	+	+	–
Rash	0[a]	++	+/–	++	0
↓WBC	—	++	0	0	0
Agranulocytosis	0	+	0	0	0
↓Platelets	–	0[b]	++	0	0
↑ Liver Enzymes	0	++	++	(+)	0
Hepatitis	0	+	+	?	0
Dizziness Ataxia Diplopia	+/–	++	+	+	+/–
Hyponatremia	–	++	0	0	0
Alopecia	+/–	0	++	?	0
Thyroid Decrements	++	+/–	+/–	+/–	0
Teratogenic	+	++	++	0	0
Malformations	Epstein's Anomaly	Spina Bifida	Spina Bifida		
Developmental Delay	0	+/–	++	+/–	0

Legend: a = psoriasis; b=with aplastic anemia; ++ = moderate to substantial or common; + = mild to less frequent; +/– = equivocal or rare; 0 = not present or no change; – = opposite effect; () = ambiguous findings

(Table 6.2.4.3 and 6.2.4.4) careful consideration of potential advantageous pharmacodynamic interactions and disadvantageous pharmacokinetic drug-drug interactions that need to be avoided or anticipated.

Mood stabilizers

Lithium

In the late 1960s and early 1970s, open and double-blind randomized treatment and discontinuation studies revealed **highly significant effects of lithium in long-term prophylaxis**. These studies followed shortly after a series of studies demonstrating lithium's acute antimanic effects in comparison with both placebo and the existing neuroleptic treatments. High rates of response were touted and lithium clinics were established with the hope that the 60 to 80 per cent response rates revealed in the controlled clinical trials would be mirrored by clinical practice.[11]

However, over the past two decades there has been increasing recognition of the inadequacy of lithium both in acute treatment and long-term prophylaxis, even when used with adjunctive treatments such as antipsychotics, benzodiazepines, and antidepressants.[12,13,14] Given this increasing cognizance of lithium's less than dramatic efficacy in many patients with bipolar illness, alternative and adjunctive treatments were sought.

The anticonvulsants carbamazepine, valproate, and lamotrigine

Carbamazepine and valproate are now well recognized as potential mood-stabilizing anticonvulsants, and initial promising data are emerging for lamotrigine as well. The data are more equivocal for oxcarbazepine and minimal for zonisamide. Importantly, **the GABA-active anticonvulsants gabapentin, tiagabine, and topiramate** are **not effective in acute mania** and thus cannot be considered mood stabilizers. Nonetheless, topiramate may be useful in some common comorbidities of bipolar illness including alcohol and cocaine abuse, bulimia, overweight, migraine, and PTSD. Similarly, gabapentin (which increases brain GABA and acts at the alpha$_2$ delta subunit of the L-type calcium channel) and its close relative pregabalin may have secondary utility in comorbid panic anxiety and social phobia, sleep disorder, alcohol withdrawal, and chronic pain syndromes.

L-type calcium-channel blockers

The dihydropyridine L-type calcium channel blockers became a focus of possible interest for lithium-intolerant and lithium-unresponsive patients based on a variety of clinical and theoretical rationales. Dubovsky *et al.*[15] found increased intracellular calcium in blood elements of patients with bipolar illness, a finding that has been replicated more than a dozen times. Dubovsky *et al.*[16] proceeded to demonstrate the potential antimanic efficacy of the L-type calcium channel blocker verapamil. Many other small

controlled studies were also positive, although less than dramatic results have recently been reported by two groups.[17,18] In addition, one controlled study found that verapamil was not an effective acute antidepressant, even though it appeared to have good antimanic properties.[19] Verapamil was never widely used in clinical practice.

Given these ambiguities with verapamil, Pazzaglia *et al.*[20] and Post *et al.*[21] at the NIMH in the U.S. chose to explore the potential antimanic and antidepressant effects of the dihydropyridine L-type calcium channel blocker nimodipine which has a very different biophysical and pharmacological profile from verapamil.[22] In contrast to the phenylalkylamine verapamil, the dihydropyridine nimodipine is a more potent anticonvulsant that blocks cocaine-induced hyperactivity, sensitization, and dopamine overflow; is positive in animal models of depression; and it improves rather than impairs cognition.

Pharmacology

Lithium

A series of sequential decade-related candidate mechanisms for lithium's psychotropic (mood-stabilizing) actions have been suggested over the last 50 years. These included:

- Effects on enzymes and biosynthetic aminergic neurotransmitter pathways (1950s)
- Effects on presynaptic adrenergic release and reuptake mechanisms (1960s)
- Effects on postsynaptic receptor modulation and impact on receptor supersensitivity (1970s)
- Effects on second messenger systems, adenylate cyclase and G-proteins (1980s)[23]
- Effects on phosphoinositol turnover, protein kinases (PKC and GSK-36), and other signal transduction pathways (1990s)[24,25]
- Most recently, effects on the nuclear level of DNA binding and gene transcription have been found that increase neurotrophic and other factors regulating neuroprotection versus apoptosis and cell death (2000s)[26,27]

Several of these candidate mechanisms are summarized in Figs 6.2.4.1 and 6.2.4.2, and are compared with the mechanisms of some of the other mood-stabilizing anticonvulsants such as carbamazepine and valproate. Lithium's effects on G proteins[20] are conceptually intriguing in relation to lithium's ability to modulate overactive systems, and preliminary support for its adenylate cyclase effects being important for acute mania are based on the finding that novel adenylate cyclase inhibitors are also effective in acute mania.[21] Similarly, recent studies implicating the ability of lithium and valproate to inhibit protein kinase C has been preliminarily validated with the demonstration of antimanic effects of the protein kinase C inhibitor tamoxifen in two double-blind studies.[28,29]

Most recently, lithium has emerged as a possible **neurotrophic and neuroprotective drug** via multiple potential pathways. It inhibits calcium influx through the N-methyl-$_D$-aspartate glutamate receptor;[23] it inhibits GSK-3B; and it increases the ratio of neurotrophic to cell death factors. For example, it increases Bcl-2 and brain-derived neuotrophic factor (BDNF) while decreasing the apoptotic factor BAX and P-53.[25,26] Much work remains to be done to implicate or rule out these changes in the wide array of psychotrophic actions of lithium. Lithium also increases white blood cell count and platelets by increasing granulocyte-macrophage colony-stimulating factor.[30] This effect is sufficient to overcome the benign white count suppression of carbamazepine.[31,32]

The potential clinical relevance of this finding is also evidenced by the data that pretreatment with lithium can decrease the size of a cerebral infarct following middle cerebral artery ligation and decrease the amount of neurological deficit.[33] Given the new findings of altered size of crucial structures involved in emotion regulation in the affective disorders, including amygdale,[34,35] hippocampus,[36] and prefrontal cortex,[4] one can only wonder whether lithium's neuroprotective effects could alter some of these putative neuronal- or glial-based deficits in central nervous system structure and function. The preliminary evidence supports this proposition because lithium increases patients' NAA, a marker of neuronal integrity and, grey matter volume as well.[37,38]

Carbamazepine, valproate, and lamotrigine

As seen in Table 6.2.4.2, since lithium, carbamazepine, and valproate share a group of effects in common, one wonders if they are related to their global antimanic/antidepressant effects. Notable differences among these agents are also present, perhaps also related to some of the differences in therapeutic targets and comorbidities seen in Table 6.2.4.1.

Given the shared effects of carbamazepine and lamotrigine in potent sodium channel blockade and subsequent decreased release of glutamate, one wonders about mechanisms that account for their difference in the epilepsies (carbamazepine exacerbates while lamotrigine improves absence seizures) and bipolar disorder (lamotrigine is a better antidepressant than antimanic). Potential candidates are the effects on different calcium channel subtypes (N,P) and ability to inhibit GABA release, but factors critical to lamotrigine's antidepressant effects remain to be elucidated.[39]

L-type calcium channel blockers

There are several subtypes of voltage-dependent blockers of calcium influx which modulate the L-type channel, i.e. one with relatively long (L) opening times. These subtypes include the widely recognized phenylalkylamines typified by verapamil, the benzodiazepines typified by diltiazem, the diphenylpiperazines typified by flunarizine, and the 1,4-dihydropyridines typified by nifedipine, nimodipine, isradipine, amlodipine, nicardipine, and nitrendipine. Remarkably, even though these agents all potently bind to this voltage-dependent calcium channel and act to inhibit calcium influx, their biochemical and physiological effects are very different, as noted above.

Verapamil and the phenylalkylamines are charged and bind at the outer portion of the calcium channel, while the dihydropyridines, which are uncharged (except amlodipine), bind deeper inside the calcium channel. These different membrane properties and binding characteristics result in a different profile of effects for the dihydropyridines compared with the phenylalkylamines, including increased lipid solubility. It is possible that these differences relate to the increased effectiveness of the dihydropyridines on depression[40] and ultra rapid cycling seen in a small subgroup of lithium-refractory bipolar patients.

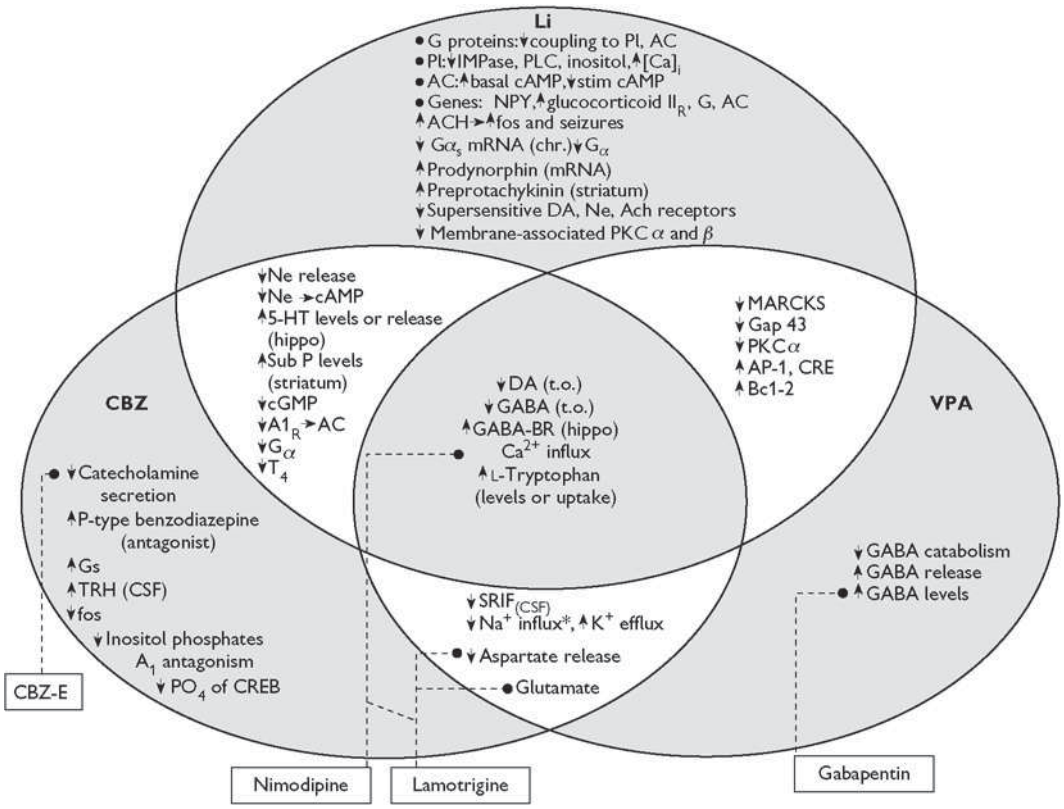


Fig. 6.2.4.1 Common and differential mechanism of mood stabilizers: PI, phosphoinositol; AC, adenylate cyclase; IMPase, inosital monophosphatase; PLC, phospholipase C; cAMP, cyclic adenosine monophosphate; NPY, neuropeptide Y; ACH, acetylcholine; $G\alpha_s$, G protein alpha (s) subunit; DA, dopaminergic; Ne, noradrenergic; PKC, protein kinase C; $A1_R$, adenosine A1 receptors; T_4, thyroxine; Gap 43, growth-associated protein 43; CRE, cyclic response element; CBZ, carbamazepine; TRH, throtrophin-releasing hormone; CREB, cyclic response element binding protein; VPA, valproate; SRIF, somatosatin; t.o., turnover.

Pharmacokinetics, dosage, and administration

Lithium

Lithium blood levels are conventionally described as being therapeutic from 0.5 to 1.2 mmol/l. Within this range, there is a wide agreement that higher doses are associated with increasing numbers of side-effects, but there is disagreement as to whether this is uniformly accompanied by a better therapeutic effect. Gelenberg *et al.*[12] reported that higher doses of lithium in the 0.8 to 1.0 mmol/l range averaging 0.83 mmol/l were 2.6 times as effective as doses achieving blood levels in the lower therapeutic range of 0.4 to 0.6, averaging 0.54 mmol/l. **This increased efficacy** (i.e. lower risk of relapse) **was achieved at the cost of a greater frequency of side effects**. Three times as many patients given doses in the high range withdrew from the study due to side effects. A series of other studies suggest that low to moderate doses of lithium may be just as effective as those in the higher range.[13]

Data of Kleindeinst *et al.*[41] further suggest a differential dose/effect relationship for lithium in mania versus depression prophylaxis. They found higher doses/blood levels more effective in preventing manic episodes while, **paradoxically, lower doses/levels were more effective in preventing depressions**. There is a considerable time lag before lithium reaches steady-state levels and some attempts at lithium loading with large doses from the outset have not been successful. A certain amount of time is needed for lithium to gain access into the central nervous system compartment and steady-state levels are not typically achieved until five half-lives or approximately six days. With the advent of magnetic resonance spectroscopy it has been found that lithium levels in the brain are half those in serum and, based on one small study, may be better correlated with the degree of clinical response than serum levels. **Older individuals have higher intracellular lithium levels**, which may account for the observations of **toxicity at apparently therapeutic blood levels**.

While lithium has traditionally been administered in two, three, or four times daily dosing regimens in attempts to achieve the most consistent and stable blood levels possible, several findings have led to changes in this pattern of dosing. Extended-release preparations are now available and suitable for twice-daily dosing. Even with the original preparations of lithium carbonate many clinicians and investigators have administered lithium in single nighttime doses in order to achieve the highest blood levels (and the potential for side effects) during sleep, and utilize the much lower trough levels at a time when the kidney, for example, can be relatively spared from continuously high lithium levels. Some investigators feel that this might be associated with lower long-term renal side effects, and preliminary data suggest that this paradigm may not be associated with any loss of clinical efficacy. Given lithium's unique anti-suicide

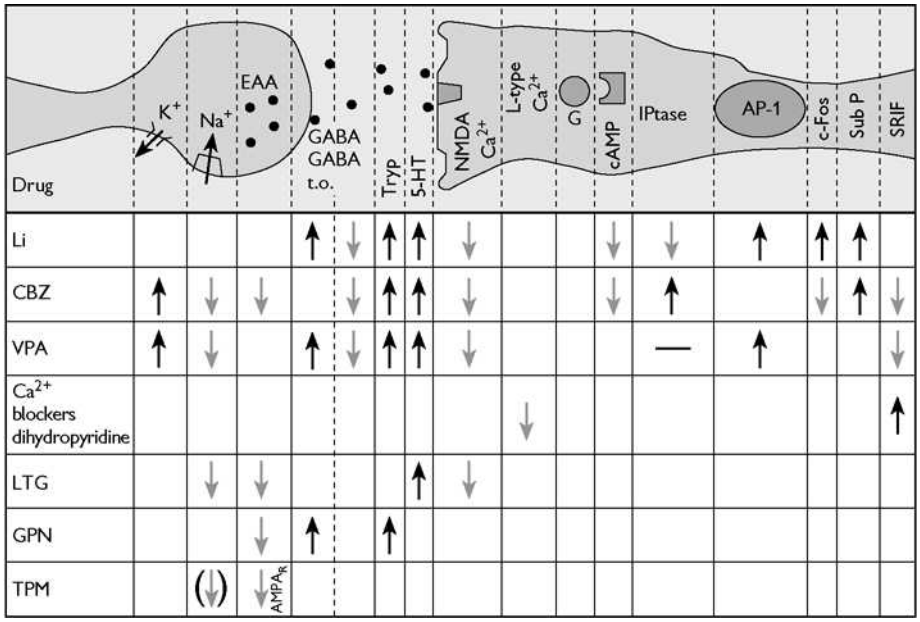


Fig. 6.2.4.2 Mechanisms of mood Stabilization. Depicted schematically at the top of the figure is a synapse with various types of channels, neurotransmitters, and proteins associated with the mechanisms of action of the mood stabilizers listed in the table below. **Row headings**: Li, lithium; CBZ, carbamazepine; VPA, valproate; LTG, lamotrigine; GPN, gabapentin; TPM, topiramate. **Column headings**; K^+ efflux; Na^+ influx; EAA, excitatory amino acids; GABA, γ-aminobutyric acid; GABA t.o., GABA turnover; Tryp, tryptophan; 5-HT, serotonin; NMDA Ca^{2+}; L-type Ca^{2+}; G, G protein; cAMP, cyclic adenosine monophosphate; IPtase, inositol phosphatase; AP-1, activator protein 1; Sub P, substance P; SRIF, somatostatin. Arrows indicate increases or decesases in substance/activity.

effects and strong evidence of neurotrophic and neuroprotective effects in animals at clinically relevant blood levels of 0.5 mmol/l, *administering lithium at whatever doses/blood levels are not associated with side effects would appear to have considerable merit*.

Carbamazepine, valproate, lamotrigine

Carbamazepine, in contrast to oxcarbazepine, is a **potent inducer of CYP 3A4** hepatic enzymes, accounting for **auto-induction after one to three weeks of treatment.**[8] With immediate release preparations, B.I.D or T.I.D. dosing is required, but with long-acting preparations such as Equetro®(Tegretol Retard®), all H.S. dosing should be considered. The dose and blood level to side-effects relationships are highly variable and individual. In non-emergency situations, 200 mg H.S. may be considered as a starting dose with slow upward titration to avoid or accommodate to side effects. In manic inpatients initial dosing of 600 mg to 800 mg on day one may be tolerated.

While blood levels of 4–12 μg/ml are touted as therapeutic in epilepsy, there is no relation of blood level to degree of clinical response across either seizure disorder or affectively ill patients. Therefore, **individualized titration to good clinical effect and minimal side-effects** burden is more appropriate than compulsive blood level monitoring.

Valproate may be given in loading doses of 10 to 15 mg/kg. Efficacy is usually observed between 50–120 mg/ml with a target of 80 mg/ml or more in mania. *Single nighttime doses can be employed* with both immediate release and extended release preparations, which may be better tolerated.

Lamotrigine must be dosed and titrated very slowly in an attempt to avoid the occurrence of a **serious rash**. A typical procedure is to start with 25 mg/day for two weeks and then 50 mg/day for two weeks, and then increase by 25 mg (preferably) to 50 mg/week thereafter. A target dose is about 200 mg/day, but increases to 400 mg/day in those showing partial responses is often tolerated. The rate of titration and target dose of lamotrigine should be halved on **valproate** (which **doubles lamotrigine levels**) and can be doubled on carbamazepine and related potent enzyme inducers.

L-type calcium channel blockers

The pharmacokinetics of the dihydropyridine L-type calcium channel blockers differ markedly. Nimodipine has a T½ of 1–2 h requiring T.I.D. dosing to peak total doses of 240 to 480 mg/day. Isradipine's T½ is 8 h, allowing B.I.D. dosing to a peak dose of 15 mg/day. Notably, amlodipine has a longer half-life and is suitable for single nighttime or twice-daily dosing in comparison with the better psychiatrically studied drugs nimodipine and isradipine. However, as long-acting preparations of these compounds become available, the importance of this half-life dissociation among the different compounds may dissipate. In our patient cohort at the NIMH, the pharmacokinetics of nimodipine (Bay e 9736) and its several metabolites (Bay o 1762, Bay m 5397, and Bay m 8922) were characterized by a rapid peaking (in 30–45 min) and decline in the second hour in response to a 60 mg challenge dose at steady-state blood levels after an overnight fast. Additional small secondary peaks occurred when usual dosing of four-times-a-day was resumed. There were no notable differences between capsule and tablet preparations. The phenylalkylamine verapamil has a 5–8 h half-life, also requiring T.I.D. dosing to a target peak daily dose of 480 mg/day.

Side effects

Lithium

Since lithium has been in use for much of the latter half of the twentieth century, its side-effects profile has been well described. Tremor and gastrointestinal distress, particularly diarrhoea, are generally dose-related, but some patients can have idiosyncratic sensitivity to these side effects, even at relatively low doses. Lithium-induced tremor can be countered with the beta-blocker propranolol in doses of 10 mg four times/day.

Side effects most likely to be associated with non-compliance or discontinuation of the drug include a sense of psychomotor slowing, cognitive dulling, acne or psoriasis, and weight gain. There is preliminary evidence that the anticonvulsants topiramate or zonisamide may help to reverse or stabilize lithium-related weight gain. Cognitive dulling could be treated with dose lowering, assessment of thyroid function, and T_3 (25–50 μg) augmentation even with normal thyroid indices, folate augmentation, and, potentially, an acetylcholinesterase inhibitor such as donepezil (Aricept®).

Lithium interferes with the actions of ADH (i.e. vasopressin) because of its ability to block vasopressin-induced adenylate cyclase. A syndrome of **reversible diabetes insipidus** is thus induced which, in most patients, is not problematic, although in a small percentage of patients, excretion of large volumes of urine can be extreme, inconvenient, and disruptive of normal social routines and sleep. This can be countered with amiloride or the thiazide diuretics (however, the latter also increase lithium levels).

Lithium is clearly able to induce thyroid dysfunction with increases in thyroid-stimulating hormone, sometimes proceeding to more full-blown evidence of chemical hypothyroidism. The threshold for treating lithium-related increases in thyroid-stimulating hormone has not been definitively identified, but with some evidence of lower levels of free thyroxine being associated with increased levels of depression and other low thyroid indices being associated with increased cognitive dysfunction, **replacement of thyroid hormones would appear indicated as thyroid-stimulating hormone begins to exceed normal levels**. Whether thyroid supplementation can reverse or prevent these lithium-related abnormalities remains to be directly assessed in prospective studies.

Some investigators suggest that long-term lithium may be associated with **slowly increasing creatinine levels** and a decrease in creatinine clearance.[40] The incidence of these glomerular filtration abnormalities in lithium-treated patients compared with age- and gender-matched controls remains controversial, as does the mode of treatment in the face of progressive changes in these indices. Given the availability of other potential mood-stabilizing agents, a reduction in lithium levels and supplementation or switching to other agents would be a conservative measure, if tolerated. However, others recommend careful monitoring of continued lithium therapy because the effects of lithium discontinuation on creatinine levels are highly inconsistent.

Severe episodes of lithium intoxication are to be avoided since they can be associated with a syndrome of irreversible cerebellar dysfunction.[42] The use of lithium with very high dose neuroleptic treatment is also to be avoided since occasional idiosyncratic and irreversible organic brain syndromes have resulted on rare occasions. Marked EEG changes and tonic-clonic seizures were also observed with the combination of lithium and clozapine. Lithium can be associated with alterations in calcium homeostasis and frant hyperparathyroidism. As mentioned above, lithium can increase white cells and platelets via its action on granulocyte-macrophage colony stimulating factor.

Valproate has many lithium-like side effects which may also be additive in combination treatment. **Valproate has a black box warning for rare hepatitis/pancreatitis and should not be given to children below two years of age**. It can also cause low platelets, and signs of bleeding tendency should be attended to. Because **valproate increases homocysteine, routine supplementation with folate (and possibly also B6 and B12 in women of child bearing age) would appear prudent**. Valproate can cause asymptomatic to symptomatic hyperammoniumemis; treatment with l-carnitine may be helpful. Zinc and selenium are anecdotally touted as preventing alopecia, but systematic evidence is lacking. Data are mixed as to whether valproate increases testosterone and causes the polycystic ovary syndrome (PCOS). Hirsutism is rarely seen in affectively ill patients and birth control pills will prevent PCOS. Valproate does increase the already high rates of menstrual irregularities seen in women with bipolar disorder.

Carbamazepine has less weight-gain liability than lithium or valproate, but has greater potential for rash and rare but serious hematological problems. **Carbamazepine** routinely causes a benign **drop in white blood cells via its effects on colony stimulating factor; this can be countered by lithium**. More problematic are agranulocytosis and aplastic anemia, estimated to occur in 1 in 20 000 to 50 000 patients. **Patients should be warned to consult their physician** if they develop a fever, sore throat, or other infection that could emanate from a low white blood count, or bleeding gums or petechiae that could reflect low platelets. **Hyponatremia is more common with oxcarbazepine** than carbamazepine, but the hyponatremia of carbamazepine can be treated or prevented with lithium or demeclocycline. Low T_4 on carbamazepine is usually not reflective of hypothyroidism because TSH is not increased, BMR is not decreased, and the degree of drop in T_4 and free T_4 may even be correlated with degree of antidepressive effect.

The **major side-effects concern of lamotrigine is that of a severe rash**, estimated to occur in 1 of 5 000 adults and 1 of 2 500 children. Even a benign rash should lead to drug discontinuation because there is no way to predict when a rash may progress to a Stevens-Johnson syndrome or Toxic Epidermal Necrolysis. Otherwise, the side-effects profile of lamotrigine fits well with bipolar depression treatment because the drug is weight-neutral, non-sedating, and without sexual dysfunction or endocrine dysregulation.

L-type calcium channel blockers

The side-effects profile of nimodipine and related dihydropyridine L-type calcium channel blockers differs considerably from that of lithium. These drugs are primarily used in cardiology for their anti-hypertensive and anti-arrhythmogenic effects. As such, they may cause **side effects related to hypotension including dizziness and tightness in the chest**. Unlike lithium, they are not typically associated with gastrointestinal distress and tend to be slightly constipating rather than associated with diarrhea. Therefore, they might be able to replace some component of lithium's therapeutic action, as noted below, potentially without exacerbating some of lithium's related side effects. Although these agents are often used in migraine prophylaxis, they can also be associated with headache on rare occasions. Redness and erythema with excessive warmth in

the pretibial areas is also an occasional side effect of these agents, as is, more rarely, edema itself.

Nimodipine and related dihydropyridine L-type calcium channel blockers do not appear to share lithium's ability to induce cognitive slowing and, in fact, these agents have been reported to improve performance in some preclinical models of learning and memory deficits as well as in some clinical studies of patients with Alzheimer's disease. This might be related to nimodipine's ability to increase somatostatin upon chronic administration,[43] although other mechanisms remain to be explored.

Also, in contrast to lithium, which is associated with a small incidence of Epstein's anomaly, verapamil is not teratogenic. While systematic data are not available for the other dihydropyridine L-type calcium channel blockers, it is hoped that they will prove to be as safe as verapamil appears to be.

Indications and contraindications

Lithium

The double-blind controlled studies of lithium in acute mania were positive several decades ago, and in the largest randomized study comparing lithium with placebo in a study primarily designed to evaluate the acute antimanic efficacy of valproate, lithium and valproate appeared to show approximately equal efficacy and both were superior to placebo.[44] However, there appear to be a number of subtypes of illness with consistently higher or lower rates of response. As summarized in Table 6.2.4.1, a useful rule of thumb is that lithium is relatively less effective, with a low effectiveness rate of around 30 per cent in acute manic syndrome characterized by anxiety and dysphoria, comorbid substance abuse, comorbid medical conditions, the pattern of illness of depression-mania-well intervals as opposed to mania-depression-well intervals, in those with a negative family history of bipolar illness in first-degree relatives, in those with evidence of **EEG and neurological dysfunction, and in those with a pattern of rapid cycling or multiple prior episodes.**[2]

Lithium used to be considered contraindicated in pregnancy, but with the recognition that major cardiac malformations such as Epstein's anomaly are rare (1 in 1 200 births), many clinicians and patients are deciding to continue lithium when it appears important to continued mood stability.

L-type calcium channel blockers

As discussed earlier, the evidence of verapamil's acute antimanic efficacy is derived from a considerable series of small double-blind studies. While the initial studies were unequivocally positive, more recent studies have not been as positive comparing either verapamil with placebo or verapamil with lithium. Only very preliminary evidence is available for the acute antimanic efficacy of nimodipine, with the open study of Brunet *et al.*[45] being positive in six of six individuals.[18,63] In placebo-controlled studies using an off-on-off-on design, 10 of 30 patients with refractory recurrent affective disorder responded and, in may instances, the antimanic efficacy was demonstrated both in the on-phase and with symptom exacerbation in the off-phase.[46] However, many of these individuals had ultra-rapid or ultra-ultra-rapid (ultradian) cycling patterns and most of the efficacy data reflected effective pharmacoprophylaxis of mania rather than acute antimanic efficacy *per se*.

Dubovsky[47] observed that a prior history of lithium response appears to be associated with a good response to verapamil. However, with the dihydropyridines we have observed some instances of response in those who were previously non-responders to lithium.[46] Among those responsive to the drug were patients with ultra-rapid and ultradian cycling and those with a pattern of recurrent brief depression. As indicated above, a number of these individuals were rechallenged and **responsiveness was confirmed in an off-on-off-on design**. Whether those bipolar patients with increased intracellular calcium would be among those responsive to the L-type calcium channel blockers also awaits completion of such a study, which is now under way.

A promising area of investigation is work using brain imaging, which has found that depressed patients with the classical pattern of relative frontal hypometabolism, and especially in the left insula, were among those who responded best to nimodipine, while equally depressed patients with left insula hypermetabolism responded to carbamazepine.[48] These data raise the possibility that regional topographies of blood flow or metabolism might ultimately help identify a subgroup of patients more responsive to the calcium-channel blocking agents. Also potentially helpful are data showing that nimodipine increases somatostatin in the CSF, and in one small study, that those with lower baseline CSF somatostatin were better responders to the drug.[49]

Interactions

Lithium

Owing to its renal excretion, lithium has renally-mediated rather than hepatically-mediated drug-drug interactions. Lithium excretion is decreased by medications such as thiazides, non-steroidal anti-inflammatory drugs, angiotensin-converting enzyme inhibitors, and, to a lesser extent, furosemide, and by physiological states such as dehydration, advanced age, and renal disease.[50] Owing to lithium's poor therapeutic index, these interactions can result in clinical lithium toxicity unless appropriate dosage adjustment is made. In contrast, lithium clearance is less consistently affected by amiloride, aspirin, and sulindac, and increased with other medications with diuretic effects such as acetazolamide, mannitol, aminophylline, caffeine, and theophylline, as well as during pregnancy.

Valproate, carbamazepine, lamotrigine

Valproate is FDA-approved for migraine prevention and treatment of acute mania. However, it is **widely considered a mood stabilizer and used in long-term prophylaxis of both mood phases**. A modicum of data support acute and prophylactic efficacy in depression. **Valproate decreases alcohol intake** in bipolar patients[51] with this comorbidity and is useful in a wide range of anxiety syndromes.

Valproate is contraindicated in pregnancy because of a several percentage risk of neural tube defects (spina bifida) and, more recently noted, a substantial (20 per cent) incidence of other adverse events,[52] as well as moderate to severe developmental delay in a sizeable number of children.

Carbamazepine is approved for trigeminal neuralgia and the long-acting preparation of carbamazepine (Equetro®/Tegretol Retard®) is approved only in acute mania, but can be considered in long-term prophylaxis for those not responsive to lithium or valproate.

Its profile of effectiveness is almost the converse of that of lithium, with better effects in those with bipolar II illness, anxiety, and substance abuse comorbidity, mood incongruent delusions,[53] and a negative family history of bipolar disorder in first-degree relatives. Carbamazepine should not be used in pregnancy because it carries about a 0.5 per cent risk of neural tube defects.

Lamotrigine is unique because it is approved for only **prevention of depressed episodes** and to a lesser extent, manic and mixed episodes. Robust efficacy in acute depression has been seen in three studies, but not in four others, although the meta analysis remains significantly positive. A small series of patients studied in Canada identified correlates of positive response, including continuous cycling pattern and a **personal and positive family history of anxiety disorders**.[54,55] Only one of five pregnancy case registries showed a significance for the increased occurrence of cleft lip and palate, and this agent clearly appears safer than valproate or carbamazepine. Its overall serious adverse event percentage of 1 per cent does not appear to be different from that of the general population.

L-type calcium channel blockers

Different calcium-channel blockers differentially affect carbamazepine levels. While verapamil and diltiazem increase carbamazepine levels substantially, potentially causing toxicity, this is not the case with nimodipine, isradipine, or amlodipine. Preliminary data from our group and others suggest that carbamazepine decreases nimodipine levels after a 60 mg challenge dose. In our study, group mean peak nimodipine levels during treatment with carbamazepine were about one-half those observed during treatment with nimodipine alone, although this finding showed only a trend level of significance, probably due to small sample size.

When calcium channel blockers are added to ß-adrenergic blocking agents, depression of ventricular function, cardiac slowing, and atrioventricular block can result. Combining calcium channel blockers with ß-adrenergic blocking agents may produce hypotension. Verapamil and nitrendipine increase plasma concentrations of digoxin and produce bradycardia, hypotension, or atrioventricular block.

Effects of withdrawal

Lithium

In addition to the variety of predictors of relative lithium nonresponsiveness from the outset, two relatively new and different mechanisms for the development of treatment resistance or loss of efficacy have been uncovered during long-term follow up in patients who are initially responsive. The first of these is the apparent development of tolerance characterized by an increasing frequency and/or severity of breakthrough episodes despite good compliance and consistent maintenance of lithium blood levels. In a group of 66 patients referred to the NIMH because of lithium nonresponsiveness, 23 patients **(34.8 per cent) displayed this apparent tolerance pattern**.[55] Although it has not been systematically studied, the initial therapeutic manoeuvres in the face of such loss of efficacy at maximum tolerated doses would appear to be augmenting lithium's effects with other putative mood-stabilizing agents with different mechanisms of action, and if lithium should be discontinued, a consideration of its reinstitution with a hope for renewal of responsivity.

In contrast with this tolerance pattern in which patients suffer breakthrough episodes despite remaining under treatment, the phenomenon of lithium-discontinuation-related refractoriness refers to a small group of patients who have done extremely well on their long-term lithium, discontinue the drug, suffer additional relapses, and then fail to re-respond once lithium is reinstituted. This phenomenon accounted for nine of the 66 patients (13.6 per cent) who presented to us as lithium-refractory.[56] The average time well on lithium was 6.6 years, substantially greater than the average well interval of 1.5 years prior to instituting lithium therapy, strongly suggesting that lithium had been effective in these individuals, and if they had remained on the drug, they might have remained well. Sadly, for each of these individuals, this did not prove to be the case.

One patient had been well on lithium for more than 16 years and tapered lithium slowly, suggesting that neither the duration of time well nor the use of a slow taper would necessarily prevent the development of discontinuation-induced refractoriness. A number of other investigative groups have observed such a phenomenon.[57] In these studies, discontinuation-induced refractoriness occurred in anywhere from 3.6 per cent to 18.6 per cent of patients, with a total of 39 of 321 (12.1 per cent), and 12 of 92 (13 per cent) in studies that tracked responders only. Even if it only occurs in about 10 per cent of patients who discontinue their lithium, it would nevertheless appear to be of considerable clinical import and should be included in the informed consent process so that the patient has all of the available data when making decisions of whether or not to continue treatment. That is, in considering the risk-benefit of stopping lithium, **a patient should not only know the very high risk of relapse** (50 per cent in the first five months after discontinuing lithium, 80–90 per cent after 1.5 years,[58] but also that there is **no guarantee that responsivity would be as rapid, robust, or complete** as previously experienced, and that a small subgroup of individuals, perhaps as many as 10 per cent, will not achieve the same good response that they had previously.

Valproate, carbamazepine, lamotrigine

Valproate will displace carbamazepine from its protein binding and inhibit the epoxide hydrolase from converting the active epoxide to the inactive diole. Thus, **carbamazepine dose will need to be decreased when used in combination with valproate**. **Valproate will double lamotrigine levels** such that dosing should be one-half that of normal.

Carbamazepine as a potent 3A4 inducer has many drug-drug interactions. It will decrease levels of oestrogen such that **high dosage forms need to be used with birth control pills**. It will lower levels of lamotrigine, haloperidol, aripiprazole, and many other compounds.

Inhibition of 3A4 will notably increase carbamazepine levels and potentially cause toxicity if a patient is at or near their side-effects threshold. **Erythromycin** and its analogues, **verapamil, and diltiazem** are such examples, and patients should be warned to check with their pharmacists about its many other interactions. Side effects can be avoided with lower carbamazepine doses in advance of such drugs being administered.

Each of these drugs can be discontinued rapidly without a major withdrawal syndrome. Tolerance to the efficacy of each of these

agents has been observed in isolated cases or small series. Increasing doses and/or switching to or adding other agents without cross-tolerance (mechanistically different) are typically used clinically [8] In instances of tolerance development a period of time off the new ineffective drug may be associated with transient renewal of efficacy. Theoretical, but unproven ways of slowing or minimizing tolerance development include: using more maximally tolerated doses, rather than minimally effective ones; holding doses stable and not decreasing them unnecessarily; using combination of drugs; and treating earlier as opposed to later in the course of illness.

L-type calcium channel blockers

Because so few patients have been studied with the L-type calcium channel blockers in long-term prophylaxis, it is uncertain to what extent patients may become tolerant to these agents. However, since tolerance has been observed to virtually every other putative mood-stabilizing agent, it is likely that this will also occur with nimodipine and related agents. Similarly, it is uncertain whether the phenomenon of discontinuation-related refractoriness observed with lithium would extend to the class of L-type calcium channel blockers; while we have not observed this phenomenon in our cohort of nimodipine responders (63), only a small group of patients has been studied to date.

Further information

Post, R.M., and Leverich, G.S. (2008). *Treatment of Bipolar Illness: A Case Book for Clinicians and Patients*. WW Norton, Inc. (in press).

Post, R.M., Altshuler, L.L. (2005). Mood disorders: treatment of bipolar disorders. In: *Comprehensive Textbook of Psychiatry*. (eds. B.J. Sadock and V.A. Sadock), pp. 1661–707. Lippincott Williams & Williams, New York.

Post, R.M., Speer, A., Leverich, G.S. (2006).Complex combination therapy: the evolution toward rational polypharmacy in lithium-resistant bipolar illness. In *Bipolar Psychopharmacotherapy: Caring for the Patient*. (eds. H. Akiskal and M. Tohen), pp. 135–67. John Wiley & Sons.

References

1. Post, R.M., and Leverich, G.S. (2007). *Treatment of Bipolar Illness: A Case Book for Clinicians and Patients*. WW Norton, Inc. (in press).
2. Post, R.M., Speer, A., Leverich, G.S., *et al.* (2006).Complex combination therapy: the evolution toward rational polypharmacy in lithium-resistant bipolar illness. In *Bipolar Psychopharmacotherapy: Caring for the Patient*. (eds. H. Akiskal and M. Tohen), pp. 135–67. John Wiley & Sons.
3. Post, R.M. and Luckenbaugh, D.A. (2003). Unique design issues in clinical trials of patients with bipolar affective disorder. *Journal of Psychiatric Research*, **37**(1), 61–73.
4. Post, R.M., Leverich, G.S., Weiss, S.R., *et al.* (2003). Psychosocial stressors as predisposing factors to affective illness and PTSD: potential neurobiological mechanisms and theoretical implications. In: *Neurodevelopmental Mechanisms in Psychopathology*. (eds. D. Cicchetti and E. Wilker). pp. 491–525. Cambridge University Press, New York.
5. Post, R.M. (2007a). Kindling and sensitization as models for affective episode recurrence, cyclicity, and tolerance phenomena. *Neuroscience and Biobehavioral Reviews*, **31**(6), 858–73.
6. Leverich, G.S., Post, R.M., Keck, P.E. Jr., *et al.* (2007). The poor prognosis of childhood-onset bipolar disorder. *Journal of Pediatrics*,**150**(5), 485–90.
7. Post, R.M. (2007b). Role of BDNF in bipolar and unipolar disorder: clinical and theoretical implications. *Journal of Psychiatric Research*, **41**, 979–90.
8. Post, R.M., Ketter, T.A., Uhde, T., *et al.* (2007). Thirty years of clinical experience with carbamazepine in the treatment of bipolar illness: principles and practice. *CNS Drugs*, **21**(1), 47–71.
9. Post, R.M. (2004). The status of the sensitization/kindling hypothesis of bipolar disorder. *Current Psychosis and Therapeutics Reports*, **2**, 135–41.
10. Kessing, L.V., and Andersen, P.K. (2004). Does the risk of developing dementia increase with the number of episodes in patients with depressive disorder and in patients with bipolar disorder? *Journal of Neurology, Neurosurgery, and Psychiatry*, **75**(12),1662–66.
11. Schou, M. (1997). Forty years of lithium treatment. *Archives of General Psychiatry*, **54**, 9–15.
12. Gelenberg, A.J., Kane, J.M., Keller, M.B., *et al.* (1989). Comparison of standard and low serum levels of lithium for maintenance treatment of bipolar disorder. *New England Journal of Medicine*, **321**, 1489–93.
13. Vestergaard, P., Licht, R.W., Brodersen, A., *et al.* (1998). Outcome of lithium prophylaxis: a prospective follow-up of affective disorder patients assigned to high and low serum lithium levels. *Acta Psychiatrica Scandinavica*, **98**, 310–15.
14. Gitlin, M.J., and Altshuler, L.L. (1997). Unanswered questions, unknown future for one of our oldest medications. *Archive General of Psychiatry*, **54**(1), 21–3.
15. Dubovsky, S.L., Murphy, J., Thomas, M. *et al.* (1992). Abnormal intracellular calcium ion concentration in platelets and lymphocytes of bipolar patients. *American Journal of Psychiatry*, **149**, 118–20.
16. Dubovsky, S.L., Franks, R.D., Allen, S. *et al.* (1986). Calcium antagonists in mania: a double-blind study of verapamil. *Psychiatry Research*, **18**, 309–20.
17. Janicak, P.G., Sharma, R.P., Pandey, G., *et al.* (1998). Verapamil for the treatment of acute mania: a double-blind, placebo controlled trial. *American Journal of Psychiatry*, **155**, 972–3.
18. Walton, S.A., Berk, M., and Brook S. (1996). Superiority of lithium over verapamil in mania: a randomized, controlled, single-blind trial. *Journal of Clinical Psychiatry*, **57**, 543–6.
19. Hoschl, C. and Kozeny, J. (1989). Verapamil in affective disorders: a controlled, double-blind study. *Biological Psychiatry*, **25**, 128–40.
20. Pazzaglia, P.J., Post, R.M., Ketter T.A., *et al.* (1993). Preliminary controlled trial of nimodipine in ultra-rapid cycling affective dysregulation. *Psychiatry Research*, **49**, 257–72.
21. Post, R.M., Pazzaglia, P.J., Ketter, T.A., *et al.* (2000). Carbamazepine and nimodipine in affective illness: effiicacy, mechanisms of action, and interactions. In *Pharmacotherapy for mood, anxiety, and cognitive disorders* (eds. S. Montgomery and U. Halbreich), American Psychiatric Press, Washington, DC.
22. Triggle, DJ. (2007). Calcium Channel Antagonists: Clinical uses-past, present, and future. *Biochemical Pharmacology*, **74,** 1–9.
23. Belmaker, R.H., Avissar, S., and Schreiber, G. (1991). Effect of lithium on human neurotransmitter receptor systems and G proteins. In *Lithium and the cell: pharmacology and biochemistry* (ed. N.J. Birch), pp. 113–19. Academic Press, London.
24. Manji, H.K., Potter, W.Z., and Lenox, R.H. (1995). Signal transduction pathways. Molecular targets for lithium's actions. *Archives of General Psychiatry*, **52**, 531–43.
25. Bebchuk, J.M., Arfken, C.L., Dolan-Manji, S., *et al.* (2000). A preliminary investigation of a protein kinase C inhibitor in the treatment of acute mania. *Archives of General Psychiatry*, **57**, 95–7.
26. Nonaka, S., Hough, C.J., and Chuang, D.M. (1998). Chronic lithium treatment robustly protects neurons in the central nervous system against excitotoxicity by inhibiting *N*-methyl-*D*-aspartate receptormediated calcium influx. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 2642–7.
27. Chen G., Zeng, W.Z., Yuan, P.X., *et al.* (1998). The mood-stabilizing agents lithium and valproate dramatically increase the levels of the neuroprotective protein BCL-2 in the CNS. *Journal of Neurochemistry*, **72**, 879–82.

28. Zarate, C.A. Jr., Singh, J.B., Carlson, P.J., *et al.* (2007). Efficacy of a protein kinase C inhibitor (tamoxifen) in the treatment of acutemania: a pilot study. *Bipolar Disorders*, **9**(6), 561–70. 29. Einat, H., Yuan, P., Szabo, S.T., *et al.* (2007). Protein kinase C inhibition by tamoxifen antagonizes manic-like behavior in rats: implications for the development of novel therapeutics for bipolar disorder. *Neuropsychobiology*, **55**(3–4), 123–31.
30. Gallicchio, V.S., Chen, M.G., and Watts, T.D. (1984). Specificity of lithium (Li+) to enhance the production of colony stimulating factor (GM-CSF) from mitogen-stimulated lymphocytes *in vitro. Cellular Immunology*, **85**, 58–66.
31. Joffe, R.T., Post, R.M., Roy-Byrne, P.P., *et al.* (1985). Hematological effects of carbamazepine in patients with affective illness. *American Journal of Psychiatry*, **142**, 1196–9.
32. Kramlinger, K.G. and Post, R.M. (1990). Addition of lithium carbonate to carbamazepine: hematological and thyroid effects. *American Journal of Psychiatry*, **147**, 615–20.
33. Chuang, D.M., Chen, R.W., Chalecka-Franaszek, E., *et al.* (2002). Neuroprotective effects of lithium in cultured cells and animal models of diseases. *Bipolar Disorders,* **4**(2):129–36.
34. Wignall, E.L., Dickson, J.M., Vaughan, P., *et al.* (2004). Smaller hippocampal volume in patients with recent-onset posttraumatic stress disorder. *Biological Psychiatry*, **56**(11), 832–6.
35. Blumberg, H.P., Fredericks, C., Wang, F., *et al.* (2005). Preliminary evidence for persistent abnormalities in amygdala volumes in adolescents and young adults with bipolar disorder. *Bipolar Disorders*, **7**(6), 570–6.
36. Sheline, YI., Sanghavi, M., Mintun, M.A., *et al.* (1999). Depression duration but not age predicts hippocampal volume loss in medically healthy women with recurrent major depression. *Journal of Neuroscience*, **19**(12), 5034–43.
37. Moore, G.J., Bebchuk, J.M., Wilds, I.B., *et al.* (2000). Lithium-induced increase in human brain grey matter. *Lancet*, **356**(9237): 1241–2.
38. Bearden, C.E., Thompson, P.M., Dalwani, M., *et al.* (2007). Greater cortical gray matter density in lithium-treated patients with bipolar disorder. *Biological Psychiatry*, **62**(1), 7–16.
39. Ketter, T.A., Wang, P.W., Becker, O.V., *et al.* (2003). The diverse roles of anticonvulsants in bipolar disorders. *Annals of Clinical Psychiatry*, **15**(2), 95–108.
40. Taragano, F.E., Bagnatti, P., Allergri, F., *et al.* (2005). A double-blind, randomized clinical trial to assess the augmentation with nimodipine of antidepressant therapy in the treatment of 'vascular depression'. *International Psychogeriatrics*, **17**, 487–498.
41. Kleindienst, N., Engel, R., Greil, W., *et al.* (2005). Which clinical factors predict response to prophylactic lithium? A systematic review for bipolar disorders. *Bipolar Disorders*, **7**(5), 404–17.
42. Kores, B. and Lader, M.H. (1997). Irreversible lithium neurotoxicity: an overview. *Clinical Neuropharmacology*, **20**, 283–99.
43. Pazzaglia, P.J., George, M.S., Post, R.M., *et al.* (1995). Nimodipine increases CSF somatostatin in affectively ill patients. *Neuropsychopharmacology*, **13**, 75–83.
44. Bowden, C.L., Brugger, A.M., Swann, A.C., *et al.* (1994). Efficacy of divalproex vs. lithium and placebo in the treatment of mania. The Depakote Mania Study Group. *Journal of the American Medical Association*, **271**, 918–24.
45. Brunet, G., Cerlich, B., Robert, P., *et al.* (1990). Open trial of a calcium antagonist, nimodipine, in acute mania. *Clinical Neuropharmacology*, **13**, 224–8.
46. Pazzaglia, P.J., Post, R.M., Ketter, T.A., *et al.* (1998). Nimodipine monotherapy and carbamazepine augmentation in patients with refractory recurrent affective illness. *Journal of Clinical Psychopharmacology*, **18**, 404–13.
47. Dubovsky, S.L. (1995). Calcium channel antagonists as novel agents for manic-depressive disorder. In *Textbook of psychopharmacology* (eds. A.E. Schatzberg and C.B. Nemeroff), pp. 377–88. American Psychiatric Press, Washington, DC.
48. Ketter, T.A., Kimbrell, T.A., George, M.S., *et al.* (1999). Baseline cerebral hypermetabolism associated with carbamazepine response and hypometabolism with nimodipine response in mood disorders. *Biological Psychiatry*, **46**, 1364–74.
49. Frye, M.A., Pazzaglia, P.J., George, M.S., *et al.* (2003). Low CSFsomatostatin associated with response to nimodipine in patents with affective illness. *Biological Psychiatry*, **53**(2), 180–3.
50. Muller-Oerlinghausen, B. (1999). Drug interactions with lithium: aguide for clinicians. *CNS Drugs*, **11**, 41–8.
51. Salloum, I.M., Cornelius, J.R., Daley, D.C., *et al.* (2005). Efficacy of valproate maintenance in patients with bipolar disorder and alcoholism: a double-blind placebo-controlled study. *Archives of General Psychiatry*, **62**(1), 37–45.
52. Meador, K.J., Baker, G.A., Finnell, R.H., *et al.* (2006). In utero antiepileptic drug exposure: fetal death and malformations. *Neurology*, **67**(3), 407–12.
53. Greil, W., Kleindienst, N., Erazo, N., *et al.* (1998). Differential response to lithium and carbamazepine in the prophylaxis of bipolar disorder. *Journal of Clinical Psychopharmacology*, **18**(6), 455–60.
54. Passmore, M.J., Garnham, J., Duffy, A., *et al.* (2003). Phenotypic spectra of bipolar disorder in responders to lithium versus lamotrigine. *Bipolar Disorders*,**5**(2), 110–4.
55. Grof, P. (2003). Selecting effective long-term treatment for bipolar patients: monotherapy and combinations. *Journal of Clinical Psychiatry*, **64** Suppl 5, 53–61.
56. Post, R.M., Leverich G.S., Altshuler, L., *et al.* (1992). Lithium-discontinuation-induced refractoriness: preliminary observations. *American Journal of Psychiatry*, **149**, 1727–9.
57. Koukopoulos, A., Reginaldi, D., Minnai, G., *et al.* (1995). The long term prophylaxis of affective disorders. *Advances in Biochemical Psychopharmacology*, Gessa, Fratta, Pani & Serra (eds)., *Depression and Mania: From Neurobiology to Treatment*, Vol 49.
58. Suppes, T., Baldessarini, R.J., Faedda, G.L., *et al.* (1991). Risk of recurrence following discontinuation of lithium treatment in bipolar disorder. *Archives of General Psychiatry*, **48**, 1082–4.

6.2.5 Antipsychotic and anticholinergic drugs

Herbert Y. Meltzer and William V. Bobo

Introduction

The discovery by Delay and Denicker in 1953 that chlorpromazine was highly effective in alleviating delusions, hallucinations, and disorganized thinking, was the seminal breakthrough in the treatment of schizophrenia, the first agent to produce sufficient relief of core psychotic symptoms to permit life outside of institutions for many patients with schizophrenia, and even a return to a semblance of function within normal limits. Chlorpromazine and the other related typical antipsychotic drugs which were introduced over the next 30 years have proven to be of immense benefit to vast numbers of people who experience psychotic symptoms as a

component of a diverse group of neuropsychiatric and medical disorders, as well as drug-induced psychoses. These drugs have been invaluable in providing clues to the aetiology of schizophrenia and other forms of mental illness with psychotic features and as tools in understanding fundamental neural processes, especially those involving dopamine, a key neurotransmitter involved in psychosis. This class of drugs has now been supplanted by the so-called atypical antipsychotic drugs, of which clozapine is the prototype. This chapter will describe the various classes of antipsychotic agents, with emphasis on the atypical antipsychotic drugs, their benefits and adverse effects, recommendations for use in clinical practice, and mechanism of action. The drugs used to treat the extrapyramidal side-effects (EPS) produced mainly by the typical antipsychotic drugs are also considered.

The classes of antipsychotic drugs

Antipsychotic drugs have been classified into two broad categories: typical and atypical.[1] Typical antipsychotic drugs are those which (typically) produce EPS at clinically effective doses, including parkinsonism (muscle rigidity, tremor, bradykinesia), acute dystonic reactions, dyskinesias, akathisia (restlessness), and tardive dyskinesia. They are also called neuroleptics because of their inhibitory effect upon locomotion activity. They are sometimes referred to as first generation antipsychotic drugs, but this has multiple problems as a class designation. The prototype of the atyipcal class of agents is clozapine which was first discovered during the early stages of the development of the drugs called first generation agents. The major mode of action of typical neuroleptics is to block dopamine D_2 receptors in the limbic system, which includes the nucleus accumbens, stria terminalis, and amygdala.

The typical antipsychotic drugs are members of a variety of chemical families (Table 6.2.5.1). They vary in affinity for the D_2 receptor, with low affinity drugs such as chlorpromazine, which require high doses for clinical efficacy, to high affinity drugs such as haloperidol, which are effective at lower doses (Table 6.2.5.1). Kapur and Seeman[3] have proposed that the rate of dissociation of all antipsychotic drugs from the D_2 receptor provides the basis for the distinction between typical and atypical antipsychotic drugs, with atypical antipsychotic drugs dissociating more rapidly. While this is true for clozapine and quetiapine, the atypical drugs risperidone, sertindole, olanzapine and asenapine dissociate no more rapidly or even slower than haloperidol. As such, 'fast dissociation' cannot provide the pharmacological basis for atypicality for most of the drugs that are considered atypical.

Low-potency typical neuroleptic agents are those in which the usual dose range in schizophrenia is equal to or greater than 200 mg/day, while mid- to high-potency agents are those in which the dose range is between 2 and 175 mg/day. In general, the low-potency drugs are more sedative and more hypotensive than the high-potency agents but also have less of a tendency to produce extrapyramidal side-effects. The typical antipsychotic drugs differ from one another with regard to potential for other side-effects, e.g. weight gain and hypotension, but have comparable efficacy as antipsychotic agents.[4]

Atypical antipsychotic drugs are those antipsychotic agents with a significantly lower propensity to produce EPS at clinically effective doses.[1] They are also characterized by a more diverse and complex pattern of pharmacological activity, including serotonin $(5\text{-hydroxytryptamine})_{2A}$ and dopamine D_2 antagonism as well as a variety of activities at other receptors whose contribution to their mode of action is still being elucidated.[2] Substituted benzamides, e.g. amisulpride, also have low EPS at clinically effective doses and may constitute another class of atypical agents. New classes of atypical antipsychotic drugs are emerging from research with considerable frequency at the current time.

The prototypical atypical antipsychotic drug is clozapine, a dibenzodiazepine (Table 6.2.5.1).[5] Others include aripiprazole,[6] olanzapine, quetiapine, paliperidone, risperidone, sertindole, ziprasidone and zotepine, while iloperidone,[7] asenapine,[8] and lurasidone[9] are in development and have a similar pharmacology to that of risperidone. These drugs are all more potent 5-HT_{2A} than D_2 receptor antagonists as well as multireceptor antagonists[9,10] except for aripiprazole, which is a dopamine D_2 receptor partial agonist. Bifeprunox is also a partial D_2 agonist. It lacks 5-HT_{2A} receptor blocking properties, relying instead on 5-HT_{1A} partial agonism to reduce serotonergic tone. Amisulpride and remoxipride are substituted benzamides. Both are selective D_2/D_3 antagonists.[2]

Table 6.2.5.1 Selected antipsychotic drugs and classification schemes

Drug name	Trade name	Chemical class	General class	D2 potency*
Aripiprazole	Abilify	Dihydrocarbostyril	Atypical	
Chlorpromazine	Thorazine	Phenothiazine	Typical	Low
Clozapine	Clozaril	Dibenzazepine	Atypical	
Droperidol	Inapsine	Butyrophenone	Typical	Mid
Fluphenazine	Prolixin	Phenothiazine	Typical	High
Haloperidol	Haldol	Butyrophenone	Typical	High
Loxapine	Loxitane	Dibenzazepine	Typical	Mid
Mesoridazine	Serentil	Phenothiazine	Typical	Low
Molindone	Moban	Dihydroindolone	Typical	Mid
Olanzapine	Zyprexa	Thiobenzodiazepine	Atypical	
Paliperidone	Invega	9-hydroxy metabolite of risperidone	Atypical	
Perphenazine	Trilafon	Phenothiazine	Typical	Mid
Pimozide	Orap	Butyrophenone	Typical	Mid
Promazine		Phenothiazine	Typical	Mid
Quetiapine	Seroquel	Dibenzothiazepine	Atypical	
Risperidone	Risperdal, Risperdal CONSTA	Benzisoxazole	Atypical	
Thioridazine	Mellaril	Phenothiazine	Typical	Low
Tiotixene	Navane	Thioxanthene	Typical	High
Trifluoperazine	Stelazine	Phenothiazine	Typical	Mid
Ziprasidone	Geodon	Benzisothiazole	Atypical	

* Classification on the basis of potency of D2 receptor binding for typical antipsychotic drugs only

Remoxipride was withdrawn shortly after its introduction because of a high rate of aplastic anaemia.

As will be discussed, the atypical antipsychotic drugs differ not only with regard to side-effects but also with regard to efficacy.[11,12] Atypical antipsychotic agents have been shown to have advantages, albeit modest, in treating negative mood symptoms[13–15] and to improve cognitive dysfunction in schizophrenia and perhaps other psychiatric disorders.[16–18]

Pharmacology

There is abundant evidence that dopamine plays a key role in the aetiology of psychosis and the action of antipsychotic drugs.[19] The antipsychotic action of the typical antipsychotic drugs is highly correlated with their affinities for D_2 receptors. Amphetamine and methamphetamine, which increase synaptic concentrations of dopamine, have been found to exacerbate delusions and hallucinations in some patients with schizophrenia This effect is believed to be due to stimulation of a subgroup of D_2 receptors in mesolimbic nuclei.[19,20] The cell bodies of mesolimbic dopamine neurones reside in the ventral tegmentum, the so-called A10 area, and have terminals in the nucleus accumbens, stria terminalis, and olfactory tubercle. The outflow of these regions to the thalamus and the cortex is believed to mediate psychotic symptoms. The firing rate of the mesolimbic dopaminergic neurones is subject to multiple influences, including stimulatory serotonergic input from the median raphe.[21] The origin of the dopamine neurones that terminate on cholinergic neurones in the basal ganglia is the substantia nigra, the so-called A9 region.[20] Blockade of striatal D_2 receptors in this pathway leads to the extrapyramidal side-effects produced by antipsychotic agents. A group of ventral tegmental dopamine neurones project to various regions of the cortex and comprise the mesocortical dopamine system. There is extensive evidence that these neurones are important for cognition, especially working memory,[22] as well as negative symptoms.[23] Neuroleptic drugs occupy 80 to 95 per cent of striatal D_2 receptors in patients with schizophrenia at clinically effective doses, though a lower blockade threshold of 60 per cent for improving positive symptoms has been identified.[24] Extrapyramidal side-effects occur above 80 per cent occupancy of these receptors. Blockade of D_2 receptors in the anterior pituitary gland is the basis for their ability to stimulate prolactin secretion.[25]

The prefrontal cortex has relatively low concentrations of D_2 receptors and has a higher density of D_1, D_3 and D_4 dopamine receptors.[20] The activation of D_1 receptors in prefrontal cortex may be especially critical for normal working memory and other executive type functions subserved by this brain region. However, no D1 agonists are available for treatment at the current time, although several are in development. Drugs which selectively block D_4 receptors have not been found to have an antipsychotic effect.[26] There are only limited data regarding the aetiologic or pharmacological significance of D_3 receptors in schizophrenia.

The typical antipsychotic drugs vary in their *in vitro* and *in vivo* affinities for receptors such as the dopamine D_1, histamine H_1, muscarinic, α-1 and α-2 adrenergic, and serotonergic receptors (Table 6.2.5.2), which mediate effects on arousal, extrapyramidal, cognitive, cardiovascular, gastrointestinal, and genitourinary function (Table 6.2.5.3).[27]

Thioridazine is a relatively potent antimuscarinic agent. Most of the low-potency antipsychotic agents are potent α_1 and H_1 antagonists. These affinities contribute to hypotension and weight gain, respectively. While some typical antipsychotic drugs have a high affinity for 5-HT_{2A} receptors, their affinities for D_2 receptors are even higher, which diminishes the beneficial effects of the 5-HT_{2A} receptor blockade. The specific receptor profile of each atypical antipsychotic is of special interest because it may account for critical differences among these compounds, especially in terms of side effect burden (Table 6.2.5.4).

The affinities of the atypical antipsychotic drug have been related to their efficacy and side effect profiles. As noted above, the most important determinant of atypicality for most of the currently available agents of this type is that they are more potent 5-HT_{2A} than D_2 receptor antagonists. An exception is aripiprazole, which combines potent 5-HT_{2A} antagonism and 5-HT_{1A} agonism, with partial D_2 receptor agonism. Another exception is amisulpiride, which is a selective $D_{2/3}$ antagonist with little pharmacological activity at 5-HT_{2A} receptors. Combined 5-HT_{2A} with less potent D2 antagnoism is the most consistent principle yet discovered to produce a separation between antipsychotic action and interference with motor function. This hypothesis arose from showing that it could distinguish clozapine, the prototypical atypical antipsychotic drug, and a series of other atypical antipsychotic compounds from those which have typical properties.[28] These studies suggested that the low potential for extrapyramidal side-effects of clozapine, and subsequently, olanzapine, quetiapine, risperidone, iloperidone, ziprasidone, paliperidone and asenapine are due, in part, to their relatively stronger 5-HT_2 antagonist and weak D_2 antagonist properties. The serotonin-dopamine interaction in the nigrostriatal and mesolimbico-cortical pathways appears to be mediated by stimulation of 5-HT_{2A} receptors, which are located on dopaminergic cell bodies, whereas antagonism of these receptors may release these neurones from tonic inhibition.

The atypical antipsychotic agents have the ability to increase prefrontal cortical dopaminergic activity compared with subcortical dopaminergic activity.[29] The ability to increase the release of dopamine in the prefrontal cortex may be important for atypical antipsychotic agents to improve cognition and negative symptoms. It may also contribute to decreasing the release of dopamine in the mesolimbic region, because prefrontal dopamine neurones modulate the activity of corticolimbic glutamatergic neurones that influence the release of dopamine from nerve terminals in the limbic region.[22] Typical neuroleptic drugs do not share this ability to increase dopamine efflux in prefrontal cortex. Clozapine and some of the other atypical antipsychotic drugs that are also potent 5-HT_{2A} antagonists, but not typical neuroleptics, also produce marked increases in prefrontal cortical and hippocampal acetylcholine efflux.[30] These atypical agents also produce marked increases in noradrenaline efflux in the prefrontal cortex which is correlated in time and magnitude with the increase in extracellular dopamine.[31] It is of interest that in rodents, combining ritanserin (a mixed 5-HT2a/2B/2C antagonist) or M-100907 (a selective 5-HT_{2A} antagonist) with a selective $D_{2/3}$ antagonist resulted in increased prefrontal dopamine release.[32,33] The combination of haloperidol and M-100907 also increased prefrontal dopamine release, with the greatest effects observed when lower doses of haloperidol were used.[34] Because reduced noradrenergic and dopaminergic function in prefrontal cortex and hippocampus has been

Table 6.2.5.2 Affinities of selected antipsychotic drugs at various neuroreceptors

Drug name	D2	5-HT$_{1A}$	5-HT$_{2A}$	5-HT$_{2C}$	α-1	α-2	H-1	M-1
Aripiprazole	0.95	5.6	4.6	181.0	25.0	74.0	29.0	>6K
Chlorpromazine	2.0	>3K	3.2	26.0	0.28	184.0	0.18	47.0
Clozapine	431.0	105.0	13.0	29.0	l.6	142.0	2.0	14.0
Droperidol	0.25 [173]	NA	NA	NA	NA	NA	NA	NA
Fluphenazine	0.54	145.0	7.4	418.0	6.4	314.0	7.3	>1K
Haloperidol	2.0	>1K	73.0	>10K	12.0	>1K	>3K	>10K
Loxapine	10.0	>2K	3.9	21.0	31.0	151.0	2.8	175.0
Molindone	63.0[43]	>3K[43]	320.0[43]	>10K[43]	>2K[43]	>1K[43]	>2K[43]	NA
Olanzapine	72.0	>2K	3.0	24.0	109.0	314.0	4.9	24.0
9-OH risperidone*	9.4	637.8	1.9	100.3	2.5	4.7	5.6	>10K
Perphenazine	1.4[43]	421.0[43]	5.6[43]	132.0[43]	10.0[43]	810.5[43]	8.0[43]	NA
Pimozide	0.65[43]	650.0[43]	19.0[43]	>3K[43]	197.7[43]	>1K[43]	692.0[43]	800.0[174]
Quetiapine	567.0	431.0	366.0	>1K	22.0	>3K	7.5	858.0
Risperidone	4.9	427.0	0.19	94.9	5.0	151.0	5.2	>10K
Thioridazine	10.0	108.0	11.0	69.0	1.3	134.0	14.0	33.0
Tiotixene	1.4	410.0	111.0	>1K	12.0	80.0	12.0	>10K
Trifluoperazine	1.3[43]	950.0[43]	13.0[43]	378.0[43]	24.0[43]	653.7[43]	63.0[43]	NA
Ziprasidone	4.0	76.0	2.8	68.0	18.0	160.0	130.0	>10K

All receptor binding affinities are reported as K_i (nM) using National Institutes of Mental Health (NIMH) Psychoactive Drug Screening Program (PDSP) certified data, available online at http://pdsp.cwru.edu/pdsp.php, unless otherwise specified. In general, the lower the K_i (nM) value, the higher the binding affinity for the drug at a given receptor site.

NA = human cloned receptor data not available

* 9-hydroxy (9-OH) risperidone is marketed as paliperidone

Table 6.2.5.3 Hypothesized therapeutic and adverse effects of receptor occupancy by antipsychotic drugs

Target receptor	Pharmacological activity	Therapeutic effect(s)	Adverse effect(s)
Dopamine D2	Antagonism or partial agonist effects	Reduction of positive symptoms	Extrapyramidal effects (EPS) Hyperprolactinemia
Serotonin (5-HT) $_{1A}$	Full or partial agonist effects	Cognitive enhancement Reduction of mood and anxiety symptoms	
5-HT $_{2A}$	Antagonism	Reduction of negative symptoms Reduction of EPS Reduction of mood and anxiety symptoms Increased deep sleep	
5-HT $_{2C}$	Antagonism	Reduced anxiety symptoms	Weight gain
Adrenergic α-1	Antagonism		Orthostatic hypotension Dizziness
Adrenergic α-2	Antagonism		Reflex tachycardia
Histamine H-1	Sedation	Sedation Drowsiness Weight gain	
Muscarinic (cholinergic) M-1	Antagonism	Reduction of EPS	Blurry vision Exacerbation of acute angle closure glaucoma Sinus tachycardia Constipation Urinary retention Memory dysfunction

Adapted from Kelly, D.L. and Love, R.C. Ziprasidone and the QTC interval: pharmacokinetic and pharmacodynamic considerations, Psychopharmacology Bulletin, **35**, 66–79, copyright 2001, MedWorks Media Global, LLC.

Table 6.2.5.4 Adverse effects of selected antipsychotic drugs

Typical antipsychotic drugs

	EPS	Tardive dyskinesia	Prolactin elevation	Sedation	Weight gain	Orthostasis	Anti-cholinergic	Diabetes exacerbation & dyslipidemia
Chlorpromazine Fluphenazine Haloperidol Loxapine Mesoridazine Molindone Perphenazine Thioridazine Tiotixene Trifluoperazine	Some (for low potency* drugs) - +++ (for high-potency* drugs)	++ - +++	++ - +++ (risk higher for high-potency drugs)	Some (for high potency drugs) - +++ (for low-potency drugs); ? least for molindone	Some (for high potency drugs) - +++ (for low-potency drugs); ? least for molindone	Some (for high potency drugs) - +++ (for low-potency drugs)	Some (for high potency drugs) - +++ (for low-potency drugs)	+ - ++

* See Table 6.2.5.1 for list of low-, mid-, and high-potency (with respect to dopamine D2 receptor blockade) antipsychotic drugs

Adapted from the International Psychopharmacology Algorithm Project (IPAP) algorithm for the treatment of schizophrenia, available at www.ipap.org, copyright 2008 International Psychopharmacology Algorithm Project (IPAP)

Atypical antipsychotic drugs

	EPS	Tardive dyskinesia	Prolactin elevation	Sedation	Weight gain	Orthostasis	Anti-cholinergic	Glucose dysregulation & dyslipidemia
Amisulpride	+	Rare	+++	+	0 - +	+	0	0
Aripiprazole	0 - +	0 - +	0	0 - +	0 - +	+ - ++	0	0
Clozapine	0	0	Transient	+++	+++	+++	+++	+++
Olanzapine	0 - + (if < 10 mg/day)	Rare	+ (if < 20 mg/day)	++	+++	+	+	+++
Quetiapine	0	Rare	0	++	+ - ++	++	0 - +	++
Risperidone	+ (less if < 4 mg/day)	Rare	+++	+	+ - ++	++	0	+
Ziprasidone	0 - +	Rare	0 - +	0 - ++	0	+ - ++	0	0

Sufficient data for paliperidone, iloperidone and asenapine are not yet available for inclusion in this table.

Adapted from the International Psychopharmacology Algorithm Project (IPAP) algorithm for the treatment of schizophrenia, available at www.ipap.org, copyright 2008 International Psychopharmacology Algorithm Project (IPAP)

associated with negative symptoms and cognitive impairment in schizophrenia,[22,35] the cortical release of these two neurotransmitters, and possibly also acetylcholine, may provide a pharmacological basis for the advantages of atypical antipsychotics over typical neuroleptic drugs in the treatment of these critical symptom domains. In patients with schizophrenia who were stabilized on typical neuroleptics, the addition of mianserin, a 5-$HT_{2A/C}$ and adrenergic α-2 antagonist, was associated with improved neurocognitive performance,[36] adding further support to a role of 5-HT_{2A} receptors in the treatment of cognitive dysfunction in schizophrenia.

The importance of serotonin receptors other than 5-HT_{2A} for the action of antipsychotic drugs has received considerable attention. Activation of 5-HT_{1A} receptors are believed to have a dopamine modulating effect similar to that of 5-HT_{2A} antagonism.[37] Under experimental conditions, 5-HT_{1A} agonists have been shown to stimulate cortical dopamine release[38,39] and, in schizophrenic patients who were stabilized on haloperidol, the addition of tandospirone, a 5-HT_{1A} partial agonist, resulted in improved neurocognitive performance.[40] This effect has also been demonstrated more recently for buspirone, another 5-HT_{1A} partial agonist.[41] Serotonin-1A receptors may be important for cognitive effects of at least some of the atypical antipsychotic drugs that are active at this receptor site. Activity at 5-HT_{1A} receptors is not shared by all antipsychotic drugs (Table 6.2.5.2), however. Antagonism of 5-HT_{2C} receptors also appears to result in cortical dopamine and norepinephrine release, as well as in the nucleus accumbens.[42] The cognitive effects of selective 5-HT_{2C} antagonists added to typical neuroleptic drugs in patients with schizophrenia have not been examined. As is the case with 5-HT_{1A} activity, not all atypical antipsychotic drugs are active at 5-HT_{2C} receptors (Table 6.2.5.2). Like antagonism at histamine H_1 receptors,[43] 5-HT_{2C} antagonist activity may be related to antipsychotic induced weight gain.[44]

Atypical antipsychotics may display regional selectivity in terms of their dopaminergic activity, relative to typical neuroleptics. For instance, atypical antipsychotic drugs appear to preferentially block cortical D_2 receptors, relative to those located in the striatum.[45,46] Haloperidol results in proportionally equivalent D_2 blockade in both brain regions.[47] The atypical antipsychotics also increase the expression of the early intermediate gene c-*fos*, in the prefrontal cortex and the shell of the nucleus accumbens, while sparing the core of the latter region and the striatum. Typical neuroleptic drugs have the opposite effect on c-*fos* expression. Sparing the dorsal

striatum is believed to be related to the low potential for extrapyramidal side-effects of these agents.[2,21]

Clozapine, olanzapine, risperidone, and quetiapine are able to block the interference in prepulse inhibition produced by d-amphetamine, apomorphine, or phencyclidine at doses that do not interfere with locomotor function. Clozapine and M100907 are able to block the effects of phencyclidine, an *N*-methyl-D-aspartate receptor antagonist, on locomotor activity in rodents. This suggests the ability of rat 5-HT_{2A}-receptor blockade to block some of the effects of phencyclidine which is one of the more important models for schizophrenia.[2,21] In a recent single photon emission tomography (SPECT) study, patients with schizophrenia who received treatment with clozapine evidenced reduced NMDA-active radiotracer binding compared with healthy controls, drug free patients with schizophrenia, and patients with schizophrenia who were treated with typical neuroleptics.[48] The extent of involvement of other atypical antipsychotic drugs relative to typical antipsychotics at NMDA receptors and other glutamatergic targets is an area of active interest. Other receptor targets that are of special interest in terms of improving cognitive functioning and selected psychotic symptoms include M1 muscarinic, α-7 nicotinic, and α-1 and α-2 adrenergic receptors.

Administration, pharmacokinetics, and dosage

Administration

(a) Typical antipsychotic drugs

The major uses of the antipsychotic drugs are for the treatment of schizophrenia, mood disorders typically with psychotic features, and senile psychoses.[4,49] Other indications are discussed elsewhere in this book in the consideration of the management of specific disorders, such as Tourette's syndrome, and aggression. The major advantage of the typical neuroleptic drugs is their ability to improve positive symptoms, i.e. delusions and hallucinations. Administration of typical neuroleptic drugs leads to the complete or nearly complete elimination of positive symptoms and disorganization of thought and affect in about 60 to 70 per cent of patients with schizophrenia and an even higher proportion of those with psychotic mania and psychotic depression.[49] The antipsychotic response in schizophrenia and mania is sometimes apparent within a few days in many patients but usually takes up to several weeks or months. A reasonable duration for a therapeutic trial with one of these agents is 4 to 6 weeks. It is not appropriate to switch medications after 1 or 2 weeks, even if a response is not apparent, unless side-effects pose a serious problem. Positive symptoms (delusions and hallucinations) do not respond to typical neuroleptic drugs in about 10 per cent of schizophrenic patients even during the first episode.[50] Another 20 per cent of patients with schizophrenia develop resistance to these agents during the subsequent course of their illnesses.[51] Development of resistance to typical neuroleptic drugs may occur at any time during the course of treatment, even after many years of control of positive symptoms. Such patients are more likely to respond to clozapine[51] or one of the other atypical antipsychotics.[50,51]

The average doses of the typical neuroleptic drugs are given in Table 6.2.5.5. The best results with these drugs in terms of efficacy and side-effects may be expected with the lowest dose needed to produce control of positive symptoms with the fewest extrapyramidal side-effects.[4,49]

There are some patients for whom higher doses are indicated, but most controlled studies have failed to find benefits from high-dose strategies of combining two or more of these agents. Increasing the dose of these agents when patients fail to respond rapidly, for example within days, is not recommended. Augmentation with a benzodiazepine may be useful to decrease anxiety until the lower doses of neuroleptic drugs produce adequate control of positive symptoms.[4,49] Patients who may require higher doses of neuroleptic drugs to respond adequately are at greater risk of hyperprolactinaemic effects, EPS, and tardive dyskinesia and are generally better treated with an atypical antipsychotic drug.

However, the improvement in positive symptoms which is often achievable with the typical antipsychotic drug is only one element in the treatment of schizophrenia and is not sufficient grounds for judging response to be adequate. Additional efficacy factors of major importance are summarized in Table 6.2.5.6.

Tolerability and safety factors, such as compliance, tardive dyskinesia, weight gain, and medical morbidity are also major elements in outcome and are influenced by the choice of a typical or atypical antipsychotic drug. Typical neuroleptic drugs are not as effective for improving primary negative symptoms of schizophrenia in the majority of patients.[52,53] There is a consensus that typical neuroleptic drugs can improve negative symptoms that are secondary to positive symptoms and depression while at the same time possibly causing secondary negative symptoms due to their ability to produce extrapyramidal side-effects.[52] Abnormalities in specific domains of cognition (Table 6.2.5.6) are present in first-episode schizophrenic patients at a moderate to severe level and show slight to moderate, rarely severe, deterioration during the course of illness.[54,55] Approximately 85 per cent of patients with schizophrenia are clinically impaired in one or more domains of cognition.[55,56] Cognition has been shown to be perhaps the most critical determinant of functional capacity among patients with schizophrenia, even more so than positive symptoms.[57] Typical neuroleptic drugs usually do not improve cognitive function.[58] Those typical neuroleptic drugs such as thioridazine and mesoridazine, which have strong antimuscarinic properties, may produce further impairment in some memory functions.[58]

All of the typical neuroleptic drugs are likely to be equally effective in treating either the initial presentation or recurrent psychosis due to breakthrough of symptoms, despite compliance, or because of having stopped medication[4,49,51] First-episode patients with schizophrenia usually require much lower doses than patients with two or more episodes, suggesting some progression of the disease process or development of tolerance to the mechanism of action of these drugs.[59] Doses for more chronic patients should be in the range of 5 to 10 mg haloperidol equivalents per day (Table 6.2.5.5) for up to 4 to 6 weeks unless there is a major need for chemical means to prevent harm to self or others, to decrease excitement, or induce sleep.[60] Auxiliary medications for anxiety and sleeplessness, for example benzodiazepines, may supplement these low doses of antipsychotics.[61]

Parenteral injections of haloperidol, chlorpromazine, or other neuroleptics may be needed for patients who refuse oral medication or where very rapid onset of action is needed to control acutely dangerous behaviours if less restrictive means either fail or cannot be utilized safely. Commonly, haloperidol (2–10 mg) with or without lorazepam (2–4 mg) is delivered intramuscularly every 30

Table 6.2.5.5 Oral dosing of antipsychotic drugs

Typical antipsychotic drugs				
	Equivalent doses (mg/day)	**Starting dose**	**Titration schedule**	**Dose range (mg/day)**
Chlorpromazine[a]	100	15–50 mg BID-QID	As clinically indicated	300–1000 (divided QD-QID)
Fluphenazine[b]	2	0.5–10 mg/day (divided Q6–8 hours)	As clinically indicated	5–20
Haloperidol[c]	2	0.5–5 mg BID	As clinically indicated	5–20
Loxapine	10	10 mg BID	As clinically indicated	30–100
Mesoridazine	50			150–400
Molindone	10	50–75 mg/day divided TID-QID	As clinically indicated	30–100
Perphenazine[d]	10	4–8 mg TID (8–16 mg BID-QID if hospitalized)	As clinically indicated	16–64
Thioridazine	100	50–100 mg TID	As clinically indicated	300–800
Tiotixene	5	2 mg TID	As clinically indicated	15–50
Trifluoperazine	5	2–5 mg BID	As clinically indicated	15–50

For elderly patients, or those with renal or hepatic problems, doses of drug may need to be reduced by one-half or more

[a] Short-acting IM formulation may be given 25–50 mg (may repeat after 1–4 hrs as required); may gradually increase dose up to 400 mg IM Q 4-6 hrs (maximum of 2000 mg/day) may be needed for severe cases

[b] Short-acting IM formulation may be given 2.5–10 mg/day in Q6–8 hr intervals; Depot IM formulation may be given 12.5–25 mg Q 3 weeks

[c] Short-acting IM formulation may be given 2–5 mg Q 1–4 hrs; Depot IM formulation may be given at approximately 10–20 times the stable oral dose Q 4 weeks

[d] Short-acting IM formulation may be given 5–10 mg Q 6 hrs (maximum of 30 mg/day)

Atypical antipsychotic drugs			
	Starting dose	**Titration schedule**	**Dose range (mg/day)**
Aripiprazole[a]	10–15 mg daily	As clinically indicated, every 2 weeks	10–30
Clozapine	12.5 mg QD-BID	Increase by 25–50 mg/day until usual effective dose of 300–450 mg/day after 2–4 weeks	150–600
Olanzapine[b]	5–10 mg daily	As clinically indicated, by 5 mg/day every 7 days	10–30
Paliperidone	6 mg/day	As clinically indicated, by 3 mg/day Q 2–4 week increments, up to 12 mg daily	6–12
Quetiapine	25 mg BID	Increase by 25–50 mg BID-TID on days 2 and 3, to target dose of 300–400 mg daily (QD – TID) by day 4. Further increases as clinically indicated by 25–50 mg BID every 2 days.	300– 800
Risperidone[c]	0.5–1 mg BID	Increase by 0.5–1 mg BID on days 2 and 3, with further dose increases thereafter by 0.5–1 mg increments Q 7 days as required	2–8
Ziprasidone[d]	20 mg BID with food	Increase by 20–40 mg BID every 2 days to target dose of 80 mg (all doses with food)	120–200

For elderly patients, or those with renal or hepatic problems, doses of drug may need to be reduced by one-half or more, and titration may be slower

[a] Short-acting IM formulation may be given at 9.75 mg, though the lower 5.25 mg dose may be indicated in some situations.

[b] Short-acting IM formulation may be given 10 mg as required (may be repeated after 2 hrs, up to 30 mg/day).

[c] Long-acting IM formulation may be initiated at 25 mg Q 2 weeks (continue oral risperidone dose for 3 weeks), with increases as clinically indicated every 4 weeks up to a dose of 50 mg Q 2 weeks

[d] Short-acting IM formulation may be given 10–20 mg as required (may be repeated Q 2–4 hrs as needed, up to 40 mg/day)

to 60 minutes as required, up to three doses. Doses of haloperidol given intramuscularly in such situations generally should not exceed 18 mg per day. Oral medication should be substituted as soon as feasible. If positive symptoms fail to respond to a single trial of a typical neuroleptic drug at adequate doses in patients with schizophrenia, there is evidence that switching to another typical antipsychotic, even of a different chemical class, is unlikely to produce greater control.[4,49,51] This is likely to be true for other indications for the use of antipsychotic agents as well.

In cases of repeated illness relapse due to poor compliance or when patients prefer it, the use of long acting (e.g. depot) injectable antipsychotic medications, typically administered once every 2–4 weeks, may be used. The use of injectable antipsychotic medication has been associated with lower rates of relapse and rehospitalization and greater global improvement compared with oral typical neuroleptics,[62] possibly as a result of ensured drug delivery. Long acting injectable drugs should not be given to ameliorate acute behavioural disturbances.

(b) Atypical antipsychotic drugs

As implied above, there are major advantages for many patients to be treated with the atypical antipsychotic drugs and it is generally

Table 6.2.5.6 Target signs and symptoms for the pharmacological management of schizophrenia

Target	Description	
Positive syndrome	Hallucinations Delusions	◆ Typically the most amenable to treatment with all antipsychotic drugs
Negative syndrome	Avolition Apathy Anhedonia Lack of responsiveness Poor rapport with others Passive social withdrawal Poverty of speech Affective flattening	◆ Robustly correlated with functional impairment in schizohprenia ◆ More difficult to treat pharmacologically, and may required longer to respond than positive signs and symptoms ◆ Pharmacological adjuncts may be needed, though under-studied ◆ Atypical antipsychotic drugs are believed to be more efficacious than typical neuroleptics
Hostility/excitement	Verbal or physical aggression	◆ Typically amenable to treatment with all antipsychotic drugs ◆ Use of parenteral formulation may be required
Mood and anxiety symptoms	Depressed mood Anxious mood Nervousness Panic symptoms Suicidal ideation	◆ Believed to be more responsive to treatment with atypical antipsychotic drugs ◆ Clozapine has demonstrated superiority for treating chronic suicidality in schizophrenia
Cognitive impairment (psychopathological definition)	Disorientation Problems with abstraction Attentional problems Preoccupations Disorganized thought processes	◆ Some domains respond favourably to antipsychotic drug treatment, though response is often incomplete
Cognitive impairment (neuropsychological testing definition)	Working memory Attention/vigilance Verbal learning/memory Visual learning/memory Problem solving Processing speed	◆ Neuropsychological deficits, like negative signs and symptoms, are robustly correlated with functional outcome in schizophrenia ◆ Very difficult to treat with medication alone ◆ Atypical antipsychotic drugs are believed to be superior to typical neuroleptics, though effect sizes are only mild to moderate for the former

recommended that, where possible, these agents be considered as the first-line treatment.[63,64] The atypical antipsychotic drugs are the dominant antipsychotic treatment for schizophrenia, mania, and psychotic depression in clinical practice in many parts of the world. However, there is considerable international variation in their usage. Cost factors may explain part of the variance in the use of these agents within and between countries. The typical neuroleptic drugs are no longer covered by patent protection and are available in inexpensive generic forms. There are a number of patients whose psychosis is adequately controlled by these agents and they (and their families and prescribers) are content to continue them even when informed of the potential advantages of the newer antipsychotic agents. When only the cost of medication is considered, it may seem that fiscal reasons argue for continuation of typical neuroleptic drug treatment since the atypical agents can cost up to 100 times more. In addition, the widely accepted notion of greater overall benefit from treatment with atypical antipsychotic drugs, as opposed to typical neuroleptics, was recently challenged by results from two effectiveness studies. The first of these demonstrated no significant difference in all-cause discontinuation from the study as the primary endpoint, as well as discontinuation for lack of efficacy, between atypical drugs, with the exception of olanzapine, and the typical neuroleptic perphenazine.[65] The latter study reported a lack of significant differences in quality of life between patients who received naturalistic treatment with typical or atypical antipsychotic drugs.[66] Methodological limitations, detailed discussion of which is beyond the scope of this chapter, limit the conclusions that can be drawn from these reports about the relative merits of one class of antipsychotics versus another, both of which are in disagreement with the majority of the clinical literature that documents differences between these broad classes of antipsychotic drugs across a wide range of outcomes. Because medication costs account for, at most, 5 per cent of the total costs of schizophrenia, with the major costs being hospitalization and indirect costs such as lost income and disability income to support patients in the community, more effective and tolerable medications may offset their greater cost.[67,68] As such, atypical antipsychotic drugs are recommended as first line treatments of schizophrenia and related psychotic disorders. Each will be discussed separately.

(i) Clozapine

Clozapine was synthesized in 1959 as part of a project to discover antipsychotic drugs with low potential for extrapyramidal side-effects. It proved to be one of the most interesting and clinically important compounds ever discovered. It was labelled as atypical because of its ability to block amphetamine-induced locomotor activity, one of the most widely accepted models for antipsychotic activity, without producing catalepsy in rodents, the leading model for causation of extrapyramidal side-effects in humans. Subsequent clinical studies showed it to have the lowest extrapyramidal side-effects of any antipsychotic drug known.[5,69] Clinical trials in the 1960s and 1970s suggested it was also superior in

efficacy with regard to control of positive symptoms, but given the standards of clinical trials of that era, these conclusions could not be relied upon.[70] In 1975, 6 years after its introduction in Europe, clozapine's ability to cause granulocytopenia or agranulocytosis was first reported. Six deaths occurred in clozapine-treated patients in a geographically restricted area of Finland over a short period of time. The role of clozapine in these deaths is still uncertain because no other such clustering has ever occurred in Finland, or elsewhere. Nevertheless, clozapine was withdrawn from general use, although it remained available for humanitarian use in patients who had previously received it, for individual cases where it seemed indicated because of its low potential for extrapyramidal side-effects, and for research purposes.[69]

Clozapine was reintroduced in 1989 after it was demonstrated to be superior to chlorpromazine to improve positive and negative symptoms in 300 patients who were resistant to the action of at least three typical neuroleptics.[71] Thirty per cent of the patients treated with clozapine responded after 6 weeks of treatment compared to 4 per cent of the chlorpromazine-treated patients. Subsequent studies have shown that up to 60 to 70 per cent of patients will respond within 6 months of treatment. Patients with shorter duration of illness tend to respond better. Some predictors of response include weight gain and absence of atrophy in the prefrontal cortex.[52] Clozapine has been reported in several studies to reduce the risk of suicide.[52,72] It has been shown in a large number of studies to improve some aspects of cognitive function, especially verbal fluency, immediate and delayed verbal learning and memory, and attention.[16–18]

Because of the side-effect profile of clozapine, it is not generally used as a first-line drug. On the other hand, monitoring the white blood count for the development of agranulocytosis or granulocytopenia, as well as improved methods of treating agranulocytosis, have made it much safer to use. Clozapine is still probably underutilized in many parts of the world. Any patient with an unsatisfactory response to the typical neuroleptics and at least one atypical antipsychotic should be considered for clozapine treatment. This amounts to at least 20 per cent of schizophrenics. Clozapine use may also be considered for patients with schizophrenia who are at high risk for suicide, even if the aforementioned threshold of inadequate response to other drugs has not yet been met.

Clozapine is usually given twice daily, but sometimes more than half of the dose or the entire dose is given at sleep time to minimize daytime sedation. The daily dosage is gradually titrated to the target range described in Table 6.2.5.5. Patients who are treatment resistant may require higher doses. Typical or non-clozapine atypical antipsychotic drugs should be discontinued either before beginning clozapine or by eliminating them over a 1- to 2-week period as the dose of clozapine is increased. Because clozapine produces only about 40 to 50 per cent occupancy of striatal D_2-receptors,[73] and some of its key advantages are believed to be related to its low D_2-receptor blockade, concomitant administration of typical neuroleptic drugs would be predicted to interfere with some of the benefits of clozapine and, thus, should not ordinarily be prescribed with clozapine. However, some patients with persistent positive symptoms despite an adequate trial of clozapine monotherapy might be expected to benefit from the addition of low-dose haloperidol, or its equivalent, to provide additional low level D_2-receptor blockade.

Determination of clozapine plasma levels is useful whenever patients are not responding adequately. If response is inadequate, various approaches to augment response have also been utilized. In addition to adding a low dose of a typical neuroleptic, as mentioned above, it may be useful to augment clozapine treatment with valproic acid or other mood stabilizer (such as lithium, carbamazepine, lamotrigine or topiramate), anxiolytic drugs, or an antidepressant.[52] The choice of augmenting agent is largely driven by symptomatic considerations, or pharmacokinetic interactions in the case of fluvoxamine. However, none of these strategies have strong empiric support. One exception may be the addition of sulpiride, which may result in a significant reduction in symptom burden when added to clozapine.[74] Electroconvulsive therapy (ECT) also resulted in a modest further reduction in symptoms when used in conjunction with clozapine, and appears to be well tolerated.[75] It is difficult to postulate a rationale for adding another atypical antipsychotic, with the exception of amisulpride, because of their similarity in pharmacology to clozapine. It should be discontinued if side-effects are intolerable, or if there is no apparent response after a 6-month trial of clozapine alone and subsequent trials with augmentation therapy. Clearly, further studies involving clozapine partial- or non-responders are urgently needed. It should be noted that discontinuation of clozapine can precipitate a severe relapse even when clozapine is slowly tapered.[76]

(ii) Risperidone

Risperidone is useful as a first-line drug for the treatment of all forms of schizophrenia, including residual schizophrenia.[77–79] Definitive data are lacking for its efficacy in patients who are neuroleptic resistant or who have failed to respond to other atypical antipsychotics, including clozapine.[80] Clinical experience is not supportive of widespread efficacy in these groups but there may be some responders. However, risperidone may be useful in patients who fail to tolerate other antipsychotic agents because of side-effects not shared by risperidone, such as anticholinergic effects. Risperidone is well-tolerated in low doses by the elderly and has been widely used in the United States for the treatment of a variety of senile psychoses.[81,82] Its efficacy against haloperidol was established in a series of multicentre trials which demonstrated advantages for risperidone in overall psychopathology in mainly chronic schizophrenic patients in an acute exacerbation at doses in the 6 to 8 mg/day range.[11,77] However, these doses have proven to be higher than is needed for most patients in clinical practice, possibly reflecting some of the problems in generalizing from controlled clinical trials. The doses for schizophrenia most often used in non-elderly adults are now 4 to 6 mg/day. First-episode patients may not tolerate higher doses (e.g. above 5 mg per day), and some may respond to as little as 1 to 2 mg/day. Some treatment-resistant patients may need doses higher than 6 mg/day; however, the results of clinical studies in this population have been mixed. Whether prolonged trials, i.e. up to 6 months, are useful in such patients, as they are in clozapine patients, is not yet known.

Beyond treatment of acute symptoms, the position occupied by risperidone as a first line treatment option is also supported by long term maintenance phase and relapse prevention studies. For instance, relative to haloperidol, risperidone has also been associated with a lower risk of relapse (34 vs. 60 per cent) over a

minimum of 12 months of treatment.[83] In another study that |retrospectively compared rates of rehospitalization for patients who received treatment with risperidone, olanzapine, or typical neuroleptics, rehosptialization rates for risperidone and olanzapine were similar, and both were significantly less than those of patients treated with typical neuroleptics.[84]

Risperidone is usually initiated at low doses (e.g. 1–2 mg daily) and is titrated into the dosage range provided in Table 6.2.5.5. The medication is often initiated in twice daily dosing; however, because its primary active metabolite, 9-OH risperidone, is pharmacologically equivalent to its parent drug and because it has a longer elimination half-life, once daily dosing is also possible. Risperidone is available in soluble wafer and liquid forms, which may be advantageous for patients who have swallowing difficulties or require taking their medication in a non-pill form for other reasons, including their own preference.

For patients who have a history of poor compliance leading to frequently relapsing illness, or for those who prefer it, a long acting injectable form of risperidone is available (Table 6.2.5.1) for administration typically every two weeks. For treatment responsive patients, response may be expected to occur in the 25–50 mg (per every two week dose) range,[85] however, oral risperidone must be continued through at least the first 3 weeks of treatment with the long acting injectable form before being slowly tapered. Supplementation with oral medication may be required when the dose of the long-acting drug is upwardly adjusted due to breakthrough psychotic symptoms. As is the case with long-acting injectable typical antipsychotics, long-acting injectable risperidone should not be used acutely to control dangerous behaviours.

Risperidone has more of a tendency to produce extrapyramidal side-effects than any of the other atypical antipsychotics but this can be minimized by using the lowest dose which controls positive symptoms and adding an anticholinergic drug, if necessary.[82] Addition of a typical neuroleptic to risperidone will increase the risk of extrapyramidal side-effects. Risperidone is not well tolerated by patients with Parkinson's disease because of extrapyramidal side-effects. There are some data suggesting the risk of tardive dyskinesia in patients with schizophrenia, and especially the elderly, with risperidone is less than that of the typical neuroleptic drugs.[86]

Among atypical antipsychotic drugs, risperidone and paliperidone appear to be the most liable in terms of increasing prolactin release. As is the case with EPS, the effect of risperidone on prolactin concentration appears to positively correlate with dose.[25] The changes may occur in both men and women; however, the greatest elevations appear to occur among women. Elevations in prolactin levels as a result of treatment with risperidone do not always translate into clinical symptoms such as sexual dysfunction or gynaecomastia in men and menstrual changes and breast discharge in women; however, patients should be monitored clinically for these effects, and prolactin concentrations measured if these symptoms occur.

The issue of whether the improvement in negative symptoms by risperidone and other atypical antipsychotic drugs is due to an effect on so-called primary negative symptoms versus secondary negative symptoms has been much debated. Data from large multicentre trials of risperidone versus clozapine show an effect on primary negative symptoms as residual change left after adjusting for improvement due to decreases in positive or depressive symptoms and extrapyramidal side-effects.[87] In addition, results from a meta-analysis of 6 studies comparing risperidone to typical neuroleptic treatment indicated greater response rate for negative symptoms (defined as achieving >20 per cent reduction in negative symptom burden) as well as greater reduction in anxious/depressive symptoms among risperidone treated patients.[88] Risperidone has a greater ability to improve cognition in schizophrenia than the typical neuroleptic drugs.[17] Improvement in working memory has been the strongest finding, while improvements in attention, executive function, and verbal learning and memory have also been reported. Risperidone has been shown to be a cost-effective treatment for schizophrenia,[89] especially in its long-acting injectable form,[90] and to improve quality of life,[91] firm conclusions about relative cost-effectiveness between atypical antipsychotic drugs for non-treatment-resistant schizophrenia are difficult to draw at this time. Further research in this important area is needed.

In summary, risperidone is a first line pharmacological treatment of schizophrenia and other forms of psychosis. It may produce significant advantages over typical neuroleptic drugs with regard to negative symptoms, cognition, and extrapyramidal side-effects, but it does produce dose dependent increases in EPS risk, and increases in serum prolactin levels resembling those of typical antipsychotic drugs. It should be used at the lower doses where possible. A long-acting injectable form of this medication should be considered a first line treatment option in cases of frequent relapse due to poor medication compliance.

(iii) Olanzapine

Olanzapine is indicated as a first-line treatment for all forms of schizophrenia[63,92] with the caveat that it has not been shown to be as effective as clozapine in neuroleptic-resistant patients at conventional doses.[8,60] However, some patients of this type do respond to olanzapine,[93] perhaps at high doses.[94,95] There are no means yet to determine which of this group of patients will respond to olanzapine (or risperidone) so some clinicians may elect a trial with either of these agents before considering clozapine.

The efficacy of olanzapine in treating psychosis and negative symptoms in patients with an acute exacerbation of schizophrenia has been firmly established in a variety of large-scale, multicentre trials.[92] In these trials, olanzapine at doses of 10 to 20 mg/day has been superior to placebo and equivalent or superior to haloperidol in some measures of total psychopathology, positive, or negative symptoms. For example, in the North American multicentre trial, high-dose olanzapine (15 ± 5 mg/day) was superior to haloperidol (15 ± 5 mg/day) in the treatment of negative symptoms.[96] The effect of olanzapine to improve negative symptoms was found to be on primary rather than secondary negative symptoms.[97]

Olanzapine has also been found to be effective as a maintenance treatment of schizophrenia.[98] The estimated relapse rates, defined as the need for hospitalization, during a 1-year period in three studies of patients receiving olanzapine for maintenance treatment were 19.6 to 28.6 per cent. These rates were significantly lower than those in patients receiving placebo, ineffective doses of olanzapine, or haloperidol.[98] Olanzapine has some efficacy in treating anxious and depressive symptoms,[99] as well as cognitive dysfunction,[17,18] associated with schizophrenia or schizoaffective disorder. Pharmacoeconomic studies and investigations of medication effects on quality of life measures indicate that olanzapine has

a beneficial cost-outcome profile. For instance, in one investigation, the higher cost of olanzapine relative to haloperidol was offset by olanzapine-treatment associated reductions in rehospitalization and overall treatment costs.[100] Olanzapine treatment has also been associated with better outcomes as assessed by overall and health-related quality of life relative to haloperidol.[101] As mentioned above, olanzapine was found to be the most effective antipsychotic drug in the recent Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) phase I study.[65]

The average clinical dose of olanzapine is 12.5 to 20 mg/day but many patients may be expected to respond to lower doses (e.g. 10 mg daily).[102] A principle advantage of olanzapine is its once daily dosing and the feasibility of starting the medication at a dose that is clinically effective for most patients. Doses higher than 20 mg/day are rarely more effective than lower doses, especially for non-refractory cases. Augmentation of olanzapine with typical neuroleptic drugs or risperidone should be done sparingly to avoid extrapyramidal side-effects and possibly compromising efficacy. Olanzapine is also available in a soluble wafer form that may be preferred to the pill form by some patients, especially those with swallowing difficulties and related problems.

For acute situations where rapid control of agitation, hostility or other dangerous behaviours is required, olanzapine is available as a short-acting injectable medication (Table 6.2.5.1).[103] Similar to short-acting typical neuroleptic drugs, the medication is delivered intramuscularly. Doses of 5 to 10 mg per injection may be given, depending on the severity of the target behaviours. A long-acting formulation is in clinical testing.

In summary, olanzapine has found wide acceptance as an atypical antipsychotic drug because of its once-a-day administration, efficacy for negative symptoms, improvement in cognitive function, and low extrapyramidal side-effect profile. Significant weight gain and other metabolic effects may be a problem for some patients, as will be discussed below.

(iv) Quetiapine

Quetiapine has been shown to be as effective as typical antipsychotics, with fewer extrapyramidal side-effects and no effect on serum prolactin levels.[104,105] Part of the reason for this may be that quetiapine and clozapine both appear to bind more loosely to striatal D2 receptors than other antipsychotic drugs, and that both drugs show antipsychotic activity at D2 receptor occupancies that are well below the 60 per cent threshold identified for most other antipsychotic drugs.[106] In spite of this similarity with clozapine, quetiapine does not appear to have efficacy comparable to clozapine for treatment-resistant patients.

The efficacy of quetiapine for acute phase schizophrenia is supported by results from several randomized, controlled trials that documented superiority of quetiapine relative to placebo across several doses, with some patients with some patients responding to 150 mg/day and others requiring 750 mg/day.[104] For instance, in one high- (750 mg/day) versus low-dosage (250 mg/day) study, both dosage groups evidenced greater reduction in positive symptoms relative to placebo; however, the differences were significant only for the high-dose group.[105] In another study that assessed multiple fixed doses of quetiapine (75 to 750 mg/day) compared with haloperidol and placebo, significant differences in improvement over placebo for quetiapine were observed in the dosage range of 150 to 750 mg/day.[104]

Quetiapine's effect on negative symptoms continue to be investigated. One placebo controlled comparison documented improvements in negative symptoms with quetiapine treatment across a wide range of doses, with the greatest improvement reported at 300 mg daily.[104] In the high- vs. low- dose study reviewed above, the high-dose group also experienced greater improvement in negative symptoms relative to placebo.[105] Like risperidone and olanzapine, quetiapine appears to improve depressive symptoms[107] and certain cognitive deficits[17,18] associated with schizophrenia or schizoaffective disorder. The improvements in cognition with quetiapine appear to be superior to those of haloperidol.[108]

These results suggest that, overall, the greatest improvement in positive and negative symptoms may occur when quetiapine is used at the higher end of its dosage range. The average clinical dose appears to be between 300 and 500 mg/day, usually given twice daily though some benefit from the medication when given only once daily. The effects of using higher doses for patients who do not respond adequately to these doses are uncertain. A titration of the dosage is required after initiating the medication. From the viewpoint of EPS and hyperprolactinaemic effects, quetiapine appears to confer only low risk. As such, it, like clozapine, appears to be well tolerated even among patients with idiopathic Parkinson's disease.[109] Sedation may be a limiting side effect for some, especially during dosage titration. Weight related, metabolic, and other adverse effects will be discussed in greater detail below.

In summary, quetiapine also appears to be effective for a wide range of schizophrenia-associated symptoms and confers a lower level of risk in terms of antidopaminergic adverse effects. The dosage range of this medication may be quite wide, though patients may have a greater chance of benefiting from the medication at the higher end of this range.

(v) Ziprasidone

Ziprasidone has a varied receptor occupancy profile. Like most atypical antipsychotic drugs, it displays high affinity 5-HT2A binding coupled with relatively lower affinity D2 receptor binding. Ziprasidone is also a 5-HT1A agonist, as well as both a serotonin and norepinephrine reuptake pump inhibitor.[110] This profile predicts a wide range of pharmacological activity against core psychotic symptoms, negative and affective symptoms, as well as neurocognition.

Ziprasidone, like quetiapine, has been shown to be superior to placebo for the reduction of total psychopathology and positive and negative symptoms.[111,112] There is limited evidence to suggest superiority over typical neuroleptics with regard to improvement in positive and negative symptoms.[111,113] Studies of multiple fixed doses of ziprasidone vs. haloperidol at conventional doses indicate that ziprasidone yields similar efficacy to haloperidol for reducing positive symptoms and global psychopathology at a dose of 160 mg/day.[113] Doses greater than 160 mg/day have not been systematically investigated.

Ziprasidone significantly improved negative symptoms and reduced the risk of relapse compared to placebo in a 1-year maintenance study in stable hospitalized chronic schizophrenic patients.[114] These maintenance phase effects were not dependent on the daily dose of ziprasidone. In a 28-week comparison with haloperidol, the two groups evidenced similar overall effects for positive symptoms; however, between groups differences were

documented favouring ziprasidone for negative symptoms and EPS.[115] Ziprasidone was effective against depressive symptoms associated with schizophrenia in one study at a dose of 160 mg/day.[116] Significant improvements in multiple cognitive domains have been reported among ziprasidone treated patients in a variety of treatment contexts.[117] Such changes appear to be unrelated to improvements in other symptoms of schizophrenia. Ziprasidone treatment has been associated with significant improvement in quality of life measures in one post hoc data analysis.[118] Further investigation of the effect of ziprasidone on health related quality of life and similar outcomes are warranted. Ziprasidone treatment of schizophrenia appears to be cost-effective relative to no treatment.[119] Further cost-benefit studies are needed.

The dose range of ziprasidone for acute treatment appears to be between 80 and 160 mg/day, higher doses within this range may be more effective (Table 6.2.5.5). Doses greater than 120 mg/day appear to be required to achieve >60 per cent dopamine D2 receptor blockade,[120] the D2 receptor occupancy threshold that appears to coincide with efficacy against positive symptoms, as presented earlier. The medication is usually given twice daily, although some may take the medication once daily at night time. A titration of the total daily dose into the recommended range is required after initiating the medication. One critical aspect of medication administration for ziprasidone is the requirement that the medication be taken with food. There appear to be profound differences in bioavailability at equivalent doses between the fed and unfed state.[121] A full meal, as opposed to a light snack, appears to be required. Therefore, patients are encouraged to take their medication with meals.

A short-acting intramuscular formulation of ziprasidone has been developed which should be useful in situations where more rapid action is needed. This formulation is available in two doses (10 and 20 mg), the preferred dosage being 20 mg due to significantly greater reduction in agitation relative to lower dose.[122] The use of the short-acting injectable form can facilitate a transition to oral medication, and may reduce that time required to titrate the daily dose of ziprasidone to one that is likely to be effective.

Ziprasidone appears to be well tolerated. Treatment-emergent EPS burden is low.[112,113] Initial problems with somnolence or behavioural activation are usually self limited, although temporary use of clonazepam or other benzodiazepine at low doses may improve tolerability, especially during the titration phase, should the latter occur. Importantly, data from both short- and long-term studies indicated that ziprasidone is not associated with clinically significant changes in weight, glycaemic measures, or markers of lipid homeostasis.[123]

Ziprasidone can result in partial blockade of the slow potassium rectifier current in the cardiac conduction system, which may result in prolongation of the QTc interval (discussed in greater detail below).[124] On the other hand, there is only one case report of ziprasidone induced *torsades de pointes*, the risk of which is believed to be increased if the QTc interval is >500 msec.[125] There have also been no reported deaths in the context of overdose with the medication. Under routine circumstances, screening electrocardiograms are not required. Nevertheless, caution may be warranted for individuals who are at risk for significant prolongation of the QTc, including patients who take medications other than ziprasidone that prolong the QTc. Concomitant use of CYP-450 3A4 inhibitors does not appear to pose a significant risk.[126]

In summary, ziprasidone appears to be a useful additional atypical antipsychotic agent because of its favourable side-effect profile, including no weight gain—a major problem with olanzapine and clozapine-and no prolactin elevation, which is a less serious side-effect of risperidone. Patients should be instructed to take the medication with food. Ziprasidone treatment may result in an increase in the QTc interval; however, in a great majority of cases, this is not clinically significant.

(vi) Aripiprazole

Among the atypical antipsychotics, aripiprazole is pharmacologically unique in that it combines partial D2 receptor agonism with high potency 5-HT2A antagonism. Because it is a partial D2 receptor agonist, it binds to the receptor with full affinity, but exerts only a fraction of the intrinsic activity at that site that would be expected of endogenous dopamine. As such, in states of relative dopamine excess, as is believed to be the case in the ventral striatum among schizophrenic patients who experience positive symptoms, aripiprazole is believed to exert relative antagonist activity at D2 receptors.[127] Conversely, it is believed to act primarily as an agonist in cases of relative hypo-dopaminergia, as may be the case in the prefrontal cortex in patients with schizophrenia.[22] For this reason, aripiprazole and other D2 partial agonists in development are sometimes referred to as 'dopamine stabilizers.' Aripiprazole also functions as a potent 5-HT1A partial agonist.[128]

The efficacy of aripiprazole in the treatment of acute schizophrenia at doses ranging between 10 and 30 mg (taken once daily) was established on the basis of four short-term randomized controlled studies.[129] Relative to placebo, efficacy against negative symptoms was also demonstrated.[130] Long term superiority of aripiprazole (vs. placebo) for relapse over 26 weeks[131] and medication compliance and symptom response (vs. haloperidol) for up to 52 weeks has also been established.[133] One study reported on the effectiveness of flexibly dosed aripiprazole (15–30 mg daily) among patients with schizophrenia with a history of resistance to treatment with olanzapine or risperidone.[133] The utility of aripiprazole in the setting of well defined treatment refractory schizophrenia requires further systematic investigation.

The overall effectivenss of aripiprazole has been evaluated in two recent studies. One study reported the effectivenss of flexibly dosed aripiprazole over 8 weeks of treatment (53 per cent response rate at mean endpoint dose = 19.9 mg/day) among a cohort of patients with chronic schizophrenia and schizoaffective disorder under routine treatment conditions in a community healthcare setting.[134] The second study documented comparable effectiveness with olanzapine over 52 weeks of treatment, with more favourable effects for aripiprazole for several metabolic adverse effects.[135] As is the case with other atypical antipsychotic drugs, early evidence indicates that aripiprazole may also have beneficial effects on neurocognitive performance in patients with schizophrenia at recommended doses.[136] Further study of the effects of this medication on cognition is indicated. Clinically relevant improvement in quality of life has been documented in one study.[137] More cost-outcome studies of aripiprazole are needed.

Treatment with aripiprazole is usually initiated with 10 to 15 mg daily, although some patients may not be able to tolerate these doses due to agitation, nausea or vomiting. The dose can be increased up to 30 mg if needed, and tolerated. An oral solution form is also available. Aripiprazole is also available in a soluble

wafer as well as an acute intramuscular form. The acute injectable form appears to be effective in the dosage range of 5.25 to 15 mg.[138] The recommended dose is 9.75 mg.

Aripiprazole is generally well tolerated, with an adverse effect profile similar to placebo in short-term studies involving patients with acute schizophrenia and in longer-term studies of chronic, stable patients.[129] As is the case with all atypical antipsychotic drugs, the EPS burden is lower than that of typical neuroleptics. Some patients, however, may encounter this effect if the dose is started too high or if the titration is too aggressive. Aripiprazole treatment does not appear to significantly increase, and may cause a slight decrease, in prolactin levels.[25] Importantly, short- and longer-term studies indicated that, similar to ziprasidone, aripiprazole is not associated with a high risk of significant changes in weight, glycaemic measures, or markers of lipid metabolism.[123]

In summary, aripiprazole appears to be effective as an acute and long-term maintenance treatment for schizophrenia and related psychotic disorders at recommended doses, though some patients may require higher doses. Aripiprazole was initiated in most studies at doses of 10 to 15 mg once daily; however, some patients may require a slower titration following a lower starting dose. This medication is available in many dosing forms, all of which appear to be very well tolerated. Important benefits from a tolerability viewpoint include very low rates of prolactin elevation, and low risk of weight gain and metabolic adverse effects.

(vii) Paliperidone

Paliperidone is the most recent antipsychotic drug to gain approval for use in the US. It is the 9-OH metabolite of risperidone, which has a longer elimination half-life than the parent compound, as reviewed above. Additionally, paliperidone, which is pharmacologically similar with regard to receptor occupancy profile to risperidone, is available commercially in an extended release form.

The short-term efficacy of paliperidone has been established on the basis of three randomized, placebo controlled studies, two of which have been published,[139,140] that investigated the clinical efficacy of 5 fixed doses (3, 6, 9, 12, and 15 mg) given once daily relative to placebo. In each of these studies, all doses of paliperidone were superior to placebo for reducing global psychopathology and positive symptoms, as well as negative symptoms, anxious/depressive symptoms associated with schizophrenia, and hostility/excitement. In addition, all doses of paliperidone were superior to placebo for improving measures of functional capacity. Paliperidone has not been investigated in the context of treatment refractory schizophrenia. Paliperidone appears to be effective for the prevention of relapse on the basis of one published study.[141]

The recommended starting dose of paliperidone in its extended release form is 6 mg, given once daily. Even though there is a suggestion of greater improvement in terms of symptom reduction from the paliperidone registration studies at higher doses, the adverse effect burden may also be greater (discussed below). Doses may be upwardly adjusted at 3 mg/day increments, up to 12 mg daily. Investigatons of doses greater than 6 mg daily for patients who do not respond adequately have not been performed. A long acting injectable form of paliperidone is currently in development.

Pooled analysis of data from the three short-term, acute phase studies indicate that paliperidone appears to be well tolerated, and that the recommended starting dose (6 mg once daily) was associated with a placebo-like overall adverse effect profile.[142] At doses higher than 6 mg, there appeared to be an increase in the reported incidence of EPS, though not to the degree at any of the doses tested that would be expected with typical neuroleptic treatment. Elevations in prolactin levels appear to be consistent with those observed with risperidone treatment, and appear to be greater in magnitude at higher doses. This effect appears to be especially pronounced among female patients. There were no significant changes from baseline in weight or measures of lipid or glucose handling. Data from long term investigations will provide a more comprehensive picture of paliperidone's tolerability profile.

In summary, paliperidone, the newest atypical antipsychotic drug, appears to be safe and effective for both short- and long-term treatment of schizophrenia. The EPS and prolactinemic adverse effect burden may resemble that of risperidone, but this notion requires prospective investigation. Paliperidone in its extended release form can be started at a clinically effective dose. A long-acting injectable form is currently in development.

(viii) Amisulpride

The efficacy of amisulpride for the treatment of positive symptoms has been established over a wide dosage range (200 to 1200 mg daily) in treatment studies of up to 12 months duration.[143] In general, it appears that higher doses (above 400 mg/day) are effective for treating patients with predominately positive symptoms, although efficacy against negative symptoms has also been demonstrated in this dosing range.[144,145] Low-dose amisulpride (≤ 300 mg/day) has been shown to be effective in treating negative symptoms in schizophrenics with predominantly negative symptoms.[146–148] Evaluation of the effect of amisulpride in patients with minimal extrapyramidal side-effects and positive symptoms suggests amisulpride is able to improve primary negative symptoms, even in patients with deficit syndrome schizophrenia.[146–148] At both dose ranges, amisulpride produces minimal extrapyramidal side-effects, but may result in increased prolactin levels. Amisulpride has been directly compared with haloperidol, and with both risperidone and olanzapine. In general, amisulpride appears to be as clinically effective as all three drugs for treating positive symptoms. Improvement in negative symptoms is superior to haloperidol and appear equivalent to olanzapine and risperidone. Improvement in depressive symptoms related to schizophrenia were also equivalent between amisulpride and olanzapine; however, in a meta-analysis of three studies, amisulpride was shown to be superior to high dose risperidone (8 mg daily) and haloperidol.[149] It is unknown if it is effective in neuroleptic-resistant patients. Because its pharmacology is quite distinct from that of the 5-HT_{2A}-based receptor antagonists previously discussed, amisulpride may be useful in patients who fail to tolerate that class of drugs. Amisulpride has also been demonstrated as being superior to haloperidol on quality of life measures and global functioning.[150] Pharmacoeconomic analyses indicate that amisulpride has a beneficial cost-outcome profile.[151]

(c) Iloperidone and asenapine

Clinical trials are currently taking place with both of these atypical agent to determine its efficacy and side-effect profile compared with typical and other atypical antipsychotic drugs. Like most other atypical antipsychotic drugs, both asenapine and iloperidone combine potent 5-HT2A antagonism with less potent D2 receptor antagonism. Asenapine is currently undergoing investigation in phase 3

clinical trials. Short-term, acute phase efficacy of iloperidone for symptom reduction relative to placebo has been demonstrated at daily doses of 20 to 24 mg daily, with less certain effects at lower doses.[152] Long-term investigations thus far indicate a low incidence of EPS, lack of effect on prolactin release, and minimal effect on body weight.[152] It has the potential to be made into a long-acting form, which would be of great value.

Pharmacokinetics, metabolism, and drug interactions

(a) Typical neuroleptics

The typical neuroleptics are well absorbed when administered orally or parenterally. Intramuscular injection leads to more rapid and higher plasma levels. Peak plasma levels are reached in 30 min after intramuscular injection and 1 to 4 h after oral injection. Steady state is achieved in 3 to 5 days. The half-life for elimination is in the range of 10 to 30 h. Substantial amounts of the antipsychotics are stored in lipids, including in the brain. There is controversy about how long these drugs persist in the system after discontinuation. By the criterion of elevations of plasma prolactin levels, the concentrations are too low to be biologically active within 48 h after discontinuing oral medication. On the other hand, some rodent and human positron emission tomography studies suggest that long-acting forms of haloperidol or fluphenazaine may persist for 1 to 3 months. Metabolism of the typical and atypical antipsychotic drugs occurs in the liver for the most part, via conjugation with glucuronic acid, hydroxylation, oxidation, demethylation and sulphoxide formation. Much of this metabolism occurs via the hepatic cytochrome (CYP)-450 enzymes, particularly the 2D6 and 3A4 sub-families for most drugs. Some metabolites have significant biological activity, for example mesoridazine, and 7-hydroxyloxapine. Dosing of typical neuroleptic medications are determined by clinical effects, less by pharmacokinetic factors.

Pharmacokinetic drug-drug interactions at the level of protein binding are expected to be minimal, even though most typical neuroleptics are tightly bound to plasma proteins. Even so, appropriate therapeutic monitoring of drugs that are also tightly bound to plasma proteins but have a narrow therapeutic index (e.g. warfarin, digoxin, phenytoin) when used in conjunction with typical neuroleptics is warranted. Interactions at the level of the CYP-450 system are also thought to be minimal for most agents. Because smoking is so common among patients with schizophrenia and because smoking can be associated with potent induction of CYP-450 1A2 isoenzymatic activity, dosage adjustments may be needed for selected antipsychotic drugs during any changes in smoking status. Other combinations with typical neuroleptics may be worth avoiding for other reasons, such as increased central nervous system affects (e.g. anxiolytics, other central nervous system depressants, anticholinergics, certain antihypertensive drugs), increased EPS (e.g. metoclopramide, D2 blocking anti-nausea drugs, caffeine), impaired cardiac conduction (certain drugs combined with typical neuroleptics known to prolong the QTc interval), and neurotoxicity (lithium), especially among individuals who are more advanced in age.

(b) Atypical antipsychotic drugs

(i) Clozapine

There are wide variations in the pharmacokinetics of clozapine in patients. The average half-life is 6 to 12 h. Plasma concentrations are higher in Chinese patients than in Caucasian patients, in non-smokers than smokers, and in females than males. The bioavailability is not affected by food intake, metabolism occurs mainly in the liver. The chief metabolite is *N*-desmethylclozapine, which has some biological activity. Clozapine is metabolized by CYP1A2, and several potential drug-drug interactions are thus possible. When agents that induce CYP-1A2 are prescribed or ingested, close monitoring of patients for a worsening of symptoms is warranted. Plasma levels of clozapine of approximately 350 ng/ml are more often associated with good response than lower levels,[153] and should be checked in such cases. Upward adjustment of the clozapine daily dosage will typically correct the problem. On the other hand, if a CYP1A2 inducer is discontinued or a potent inhibitor is added, this may result in a rise in clozapine concentration, and an increase in adverse effect risk.[154] Caution may also be warranted for drugs that are potent inhibitors of CYP 2C19 and CYP3A4.[155] In addition, caution is warranted when considering concomitant use of drugs which can also cause bone marrow suppression (e.g. carbamazepine) or precipitously drop seizure threshold.

(ii) Risperidone

Risperidone is well absorbed from the gut and is extensively metabolized in the liver by CYP2D6 to 9-hydroxyrisperidone in approximately 92 to 94 per cent of Caucasians.[156] Thus, 9-0H risperidone is an active species in the majority of patients. About 6 to 8 per cent of Caucasians and a small proportion of Asians have a polymorphism of the CYP2D6 gene, which leads to poor metabolism of risperidone. For poor metabolizers of risperidone, the active moiety is mainly the parent compound. The half-life of the 9-hydroxy metabolite is about 21 h whereas the half-life of risperidone is about 3 h. Thus, risperidone can be used on a once-a-day schedule for normal metabolizers whereas multiple doses are needed for those who are poor metabolizers. Risperidone should be titrated from 2 to 5 mg/day over at least a 3-day period to minimize hypotensive and neuro muscular side-effects. Drugs known to induce or inhibit CYP2D6 and 3A4 may alter plasma levels of risperidone; thus, close monitoring is advised when such agents are added to ongoing risperidone treatment.

(iii) Olanzapine

Olanzapine has a half-life of 24 to 30 h, which indicates that single daily administration is adequate.[92] The metabolic pathways of olanzapine involves CYP2D6, CYP1A2 and flavin-containing mono-oxygenases, as well as *N*-glucuronidation. It has a low potential for drug-drug interactions and requires extremely high concentrations not likely to be achieved under clinical conditions to inhibit cytochrome P-450 systems. Plasma levels of approximately 9.3 mg/ml have been reported to predict better clinical response to olanzapine in inpatients with an acute exacerbation.[157] Drugs that are known inducers or inhibitors of CYP1A2 may significantly affect plasma levels of olanzapine and alter its clinical effects at a given dosage; thus, active monitoring of symptoms and adverse effects is indicated if such agents are added. As is the case with clozapine, gender and smoking status may influence olanzapine levels leading to adjustment in dosage.[153]

(iv) Quetiapine

Quetiapine is well absorbed and is approximately 83 per cent protein bound.[158] Quetiapine is absorbed better after eating.[158] It has a half-life of 6 h. It is metabolized in the liver by CYP3A4 to inactive metabolites. Quetiapine has significant interactions with several inducers and inhibitors of CYP3A4. Co-administration with

these agents may require dosage adjustment. Thioridazine may also significantly increase the clearance of quetiapine,[159] thus necessitating dosage adjustment. Despite the short half-life, a clinical trial compared three dosing regimens (450 mg/day given in two or three divided doses, and 50 mg/day given twice daily). Both of the higher-dose groups were superior to the low-dose group and there were no differences between the two high-dose schedules. Once daily dosing, which is also a common dosing strategy for quetiapine, is also supported in the literature.[160] The feasibility of such a dosing schedule, which does not seem to be predicted by peripheral pharmacokinetic parameters, is possible because quetiapine appears to interact centrally with both D2 and 5-HT2A receptors much longer than its 6 h elimination half-life.[161]

(v) Ziprasidone

Ziprasidone has a half-life of 4 to 10 h. Twice-daily administration is possible despite this relatively short half-life. Clinically, many patients are prescribed this medication only once daily. Regardless, ziprasidone should always be taken after eating in order to facilitate absorption. About two-thirds of ziprasidone is metabolized by aldehyde oxidase into inactive metabolites. The remainder is metabolized by CYP3A4 and CYP1A2 into inactive metabolites. At the current time, there are no known drug interactions with ziprasidone at the level of aldehyde oxidase, since enzymatic activity does not appear to be altered by coadministered drugs. Although CYP3A4 appears to play only a minor role in the metabolism of ziprasidone, potent inhibitors or inducers of CYP3A4 may significantly alter plasma concentrations of ziprasidone,[162] and may thus necessitate an adjustment in dosage. The use of concomitant medications that may prolong the QTc interval should be avoided. Ziprasidone is contraindicated for patients with a history of known QT prolongation, recent acute myocardial infarction, or uncompromised heart failure.

(vi) Aripiprazole

Aripiprazole is well absorbed from the gut, and has an elimination half-life of 75 hours. It is metabolized primarily by CYP3A4 and 2D6 isoenzymes into an active metabolite, dehydro-aripiprazole, which has a half-life of 94 hours. This pharmacokinetic pattern supports once daily dosing. Because aripiprazole is metabolized by CYP3A4 and 2D6, known inhibitors or inducers of these isoenzymes may result in increased or decreased clearance of aripiprazole and dehydro-aripiprazole.[163]

(vi) Paliperidone, Iloperidone, and Amisulpride

Paliperidone is currently marketed in the US and abroad only in an osmotically controlled extended release formulation, which results in steady release of active drug over a 24 hr period. Hepatic metabolism is not considered a major route of clearance. Paliperidone is converted into metabolites that are not believed to contribute significantly to its overall pharmacological activity. Few significant drug-drug interactions at the level of the CYP450 system are therefore anticipated. Even so, the plasma concentration of paliperidone may be altered by drug interactions at CYP3A4.[164]

Iloperidone has a half-life of 12 to 15 h. Its absorption is not affected by food. It should be titrated slowly because of orthostatic hypotension, and close monitoring is warranted when it is combined with antihypertensive drugs or drugs that are associated with orthostatic effects. The optimal dose has not yet been established but is likely to be in the 5 to 10 mg/day range. Amisulpride has a half-life of 10 to 15 h. It is well tolerated. As yet, there are no known drug interactions. More information regarding the metabolic handling and potential for drug-drug interactions for both of these medications is anticipated as they continue to be further developed.

Side-effects

Typical neuroleptics

The adverse effects that are most routinely concerning for antipsychotic drug treatment are extrapyramidal adverse effects (EPS), especially for typical neuroleptic mediations. For typical neuroleptics, high-potency drugs such as haloperidol and fluphenazine are more likely to produce EPS than low-potency agents such as chlorpromazine and thioridazine. The latter may have lower potential for extrapyramidal side-effects than other typical neuroleptics because of its relatively higher affinity for muscarinic receptors. Atypical antipsychotic drugs are less likely to cause EPS during acute and long term tretment. There are a wide range of extrapyramidal side-effects produced by the typical neuroleptics, including dystonic reactions when first administered, akathisia during the first 2–3 weeks, parkinsonism during the first several weeks with variable persistence, neuroleptic malignant syndrome at any time point but usually in the initial weeks, and tardive dyskinesia.

Dystonic reactions due to neuroleptic drugs can be treated with parenteral anticholinergic agents or diphenhydramine, an antihistamine with some anticholinergic properties. The use of anticholinergic and other agents to manage parkinsonism due to typical neuroleptic drugs will be discussed subsequently.

Akathisia may be the most common of the EPS effects, occurring in up to 70 per cent of patients treated long term with haloperidol.[165] The term refers to a subjective uncomfortable experience of motor restlessness which is relieved by movement. As such, patients will complain of discomfort, and manifest increases in psychomotor behaviour. These symptoms can be so distressing as to increase the risk of agitation or even suicidal behaviours.[166] Although patient age does not seem to influence risk of developing akathisia, women are believed to be at higher risk. Accurate diagnosis of this condition is necessary in order to prevent inadvertent increases in neuroleptic dose from a belief that the patient's discomfort from akathisia is due instead to worsening psychosis. This effect may be managed by reduction in dosage or switching medications to an atypical antipsychotic drug or a drug that is less likely to cause akathisia. When these strategies are not feasible, the symptoms may respond to anticholinergic medications, usually within 3–7 days. Other options include low doses of benzodiazepines or beta-adrenergic blockers, assuming no contraindications to either.

Parkinsonism caused by antipsychotic drugs resembles idiopathic parkinsonism. Diagnostically, severe neuroleptic induced parkinsonism may resemble depression or negative symptoms of schizophrenia; however, the associated motor signs and time course of symptoms in relation to starting antipsychotic treatment distinguish the former. Like akathisia, the onset and severity of antipsychotic induced parkinsonism is related to medication

dosage; thus, a lowering of the dose or switching to a medication that is less likely to cause this effect may provide significant relief, or ameliorate the parkinsonian signs and symptoms altogether. When this is not feasible, anticholinergic medications may provide relief, typically within 3–7 days. The response to anticholinergic medication is quite variable, however.

Tardive dyskinesia emerges at various rates depending upon age, sex, and diagnosis.[167,168] The rate in younger patients is between 3 and 5 per cent per year. It is higher in bipolar than schizophrenic patients and much higher in people above the age of 60. It is related to dose and will be less likely with lower doses of typical neuroleptics. Tardive dyskinesia is ordinarily reversible, although irreversible and/or extremely severe and rarely life-threatening forms can occur. The best way to minimize its occurrence is to use an atypical antipsychotic drug in lieu of a typical agent, since these drugs as a class are associated with a much lower risk of tardive dyskinesia.[167] Patients with mood disorders should generally not receive maintenance treatment with typical antipsychotic drugs unless mood stabilizers alone prove insufficient because they are at greater risk for tardive dyskinesia. There are no definitive treatments for tardive dyskinesia. Generally, the best strategy is prevention through the use of atypical antipsychotic drugs, and periodic screening with a structured assessment tool such as the Abnormal Involuntary Movement Scale (AIMS). There is some suggestion in the literature that continuation of antipsychotic treatment does not worsen tardive dyskinesia, and may eventually result in a stabilization and improvement of tardive symptoms. Switching to clozapine appears to be helpful, although such an effect is not invariable.

Neuroleptic malignant syndrome is a rare life-threatening side-effect related to an apparent compromise of the neuromuscular and sympathetic nervous systems.[169] It usually occurs at the initiation of treatment with a high-potency agent but may occur with any of the typical (or atypical agents) at any point. Immediate discontinuation of the medication is essential. The condition is characterized by muscle rigidity, breakdown of muscle fibres leading to large increases in plasma creatine kinase activity, fever, autonomic instability, changing levels of consciousness, and sometimes death. It may be treated by discontinuing all antipsychotic drug treatment, applying external hypothermia, supporting blood pressure, and administering a direct-acting dopamine agonist such as bromocriptine or pergolide, and dantrolene sodium, which blocks the release of intracellular stored calcium ions. After its successful treatment, an atypical antipsychotic should be used even though these agents, including clozapine, may also induce neuroleptic malignant syndrome.

The typical neuroleptic drugs produce a wide variety of **other side-effects**, including weight gain, seizures (especially pimozide), sedation, hypotension, elevated liver enzymes, retinitis pigmentosa (thioridazine), orthostatic hypotension, prolongation of the QTc interval (low potency phenothiazines, pimozide) and anticholinergic effects (mesoridazine, chlorpromazine, thioridazine). All the typical neuroleptic drugs produce marked increases in serum prolactin levels, with the increases being greater in females than males.[25] Prolactin elevations may affect sexual function in both males and females, with difficulty achieving erection or orgasm among the most common side-effects.[25]

Atypical antipsychotic drugs

(a) Clozapine

(i) Agranulocytosis

It has now been reliably established that clozapine produces agranulocytosis in slightly less than 1 per 100 patients.[170, 171] This is 15 to 30 times the rate associated with the phenothiazines and possibly higher than that for the butyrophenones. The peak of agranulocytosis with clozapine occurs between 4 and 18 weeks, and then falls off sharply. Weekly monitoring of the white cell or absolute neutrophil count is required for 26 weeks in most countries, with the frequency decreasing to biweekly or monthly thereafter, sometimes on a voluntary basis. In the US, monthly monitoring is required assuming no hematological abnormalities after one year of treatment. The cost-effectiveness of monitoring after a year has not been studied but it is probably in the range that would lead to its abandonment by current standards. With monitoring, agranulocytosis can usually be detected before infection sets in or becomes overwhelming. Discontinuation of clozapine, beginning treatment with colony cell stimulating factors, and the usual procedures for treating an infection are usually effective in restoring the white cell line.

(ii) Other side-effects

Clozapine produces a wide range of side-effects.[171] These can generally be managed by dose adjustment and concomitant medications. Clozapine produces hypotension because of its potent α_1-adrenoceptor antagonism and must be slowly titrated in most patients. Low-dose glucocorticoid treatment may be helpful in some patients with severe hypotension. Clozapine rarely if ever produces significant extrapyramidal side-effects, although some cases of akathisia and neuroleptic malignant syndrome have been reported.

Major motor seizures are another important side-effect of clozapine. They are dose related, with the incidence being about 2 per cent in patients at low doses and 6 per cent at doses greater than 600 mg/day. They are sometimes preceded by myoclonic jerks. Valproic acid and dose reductions are usually effective in preventing the progression of myoclonic jerks or treating major motor seizures. Other anticonvulsants can be combined with clozapine if needed, though caution would be clearly warranted with the use of carbamazepine due to its potential for bone marrow suppression.

Hypersalivation is another side-effect. It usually responds to anticholinergic therapy or to clonidine. Exacerbation of obsessive-compulsive symptoms has been reported with clozapine. Augmentation with an SSRI or lithium carbonate is usually effective.

Weight gain is a frequent side-effect of clozapine, with about 30 per cent of patients gaining more than 7 per cent of body weight.[171] Diet and exercise are useful in minimizing this effect. A related problem is the emergence of insulin resistance or type II diabetes, or exacerbation of existing diabetes, with or without atherogenic changes in serum lipid profile. There have also been reports of diabetic ketoacidosis that emerged in the context of clozapine treatment. Of the atypical antipsychotic drugs, clozapine and olanzapine are associated with the highest risk for clinically significant weight gain, as well as abnormalities in glycaemic control and lipid homeostasis.[123]

Somnolence, tachycardia, hypertension, constipation and stuttering are also produced by clozapine. Tachycardia is treated only when the pulse is greater than 100 beats/minute. β-Blockers are effective to reduce the heart rate, but may also result in synergism of hypotensive effects.[171]

There have been reports of clozapine-associated myocarditis and cardiomyopathy.[172] The presence of eosinophilia accompanied by cardiotoxic signs such as tachycardia, fatigue, orthostasis, or respiratory problems (many of which are adverse effects of clozapine) should alert the clinician to the possibility of myocarditis and the need for medical evaluation.

Finally, treatment with clozapine may not uncommonly result in an asymptomatic mild elevation in hepatic transaminase levels; however, there have also been reports of hepatotoxicity in the setting of clozapine treatment. Polypharmacy appears to be a risk factor. Cases of fulminant hepatotoxicity leading to liver failure are rare.

(iii) Risperidone

Risperidone is associated with moderate weight gain, comparable to that of typical neuroleptic drugs in most cases, and less than that of clozapine and olanzapine.[123,173] Risperidone also produces some postural hypotension because of its α_1-adrenoceptor blocking properties. Risperidone produces greater increases in serum prolactin secretion than any of the other atypical antipsychotic drugs.[25] The increases appear to be at least comparable to those of typical neuroleptics.[173] At higher doses, particularly above 6 mg daily in most adults, the incidence of EPS also increases,[77] though typically not to the degree observed when using typical neuroleptic drugs in clinical practice. Risperidone, like clozapine and other agents of this type, can sometimes exacerbate or induce symptoms of obsessive-compulsive disorder and tics, probably due to its antiserotonergic properties. This can be counteracted in some patients by the addition of an SSRI. Risperidone is not associated with agranulocytosis or increased risk of seizures. Because of its low affinity for muscarinic receptors, risperidone treatment is not associated with significant anticholinergic effects.

(iv) Olanzapine

Olanzapine also produces dose-dependent extrapyramidal side-effects, including some dystonic reactions in patients with schizophrenia, but these are less frequent and severe than those produced by typical neuroleptic drugs or risperidone.[173] Olanzapine is less well tolerated than clozapine in patients with Parkinson's disease. Olanzapine, like other atypical antipsychotic drugs, is associated with a lower risk of tardive dyskinesia than typical neuroleptics.

The major side-effect of olanzapine is weight gain.[173] Large weight gains due to increased appetite occur in 10 to 15 per cent of olanzapine-treated patients during the first 6 months of treatment. Another 20 to 35 per cent gain between 7 and 10 per cent of body weight. These gains tend to become permanent for as long as patients continue the medication. Like clozapine, olanzapine is also associated with higher risk of insulin resistance, glycaemic changes, and development of atherogenic changes in lipid profile.[123] Cases of diabetic ketoacidosis associated with olanzapine treatment have been reported.

Olanzapine is also associated with some increase in liver enzymes, orthostatic hypotension, anticholinergic side-effects, and sedation. Many of these adverse effects are time limited and reduce in intensity or resolve over the first few weeks of treatment with continuous use. Olanzapine produces transient increases in serum prolactin levels, which are smaller in magnitude than those produced by typical neuroleptic drugs or risperidone.[25, 173]

Olanzapine, like other agents of this type, can occasionally exacerbate or induce symptoms of obsessive-compulsive disorder and tics, probably due to its antiserotonergic properties. This can be counteracted in some patients by the addition of an SSRI. Olanzapine is not associated with agranulocytosis or increased risk of seizures.

(v) Quetiapine

Quetiapine appears to have fewer extrapyramidal side-effects than either risperidone or olanzapine.[158,173] Quetiapine is tolerated in patients with Parkinson's disease to a much greater extent than risperidone or olanzapine. The incidence of extrapyramidal side-effects with quetiapine in schizophrenic patients appears to be comparable to placebo. The major side-effects with quetiapine are headache, agitation, dry mouth, dizziness, weight gain, and postural hypotension.[173]

With regard to weight gain and other metabolic effects, quetiapine treatment appears to confer moderate risk—similar to that of risperidone, but less than that associated with clozapine or olanzapine treatment.[123] Far less is known about the long term effects of quetiapine on markers of glycaemic and lipid homeostasis. Nevertheless, clinically significant changes in serum lipids have been reported.

Decreased serum thyroid hormone levels, increased hepatic transaminases and elevated serum lipids have been reported. Decreases in total and free thyroxine, when they occur, are mild, non-progressive, and are not believed to be clinically signficiant. The effect may be dose dependent. Similar to clozapine, asymptomatic elevations in hepatic transaminases may be encountered early in the course of treatment, followed by a return to baseline values. Animal studies suggest an increased risk of cataracts.[173] Periodic ophthalmological screening for lenticular opacities is recommended by the manufacturer, though no causal relationship between the use of quetiapine and the development of cataracts has been demonstrated to date.

(vi) Ziprasidone

Ziprasidone does not increase serum prolactin levels and is virtually devoid of extrapyramidal side-effects, weight gain, and changes in markers of glucose handling and lipid metabolism. Its major side-effects are nasal congestion and somnolence,[173] the latter of which is usually transient. There has been some concern of cardiovascular side-effects, for example increased QTc interval; however, perusal of the available data does not reveal a significant problem in this regard. However, caution is warranted when considering the coadministration of ziprasidone with other drugs that are known to prolong the QTc interval, since ziprasidone has been associated with a significant increase in the QTc interval of 16.6 msec, which was greater than that of other atypical antipsychotics and haloperidol, but less than thioridazine.[174] Screening for electrolyte abnormalities and cardiac disease (including recent myocardial infarction, congestive heart failure symptoms and arrythmias with or without syncope) may be indicated prior to starting ziprasidone.

(vii) Aripiprazole and paliperidone

Aripiprazole is well tolerated, and does not appear to routinely cause EPS or hyperprolactinaemic changes at recommended dosages. This also appears to be the case for higher than recommended doses. Aripiprazole is also not associated with clinically significant increases in weight, or changes in markers of glucose handling or lipid homeostasis. Both aripiprazole and ziprasidone are therefore believed to be the atypical antipsychotic drugs with the most advantageous metabolic risk profile.

Paliperidone in its extended release form also appears to be well tolerated during short term, acute phase treatment. In these studies, the most common side effect was tachycardia. Rates of discontinuation due to adverse effect burden were also very low. The risk of hyperprolactinaemic changes with paliperidone appears to resemble those of risperidone, although no head-to-head comparisons have been carried out. The changes in prolactin levels may be dose related. The EPS burden associated with paliperidone during the short term studies was low for the 6 mg dose; however, at higher doses, the incidence of EPS appears to be higher. Measures of weight and metabolic effects during 6 week treatment with paliperidone showed no significant changes from baseline. Similar results were found for paliperidone during medium-term treatment.[141] Future long term studies will add greatly to our understanding of paliperidone's adverse effect profile.

Indications and contraindications

The main indication for the antipsychotic drugs is the treatment of all phases of schizophrenia, including acute, florid symptoms of psychosis, prevention of relapse, and deficit symptoms. Important other uses include the psychotic phase and prophylaxis of mania, depression with psychotic features, the psychosis, agitation, and aggression of various dementias, the treatment of psychoses due to l-dopa or other dopamine agonists in Parkinson's disease, Tourette's syndrome, treatment-resistant obsessive-compulsive disorder, self-injurious behaviour, porphyria, antiemesis, intractable hiccoughs, and as antipruritics. Some current research has suggested that the antipsychotic drugs may be of use to prevent the onset of schizophrenia by administering them to individuals who are in the prodromal phase of the illness. The atypical antipsychotics may be effective for augmenting antidepressants in patients with treatment-resistant non-psychotic depression, and are being tried on an experimental basis for various character disorders such as borderline, schizoid, and schizotypal personality disorders. Clozapine, which has the lowest incidence of extrapyramidal side-effects of any of the antipsychotic drugs, has some special applications in neurological conditions such as essential tremor and the treatment of the water intoxication syndrome in schizophrenic patients. The uses of the classical antipsychotics such as chlorpromazine and haloperidol have been limited by their side-effects, especially parkinsonism and tardive dyskinesia, a slowly developing, sometimes irreversible series of abnormal involuntary movements involving facial, limb, and girdle muscles. As has been discussed, the atypical antipsychotic drugs such as clozapine, olanzapine, quetiapine, and risperidone, as well as iloperidone and ziprasidone, which are in development, have significant advantages with regard to parkinsonism. Clozapine definitely has a vastly reduced risk of tardive dyskinesia and the other atypical agents most likely have a risk that is less than that of the typical neuroleptic drugs but more than clozapine. Uses in other psychiatric and neurological conditions may be expected to emerge as the safety profile of these agents is better described.

Antiparkinsonian agents

Anticholinergic drugs

Antiparkinsonian medications, including anticholinergic, antihistaminic, benzodiazepines, dopamine agonists, and β-blockers are of importance in the management of extrapyramidal side-effects. They are usually needed with the typical neuroleptic drugs but some patients will require antiparkinsonian treatment with olanzapine, risperidone, or quetiapine. The anticholinergics and the antihistaminics (e.g. diphenhydramine) are used to treat acute dyskinesias and dystonias, pseduoparkinsonian symptoms (tremor, rigidity, bradykinesia, shuffling gait), and akathisia. These agents act centrally in the basal ganglia to block the effects of increased acetylcholine release due to D_2-receptor blockade. The most widely used anticholinergic drugs are benztropine, biperiden, procyclidine, and trihexyphenidyl. Benztropine is given in doses of 1 to 6mg/day usually in divided doses. Biperiden is given in doses of 2 to 16 mg/day in two or three doses. Procyclidine is given in divided doses of 5 to 30 mg/day. Trihexylphenidyl is given in doses of 1 to 15 mg/day, in a single or divided dose.

These agents are competitive antagonists of the five subtypes of muscarinic receptors that have been identified and which are labelled M_1 to M_5. They have minimal antagonist effect at nicotinic cholinergic receptors. Blockade of cholinergic receptors on intrastriatal neurones by these agents restores the cholinergic balance, which is disrupted by blockade of D_2 dopamine receptors by some antipsychotic agents. Other central effects include impairment of various forms of memory. Elderly patients in particular may develop anticholinergic-induced agitation, irritability, disorientation, hallucinations and delirium because of the natural loss of cholinergic neurones with aging.

Side-effects

These agents have some preference for the central nervous system but some peripheral anticholinergic effects are to be expected. Blockade of vagal tone in the heart produces tachycardia. Other adverse effects include decreased bladder function and urinary retention and decreased bowel motility leading to constipation and impaction. Decreased saliva and bronchial secretion contribute to dry mouth and increased dental caries while decreased sweating increases the risk of heat stroke. Blockade of muscarinic receptors in the eye cause pupillary dilation and inhibition of accommodation, leading to photophobia and blurred vision. Rarely, narrow-angle glaucoma may ensue. The muscarinic receptors in the basal ganglia are predominantly M_2 whereas those in the periphery are M_1. The rank order of the anticholinergic drugs for relative selectivity for the M_2 receptor is biperiden, procyclidine, trihexylphenidyl, and benztropine. All these agents can cause dry mouth, blurred vision, urinary retention, constipation, and increased intraocular pressure. They may cause anticholinergic delirium in elderly patients or after taking high doses. Biperiden is less likely to cause peripheral anticholinergic effects. Benztropine, biperiden, and trihexyphenidyl may cause euphoria because of their ability to inhibit dopamine reuptake and may be subject to abuse.

Indications

The anticholinergic drugs or the antihistamine diphenhydramine are given intramuscularly for the treatment of acute dystonic reactions. They are usually effective within minutes and may have to be repeated. It is usually not necessary to prescibe an oral anticholinergic following a dystonic reaction, though some may require their brief use depending on which anipsychotic is prescribed. These agenst should not be given prophylactically unless the patient is at established risk for EPS at the dose of antipsychotic which is being started. If akathisia or parkinsonism develops following treatment with a typical neuroleptic drug, the first consideration should be whether to continue to use the offending agent and drop the dosage or to substitute an atypical antipsychotic drug. If decreasing the dose of antipsychotic drugs does not suffice or is not clinically feasible, substituting an atypical agent is the clearly the recommended choice since it avoids all the unpleasant side-effects of the anticholinergic agents.

Other drugs

Amantadine, which also has antiviral actions, is able to increase the release of dopamine in the basal ganglia, which diminishes the release of acetylcholine. It may improve acute dystonias, akathisia, akinesia, parkinsonism, and tardive dyskinesia. It has also been reported to improve sexual function and decrease weight gain due to neuroleptic drugs. It may cause increased arousal, agitation, and indigestion, however. The usual oral dose is 100 to 400 mg/day.

β-Blockers such as propranolol, atenolol, and pindolol are useful for treating akathisia and tremor. They may cause bradycardia, and particularly immediate-release forms should not be stopped abruptly due to rebound tachycardia.

Benzodiazepines, such as clonazepam, lorazepam, and diazepam, are useful for treating akathisia, acute dystonias, and acute dyskinesias. They can cause drowsiness and lethargy, and have abuse potential.

Conclusions

Antipsychotic drugs are invaluable tools in treating a large variety of patients with schizophrenia and other conditions. Their main benefits are, in fact, to treat psychotic symptoms, but the newer agents in particular may improve negative symptoms, cognition, mood, anxiety, and aggression as well. The evidence for atypical antipsychotic drugs to improve cognition is steadily increasing and this should be one of the driving forces behind the substitution of these agents for the typical antipsychotic drugs. Recent evidence of volumetric increases in cerebral cortical gray matter associated with atypical, but not typical, antipsychotic drugs may be related to improvement in such symptoms.[175] As such, atypical antipsychotic drugs may produce 'disease modifying' rather than just 'symptomatic' effects, a matter that is of considerable current interest.

Antipsychotic drugs are useful as both acute and maintenance treatments to prevent the recurrence of psychotic symptoms. The extrapyramidal side-effects and greater tardive dyskinesia risk of the typical antipsychotics, coupled with their lesser efficacy to improve negative symptoms and cognition suggest that newer agents are preferred. Clozapine, despite its risk of agranulocytosis, is the treatment of choice for patients who fail to respond to other typical or atypical antipsychotic agents. Risperidone, olanzapine, quetiapine, ziprasidone, aripiprazole and paliperidone have somewhat different pharmacologic profiles. It is not clear which of these agents should be tried in a given patient but on going research may clarify that. These agents, and clozapine, appear to differ significantly in their propensity for causing clinically significant changes in weight and markers of metabolic status. Amisulpride has a mechanism of action different from that of the other atypical agents, with some preference for treating negative symptoms. These compounds, as well as others expected to be approved for use in the near future, for example iloperidone and asenapine, will need to be compared with each other to determine if differential indications exist. Side-effect differences among these drugs as well as the availability of long-acting preparations may help clinicians choose among them. Cost-effective analyses currently favour use of the atypical antipsychotic drugs because of better compliance leading to less frequent relapses and shorter hospital stays. They also facilitate retention of work skills and return to work which decreases the indirect costs of illness in patients still young enough to be able to work. As long as the typical antipsychotics remain in use, and for some patients who receive atypical agents, anticholinergic and other antiparkinsonian drugs will continue to be necessary to treat extrapyramidal side-effects.

Because of the compliance problem, which is less with the atypical than the typical antipsychotics, it is important to develop more long acting atypical drugs. Risperidone is currently available in such a form, and paliperidone and olanzapine will also be in the near future. While the current group of atypical antipsychotic drugs is predominantly characterized by relatively more potent 5-HT_{2A} than D_2 receptor antagonism, it is likely that a number of different strategies will emerge for compounds which produce fewer extrapyramidal side-effects than the typical neuroleptics. Because these compounds are so effective in that regard, the real challenge is to develop agents which address other key features of schizophrenia, especially cognitive impairment and negative symptoms, without the side-effect burden of this group of compounds.

Further information

Davis, K.L., Charney, D., Coyle, J., *et al.* (eds.) (2001). *Neuropsychopharmacology: A Fifth Generation of Progress*. New York: Raven Press.

Breier, A., Tran, P.V., Herrera, J.M., *et al.* (2001). *Current Issues in the Pharmacology of Schizophrenia*. Philadelphia: Lippincott Williams & Wilkins.

References

1. Meltzer, H. (1995). The concept of atypical antipsychotics. In: *Advances in the neurobiology of schizophrenia* (eds. J.A. den Boer, H.G.M. Westenberg, and H.M. van Praag), pp. 265–73. Wiley, Chichester.
2. Arndt, J. and Skarsfeldt, T. (1998). Do novel antipsychotics have similar pharmacological characteristics? A review of the evidence. *Neuropsychopharmacology*, **181**, 63–101.
3. Kapur, S. and Seeman, P. (2001). Does fast dissociation from the dopamine D(2) receptor explain the action of atypical antipsychotics? A new hypothesis. *American Journal of Psychiatry*, **158**, 360–9.
4. Dixon, L., Lehman, A., and Levine, J. (1995). Conventional antipsychotic medications for schizophrenia. *Schizophrenia Bulletin*, **21**, 567–78.

5. Fitton, A. and Heel, R. (1990). Clozapine: a review of its pharmacological properties, and therapuetic use in schizophrenia. *Drugs*, **40**, 722–47.
6. Burris, K.D., Molski, T.F., Xu, C., *et al.* (2002). Aripiprazole, a novel antipsychotic, is a highaffi nity partial agonist at human dopamine D2 receptors. *Journal of Pharmacology & Experimental Therapeutics*, **302**(1), 381–9.
7. Szewezak, M., Corbett, R., Rusk, D., *et al.* (1995). The pharmacological profi le of iloperidone, a novel atypical antipsychotic agent. *Journal of Pharmacology and Experimental Therapeutics*, **274**, 1404–13.
8. Alphs, L., Panagides, J., Lancaster, S. (2007). Asenapine in the treatment of negative symptoms of schizophrenia: clinical trial design and rationale. *Psychopharmacology Bulletin*, **40**, 41–53.
9. Schotte, A., Janssen, P.F., Gommeren, W., *et al.* (1996). Risperidone compared with new and reference antipsychotic drugs: in vitro and in vivo receptor binding. *Psychopharmacology*, **124**(1–2), 57–73
10. Meltzer, H. and Fatemi, S. (1996). The role of serotonin in schizophrenia and the mechanism of action of anti–psychotic drugs. In *Serotonergic mechanisms in antipsychotic treatment* (eds. J.M. Kane., H.–J. Moller, and F. Awouters), pp. 77–107. Dekker, New York.
11. Anonymous (1998). Adverse effects of the atypical antipsychotics. Collaborative Working Group on Clinical Trial Evaluations. *Journal of Clinical Psychiatry*, **59**, 17–22.
12. Luft, B., Taylor, D. (2006). A review of atypical antipsychotic drugs versus conventional medication in schizophrenia. *Expert Opinion on Pharmacotherapy*. **7**(13), 1739–48.
13. The Collaborative Working Group (1998). Atypical antipsychotics for treatment of depression in schizophrenia and affective disorders. *Journal of Clinical Psychiatry*, **12**, 41–6.
14. McElroy, S.L., Fry, M., Denicoff, K., *et al.* (1998). Olanzapine in treatment–resistant bipolar disorder. *Affective Disorders*, **49**, 119–22.
15. Calabrese, J., Kimmel, S., Woyshville, M., *et al.* (1996). Clozapine in treatment–refractory mania. *American Journal of Psychiatry*, **153**, 759–64.
16. Keefe, R.S., Silva, S.G., Perkins, D.O., *et al.* (1999). The effects of atypical antipsychotic drugs on neurocognitive impairment in schizophrenia: a review and meta-analysis *Schizohrenia Bulletin*, **25**, 201–22.
17. Woodward, N.D., Purdon, S.E., Meltzer, H.Y., *et al.* (2005). Meta-analysis of neuropsychological change to clozapine, olanzapine, quetiapine, and risperidone in schizophrenia. *International Journal of Neuropsychopharmacology*, **8**, 457–72.
18. Keefe, R., Silva, S., Perkins, D., *et al.* (1999). The effects of atypical antipsychotic drugs on neurocognitive impairment in schizohprenia. *Schizophrenia Bulletin*, **25**, 201–32.
19. Meltzer HY. and Stahl, S.M (1976). The dopamine hypothesis of schizophrenia: A review. *Schizophrenia Bulletin* **2**, 19–76.
20. Davis, K., Kahn, R., Ko, G., *et al.* (1992). Dopamine in schizophrenia: a review and reconceptualization. *American Journal of Psychiatry*, **148**, 1474–86.
21. Meltzer, H. (1999). The role of serotonin in antipsychotic drug action. *Neuropsychopharmacology*, **21**, 106–15.
22. Goldman–Rakic, P. and Selemon, L. (1997). Functional and anatomical aspects of prefrontal pathology in schizophrenia. *Schizophrenia Bulletin*, **23**, 437–58.
23. Abi-Dargham, A., Moore, H. (2003). Prefrontal DA transmission at D1 receptors and the pathology of schizophrenia. *Neuroscientist*, **9**, 404–16.
24. Farde, L., Nordstrom, A., Wiesel, F., *et al.* (1992). Positron emission tomographic analysis of central D1 and D2 dopamine receptor occupancy in patients treated with classical neuroleptics and clozapine. Relation to extrapyramidal side effects. *Archives of General Psychiatry*, **49**, 538–44.
25. Haddad, P.M., Wieck, A. (2004). Antipsychotic-induced hyperprolactinemia: mechanisms, clinical features and management. *Drugs*, **64**, 2291–314.
26. Kramer, M., Last, B., *et al.*, and the D4 Dopamine Antagonist Group (1997). The effects of a selective D4 dopamine receptor antagonism. (L–745, 870) in acutely psychotic inpatients with schizophrenia. *Archives of General Psychiatry*, **54**, 567–72.
27. Richelson, E. (1988). Neuroleptic binding to human brain receptors: relation to clinical effects. *Annals of the New York Academy of Sciences*, **537**, 435–42.
28. Meltzer, H.Y., Matsubara, S., and Lee, M. (1989). Classification of typical and atypical antipsychotic drugs on the basis of dopamine D-1, D-2 and serotonin 2 pKi values. *Journal of Pharmacology and Experimental Therapeutics*, **251**, 238–46.
29. Kuroki, T., Meltzer, H., and Ichikawa, J. (1998). Effects of antipsychotic drugs on extracellular dopamine levels in rat medial prefrontal cortex and nucleus accumbens. *Journal of Pharmacology and Experimental Therapeutics*, **288**, 774–81.
30. Parada, M., Hernande, L., Puig de Parada, M., *et al.* (1997). Selection action of acute systemic clozapine on acetylcholine release in the rat prefrontal cortex by reference to the nucleus accumbens and striatum. *Journal of Pharmacology and Experimental Therapeutics*, **281**, 582–8.
31. Li, X., Perry, K., Wong, D., *et al.* (1998). Olanzapine increases *in vivo* dopamine and norepinephrine release in rat prefrontal cortex, nucleus accumbens and striatum. *Psychopharmacology*, **136**, 153–61.
32. Andersson, J.L., Nomikos, G.G., Marcus, M., *et al.* (1995). Ritanserin potentiates the stimulatory effects of raclopride on neuronal activity and dopamine release selectivity in the mesolimbic dopaminergic system. *Naunyn- Schmiedeberg's Archives of Pharmacology*, **352**, 374–85.
33. Westerink, B.H., Kawahara, Y., De Boer, P., *et al.* (2001). Antipsychotic drugs classified by their effects on the release of dopamine and noradrenaline in the prefrontal cortex and striatum. *European Journal of Pharmacology*, **412**, 127–38.
34. Liegois, J.F., Ichikawa, J., Meltzer, H.Y. (2002). 5HT2A receptor antagonism potentiates haloperidol-induced dopamine release in rat medial prefrontal cortex and inhibits that in the nucleus accumbens in a dose-dependent manner. *Brain Research*, **947**, 1547–65.
35. Armsten, A.F. (2004). Adrenergic targets for the treatment of cognitive deficits in schizophrenia. *Psychopharmacology (Berl)*, **174**, 25–31.
36. Poyurovsky, M., Koren, D., Gonopolsky, I., *et al.* (2003). Effects of the 5-HT2 antagonist mianserin on cognitive dysfunction in chronic schizophrenia patients: an add-on, double-blind, placebo-controlled study. *European Neuropsychopharmacology*, **13**, 123–8.
37. Araneda, R. and Andrade, R. (1991). 5-Hydroxytryptamine2 and 5-hydroxitryptamine1A receptors mediate opposing responses on membrane excitability in rat association cortex. *Neuroscience*, **40**, 399–412.
38. Ichikawa, J., Meltzer, H.Y. (1999). R(+)-8-OH-DPAT, a serotonin1A receptor agonist, potentiated S(-) sulpiride-induced dopamine release in rat medial prefrontal cortex and nucleus accumbens but not striatum. *Journal of Pharmacology and Experimental Therapeutics*, **291**, 1227–32.
39. Sakaue, M., Somboonthum, P., Nishihara, B., *et al.* (2000). Postsynaptic 5-hydroxytryptamine(1A) receptor activation increases in vivo dopamine release in rat prefrontal cortex. *British Journal of Pharmacology*, **129**, 1028–34.
40. Sumiyoshi, T., Matsui, M., Yamashita, I., *et al.* (2001). Enhancement of cognitive performance in schizophrenia by addition of tandospirone to neuroleptic treatment. *American Journal of Psychiatry*, **158**, 1722–5.
41. Sumiyoshi, T., Jayathilake, K., Roy, A., *et al.* (2006). Effect of buspirone, a serotonin(1A) partial agonist, on cognitive function in schizophrenia. [abstract]. *International Journal of Neuropsychopharmacology*, **9** (suppl 1), S248.

42. Bonaccorso, S., Meltzer, H.Y., Li, Z., *et al.* (2002). SR46349-B, a 5-HT(2A/2C) receptor antagonist, potentiates haloperidol-induced dopamine release in rat medial prefrontal cortex and nucleus accumbens. *Neuropsychopharmacology*, **27**, 430–41.
43. Kroeze, W.K., Hufeisen, S.J., Popadak, B.A., *et al.* (2003). H1-histamine receptor affinity predicts short-term weight gain for typical and atypical antipsychotic drugs. *Neuropsychopharmacology*, **28**, 519–26.
44. Reynolds, G.P., Templeman, L.A., Zhang, Z.J. (2005). The role of 5-HT2C receptor polymorphisms in the pharmacogenetics of antipsychotic drug treatment. *Progress in Neuropsychopharmacology and Biological Psychiatry*, **29**, 1021–8.
45. Bigliani, V., Mulligan, R.S., Acton, P.D., *et al.* (2000). Striatal and temporal cortical D2/D3 receptor occupancy by olanzapine and sertindole in vivo: a [123I] epidepride single photon emission tomography (SPET) study. *Psychopharmacologia*, **150**, 132–40.
46. Stephenson, C.M., Bigliani, V., Jones, H.M., *et al.* (2000). Striatal and extra-striatal D(2)/D(3) dopamine receptor occupancy by quetiapine in vivo. *British Journal of Psychiatry*, **177**, 408–15.
47. Xiberas, X., Martinot, J.L., Mallet, L., *et al.* (2001). Extrastriatal and striatal D(2) dopamine receptor blockade with haloperidol or new antipsychotic drugs in patients with schizophrenia. *British Journal of Psychiatry*, **179**, 503–8.
48. Bressan, R.A., Erlandsson, K., Stone, J.M., *et al.* (2005). Impact of schizophrenia and chronic antipsychotic treatment on [123I] CNS-1261 binding to N-methyl-Daspartate receptors in vivo. *Biological Psychiatry*, **58**, 41–6.
49. Kane, J. and Marder, S. (1993). Psychopharmacologic treatment of schizophrenia. *Schizophrenia Bulletin*, **19**, 287–302.
50. Lieberman, J., Jody, D., Geisler, S., *et al.* (1993). Time course and biologic correlates of treatment response in fi rst–episode schizophrenia. *Archives of General Psychiatry*, **50**, 369–76.
51. Meltzer, H., Lee, M., and Colal, P. (1998). The evolution of treatment resistance. Biological implications. *Journal of Clinical Psychopharmacology*, **18**, 5–11.
52. Meltzer, H. (1997). Treatment-resistant schizophrenia: the role of clozapine. *Current Medical Research Opinion*, **14**, 1–20.
53. Tandon, R., Ribeiro, S., DeQuardo, J., *et al.* (1993). Covariance of positive and negative symptoms during neuroleptic treatment in schizophrenia: a replication. *Biological Psychiatry*, **34**, 495–7.
54. The Collaborative Working Group (1998). Assessing the effects of a typical antipsychotics on negative symptoms. *Journal of Clinical Psychiatry*, **12**, 28–35.
55. Saykin, A., Shtasel, D., Gur, R., *et al.* (1994). Neuropsychological deficits in neuroleptic naïve patients with first–episode schizophrenia. *Archives of General Psychiatry*, **51**, 124–31.
56. Palmer, B., Heaton, R., Paulsen, J., *et al.* (1997). Is it possible to be schizophrenic yet neuropsychologically normal? *Neuropsychology*, **11**, 437–46.
57. Green, M. (1996). What are the functional consequences of neurocognitive deficits in schizophrenia? *American Journal of Psychiatry*, **153**, 321–30.
58. Meltzer, H.Y., and McGurk, S. (1999). The effect of clozapine, risperidone and olanzapine on cognitive function in schizophrenia. *Schizophrenia Bulletin*, **25**, 233–56.
59. Sheitman, B., Lee, H., Strauss, R., *et al.* (1997). The evaluation and treatment of first–episode psychosis. *Schizophrenia Bulletin*, **23**, 653–61.
60. McEvoy, J., Hogarty, G., and Steingard, S. (1991). Optimal dose of neuroleptic in acute schizophrenia. A controlled study of the neuroleptic threshold and higher haloperidol dose. *Archives of General Psychiatry*, **48**, 739–45.
61. Carpenter, W., Buchanan, R., Kirkpatrick, B., *et al.* (1999). Diazepam treatment of early signs of exacerbation in schizophrenia. *American Journal of Psychiatry*, **156**, 299–303.
62. Schooler, N.R. (2003). Relapse and rehospitalization: comparing oral and depot antipsychotics. *Journal of Clinical Psychiatry*, **64** (suppl 16), 14–7.
63. Lehman, A.F., Kreyenbuhl, J., Buchanan, R.W., *et al.* (2004). The Schizophrenia Patient Outcomes Research Team (PORT): updated treatment recommendations 2003. *Schizophrenia Bulletin*, **30**, 193–217.
64. Lehman, A.F., *et al*. American Psychiatric Association; Steering Committee on Practice Guidelines. (2004). Practice guideline for the treatment of patients with schizophrenia, second edition. *American Journal of Psychiatry*, **161** (2 suppl), 1–56.
65. Lieberman, J.A., Stroup, T.S., *et al.* Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Investigators. (2005). Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New England Journal of Medicine*, **353**, 1209–23.
66. Jones, P.B., Barnes, T.R., Davies, L., *et al*. (2006). Randomized controlled trial of the effect on Quality of Life of second- vs first generation antipsychotic drugs in schizophrenia: Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1). *Archives of General Psychiatry*, **63**, 1079–87.
67. Revicki, D. (1999). Pharmacoeconomic studies of atypical antipsychotic drugs for the treatment of schizophrenia. *Schizophrenia Research*, **35**, 101–9.
68. Bobes, J., Canas, F., Rejas, J., *et al.* (2004). Economic consequences of the adverse reactions related with antipsychotics: an economic model comparing tolerability of ziprasidone, olanzapine, risperidone, and haloperidol in Spain. *Progress in Neuropsychopharmacology and Biological Psychiatry*, **28**, 1287–97.
69. Meltzer, H. (1979). The clozapine story. In: *The handbook of psychopharmacology trials* (eds. M. Hertzman and D. Feltner), pp. 137–56. New York University Press.
70. Baldessarini, R. and Frankenberg, F. (1991). Clozapine: a novel antipsychotic agent. *New England Journal of Medicine*, **324**, 746–54.
71. Kane, J., Honigfeld, G., *et al.* Clozaril Collaborative Study Group (1988). Clozapine for the treatmentresistant schizophrenic: a double–blind comparison with chlorpromazine. *Archives of General Psychiatry*, **45**, 789–96.
72. Meltzer, H.Y., Alphs, L., *et al.* International Suicide Prevention Trial Study Group. (2003). Clozapine treatment for suicidality in schizophrenia: International Suicide Prevention Trial (InterSePT). *Archives of General Psychiatry*, **60**, 82–91.
73. Nordstrom, A., Farde, L., Nyber, S., *et al.* (1995). D1, D2, and 5-HT2 receptor occupancy in relation to clozapine serum concentration: a PET study of schizophrenic patients. *American Journal of Psychiatry*, **152**, 1444–9.
74. Shiloh, R., Zemishlany, Z., Aizenberg, D., *et al.* (1997). Sulpiride augmentation in people with schizophrenia partially responsive to clozapine. A double-blind, placebo-controlled study. *British Journal of Psychiatry*, **171**, 569–73.
75. Havaki-Kontaxaki, B.J., Ferentinos, P.P., Kontaxakis, V.P., *et al.* (2006). Concurrent administration of clozapine and electroconvulsive therapy in clozapine-resistant schizophrenia. *Clinical Neuropharmacology*, **29**, 52–6.
76. Meltzer, H., Lee, M., Ranjan, R., *et al.* (1996). Relapse following clozapine withdrawal: effect of cyproheptadine plus neuroleptic. *Psychopharmacology*, **124**, 176–87.
77. Marder, S. and Meibach, R. (1994). Risperidone in the treatment of schizophrenia. *American Journal of Psychiatry*, **151**, 825–35.
78. Pajonk, F.G. (2004). Risperidone in acute and long-term therapy of schizophrenia—a clinical profile. *Progress in Neuropsychopharmacology and Biological Psychiatry*, **28**, 15–23.
79. Leucht, S., Pitschel–Walz, G., Abraham, D., *et al.* (1999). Efficacy and extrapyramidal side–effects of the new antispychotics olanzapine, quetiapine, risperidone, and sertindole compared to conventional antipsychotics and placebo. A meta–analysis of randomized controlled trials. *Schizophrenia Research*, **35**, 51–68.

80. Citrome, L., Bilder, R.M., Volavka, J. (2002). Managing treatmentresistant schizophrenia: evidence from randomized controlled trials. *Journal of Psychiatric Practice*, **8**, 205–15.
81. Kumar, V. and Brecher, M. (1999). Psychopharmacology of atypical antipsychotics and clinical outcomes in elderly patients. *Journal of Clinical Psychiatry*, **60**, 10–16.
82. Simpson, G. and Lindenmayer, J. (1997). Extrapyramidal symptoms in patients treated with risperidone. *Journal of Clinical Psychopharmacology*, **17**, 194–201.
83. Csernansky, J.G., Mahmoud, R., Brenner, R. (2002). A comparison of risperidone and haloperidol for the prevention of relapse in patients with schizophrenia. *New England Journal of Medicine*, **346**, 16–22.
84. Rabinowitz, J., Lichtenberg, P., Kaplan, Z., *et al.* (2001). Rehospitalization rates of chronically ill schizohprenic patients discharged on a regimen of risperidone, olanzapine, or conventional antipsychotics. *American Journal of Psychiatry*, **158**, 266–9.
85. Ehret, M.J., Fuller, M.A. (2004). Long-acting injectable risperidone. *Annals of Pharmacotherapy*, **38**, 2122–7.
86. Jeste, D., Lacro, J., Bailey, A., *et al*. (1999). Lower incidence of tardive dyskinesia with risperidone compared with haloperidol in older patients. *Journal of the American Geriatrics Society*, **47**, 716–19.
87. Moller, H., Muller, H., Borison, R., *et al*. (1995). A path-analytical approach to differentiate between direct and indirect drug effects on negative symptoms in schizophrenic patients (a re–evaluation of the North American risperidone study). *European Archives of Psychiatry Clinical Neuroscience*, **245**, 45–9.
88. Carman, J., Peuskens, J., Vangeneugden, A. (1995). Risperidone in the treatment of negative symptoms of schizophrenia: a meta-analysis. *International Clinical Psychopharmacology*, **10**, 207–13.
89. Glennie, J. (1997). Technology overview: pharmaceuticals: pharmacoeconomic evaluations of clozapine in treatment-resistant schizophrenia and risperidone in chronic schizophrenia. Ottawa (ON): Canadian Coordinating Office for Health Technology Assessment (CCOHTA).
90. Laux, G., Heeg, B., van Hout, B.A., *et al*. (2005). Costs and effects of long-acting risperidone compared with oral atypical and conventional depot formulations in Germany. *Pharmacoeconomics*, **23** (suppl 1), 49–61.
91. Awad, A.G. and Voruganti, L.N.P. (2004). Impact of atypical antipsychotics on quality of life in patients with schizophrenia. *CNS Drugs*, **18**, 877–93.
92. Tollefson, G. and Kuntz, A. (1999). Review of recent clinical studies with olanzapine. *British Journal of Psychiatry*, **37**, 30–35.
93. Conley, R., Kelly, D., and Gale, E. (1998). Olanzapine response in treatment-refractory schizophrenic patients with a history of substance abuse. *Schizophrenia Research*, **33**, 95–101.
94. Tollefson, G.D., Birkett, M.A., *et al*., Lilly Resistant Schizophrenia Study Group. (2001). Double-blind comparison of olanzapine versus clozapine in schizohprenic patients clinically eligible for treatment with clozapine. *Biological Psychiatry*, **49**, 52–63.
95. Dinakar, H.S., Sobel, R.N., Bopp, J.H., *et al*. (2002). Efficacy of olanzapine and risperidone for treatment-resistant schizophrenia among long-stay state hospital patients. *Psychiatric Services*, **53**, 755–7.
96. Beasley, C., Tollefson, G., Tran, P., *et al*. (1996). Olanzapine versus placebo and haloperidol acute phase results of the North American double-blind olanzapine trial. *Neuropsychopharmacology*, **14**, 111–23.
97. Tollefson, G. and Sanger, T. (1997). Negative symptoms, a path analytic approach to a double-blind, placebo- and haloperidol-controlled clinical trial with olanzapine. *American Journal of Psychiatry*, **54**, 466–74.
98. Tran, P., Dellva, M., Tollefson, G., *et al.* (1998). Oral olanzapine versus oral haloperidol in the maintenance treatment of schizophrenia and related psychoses. *British Journal of Psychiatry*, **172**, 499–505.
99. Tollefson, G., Sanger, T., Lu, Y., *et al*. (1998). Depressive signs and symptoms in schizophrenia: a prospective blinded trial of olanzapine and haloperidol. *Archives of General Psychiatry*, **55**, 250–8.
100. Almond, S., O'Donnell, O. (1998). Cost analysis of the treatment of schizophrenia in the UK: a comparison of olanzapine and haloperidol. *Pharmacogenomics*, **13**, 575–88.
101. Hamilton, S.H., Revicki, D.A., Edgell, E.T., *et al.* (1999). Clinical and economic outcomes of olanzapine compared with haloperidol for schizophrenia: results from a randomized clinical trial. *Pharmacoeconomics*, **15**, 469–80.
102. Kinon, B.J., Ahl, J., Stauffer, V.L., *et al*. (2004). Dose response and atypical antipsychotics in schizophrenia. *CNS Drugs*, **18**, 597–616.
103. Wagstaff, A.J., Easton, J. and Scott, L.J. (2005). Intramuscular olanzapine: a review of its use in the management of acute agitation. *CNS Drugs*, **19**, 147–64.
104. Arvanitis, L. and Miller, B. (1997). Multiple fixed dose of 'Seroquel' (quetiapine) in patients with acute exacerbation of schizophrenia: a comparison with haloperidol and placebo. The Seroquel Trial 13 Study Group. *Biological Psychiatry*, **42**, 233–46.
105. Small, J., Hirsch, S., *et al*. and the Seroquel Study Group (1997). Quetiapine in patients with schizophrenia: a high- and low-dose double-blind comparison with placebo. *Archives of General Psychiatry*, **54**, 549–57.
106. Gefvert, O., Lundberg, T., Wieselgren, I.M., *et al.* (2001). D(2) and 5HT(2A) receptor occupancy of different doses of quetiapine in schizophrenia: a PET study. *European Neuropsychopharmacology*, **11**, 105–10.
107. Emsley, R.A., Buckley, P., Jones, A.M., *et al.* (2003). Differential effect of quetiapine on depressive symptoms in patients with partially responsive schizophrenia. *Journal of Psychopharmacology*, **17**, 210–15.
108. Velligan, D.I., Newcomer, J., Pultz, J., *et al.* (2002). Does cognitive function improve with quetiapine in comparison to haloperidol? *Schizophrenia Research*, **53**, 239–48.
109. Fernandez, H.H., Trieschmann, M.E., Friedman, J.H. (2003). Treatment of psychosis in Parkinson's disease: safety considerations. *Drug Safety*, **26**, 643–59.
110. Stahl, S.M. and Shayegan, D.K. (2003). The psychopharmacology of ziprasidone: receptor-binding properties and real-world psychiatric practice. *Journal of Clinical Psychiatry*, **64** (Suppl 19), 6–12.
111. Davis, R. and Markham, A. (1997). Ziprasidone. *CNS Drugs*, **8**, 153–9.
112. Tandon, R., Harrigan, E., and Zorn, S. (1997). Ziprasidone: a novel antipsychotic with unique pharmacology and therapeutic potential. *Journal of Serotonin Research*, **4**, 159–77.
113. Swainston, H.T. and Scott, L.J. (2006). Ziprasidone: a review of its use in schizophrenia and schizoaffective disorder. *CNS Drugs*, **20**, 1027–52.
114. Arato, M., O'Connor, R., and Meltzer, H. The Ziprasidone Extended Use in Schizophrenia (Zeus) study: a prospective, double-blind, placebo-controlled, 1-year clinical trial. Submitted for publication.
115. Hirsch, S.R., Kissling, W., Bauml, J., *et al*. (2002). A 28-week comparison of ziprasidone and haloperidol in outpatients with stable schizophrenia. *Journal of Clinical Psychiatry*, **63**, 516–23.
116. Daniel, D.G., Zimbroff, D.L., Potkin, S.G., *et al*. (1999). Ziprasidone 80 mg/day and 160 mg/day in the acute exacerbation of schizophrenia and schizoaffective disorder: a 6-week placebo-controlled trial. Ziprasidone Study Group. *Neuropsychopharmacology*, **20**, 491–505.
117. Harvey, P.D. (2003). Ziprasidone and cognition: the evolving story. *Journal of Clinical Psychiatry*, **64** (suppl 19), 33–9.
118. Phillips, G.A., Van Brunt, D.L., Roychowdhury, S.M., *et al.* (2006). The relationship between quality of life and clinical efficacy from a randomized trial comparing olanzapine and ziprasidone. *Journal of Clinical Psychiatry*, **67**, 1397–403.
119. Bernardo, M., Ramon Azanza, J., Rubio-Terres, C., *et al.* (2006). Cost-effectiveness analysis of schizophrenia relapse prevention: an economic evaluation of the ZEUS (Ziprasidone-Extended-Use-In-Schizophrenia) study in Spain. *Clinical Drug Investigation*, **26**, 447–57.

120. Mamo, D., Kapur, S., Shammi, C.M., *et al.* (2004). A PET study of dopamine D2 and serotonin 5-HT1 receptor occupancy in patients with schizophrenia treated with therapeutic doses of ziprasidone. *American Journal of Psychiatry*, **161**, 818–25.
121. Hamelin, B.A., Allard, S., Laplante, L., *et al.* The effect of timing of a standard meal on the pharmacokinetics and pharmacodynamics of the novel atypical antipsychotic agent ziprasidone. *Pharmacotherapy*, **18**, 9–15.
122. Daniel, D.G., Potkin, S.G., Reeves, K.R., *et al.* (2001). Intramuscular (IM) ziprasidone 20 mg is effective in reducing acute agitation associated with psychosis: a double-bline, randomized trial. *Psychopharmacology*, **155**, 128–34.
123. Newcomer, J.W. and Haupt, D.W. (2006). The metabolic effects of antipsychotic medications. *Canadian Journal of Psychiatry*, **51**, 480–91.
124. Haddad, P.M. and Anderson, I.M. (2002). Antipsychotic-related QTc prolongation, torsade de pointes and sudden death. *Drugs*, **62**, 1649–71.
125. Heinrich, T.W., Biblo, L.A. and Schneider, J. (2006). Torsades de pointes associated with ziprasidone. *Psychosomatics*, **47**, 264–8.
126. Harrigan, E.P., Miceli, J.J., Anziano, R., *et al.* (2004). A randomized evaluation of the effects of six antipsychotic agents on QTc, in the absence and presence of metabolic inhibition. *Journal of Clinical Psychopharmacology*, **24**, 62–9.
127. Burris, K.D., Molski, T.F., Xu, C., *et al.* (2002). Aripiprazole, a novel antipsychotic, is a high-affinity partial agonist at human dopamine D2 receptors. *Journal of Pharmacology and Experimental Therapeutics*, **302**, 381–9.
128. Jordan, S., Koprivica, V., Chen, R., *et al.* (2002). The antipsychotic aripiprazole is a potent, partial agonist at the human 5-HT1A receptor. *European Journal of Pharmacology*, **441**, 137–40.
129. El-Sayeh, H.G., Morganti, C. and Adams, C.E. (2006). Aripiprazole for schizophrenia. Systematic review. *British Journal of Psychiatry*, **189**, 102–8.
130. Potkin, S.G., Saha, A.R., Kujawa, M.J., *et al.* (2003). Aripiprazole, an antipsychotic with a novel mechanism of action, and risperidone vs placebo in patients with schizophrenia and schizoaffective disorder. *Archives of General Psychiatry*, **60**, 681–90.
131. Piggott, T.A., Carson, W.H., *et al.*, Aripiprazole Study Group. (2003). Aripiprazole for the prevention of relapse in stabilized patients with chronic schizophrenia: a placebo-controlled 26-week study. *Journal of Clinical Psychiatry*, **64**, 1048–56.
132. Kasper, S., Lerman, M.N., McQuade, R.D., *et al.* (2003). Efficacy and safety of aripiprazole vs. haloperidol for long-term maintenance treatment following acute relapse of schizophrenia. *International Journal of Neuropsychopharmacology*, **6**, 325–37.
133. Kane, J.M., Meltzer, H.Y., *et al.* Aripiprazole Study Group. (2007). Aripiprazole for treatment-resistant schizophrenia: results of a multicenter, randomized, double-blind, comparison study versus perphenazine. *Journal of Clinical Psychiatry*, **68**, 213–23.
134. Tandon, R., Marcus, R.N., Stock, E.G., *et al.* (2006). A prospective, multicenter, randomized, parallel-group, openlabel study of aripiprazole in the management of patients with schizophrenia or schizoaffective disorder in general psychiatric practice: Broad Effectiveness Trial With Aripiprazole (BETA). *Schizophrenia Research*, **84**, 77–89.
135. Chrzanowski, W.K., Marcus, R.N., Torbeyns, A., *et al.* (2006). Effectiveness of long-term aripiprazole therapy in patients with acutely relapsing or chronic, stable schizophrenia: a 52-week, open-label comparison with olanzapine. *Psychopharmacology*, **189**, 259–66.
136. Kern, R.S., Green, M.F., Cornblatt, B.A., *et al.* (2006). The neurocognitive effects of aripiprazole: an open-label comparison with olanzapine. *Psychopharmacology*, **187**, 312–20.
137. Kerwin, R., Millet, B., Herman, E., *et al.* (2007). A multicentre, randomized, naturalistic, open-label study between aripiprazole and standard of care in the management of community treated schizophrenic patients Schizohprenia Trial of Aripiprazole: (STAR) study. *European Psychiatry*, in press.
138. Tran-Johnson, T.K., Stack, D.A., Marcus, R.N., *et al.* (2007). Efficacy and safety of intramuscular aripiprazole in patients with acute agitation: a randomized, double-blind, placebo-controlled trial. *Journal of Clinical Psychiatry*, **68**, 111–9.
139. Kane, J., Canas, F., Kramer, M., *et al.* (2007). Treatment of schizophrenia with paliperidone extended-release tablets: a 6-week placebo-controlled trial. *Schizophrenia Research*, **90**, 147–61.
140. Davidson, M., Emsley, R., Kramer, M., *et al.* (2007). Efficacy, safety and early response of paliperidone extended-release tablets (paliperidone ER): Results of a 6-week, randomized, placebo-controlled study. *Schizophrenia Research*, **93**, 117–30.
141. Kramer, M., Simpson, G., Maciulis, V., *et al.* (2007). Paliperidone extended-release tablets for prevention of symptom recurrence in patients with schizophrenia: a randomized, double-blind, placebo-controlled study. *Journal of Clinical Psychopharmacology*, **27**, 6–14.
142. Meltzer, H.Y., Bobo, W.V., Nuamah, I.F., *et al.* (2008). Efficacy and tolerability of oral paliperidone extended-release tablets in the treatment of acute schizophrenia: pooled data from three 6-week, placebo-controlled studies. *Journal of Clinical Psychiatry*, **69**, 817–29.
143. McKeage, K., Plosker, G.L. (2004). Amisulpride: a review of its use in the management of schizophrenia. *CNS Drugs*, **18**, 933–56.
144. Möller, J., Boyer, P., *et al.* and the PROD–ASLP Study Group. (1997). Improvement of acute exacerbations of schizophrenia with amisulpride: a comparison with haloperidol. *Psychopharmacology*, **132**, 396–401.
145. Freeman, H. (1997). Amisulpride compared with standard neuroleptics in acute exacerbations of schizophrenia: three efficacy studies. *International Clinical Psychopharmacology*, **12**, 11–17.
146. Paillere–Martinot, M., Lecrubier, Y., Martinot, J., *et al.* (1995). Improvement of some schizophrenic defi cit symptoms with low doses of amisulpride. *American Journal of Psychiatry*, **152**, 130–4.
147. Boyer, P., Lecrubier, Y., Pucch, A., *et al.* (1995). Treatment of negative symptoms in schizophrenia with amisulpride. *British Journal of Psychiatry*, **166**, 68–72.
148. Perry, P., Miller, D., Arndt, S., *et al.* (1992). Clozapine and norclozapine plasma concentrations and clinical response of treatment–refractory schizophrenic patients. *American Journal of Psychiatry*, **148**, 231–5.
149. Peuskens, J., Moller H.J. and Puech, A. (2002). Amisulpride improves depressive symptoms in acute exacerbations of schizophrenia: comparison with haloperidol and risperidone. *European Neuropsychopharmacology*, **12**, 305–10.
150. Colonna, L., Saleem, P., Dondey-Nouvel, L., *et al.* (2000). Long-term safety and effi cacy of amisulpride in subchronic or chronic schizophrenia. Amisulpride Study Group. *International Clinical Psychopharmacology*, **15**, 13–22.
151. Surguladze, S., Patel, A., Kerwin, R.W., *et al.* (2005). Cost analysis of treating schizophrenia with amisulpride: naturalistic mirror image study. *Progress in Neuropsychopharmacology and Biological Psychiatry*, **29**, 517–22.
152. Kelleher, J.P., Centorrino, F., Albert, M.J., *et al.* (2002). Advances in atypical antipsychotics for the treatment of schizophrenia: new formulations and new agents. *CNS Drugs*, **16**, 249–61.
153. Bell, R., McLaren, A., Glanos, J., *et al.* (1998). The clinical use of plasma clozapine levels. *Australian and New Zealand Journal of Psychiatry*, **32**, 567–74.
154. Prior, T.I., Chue, P.S., Tibbo, P., *et al.* (1999). Drug metabolism and atypical antipsychotics. *European Neuropsychopharmacology*, **9**, 301–9.
155. Olesen, O.V., Linnet, K. (2001). Contributions of five human cytochrome P450 isoforms to the N-demethylation of clozapine in vitro at low and high concentrations. *Journal of Clinical Pharmacology*, **41**, 823–32.
156. He, H. and Richardson, J. (1995). A pharmacological, pharmacokinetic and clinical overview of risperidone, a new antipsychotic that blocks serotonin 5–HT2 and dopamine D2 receptors. *International Clinical Psychopharmacology*, **10**, 19–30.

157. Aravagiri, M., Ames, D., Wirshing, W., *et al.* (1997). Plasma level monitoring of olanzapine in patients with schizophrenia: determination by high–performance liquid chromatography with electrochemical detection. *Therapeutic Drug Monitoring*, **19**, 307–13.
158. Casey, D. (1996). 'Seroquel'(quetiapine): preclinical and clinical findings of a new atypical antipsychotic. *Experimentation, Opinion and Investigation of Drugs*, **5**, 939–57.
159. Potkin, S.G., Thyrum, P.T., Alva, G., *et al.* (2002). The safety and pharmacokinetics of quetiapine when coadministered with haloperidol, risperidone, or thioridazine. *Journal of Clinical Psychopharmacology*, **22**, 121–30.
160. Chengappa, K.N., Parepally, H., Brar, J.S., *et al.* (2003). A random-assigment, double-blind, clinical trial of once- vs twice-daily administration of quetiapine fumarate in patients with schizophrenia or schizoaffective disorder: a pilot study. *Canadian Journal of Psychiatry*, **48**, 198–94.
161. Gefvert, O., Bergstrom, M., Langstrom, B., *et al.* (1998). Time course of central nervous dopamine-D2 and 5-HT2 receptor blockade and plasma drug concentrations after discontinuation of quetiapine (Seroquel) in patients with schizophrenia. *Psychopharmacology*, **135**, 119–26.
162. Beedham, C., Miceli, J.J., Obach, R.S. (2003). Ziprasidone metabolism, aldehyde oxidase, and clinical implications. *Journal of Clinical Psychopharmacology*, **23**, 229–32.
163. Swainston, H.T. and Perry, C.M. (2004). Aripiprazole: a review of its use in schizophrenia and schizoaffective disorder. *Drugs*, **64**, 1715–36.
164. Jung, S.M., Kim, K.A., Cho, H.K., *et al.* (2005). Cytochrome P450 3A inhibitor itraconazole affects plasma concentrations of risperidone and 9-hydroxyrisperidone in schizophrenic patients. *Clinical Pharmacology and Therapeutics*, **78**, 520–8.
165. Sachdev, P. (1995). The epidemiology of drug-induced akathisia: Part I. Acute akathisia. *Schizophrenia Bulletin*, **21**, 431–49.
166. Atbasoglu, E.C., Schultz, S.K. and Andreasen, N.C. (2001). The relationship of akathisia with suicidality and depresonalization among patients with schizophrenia. *Journal of Neuropsychiatry and Clinical Neuroscience*, **13**, 336–41.
167. Cavallaro, R. and Smeraldi, E. (1995). Antipsychotic–induced tardive dyskinesia. *CNS Drugs*, **4**, 278–93.
168. Kane, J.M. (2004). Tardive dyskinesia rates with atypical antipsychotics in adults: prevalence and incidence. *Journal of Clinical Psychiatry*, **65** (suppl 9), 16–20.
169. Velamoor, V. (1998). Neuroleptic malignant syndrome. Recognition, prevention and management. *Drug Safety*, **19**, 73–81.
170. Alvir, J., Lieberman, J., Safferman, A., *et al.* (1993). Clozapine–induced agranulocytosis: incidence and risk factors in the United States. *New England Journal of Medicine*, **329**, 162–7.
171. Lieberman, J. and Safferman, A. (1992). Clinical profile of clozapine: adverse reactions and agranulocytosis. *Psychiatric Quarterly*, **63**, 51–70.
172. Merrill, D.B., Dec, G.W. and Goff, D.C. (2005). Adverse cardiac effects associated with clozapine. *Journal of Clinical Psychopharmacology*, **25**, 32–41.
173. Collaborative Working Group. (1998). Adverse effects of the atypical antipsychotics. *Journal of Clinical Psychiatry*, **12**, 17–22.
174. Kelly, D.L., Love, R.C. (2001). Ziprasidone and the QTc interval: pharmacokinetic and pharmacodynamic considerations. *Psychopharmacology Bulletin*, 35, 66–79.
175. Garver, D.L., Holcomb, J.A. and Christensen, J.D. (2005). Cerebral cortical gray expansion associated with two second-generation antipsychotics. *Biological Psychiatry*, **58**, 62–6.
176. Marder, S.R., Kramer, M., Ford, L., *et al.* (2007). Efficacy and safety of paliperidone extended-release tablets: results of a 6-week, randomized, placebo-controlled study. Biological Psychiatry, **62**, 1363–70.
177. Papakostas, G.I., Shelton, R.C., Smith, J., *et al.* (2007). Augmentation of antidepressants with atypical antipsychotic medication for treatment-resistant major depressive disorder: a meta-analysis. *Journal of Clinical Psychiatry*, **68**, 826–31.

6.2.6 Antiepileptic drugs

Brian P. Brennan and Harrison G. Pope Jr.

Introduction

Several drugs originally developed to treat epilepsy have been found effective in certain psychiatric disorders. This chapter reviews the antiepileptic drugs most extensively studied in psychiatric disorders: valproate, carbamazepine, lamotrigine, and topiramate. We then briefly mention six other antiepileptics currently under investigation in various psychiatric disorders, but not yet extensively studied: gabapentin, oxcarbazepine, levetiracetam, tiagabine, zonisamide, and pregabalin. The antiepileptic drug phenytoin is rarely used in psychiatric disorders, and is therefore not included in this chapter. The benzodiazepines, which have antiepileptic properties, are also omitted here, as they are discussed in Chapter 6.2.2. We briefly list studies documenting the efficacy of these various agents in psychiatric disorders, but the reader is referred to the individual chapters on specific disorders for a more detailed discussion of treatment strategies.

Valproate[1]

Introduction

Valproate (valproic acid) is a simple branched-chain carboxylic acid, first used as an organic solvent in the late 1800s (see Fig. 6.2.6.1). Its antiepileptic properties were discovered serendipitously in 1963, and its clinical use as an antiepileptic drug began in 1964. As early as 1966, valpromide (the amide precursor of valproate) was reported to be effective in the treatment of bipolar disorder.(1) Since then, valproate has been used effectively in the treatment of numerous psychiatric and neurologic conditions, and is now widely used as a mood stabilizer in the treatment of bipolar disorder.

Valproate is currently available as five different preparations: valproate (Depakene), sodium valproate (Depakene syrup), divalproex sodium (Depakote) (which is an equal proportion of sodium valproate and valproic acid), divalproex sodium sprinkle capsules (Depakote sprinkle capsules), and valpromide (the amide precursor of valproate, which is available in Europe, but not in the United States).

Pharmacology

The mechanism of action of valproate in the treatment of epilepsy is unclear, but appears to be related to increased levels of gamma-aminobutyric acid (GABA) in the brain. It inhibits the breakdown and turnover of GABA, increases its release, and increases the density of the GABA-ββ receptor subtype.(2) Its mechanism of action in treating psychiatric disorders is unknown.

[1] Valproate is marketed in the British Commonwealth as 'valproic acid' and as 'sodium valproate' in the US, but these are effectively interchangeable as they both yield valproate in the bloodstream.

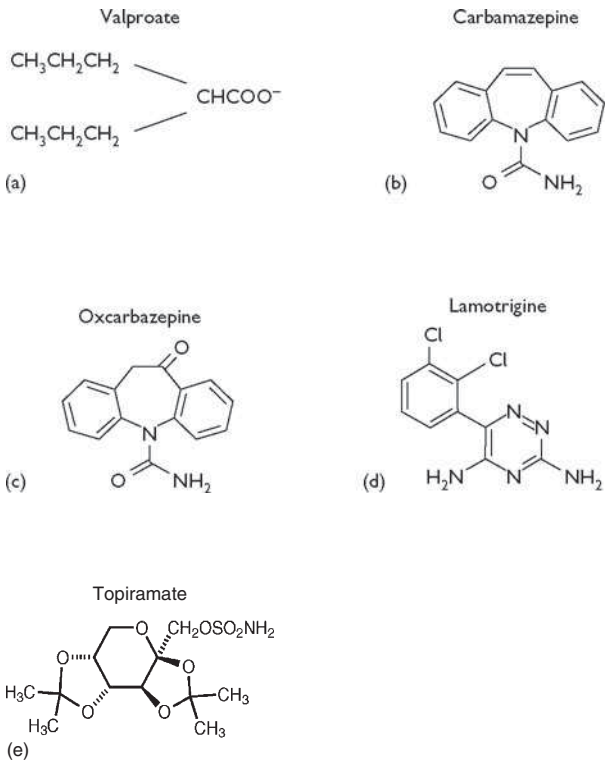


Fig. 6.2.6.1 Molecular structures of selected antiepileptic drugs. (a) Valproate, (b) Carbamazepine, (c) Oxcarbazepine, (d) Lamotrigine, and (e) Topiramate.

Pharmacokinetics

All preparations of valproate are completely absorbed after oral administration. The rate of absorption varies slightly with the different preparations, but these differences are probably not clinically significant. Co-administration with food can delay absorption. Valproate is approximately 90 per cent protein-bound. Only the unbound drug crosses the blood brain barrier and is pharmacologically active in the CNS. As total serum valproate concentration increases, the unbound portion of valproate is disproportionately increased, presumably due to saturation of the protein-binding sites. Therefore, at higher serum concentrations, small increases in dose may result in significant changes in efficacy and side effects. Valproate is metabolized by the liver to a glucuronide conjugate or one of several metabolites, some having antiepileptic activity. The half-life of valproate ranges from 6 to 17 h. Enzyme-inducing antiepileptic drugs, such as carbamazepine and phenytoin, shorten the half-life of valproate (see Interactions).[3]

Side effects

Valproate is often associated with minor side effects, but can rarely cause life-threatening, idiosyncratic reactions. Common side effects include gastrointestinal symptoms, such as nausea, vomiting, abdominal pain, and diarrhoea; and neurological symptoms, such as tremor, somnolence, and dizziness. Weight gain is also common. Hair loss occurs in some patients, but is often transient and reversible upon discontinuation of the drug. Rare, but potentially fatal, idiosyncratic reactions include hepatic failure, acute haemorrhagic pancreatitis, and agranulocytosis.[2,4] Known risk factors for irreversible hepatic failure include young age (especially less than 2 years old), developmental delay, a metabolic disorder, or concomitant administration of other antiepileptic drugs.[5] Because of this risk, liver function tests are recommended prior to initiation of therapy and periodically thereafter (see Dosage and Administration).

Toxic effects

(a) Overdose

Overdose with valproate can result in heart block, coma, and death. Haemodialysis may be useful in clearing the drug rapidly, and naloxone may reverse the CNS depressant effects.[3]

(b) Pregnancy

Valproate increases the risk of neural tube defects (such as spina bifida) to approximately 1–2 per cent of pregnancies when administered in the first trimester. Other reported congenital anomalies include craniofacial defects and cardiovascular malformations. Valproate is found in human breast milk, at approximately 1–10 per cent of serum concentrations, but its effects on the nursing child are unknown.[6]

Indications and contraindications

Controlled trials confirm that valproate is effective in the treatment of multiple seizure types, including complex partial, simple and complex absence, generalized tonic–clonic, and myoclonic seizures.[3] Several controlled studies indicate efficacy in the treatment of acute mania,[7–13] mixed episodes,[12,14] and in the prophylaxis of recurrent mood episodes.[13,15,16] One small controlled study has offered limited evidence for the efficacy of valproate in the treatment of bipolar depression.[17] There is growing support from controlled studies for the efficacy of valproate, combined with both typical and atypical antipsychotics, in the treatment of acute exacerbations of schizophrenia[18,19]; particularly when the presentation includes agitation and hostility.[20] There are also several small controlled studies demonstrating the benefit of valproate in the treatment of mood instability and impulsivity associated with borderline personality disorder.[21–23] Other conditions for which valproate may sometimes be useful include pain syndromes, anxiety disorders, alcohol and sedative withdrawal syndrome, impulse control disorders, and behavioural and affective disturbances associated with intellectual disability and dementia.[2]

Valproate is contraindicated in patients with known hypersensitivity to the drug. It should be used cautiously in patients with significant hepatic disease.

Interactions

In general, valproate can be combined safely with other psychotropic medications and antiepileptic drugs. However, given that valproate is highly protein-bound and can inhibit hepatic enzymes, some drug–drug interactions have been identified.[3] **Aspirin**, which is highly protein-bound, elevates the free fraction of valproate, resulting in increased effects of valproate on the CNS. Valproate can displace **diazepam**, **phenytoin**, **carbamazepine**, and **warfarin** from protein-binding sites, resulting in increased activity of these drugs. Co-administration of valproate with **lamotrigine** significantly increases the half-life of the latter and can increase the risk of lamotrigine-induced rashes. When administered with **carbamazepine**, three potential interactions may occur: (1) valproate can increase the concentration of carbamazepine's metabolite, carbamazepine-10,11-epoxide, by inhibiting its further metabolism; (2) carbamazepine may lower the valproate level; and (3) valproate may increase the carbamazepine level.[24] Therefore, close monitoring of serum

concentrations of both drugs is important when they are combined. **Amitriptyline** and **fluoxetine** may increase serum valproate concentrations, possibly by inhibition of valproate metabolism.

Effects of withdrawal

As with other antiepileptic drugs, valproate should be tapered gradually over several weeks to minimize the risk of rebound seizures.

Dosage and administration

Before initiating treatment with valproate, it is advisable to obtain a baseline complete blood count (CBC), liver function tests (LFTs), and if appropriate, a pregnancy test. CBC and liver function tests should be performed monthly for the first 3 months, and, if no abnormalities are found, every 6 to 12 months thereafter. If hepatic transaminase levels increase to more than three times normal, valproate should be discontinued. If the transaminase levels eventually return to baseline and the patient responded to valproate previously, re-challenge can be considered. If hepatic transaminase levels increase, but are less than three times normal, monitoring should be increased to once every 1–2 weeks until transaminase levels stabilize, and then monthly thereafter.[2]

The initial starting dose of valproate in adults is 250 to 1000 mg per day, given in two or three divided doses (see Table 6.2.6.1 for dosage forms). The dose may be increased every 1–3 days depending on the patient's response and tolerance. The usual therapeutic concentration is between 50–100 μg/ml (drawn 12 h after the last dose) for both psychiatric and neurological disorders. Some clinicians give the entire daily dose of valproate at bedtime. In patients with seizure disorders or acute mania, an oral loading strategy can be used.[25] In this situation, the patient receives 20 mg/kg as a bolus on the first day, resulting in rapid achievement of therapeutic levels. However, psychiatric patients who are not acutely manic usually have difficulty tolerating the oral loading strategy.

Table 6.2.6.1 Available dosage forms of antiepileptic drugs

Drug (proprietary name)	Preparation
Valproate	
Valproate (Depakene)	250 mg capsule
Valproate syrup (Depakene syrup)	250 mg/5 ml
Divalproex sodium (Depakote)	125, 250, and 500 mg tablets
Divalproex sodium extended release (Depakote ER)	250, 500 mg tablets
Divlaproex sodium sprinkle capsules	125 mg capsule
Carbamazepine	
Carbamazepine (Tegretol®)	100, 200 mg tablets; 100 mg chewable tablets; suspension of 100 mg/5 ml
Carbamazepine extended-release tablets	
(Tegretol XR®)	100, 200, and 400 mg tablets
(Carbatrol®)	100, 200, and 300 mg capsules
(Equetro®)	100, 200, and 300 mg capsules
Oxcarbazepine (Trileptal®)	150, 300 and 600 mg tablets; suspension of 300 mg/5 ml
Lamotrigine (Lamictal®)	25, 100, 150, and 200 mg tablets; 2, 5, and 25 mg chewable tablets

Carbamazepine

Introduction

Carbamazepine (Tegretol®) is an iminostilbene derivative with a structure similar to the tricyclic antidepressant imipramine (see Fig. 6.2.6.1). It was initially developed as a potential antidepressant in the 1950s, but was found to have antiepileptic and analgesic properties, and has been marketed for the treatment of seizures and pain syndromes since 1963. For many clinicians, it has been the preferred treatment for partial and generalized tonic–clonic seizures, as well as neuropathic pain. Its clinical use in affective disorders began in the early 1970s; since then, it has become widely used in psychiatry.

Pharmacology

Carbamazepine's mechanism of action in the treatment of seizures and pain syndromes is controversial, but probably results from blockade of voltage-sensitive sodium channels or enhancement of gamma-aminobutyric acid (GABA) activity. Its mechanism of action in psychiatric disorders is unknown, and may be different, given that it affects numerous neurotransmitter systems.[26,27]

Pharmacokinetics

Carbamazepine is absorbed slowly, with peak plasma levels occurring 4–5 h after administration of the tablets. Absorption is faster for the carbamazepine liquid, and slower for carbamazepine extended-release tablets. Oral bioavailability is about 80 per cent; plasma protein binding is approximately 75 per cent. The half-life of carbamazepine is variable, as it induces its own metabolism with chronic administration (autoinduction). Initially, the half-life ranges from 18–65 h, but after autoinduction is complete (usually 3–5 weeks), it is decreased to 5–25 h. Children metabolize carbamazepine more rapidly than adults, and therefore require higher doses to achieve similar levels. Carbamazepine is metabolized in the liver by the cytochrome P_{450} system to a wide variety of metabolites, some with antiepileptic activity. The predominant metabolite, carbamazepine-10, 11-epoxide (CBZ-E), is further metabolized by epoxide hydrolase to an inactive form. Most of carbamazepine's metabolites are excreted as glucuronide conjugates in the urine.[28,29]

Side effects

Carbamazepine is generally well tolerated, with less than 5 per cent of patients discontinuing the medication because of adverse effects. Common side effects seen during initiation of treatment include dizziness, ataxia, sedation, nausea, and diplopia. These are often mild in severity, and frequently resolve with continued treatment.

(a) Haematological side effects

Carbamazepine commonly causes a benign suppression of white blood cell count, but in rare cases may cause severe and potentially fatal blood dyscrasias, including agranulocytosis, pancytopenia, and aplastic anaemia. The incidence of these non-dose-related, idiosyncratic reactions has been estimated to range between 1 in 10 000 to 1 in 300 000.[4]

(b) Hepatic toxicity

Carbamazepine is frequently associated with benign transaminase elevations. Very rarely, a non-dose-related, idiosyncratic reaction causes hepatic failure, which can be fatal.

(c) Cardiovascular effects

Carbamazepine slows intracardiac conduction, and is relatively contraindicated in patients with heart block.

(d) Dermatologic effects

Rashes occur in 5–15 per cent of patients. These are usually benign, but rarely lead to exfoliative dermatitis, Stevens–Johnson syndrome, or toxic epidermal necrolysis. Therefore, it is usually recommended that the drug be discontinued if any rash develops.

(e) Endocrine effects

Carbamazepine can exert antidiuretic effects, which result in hyponatremia in 5–40 per cent of patients.[30] Usually, this effect is clinically insignificant.

Carbamazepine can result in decrease in free T_3 and T_4, but clinical hypothyroidism is extremely rare.

Toxic effects

(a) Overdose

Carbamazepine overdose can be fatal. Common symptoms include nystagmus, tremor, ophthalmoplegia, and myoclonus. Life-threatening effects include atrioventricular block, coma, seizures, and respiratory depression.[31]

(b) Pregnancy

Carbamazepine exposure in the first trimester results in neural tube defects in approximately 1 per cent of infants. Craniofacial abnormalities and developmental delay have been reported as well. Carbamazepine is found in breast milk, but its effects on the nursing infant are unknown.[6,32]

Indications and contraindications

Carbamazepine is indicated for the treatment of simple partial, complex partial, and generalized tonic–clonic seizures. It is ineffective against absence seizures, and may even exacerbate them. Carbamazepine is also indicated in the treatment of trigeminal neuralgia and other neuropathic pain syndromes. Several double-blind, placebo-controlled trials confirm carbamazepine's efficacy in treating both the manic and mixed phase of bipolar disorder.[33,34] There is limited evidence demonstrating efficacy in the treatment of either bipolar or unipolar depression. Uncontrolled reports also suggest that carbamazepine may be useful in the treatment of personality disorders, impulse control disorders, and alcohol/sedative withdrawal syndrome.

Carbamazepine is contraindicated in patients with a history of previous bone marrow depression, hypersensitivity to the drug, or hypersensitivity to any of the tricyclic antidepressants (given its structural similarity to imipramine). Its use with monoamine oxidase inhibitors is not recommended, and carbamazepine should be used with caution in patients with cardiac disease.

Interactions

Given that carbamazepine is extensively metabolized by the liver and induces hepatic enzymes, it produces many significant drug–drug interactions (Table 6.2.6.2).[35–37] Many drug levels are reduced by carbamazepine and can become subtherapeutic. Therefore, it is important to monitor concomitantly administered medications, as dosage adjustments may be necessary.

Table 6.2.6.2 Carbamazepine (CBZ)-drug interactions

CBZ decreases drug levels	Drugs that increase CBZ levels
Alprazolam	Acetazolamide
Clobazam	Cimetidine
Clonazepam	Clarithromycin
Clozapine	Danazol
Dicoumarol	Dextropropoxyphene
Doxycycline	Diltiazem
Ethosuximide	Fluoxetine
Fentanyl	Gemfibrozil
Haloperidol	Isoniazid
Imipramine	Itraconazole
Lamotrigine	Ketaconazole
Mesuximide	Loratadine
Methadone	Macrolide antibiotics
Methylprednisolone	Metronidazole
Oral contraceptives (can result in contraceptive failure)	Nicotinamide
Pancuronium	Nicotinic acid
Paracetamol	Propoxyphene
Phensuximide	Remacemide
Phenytoin (can either increase or decrease)	Rifampicin
Prednisolone	Stiripentol
Primidone	Terfenadine
Remacemide	Valproate
Theophylline	Verapamil
Tiagabine	Viloxazine
Topiramate	
Valproate	
Vecuronium	
Warfarin	
CBZ increases drug levels	**Drugs that decrease CBZ levels**
Clomipramine (possibly)	Cisplatin
Phenytoin (can either increase or decrease)	Doxorubicin
Primidone	Felbamate
	Rifampicin
	Phenobarbital
	Phenytoin
	Primidone
	Theophylline

Effects of withdrawal

As with other antiepileptic drugs, carbamazepine should be gradually tapered over several weeks in order to avoid rebound seizures.

Dosage and administration

Carbamazepine is generally initiated at a starting dose of 100–400 mg, taken either as a single dose or two divided doses (see Table 6.2.6.1 for dosage forms). The dose is gradually increased by 100 or 200 mg every 2 weeks as the patient tolerates. The usual therapeutic serum concentration is 4–12 mg/l (20–50 µmol/l), which is measured before the first morning dose. The half-life of

carbamazepine will decrease with chronic administration due to autoinduction, necessitating frequent monitoring of the serum carbamazepine concentrations and continued dosage adjustment in the first 2 months of therapy.

Laboratory screening

Given the risk of severe blood dyscrasias and hepatic failure, some authorities recommend obtaining a CBC and LFTs at the initiation of treatment. These tests are often repeated every 2 weeks for the first few months of treatment, and then every 3 to 6 months thereafter. However, some authorities argue that testing is unnecessary, since idiosyncratic reactions are rare and may occur too rapidly to be detected by routine laboratory monitoring.

Lamotrigine

Introduction

Lamotrigine (Lamictal®) is a phenyltriazine compound, structurally unrelated to other antiepileptic drugs (see Fig. 6.2.6.1). It was introduced in Ireland in 1993 and in the United Kingdom and the United States in 1994.

Pharmacology

Lamotrigine is thought to act by blocking voltage-sensitive sodium channels, and by inhibiting the release of glutamate. In experimental animal seizure models, it has an antiepileptic profile similar to that of phenytoin and carbamazepine.[38]

Pharmacokinetics

The oral bioavailability of lamotrigine approaches 100 per cent, and absorption is unaffected by food. Peak plasma concentrations are reached 2–3 h after an oral dose. The half-life of lamotrigine is approximately 30 h, but is altered by the presence of other antiepileptic drugs (see Interactions). Plasma protein binding is approximately 55 per cent. Lamotrigine is metabolized by the liver to an inactive glucuronide conjugate, and then excreted in the urine. The clearance of lamotrigine may be reduced in patients with renal impairment and Gilbert's syndrome, and these individuals may benefit from dosage reduction.[39]

Side effects

In general, lamotrigine has few side effects and is better tolerated than other antiepileptic drugs. The most common side effects include dizziness, headache, diplopia, ataxia, blurred vision, nausea, somnolence, and rash. The most concerning side effect is skin rash, which can be life-threatening. Approximately 10 per cent of adults develop a rash while taking lamotrigine, but the majority of these are benign. However, about 1 in 1000 will develop a life-threatening rash, such as Stevens–Johnson syndrome or toxic epidermal necrolysis. The incidence of rash is much higher in paediatric patients, occurring in 1 in 50 to 1 in 100 patients; therefore, lamotrigine should be used with caution in patients less than 16 years of age. Starting at a low dose and slowly increasing it can minimize the risk of rash. Co-administration of valproate can increase the risk of rash. Given the difficulty in predicting who will develop a life-threatening rash, lamotrigine is usually discontinued at the first sign of any rash.

There are a few reports of possible idiosyncratic reactions in patients taking lamotrigine. These include disseminated intravascular coagulation, multiorgan failure, and acute hepatic necrosis.[40] It is unclear, however, if these conditions were actually caused by the drug itself.

Toxic effects

(a) Overdose

The few reported cases of overdose on lamotrigine (at doses up to 4000 mg) were not fatal, but resulted in symptoms such as excessive sedation, dizziness, and headache.

(b) Pregnancy

The effects of lamotrigine on human pregnancy and breast-fed infants are unknown.

Indications and contraindications

Several double-blind, placebo-controlled, add-on trials confirm lamotrigine's efficacy in treating some patients with partial or generalized tonic–clonic seizures.[39] Clinical trials also suggest efficacy against absence, atypical absence, and myoclonic seizures, as well as seizures associated with the Lennox–Gastaut syndrome.[41] Lamotrigine demonstrated efficacy in the treatment of bipolar depression in a large placebo-controlled trial.[42] However, it does not appear to be beneficial in the treatment of acute mania, largely due to the drug's long titration schedule. Several controlled studies have also demonstrated efficacy in the maintenance treatment of bipolar disorder.[39,43–45] and in the rapid-cycling subtype of bipolar disorder.[46] In addition to bipolar depression, lamotrigine has shown some benefit in two small placebo-controlled studies when added to selective serotonin reuptake inhibitors (SSRIs) as an augmentation strategy in the treatment of unipolar depression.[47,48] Preliminary evidence from two small placebo-controlled studies also indicates that lamotrigine may have benefit as an augmentation therapy with conventional and atypical antipsychotics in treatment-resistant schizophrenia.[49,50]

Interactions

Lamotrigine does not appear to affect the kinetics of other antiepileptic drugs or oral contraceptives, but its own kinetics are markedly affected by the concomitant administration of other antiepileptic drugs. Valproate inhibits the metabolism of lamotrigine, resulting in a doubling of the half-life to almost 60 h. The enzyme-inducing antiepileptic drugs, such as carbamazepine and phenytoin, decrease the half-life to approximately 15 h. These interactions necessitate dosage adjustments when starting lamotrigine (see Dosage and Administration).[38]

Effects of withdrawal

As with other antiepileptic drugs, lamotrigine should be tapered gradually over several weeks in order to avoid rebound seizures.

Dosage and administration

Lamotrigine must be started at a low dose and increased with caution to a therapeutic dosage in order to minimize the risk of rash. The patient should be informed of the risk of developing a rash, and instructed to contact the physician immediately if one appears. The starting dose depends upon the concomitant administration of other antiepileptic drugs. See Tables 6.2.6.3–6.2.6.5 for the appropriate lamotrigine titration schedules. Table 6.2.6.1 shows dosage forms.

Table 6.2.6.3 Suggested lamotrigine titration schedule for patients taking carbamazepine, phenytoin, phenobarbital, primidone, or rifampicin and not taking valproate

Week 1	25 mg daily
Week 2	50 mg daily
Week 3	100 mg daily, in divided doses
Week 4	100 mg daily, in divided doses
Week 5	200 mg daily, in divided doses
Week 6	300 mg daily, in divided doses
Week 7	Up to 400 mg daily, in divided doses

Table 6.2.6.4 Suggested lamotrigine monotherapy for patients not taking carbamazepine, phenytoin, phenobarbital, primidone, or rifampicin and not taking valproate

Week 1	12.5 mg daily
Week 2	25 mg daily
Week 3	50 mg daily
Week 4	50 mg daily
Week 5	100 mg daily
Week 6	150 mg daily
Week 7	200 mg daily

Table 6.2.6.5 Suggested lamotrigine titration schedule for patients taking valproate

Week 1	12.5 mg every *other* day
Week 2	25 mg every *other* day
Week 3	25 mg daily
Week 4	25 mg daily
Week 5	50 mg daily
Week 6	75 mg daily
Week 7	100 mg daily

Topiramate

Introduction

Topiramate (Topamax®) is a sulfamate-substituted monosaccharide originally developed as an oral hypoglycemic agent. It was subsequently found to have antiepileptic effects. It is currently approved for epilepsy in over 60 countries and for migraine prophylaxis in more than 20 countries.

Pharmacology

Topiramate is believed to act through several different mechanisms including: (1) inhibition of sodium channel conductance; (2) inhibition of L-type calcium channels; (3) increase in GABA release through an unknown mechanism; (4) decrease in glutamate-mediated excitation through blockade of kainate receptors; and (5) inhibition of carbonic anhydrase.[51]

Pharmacokinetics

There is almost complete, linear, and rapid absorption of topiramate across dose ranges.[52] The absorption of topiramate is not affected by food. It is not significantly protein-bound. Hepatic metabolism (~20 per cent) is less important than renal clearance (~80 per cent) and inactive metabolites comprise less than 5 per cent the administered dose.[52] The normal half-life of topiramate is 19–23 h; this is not influenced by the administration of hepatic enzyme-inducing medications such as carbamazepine or phenytoin.[52]

Side effects

The most common side effects of topiramate are memory and concentration difficulties, paresthesias, somnolence, dizziness, anorexia, and weight loss. In clinical trials with topiramate, kidney stones are reported 2–4 times more frequently than in the general population, probably as a result of the drug's inhibition of carbonic anhydrase. Patients predisposed to developing kidney stones should maintain good hydration during topiramate therapy to minimize the risk of renal stone formation. Rarely, topiramate has been associated with acute myopia precipitating secondary angle closure glaucoma and with oligohidrosis leading to hyperthermia.

Toxic effects

(a) Overdose

Several cases of topiramate overdose have been recorded, with signs and symptoms including severe metabolic acidosis, convulsions, drowsiness, speech disturbance, blurred vision, diplopia, lethargy, tupor, hypotension, abdominal pain, agitation, and dizziness. The clinical consequences in most cases were not severe. However, deaths have occurred after poly-drug overdoses involving topiramate.

(b) Pregnancy

The effects of topiramate on human pregnancy and breast-fed infants are unknown.

Indications and contraindications

Topiramate has confirmed efficacy as monotherapy for partial-onset seizures or primary generalized tonic–clonic seizures for patients 10 years and older and as adjunctive therapy for the same seizure types, including seizures associated with Lennox–Gastault syndrome, in patients of age 2 and older.[52] In addition, several large controlled studies have demonstrated benefit in the prophylaxis of migraines.[53–55] It may also be helpful in the treatment of essential tremor and chronic pain syndromes. There are no large placebo-controlled studies investigating the use of topiramate as monotherapy for bipolar disorder. However, there is limited evidence from several small open-label studies for topiramate as an adjunctive treatment for both the depressive[56] and manic[57,58] phases of bipolar disorder. Controlled studies have demonstrated efficacy in eating disorders including binge eating disorder[52] and bulimia nervosa.[59] Given its ability to reduce weight, the utility of topiramate may be greatest as an adjunctive treatment in bipolar disorder with comorbid obesity. Several small placebo-controlled studies have also suggested that topiramate may have benefit in the treatment of alcohol dependence,[60] impulse control disorders such as pathological gambling,[61] and borderline personality disorder.[62]

Interactions

Topiramate does not typically alter the metabolism of any other drugs. One important exception to this is that the effectiveness of

the ethinylestradiol component of oral contraceptives is reduced when more than 200 mg of topiramate is prescribed daily.[52] Topiramate occasionally leads to a modest increase in phenytoin concentrations (0–25 per cent) and to modest increases in the clearance of risperidone, pioglitazone, and lithium.[52] Hydrochlorothiazide can lead to modest increases in serum concentration of topiramate.[52] Concomitant administration of topiramate and valproate has been associated with hyperammonemia with or without encephalopathy in patients who have previously tolerated either drug alone.

Effects of withdrawal

As with other antiepileptic drugs, topiramate should be tapered gradually over several weeks in order to avoid rebound seizures.

Dosage and administration

The recommended dose for monotherapy treatment of epilepsy in adults and children 10 years of age or older is 400 mg/day in two divided doses. It is recommended that therapy be initiated at 25–50 mg/day followed by titration to an effective dose in increments of 25–50 mg/week. The recommended total daily dose for migraine prophylaxis is 100 mg/day administered in two divided doses. In patients with renal impairment one-half of the usual adult dose is recommended.

Other antiepileptics

Several other drugs with antiepileptic activity are currently under investigation in various psychiatric disorders, but have not as yet shown well-documented efficacy. Gabapentin (Neurontin®), a structural analog of gamma-aminobutyric acid (GABA), has been marketed for use as adjunctive therapy in the treatment of epilepsy since 1993 and also carries a clinical indication for postherpetic neuralgia. It displays a pharmacological profile similar to phenytoin and carbamazepine in animal seizure models.[63]

Several placebo-controlled trials confirm gabapentin's efficacy as adjunctive therapy in some patients with partial seizures, especially complex partial seizures and partial seizures with secondary generalization.[42] Additionally, several controlled studies have demonstrated benefit in the treatment of various pain syndromes, particularly neuropathic pain.[64,65] In a large, multicentre, controlled study gabapentin failed as an adjunctive treatment with lithium and/or valproate in the treatment of acute manic or mixed episodes, and actually performed significantly more poorly than placebo.[66] Since that time, however, a small controlled study has suggested that gabapentin may have efficacy as an adjunctive maintenance treatment for bipolar disorder.[67] Two small controlled studies have also demonstrated efficacy for gabapentin in the treatment of panic disorder[68] and social anxiety[69] respectively.

Oxcarbazepine (Trileptal®), the 10-keto analog of carbamazepine, has an antiepileptic effect similar to carbamazepine, but has fewer side effects, and has not been associated with severe blood dyscrasias. The drug also has fewer effects on hepatic enzyme activity, and hence causes fewer drug interactions. Given its similarity to carbamazepine, oxcarbazepine might be expected to be effective for the manic phase of bipolar disorder. However, despite some promising uncontrolled observations, there are as yet no large controlled trials of oxcarbazepine as monotherapy in adult bipolar disorder. A large double-blind study of oxcarbazepine in children with manic or mixed episodes did not demonstrate efficacy.[70] Preliminary evidence from one small open-label study suggests that oxcarbazepine may also reduce the risk of relapse prevention in alcohol abuse.[71]

Zonisamide (Zonegran®) is an antiepileptic that causes carbonic anhydrase inhibition; it is somewhat similar in its pharmacologic profile to topiramate. Like topiramate, it frequently causes weight loss, and has been shown effective for weight loss in a controlled study of obese individuals with binge eating disorder.[72] It also increases the risk of kidney stones to at least the same degree as topiramate. Small open-label studies have suggested that zonisamide may be effective in the treatment of both the manic and depressed phases of bipolar disorder.[73,74]

Tiagabine (Gabitril®) is believed to exert its anticonvulsant effects by enhancing the effects of GABA. This is thought to occur through blockade of reuptake of GABA into presynaptic neurones. Tiagabine showed efficacy in the treatment of generalized anxiety disorder in a large double-blind study.[75] In addition, several small open-label studies have demonstrated preliminary evidence benefit in the treatment of posttraumatic stress disorder,[76] major depressive disorder with comorbid anxiety,[77] and as an augmentation therapy for generalized anxiety disorder.[78] Tiagabine demonstrated limited efficacy as an add-on treatment for treatment-refractory bipolar disorder in one small clinical case series.[79]

Levetiracetam (Keppra®) is a novel anticonvulsant with an unclear mechanism of action. It has demonstrated benefit in the treatment of social anxiety disorder in one small controlled study, warranting further investigation into this potential use.[80]

Pregabalin (Lyrica®) is a structural derivative of GABA. However, it does not act by altering GABA levels or by binding to the GABA receptor. The exact mechanism of action of pregabalin is unknown, although it is most likely related to its high affinity for the alpha$_2$-delta site on voltage-gated calcium channels. Pregabalin has shown promise as an anxiolytic. Three separate large placebo-controlled studies have demonstrated benefit with pregabalin in the treatment of generalized anxiety disorder.[81–83] Pregabalin also demonstrated efficacy in the treatment of social anxiety disorder in one large controlled study.[84]

It is important to note that soon after the completion of this text, the United States Food and Drug Administration released an alert recommending that patients taking antiepileptic drugs be closely monitored for changes in behavior that could indicate worsening of depression or the emergence of suicidal thoughts or behavior. This warning was based on a meta-analysis of placebo-controlled trials involving eleven antiepileptic medications, including both trials in epilepsy and trials in psychiatric disorders. This meta-analysis revealed that patients receiving antiepileptic drugs had approximately twice the risk of suicidal behavior or ideation (0.43%) compared to patients receiving placebo (0.22%).[85] Further investigation and analyses of this possible association are ongoing.

Further information

McElroy, S.L. and Pope, H.G. Jr. (1988). *Use of anticonvulsants in psychiatry: recent advances*. Oxford Health Care, Clifton, NJ.

Muzina, D.J., El-Sayegh, S., and Calabrese, J.R. (2002). Antiepileptic drugs in psychiatry: focus on randomized controlled trials. *Epilepsy Research*, **50**(1–2), 195–202.

Weisler, R.H., Cutler, A.J., Ballenger, J.C., *et al.* (2006). The use of antiepileptic drugs in bipolar disorders: a review based on evidence from controlled trials. *CNS Spectrums*, **11**(10), 788–99.

Ovsiew, F. (2004). Antiepileptic drugs in psychiatry. *Journal of Neurology, Neurosurgery, and Psychiatry*, **75**(12), 1655–8.

References

1. Lambert, P.A., Carraz, G., Borselli, S., *et al.* (1966). Action neuropsychotrope d'un nouvel anti-épileptique: le dépamide. *Annales Medico-psychologiques*, **124**, 707–10.
2. Pope, H.G. Jr. and McElroy, S.L. (1995). Valproate. In *Comprehensive textbook of psychiatry/VI* (6th edn) (eds. H. Kaplan and B. Sadock), pp. 2112–20. Williams & Wilkins, Baltimore.
3. Davis, R., Peters, D.H., and McTavish, D. (1994). Valproate. A reappraisal of its pharmacological properties and clinical efficacy in epilepsy. *Drugs*, **47**, 332–72.
4. Tohen, M., Castillo, J., Baldessarini, R.J., *et al.* (1995). Blood dyscrasias with carbamazepine and valproate: a pharmacoepidemiological study of 2,228 patients at risk. *The American Journal of Psychiatry*, **152**, 413–18.
5. Bryant, A.E. III and Dreifuss, F.E. (1996). Valproate hepatic fatalities. III. U.S. experience since 1986. *Neurology*, **46**, 465–9.
6. Chang, S.I. and McAuley, J.W. (1998). Pharmacotherapeutic issues for women of childbearing age with epilepsy. *The Annals of Pharmacotherapy*, **32**, 794–804.
7. American Psychiatric Association. (2002). Practice guideline for the treatment of patients with bipolar disorder. *The American Journal of Psychiatry*, **159**(Suppl. 4), 1–50.
8. Pope, H.G. Jr, McElroy, S.L., Keck, P.E. Jr., *et al.* (1991). Valproate in the treatment of acute mania. A placebo-controlled study. *Archives of General Psychiatry*, **48**, 62–8.
9. Kravitz, H. M. and Fawcett, J. (1994). Efficacy of divalproex vs lithium and placebo in mania. *The Journal of the American Medical Association*, **272**, 1005–6.
10. Bowden, C.L., Brugger, A.M., Swann, A.C., *et al.* (1994). Efficacy of divalproex vs lithium and placebo in the treatment of mania. The Depakote mania study group. *The Journal of the American Medical Association*, **271**, 918–24.
11. McElroy, S.L., Keck, P.E., Stanton, S.P., *et al.* (1996). A randomized comparison of divalproex oral loading versus haloperidol in the initial treatment of acute psychotic mania. *The Journal of Clinical Psychiatry*, **57**, 142–6.
12. Freeman, T.W., Clothier, J.L., Pazzaglia, P., *et al.* (1992). A double-blind comparison of valproate and lithium in the treatment of acute mania. *The American Journal of Psychiatry*, **149**, 108–11.
13. Tohen, M., Ketter, T.A., Zarate, C.A., *et al.* (2003). Olanzapine versus divalproex sodium for the treatment of acute mania and maintenance of remission: a 47-week study. *The American Journal of Psychiatry*, **160**(7), 1263–71.
14. Swann, A.C., Bowden, C.L., Morris, D., *et al.* (1997). Depression during mania. Treatment response to lithium or divalproex. *Archives of General Psychiatry*, **54**(1), 37–42.
15. Bowden, C.L., Calabrese, J.R., McElroy, S.L., *et al.* (2000). A randomized, placebo-controlled 12-month trial of divalproex and lithium in treatment of outpatients with bipolar I disorder. Divalproex Maintenance Study Group. *Archives of General Psychiatry*, **57**(5), 481–9.
16. Gyulai, L., Bowden, C.L., McElroy, S.L., *et al.* (2003). Maintenance efficacy of divalproex in the prevention of bipolar depression. *Neuropsychopharmacology*, **28**(7), 1374–82.
17. Davis, L.L., Bartolucci, A., and Petty, F. (2005). Divalproex in the treatment of bipolar depression: a placebo-controlled study. *Journal of Affective Disorders*, **85**(3), 259–66.
18. Casey, D.E., Daniel, D.G., Wassef, A.A., *et al.* (2003). Effect of divalproex combined with olanzapine or risperidone in patients with an acute exacerbation of schizophrenia. *Neuropsychopharmacology*, **28**(1), 182–92.
19. Wassef, A.A., Dott, S.G., Harris, A., *et al.* (2000). Randomized, placebo-controlled pilot study of divalproex sodium in the treatment of acute exacerbations of chronic schizophrenia. *Journal of Clinical Psychopharmacology*, **20**(3), 357–61.
20. Citrome, L., Casey, D.E., Daniel, D.G., *et al.* (2004). Adjunctive divalproex and hostility among patients with schizophrenia receiving olanzapine or risperidone. *Psychiatric Services*, **55**(3), 290–4.
21. Hollander, E., Allen, A., Lopez, R.P., *et al.* (2001). A preliminary double-blind, placebo-controlled trial of divalproex sodium in borderline personality disorder. *The Journal of Clinical Psychiatry*, **62**(3), 199–203.
22. Frankenburg, F.R. and Zanarini, M.C. (2002). Divalproex sodium in the treatment of women with borderline personality disorder and bipolar II disorder: a double-blind placebo-controlled pilot study. *The Journal of Clinical Psychiatry*, **63**(5), 442–6.
23. Hollander, E., Swann, A.C., Coccaro, E.F., *et al.* (2005). Impact of trait impulsivity and state aggression on divalproex versus placebo response in borderline personality disorder. *The American Journal of Psychiatry*, **162**(3), 621–4.
24. Bernus, I., Dickinson, R.G., Hooper, W.D., *et al.* (1997). The mechanism of the carbamazepine-valproate interaction in humans. *British Journal of Clinical Pharmacology*, **44**, 21–7.
25. Keck, P.E. Jr., McElroy, S.L., Tugrul, K.C., *et al.* (1993). Valproate oral loading in the treatment of acute mania. *The Journal of Clinical Psychiatry*, **54**, 305–8.
26. McElroy, S.L. and Keck, P.E. (1995). Antiepileptic drugs. In *The American psychiatric press textbook of psychopharmacology* (eds. A.F. Schwartzberg and C.B. Nemeroff), pp. 351–75. American Psychiatric Press, Inc., Washington, DC.
27. Post, R.M. (1995). Carbamazepine. In *Comprehensive textbook of psychiatry/VI* (6th edn) (eds. H. Kaplan and B. Sadock), pp. 1964–72. Williams & Wilkins, Baltimore.
28. Maxmen, J.S. and Ward, N.G. (1995). *Psychotropic drugs: fast facts* (2nd edn). W.W. Norton and Company, Inc., New York.
29. Levy, R.H. and Kerr, B.M. (1988). Clinical pharmacokinetics of carbamazepine. *The Journal of Clinical Psychiatry*, **49**(Suppl. 4), 58–61.
30. Van Amelsvoort, T., Bakshi, R., Devaux, C., *et al.* (1994). Hyponatremia associated with carbamazepine and oxcarbazepine therapy: a review. *Epilepsia*, **35**, 181–8.
31. Spiller, H.A., Krenzelok, E.P., and Cookson, E. (1990). Carbamazepine overdose: a prospective study of serum levels and toxicity. *Journal of Toxicology. Clinical Toxicology*, **28**, 445–58.
32. Hansen, D.K., Dial, S.L., Terry, K.K., *et al.* (1996). In vitro embryotoxicity of carbamazepine and carbamazepine-10,11-epoxide. *Teratology*, **54**, 45–51.
33. Weisler, R.H., Kalali, A.H., Ketter, T.A., and SPD417 Study Group. (2004). A multicenter, randomized, double-blind, placebo-controlled trial of extended-release carbamazepine capsules as monotherapy for bipolar disorder patients with manic or mixed episodes. *The Journal of Clinical Psychiatry*, **65**(4), 478–84.
34. Weisler, R.H., Keck, P.E. Jr, Swann, A.C., *et al.* (2005). Extended-release carbamazepine capsules as monotherapy for acute mania in bipolar disorder: a multicenter, randomized, double-blind, placebo-controlled trial. *The Journal of Clinical Psychiatry*, **66**(3), 323–30.
35. Ketter, T.A., Post, R.M., and Worthington, K. (1991). Principles of clinically important drug interactions with carbamazepine. I. *Journal of Clinical Psychopharmacology*, **11**, 198–203.
36. Ketter, T.A., Post, R.M., and Worthington, K. (1991). Principles of clinically important drug interactions with carbamazepine. II. *Journal of Clinical Psychopharmacology*, **11**, 306–13.
37. Spina, E., Pisani, F., and Perucca, E. (1996). Clinically significant pharmacokinetic drug interactions with carbamazepine. An update. *Clinical Pharmacokinetics*, **31**, 198–214.

38. Fitton, A. and Goa, K. (1995). Lamotrigine: an update of its pharmacology and therapeutic use in epilepsy. *Drugs*, **50**, 691–713.
39. Goodwin, G.M., Bowden, C.L., Calabrese, J.R., *et al.* (2004). A pooled analysis of 2 placebo-controlled 18-month trials of lamotrigine and lithium maintenance in bipolar I disorder. *The Journal of Clinical Psychiatry*, **65**(3), 432–41.
40. Calabrese, J.R., Bowden, C.L., Sachs, G.S., *et al.* (1999). A double-blind placebo-controlled study of lamotrigine monotherapy in outapatients with bipolar I depression. Lamictal 602 Study Group. *The Journal of Clinical Psychiatry*, **60**(2), 79–88.
41. Bowden, C.L., Calabrese. J.R., Sachs, G., *et al.* (2003). A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently manic or hypomanic patients with bipolar I disorder. *Archives of General Psychiatry*, **60**(4), 392–400.
42. Calabrese, J.R., Bowden, C.L., Sachs, G., *et al.* (2003). A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently depressed patients with bipolar I disorder. *The Journal of Clinical Psychiatry*, **64**(9), 1013–24.
43. McElroy, S.L., Zarate, C.A., Cookson, J., *et al.* (2004). A 52-week, open-label continuation study of lamotrigine in the treatment of bipolar depression. *The Journal of Clinical Psychiatry*, **65**(2), 204–10.
44. Goodwin, G.M., Bowden, C.L., Calabrese, J.R., *et al.* (2004). A pooled analysis of 2 placebo-controlled 18-month trials of lamotrigine and lithium maintenance in bipolar I disorder. *The Journal of Clinical Psychiatry*, **65**(3), 432–41.
45. Chattergoon, D.S., McGuigan, M.A., Koren, G., *et al.* (1997). Multiorgan dysfunction and disseminated intravascular coagulation in children receiving lamotrigine and valproic acid. *Neurology*, **19**, 1442–4.
46. Calabrese, J.R., Suppes, T., Bowden, C.L., *et al.* (2000). A double-blind, placebo-controlled, prophylaxis study of lamotrigine in rapid-cycling bipolar disorder. Lamictal 614 Study Group. *The Journal of Clinical Psychiatry*, **61**(11), 841–50.
47. Normann, C., Hummel, B., Scharer, L.O., *et al.* (2002). Lamotrigine as adjunct to paroxetine in acute depression: a placebo-controlled, double-blind study. *Journal of Clinical Psychiatry*, **63**(4), 337–44.
48. Barbosa, L., Berk, M., and Vorster, M. (2003). A double-blind, randomized, placebo-controlled trial of augmentation with lamotrigine or placebo in patients concomitantly treated with fluoxetine for resistant major depressive episodes. *The Journal of Clinical Psychiatry*, **64**(4), 403–7.
49. Tiihonen, J., Hallikainen, T., Ryynanen, O.P., *et al.* (2003). Lamotrigine in treatment-resistant schizophrenia: a randomized placebo-controlled trial. *Biological Psychiatry*, **54**(11), 1241–8.
50. Kremer, I., Vass, A., Gorelik, I., *et al.* (2004). Placebo-controlled trial of lamotrigine added to conventional and atypical antipsychotics in schizophrenia. *Biological Psychiatry*, **56**(6), 441–6.
51. van Passel, L., Arif, H., and Hirsch, L.J. (2006). Topiramate for the treatment of epilepsy and other nervous system disorders. *Expert Review of Neurotherapeutics*, **6**(1), 19–31.
52. McElroy, S.L., Arnold, L.M., Shapira, N.A., *et al.* (2003). Topiramate in the treatment of binge eating disorder associated with obesity: a randomized, placebo-controlled trial. *The American Journal of Psychiatry*, **160**(2), 255–61.
53. Diener, H.C., Tfelt-Hansen, T., Dahlof, C., *et al.* MIGR-003 study group. (2004). Topiramate in migraine prophylaxis-results from a placebo-controlled trial with propranolol as an active control. *Journal of Neurology*, **251**(8), 943–50.
54. McIntyre, R.S., Mancini, D.A., McCann, S., *et al.* (2002). Topiramate versus bupropion SR when added to mood stabilizer therapy for the depressive phase of bipolar disorder: a preliminary single-blind study. *Bipolar Disorder*, **4**, 207–13.
55. Chengappa, K.N., Rathore, D., Levine, J., *et al.* (1999). Topiramate as add-on treatment for patients with bipolar mania. *Bipolar Disorder*, **1**, 42–53.
56. Bahk, W.M., Shin, Y.C., Woo, J.M., *et al.* (2005). Topiramate and divalproex in combination with risperidone for acute mania: a randomized, open-label study. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, **29**, 115–21.
57. McElroy, S.L., Arnold, L.M., Shapira, N.A., *et al.* (2003). Topiramate in the treatment of binge eating disorder associated with obesity: a randomized, placebo-controlled trial. *The American Journal of Psychiatry*, **160**(2), 255–61.
58. Silberstein, S.D., Schmitt, J., Neto, W., *et al.* (2004). Topiramate in migraine prevention: results of a large controlled trial. *Archives of Neurology*, **61**(4), 490–5.
59. Hoopes, S.P., Reimherr, F.W., Hedges, D.W., *et al.* (2003). Treatment of bulimia nervosa with topiramate in a randomized, double-blind, placebo-controlled trial, 1. Improvement in binge and purge measures. *The Journal of Clinical Psychiatry*, **64**(11), 1335–41.
60. Johnson, B.A., Ait-Daoud, N., Akhtar, F.Z., *et al.* (2004). Oral topiramate reduces the consequences of drinking and improves the quality of life of alcohol-dependent individuals: a randomized controlled trial. *Archives of General Psychiatry*, **61**(9), 905–12.
61. Dannon, P.N., Lowengrub, K., Gonopolski, Y., *et al.* (2005). Topiramate versus fluvoxamine in the treatment of pathological gambling: a randomized, blind-rater comparison study. *Clinical Neuropharmacology*, **28**(1), 6–10.
62. Nickel, M.K., Nickel, C., Mitterlehner, F.O., *et al.* (2004). Topiramate treatment of aggression in female borderline personality disorder patients: a double-blind, placebo -controlled study. *The Journal of Clinical Psychiatry*, **65**(11), 1515–19.
63. Goa, K.L. and Sorkin, E.M. (1993). Gabapentin. *Drugs*, **46**, 409–27.
64. Serpell, M.G. and Neuropathic Pain Study Group. (2003). Gabapentin in neuropathic pain syndromes: a randomized, double-blind, placebo-controlled trial. *Pain*, **99**, 557–66.
65. Morello, C.M., Leckband, S.G., Stoner, C.P., *et al.* (1999). Randomized, double-blind study comparing the efficacy of gabapentin with amitriptyline on diabetic peripheral neuropathy pain. *Archives of Internal Medicine*, **159**, 1931–7.
66. Pande, A.C., Crockatt, J.G., Janney, C.A., *et al.* (2000). Gabapentin in bipolar disorder: a placebo-controlled trial of adjunctive therapy. *Bipolar Disorders*, **2**, 249–55.
67. Vieta, E., Manuel-Goikolea, J., Martinez-Aran, A., *et al.* (2006). A double-blind, randomized, placebo-controlled, prophylaxis study of adjunctive gabapentin for bipolar disorder. *The Journal of Clinical Psychiatry*, **67**, 473–7.
68. Pande, A.C., Pollack, M.H., Crockatt, J., *et al.* (2000). Placebo-controlled study of gabapentin treatment of panic disorder. *Journal of Clinical Psychopharmacology*, **20**(4), 467–71.
69. Pande, A.C., Davidson, J.R., Jefferson, J.W., *et al.* (1999). Treatment of social phobia with gabapentin: a placebo-controlled study. *Journal of Clinical Psychopharmacology*, **19**(4), 341–8.
70. Wagner, K.D., Kowatch, R.A., Emslie, G.J., *et al.* (2006). A double-blind, randomized, placebo-controlled trial of oxcarbazepine in the treatment of bipolar disorder in children and adolescents. *The American Journal of Psychiatry*, **163**(7), 1179–86.
71. Croissant, B., Diehl, A., Klein, O., *et al.* (2006). A pilot study of oxcarbazepine versus acamprosate in alcohol-dependent patients. *Alcoholism, Clinical and Experimental Research*, **30**(4), 630–5.
72. McElroy, S.L., Kotwal, R., Guerdjikova, A.I., *et al.* (2006). Zonisamide in the treatment of binge eating disorder with obesity: a randomized controlled trial. *The Journal of Clinical Psychiatry*, **67**(12), 1897–906.
73. McElroy, S.L., Suppes, T., Keck, P.E. Jr., *et al.* (2005). Open-label adjunctive zonisamide in the treatment of bipolar disorders: a prospective trial. *The Journal of Clinical Psychiatry*, **66**(5), 617–24.
74. Ghaemi, S.N., Zablotsky, B., Filkowski, M.M., *et al.* (2006). An open prospective study of zonisamide in acute bipolar depression. *Journal of Clinical Psychopharmacology*, **26**(4), 385–8.
75. Pollack, M.H., Roy-Byrne, P.P., Van Amerigen, M., *et al.* (2005). The selective GABA reuptake inhibitor tiagibine for the treatment of

generalized anxiety disorder: results of a placebo-controlled study. *The Journal of Clinical Psychiatry*, **66**(11), 1401–8.
76. Connor, K.M., Davidson, J.R., Weisler, R.H., *et al.* (2006). Tiagabine for posttraumatic stress disorder: effects of open-label and double-blind discontinuation treatment. *Psychopharmacology*, **184**(1), 21–5.
77. Carpenter, L.L., Schecter, J.M., Tyrka, A.R., *et al.* (2006). Open-label tiagabine monotherapy for major depressive disorder with anxiety. *The Journal of Clinical Psychiatry*, **67**(1), 66–71.
78. Schwartz, T.L., Azhar, N., Husain, J., *et al.* (2005). An open-label study of tiagabine as augmentation therapy for anxiety. *Annals of Clinical Psychiatry*, **17**(3), 167–72.
79. Suppes, T., Chisholm, K.A., Dhavale, D., *et al.* (2002). Tiagabine in treatment refractory bipolar disorder: a clinical case series. *Bipolar Disorders*, **4**(5), 283–9.
80. Zhang, W., Connor, K.M., and Davidson, J.R. (2005). Levetiracetam in social phobia: a placebo controlled pilot study. *Journal of clinical Psychopharmacology*, **19**(5), 551–3.
81. Feitner, D.E., Crockatt, J.G., Dubovsky, S.J., *et al.* (2003). A randomized, double-blind, placebo-controlled, fixed-dose, multicenter study of pregabalin in patients with generalized anxiety disorder. *Journal of Clinical Psychopharmacology*, **23**(3), 240–9.
82. Rickels, K., Pollack, M.H., Feitner, D.E., *et al.* (2005). Pregabalin for treatment of generalized anxiety disorder: a 4-week, multicenter, double-blind, placebo-controlled trial of pregabalin and alprazolam. *Archives of General Psychiatry*, **62**(9), 1022–30.
83. Montgomery, S.A., Tobias, K., Zornberg, *et al.* (2006). Efficacy and safety of pregabalin in the treatment of generalized anxiety disorder: a 6-week, multicenter, randomized, double-blind, placebo-controlled comparison of pregabalin and venlafaxine. *The Journal of Clinical Psychiatry*, **67**(5), 771–82.
84. Pande, A.C., Feitner, D.E., Jefferson, J.W., *et al.* (2004). Efficacy of the novel anxiolytic pregabalin in social anxiety disorder: a placebo-controlled, multicenter study. *Journal of Clinical Psychopharmacoly*, **24**(2), 1419.
85. United States Food and Drug Administration (January 31, 2008). *Information for Healthcare Professionals - Suicidality and Antiepileptic Drugs.*

6.2.7 Drugs for cognitive disorders

Leslie Iversen

Introduction

Cognitive disorders are among the most difficult of all nervous system illnesses to treat as they affect the most complex and least clearly understood aspects of brain function. Animal studies cannot accurately mirror the complexities of human cognition, and there are few, if any, animal models of human cognitive illnesses. As so few drugs have been found to exert clinically significant effects, animal models for testing novel cognition-enhancing agents have unknown predictive value. However, progress has been made in recent years with improved international agreement on the criteria used to approve new cognition-enhancing drugs, and the introduction of new drugs for the treatment of dementia.

Alzheimer's disease

It is important to define the objective of drug treatment in this, the most common of all forms of senile dementia. Alzheimer's disease (**AD**) is a progressive illness; drug treatment could treat the symptoms without influencing the course of the disease, or it might seek to delay or arrest the progressive cognitive deterioration which such patients suffer. Although the latter aim is the subject of intensive research in academic and industrial laboratories,[1] there are no drugs that target the underlying pathology, and only palliative treatments are, as yet, available.

The approval of new medicines for the symptomatic treatment of AD in recent years has led regulatory agencies to define more clearly what criteria should be used in assessing the clinical benefits derived from drug treatment. AD is a disease characterized by disturbances in higher cortical function, including disorders of recent memory, language function, praxis, visual perception, abstract thinking, and decision making. A variety of composite dementia assessments designed to provide an overall summary of cognitive status, for example the Mini Mental State Examination (**MMSE**), Alzheimer's Disease Assessment Scale-Cognition (**ADAS-Cog**), and the Brief Cognitive Rating Scale are used.[2] Most studies with cholinesterase inhibitors in AD have used ADAS-Cog (a 70-point scale), and a two to three-point improvement for the drug-treated group versus placebo at 6 months has generally been accepted. However, statistically significant, but small, drug-induced improvements in cognitive assessment scores do not necessarily represent a clinically significant improvement to the patient or to their doctor; they must be supplemented by evidence of clinical improvement, using some form of Clinical Global Impression of Change as an outcome measure, usually rated by a clinician on a seven-point scale.

The development of agreed scientific and clinical standards for the approval of new drugs has largely eclipsed most of the older drugs that had been used in the treatment of AD and other dementias, since none of them can meet these standards. The older drugs include a range of cerebral vasodilators (e.g. dihydroergotoxin, papaverine, isoxsuprine, cinnarizine) and the so-called 'nootropics' (e.g. piracetam, oxiracetam, aniracetam), which were widely used in some European countries, as well as the so-called 'metabolic enhancers' (e.g. idebenone and indeloxazine) which were popular for a while in Japan.

Cholinergic agents

Attention has focused instead on the cholinergic agents. The 'cholinergic hypothesis' of dementia. was boosted by the discovery in the 1970s that cholinergic neurones are particularly damaged or absent from the brains of patients dying with AD, and that the extent of damage to the cholinergic system correlates with the severity of dementia in life.[3] In AD the damage appears to be particularly severe in the system of cholinergic neurones located deep in the forebrain in the nucleus basalis of Meynert, whose fibres branch extensively and innervate most areas of the cerebral cortex. This neuronal system forms part of the ascending reticular activating system, which plays a key role in the process of selective attention—essential for the laying down of new memories. Consequently, there has been considerable interest in the possibility that 'cholinergic replacement therapy' might relieve the symptoms of AD, in the same way that dopamine replacement therapy has successfully been employed in the treatment of Parkinson's disease. The most successful approach so far has been the use of inhibitors of the enzyme acetylcholinesterase.

Inhibitors of acetylcholinesterase have been known since the nineteenth century with the discovery of physostigmine, a plant

product used as an arrow-tip poison. Irreversible organophosphate inhibitors of acetylcholinesterase were later developed as chemical warfare agents ('nerve gases'), and for more peaceful uses as insecticides. Despite their colourful past, low doses of this class of compounds have proved effective as cognitive enhancers in a wide range of animal tests, including those in which cholinergic function is deliberately impaired.[4] The first clinical trials in patients with AD were performed with physostigmine, and confirmed that the drug had significant beneficial effects on cognitive performance in AD patients.[5] However, it has limited usefulness because, although it is absorbed rapidly, it has only a very short half-life in plasma. This means that to obtain any sustained cognitive benefit it has to be given in doses that are sufficiently high to elicit a number of adverse side-effects; thus, the therapeutic window was very narrow.

Subsequently four other cholinesterase inhibitors with improved profiles have gained approval for use in AD: **tacrine**, **donepezil, rivastigmine, and galantamine**, but tacrine is no longer actively marketed owing to liver toxicity. Clinical data from several thousand patients with AD involved in trials with these cholinesterase inhibitors are now available.[6–8] The first of these to gain approval in 1997 was donepezil. Results of large-scale clinical trials with donezepil and the other cholinesterase inhibitors over periods of 15 and 24 weeks have yielded similar results for the three compounds in patients with mild to moderately severe AD. The drugs caused small but significant improvements in the ADAS-Cog, CIBIC, and MMSE scores. The most common side-effects were transient mild nausea, insomnia, and diarrhoea. Not all patients with AD will benefit from treatment with cholinesterase inhibitors; the proportion ranges from 30 to 50 per cent; although the clinical benefits of drug treatment in patients showing a response can persist for up to 24 months.

The approval of cholinesterase inhibitors for the treatment of Alzheimer's disease was an important landmark. They are reasonably well tolerated and produce significant, if modest, beneficial effects in patients with mild to moderately severe AD. However, they have not gained immediate and universal acceptance. In some countries (e.g. the United Kingdom) it has been argued that the drugs are too costly and provide at best only a modest improvement.

Some studies have also found the cholinesterase inhibitors to be effective in treating the cognitive deficits in vascular dementias, but the effects are small and less consistent. Rivastigmine has also been shown to have beneficial effects in treating the cognitive deficits in Parkinson's disease with dementia.[9] The cholinesterase inhibitors may thus find other applications as cognitive enhancers in conditions other than AD.

An alternative approach to cholinergic replacement therapy has been to develop **drugs that mimic acetylcholine** and act as agonists at the muscarinic cholinergic receptors in brain, but which, unlike acetylcholine itself, are bioavailable and brain-penetrant. Attention has focused on the discovery and development of muscarinic agonists that show selectivity for the m_1-receptor subtype, which is the predominant form present in the cerebral cortex. The most thoroughly studied cholinomimetic to date is xanomeline, a compound that acts as a highly potent and selective m_1-receptor agonist. Clinical effects were assessed in a multicentre study of 343 patients with AD.[10] Patients on the highest dose showed significant improvement when assessed using the ADAS-Cog scale and also showed a significant overall global improvement using CIBIC. In addition to cognitive improvements, patients receiving xanomeline also exhibited significant behavioural improvement, with dose-dependent reductions in vocal outbursts, suspiciousness, delusions, agitation, and hallucinations. Xanomeline is unlikely to be used in the treatment of AD because of its relatively short duration of action, but these results suggest that further research on cholinomimetics may still be justified.

An entirely different pharmacological approach is exemplified by the drug **memantine**, the first to be approved for the treatment of moderate to severe AD. Memantine is thought to act by virtue of its ability to block the NMDA sub-type of glutamate receptors in the brain. Clinical trials in AD showed small but significant beneficial effects on cognitive tests and in global clinical outcome, but curiously the effects were most notable in patients with advanced stage disease and less in patients with mild to moderate AD.[11] Although the effects of memantine are small, it remains the only effective treatment for advanced stage AD.

Attention-deficit hyperactivity disorder

Attention-deficit hyperactivity disorder (**ADHD**) is one of the most thoroughly studied disorders in child psychiatry, and the increasingly common use of stimulant drugs to treat this disorder has become the focus of much public attention and debate in recent years.[12] ADHD is defined in terms of three key features: lack of sustained attention, impulsivity, and hyperactivity. According to the DSM-IV definition the diagnosis of ADHD now includes more than 10 per cent of children.[13]

Because of the interest in the drug treatment of ADHD, a number of assessment tools have been developed. These include the widely used Conners' Teacher Rating Scale, the Conners' Parent Rating Scale, and a variety of tests designed to measure hyperactivity, problem behaviour, attention, and other aspects of cognition, as well as academic performance.[14]

The most commonly used drugs are the psychostimulants **amfetamine** and **methylphenidate**. Methylphenidate (Ritalin®) is by far the most widely prescribed. In more than 100 published trials these drugs have been found to have significant beneficial effects on all three key symptoms of ADHD in approximately 70 per cent of the treated children, and also in an adult form of ADHD.[15] Amfetamine (Adderall®) is approved for treatment of adult ADHD (not licensed in UK).

The mechanism of action of all three agents is similar; they act principally as inhibitors of the dopamine-uptake mechanism in the brain and promote the release of this neurotransmitter, thus stimulating dopaminergic mechanisms. The drugs also act to an important extent on noradrenaline-containing neurones to promote an increased release of this monoamine.[16] This may be relevant;a selective inhibitor of the noradrenaline transporter in brain, **atomoxetine**, has been approved as the first non-scheduled stimulant for the treatment of ADHD in both children and adults.[17]

Another non-amfetamine **modafinil** is also non-scheduled and is widely used in the United States for the treatment of ADHD,[18] although the drug has not yet gained regulatory approval from FDA because of possible serious adverse skin reactions. The mode of action of modafinil is unknown but it is used for the treatment of narcolepsy.

It is paradoxical that stimulant drugs, whose actions include an ability to promote hyperactivity, should have a calming effect

on hyperactive children. One explanation is that actions of these amfetamine derivatives show the 'rate dependency' typical of other central nervous system agents, i.e. they tend to stimulate low rates of behaviour and to suppress high rates.[19] An alternative view is that the relatively low doses of amphetamines used in the treatment of ADHD would not have stimulant effects even in normal healthy adults. There are few animal models that can be used in the study of psychostimulant use or ADHD. Mice that are genetically engineered to delete the genes for the dopamine and other monoamine transporters have proved valuable.[20] Animals which lack the dopamine transporter have elevated levels of dopamine in their brains and are behaviourally hyperactive. Paradoxically, d-amfetamine decreases activity in these animals, in keeping with the 'rate-dependency' hypothesis.[20]

The use of amfetamines, particularly methylphenidate, has increased rapidly during the past 30 years, particularly in the United States where in some states more than 10 per cent of school-age boys receive the drug.[12] The use of amfetamines in Europe has been at a much lower level so far, although their use in ADHD has also been increasingly rapidly.[12] In turn, such widespread use of psychostimulants creates problems about diversion and abuse.[12]

Further information

'Cognition Enhancers' in Drugs Futures 2025—[www.foresight.gov.uk/Brain_Science_Addiction].

'Drug Treatment of Alzheimer's Disease', Royal College of Psychiatry—[www.rcpsych.ac.uk/mentalhealthinformation/olderpeople/drugtreatmentofalzheimers.aspk].

'Methylphenidate, atomoxetine and dexamfetamine for attention deficit hyperactivity disorder (ADHD) in children and adolescents—guidelines.' National Institute for Clinical Excellence UK, March 22, 2006. [http://www.nice.org.uk].

References

1. Moreira, P.I., Zhu, X., Nunomura, A., *et al.* (2006). Therapeutic options in Alzheimer's disease. *Expert Review of Neurotherapeutics*, **6**, 897–910.
2. Gershon, S., Ferris, S.H., Kennedy, J.S., *et al.* (1994). Methods for the evaluation of pharmacological agents in the treatment of cognitive and other deficits in dementia. In *Clinical evaluation of psychotropic drugs: principles and guidelines* (eds. R.F. Prien and D.S. Robinson), pp. 467–99. Raven Press, New York.
3. Bartus, R.T., Dean, R.L., Beer, B., *et al.* (1982). The cholinergic hypothesis of geriatric memory dysfunction. *Science*, **217**, 408–17.
4. Rupniak, N.M.J., Steventon, M.J., Jennings, C.A., *et al.* (1989). Comparison of the effects of four cholinomimetic agents on cognition in primates following disruption by scopolamine. *Psychopharmacology*, **99**, 189–95.
5. Giacobini, E. (1998). Cholinesterase inhibitors for Alzheimer's disease therapy: from tacrine to future applications. *Neurochemistry International*, **32**, 413–19.
6. Birks, J. and Harvey, R.J. (2006). Donepezil for dementia due to Alzheimer's disease. *Cochrane Database Systematic Reviews*, (1), CD001190.
7. Loy, C. and Schneider, L. (2004). Galantamine for Alzheimer's disease. *Cochrane Database Systematic Reviews*, (4), CD001747.
8. Birks, J., Grimley Evans, J., Lakovidou, V., *et al.* (2006). Rivastigmine for Alzheimer's disease. *Cochrane Database Systematic Reviews*, (4), CD001191.
9. Maidment, I., Fox, C., and Boustani, M. (2006). Cholinesterase inhibitors for Parkinson's disease dementia. *Cochrane Database Systematic Reviews*, (1), CD004747.
10. Bodick, N.C., Offen, W.W., Levey, A.I., *et al.* (1997). Effects of xanomeline, a selective muscarinic receptor agonist, on cognitive function and behavioral symptom in Alzheimer's disease. *Archives of Neurology*, **54**, 465–73.
11. McShane, R., Areosa Sastre, A., and Minakaran, N. (2006). Memantine for dementia. *Cochrane Database Systematic Reviews*, (2), CD003154.
12. Iversen, L.L. (2006). *Speed, ecstasy, ritalin: the science of amphetamines*. Oxford University Press, Oxford.
13. American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th edn, Text Revision) (DSM-IV-TR). American Psychiatric Association, Washington, DC.
14. Conners, C.K. (1998). Rating scales in attention deficit hyperactivity disorder: use in assessment and treatment monitoring. *The Journal of Clinical Psychiatry*, **59**(Suppl. 7), 24–30.
15. Wender, P.H. (2001). *ADHD: attention-deficit hyperactivity disorder in children, adolescents and adults*. Oxford University Press, Oxford.
16. Solanto, M.V. (1998). Neuropsychopharmacological mechanism of stimulant drug action in attention-deficit hyperactivity disorder: a review and integration. *Behavioural Brain Research*, **94**, 127–52.
17. Kratochvil, C.J., Vaughan, B.S., Daughton, J.M., *et al.* (2004). Atomoxetine in the treatment of attention deficit hyperactivity disorder. *Expert Review of Neurotherapeutics*, **4**, 601–11.
18. White, R.F. and Giorgadze, A. (2006). A randomized, double-blind, placebo-controlled study of modafinil film-coated tablets in children and adolescents with attention-deficit hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, **45**, 503–11.
19. Robbins, T.W. and Sahakian, B.J. (1979). Paradoxical effects of psychomotor stimulant drugs in hyperactive children from the standpoint of behavioural pharmacology. *Neuropharmacology*, **18**, 931–50.
20. Gainetdinov, R.R., Sotnikova, T.D., and Caron, M.G. (2002). Monoamine transporter pharmacology and mutant mice. *Trends in Pharmacological Sciences*, **23**, 367–73.

6.2.8 Drugs used in the treatment of the addictions

Fergus D. Law and David J. Nutt

Medical treatment of the addictions remains controversial, with addiction itself viewed as a lifestyle problem, a hijacking of brain systems by drugs, or as a medical illness. Many of these controversies may be avoided by taking a goal-oriented approach to treatment, in which clinical objectives are defined, and both medications and psychological interventions are used to facilitate progress towards these. The effectiveness of medications is maximized when they are used as one component of a comprehensive treatment plan.

There are no 'magic bullets' in addiction treatment—the same pharmacological principles apply to these drug treatments as to any other. Drugs need to be given in effective doses, at appropriate intervals, allowed time to reach steady state, and also to dissipate when terminated on the basis of their half-life. Some drugs also have an abuse potential of their own (e.g. opiates, sedative-hypnotics) especially those with a rapid onset of action, and such

drugs need to be particularly closely monitored and controlled, to minimize their diversion and misuse.

Medications in perspective

The clinical goal-oriented approach requires clarity about the clinical objectives at each phase of the treatment process. A typical treatment plan involves three primary clinical objectives:

- drug and psychosocial stabilization
- detoxification when appropriate
- prevention of relapse or recurrence.

Stabilization with a substitution treatment (e.g. methadone or buprenorphine in opiate addiction) involves prescribing a pharmacological equivalent to the abused drug to stop illicit use, crime, etc. It allows time for stabilization to occur and to consider later objectives. Stabilization itself may be either short-term with the primary goal of terminating illicit drug use 'on top' of the prescription, or longer-term where it is commonly known as maintenance. The goals of maintenance are either psychosocial stabilization in preparation for detoxification, or harm reduction in patients where abstinence is not practicable or safe. Unless a sufficient degree of stabilization has been achieved prior to detoxification, the chances of success are strictly limited. The harm-reduction goal has generated much controversy, but one of its major benefits is the reduction in the spread of HIV among injecting drug users. In this group stable long-term maintenance treatment is preferable to repeated cycles of premature discontinuation followed by relapse to uncontrolled drug use with its attendant elevated risk of death from overdose as tolerance wanes. Monitoring by using drug screens and clinical assessments is required to ensure that patients do not use 'on top' of their prescription.

Specific drugs used in addiction treatment

This chapter deals with methadone, levacetylmethadol (LAAM), codeine and dihydrocodeine tartrate, buprenorphine, clonidine, lofexidine, naltrexone, naloxone, acamprosate, disulfiram, and clomethiazole (chlormethiazole), and covers addiction indications only. Many of these drugs are not licensed for use in addiction treatment (e.g. methadone tablets and injection), or are currently licensed in only one or a few countries (e.g. clonidine in Germany, lofexidine in the United Kingdom).

Methadone[1,2] (Methadose®, Physeptone®, Synastone®)

This is a long-acting opioid analgesic which has been the mainstay of opioid substitution treatment, but is often difficult to stop due to its prolonged withdrawal syndrome.

Pharmacology: it is a strong full μ-opioid agonist.

Types of compounds available: it is available in liquid, injectable, and tablet formulations.

Pharmacokinetics: t_{max} 2 to 4 h after oral dosing; 1 h after intramuscular injection; its half-life is 25 h.

Side-effects: as with other μ-opioids, its side-effects include mental blunting, sweating, constipation, nausea, and analgesia.

Toxic effects: acute overdose leads to respiratory depression and pulmonary oedema.

Indications: opiate maintenance, stabilization, and detoxification; it is also indicated for use during pregnancy.

Contraindications: respiratory or severe liver disease, monoamine oxidase inhibitors; caution should be exercised in elderly people.

Interactions: respiratory depression especially in combination with other sedative drugs; metabolism affected by hepatic enzyme induction and inhibition; plasma levels are affected by HAART drugs.

Effects of withdrawal: these include moderate but prolonged abstinence syndrome, especially poor sleep.

Dosage and administration: single daily dose, occasionally twice daily; dose depends on the level of dependence—if unknown, initially 10 to 20 mg daily. The minimum dose that covers withdrawal symptoms for 24 h should be given, and increased by 5 to 10 mg as necessary. Close monitoring is necessary by clinical assessment and drug screen. Some centres use intravenous preparations in those who don't respond to oral preparations.

LAAM[3] (ORLAAM®)

LAAM is a methadone variant with a much longer half-life, requiring only three visits a week for full supervision of medication. However, take-home medication is not allowed and its use is restricted to specialist clinics and is reserved for patients who have failed other treatments. Its licence in Europe has been withdrawn due to QTc prolongation.

Pharmacology: it is a synthetic μ-opioid agonist with active metabolites which are more potent than the parent drug.

Types of compounds available: aqueous solution.

Pharmacokinetics: t_{max} 2 to 4 h; duration of action is 48 to 72 h; half-lives for LAAM and its metabolites are 2 to 4 days; it takes 2 weeks to reach steady state.

Side-effects: too rapid escalation of the dose may result in sedation, orthostatic hypotension, poor concentration, and overdose.

Toxic effects: as for methadone; QTc prolongation; overdose occurs with too frequent (daily) dosing, use of multiple drugs, or 'on-top' use due to impatience with its slow onset of action.

Indications: opiate maintenance, stabilization, and detoxification.

Contraindications: pregnancy (transfer to methadone); QTc prolongation prior to induction of treatment; dose should be reduced in elderly people, and in renal and hepatic impairment.

Interactions: as for methadone; other drugs prolonging QTc interval.

Effects of withdrawal: as for methadone, but with a milder withdrawal syndrome due to longer $t_{1/2}$.

Dosage and administration: pre-treatment ECG to identify prolonged QTc intervals, repeated 12–14 days after initiating treatment and periodically thereafter to rule out alterations to the QTc; give three times a week or on alternate days; increase dose by 20 to 40 per cent when transferring form 48 h to 72 h dosing interval. Transfer methadone to LAAM by giving 1.2 to 1.3 times the daily methadone dose; and LAAM to methadone by waiting at least 48 h and then giving 0.8 times the LAAM dose. If low or unknown tolerance, the initial dose is 20 to 40 mg three times weekly. Adjust dose in 5- to 10-mg steps, but no more frequently than

weekly at the most. Strongly warn patients of the risk of supplementation with street drugs especially prior to steady state. LAAM is detected by urine screens for methadone.

Buprenorphine[4,5] (Subutex®, Suboxone®, Temgesic®, Buprenex®)

Advantages over methadone and other full μ-agonists are its safety in overdose, the attenuation of the drug 'high' during on-top use, and its low levels of psychological reinforcement and withdrawal symptomatology during detoxification.

Pharmacology: a partial μ-opioid agonist, which explains the ceiling on respiratory depression; slow onset of action; dissociates slowly from the μ-receptor.

Types of compounds available: 0.2-, 0.4-, 2- or 8-mg sublingual tablets, or 0.3-mg ampoules for injection; it is also available in combination with naloxone in a 1:4 ratio (Suboxone®) to reduce misuse if diverted.

Pharmacokinetics: sublingual tablets absorbed rapidly into the buccal mucosa and released slowly into the blood stream; t_{max} 2 to 6 h.

Side-effects: withdrawal symptoms if either too little or too much is given; nausea and vomiting are rare in addicts.

Toxic effects: as for methadone, but less constipation and respiratory depression.

Indications: opiate maintenance, stabilization, and detoxification, including in pregnancy; may be especially suitable for opioid antagonist-assisted withdrawal; no dosage adjustment needed in renal failure or elderly people.

Contraindications: severe respiratory disease; use with care in severe liver disease.

Interactions: rare; sedation with benzodiazepines.

Effects of withdrawal: there is a mild but delayed withdrawal syndrome.

Dosage and administration: initial dose is 0.8–4 mg increasing by 4 to 8 mg daily until the required dose level is reached; usual daily dose 8 to 32 mg; doses above 12 mg may be given on alternate days; minimize withdrawal symptoms after long-term use by reducing by 1 mg every 3 to 4 days or less often; buprenorphine-assisted heroin detoxification by rapid reduction over 5 to 10 days; the injectable form is not recommended for use in addiction treatment. Monitor clinical state and perform drug screens for compliance and on-top use.

Codeine phosphate and dihydrocodeine tartrate[6,7] (DF118 Forte®, DHC Continus®)

Advantages over methadone occur in situations where long-acting opioids may be inappropriate. These are often preferred by patients, and by doctors treating the young, low-dose users, and in acute situations (e.g. in police custody). Disadvantages are its ease of misuse, high levels of psychological reinforcement due to its rapid onset of action, and its unfavourable side-effect profile.

Pharmacology: it is a weak short-acting μ-opioid agonist.

Types of compounds available: Codeine: 15-, 30- and 60-mg oral tablets, linctus, syrup and injection; dihydrocodeine tartrate: 30- or 40-mg oral tablets for use three to six times a day, and a 60-, 90-, or 120-mg slow-release preparation (DHC Continus) for use every 12 h; also parenteral preparation and elixir.

Pharmacokinetics: peak plasma levels at 1 to 2 h; half-life 3.5 to 4.5 h.

Side-effects: it is more likely to cause sedation, dizziness, stimulation, euphoria, constipation, histamine release, psychomimetic effects, and disturbing dreams than other opioids.

Toxic effects: precipitation of life-threatening exacerbations of asthma; in overdose, coma with myotonic twitching, grand mal convulsions, and rarely rhabdomyolysis may occur.

Indications: opiate maintenance, stabilization, and detoxification

Contraindications: acute exacerbations of asthma, lower respiratory tract infection, respiratory depression, and hepatic failure, increased intracranial pressure; caution should be exercised in renal impairment and elderly people.

Interactions: as for methadone; it may enhance the effects of warfarin.

Effects of withdrawal: mild withdrawal syndrome.

Dosage and administration: dose depends on level of dependence; monitor clinical state and urine screen to confirm compliance and termination of on-top use.

Naltrexone[8–11] (Nalorex®, Opizone®)

Naltrexone is used to maintain abstinence in detoxified opiate addicts during the period of highest vulnerability to relapse following detoxification. It blocks the 'high' produced by opiates and promotes the extinction of conditioned responses. It also has a role in alcohol misuse and ultra-rapid opioid detoxification. Nalmefene is a related long-acting μ receptor antagonist that has a licence for alcoholism in some countries.

Pharmacology: it is a long-acting non-selective opioid antagonist.

Type of compound available: 50-mg oral tablet.

Pharmacokinetics: t_{max} 1 h; duration of action is dose related, and a single dose can be effective for up to 48 h.

Side-effects: opiate withdrawal syndrome may occur on induction; occasionally, gastrointestinal irritation, headaches, arthralgia, flattening of mood, and rash occur.

Toxic effects: severe opioid withdrawal in dependent addicts lasting 2 days; reversible liver toxicity at high dose in obese and elderly people; liver function tests should be monitored especially if baseline tests are impaired.

Indications: prevention of impulsive relapse following detoxification in opioid users; opioid antagonist-assisted withdrawal; it reduces the reinforcing effects of alcohol.

Contraindications: acute hepatitis or liver failure, and active peptic ulcer; caution should be exercised in hepatic or renal impairment.

Interactions: competitive opioid blockade, so potentially can be overcome using very high opiate doses.

Effects of withdrawal: none, but risk of opioid overdoes following withdrawal (loss of toerance).

Dosage and administration: treatment initiated following LFTs and opioid-negative urine screen (or a negative naloxone challenge). Twenty-five mg is given on the first day, and then 50 mg daily for 3 to 6 months. Thrice weekly dosing (100/100/150 mg) may occasionally improve compliance. Supervision of consumption by a supportive person, urine tests to monitor compliance, and regular reviews are very important in maximizing effectiveness.

Naloxone[12] (Narcan®)

Naloxone is a short-acting antagonist used in the treatment of opioid overdose, during detoxification, and naltrexone induction. Take home naloxone may be used to treat opiate overdoses in the community by patients trained in its use.

Pharmacology: it is a short-acting competitive opioid antagonist.

Types of compounds available: it is available in injectable form for intramuscular, intravenous, or subcutaneous use.

Pharmacokinetics: half-life 1 to 2 h.

Side-effects: withdrawal in opiate-dependent subjects; nausea and vomiting; rarely, high blood pressure and pulmonary oedema can occur.

Toxic effects: very occasional deaths due to acute pulmonary oedema, extreme hypertension, and ventricular arrhythmias have occurred in those with known myocardial disease.

Indications: naloxone reverses the effects of opioid overdose, and in high doses may help in overdose due to alcohol and benzodiazepines. Naloxone is also occasionally used as the diagnostic test of opioid dependence, and as a challenge test prior to naltrexone initiation. It is also used for opioid antagonist-assisted withdrawal, and in combination with oral opiate agonists to reduce misuse by the injectable route (e.g. Suboxone®). In neonates it is used for the reversal of the effects of opioids given to mothers during labour.

Contraindications: none if not opioid dependent (safe in neonates, children, pregnancy, elderly people); caution is advisable in opioid dependence, painful conditions, and cardiovascular disease.

Interactions: severe hypertension following reversal of coma due to clonidine overdose.

Effects of withdrawal: none.

Dosage and administration: in opioid overdose, give 0.4 to 2 mg intravenously (5 to 10 μg/kg in neonates and children) and repeat at 2- to 3-min intervals until desired response; may also be given intramuscularly in overdose; lower doses are used in adults (0.1–0.2 mg) to reverse opioid-induced respiratory depression, but higher doses are needed with buprenorphine and σ-receptor agonists. Continue naloxone by infusion or repeated injection if necessary to maintain recovery. During naltrexone induction, give 0.2 mg parenterally followed by 0.8 mg 30 min later (or 0.6 mg 30 s later if given intravenously). In equivocal cases give 1.6 mg.

Clonidine hydrochloride[13, 14] (Catapres®, Dixarit®)

Clonidine is an α_2-adrenoceptor agonist used to suppress some symptoms of opioid withdrawal, especially methadone-assisted withdrawal. It is ineffective for subjective symptoms, muscle/bone aches, stomach cramps, and insomnia.

Pharmacology: clonidine is an antihypertensive agent that decreases central and peripheral (sympathetic) noradrenergic activity by stimulating presynaptic receptors in the locus coeruleus.

Types of compounds available: it is available as tablet, liquid, sustained release capsule, and transdermal preparation; the tablet is licensed in Germany.

Pharmacokinetics: t_{max} 90 min; half-life of 20 to 25 h

Side-effects: it causes hypotension, sedation, dry mucous membranes, bradycardia, depression, impotence, constipation and diarrhoea, sleep disturbance, fluid retention, headache, euphoria, and Raynaud's phenomenon.

Toxic effects: a clonidine withdrawal syndrome may occur on abrupt withdrawal or non-compliance; paralytic ileus, psychotic features, or depression can also occur; coma or severe sedation can occur on acute overdose.

Indications: rapid opiate withdrawal and opiate antagonist-assisted withdrawal; it is also used as adjunct for alcohol, benzodiazepine, and nicotine withdrawal.

Contraindications: low baseline blood pressure, disorders of cardiac pacemaker activity and conduction, cardiovascular and cerebrovascular disease, and porphyria; caution is necessary in renal and hepatic impairment, peripheral vascular disease, and where there is a history of depression or psychosis.

Interactions: combinations with sedative drugs; phenothiazines may increase hypotension. Tricyclic antidepressants may block its effects.

Effects of withdrawal: an increase in sympathetic activity with symptoms mimicking the opiate withdrawal syndrome may occur 18 to 72 h after the last dose on abrupt termination. Blood pressure rebound is rare when used for less than 1 month. The effects are minimized by gradual withdrawal.

Dosage and administration: expertise is needed to monitor cardiovascular signs and adjust dose during a clonidine detoxification over 1 to 3 weeks. Start clonidine after discontinuation of the opioid. Following a test dose, 0.1 mg tablets are given four to six times daily building up to 2 mg daily over a few days in inpatients but half this dose in outpatients. Patches applied once weekly and supplemented by tablets if withdrawal symptoms occur. Frequent monitoring for hypotension and bradycardia is needed.

Lofexidine[13–15] (BritLofex®)

Lofexidine is an analogue of clonidine, but easier to use because there is less hypotension and sedation.

Pharmacology: as for clonidine; differences occur possibly because it is more potent at the A subtype of α_2-adrenoceptors.

Types of compounds available: 0.2-mg oral tablets; these are licensed in the United Kingdom only.

Pharmacokinetics: t_{max} 3 h; half-life of 15 h.

Side-effects: these are the same as for clonidine, but with markedly less hypotension and other side-effects.

Toxic effects: as for clonidine; it has little rebound effect on blood pressure, and no psychiatric complications or misuse has been reported.

Indications and contraindications: as for clonidine.

Interactions: as for clonidine.

Effects of withdrawal: as for clonidine, but less rebound.

Dosage and administration: treat for 1 to 3 weeks; dose may be started before opiate is stopped; on day 1, give 0.8 mg in divided doses, and build up by 0.4 to 0.8 mg daily. Aim for a minimum of 1.6 mg daily in four divided doses, increasing to a maximum daily dose of 2.4 mg. Plan for a peak in the dose when the peak of withdrawal symptoms are expected. Blood pressure should be monitored 2 h after the initial dose, and daily as the dose is increasing. After peak opiate withdrawal lofexidine should be withdrawn gradually over at least 2–4 days. The dose should be reduced by 0.2–0.4 mg daily.

Acamprosate (calcium acetylhomotaurinate)[16,17] (Campral EC®)

Acamprosate is a non-aversive agent used to maintain abstinence in alcohol-dependent patients during the most vulnerable period following detoxification, which may work as an anticraving agent.

Pharmacology: this is not fully understood. It affects the brain's γ-aminobutyric acid (inhibitory) and glutamate (excitatory) systems.

Types of compounds available: oral enteric-coated tablets containing 333 mg acamprosate.

Pharmacokinetics: t_{max} 5 h; half-life 21 h; steady state after 7 days.

Side-effects: there are a range of dose-related side-effects which are mainly mild and transient, including diarrhoea and other gastrointestinal effects, pins and needles in the limbs, skin pruritus, confusion, and sexual effects. Transient reductions in blood pressure occur in those with alcohol-induced hepatic cirrhosis.

Toxic effects: hypercalcaemia is a theoretical possibility following overdose.

Indications: maintenance of abstinence following alcohol detoxification; it is suitable for use in those with liver dysfunction.

Contraindications: renal impairment and severe hepatic failure.

Interactions: concomitant food decreases bioavailability.

Effects of withdrawal: none.

Dosage and administration: begin treatment as soon as possible after detoxification, and continue for 1 year. Four tablets a day (2–1–1) with meals if body weight is less than 60 kg, but six tablets a day (2–2–2) if over 60 kg. The drug should be continued during alcohol relapses.

Disulfiram[18–20] (Antabuse®)

An unpleasant disulfiram-ethanol reaction occurs when alcohol is consumed, acting as a deterrent or punishment if drinking occurs. Disulfiram is used under specialist supervision during periods of vulnerability to relapse.

Pharmacology: disulfiram is an aldehyde dehydrogenase inhibitor leading to the accumulation of acetaldehyde after ethanol consumption.

Types of compounds available: 200-mg oral tablets.

Pharmacokinetics: inhibition of alcohol dehydrogenase develops slowly over 12 to 24 h and peaks at 48 h.

Side-effects: relatively non-toxic on its own, but may cause drowsiness, fatigue, halitosis, nausea, vomiting, and a decrease in libido. With alcohol, disulfiram causes nausea, vertigo, anxiety, blurred vision, hypotension, chest pain, palpitations, tachycardia, facial flushing, and throbbing headache. Symptoms can last 3 to 4 days, but may persist for 1 week. Symptoms may occur even with small amounts of alcohol, but 25 to 50 per cent of patients experience little or no reaction at standard doses.

Toxic effects: the disulfiram-ethanol reaction may be very severe with respiratory depression, cardiovascular collapse, cardiac arrhythmias, coma, cerebral oedema, hemiplegia, convulsions, and death. Chronic treatment and overdose may cause high blood pressure, hepatotoxicity, and neuropsychiatric complications.

Indications: it is used as a deterrent to the use of alcohol and maintenance of abstinence, especially if there is high motivation and good compliance in the patient.

Contraindications: cardiac failure, cardiovascular or cerebrovascular disease, hypertension, peripheral neuropathy, psychosis, severe personality disorder, suicide risk, pregnancy, and breast feeding. It should be used with caution in hypertension, diabetes mellitus, epilepsy, impaired hepatic or renal function, respiratory disorders, cerebral damage, and hypothyroidism, as it may exacerbate these conditions. Caution should be exercised in the elderly.

Interactions: caution with phenytoin, diazepam, chlordiazepoxide, theophylline, warfarin, and caffeine metabolism; acute psychosis or confusional state may occur with metronidazole. Concurrent tricyclic antidepressants may exacerbate the disulfiram-ethanol reaction and cause a toxic confusional state.

Effects of withdrawal: none but restoration of alcohol dehydrogenase depends on *de novo* enzyme synthesis which occurs over 6 or more days.

Dosage and administration: after 24 h without alcohol, give 800 mg as a single dose on day 1, then reduce dose over 5 days from 100 to 200 mg daily. Effectiveness is dose related. Blood pressure should be monitored regularly if the patient is taking over 500 mg/day. Compliance is improved with monitoring (carbon disulphide breath test) and supervision. Patients should be warned not to ingest any alcohol, including alcohol in food, liquid medicines, and even toiletries. An alcohol challenge test may be done in specialist centres. The patient should be reviewed every 6 months at a minimum. Alcohol should be avoided for at least 1 week on terminating disulfiram.

Clomethiazole (edisylate)[21] (Heminevrin®)

Clomethiazole (previously known as chlormethiazole) is a sedative-hypnotic-anxiolytic which also inhibits the metabolism of alcohol resulting in a more gradual elimination of alcohol from the body.

Pharmacology: it is an agonist at the picrotoxin/barbiturate site of the GABA-A receptor, a glutamate antagonist, and an inhibitor of alcohol dehydrogenase.

Types of compounds available: oral and parenteral forms.

Pharmacokinetics: t_{max} 1 h (oral dosing); plasma half-life is 4 h but is double this in elderly people.

Side-effects: conjunctival irritation, nasal congestion, tingling in the nose, headaches, and reversible elevation of liver function tests.

Toxic effects: respiratory depression, sudden fall in blood pressure, anaphylactic reactions, and death (often due to combination with alcohol).

Indications: acute alcohol withdrawal and delirium tremens in inpatients.

Contraindications: alcohol addicts who continue to drink, pulmonary insufficiency, pregnancy, and lactation; caution is advised in renal impairment, severe liver damage, and cardiac and respiratory disease.

Interactions: alcohol and diazoxide may cause severe respiratory depression. Plasma levels are increased by cimetidine. It causes severe bradycardia with propranolol.

Effects of withdrawal: rebound insomnia and anxiety (as with other sedative drugs).

Dosage and administration: titrate using three or four daily doses according to patient response. Initially give 2 to 4 capsules, then 9 to 12 in divided doses over the next 24 h, 6 to 8 capsules on

day 2, 4 to 6 on day 3, reducing it to 0 by day 6 to 9 in order to avoid dependency. Only use it by infusion where resuscitation facilities are available; initially give 3 to 7.5 ml/min, then reduce dosage to 0.5 to 1 ml/min (infusion no longer available in UK).

Further information

British National Formulary (BNF) www.bnf.org.
NIDA (National Institute on Drug Abuse) www.nida.nih.gov.
SAMHSA (Substance Abuse and mental health services administration). www.samhsa.gov

References

1. Lingford-Hughes, A.R., Welch, S., and Nutt, D.J. (2004). Evidence-based guidelines for the pharmacological management of substance misuse, addiction and comorbidity: recommendations from the British Association for Psychopharmacology. *Journal of Psychopharmacology*, **18**, 293–335.
2. NICE. (2007). *Methadone and buprenorphine for the management of opioid dependence. NICE technology appraisal guidance 114*. National Institute for Health and Clinical Excellence, London.
3. Clark, N., Lintzeris, N., Gijsbers, A., *et al.* (2002). LAAM maintenance vs methadone maintenance for heroin dependence. *The Cochrane Database of Systematic Reviews*, (2): CD002210.
4. Law, F.D., Myles, J.S., Daglish, M.R.C., *et al.* (2004). The clinical use of buprenorphine in opiate addiction: evidence and practice. *Acta Neuropsychiatrica*, **16**, 246–74 (with erratum at 2004, **16**, 326).
5. Nutt, D.J. (1997). Receptor pharmacology of buprenorphine. *Research and Clinical Forums*, **19**, 9–15.
6. Krausz, M., Verthein, U., Degwitz, P., *et al.* (1998). Maintenance treatment of opiate addicts in Germany with medications containing codeine—results of a follow up study. *Addiction*, **93**, 1161–7.
7. Robertson, J.R., Raab, G.M., Bruce, M., *et al.* (2006). Addressing the efficacy of dihydrocodeine versus methadone as an alternative maintenance treatment for opiate dependence: a randomized controlled trial. *Addiction*, **101**, 1752–9.
8. Gonzalez, J.P. and Brogden, R.N. (1988). Naltrexone. A review of its pharmacodynamic and pharmacokinetic properties and therapeutic efficacy in the management of opioid dependence. *Drugs*, **35**, 192–213.
9. NICE. (2007). *Naltrexone for the management of opioid dependence. NICE technology appraisal guidance 115*. National Institute for Health and Clinical Excellence, London.
10. Srisurapanont, M. and Jarusuraisin, N. (2005). Opioid antagonists for alcohol dependence. *The Cochrane Database of Systematic Reviews*, (1): CD001867.
11. Minozzi, S., Amato, L., Vecchi, S., *et al.* (2006). Oral naltrexone maintenance treatment for opioid dependence. *The Cochrane Database of Systematic Reviews*, (1): CD001333.
12. Handal, K.A., Schauben, J.L., and Salamone F.R. (1983). Naloxone. *Annals of Emergency Medicine*, **12**, 438–45.
13. Gowing, L., Farrell, M., Ali, R., *et al.* (2004). Alpha2 adrenergic agonists for the management of opioid withdrawal. *The Cochrane Database of Systematic Reviews*, (4): CD002024.
14. Kahn, A., Mumford, J.P., Rogers, GA., *et al.* (1997). Double-blind study of lofexidine and clonidine in the detoxification of opiate addicts in hospital. *Drug and Alcohol Dependence*, **44**, 57–61.
15. Strang, J., Bearn, J., and Gossop, M. (1999). Lofexidine for opiate detoxification: a review of recent randomised and open controlled trials. *American Journal on Addictions*, **8**, 337–48.
16. Littleton, J. (1995). Acamprosate in alcohol dependence: how does it work? *Addiction*, **90**, 1179–88.
17. Wilde, M.I. and Wagstaff, A.J. (1997). Acamprosate: a review of its pharmacology and clinical potential in the management of alcohol dependence after detoxification. *Drugs*, **53**, 1038–53.
18. Banys, P. (1988). The clinical use of disulfiram (Antabuse): a review. *Journal of Psychoactive Drugs*, **20**, 243–60.
19. Heather, N. (1993). Disulfiram treatment for alcohol problems: is it effective and, if so, why? In *Treatment options in addiction: medical management of alcohol and opiate abuse* (ed. C. Brewer), pp. 1–18. Gaskell, London.
20. Brewer, C. (1984). How effective is the standard dose of disulfiram? A review of the alcohol-disulfiram reaction in practice. *The British Journal of Psychiatry*, **144**, 200–2.
21. Majumdar, S.K. (1990). Chlormethiazole: current status in the treatment of the acute ethanol withdrawal syndrome. *Drug and Alcohol Dependence*, **27**, 201–7.

6.2.9 Complementary medicines

Ursula Werneke

Complementary medicines pose a particular challenge to medical practitioners who may feel that their patients need conventional treatment but often find themselves out of their depth when patients ask about complementary therapies. Pharmacological options include herbal medicines, certain foods, and nutritional supplements such as vitamins and minerals. Physical treatments include acupuncture, massage, and osteopathy to name a few. Treatments, which purport to achieve their effects through changes in internal 'energy flow' include reiki, reflexology, healing, and therapeutic touch, and also homeopathy and traditional Chinese acupuncture. All these treatments are either used alternatively, i.e. instead of, or complementary, i.e. in addition to, conventional medicine. In patients with mental health problems, depending on the definition and inclusion criteria, estimates of the prevalence of complementary medicine use range from 8 per cent to 57 per cent. Depression and anxiety seem to be the most common indications.[1]

Herbal remedies and supplements

Principles of treatment

Herbal remedies and supplements may come in many different forms and formulations. Since they are currently not subject to the same regulatory requirements as conventional medicines they can vary substantially in contents and dose even if they purport to contain the same ingredients. The pharmacological properties of an extract or a supplement may depend on many different factors (Table 6.2.9.1).

Condition-specific remedies

Cognitive enhancers

Cognitive enhancers are either used in the treatment of dementia to enhance mental performance or prevent cognitive decline in healthy people. One strategy aims at increasing choline availability, e.g. by inhibiting acetylcholine esterase. Alternative non-cholinergic neuroprotective strategies have been postulated. These rely on antioxidants scavenging free radicals thereby reducing neurotoxicity or anti-coagulants and increasing cerebral blood flow.[4] Suggested herbal cognitive enhancers for which some positive trial evidence has been collated include ginkgo (*Ginkgo biloba*), panax

Table 6.2.9.1 Determinants of pharmacological properties of complementary medicines

Factor	Problem	Example
Material production	Quality may depend on plant material used, time of harvesting, geographical location, or other environmental factors	St John's wort extracts prepared from the flowers are more potent than extracts prepared from the leaves
Extraction method	Determines remedy composition	Alcoholic valerian extracts may be safer than aqueous extracts such as teas because harmful volatile substances (valepotriates) are eliminated more easily. The resulting extracts may be less potent though, since valepotriates have GABAergic properties.[2] Conversely, aqueous kava extracts may be safer than alcoholic extracts because liver protective substances such as glutathione are retained
Standardization	Difficult to achieve if active ingredient is unknown	St John's wort is based on the extract traditionally standardized on hypericin, a photosensitive red pigment. However, current evidence suggests that standardization should be based on hyperforin, which inhibits the reuptake of monoamines[1]
Dosing	Depends on standardization	300 mg of St John's wort extract standardized on 0.5% or 5% hyperforin most likely have different pharmacodynamic effects
Contamination	Increased and sometimes unexpected toxicity	Contamination with e.g. fertilizer residuals or heavy metals. Association of eosinophilicmyalgic syndrome and some 5-hydroxy-tryptophan products may be at least in part due to contamination of some batches
Adulteration	May lead to serious side effects and drug interactions falsely ascribed to the remedy *per se*	Adulterants include steroids, NSAIDs, anticonvulsants, benzodiazepines, hypoglycemic agents, erectogenic agents, and warfarin[3]

ginseng (*Panax ginseng*), hydergine (*Claviceps purpurea*), sage (*Salvia officinalis*), and vitamin E. The potential side effects can be derived from the purported mechanisms of action. For example, remedies increasing the cerebral blood flow such as ginkgo may increase the risk of cerebral haemorrhage. *Panax ginseng* has been associated with manic episodes and hydergine can lead to ergot poisoning unless dosed carefully. Some sage species can lower the seizure threshold. Others contain camphor, which can be toxic in high doses.[2] Also, evidence is emerging that taking vitamin E above the recommended level may increase all-cause mortality.[5]

Anxiolytics and sedatives

Drugs considered to be either anxiolytics or sedatives essentially have the same underlying mechanisms of action. The stronger an agent the more sedating it will be, leading to coma in extreme cases. Four mechanisms of action have been implicated; binding to γ-aminobutyric acid (GABA) receptors leading to hyperpolarization of the cell membrane through increased influx of chlorine anions; inhibition of excitatory amino acids (EAA) thereby also impairing the ability to form new memories; sodium channel blockade, reducing depolarization of the cell membrane; and calcium channel blockade, reducing the release of neurotransmitters into the synaptic cleft.[4] The most commonly used CAMs for anxiolysis and sedation, such as valerian (*Valeriana officinalis*), passion flower (*Passiflora incarnata*), kava (*Piper methysticum*), and German chamomile (*Matricaria recutita*) are GABAergic. Lemon balm (*Melissa officinalis*) has cholinergic and GABAergic properties. For other plant remedies, including hops (*Humulus lupulus*), oats (*Avena sativa*), lavender (*Lavendula angustifolia*), and starflower also known as borage (*Borago officinalis*), the actual mechanism of action remains unknown. Melatonin regulates the circadian rhythm and also has some GABAergic properties although trial evidence remains inconclusive. Some of these remedies can potentially have serious side effects. For instance, kava extracts have been associated with significant and potentially fatal hepatotoxicity. Starflower contains γ-linolenic acid that may lower the seizure threshold. Some passion flower extracts may contain cyanide components. As expected, all remedies in this class can lead to drowsiness when taken in high doses and can potentiate the effect of synthetic sedatives.

Antidepressants and mood stabilizers

Most complementary antidepressants are thought to work through serotonergic and noradrenergic pathways. The most robust clinical data are available for St John's wort (*Hypericum perforatum*), having been extensively reviewed in meta-analyses.[1] Hyperforin, inhibiting the reuptake of monoamines, is thought to be the most likely active component. Supplements, such as *S*-adenosylmethionine (SAMe), folic acid, L-tryptophan, and 5-hydroxytryptophan are components or co-factors in the serotonin synthesis. For SAMe, equivalence to tricyclic antidepressants has been demonstrated. However, SAMe is very expensive and a suitable oral formulation may be difficult to obtain. Selenium has also been suggested but the mechanism of action, albeit still unclear, seems to be different. Its antioxidant properties may reduce nerve cell damage. Selenium also facilitates conversion from thyroxin (T4) to thyronine (T3), and T3 substitution is one possible augmentation strategy for antidepressants. As for lithium, the therapeutic index is narrow. Omega-3 fatty acids are known to stabilize membranes and to facilitate monoaminergic, serotonergic, and cholinergic neurotransmission. The currently available evidence supports the use of eicosapentaeonic acid on its own or in combination with docosahexaonic acid as adjunctive treatment.[6] Serotonergic remedies should not be combined with each other or with conventional antidepressants because of the increased risk of serotonin syndrome. Equally, herbal antidepressants may induce mania in vulnerable patients although current evidence relies on case reports only. Finally L-tryptophan and 5-hydroxytryptophan should be avoided until the associated risk of eosinophilic myalgic syndrome is fully explained.

Remedies for psychosis

Rauwolfia (*Rauvolfia serpentina*) extracts were traditionally used before synthetic antipsychotics became widely available. Several alkaloid derivatives including reserpine were introduced in the 1950s. They block vesicular storage of monoamines so that the presence of monoamines in the cytoplasm is prolonged, enabling them to be more easily degraded by monoamine oxidases. In consequence, the amount of neurotransmitter available on depolarization of the cell membrane is reduced.[4] On the one hand, this may lead to a reduction of dopamine and the resolution of psychotic symptoms. On the other hand, less serotonin and noradrenaline will be available, which explains why drugs such as reserpine may precipitate depression. An alternative strategy is the augmentation of antipsychotic treatment with omega-3-fatty acids, but the results of clinical trials remain inconclusive and larger trials will be needed to clarify effectiveness.[6,7]

Remedies for movement disorders

Attempts have been made to treat tardive dyskinesia with antioxidants. This approach relies on the assumption that tardive dyskinesia is not only due to dopamine receptor super-sensitivity but is also related to oxidative tissue damage induced by antipsychotics. Clinical trials suggest that vitamin E may prevent the progression of tardive dyskinesia. One trial found actual improvement. However, the benefits have to be offset against taking vitamin E long-term, particularly when higher than recommended daily doses are used.[5] A far more powerful antioxidant than vitamin E is melatonin attenuating dopaminergic activity in the striatum as well as hypothalamic dopamine release.[8]

Remedies for the treatment of addiction

Only few plants have been identified as having the potential to counter addiction. Such may be ibogaine, derived from the West African shrub *Tabernanthe iboga*. It has hallucinogenic properties, and has been used to counter nicotine, cocaine, and opiate addiction. It causes dose-dependent CNS stimulation ranging from mild excitation and euphoria to visual and auditory hallucinations. The therapeutic value of ibogaine is limited since it is highly neurotoxic and can cause irreversible cerebellar damage. A synthetic derivative with similar reported effects, but without cerebellar toxicity is 18-methoxycoronaridine (18-MC).[4] Between 1990 and 2006, twelve deaths after ibogaine use were reported. More deaths may have occurred but may not have been reported due to the 'underground nature of ibogaine treatment'. Passion flower and valerian, by virtue of their GABAergic properties, may ameliorate withdrawal symptoms. Kudzu, Japanese arrowroot (*Pueraria lobata*), has traditionally been used for the treatment of alcohol hangover. The active ingredient, puerarin, counteracts the anxiogenic effects associated with alcohol withdrawal.[9]

Examples of remedies commonly used for chronic somatic conditions

Many different remedies are available for somatic conditions. Their use may be problematic in chronic conditions such as cancer or HIV where the therapeutic margin of conventional medicines is narrow.[10] For example, echinacea (*Echinacea purpurea*) is used to boost immune system. This may be detrimental where immunosuppression is desired since echinacea may potentially stimulate the growth of malignant or infectious cell lines. Patients with breast cancer may often resort to phytoestrogens such as soy (*Glycine max*), wild yam (*Dioscorea alata*), or liquorice (*Glyccyrhiza glabra*) to reverse the effects of antiestrogenic therapies such as tamoxifen. Phytoestrogens, however, can theoretically stimulate breast cancer cells and thus should be advised against in this patient group. Liquorice is a popular ingredient of many traditional Chinese medicines and may cause hypokalemia if used excessively. Evening primrose (*Oenothera biennis*) oil is a popular remedy for premenstrual syndrome and mastalgia. Like starflower it contains γ-linolenic acid and may lower the seizure threshold or reduce the efficacy of antiepileptic drugs.

Drug interactions

Determining interactions between complementary and conventional medicines can be extremely difficult. In the first instance, the clinician must be prepared to consider such a possibility and take a corresponding history. As often, association does not prove causality. Drug interactions can be distinguished into pharmacodynamic and pharmacokinetic interactions. Pharmacodynamic interactions occur when remedies act as agonists, antagonists or inverse agonists to conventional medicines. Additive toxicity, e.g. hepatotoxicity due to pyrrolizidine alkaloids, or increase of coagulability due to coumarinic constituents may also be of concern. Pharmacokinetic interactions include interactions with the cytochrome microsomal enzyme system (CYP) or membrane transporter proteins expressed through the ABC cassette genes (Table 6.2.9.2).[1,2,11,12]

The CYP system

The pharmacokinetics of most anticancer drugs is highly variable and may be genetically determined. For instance, the oxidative metabolism depends on the CYP system. The effects of CYP inducers and inhibitors are essentially differential depending on whether metabolites are more or less active (Table 6.2.9.2). If metabolites are less active than the original agent, CYP inhibitors increase whereas inducers reduce therapeutic effectiveness. Conversely, if metabolites are more active than the original agent, CYP inhibitors reduce whereas as inducers increase therapeutic effectiveness. Often such interactions have only been studied *in vitro* and it remains unclear whether they translate into tangible clinical effects.[11,12] In clinical practice, it is often possible to monitor combination of medicines more closely or to adjust the doses of conventional drugs in the required direction rather than to advise discontinuation of complementary remedies. Interactions with CYP 3A4 are of particular concern, since this enzyme metabolizes up to 60% of all clinically used drugs including HIV protease inhibitors, HIV non-nucleoside reverse transcriptase inhibitors, warfarin, ciclosporin, oral contraceptives, digoxin, theophylline, anticonvulsants, and various psychoactive drugs.[13]

ABC transporters

The ABC cassette genes represent proteins binding to ATP and use this energy to drive various molecules through cell membranes. The transport is mostly unidirectional. The ABC genes have been mainly explored for their capacity to cause multi-drug resistance in cancer chemotherapy.[12] Thus remedies exerting such effects may be of particular interests to the liaison psychiatrist. The most commonly known transporter is p-glycoprotein involved in the

Table 6.2.9.2 Examples of potential drug interactions of commonly used psychotropic remedies

Remedy	Pharmacodynamic	Pharmacokinetic
Ginkgo	Antithrombolytic agents	1A1, 1A2, 2B1/2, 2C9, **2C19***, 3A1, **3A4** inhibition
Panax ginseng	Insulin and oral hypoglycaemics, antithrombolytic agents, MAOIs (phenelzine), loop diuretics	1A1, 1A2, 1B1, 2C9, 2C19, **2D6**, 2E1 inhibition3A4 inconclusive; p-glycoprotein inhibition
Hydergine	Serotonergic antidepressants, choline-esterase inhibitors	
Vitamin E	Anticoagulants and antiplatelet drugs; prevention of nitrate tolerance possible; ↑ effect of sildenafil and related phophodiesterase-5 inhibitors possible; ↓ effect of chemotherapies relying on oxidative stress	CYP 3A11 induction
Valerian	↑ Effect of sedatives	**CYP 3A4** and p-glycoprotein inhibition
Passion flower	Anticoagulants, ↑ effect of sedatives	CYP 3A4 inhibition
Kava	↓ Effect of levodopa	Potentiation of liver toxicity of other drugs CYP 1A2, 2C9, 2C19, 2D6, 3A4 and 4A9/11 inhibition
Melatonin	Anticoagulants, ↑ effect of sedatives; ↓ effect of chemotherapies relying on oxidative stress	
St John's wort	Serotonergic antidepressants	CYP 3A4, 1A2, 2C9, **3A4**and **2E1** induction:, p-glycoprotein induction
Omega 3 fatty acids	↑ Effect of warfarin, aspirin and non- steroidal anti- inflammatory drugs	CYP 3A4 and p-glycoprotein inhibition
Rauwolfia	↑ Effect of anti- psychotics and barbiturates; ↓ effect of levodopa; severe bradycardia with digitalis glycosides; hypertension in combination with sympathomimetics	
Iboga	Cholinergic and anticholinergic drugs	
Echinacea	↓ Effect of immunosuppressants	CYP **2A1**, 2C9, and **3A4** inhibition, CYP 3A4 induction also possible depending on extract
Evening primrose oil	Other drugs reducing seizure threshold, anticoagulants	1A2, 2C9, 2C19, 2D6, 3A4 inhibition

*Bold font: *In vivo* evidence available.

transport of many psychotropic drugs through the blood brain barrier. St John's wort, valerian, and panax ginseng are remedies shown to change p-glycoprotein activity (Table 6.2.9.2).[11,12] However, whether such effects are sufficiently powerful to affect conventional treatments remains unclear.[14]

Conclusions

At present, the evidence base for the use of psychotropic complementary medicines is extremely limited. Due to the large variability of formulations it can be extremely difficult to conduct clinical trials with replicable results even if a candidate plant has been identified. Pooling results of existing trials in meta-analyses may be unhelpful if the trials are too small or heterogenous or if the analysis is not adjusted for the extract types used. Equally, systematic pharmacovigilance is difficult to implement in the absence of a regulatory framework.

Clinicians need to be aware that patients may use complementary therapies regardless of the evidence available and should inquire about such forms of self-medication. Pattern of use may vary with cultural background and health beliefs. Given the complex pattern of potential interactions, conventional health care professionals should not be afraid to discuss complementary use with their patients. For instance, complementary medicines should be considered a potential cause when the clinical presentation, the treatment result, adverse effects, or even diagnostic investigations are unusual or unexpected. Equally, patients should be encouraged to disclose information about complementary medicines to health care professionals. On the one hand, discussions need to be conducted sensitively in order to avoid alienating patients who may feel that they have not been taken seriously or have been criticized for using complementary medicines. On the other hand, uncritical encouragement of potentially harmful or inappropriate use of complementary medicines may possibly lead to litigation.(15) In most cases, remedies may not have to be discontinued if conventional treatments are closely monitored and adjusted. A constructive discussion about complementary medicines may potentially be a gateway towards enhancing compliance with conventional treatments.

Further information

Memorial Sloan-Kettering Cancer Center: Cancer Information: Integrative Medicine: www.mskcc.org. Keyword: herbs.

National Centre for Complementary Alternative Medicines / National Institute of Health: http://nccam.nih.gov.

Royal Botanic Gardens, Kew: Education: Resources: Information Sheets: www.rbgkew.org.uk/ksheets/.

Royal College of Psychiatrists: Mental Health Information: Therapies: Complementary and Alternative Medicines 1 & 2: www.rcpsych.ac.uk/mentalhealthinformation/therapies.aspx.

The Prince of Wales Foundation for Integrated Health: http://www.fih.org.uk/.

References

1. Werneke, U., Turner, T., and Priebe, S. (2006). Complementary alternative medicine in psychiatry: a review of effectiveness and safety. *The British Journal of Psychiatry*, **188**, 109–21.
2. Natural Medicines Comprehensive Database. (2007). Keyword: product search. www.naturalmedicines.com.
3. Ernst, E. (2002). Adulteration of Chinese herbal medicines with synthetic drugs: a systematic review. *Journal of Internal Medicine,* **252**, 107–13.
4. Spinella, M. (2001). *The psychopharmacology of herbal medicine*. MIT Press, Cambridge.
5. Food Standards Agency. Expert Group on Vitamins and Minerals. (2003). *Safe upper levels for vitamins and minerals*. www.foodstandards.gov.uk.
6. Freeman, M.P., Hibbelen, J.R., Wisner, L.K., *et al.* (2006). Omega-3 fatty acids: evidence base for treatment and future research in psychiatry. *The Journal of Clinical Psychiatry*, **67**, 1954–67.
7. Joy, C.B., Mumby-Croft, R., and Joy, L.A. (2003). *Polyunsaturated fatty acid supplementation for schizophrenia* (*Cochrane review*). In *The* Cochrane Library, Issue 4, John Wiley & Sons, Ltd., Chichester, UK.
8. Lohr, J.B., Kuczenski, R., and Niculescu, A.B. (2003). Oxidative mechanisms and tardive dyskinesia. *CNS Drugs*, **17**, 47–62.
9. Overstreet, D.H., Keung, W.M., Rezvani, A.H., *et al.* (2003). Herbal remedies for alcoholism: promises and possible pitfalls. *Alcoholism: Clinical and Experimental Research*, **27**, 177–85.
10. Labriola, D. and Livingston, R. (1999). Possible interactions between dietary antioxidants and chemotherapy. *Oncology* (*Williston Park*), **13**, 1003–8.
11. Sparreboom, A., Cox, M.C., Acharya, M.R., *et al.* (2004). Herbal remedies in the United States: potential adverse interactions with anticancer agents. Journal of Clinical Oncology, **22**, 2489–503.
12. Sparreboom, A., Danesi, R., Ando, Y., *et al.* (2003). Pharmacogenomics of ABC transporters and its role in cancer chemotherapy. *Drug Resistance Updates*, **6**, 71–84.
13. Committee of Safety in Medicine & Medicines Control Agency. (2000). Reminder: St John's wort (*Hypericum perforatum*) interactions. *Current Problems in Pharmacovigilance*, **26**, 6–7.
14. Morris, M.E. and Zhang, S. (2006). Flavonoid-drug interactions: effects of flavonoids on ABC transporters. *Life Sciences*, **78**, 2116–30.
15. Cohen, M.H. and Eisenberg, D.M. (2002). Potential physician malpractice liability associated with complementary and integrative medicinal therapies. *Annals of Internal Medicine*, **136**, 596–603.

6.2.10 Non-pharmacological somatic treatments

Contents

6.2.10.1 Electroconvulsive therapy

Max Fink

Introduction

Convulsive therapy (ECT or electroshock) is an effective treatment for those with severe and persistent emotional disorders. It is safe for patients of all ages, for those with debilitating systemic illnesses and during pregnancy. It relieves symptoms in a briefer time than do psychotropic drugs. To achieve remission, treatments are usually given three times a week for two to seven weeks. To sustain recovery, treatments are continued either weekly or biweekly for several months. The overall duration of the treatment course is similar to that of the psychotropic medications frequently used for the same conditions.

The treatment is severely stigmatized and its use is discouraged, even interdicted, in the belief that the electricity or the seizures irreversibly damage the brain.[1–5] Few physicians are tutored in its use and facilities are limited making ECT unavailable to many who would benefit. The ease in the use of psychotropic medications, and neither greater efficacy nor greater safety, encourages their preferential use as ECT is relegated to the 'last resort.' In countries where psychotropic medications are expensive, ECT is prescribed, but the expense for anesthetics limits its use to its unmodified form.

Despite these hurdles of stigma, expense and lack of training, its use has persisted for more than 70 years. Indeed, its use is increasing. Whole societies where it was interdicted at the end of the 20th century, as in the Netherlands, Germany, Austria, Italy, and Japan, interest and usage has increased, texts have been written or translated, and local psychiatric societies formed to encourage its use.[4–6]

Origins

At the turn of the century when malarial fevers were used to treat patients with neurosyphilis, it was deemed possible to treat one illness by developing another. Reports that patients with dementia praecox were relieved of their psychosis after suffering convulsions supported a concept of an antagonism between epilepsy and psychosis. An explanation was seen in the reports that the concentrations of brain glial cells in patients with dementia praecox were low and in those with epilepsy were high.[7] Was it possible that the root of schizophrenia lay in the paucity of glia and would their increase relieve the illness?

After testing ways to induce a grand mal seizure in animals, Ladislas Meduna, a Hungarian neuropathologist and psychiatrist, on January 24, 1934, injected camphor-in-oil into a man with the catatonic form of schizophrenia. The patient seized and recovered without incident. Following the model of malarial fever therapy, Meduna repeated the injections every three days, and after the fifth seizure, the patient, for the first time in four years, talked spontaneously and fed and cared for himself. After three additional treatments he was discharged home, returned to work, and was well when Meduna left Hungary in 1939.[7]

Chemically induced seizures, either with camphor or pentylenetetrazol (Metrazol), were rapidly adopted worldwide as the treatment for dementia praecox, but the treatments were painful and

frightening, so alternative means were sought. In 1938, the Roman psychiatrists Ugo Cerletti and Luigi Bini demonstrated the ease of administration and the efficiency of electrically induced seizures. Quickly, 'electroconvulsive therapy' (ECT, electroshock) became the commonest method of inducing seizures and is the standard induction today.[4–5,8]

For whom is ECT effective?

Established DSM diagnoses are usually cited as the indications for ECT. The diagnoses are imprecise, however, offering heterogeneous population samples. A syndromic view offers more homogeneous populations for treatment.[9–11]

Defined by DSM Classification. The DSM defined conditions for which ECT is prescribed are cited in established texts[12–15] (Table 6.2.10.1.1). The breadth of its clinical efficacy across major DSM diagnostic classes is striking, reflecting commonalities in the pathophysiology of different disorders. This experience challenges the concept that DSM classified disorders are distinct biological abnormalities, and supports the 19th century concept of a unitary psychosis.[16]

ECT is ***not*** useful for a patient with neurosis, situational maladjustment, personality disorder (character pathology), or drug dependence. It is of limited benefit for anyone with a lifelong history of mental and emotional dysfunction, unless the onset of the present illness is acute and well defined, or affective, psychotic, or catatonic features dominate the presentation[12–15] (Table 6.2.10.1.2).

Defined by syndrome. DSM-III and DSM-IV classify illnesses based on the check-off of symptoms modified by duration criteria. The DSM criteria identify heterogeneous populations that do not support useful treatment algorithms or the search for biological roots of the illnesses. Clinical syndromes describe more homogeneous populations, often substantiated by biological tests and/or a high specificity of interventions. Melancholia, psychotic depression, catatonia, delirious mania, and acute schizophrenia are syndromes that are particularly responsive to ECT (Table 6.2.10.1.3). These syndromes are not readily identified in the established classification systems. Summary descriptions are offered here; the interested reader will find more extensive descriptions in the cited literature.

(a) Depressive mood disorders

Convulsive therapy is most effective against mood disorders, depression and mania. Depressive mood disorders are dominated by sadness, hopelessness, fears, and thoughts that life is no longer worth-living. Variants are recognized, dominated by vegetative and motor abnormality (melancholia), by delusions (psychotic depression), by severe cognitive deficit (pseudodementia), or by catatonia.[14]

While all variants respond to induced seizures, some also respond to other specific treatments. Melancholic and pseudodementia patients respond to tricyclic antidepressants. Psychotic depressed patients require high doses of both antidepressant and antipsychotic medications.[11] Anticonvulsant sedative drugs, the barbiturates and the benzodiazepines, are useful in catatonic patients.[9]

(i) Melancholia

Motor signs (retardation or agitation) and vegetative symptoms of inability to sleep, feeding, and weight loss are its features. Work, sex, and family are disregarded. Thoughts of suicide are prominent.[10–11]

Table 6.2.10.1.1 DSM defined clinical diagnoses in which ECT is effective[11]

Major depressive disorder	
single episode	[296.2x]**
recurrent	[296.3x]**
Bipolar disorder	
mania	[296.4x]**
depressed	[296.5x]**
mixed type	[296.6x]**
not otherwise specified	[296.70]**
Atypical psychosis	**[298.90]**
Schizophrenia	
catatonia	[295.2x]
schizophreniform	[295.40]
schizo-affective	[295.70]
Catatonia	
Schizophrenia, catatonic type	[295.2x]
Catatonic disorder due to a medical condition	[293.89]
Malignant catatonia	[293.89]
Neuroleptic malignant syndrome	[333.92]
Delirium	
Due to a general medical condition	[293.0]
Due to substance intoxication	

* from Fink, 1999

** specifier for psychosis

Hypercortisolemia is characteristic of the syndrome.[17] Cortisol metabolism is influenced by hypothalamic, pituitary, and adrenal interactions. Melancholic patients exhibit elevated serum levels of cortisol, obtunded diurnal rhythmicity, and serum levels remain elevated despite an administered dose of dexamethasone. The abnormality is measurable by the dexamethasone suppression test (DST) or its variant, the dexamethasone/corticotrophin releasing factor test (Dex/CRH). Elevated cortisol levels normalize with treatment and become abnormal again with relapse. In the 1980s, the specificity of the DST was considered poor for the major depressions defined by DSM-III and the test was discarded. But the re-assessment of the literature and recent reports find the test as

Table 6.2.10.1.2 DSM diagnoses in which ECT is ineffective[11]

Dementia and Amnestic Disorders	[293.0, 290.xx, 294.xx]
Substance-related Disorders	[303.xx, 291,x, 304.x, 292.x]
Anxiety and Somatiform Disorders	[300.xx]
Factitious Disorders	[300.xx]
Dissociative Disorders	[300.1x, 300.6]
Sexual Dysfunctions	[302.xx, 625.8, 608.89, 607.84, 608.89, 625.8]
Sleep Disorders	[307.xx, 780.xx]
Impulse disorders	[312.3x]
Adjustment disorders	[309.xx]
Personality disorders	[301.xx]

Table 6.2.10.1.3 ECT responsive syndromes

Mood disorders
Depression
Melancholia
Psychotic (Delusional) depression
Mania
Mixed States (mania, depression)
Rapid cycling mania
Depressive phase of bipolar disorder
Delirious mania
Psychosis
Acute schizophrenia
Postpartum psychosis
Catatonia
Hypokinetic catatonia (Kahlbaum Syndrome)
Excited catatonia (delirious mania, oneiroid state)
Malignant catatonia (NMS, TSS)
Other
Delirium
Suicide risk
Status epilepticus (SE< NCSE)

* from Fink, 1999.

both sensitive and specific for melancholic depression, where it has a positive predictive value.[11,18]

After an extensive review of the literature, Taylor and Fink (2006) concluded that classifying mood disorder patients as either melancholic or non-melancholic offered more homogeneous populations with better outcomes with TCAs and ECT than did the DSM classification of major depression and bipolar disorder.[11] In their formulation, melancholia is a syndrome of depressive mood, with motor and vegetative abnormalities and with evidence of cortisol abnormality.

(ii) Delusional (psychotic) depression

Overwhelmed by feelings of helplessness, hopelessness, and worthlessness, the patient believes others are watching or talking about him, reporting voices when no one is present. He imagines that events depicted on a television or movie screen apply directly to him. This form is labelled psychotic depression and is remarkably responsive to ECT.

In 1975, Glassman and his associates at Columbia University reported that only three of 13 delusional depressed patients (23 per cent) improved when they were treated with high doses of imipramine, while 14 of 21 non-delusional patients (66 per cent) improved under the same treatment.[18] Nine of the 10 unimproved delusional patients responded well to ECT. These findings have been repeatedly verified.[11,12]

In a study of 437 depressed hospitalized patients treated with imipramine in doses of 200 to 350 mg/day for 25 days or longer, 247 (57 per cent) were evaluated as recovered and were discharged.[19] When the 190 unimproved patients were treated with bilateral ECT, 156 (72 per cent) were recovered. Most of the depressed patients who had not improved with imipramine were delusional as well as depressed.

Only a third of delusional depressed patients recover when treated with antidepressant drugs alone and half recover with antipsychotic drugs alone.[11, 19–20] Two-thirds of those treated with ECT or with high doses of both antidepressant and antipsychotic drugs regain their health.

In a two-year study of late-life depression, 47 per cent of the delusional depressed patients treated with medication relapsed earlier and more often than the nondelusional depressed (15 per cent), indicating that delusional depression is particularly resistant to medication.[21] It is, however, so amenable to ECT that it is considered a primary indication for its use.[11–14] But the condition is difficult to diagnose making inadequate treatment common. In a three-hospital research study of ECT and continuation medications, only 2 of 52 delusional depressed patients had adequate courses of medication treatment before they were referred for ECT.[22] The same failure was found in another multi-center study with only 5 of 106 patients failing adequate courses of treatment before referral to ECT.[23]

Many reviews find psychotic depression to have a more severe pathophysiology and just using the same treatments as for non-psychotic depression, even at much higher doses is not adequate.[11,12] Yet, bilateral ECT is remarkably effective. In a multi-site collaborative ECT study, of 253 patients with unipolar major depression, 77 were psychotic depressed. Their remission rate was 95 per cent compared to 83 per cent for the non-psychotic depressed, with the speed of response faster for the psychotic depressed patients.[24, 25]

(iii) Pseudodementia (reversible dementia)

Because the depressed patient ignores daily events, little of what happens to him is registered and memory is compromised. The condition is hardly distinguishable from Alzheimer's dementia. The onset is usually more rapid and severe compared to the onset of a structural dementia, and patients often report a history of prior depressive episodes.[11,14]

Because the syndrome is not well known, patients are often sent to nursing home care. An example of a 58-year old woman who developed a reversible dementia and was not adequately treated for eight years is reported. Once the diagnosis was considered, antidepressant treatment relieved the syndrome and returned the patient to a more normal family life.[14]

(iv) Catatonia

When the patient is mute, sitting rigidly in a chair or lying motionless on his bed, and unresponsive to questions and commands, he appears as in a stupor. The state is called *catatonia* or *depressive stupor*. Catatonia is seen among patients with many DSM diagnoses.[9] It is discussed in detail below.

(b) Manic mood disorders

A mood disorder dominated by grandiosity, expansiveness, feelings of increased power and energy, and excitement, can last for hours, days, weeks, or months. Even after it is relieved, it may recur or alternate or combine with episodes of depression. When the switches occur within one or a few days, the experience is labelled *rapid cycling*, a malignant form of the illness. *Bipolar disorder* is the label applied to both mania and mixed forms of the illness.[11,26]

Disturbances in eating and sleeping, thinking, memory, and movement are features of mania. The patient does not sleep, eats poorly, loses weight, and concentrates thoughts poorly. Memory is impaired, often severely; he may be so disorganized as to appear demented and delirious. Melancholia, psychosis, pseudodementia, and catatonia variations are commonly seen.

Delirious mania is a striking form of mania. A normal person suddenly becomes excited, restless, and sleeps poorly, fears that neighbors are watching him, and is easily frightened. He may hide in the house or leave it abruptly, dressed inappropriately, sometimes naked, and wander about the streets. His hallucinations are vivid, his thoughts disorganized. Confusion alternates with mutism, posturing, rigidity, and stereotyped repetitive movements. Physical exhaustion even to the point of death occurs.[11,27]

Before ECT, patients were sedated with opiates, bromides, or chloral and many died of poor care, inanition, and pneumonia. A 1994 summary of the reports of manic patients treated with ECT finds 371/562 (66 per cent) remitted or showed marked clinical improvement.[28] The introduction of chlorpromazine and other sedative drugs quickly replaced ECT for efficacy and ease of use. But when chlorpromazine and other antipsychotic drugs were used in place of ECT, the doses often carried the risks of sudden death and neuroleptic malignant syndrome, as well as tardive dyskinesia and tardive dystonia.[9]

Anticonvulsant drugs are now preferentially recommended, even though the evidence for their efficacy is poor. Many authors encourage the use of lithium for immediate relief and for prophylaxis. In 438 manic patients treated with ECT or lithium, 78 per cent of the ECT treated group showed marked improvement compared to 62 per cent of those treated with adequate doses of lithium and 56 per cent of those treated with inadequate doses.[29] The group receiving neither ECT nor lithium fared least favourably with only 37 per cent improved.

No matter the array of medications and polypharmacy for mania, ECT is an effective alternative.

(c) Catatonia

Muscular rigidity, posturing, negativism, mutism, echolalia, echopraxia, and stereotyped mannerisms, the signs of catatonia, appear suddenly and immobilize patients.[9] When the disorder is transient, it may be disregarded, but when it persists, it threatens life. Patients undergo forced feeding and develop bedsores, muscular atrophy and pulmonary embolization. Repeated bladder catheterizations induce infections.

Catatonia is recognized in patients with affective illnesses, both depression and mania, in patients with systemic disorders, and in those with toxic brain states caused by hallucinogenic drugs. For decades, the prevailing belief was that each instance of catatonia represented schizophrenia. The major classification systems in psychiatry — DSM-III and IIIR of the American Psychiatric Association and the International Classification of Diseases (ICD-IX, ICD-X) — assigned patients with catatonia to the diagnosis of schizophrenia, catatonic type. Few patients were treated with anticonvulsant sedatives or ECT, despite their known efficacy, because neither was recommended for schizophrenia. This short-sighted view was somewhat corrected in the 1994 classification system of the American Psychiatric Association (DSM-IV), which recognized catatonia as secondary to systemic illness in the class of "*Catatonic disorder due to . . . . (Indicate the General Medical Condition)* [293.89]".[30] The experience that catatonia is not limited to patients with "schizophrenia" has led to the call for a separate category in DSM-V.[9,31]

Catatonia is defined by the persistence of two or more characteristic motor signs for more than 24 h in a patient with a mental disorder.[9,31] Posturing and staring can be observed, but most signs require elicitation in the examination. The accepted motor signs and a formal examination are cited in catatonia rating scales.[9] An intravenous challenge of lorazepam or amobarbital verifies the diagnosis in more than 2/3 patients with catatonia, and a positive test response augurs well for high dose benzodiazepine therapy. When this treatment fails, ECT is effective, although the treatment schedule may require daily treatments.

Catatonia may be transitory or may persist for months or years. It appears in many guises.[9,32] Prominent examples are *malignant (pernicious) catatonia* (MC) with a high risk of death and the *neuroleptic malignant syndrome* (NMS) that follows on the administration of neuroleptic drugs.

(i) Malignant catatonia

Descriptions of patients who develop an acute febrile delirium with excitement or stupor dot the literature. They often exhibit signs of catatonia. Vegetative dysregulation is often severe and death was a frequent feature before the introduction of ECT. Descriptions by Bell (1849), Stauder (1934), and Bond (1950) highlight the lethal nature of the syndrome. In 1952, Arnold and Stepan described patients in whom ECT rapidly relieved malignant catatonia, but to avoid mortality it had to be used within the first five days.[9]

(ii) Neuroleptic malignant syndrome (NMS)

A toxic response to neuroleptic drugs evinced by fever, motor rigidity, negativism, mutism, and cardiovascular and respiratory instability is a toxic response to neuroleptic drugs. It is indistinguishable from malignant catatonia.[9,32] It is an MC variant as the diagnostic criteria and effective treatments are the same as for MC. MC occurs with almost all neuroleptics, most commonly with the high-potency agents like haloperidol, fluphenazine, and thiothixene, but also with atypical neuroleptics.

One hypothesis explains the syndrome as a consequence of an excessive reduction in the amount of brain dopamine. Those who believe this association prescribe the dopamine agonists bromocriptine or levodopa and relieve muscular rigidity by prescribing the muscle relaxant dantrolene. Neither of these treatments has proved effective and dantrolene use is associated with considerable toxicity.[33] These are best not used and patients are best treated with sedative anticonvulsants and ECT.

(iii) Toxic serotonin syndrome (TSS)

A toxic syndrome is occasionally described in association with the SSRI antidepressant drugs. TSS is similar to MC with prominent gastrointestinal symptoms. The diagnosis and treatment follows the protocol for MC.[9]

(d) Psychosis

A severe impairment of thought characterized by delusions is a feature of many psychiatric conditions, notably manic delirium, psychotic depression, post partum depression, and toxic psychosis. It is broadly defined as a psychosis and diagnosed within the major class of psychoses as schizophrenia. In this class ECT is hardly considered. But when we consider the efficacy of ECT in the psychotic variants of the mood disorders, we appreciate that ECT is an effective treatment of psychosis.[34]

Convulsive therapy was introduced for the treatment of dementia praecox and was widely and quickly adopted. Comparisons with chlorpromazine found both treatments effective in acute and severe short-term illnesses, but neither was useful in chronic states. Chlorpromazine was favoured since its cost is considerably less and

its image better. As more patients failed to respond to medications, however, a cadre of 'medication resistant' psychotic patients developed. Families asked whether anything else could be done to better the patients' lives. Friedel (1986) augmented a failed course of thiothixene therapy with ECT, returning each of nine patients to community life. The finding was replicated in the successful augmentation in 8/9 psychotic patients.[34]

Clozapine was described as a treatment for psychotic patients who had failed to respond to two different antipsychotics. As the experience with this treatment grew, clinicians were again faced with treatment failures and ECT augmentation was tried. A synergy for ECT and clozapine was described and offers an effective treatment for patients who have failed conventional antipsychotics and clozapine.[34]

It is reasonable to consider ECT in the treatment of psychosis, whether in an affective illness or in schizophrenia. For the affective illnesses, ECT is used alone. In schizophrenia, ECT is effective alone or in augmenting neuroleptics.[34]

(e) Delirium

Acutely ill psychotic patients often exhibit disturbances in consciousness and are confused. Delirium is common in toxic states, either drug induced (alcohol being the most common), or secondary to drug withdrawal, or associated with systemic illnesses. Delirium is a feature of acute manic states (e.g. delirious mania) and the confusional state described as oneirophrenia. With few resources to treat acute psychoses, ECT was applied with favourable results.[14,35] The relief of delirium by ECT is an unrecognized effect that warrants consideration as an alternative to the risks of high potency neuroleptic drugs inducing NMS (MC).

(f) Neurological syndromes

ECT is well appreciated in catatonia, but it is also useful in status epilepticus (SE), non-convulsive status epilepticus (NCSE), and Parkinsonism.

(i) Status epilepticus

SE and NCSE are emergency conditions with high mortality rates. The pathophysiology is the persistence of seizures as biochemical inhibitory mechanisms fail to terminate a seizure.[36] Despite ever larger doses of anticonvulsant medications, proceeding from lorazepam to phenytoin, phenobarbital, and general anesthesia with midazolam, propofol, or barbiturates, patients persist in SE and NCSE.

ECT is another effective intervention. During the course of electroconvulsive therapy, the seizure threshold rises, encouraging seizure termination. The first report of the relief of intractable epilepsy by ECT in 1943 has been sporadically verified.[37]

An explanation for this application is physiologically interesting. The strength of a seizure can be judged by the immediate rise in serum prolactin after a sustained epileptic seizure. Within the hour after a seizure, the level of serum prolactin indicates whether the seizure is a cerebral grand mal event or a pseudoseizure. Serum prolactin levels do not rise in SE but remain normal. This suggests that the SE seizures are partial or incomplete and that they fail to stimulate an inhibitory termination process. But even in patients in SE, ECT elicits maximal seizures, making it a reasonable alternative to general anesthesia as a treatment for intractable seizures.

(ii) Parkinsonism

In treating older depressed patients with concurrent Parkinsonism with ECT, motor and facial rigidity were also relieved. In Parkinsonism, brain dopamine levels are reduced, making dopamine agonists effective treatments. In ECT, brain and CSF levels of dopamine increase. Experiments in Parkinsonian patients without mood disorder found motor rigidity to be relieved.[38] For those patients who are not relieved by conventional treatments, periodic ECT has been helpful. Continuation treatments, like continuation pharmacotherapy, are necessary to sustain the benefit.

(g) Suicide

All psychiatric disorders carry the risk of suicide. ECT reduces this drive. The impact of medications on suicide risk is not well defined but compared to ECT, the efficacy is less favourable.[6,11] Comparisons of ECT and TCAs across different treatment eras find the frequency of suicides decreased in the ECT era. A study of the psychiatric status of 519 patients six months after discharge from hospital treatment for depression found 0.8 per cent of the ECT treated patients had made a subsequent suicide attempt compared to 4.2 per cent for those rated as receiving adequate and 7 per cent of those receiving inadequate courses of antidepressant drugs. At the 6-month follow-up no suicides were reported in 34 women treated with ECT, but two suicides occurred in the 84 patients treated with antidepressants (2.4 per cent).[39]

In a study of the expressed suicide intent (changes in Item 3 of the HAMD rating scale) in 148 patients treated with ECT, the baseline average score was 1.8. It reduced to 0.1 in 72 responders and to 0.9 in 76 non-responders. For the total sample, there was a greater decrease in the suicide item scores than in the overall HAMD scores.[40]

In another study of 444 patients referred for ECT, 131 had high expressed suicide intent scores.[6] The scores dropped to zero in 106 (80.9 per cent) with treatment, occurring in 38.2 per cent (50/131) after 3 ECT (one week), in 61.1 per cent (80/131) after 6 ECT (two weeks); and in 76.3 per cent (100/131) after nine ECT (three weeks).

ECT's effect on the death rate in the mentally ill, particularly those with mood disorders, must be a major consideration in treatment recommendations.

Principles of treatment

When to consider ECT? Psychotropic drugs and psychotherapy are the first treatments of the psychiatrically ill, with referral to ECT when these treatments fail. Since ECT is effective in medication treatment failures, would it not be wise to spare patients a prolonged illness and risks of suicide by offering ECT as the initial treatment? ECT is indeed considered the first treatment when there is a need for a rapid, definitive response, as in suicidal patients who require constant observation and restraint, in hyperactive patients who may be at risk of harm to themselves or others, in those with malignant disorders as malignant catatonia, neuroleptic malignant syndrome, or delirious mania, or in those whose lives are in jeopardy from systemic illness. It is also preferred in those patients who have had a prior illness that responded well to ECT or who have had a poor experience with medications.[11–15]

How many failed trials of medications are reasonable before ECT is considered? For some patients, especially those whose practitioners are not knowledgeable about ECT, medication trials

become interminable and ECT is considered only when the patient seeks care elsewhere. A reasonable guideline is derived from the experimental trials with clozapine, an agent with life-threatening risks.[41] To put patients at risk and yet obtain the possible benefits of clozapine, the researchers decided that patients should not be offered clozapine unless they had experienced two unsuccessful courses of neuroleptic treatment. A similar standard seems reasonable for recommending ECT. After patients have failed two different courses of medications at adequate doses and for adequate periods, ECT is to be considered.

Financial considerations affect the decision. If the patient is severely ill and has only a limited ability to pay for extended care, repeated unsuccessful medication trials are unwarranted. All practitioners should balance the cost of medication trials and the effective use of ECT.

Consideration of age. ECT is an accepted treatment for adults. For decades, the attitudes of child and adolescent psychiatrists precluded consideration of ECT for their patients except the most devastatingly ill. The acknowledged safety of ECT in adults relaxed prejudices against its use and led to more treatment trials. Once it became clear that the response of adolescents was similar to that of adults, the attitude changed and ECT is now an accepted treatment for adolescents with the same illnesses that are successfully treated in adults.[42]

ECT is probably effective in similar conditions in children, but their expression of mood and psychotic disorders is different than in adults and difficult to interpret. The published experience in the few children treated with ECT finds that conditions that respond in adults and adolescents also respond in children.

ECT is widely used in geriatric patients. Indeed, it is increasingly called on when the side effects of medications become intolerable and when medication trials fail. The safety of modern ECT is such that even the frailest and systemically ill elderly can be safely treated with ECT. We acknowledge no absolute contraindication to ECT other than the lack of skill of the clinicians.[11–15]

The treatment process

Consent. The referral of a patient to an ECT service starts the treatment process. As in surgical treatment, the patient and family members are educated as to the risks and processes of the treatment course, and a signed voluntary consent, witnessed by a family member if possible, is obtained. In response to the turmoil of the 1970s when a draft for an unwelcome war led to widespread questioning of authority, attacks on ECT as a forced involuntary treatment led the profession to suggest procedures for informed voluntary consent. These procedures are well established.[1,2,4,12]

An explanation of why the treatment is recommended, specific anticipated benefits and risks, the names of the responsible physicians, and a statement that the patient may, at any time, discontinue the treatment are elements of a valid consent.[4,13,14] Although voluntary consent is the basis for ECT in almost all Western countries, provisions for involuntary treatment for patients who may not be able to understand the severity of their illness nor the need for treatment is provided in state laws with courts authorizing treatment. In a few venues, surrogate consent by family members is accepted. Educational videotapes and books for laymen support the consent process.[4,12,43]

Procedures. Treatments are usually given in an equipped room with access to the in-patient wards. Increasingly, as more than half the treatments are given to out-patients, units are established with ready access to the community.[12,13]

Prior to treatment, systemic medical examinations usually advised for general anaesthesia are completed. These include complete blood count, electrocardiogram, and urinalysis. If systemic illness is present, the treatment is optimized. Often, an anaesthesiologist will examine the patient and the record before treatment, obtaining a separate anaesthesia consent. Although no medical examinations relative to the ECT process are required, some centers unnecessarily insist on pre-treatment brain scans and EEG for all patients.

Anaesthesia. When curare and succinylcholine were introduced to modify the convulsion, patients thought they were suffocating as respiratory muscles relaxed. Momentary amnesia was provided by a barbiturate and the combination of barbiturate-induced amnesia and succinylcholine muscle relaxation became standard procedure.[8,12] When psychiatric practice changed from an office to a hospital venue, and anaesthesiologists administered medications, misunderstanding of the role of anaesthesia ensued and the benefits of treatments were reduced by high anaesthetic doses that made effective seizures difficult. Present practice is detailed in anaesthesiology texts.[44]

Monitoring and electrode placement. To monitor the physiologic effects of induced seizures, EEG and ECG electrodes are applied. To monitor the motor seizure, a blood pressure cuff is usually applied to the calf of one leg, inflated before the administration of a motor relaxant to observe the motor seizure duration. Two stimulating electrodes are required for ECT. In the early years, the electrodes were applied to both temples, with the maximum energies passing through the intervening brain tissues, especially the centrencephalic structures of the hypothalamus and pituitary. Relocating the electrodes on one side of the head to avoid stimulating the dominant temporal lobe led to seizures with less immediate impact on cognition. 'Right unilateral ECT' (RUL-ECT) became popular until clinicians realized that the efficacy of such treatments was significantly less than through bilateral electrodes (BL-ECT).[8,12]

At one time we believed that any seizure was therapeutic, but we now know that this is not so. A seizure with EEG or motor durations under 20 sec rarely develops a full grand mal convulsion. At first, effective treatments were characterized as those with a motor seizure of at least 25 sec. But not all seizures of such length are effective. Seizures induced through unilateral electrodes at near-threshold energies (experimentally identified as 1.5 and 2.5 times the calibrated seizure threshold) are not as effective as seizures induced through bilateral electrode placements.[45] Energies for seizure inductions in unilateral ECT must be at least 6 to 8 times the calibrated seizure threshold to achieve equal efficacy; at such high energies the advantage in minimizing immediate memory effects is lost.[46] As there is a linear relationship between age and seizure threshold, the energy levels with modern devices that deliver brief pulse electrical currents for BL-ECT is estimated by the half-age formula.[47] In devices that deliver 500 mC of energy at 100 per cent, the energy level for the first induction is set at half the patient's age. The quality and duration of the EEG seizure are a guide to later induction energies. In present clinical practice, electrodes are applied to both temples (BT-ECT) or over the outer canthus of each eye in 'bifrontal' (BF-ECT) placement. While the

advantages of BF-ECT and BT-ECT are being assessed in large studies, their efficacy seems equivalent. There is little justification for the use of RUL-ECT in clinical practice.

We now rely on the ictal EEG to define an effective treatment, and modern ECT devices record either one or two channels of brain electrical activity. The typical ictal EEG presents a build-up of energies, then high-voltage spike activity mixed with high-voltage slow waves (3–6 Hz), followed by trains of lesser voltage slow waves, and an abrupt end to the electrical activity with electrical silence. Such EEG patterns, generally of 35–130 sec in duration, are associated with motor seizures that are 10–20 per cent shorter. If seizures do not show these well-defined phases, we repeat the treatment at different energy settings until a robust EEG sequence is elicited.[11–15]

A rise in the post-ictal serum prolactin is another index of seizure adequacy. Grand mal seizures release brain peptides into the CSF and blood. Serum prolactin, easily measured, rises rapidly reaching a peak at about 25–30 min, and falls to a baseline level within 2 h. The absence of a dramatic rise in serum prolactin is a sign of inadequate treatment.[48]

The ECT course. Occasionally a single treatment relieves a disorder, but such instances are so rare as to be noteworthy. The basic course is more often between 6 and 20 treatments. These are usually given three times a week at the onset and, after the symptoms show some relief, are reduced to twice or once a week. The resolution of catatonia (MC, NMS) is frequently accomplished in three to five treatments but these are best administered daily. Depressive disorders require 6–12 treatments for resolution. Manic and psychotic disorders require 20 or more treatments.

Discontinuing treatment at the point of immediate resolution of symptoms is associated with high relapse rates. Continuation treatment, often continuation ECT, is as essential a part of ECT management as it is for pharmacotherapy.[49,50]

Continuation treatments. High relapse rates are the most common complaint in ECT practice. When patients are given a short course of treatments, early relapse is common. Because ECT is complex, frightening, and expensive, patients seek the shortest course of treatment, and physicians accede by prescribing a limited number of treatments on referral or at the time the patient signs the consent.[4]

Short courses of treatments may relieve symptoms but relapse is quick.[49,50] Continuation treatment is necessary. Two recent studies guide present practice. In a 3-hospital collaborative study of depressed patients referred for ECT, remitted patients were randomly assigned to 6-month courses of medication. Relapse rates were 84 per cent for placebo, 60 per cent for nortriptyline and 39 per cent for the combination of lithium and nortriptyline under serum level control.[50] In the 4-hospital collaborative study, depressed patients treated with bitemporal ECT were randomly assigned to continuation with ECT or the same lithium and nortriptyline combination. The 6-month relapse rates were 32 per cent for continuation medication and 37 per cent for continuation ECT.[49] These rates are statistically indistinguishable in the two studies.

ECT and psychotropic drugs.[12, 51] With the exception of antipsychotics, we lack evidence of synergy between psychotropic drugs and ECT.[11,12,14] TCA, MAOI, and SSRI antidepressants are usually discontinued during an ECT course. Anticonvulsants and sedative drugs affect seizure thresholds and may interfere with efficacy. ECT augmentation of antipsychotics is seen as safe and effective.

When ECT is administered to a patient with clinically effective serum lithium levels, generally seen as 0.8–1.2 mEq/l, there is the risk of a post-seizure delirium. If lithium treatment is sustained during ECT, the dosages are reduced so that the serum lithium levels do not rise above 0.6 mEq/l on treatment days.

Systemic drugs, especially those used to treat cardiovascular disorders, may put the patient at risk for hypotension, ataxia, or exaggerated cognitive deficits, but these effects can be easily managed, so they are usually continued during ECT.

Inducing adequate seizures in patients who have been receiving benzodiazepines may be difficult. Intravenous flumazenil, the benzodiazepine antagonist, effectively minimizes the inhibiting effects of benzodiazepines. Such use is encouraged for patients with catatonia or mania who have been treated with benzodiazepines.[9]

Risks and contraindications

Bone fractures, tardive seizures, and cardiac arrhythmias were common risks of early ECT, but the routine use of muscle relaxation with succinylcholine markedly reduced them.[8,12] Headache, tongue injury, and post-seizure delirium continue to be systemic risks. Headaches respond to analgesics, delirium to benzodiazepines, and tongue injury can be prevented by the proper application of bite-blocs. The principal risks of ECT today are cognitive effects and unacceptable relapse rates.

In a post-seizure delirium, which occurs in about 10 per cent of the treatments, the patient is poorly aware of where he is and may thrash about and be confused. It is more common in the first and second treatments than in later ones. Reassurance, calm talk, and gentle handling of movements that might be harmful can usually allay such states. If the restlessness does pose risks, it can be calmed by intravenous diazepam.

Persistent amnesia is the most dreaded risk of ECT.[12–15] Patients usually forget the personal events that occurred during the illness and treatment. On treatment days, both the anesthesia and the seizure alter cognition, temporarily interfering with the memory of events. In the first decades of ECT use, adequate ventilation was not assured and untoward effects on cognition were profound and frequent. But changes in practice have reduced these effects. Ventilation with pure oxygen, changes in the type of electrical current and the amounts of energy, and selected electrode placement reduced the effects on cognition, so that within a few weeks after the course is over, the patients' performance on memory tests usually surpasses their pre-treatment abilities.

Contraindications. There are no systemic illnesses that preclude the administration of ECT when the treatment is clearly warranted. Some conditions — severe hypertension, uncontrolled cardiac arrhythmia, bleeding tendencies, recent myocardial infarction, increased intracranial pressure, and a brain or cerebrovascular lesion — call for special care. The case literature offers suggestions for the appropriate treatment with ECT of patients with these conditions.[12–15]

Mechanism of action

When convulsive therapy was introduced, its most prominent side effect was amnesia, and much debate centered on whether amnesia

was, in fact, the mechanism for improving thought and mood. Experiments with different electrode placements discouraged this explanation.[4,12]

Others focused on the physiologic effects of seizures, especially the changes in the interseizure EEG. Such changes were found to be necessary, but not sufficient, for recovery.[8,52] Interest in this hypothesis is revived by recent studies.[53]

Explanations based on neurohumours and their receptors are important in our present views of the action of drug therapies. These are also cited to explain the benefits of ECT. The experimental data fail to support these explanations.[8,12,54]

Meduna thought that the concentration of glia was a factor in illness and that seizures elicited increased gliosis and recovery.[6] Recent reports cite increased neurogenesis as an active brain response to induced seizures.[55]

My view is that the neuroendocrine system is the most likely agent for the clinical changes brought about by induced seizures.[8,56] Neuroendocrine dysregulation is prominent in patients with the mental disorders for which ECT is effective. Thyroid, adrenal, sex gland, and hypothalamic dysfunction are common in patients with disorders in mood, thought, motor activity, feeding, sleep, sex, growth, and maturation. Indeed, every aspect of body physiology and mental activity is affected by these glands, as exemplified by the action of the adrenal glands in depressive mood disorders.

In the severely depressed patients, the adrenal glands produce too much cortisol.[17, 18, 56] The high blood levels disrupt the normal diurnal rhythms of other glandular discharges, and the glands do not respond to the usual feedback mechanisms. The most prominent features of depression — failure to eat, loss of weight, inability to sleep, loss of interest in sex, inability to concentrate thoughts, and difficulties in memory — are distortions of the functions regulated by the neuroendocrine glands in a self-adjusting feedback.

Each seizure stimulates the hypothalamus to discharge its hormones, which causes the pituitary gland to discharge its products, which then affects the level of cortisol. The first effects of this cascade are transitory, but repeated seizures restore the normal interactions of the hypothalamic-pituitary-adrenal axis. Feeding and sleep become normal, followed by motor activity, mood, memory, and thought.

How does a seizure elicit such profound changes in physiology? In ECT, the currents from the stimulating electrodes on each temple pass through the central parts of the brain, stimulating both the hypothalamus to discharge its hormones and the centrencephalic structures to produce a bilateral grand mal seizure. (One of the flaws in unilateral electrode placement is that the currents have to take indirect routes to affect the pertinent areas of the brain.) The massive amounts of hypothalamic and pituitary hormones that enter the bloodstream during ECT are measurable within a few minutes. They circulate throughout the body, affecting all the body's cells — a compelling and welcome sign of recovery.

After some courses of ECT, the return to normal endocrine function persists. At other times, the glands revert to their abnormal activities and the mental disorder becomes evident again. In these cases, repeated stimulation of the hypothalamus and the pituitary by continuation ECT restore and sustain normal glandular functions and support a normal mental state.

Suggested replacements for ECT

Although Meduna's experiments and numerous studies of ECT and sham ECT support the seizure as evidence of the brain changes essential to a therapeutic benefit, the introduction of electricity focussed attention away from the seizure and onto electricity as the medium for the treatment's efficacy. This interest is not new. Soon after Galvani and Volta demonstrated that electric currents could stimulate nerves and muscles, medical applications were enthusiastically sought. The first electrical experiments in the mentally ill are ascribed to Gale in New York State in 1802 and Aldini in Italy in 1803.[3,16,54] Electrical experiments were publicly demonstrated by Franklin, Mesmer, and Marat in the first years of the 19th century. Little benefit was recorded and most efforts are best considered quaint explorations.[3] At the time of World War I, faradization was a treatment for hysteria and applied in the military.[57]

During the second half of the 20th century, many techniques have been suggested as replacements for ECT, the latest being transcranial magnetic stimulation (rTMS), vagus nerve stimulation (VNS), and deep brain stimulation (DBS).[58] In rTMS rapidly alternating magnetic fields are delivered to stimulate the brain. At very high intensities, a seizure may be induced and some experiments have been undertaken to compare the seizure induced by magnetic currents with those induced by electric currents. The technique, called magnetic seizure therapy (MST), is reported to have a mild antidepressant effect.[58]

In VNS an electrical stimulator is implanted in the chest wall and electrodes are threaded through the neck to the left vagus nerve. The stimulator is similar to that used to reduce seizures in patients with severe epilepsy. The side effects of hoarseness, nausea, and vomiting are common. In DBS, the stimulating electrodes are placed in the brain, a technique occasionally used in severe Parkinsonism. We lack sufficient evidence for the efficacy of rTMS, VNS and DBS in psychiatric disorders to warrant their routine clinical use.

Device manufacturers who seek a market for their products encourage the technologies. The bad image and the stigma of ECT make its replacement the basis for exploration. At the time of this writing (Spring, 2007), no evidence has been published that any of these techniques have persistent therapeutic effects, and none are replacements for ECT.

The future in ECT

Induced seizures effectively allay severe psychiatric disorders. The treatment's stigma, however, inhibits its use and research into its mechanism. When neuroscientists recognize the unique nature of the seizure — a phenomenon that is ubiquitous in animal life — and seek to understand its biology, they will then seek ways to replace the gross process of induced seizures by more acceptable interventions. Understanding the mechanism will clarify the aetiology of psychiatric disorders. ECT will be replaced when we understand its mechanism better; for the present, continued usage is assured since no alternative intervention with its efficacy and safety is in our *materia medica*.

Further information

Abrams, R. (2001). *Electroconvulsive Therapy* (4th edn.). Oxford University Press, New York

Fink, M. (1999). *Electroshock: Restoring the Mind*. Oxford University Press, New York.

Taylor, M.A. Fink, M. (2006). *Melancholia: Diagnosis, Pathophysiology and Treatment of Depressive Illness*. Cambridge University Press, Cambridge UK.

American Psychiatric Association. (2001). *Electroconvulsive Therapy: Recommendations for Treatment, Training and Privileging*. Washington DC.

Scott, A.I.F. (Ed.) (2004). *The ECT Handbook*, (2nd edn.). Royal College of Psychiatrists, London.

Ottosson, J-,O, Fink, M. (2004). *Ethics in Electroconvulsive Therapy. Brunner-Routledge,* New York.

Shorter, E. and Healy, D. (2007). H*istory of the Shock Therapies*. Rutgers University Press, New Brunswick, NJ.

References

1. Fink, M. (1991). Impact of the antipsychiatry movement on the revival of electroconvulsive therapy in the United States. *Psychiatric Clinics of North America*, **14**(4), 793–801.
2. Fink, M. (1997). Prejudice against ECT: competition with psychological philosophies as a contribution to its stigma. *Convulsive Therapy*, **13**(4), 253–65; discussion 66–8.
3. Kneeland, T.W., and Warren, C.A.B. (2002). *Pushbutton psychiatry: a history of electroshock in America*. Westport, Conn.: Praeger.
4. Ottosson, J.-O., Fink, M. (2004). *Ethics in electroconvulsive therapy.* Brunner–Routledge, New York.
5. Shorter, E., and Healy, D. (2007). *Shock Therapy: The History of Electroconvulsive Treatment in Mental Illness.* Rutgers UP; New Brunswick, in press.
6. Kellner, C.H., Fink, M., Knapp, R., *et al.* (2005). Relief of expressed suicidal intent by ECT: a consortium for research in ECT study. *American Journal of Psychiatry*, **162**(5), 977–82.
7. Meduna, L. (1985). Autobiography. *Convulsive Therapy*, **1,** 43–57; 121–38.
8. Fink, M. (1979). *Convulsive therapy: theory and practice.* Raven Press, New York.
9. Fink, M., and Taylor, M.A. (2003). *Catatonia: a clinician's guide to diagnosis and treatment.* Cambridge University Press, Cambridge, New York.
10. Parker, G., and Hadzi–Pavlovic, D. (1996). *Melancholia: a disorder of movement and mood: a phenomenological and neurobiological review.* Cambridge Unniversity Press, Cambridge, New York, USA.
11. Taylor, M.A., Fink, M. (2006). *Melancholia: the diagnosis, pathophysiology, and treatment of depressive illness.* Cambridge; New York: Cambridge University Press; 2006.
12. Abrams, R. (2002). *Electroconvulsive therapy. (*4th ed.) Oxford Unversity Press, Oxford, New York.
13. American Psychiatric Association. (2001). *Committee on Electroconvulsive Therapy.* Weiner, R.D. The practice of electroconvulsive therapy: recommendations for treatment, training, and privileging: a task force report of the American Psychiatric Association. (2nd edn.) American Psychiatric Association, Washington, DC.
14. Fink, M. (1999). *Electroshock: restoring the mind.* Oxford University Press, New York.
15. Scott, Ae. (2004). *The ECT Handbook*. (2nd edn.) London: Royal College of Psychiatrists.
16. Shorter, E. (1997). *A history of psychiatry: from the era of the asylum to the age of Prozac.* John Wiley & Sons, New York.
17. Carroll, B.J., Curtis, G.C., Mendels, J., *et al.* (1976). Neuroendocrine regulation in depression. II. Discrimination of depressed from nondepressed patients. *Archives of General Psychiatry*, **33**(9), 1051–8.
18. Fink, M. (2005). Should the dexamethasone suppression test be resurrected? *Acta Psychiatrica Scandinavica*, **112**(4), 245–9.
19. Avery, D., and Lubrano, A. (1979). Depression treated with imipramine and ECT: the DeCarolis study reconsidered. *American Journal of Psychiatry*, **136**(4B), 559–62.
20. Kroessler, D. (1985). Relative Efficacy Rates for Therapies of Delusional Depression. *Convulsive Therapy*, **1**(3), 173–82.
21. Flint, A.J., and Rifat, S.L. (1998). Two–year outcome of psychotic depression in late life. *American Journal of Psychiatry*, **155**(2), 178–83.
22. Mulsant, B.H., Haskett, R.F., Prudic, J., *et al.* (1997). Low use of neuroleptic drugs in the treatment of psychotic major depression. *American Journal of Psychiatry*, **154**(4), 559–61.
23. Rasmussen, K.G., Mueller, M., Kellner, C.H., *et al.* (2006). Patterns of psychotropic medication use among patients with severe depression referred for electroconvulsive therapy: data from the Consortium for Research on Electroconvulsive Therapy. *Journal of ECT*, **22**(2), 116–23.
24. Husain, M.M., Rush, A.J., Fink, M., *et al.* (2004). Speed of response and remission in major depressive disorder with acute electroconvulsive therapy (ECT): a Consortium for Research in ECT (CORE) report. *Journal of Clinical Psychiatry*, **65**(4), 485–91.
25. Petrides, G., Fink, M., Husain, M.M., *et al.* (2001). ECT remission rates in psychotic versus nonpsychotic depressed patients: a report from CORE. *Journal of ECT*, **17**(4), 244–53.
26. Goodwin, F.K., and Jamison, K.R. (1990). *Manic–depressive illness*. Oxford University Press, New York.
27. Fink, M. (1999). Delirious mania. *Bipolar Disorders*, **1**(1), 54–60.
28. Mukherjee, S., Sackeim, H.A., Schnur, D.B., *et al.* (1994). Electroconvulsive therapy of acute manic episodes: a review of 50 years'experience. *American Journal of Psychiatry*, **151**(2), 169–76.
29. Black, D.W., Winokur, G., Nasrallah, A., *et al.* (1987). Treatment of mania: a naturalistic study of electroconvulsive therapy versus lithium in 438 patients. *Journal of Clinical Psychiatry*, **48**(4), 132–9.
30. American Psychiatric Association. (1994). *Task Force on DSM–IV. Diagnostic and statistical manual of mental disorders: DSM–IV.* (4th edn.) Washington, DC: American Psychiatric Association.
31. Taylor, M.A., and Fink, M. (2003). Catatonia in psychiatric classification: a home of its own. *American Journal of Psychiatry*, **160**(7), 1233–41.
32. Fink, M., and Taylor, M.A. (2001). The many varieties of catatonia. *European Archives of Psychiatry and Clinical Neuroscience.*, **251** (Suppl 1), I8–13.
33. Caroff, S.N. (2004). *Catatonia: from psychopathology to neurobiology.* (1st ed.) American Psychiatric Pub, Washington, DC.
34. Fink, M., and Sackeim, H.A. (1996). Convulsive therapy in schizophrenia? *Schizophrenia Bulletin*, **22**(1), 27–39.
35. Fink, M. (2000).The interaction of delirium and seizures. *Seminars in Clinical Neuropsychiatry*, **5**(2):93–7.
36. Lowenstein, D.H., and Alldredge, B.K. (1998). Status epilepticus. *New England Journal of Medicine.*, **338**(14), 970–6.
37. Fink, M., Kellner, C.H. and Sackeim, H.A. (1999). Intractable seizures, status epilepticus, and ECT. *Journal of ECT*, **15**(4), 282–4.
38. Fall, P.A., Ekman, R., Granerus, A.K., *et al.* (1995). ECT in Parkinson's disease. Changes in motor symptoms, monoamine metabolites and neuropeptides. *Journal of Neural Transmission. Parkinson's Disease and Dementia Section*, **10**(2–3), 129–40.
39. Avery, D., and Winokur, G. (1978). Suicide, attempted suicide, and relapse rates in depression. *Archives of General Psychiatry*, **35**(6), 749–53.
40. Prudic, J., and Sackeim, H.A. (1999). Electroconvulsive therapy and suicide risk. *Journal of Clinical Psychiatry*, **60** Suppl 2, 104–10; discussion 11–6.
41. Kane, J., Honigfeld, G., Singer, J., *et al.* (1988). Clozapine for the treatment–resistant schizophrenic. A double–blind comparison with chlorpromazine. *Archives of General Psychiatry*, **45**(9), 789–96.
42. Rey, J.M., and Walter, G. (1997). Half a century of ECT use in young people. *American Journal of Psychiatry*, **154**(5), 595–602.

43. Fink, M. (1986). Informed ECT for Patients and Families. Lake Bluff: Somatics, Inc.
44. Folk, J.W., Kellner, C.H., Beale, M.D., *et al.* (2000). Anesthesia for electroconvulsive therapy: a review. *Journal of ECT*, **16**(2), 157–70.
45. Abrams, R. (2002). Stimulus titration and ECT dosing. *Journal of ECT*, **18**(1), 3–9; discussion 14–5.
46. McCall, W.V., Reboussin, D.M., Weiner, R.D., *et al.* (2000). Titrated moderately suprathreshold vs fixed high–dose right unilateral electroconvulsive therapy: acute antidepressant and cognitive effects. *Archives of General Psychiatry*, **57**(5), 438–44.
47. Petrides, G., and Fink, M. (1996). The 'half–age' stimulation strategy for ECT dosing. *Archives of General Psychiatry*, **12**(3), 138–46.
48. Abrams, R., and Swartz, C. (1990). *The Technique of ECT*. Lake Bluff: Somatics, Inc.
49. Kellner, C.H., Knapp, R.G., Petrides, G., *et al.* (2006). Continuational ectroconvulsive therapy vs pharmacotherapy for relapse prevention in major depression: a multisite study from the Consortium for Research in Electroconvulsive Therapy (CORE). *Archives of General Psychiatry*, **63**(12), 1337–44.
50. Sackeim, H.A., Haskett, R.F., Mulsant, B.H., *et al.* (2001). Continuation pharmacotherapy in the prevention of relapse following electroconvulsive therapy: a randomized controlled trial. *Journal of the American Medical Association*, **285**(10), 1299–307.
51. Fink, M., and Kellner, C.H. (1993). ECT and Drugs: Concurrent Administration. *Convulsive Therapy*, **9**(4), 237–40.
52. Fink, M., and Kahn, R.L. (1957). Relation of electroencephalographic delta activity to behavioral response in electroshock; quantitative serial studies. *A. M. A. Archives of Neurology and Psychiatry*, **78**(5), 516–25.
53. Sackeim, H.A., Luber, B., Katzman, G.P., *et al.* (1996). The effects of electroconvulsive therapy on quantitative electroencephalograms. Relationship to clinical outcome. *Archives of General Psychiatry*, **53**(9), 814–24.
54. Shorter, E., and Healy, D. (2007). *History of the Shock Therapies.* Rutgers University Press, New Brunswick.
55. Fink, M. (2004). Induced seizures as psychiatric therapy: Ladislas Meduna's contributions in modern neuroscience. *Archives of General Psychiatry*, **20**(3), 133–6.
56. Fink, M. (2000). Electroschock revisited. *American Scientist*, **88**(2), 162–7.
57. Eissler, K.R. (1986). *Freud as an expert witness: the discussion of war neuroses between Freud and Wagner-Jauregg.* International Universities Press, New York.
58. Lisanby, S.H. (2004). *Brain stimulation in psychiatric treatment.* American Psychiatric Pub, Washington, DC.

6.2.10.2 Phototherapy

Philip J. Cowen

Introduction

Phototherapy or artificial bright-light treatment, has been used in the management of a number of medical disorders including psoriasis and hyperbilirubinaemia of the newborn. From the point of view of psychiatric treatment, the notion that light might help people with certain psychological symptoms has an ancient lineage. For example, Wehr and Rosenthal[1] cite Aretaeus who suggested in the second century AD that 'lethargics are to be laid in the light and exposed to the rays of the sun (for the disease is gloom)'. In 1898, a ship's physician named Frederick Cook recorded that the 'languor' which affected members of an Antarctic expedition during the winter darkness could be relieved with bright artificial light.[1]

The first systematic study of phototherapy as a psychiatric treatment was carried out in 1984 by Rosenthal *et al.*[2] who used bright artificial light to treat patients with the newly identified syndrome of seasonal affective disorder. Seasonal affective disorder is a recurrent mood disorder in which patients experience regular episodes of depression in autumn and winter with remission in spring and summer. Since then phototherapy has become the mainstay of the treatment of seasonal affective disorder, particularly in patients with atypical depressive features such as hyperphagia and hypersomnia. Phototherapy has also been used as an investigational treatment in other psychiatric disorders but the evidence for its efficacy in these conditions less established.

Mechanism of action

Light and seasonal and circadian rhythms

Animals and humans show circadian and seasonal rhythms in aspects of their physiology and behaviour that are influenced by environmental cues or *zeitgebers*. The light–dark cycle is believed to be one of the most important *zeitgebers* regulating circadian and seasonal rhythmicity in mammals. Mammalian circadian rhythms are driven by an 'oscillator' in the suprachiasmatic nucleus of the hypothalamus. Environmental light influences the activity of this nucleus via a neuronal pathway which runs from the retina to the hypothalamus. Thus appropriately timed bright light is able to advance or delay endogenous circadian rhythms.[3]

Lewy *et al.*[4] suggested that in patients with seasonal affective disorder the delayed onset of dawn in the autumn causes endogenous circadian rhythms to become phase-delayed with respect to clock time and the sleep–wake cycle. Bright-light treatment is able to correct this abnormality by phase advancing circadian rhythms, thereby re-synchronizing them with the sleep–wake cycle. This proposal is supported by the fact that controlled trials show that in most patients morning phototherapy is more effective than evening phototherapy.[5] While this hypothesis gives a good account of how bright-light treatment might ameliorate the symptoms of seasonal affective disorder, its possible efficacy in other conditions such as non-seasonal depression is difficult to explain by this mechanism.

Light treatment and monoamines

It is possible that bright-light treatment, through its interaction with the hypothalamus, could alter the circadian activity of the monoamine neurotransmitters involved in mood regulation. For example, some studies have shown that the antidepressant effects of phototherapy can be reversed by treatments that diminish both catecholamine and serotonin neurotransmission.[6] This has been taken as evidence that the antidepressant effects of bright light are mediated via activation of serotonin and catecholamine pathways. An alternative explanation is that in the absence of concomitant drug treatment, recovered depressed patients are vulnerable to depletion of these neurotransmitters in any case. However, effects of bright light on monoamines could account for the therapeutic effects of light in mood disorders other than winter depression.

Forms of phototherapy

The most common form of phototherapy uses a light box, which contains fluorescent tubes mounted behind a translucent plastic-diffusing screen. Depending on the fluorescent tubes employed, the light emitted is either full spectrum, which contains a little ultraviolet light, or cool white light which has no ultraviolet. The light box usually rests on a table or desk at about the eye level of a seated subject. The output of different light boxes varies but is usually between 2500 and 10 000 lux. Light sources producing 10 000 lux are more expensive but allow a reduced duration of exposure (30 min compared with 120 min) to secure a therapeutic effect.[7]

Phototherapy has also been administered using head-mounted units or light visors. These instruments are attached to the head and project light into the eyes allowing subjects to remain mobile while receiving treatment. While light visors are more convenient to use than light boxes, results from placebo-controlled trials have not been encouraging.[7]

Another form of light therapy involves the use of dawn-simulating alarm clocks. These clocks are programmed to simulate the illumination that would be experienced out of doors during sunrise on a spring day.[8] In practice, the clocks begin a gradual illumination of the bedroom about 2 h before normal wake time, increasing to a maximum of about 250 lux at the point of waking. Overall the effects of dawn-simulation in the treatment of winter depression seem equivalent to those of bright-light treatment[9] and patients often find dawn-simulating clocks more convenient (although a partner sleeping in the same room may not).

Adverse effects

Generally phototherapy is well tolerated although mild side effects occur in up to 45 per cent of patients early in treatment. These include headache, eye strain, blurred vision, eye irritation, and increased tension. Insomnia can occur particularly with late-evening treatment. Rare adverse events that have been reported include manic mood swings and suicide attempts, the latter putatively through light-inducing alerting and energizing effects prior to mood improvement. Whether these rare events are actually adverse reactions to the light is uncertain. There is no evidence that phototherapy employed in recommended treatment schedules causes ocular or retinal damage.

Indications and contraindications to light treatment

Seasonal affective disorder

The best established indication for light treatment is seasonal affective disorder where patients experience autumn and winter depressions. Clinical predictors of a response to light treatment include the following:

- Increased sleep
- Increased appetite and winter weight gain
- Carbohydrate craving
- Afternoon slump in energy
- Complete remission of symptoms in the summer

Several controlled trials have assessed the efficacy of bright-light treatment in the treatment of winter depression. In a meta-analysis of nine randomized studies, Golden *et al.*[9] found a significant benefit of bright light over dim light control with an effect size of 0.84 (95 per cent confidence interval, 0.6–1.08). A similar benefit was apparent for six studies of dawn simulation which had a mean effect size of 0.73 (95 per cent confidence interval, 0.37–1.08). While these data are compelling it needs to be remembered that it is often difficult to arrange a placebo treatment that will match the therapeutic expectation of bright light or dawn simulation.

Other mood disorders

Patients with more typical melancholic symptoms (e.g. weight loss and insomnia) do less well with bright-light treatment, even when the disorder is seasonal in nature. However, bright light has also been used in the treatment of non-seasonal depression both as a sole treatment as an adjunct to more conventional therapy. The evidence for the efficacy of bright light for this indication is less established but a Cochrane review[10] suggested that morning light treatment was significantly better than control treatment when applied as an adjunct to drug treatment or sleep deprivation. Most of these studies were of short-term duration and there are suggestions that the added benefit of light therapy does not persist when treatment stops.[11] In these studies, hypomania was more common in light-treated subjects. Phototherapy may also be of benefit in other conditions characterized by depressed mood and overeating (e.g. premenstrual dysphoria and bulimia nervosa). The literature contains reports of a number of controlled trials in such disorders where light treatment has improved depression ratings. However, the difficulty of distinguishing the specific and placebo effects of bright-light treatment relative to dim light control makes the current data difficult to interpret.

Circadian rhythm disorders

Because bright light is an effective *zeitgeber* for circadian rhythms it may also have a useful place in the treatment of disorders characterized by circadian rhythm disturbances. Such disorders encompass a range of conditions including phase-delayed or phase-advanced sleep disorder, jet lag, and problems related to shift work. In addition, disturbances of the sleep–wake cycle are common in older people with cognitive impairment. There are several reports of the utility of light treatment in these conditions; however, there is a paucity of randomized trial data.[7]

Contraindications

There are no absolute contraindications to phototherapy, except the obvious caveat that since the therapeutic effect depends on retinal activation, subjects must have sufficient visual function to allow this to occur. Otherwise it would seem prudent to avoid phototherapy in patients with pronounced and untreated agitation because this symptomatology could be worsened. In addition, evening phototherapy may worsen insomnia.

A substantial minority of patients with seasonal affective disorder meet criteria for bipolar II disorder, raising the concern that phototherapy may trigger hypomania in such individuals. Particular caution might be needed in patients with a bipolar I syndrome. Some regimes of phototherapy might lead to a degree of sleep deprivation which could also destabilize mood in bipolar patients.

Interactions

One of the advantages of phototherapy in seasonal affective disorders is that the use of antidepressant drugs may be avoided. Despite this, many patients with winter depression use phototherapy concomitantly with antidepressant medication without an obvious potentiation of adverse effects. However, a case report described apparent serotonin toxicity where phototherapy was combined with selective serotonin re-uptake inhibitors.[7]

Like bright light, the pineal hormone, melatonin, also has the ability to shift the timing of circadian rhythms[2] and theoretically melatonin taken at an inappropriate time of day could offset the antidepressant effect of light. It is also possible that bright-light treatment could exacerbate the ability of some drugs (e.g. chlorpromazine, St John's Wort) to cause skin photosensitivity reactions.[7]

Effects of withdrawal of phototherapy

If a patient with seasonal affective disorder responds to light treatment, withdrawal of treatment during the period of seasonal vulnerability leads to a return of symptomatology within a few days. It may be possible, however, to lessen the daily duration of treatment particularly towards the end of winter without inducing relapse. Otherwise cessation of light treatment does not seem to cause a specific withdrawal syndrome.

Administration of phototherapy

Since the best established indication for phototherapy in psychiatry is seasonal affective disorder, the following account will describe the use of bright-light treatment in winter depression. One of the major practical difficulties in phototherapy is the time needed to administer treatment. For this reason a 10 000 lux light box may be preferred because the daily duration of treatment can be reduced to 30 min. It seems likely that cool-white light and full-spectrum light have equivalent clinical efficacy, but because cool-white light is free of ultraviolet light it is theoretically safer and should be preferred.

The balance of evidence suggests that bright-light treatment of winter depression is most effective when administered in the early morning.[4] However, treatment given later in the day may be effective for some patients. In an initial trial, therefore, it is best to recommend early morning treatment but to advise the patient that the timing of therapy can eventually represent a balance of therapeutic efficacy and practical convenience. Treatment in the late evening should be avoided because of the possibility of sleep disruption.

Early-morning phototherapy should start a few minutes after waking. Subjects should allow themselves a 30 min duration of treatment with a 10 000 lux light source. They should seat themselves about 30 to 40 cm away from the light box screen. They should not gaze at the screen directly but face it an angle of 45° and glance across it once or twice each minute.

The antidepressant effect of light treatment usually appears in a few days but in controlled trials up to 3 weeks can be needed before the therapeutic effects of bright light exceed those of placebo treatment. If no benefit is noted after the third week of therapy, light treatment should probably be abandoned. As noted above mild side effects are common in the early stages of treatment but usually settle without specific intervention. If they are persistent and troublesome the patient can sit a little further away from the light source or reduce the duration of exposure. Exposure should also be reduced or stopped if elevated mood occurs.

Once a therapeutic response has occurred it is usually necessary to continue phototherapy up to the usual time of natural remission, otherwise relapse will occur. It may be possible, however, to lower the daily duration of treatment. Phototherapy can also be started in advance of the anticipated episode of depression as this may have a preventative effect; however, the evidence for this is limited and doubts have been expressed.[7,12]

Further information

Centre for Environmental Therapeutics. http://www.cet.org.

Seasonal Affective Disorder Association. http://www.sada.org.uk.

Eagles, J.M. (2003). Reading about seasonal affective disorder. *The British Journal of Psychiatry*, **182**, 174–6.

References

1. Wehr, T.A. and Rosenthal, N.E. (1989). Seasonality and affective illness. *The American Journal of Psychiatry*, **146**, 829–39.
2. Rosenthal, N.E., Sack, D.A., Gillin, C., *et al.* (1984). Seasonal affective disorder: a description of the syndrome and preliminary findings with light therapy. *Archives of General Psychiatry*, **41**, 72–80.
3. Arendt, J. and Broadway, J. (1987). Light and melatonin as *zeitgebers* in man. *Chronobiology International*, **4**, 273–82.
4. Lewy, A.J., Sack, R.L., Miller, S.L., *et al.* (1987). Antidepressant and phase-shifting effects of light. *Science*, **206**, 710–13.
5. Terman, J.S., Terman, M., Lo, E.S., *et al.* (2001). Circadian time of morning light administration and therapeutic response in winter depression. *Archives of General Psychiatry*, **58**, 69–75.
6. Neumeister, A., Turner, E.H., Matthews, J.R., *et al.* (1998). Effects of tryptophan depletion versus catecholamine depletion in patients with seasonal affective disorder in remission with light therapy. *Archives of General Psychiatry*, **55**, 524–30.
7. Eagles, J.M. (2004). Light therapy and the management of winter depression. *Advances in Psychiatric Treatment*, **10**, 233–40.
8. Terman, M. and Terman, J.S. (2006). Controlled trial of naturalistic dawn simulation and negative air ionization for seasonal affective disorder. *The American Journal of Psychiatry*, **163**, 2126–33.
9. Golden, R.N., Gaynes, B.N., Ekstrom, R.D., *et al.* (2005). The efficacy of light therapy in the treatment of mood disorders: a review and meta-analysis of the evidence. *The American Journal of Psychiatry*, **162**, 656–62.
10. Tuunainen, A., Kripke, D.F., and Endo, T. (2007). Light therapy for non-seasonal depression. http://www.cochrane.org.reviews/en/ab004050.html.
11. Martiny, K., Lunde, M., Unden, M., *et al.* (2006). The lack of sustained effect of bright light in non-seasonal major depression. *Psychological Medicine*, **36**, 1247–52.
12. Partonen, T. and Lonnqvist, J. (1996). Prevention of winter seasonal affective disorder by bright-light treatment. *Psychological Medicine*, **26**, 1075–80.

6.2.10.3 Transcranial magnetic stimulation

Declan McLoughlin and Andrew Mogg

Introduction

Transcranial magnetic stimulation (TMS) is a means of non-invasively stimulating the cerebral cortex using a hand-held coil applied to the scalp. In recent years TMS has been increasingly used to target neuronal circuitry implicated in neuropsychiatric disorders.

A key milestone in the development of TMS occurred in 1831 when Michael Faraday discovered the phenomenon of electromagnetic induction whereby a time-varying magnetic field can induce electrical currents through a conductor lying in proximity to the field. The French biophysicist D'Arsonval in 1896 induced phosphenes, vertigo, and syncope in human subjects by placing an induction coil around their heads. In the late 1950s, Kolin stimulated peripheral nerves (the frog sciatic nerve) with a magnetic field and a few years later the same technique was used in human subjects, inducing muscle twitching by applying a pulsed magnetic field over the ulnar, peroneal, and sciatic nerves.

In the mid-1980s, Barker and colleagues in Sheffield developed a magnetic stimulator to directly stimulate the human motor cortex.[1] They applied a circular coil, through which a large (4000 A), brief (110 μs) current was passed to the scalp. The resulting pulsed magnetic field was used to stimulate the motor cortex, evoking movements in the contralateral limbs and is known as transcranial magnetic stimulation. This ability to non-invasively stimulate the motor cortex with a magnetic field soon replaced high-voltage transcutaneous electrical stimulation for assessing central motor conduction times and mapping corticospinal pathways in a variety of neurological conditions. In the late 1980s machines capable of delivering multiple TMS pulses were developed. Repetitive TMS (rTMS), unlike single pulses of TMS can produce effects that last after the period of stimulation. For example, it has been shown that rapid rTMS (at frequencies of 5 Hz and greater) enhances motor excitability whereas slow rTMS (at 1 Hz or less) transiently depresses excitability.[2] The underlying principle of rTMS treatment is that the normal balance of excitatory and inhibitory processes within certain neuronal pathways may be disrupted in psychiatric conditions such as depression. Stimulating the brain using rTMS provides a means of increasing and decreasing excitation and inhibition in these pathways, having a neuromodulatory effect and allowing a focal targeting of specific neuronal circuitry.

Mechanism of action

The underlying mechanisms of the effects of rTMS remain poorly understood. This is in part because, as with attempting to understand the mode of action of psychotropic medication, it is difficult to establish links between cellular and physiological changes and alterations in emotion, thinking, and behaviour. Techniques used to try to better understand the molecular and physiological effects of rTMS have included neuroimaging and animal studies.

Neuroimaging has demonstrated that rTMS may exert effects on the brain at a considerable distance from the site of stimulation. For example, serial positron emission tomography scanning has been used to measure regional cerebral blood flow in medication-free patients with major depression before and after courses of fast and slow rTMS administered over the left prefrontal cortex.[3] It has been demonstrated that fast rTMS causes increases in regional cerebral blood flow in bilateral frontal, limbic, and paralimbic areas whereas slow rTMS caused decreases in blood flow in the right prefrontal cortex, left medial temporal cortex, and left basal ganglia and amygdala.

It has been suggested that rTMS of the left prefrontal cortex may modulate brain function by an effect on dopamine release. Elevated extracellular dopamine concentrations in the dorsal hippocampus have been demonstrated in the brains of rats who received rTMS. However, one of the problems with using animal models of rTMS is that currently small rTMS coils are not available and it is therefore impossible to focally stimulate one particular area of the small rodent brain. In humans it has been shown that rTMS to the dorsolateral prefrontal cortex can induce the release of dopamine in the ipsilateral caudate nucleus.[4]

Side-effects

Being non-invasive and not requiring a general anaesthetic, rTMS is considered to be a relatively safe treatment and few side-effects have been reported. The most significant potential side-effect is the risk of unintended seizure induction. There have been six reports to date of seizure induction in healthy volunteers. In half of these, very high stimulation intensities and frequencies were used. There have only been three reports of seizures in patients receiving rTMS and one of these patients had a pre-existing diagnosis of temporal lobe epilepsy. Researchers generally follow safety guidelines that exclude high-risk patients, (e.g. those with a stroke, brain tumour or pre-existing epilepsy) from receiving rTMS. These guidelines also suggest limits to the intensity, frequency, and stimulus duration of the rTMS used.[5]

The most common side-effect of rTMS is headache or facial discomfort that is the result of direct stimulation of muscle and nerves in proximity to the coil. Approximately 10–30 per cent of subjects experience these symptoms, which are generally short-lived and well-tolerated.

Technique

rTMS equipment comprises a stimulator unit, booster modules, a laptop computer, and a figure-of-eight coil. The stimulator unit contains the charging circuitry, energy storage capacitors, control electronics and discharge, and safety circuitry. It is connected to booster modules, which charge the high-voltage capacitors, enabling trains of high-intensity magnetic stimulation to be produced. The stimulating coil consists of tightly wound copper wire in a figure-of-eight through which a rapidly alternating electric current passes to produce a pulsed magnetic field. Various stimulation parameters including train duration, frequency of stimulation, stimulus intensity, and length of inter-train interval can be altered using computer software.

rTMS treatment is delivered via the figure-of-eight coil applied to the scalp surface. Typically, prior to treatment, TMS will be used

to map the motor area of the right abductor pollicis brevis (APB), and measure its motor threshold. The stimulus intensity delivered during treatment is then calculated in relation to this motor threshold. The main method of localizing the stimulation site has been to use a fixed point in anatomical relation to a specific motor area, for example the dorsolateral prefrontal cortex has generally been defined as the point 5 cm anterior to the APB motor area in the parasaggital plane. More recently some studies have used magnetic resonance imaging to more accurately delineate the area to be stimulated.

rTMS and depression

Transcranial magnetic stimulation was first postulated to have potential applications in psychiatry by Bickford and colleagues who noted transient elevation in mood in several healthy subjects who had received single pulses of TMS to the motor cortex.[6] Several small open studies followed that suggested that low frequency rTMS over the vertex may have antidepressant effects. Since the mid 1990s most interest has focussed on high-frequency rTMS applied to the left dorsolateral prefrontal cortex (LDLPFC), a region reported to be underactive in depression. To date there have been approximately 30 randomized trials of real and placebo rTMS in depression. In addition there have been several published meta-analyses including a Cochrane review.[7] This reviewed 16 trials, 14 of which were suitable for quantitative analysis. They found that high-frequency rTMS to the LDLPFC and low-frequency rTMS to the right dorsolateral prefrontal cortex (RDLPFC) were both superior to sham treatment but only for one measure (the Hamilton Depression Rating Scale) and at one time point (immediately after 2 weeks of treatment). The difference between real and sham was not large, leading to the conclusion that at this stage there was not strong evidence to support the use of rTMS as an antidepressant therapy.

There has been considerable heterogeneity between studies. Nearly all the trials have comprised patients with major depressive disorder defined using DSM-IV criteria. However there has been considerable variability with respect to pharmacotherapy received, with some trials specifying treatment resistance (variously defined) and some specifying medication-free participants. In two of the studies patients were started on antidepressant treatment either shortly before or simultaneously with the rTMS treatment.

The choice of appropriate sham condition is an important methodological consideration. There are two main approaches. Most studies have relied on tilting the active coil (usually through 45° or 90° with one or both wings of the coil touching the scalp). However intracerebral voltage measurements in a rhesus monkey have shown that, depending on how the coil is tilted, sham conditions obtained by coil tilting can induce voltages in the brain to levels only 24 per cent below active rTMS.[8] The fact that some 'sham' coils produce significant cortical stimulation may account for some of the benefit seen in those receiving sham stimulation and may underestimate the difference between real and placebo treatment. The other approach is to use a specially designed placebo coil. This looks identical to the real coil and makes the same noise but does not cause any cortical stimulation. However, neither does it cause sensation to the scalp, meaning that subject blinding may still be less effective. Indeed the problem of maintaining blinding in studies with rTMS continues to be a major methodological issue.

Most studies have given high-frequency rTMS to the LDLPFC, probably as a result of the positive early studies when this area was stimulated. Several investigators have used low-frequency rTMS to the RDLPFC. Low-frequency rTMS is much less likely to induce a seizure and is probably better tolerated by patients. Since slow rTMS has an inhibitory effect in contrast to the excitatory effect of fast rTMS and since there is considerable evidence that the left and right hemispheres have contrasting functions in regulating mood, it could be speculated that slow rTMS to the right cortex may have a similar effect to fast rTMS to the left.

Studies of high-frequency rTMS in depression have generally used stimulation frequencies of 5 to 20 Hz. There is a suggestion from animal studies that higher frequency stimulation may have a greater antidepressant effect but so far the numbers of subjects in human studies have been too low to show if a difference in effect of varying stimulation frequency exists. Likewise, the optimal stimulus intensity, length of treatment course, and total number of stimulations is not yet clear from the published data. However, longer trials with an increase number of stimulations appear to make little difference.[9]

Most of the rTMS studies in depression have been small, the largest until recently having 70 patients. However, recently a much larger industry-sponsored (Neuronetics) trial submitted their findings to the US Food and Drug Administration (FDA), seeking licensing approval for an rTMS device. In this study 301 patients were randomized to real or sham rTMS. Participants received 10 Hz rTMS of the left dorsolateral prefrontal cortex, 3000 pulses per day for 20 days. Although there was a marginal difference between the groups in favour of rTMS at the end of treatment, there was no significant group difference on an intention-to-treat analysis of the primary outcome measure (Montgomery-Åsberg Depression Rating Scale). In January 2007 the FDA Neurological Devices panel considered Neuronetic's application to have its rTMS equipment licensed for therapeutic use. The panel felt that there was insufficient evidence to support its efficacy. The final FDA decision is expected in summer 2007 (website: http://www.fda.gov/cdrh/panel/summary/neuro-012607.html, accessed: 5 June 2007).

In the United Kingdom the National Institute for Clinical Excellence (NICE) has issued recommendations stating 'Current evidence suggests there are no major safety concerns associated with transcranial magnetic stimulation for severe depression but there is no evidence that the procedure has clinically useful efficacy' (website: http://www.nice.org.uk/article.aspx?o=ip346consultation, accessed: 5 June 2007).

Comparisons with ECT

In addition to comparisons with placebo treatment, rTMS has also been directly compared with ECT in several studies. While ECT is the most effective treatment for severe depression in the short-term its use is limited by several issues, including acceptability to patients, the requirement to be anaesthetized, and the occurrence of side-effects, particularly cognitive side-effects. rTMS could be a potential alternative if it proved effective. In total there have been six published randomized controlled trials to date comparing ECT and rTMS. These trials have all had relatively small numbers of

patients, particularly when compared with trials of antidepressant medications. They have either shown rTMS to be less effective or not statistically different from ECT. The most recent and largest trial to date included 46 patients and compared 3 weeks of treatment with rTMS to a course of ECT.[10] The mean reduction in the Hamilton Depression Rating Scale achieved at the end of treatment was 14.1 points in the ECT group, compared with 5.4 points for the rTMS group, translating into a mean percentage reductions from baseline of 58 and 22 per cent, respectively. Overall, ECT was shown to be substantially more effective as a short-term treatment of depression than rTMS.

rTMS and schizophrenia

While most studies of rTMS within psychiatry have focussed on depression, there has been a growing interest in using rTMS as a possible treatment for schizophrenia. It has been used to treat both auditory hallucinations and to alleviate negative symptoms of schizophrenia.

Auditory hallucinations occur in approximately 70 per cent of patients with schizophrenia and in about a quarter of cases respond poorly if at all to antipsychotic medication. Recent advances in neuroimaging have enabled measurement of neural activity while hallucinations are being experienced and it has been demonstrated that auditory hallucinations are associated with activation in a number of brain areas, including the temporal cortex bilaterally. This area has been targeted in several rTMS studies using slow rTMS to reduce excitability.

The first account of using rTMS to treat auditory hallucinations reported improvement in the severity of hallucinations of three patients with schizophrenia who had 40 min/day of 1 Hz rTMS over 4 days.[11] There have now been 15 published treatment studies of rTMS targeting auditory hallucinations in schizophrenia. Ten sham-controlled trials (involving 212 patients) were included in a recent meta-analysis[12] which concluded that overall rTMS was significantly better than sham stimulation in the treatment of auditory hallucinations.

Negative symptoms of schizophrenia include alogia, avolition, anhedonia, and affective flattening and are associated with attentional impairment and executive dysfunction. Negative symptoms are often resistant to neuroleptic medication and are associated with poor clinical outcome. There is increasing evidence that negative symptoms are related to reduced cortical activation, particularly involving the left prefrontal cortex. Therefore one treatment approach has been to attempt to increase activation within this region. There have been four published randomized controlled studies comparing real and sham rTMS of the left dorsolateral prefrontal cortex to target negative symptoms of schizophrenia, of which three found no difference between real and sham treatments and one found 2 weeks of high-frequency rTMS significantly improved negative symptoms. Novak *et al.*[13] additionally performed a battery of neuropsychological tests and follow-up patients for 6 weeks after treatment but found no significant differences between treatment groups at either time point for primary or secondary outcome measures. The most recent study[14] did not provide evidence that rTMS to the DLPFC improved negative symptoms of schizophrenia in patients with prominent negative symptoms but did suggest that rTMS may improve cognitive functioning in this patient group, at least in the short-term. However larger studies with longer periods of follow-up will be required to further examine this preliminary finding.

rTMS and obsessive–compulsive disorder

There have been several studies that have attempted to treat symptoms of obsessive–compulsive disorder (OCD) by modulating activity in prefrontal and motor circuits using rTMS. The earliest blinded trial of rTMS for OCD included 12 patients and found that a single session of right prefrontal high-frequency cortical stimulation significantly decreased compulsive urges for over 8 h.[15] Obsessive thoughts did not change significantly. This study suggested that rTMS may be a useful probe of neuronal circuitry associated with symptoms of OCD. However a number of subsequent studies have failed to replicate these findings. A Cochrane review in 2003 examined three randomized controlled trials and concluded that there were currently insufficient data to draw conclusions about the efficacy of rTMS in the treatment of OCD.[16]

The most recent, and largest trial of TMS in OCD to date randomly allocated 33 patients with OCD to receive 10 sessions of either active or sham low-frequency (1 Hz) rTMS over the LDLPFC.[17] This study did not demonstrate any difference between real and sham treatments.

rTMS and other neuropsychiatric disorders

rTMS has been postulated as a potential treatment in a variety of other neuropsychiatric disorders. There is emerging evidence that it may improve some of the motor symptoms of Parkinson's disease. In a recent study six daily sessions of high-frequency rTMS were given to 55 unmedicated patients with Parkinson's disease.[18] Patients received either 10 Hz or 25 Hz rTMS bilaterally to the motor cortex arm and leg areas or to the occipital cortex (control group). It was found that stimulation to the motor areas improved all measures, e.g. walking time, key-tapping speed, and self-assessment and that 25 Hz stimulation yielded greater improvement than 10 Hz. The effect was sustained for a month after treatment and restored by further booster sessions. The authors concluded that 25 Hz rTMS can lead to cumulative and long-lasting benefits on motor performance in Parkinson's disease.

A recent study explored the effect of low-frequency rTMS to the left temperoparietal region on chronic tinnitus.[19] Patients received 1200 stimuli per day for 5 days of either real or placebo treatment in a randomized controlled crossover trial. Overall active rTMS induced a transient, but significant improvement in the symptoms of tinnitus.

rTMS has also been used in an attempt to provide relief from chronic neuropathic pain.[20] A recent review summarized that high-frequency rTMS to the motor cortex is able to produce pain relief but that the effect is brief and that research in this area is required.

rTMS has also been used as a probe of neuronal circuits in dementia and in attention deficit hyperactivity disorder although any possible therapeutic role for it in these conditions appears some way off.

Summary

rTMS has an increasing role as a useful investigational tool for probing neuronal circuitry in a variety of neuropsychiatric disorders. However its therapeutic value is at present less certain.

The antidepressant efficacy of rTMS has now been investigated for over 15 years and despite initial early enthusiasm there is still not clear evidence for its usefulness as a treatment in depression, reflected in the recent FDA and NICE decisions. Further research is required to identify specific brain regions in specific conditions that may be appropriate targets for treatment with rTMS, allowing tailoring of treatments for individual patients. The recent development of neuronavigational techniques using MRI imaging should aid treatment site localization. Other future research should be directed at establishing optimal rTMS parameters, e.g. the intensity, frequency and number of treatments.

Further information

http://www.ists.unibe.ch/ is the homepage of the International Society for Transcranial Stimulation. It also provides links to other websites of interest.

George, M.S. and Belmaker, R.H. (eds.) (2007). *Transcranial magnetic stimulation in clinical psychiatry*, p. 289. American Psychiatric Publishing. Arlington, VA. Provides a useful summary of the theoretical and practical aspects of TMS research.

Lisanby, S.H. (ed.) (2004). *Brain stimulation in psychiatric treatment*, p. 153. American Psychiatric Publishing, Washington, DC. Reviews transcranial magnetic stimulation and also other brain stimulation techniques including deep brain stimulation, magnetic seizure therapy and vagus nerve stimulation.

References

1. Barker, A.T., Jalinous, R., and Freeston, I.L. (1985). Non-invasive magnetic stimulation of human motor cortex. *Lancet*, **1**, 1106–7.
2. Hallett, M. (2000). Transcranial magnetic stimulation and the human brain. *Nature*, **406**,147–50.
3. Speer, A.M., Kimbrell, T.A., Wassermann, E.M., *et al.* (2000). Opposite effects of high and low frequency rTMS on regional brain activity in depressed patients. *Biological Psychiatry*, **48**, 1133–41.
4. Strafella, A.P., Paus, T., Barrett, J., *et al.* (2001). Repetitive transcranial magnetic stimulation of the human prefrontal cortex induces dopamine release in the caudate nucleus. *The Journal of Neuroscience*, **21**, RC157.
5. Wassermann, E.M. (1998). Risk and safety of repetitive transcranial magnetic stimulation: report and suggested guidelines from the international workshop on the safety of repetitive transcranial magnetic stimulation, June 5–7, 1996. *Electroencephalography and Clinical Neurophysiology*, **108**, 1–16.
6. Bickford, R.G., Guidi, M., Fortesque, P., *et al.* (1987). Magnetic stimulation of human peripheral nerve and brain: response enhancement by combined magnetoelectrical technique. *Neurosurgery*, **20**, 110–6.
7. Martin, J.L., Barbanoj, M.J., Schlaepfer, T.E., *et al.* (2002). Transcranial magnetic stimulation for treating depression. *Cochrane Database of Systematic Reviews*, **2**, CD003493.
8. Lisanby, S.H., Gutman, D., Luber, B., *et al.* (2001). Sham TMS: intracerebral measurement of the induced electrical field and the induction of motor-evoked potentials. *Biological Psychiatry*, **49**, 460–3.
9. Loo, C.K., Mitchell, P.B., McFarquhar, T.F., *et al.* (2007). A sham-controlled trial of the efficacy and safety of twice-daily rTMS in major depression. *Psychological Medicine*, **37**, 341–9.
10. Eranti, S., Mogg, A., Pluck, G., *et al.* (2007). A randomized, controlled trial with 6-month follow-up of repetitive transcranial magnetic stimulation and electroconvulsive therapy for severe depression. *The American Journal of Psychiatry*, **164**, 73–81.
11. Hoffman, R.E., Boutros, N.N., Berman, R.M., *et al.* (1999). Transcranial magnetic stimulation of left temporoparietal cortex in three patients reporting hallucinated. "voices" *Biological Psychiatry*, **46**, 130–2.
12. Aleman, A., Sommer, I.E., and Kahn, R.S. (2007). Efficacy of slow repetitive transcranial magnetic stimulation in the treatment of resistant auditory hallucinations in schizophrenia: a meta-analysis. *The Journal of Clinical Psychiatry*, **68**, 416–21.
13. Novak, T., Horacek, J., Mohr, P., *et al.* (2006). The double-blind sham-controlled study of high-frequency rTMS (20Hz) for negative symptoms in schizophrenia: negative results. *Neuroendocrinology Letters*, **27**, 209–13.
14. Mogg, A., Purvis, R., Eranti, S., *et al.* (2007). Repetitive transcranial magnetic stimulation for negative symptoms of schizophrenia: a randomized controlled pilot study. *Schizophrenia Research*, **93**, 221–8.
15. Greenberg, B.D., George, M.S., Martin, J.D., *et al.* (1997). Effect of prefrontal repetitive transcranial magnetic stimulation in obsessive-compulsive disorder: a preliminary study. *The American Journal of Psychiatry*, **154**, 867–9.
16. Martin, J.L., Barbanoj, M.J., Perez, V., *et al.* (2003). Transcranial magnetic stimulation for the treatment of obsessive-compulsive disorder. *Cochrane Database of Systematic Reviews*, **3**, CD003387.
17. Prasko, J., Paskova, B., Zalesky, R., *et al.* (2006). The effect of repetitive transcranial magnetic stimulation (rTMS) on symptoms in obsessive-compulsive disorder. A randomized, double blind, sham controlled study. *Neuroendocrinology Letters*, **27**, 327–32.
18. Khedr, E.M., Rothwell, J.C., Shawky, O.A., *et al.* (2006). Effect of daily repetitive transcranial magnetic stimulation on motor performance in Parkinson's disease. *Movement Disorders*, **12**, 2201–5.
19. Rossi, S., De Capua, A., Ulivelli, M., *et al.* (2007). Effects of repetitive transcranial magnetic stimulation on chronic tinnitus. A randomised, cross over, double blind, placebo-controlled study. *Journal of Neurology, Neurosurgery, and Psychiatry*, **78**, 857–63.
20. Leo, R.J and Latif, T. (2007). Repetitive transcranial magnetic stimulation (rTMS) in experimentally induced and chronic neuropathic pain: a review. *The Journal of Pain*, **8**, 453–9.

6.2.10.4 Neurosurgery for psychiatric disorders

Keith Matthews and David Christmas

Ablative neurosurgery

Definition

Historical definitions of Neurosurgery for Mental Disorder (NMD), previously known as 'psychosurgery', have either made distinctions between neurosurgery for psychiatric or 'psychological' illness and disorders assumed to have a clearer 'biological' origin (e.g. epilepsy, Parkinsons' disease); or, have emphasized control of behaviour as a therapeutic objective rather than the control of symptoms. A more recent definition, and the one used throughout this chapter is that provided by the UK Royal College of Psychiatrists:[1]

> *A surgical procedure for the destruction of brain tissue for the purposes of alleviating specific mental disorders carried out by a stereotactic or other method capable of making an accurate placement of the lesion.*

Historical overview

The first attempt at treating psychiatric illness by surgical methods is commonly attributed to Gottlieb Burckhardt, a Swiss psychiatrist, who in 1888 performed 'temporal topectomy' on six patients who were most probably suffering from schizophrenia. His intention

was to sever the connections between the frontal lobes and the rest of the brain. Results were mixed: one patient was reported as improved; two were 'quieter'; and two showed no change. However, one patient died; another developed epilepsy; and a further had motor weakness. His results were met with a mixture of ridicule and hostility and he never again wrote on the subject.

In 1935, James Fulton and Carlyle Jacobsen operated on the frontal lobes of two chimpanzees named Becky and Lucy after first studying their responses to frustration in behavioural experiments. They found their behaviour dramatically changed after the surgery. Becky's previous agitated responses to frustration became more passive whilst Lucy was much more agitated. At a London meeting in 1935, they presented their findings to an audience which included Egas Moniz, a Portuguese neurologist.

Moniz teamed up with Almeida Lima, a neurosurgeon, and in a 30-min operation in November 1935, they performed frontal leucotomy on their first patient. The procedure first involved injecting alcohol into the white matter tracts of the frontal lobes, but they later would change to using an instrument of their own design, the leucotome, to extirpate 'cores' of tissue. In 1936, they published their report on the outcomes of 20 patients who were probably suffering from depression, panic disorder, and schizophrenia. One-third were better, one-third were worse, and one-third were unchanged.

Shortly after their paper was published, Walter Freeman, a US neurologist, wrote an enthusiastic review and quickly secured the collaboration of a neurosurgeon, James Watts. They modified the procedure slightly and began practising what Freeman termed bilateral frontal lobotomy. Over the next decade, Freeman became frustrated with the cumbersome requirements of a neurosurgical theatre and team. Adapting a technique first described in the 1930s, Freeman infamously developed the transorbital lobotomy in 1946. Notorious for the initial use of an ice-pick, the procedure involved forcing a tool (an 'orbitoclast') under the upper eyelid and through the base of the skull into the frontal lobes. Also known as the 'ice-pick lobotomy', the relative ease with which the procedure could be performed resulted in the widespread adoption of the technique throughout the United States and Europe. However, it was the indiscriminate overuse of such 'freehand' procedures and the associated adverse effects that occurred in many patients that led to public and professional antipathy towards neurosurgery for psychiatric illness, which peaked in the late 1950s. The introduction of chlorpromazine in 1954 also meant that for the first time there was a non-surgical treatment for schizophrenia.

Despite a reduction in the use of neurosurgery for mental disorder in the late 1950s and early 1960s, the development of stereotactic techniques (which had been demonstrated in 1908 by Horsley and Clarke and adapted for human use in 1947 by Spiegel and Wycis) meant that greater accuracy and greater consistency in neurosurgery could deliver better outcomes in selected patients. Procedures became more selectively and reliably targeted and lesions became more discrete.

Procedures

All NMD procedures have targeted one or more of three main regions: (i) fronto-limbic connections within the orbital or cingulate cortices; (ii) subcortical limbic circuitry; and (iii) limbic cortex, including the amygdala and cingulate cortex.

Four 'modern' stereotactic procedures have been described with only two remaining in regular usage in the Western World. Anterior capsulotomy is still performed in Cardiff (UK), Spain, Belgium, and Scandinavia, whilst anterior cingulotomy is the procedure of choice in Dundee (UK), Poland, South Korea, and North America.

(a) Subcaudate tractotomy (SST)

Developed in the United Kingdom by Geoffrey Knight in 1965, lesions were originally created using radioactive Yttrium^{90} rods. SST targets the white matter tracts of the 'substantia innominata' connecting the orbital cortex to limbic regions, and probably involved lesioning the nucleus accumbens.

(b) Anterior capsulotomy (ACAPS)

Described by Jean Talairach in 1949 and further developed by Lars Leksell for the treatment of chronic pain, ACAPS places lesions in the anterior limb of the internal capsule—a large white matter bundle connecting the frontal cortex with the thalamus and limbic structures. Lesions are generated using focused gamma radiation (gammacapsulotomy) or thermal damage (thermocapsulotomy). (See figure 6.2.10.4.1 for typical ACAPS lesions.)

(c) Anterior cingulotomy (ACING)

The cingulate gyrus was first proposed as a target by John Fulton in the late 1940s. Hugh Cairns, an English neurosurgeon and a friend of Fulton's, performed 'cingulectomy' in 1948. The less destructive 'cingulotomy' was first performed by Eldon Foltz and Lowell White in 1962 for the treatment of pain. (See figure 6.2.10.4.2 for typical ACING lesions.)

(d) Limbic leucotomy

First developed by Desmond Kelly in 1973, it represented a combination of cingulotomy and subcaudate tractotomy lesions.

Indications

During the 1960s and 1970s, a variety of other operations were explored as treatments for hypersexuality, aggression, and criminality. These included hypothalamotomy and amygdalotomy. Reports of outcomes from such interventions were often favourable, but interpretation has been complicated by issues of diagnosis, patient selection, and assessment. Such clinical presentations are not considered appropriate indications for surgery today.

Three main indications exist for modern NMD: obsessive-compulsive disorder (OCD); anxiety disorders; and major depression. Only individuals who have experienced chronic, disabling symptoms that have failed to respond after the diligent pursuit of available treatments (pharmacological and psychological) should be considered for ablative neurosurgery.

There are few absolute contraindications to NMD. There is no evidence to support the use of NMD as a treatment for eating disorders, schizophrenia, or personality disorders. However, where these exist as significant comorbid conditions alongside depression or OCD, these do not represent absolute contraindications and careful consideration is required.

Ethical considerations

One of the most persistent concerns for the public and for health professionals is that NMD is used to treat patients in the absence of informed consent. Although the absence of informed consent was likely to be an issue for early procedures, to the best of our knowledge,

all contemporary centres performing NMD today insist upon the patient's ability to give informed consent. In Scotland, the Mental Welfare Commission must authorize any proposed NMD as being in the patient's best interests and must confirm that the patient is capable of providing informed consent. In England and Wales, the Mental Health Act Commission has a duty to set-up multi-disciplinary panels to authorize NMD for consenting patients under Section 57 of the Mental Health Act, 1983. Similarly, in most other countries where the procedure is available, the procedure can only go ahead with the approval of an independent review board.

Criteria for NMD

General criteria for suitability show little variation between centres. Key inclusion and exclusion criteria for NMD in Dundee are shown below in Table 6.2.10.4.1.

Outcomes from NMD

Whilst placebo-controlled trials are considered the ideal assessment for intervention trials, they have frequently been described as unethical in the case of NMD. Despite this, there have been three isolated double-blind trials, involving a total of 6 patients. Despite suggestions of non-response in all cases, follow-up was brief and there is inadequate detail to make informed judgements of outcome.

(a) Combined outcomes for different procedures

Comparisons across studies are of limited value due to differences in procedure, patient characteristics, and the use of different rating scales. However, Spangler *et al.*[2] reviewed outcomes from different procedures and for different indications, defining a positive outcome as being a score of 1 or 2 on the Clinical Global Impression (Improvement) scale.[(2)] Positive outcome rates were: ACAPS (67 per cent); ACING (61 per cent); SST (37 per cent); and limbic leucotomy (67 per cent). The most effective procedures for affective disorder and OCD (respectively) were limbic leucotomy followed by ACING. The least effective procedure for both disorders (and overall) was SST.

Table 6.2.10.4.1 Inclusion and exclusion criteria for NMD

Inclusion criteria	Exclusion criteria
1. Age ≥20 years	1. Age <20 years
2. *Legal status*: both formal and informal patients can be considered	2. Failure to fulfil ICD-10 criteria for a suitable indication
3. *ICD-10 diagnosis of*: severe depressive episode; recurrent depressive disorder, current episode moderate to severe; bipolar affective disorder, current episode severe depression	3. *Primary diagnosis of*: substance misuse; organic brain syndrome; adult personality disorder; pervasive developmental disorder
4. *Duration of episode of illness*: minimum of 3 years, with at least 2 years of unremitting symptoms despite active treatment. Only in exceptional circumstances would a duration<5 years be considered	4. Absence of evidence of an adequate therapeutic trial of psychological treatment
5. *Consent*: the patient must be capable of providing sustained, informed consent	5. Absence of evidence of extensive trials of adequate pharmacological treatment

(b) Anxiety disorders

The crude rate of improvement following NMD (all procedures) for anxiety disorders ($n = 290$) is 77 per cent.[(1)] More recent reports of ACAPS would support a claim to effectiveness but this may be at the expense of significant adverse effects (apathy and dysexecutive symptoms).

(c) Obsessive-compulsive disorder

The combined rate of 'Completely Improved' or 'Improved' outcomes following SST is 52 per cent. More recent studies involving ACAPS or ACING report improvements on the Yale-Brown Obsessive-Compulsive scale (Y-BOCS) in the region of 30 per cent. Despite this relatively low figure, approximately 85 per cent of patients ($n = 478$) will have a marked or lesser improvement following NMD for OCD.[(1)]

(d) Depression

There is only one report of outcomes from ACAPS for depression. Herner (1961) described outcomes for 19 patients with a 'depressive state'.[(3)] Outcomes included: 'permanent improvement' (74 per cent); unchanged (5 per cent); and worse (5 per cent). In all 75 per cent experienced permanent side effects.

Spangler *et al.* (1996) reported that 53 per cent of those with affective disorder ($n = 10$) responded to ACING, with a 60 per cent response rate in unipolar depression.[(2)] Dougherty *et al.* (2003) reported a mean reduction in Beck Depression Inventory (BDI) score of 33 per cent in 13 patients following ACING.[(4)] With regards to limbic leucotomy, there are few studies looking at outcomes solely in depression but Mitchell-Heggs (1976) reported that 7 of 9 patients with depression improved after surgery.[(5)]

(e) Bipolar disorder

There are only two reports of NMD for bipolar disorder, both following SST.[(6,7)] Each involved small numbers ($n = 9$) but described improvements in cycle frequency with a greater effect on manic episodes than depression. Improved drug responsiveness was also alleged.

Mechanism of action

The mechanism of action of NMD is unknown, but almost all neurosurgical procedures involve lesioning white matter tracts connecting the prefrontal cortex with the thalamus, cingulate gyrus, and areas of the limbic system such as the amygdala and hippocampus. Neuroimaging studies have suggested that this circuitry is dysfunctional in depression and interrupting parts of these circuits may rectify emotional and cognitive processing within these brain areas. In the case of OCD, there is compelling evidence that symptoms arise from functional circuits connecting the frontal cortex, thalamus, and basal ganglia and that lesioning parts of this circuit may serve to eradicate many of the symptoms.

Adverse effects

(a) General adverse effects

Transient adverse effects such as headache are relatively common and tend to resolve in the first week after surgery. Post-operative confusion can occur in 3–10 per cent of patients with higher rates following SST. Incontinence is relatively uncommon with modern procedures, but the reported rates are: 1.1 per cent after SST;

5.5 per cent after ACING; and 9.5 per cent with limbic leucotomy. Apathy has been reported to occur in up to 24 per cent of patients following limbic leucotomy, but not all studies report its occurrence and this is likely to be a relatively high estimate. The incidence of weight gain varies greatly: 6.2 per cent after SST; 65.5 per cent after ACAPS; and 5.5–21.4 per cent after ACING. Similarly, there is wide variation in the reported rates of seizures: 1.6–3.3 per cent after SST; 0–7.7 per cent after ACAPS; 1–9 per cent after ACING; and 14.2 per cent after limbic leucotomy. Finally, suicide rates range from 1 per cent after SST to 12 per cent after ACING but these rates may reflect differing severities of illness.

(b) Effects on personality

It is certainly the case that early procedures such as leucotomy had marked effects on the personality and behaviour of large numbers of patients. However, with the advent of stereotactic procedures and more focused lesions, the effects on personality appear to be mild, sometimes even absent. It is acknowledged, however, that it is difficult to make robust appraisals of personality without assessment tools designed for such a purpose and in the context of symptom reduction in chronic illnesses.

Most published studies have reported normalization of personality traits following modern procedures such as ACING and ACAPS. In addition, there is a trend towards reductions in neuroticism and increases in extraversion. There are some recent reports of adverse effects on executive function following ACAPS for anxiety disorders[8] but many such studies lack preoperative assessments of personality making conclusions difficult to draw.

(c) Effects on neuropsychological function

As with personality changes, the detrimental effects of earlier procedures upon neuropsychological functioning were probably significant. However, the majority of studies reporting neuropsychological outcomes from ACAPS, ACING, and limbic leucotomy from the early 1970s onwards report either no deterioration on general measures (such as IQ, attention, memory) post-operatively or, more frequently, improvement. It is likely that improvements in performance are mediated through symptom reduction.

Vagus nerve stimulation (VNS)

Overview

The vagus nerve is the longest of the 12 cranial nerves and 80 per cent of its fibres are sensory afferents. These fibres terminate in the nucleus tractus solitarius, sending ascending fibres to the forebrain via the locus coeruleus and parabrachial nucleus. The vagus nerve, therefore, provides an access route to modify information which is processed in brain regions involved in mood regulation.

VNS involves the subcutaneous implantation of a programmable pulse generator in a location similar to a cardiac pacemaker. Electrodes connect the generator to the left cervical portion of the vagus nerve. Stimulation is delivered in an intermittent pattern (typically 30 s every 5 min) but parameters are changed using a palmtop computer and a programming wand held over the pulse generator.

VNS was first used to treat epilepsy in 1988 and became available for the treatment of refractory partial seizures in 1994. The first trials in depression began in 1998.

Outcomes in depression

There are a number of short, open trials of VNS which typically report a 3-month response rate (≥50 per cent reduction in the 24-item Hamilton Rating scale for Depression; $HRSD_{24}$) of 30–40 per cent, and a remission rate ($HRSD_{24} \leq 9$) of approximately 15 per cent.

Larger, 12-month trials have demonstrated 12-month response rates of 27.2–46 per cent and remission rates of 15.8–29 per cent.[9,10] In a 12-month controlled comparison of VNS versus Treatment-As-Usual (TAU), George *et al.*[11] reported response rates of 27 per cent for VNS + TAU versus 13 per cent for TAU. Such improvements appear to be maintained at 2 years, with response rates of 42 per cent and remission rates of 22 per cent.[12] In the only randomized, controlled trial of VNS, Rush *et al.*[13] reported 10-week response rates on the $HRSD_{24}$ of 15.2 per cent in the VNS group versus 10.0 per cent in the placebo group, changes which were non-significant.[13] Despite positive results in uncontrolled trials, definitive evidence of efficacy remains elusive.

Adverse effects

Most adverse effects are related to stimulation, and in most people they are fairly mild and improve over time. Many can be managed by altering the stimulation parameters. In the initial stages, common effects are: hoarse voice (53 per cent); headache (23 per cent); neck pain (17 per cent); cough (13 per cent); and dyspnoea (17 per cent). At 12-months, the only adverse effect to persist at rates higher than 10 per cent is hoarse voice (21 per cent).[9]

Deep brain stimulation (DBS)

Overview

As with VNS, DBS has evolved from a treatment for neurological disorders to a putative intervention for psychiatric illness. The most

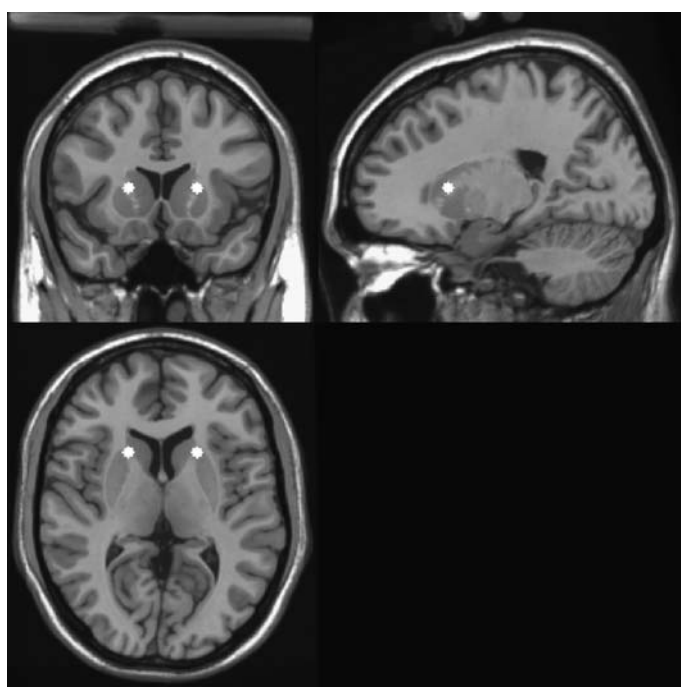


Fig. 6.2.10.4.1 Typical locations of anterior capsulotomy lesions superimposed upon normalized T1 MRI scan.
Lesions not to scale.

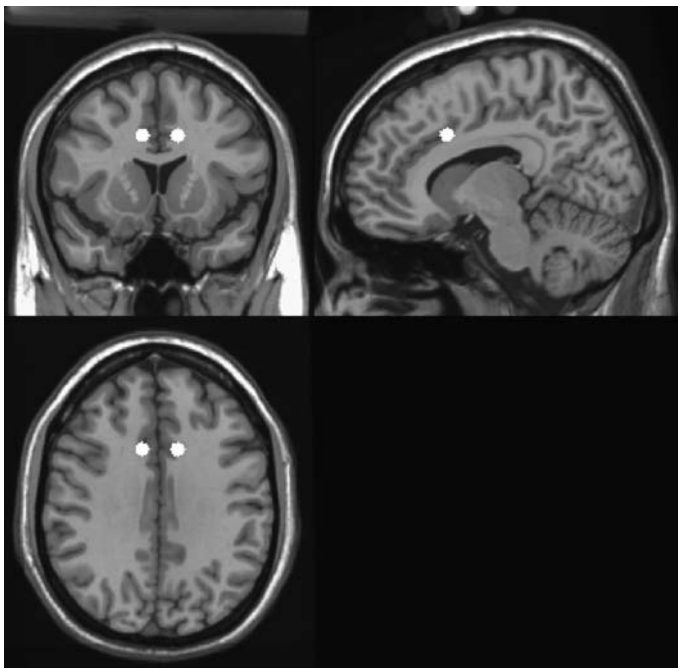


Fig. 6.2.10.4.2 Typical locations of anterior cingulotomy lesions superimposed upon normalized T1 MRI scan. Lesions not to scale.

effective targets for OCD and depression have yet to be determined but a number of possible locations for stimulation exist. However, DBS should be considered as an experimental treatment for both disorders.

The procedure involves the bilateral implantation of electrodes using stereotactic guidance with post-operative confirmation of location using MR scanning. The electrodes are connected to a generator typically implanted in the abdomen. Following surgery, the stimulation settings are programmed using immediate/short-term changes in symptoms as a guide.

DBS for obsessive-compulsive disorder

The most common target for DBS for OCD, thus far, has been ventral portion of the anterior limb of the internal capsule, the same site as ablative anterior capsulotomy. In a double-blind cross-over trial of anterior capsular stimulation in four patients Nuttin *et al*. (2003) reported reductions in symptoms of 36.8 per cent which were maintained after 21 months. In a small case series of four patients Abelson *et al.*[14] described marked improvements in one patient, with a lesser improvement in another. Greenberg *et al*. (2006) reported responses (≥35 per cent reduction in Y-BOCS score) in 4 out of 8 patients with DBS in the internal capsule.

Other proposed targets include the nucleus accumbens and the ventral caudate nucleus, but all targets may involve stimulation of a common anatomical area.

DBS for depression

In the only published report of DBS for depression, Mayberg *et al.*[15] stimulated the white matter tracts of the subgenual cingulate gyrus in six patients. Four patients were responders whilst two showed no change. Randomized on-off-on-off trials confirmed a stimulation-related improvement which was associated with a reduction in local cerebral blood flow in the subgenual cingulate and dorsolateral prefrontal cortex.

Adverse effects

Adverse effects that have been reported include: throbbing or buzzing sensations; nausea; and jaw tingling. A number of reports have described problems with battery life with the stimulators being replaced every 5 to 12 months. Battery failure has often been associated with a recurrence of symptoms over a few days which has been associated with marked depressive symptomatology and suicidal ideation. One study has reported a suicide, but commented that this was unrelated to stimulation. Numbers of reported cases are too small to determine if this is indeed the case.

Conclusions

Despite its chequered past, modern NMD bears little similarity to historical freehand procedures. Advances in neuroimaging mean that anatomical substrates for depression and OCD are being elucidated. Ablative procedures such as ACAPS and ACING are unlikely to undergo randomized controlled trials but prospective clinical audit suggests that such procedures may offer improvement to selected patients with treatment-refractory depression and OCD. Interventions such as VNS and DBS offer the possibility of double-blind testing but as yet there is insufficient evidence to suggest that such procedures offer greater effectiveness.

Further information

Freeman, C., Crossley, D., and Eccleston, D. (2000). *Neurosurgery for mental disorder. Report from the Neurosurgery Working Group of the Royal College of Psychiatrists.* Royal College of Psychiatrists, London.

Binder, D.K. and Iskandar, B.J. (2000). Modern neurosurgery for psychiatric disorders. *Neurosurgery*, **47**, 9–21.

Feldman, R.P., Alterman, R.L., and Goodrich, J.T. (2001). Contemporary psychosurgery and a look to the future. *Journal of Neurosurgery*, **95**, 944–56.

George, M.S., Rush, A.J., Sackeim, H.A., *et al.* (2003). Vagus nerve stimulation (VNS): utility in neuropsychiatric disorders. *International Journal of Neuropsychopharmacology*, **6**, 73–83.

Malone, D.A., Greenberg, B.D., and Rezai, A.R. (2004). The use of deep brain stimulation in psychiatric disorders. *Clinical Neuroscience Research*, **4**, 107–12.

References

1. Freeman, C., Crossley, D., and Eccleston, D. (2000). *Neurosurgery for mental disorder. Report from the Neurosurgery Working Group of the Royal College of Psychiatrists.* Royal College of Psychiatrists, London.
2. Spangler, W.J., Cosgrove, G.R., Ballantine, H.T. Jr., *et al.* (1996). Magnetic resonance image-guided stereotactic cingulotomy for intractable psychiatric disease. *Neurosurgery*, **38**, 1071–6.
3. Herner, T. (1961). Treatment of mental disorders with frontal stereotaxic thermo-lesions: a follow-up study of 116 cases. *Acta Psychiatrica Scandinavica*, **36**(Suppl. 158), 1–140.
4. Dougherty, D.D., Weiss, A.P., Cosgrove, G.R., *et al.* (2003). Cerebral metabolic correlates as potential predictors of response to anterior cingulotomy for treatment of major depression. *Journal of Neurosurgery*, **99**, 1010–7.
5. Mitchell-Heggs, N., Kelly, D., and Richardson, A. (1976). Stereotactic limbic leucotomy—a follow-up at 16 Months. *The British Journal of Psychiatry*, **128**, 226–40.

6. Lovett, L.M. and Shaw, D.M. (1987). Outcome in bipolar affective disorder after stereotactic tractotomy. *The British Journal of Psychiatry*, **151**, 113–16.
7. Poynton, A., Bridges, P.K., and Bartlett, J.R. (1988). Resistant bipolar affective disorder treated by stereotactic subcaudate tractotomy. *The British Journal of Psychiatry*, **152**, 354–8.
8. Rück, C., Andreewitch, S., Flyckt, K., *et al.* (2003). Capsulotomy for refractory anxiety disorders: long-term follow-up of 26 patients. *The American Journal of Psychiatry*, **160**, 513–21.
9. Marangell, L.B., Rush, A.J., George, M.S., *et al.* (2002). Vagus nerve stimulation (VNS) for major depressive episodes: one year outcomes. *Biological Psychiatry*, **51**, 280–7.
10. Rush, A.J., Sackeim, H.A., Marangell, L.B., *et al.* (2005). Effects of 12 months of vagus nerve stimulation in treatment-resistant depression: a naturalistic study. *Biological Psychiatry*, **58**, 355–63.
11. George, M.S., Rush, A.J., Marangell, L.B., *et al.* (2005). A one-year comparison of vagus nerve stimulation with treatment as usual for treatment-resistant depression. *Biological Psychiatry*, **58**, 364–73.
12. Nahas, Z., Marangell, L.B., Husain, M.M., *et al.* (2005). Two-year outcome of vagus nerve stimulation (VNS) for treatment of major depressive episodes. *The Journal of Clinical Psychiatry*, **66**, 1097–104.
13. Rush, A.J., Marangell, L.B., Sackeim, H.A., *et al.* (2005). Vagus nerve stimulation for treatment-resistant depression: a randomized, controlled acute phase trial. *Biological Psychiatry*, **58**, 347–54.
14. Abelson, J.L., Curtis, G.C., Sagher, O., *et al.* (2005). Deep brain stimulation for refractory obsessive-compulsive disorder. *Biological Psychiatry*, **57**, 510–6.
15. Mayberg, H.S., Lozano, A.M., Voon, V., *et al.* (2005). Deep brain stimulation for treatment-resistant depression. *Neuron*, **45**, 651–60.

6.3

Psychological treatments

Contents

6.3.1 Counselling

Diana Sanders

Introduction

People seek counselling for many reasons. Sometimes those who have had no previous need for mental health services are literally stopped in their tracks by life events—illness, family breakdown, intolerable stresses. People with long-term difficulties may turn to counselling when they feel the statutory services are not able to meet their needs, or as an adjunct to health care provision. With greater social mobility and the separation of family members, counselling increasingly provides the care and support previously offered within local communities. The provision and acceptability of counselling is on the increase. Counselling is possibly the most commonly delivered form of psychological therapy[1,2] and the British Association for Counsellors and Psychotherapists have over 30 000 members, with equivalent numbers in other countries. Professional training programmes in counselling have mushroomed in response to demand. Counsellors are found in many statutory and voluntary settings—mental health, primary care and medical settings, workplaces, drug and alcohol services, voluntary and charitable organizations, trauma services, and educational settings—as well as in private practice.

But what exactly is counselling? What do counsellors do? Is counselling the same as psychotherapy? And, is it an effective form of treatment? Although counselling is a major growth area within mental health, it can be difficult for consumers and purchasers of counselling services to know what kind of counselling and counsellor to use, with lack of clarity about what works for whom. There are many different models of counselling, types of counsellor and many different training courses. It is difficult to make clear distinctions between counselling and psychotherapy. Much of the work of counsellors has not historically been amenable to standard methods of evaluation, and research is relatively new. Currently there is no statutory regulation for the term 'counsellor', which means that people are able to practise as counsellors without registration or accreditation. By definition, people who seek counselling are likely to be vulnerable, and the issue of public protection is paramount.

The aim of this chapter is to clarify these issues and examine the place of counselling in psychiatry. The chapter begins by looking at the definition of counselling, and how counselling is both similar to, and distinct from, psychotherapy. The chapter goes on to look at the key features of counselling, and different models of counselling. Although counselling can and is used for many psychological difficulties, the chapter selects specific problems where there is evidence that it is an effective intervention: mild to moderate depression, adjustment difficulties, bereavement, trauma, and relationship problems. I then consider counselling in different settings, again selecting a few which illustrate the work of counsellors—primary care, mental health settings, student counselling, and the workplace—looking at the way counselling can be adapted according to the needs of the service. The chapter concludes by looking at issues of training, quality, and standards, commenting on the need for the control of an ever-developing profession without loss of the growing availability of effective counselling services to those in need.

Defining counselling

No single definition of either counselling or psychotherapy exists in spite of many attempts in Britain, North America, and elsewhere to arrive at one.[3] Currently, neither the British Association of Counselling and Psychotherapy (BACP) nor the American Counseling Association has either proprietary rights of the terms or even official definitions, although, as discussed below, the move towards statutory regulation for counselling and psychotherapy may ensure greater clarity for practitioners and consumers.

At its broadest, counselling is conceptualized as a way of helping or assisting others to make their own adjustment and decisions in the face of life problems. Counselling aims to offer a safe relationship within which the individual can explore personal difficulties and, through developing a deeper understanding of themselves, move towards change. The Department of Health defines counselling as . . . a *form of psychological therapy that gives individuals an opportunity to explore, discover, and clarify ways of living more resourcefully, with a greater sense of well being. Counselling may be concerned with addressing and resolving specific problems, making decisions, coping with crises, working through conflict, or improving relationships with others.*[4]

In contrast with other forms of psychological therapy, where the focus may be on treating specific problems, many counselling models give equal if not more weight to the *process* of change. The journey through which a client goes—to greater understanding, awareness, and resolution—is as important as the outcome. Counsellors practise within all therapeutic approaches, strongly influenced by humanistic, experiential, and psychodynamic principles. They tend not to link their work to diagnostic categories, preferring to see each client as an individual, and using an approach matching the client's needs rather than diagnosis. Counsellors may define themselves by the model they practise—for example, humanistic or psychodynamic—and/or by the type of problems they work with, such as bereavement or relationship counsellors.

Counselling may also be defined in terms of *key elements* and *goals*,[5] as shown in Table 6.3.1.1. Again, the key elements and goals illustrate the variation within counselling. For example, what is meant by counselling ranges from providing a safe arena for people to gain understanding and insight, to offering more direction and guidance leading to decision-making and problem-solving.

Table 6.3.1.1 The key elements and goals of counselling (Reproduced from Feltham, C. What are counselling and psychotherapy? In The Sage Handbook of Counselling and Psychotherapy, (eds. C. Feltham and I. Horton), pp. 3–10, copyright 2006, Sage Publications.)

The key elements of counselling	The goals of counselling
◆ Listening and talking methods of addressing psychological and psychosomatic problems and change ◆ An unstructured and non-directive form of therapy, using the therapeutic relationship as an active ingredient in promoting change ◆ Counselling operates largely without medication or other physical interventions ◆ Counselling may be concerned not only with mental health but with social, spiritual, philosophical and other aspects of living ◆ Professional forms of counselling are based on formal training, accreditation and on-going supervision and professional development	◆ Support, psycho-education and guidance ◆ Insight and understanding ◆ Self actualization and personality change ◆ Adjustment, symptom reduction and 'cure' ◆ Problem-solving and decision-making ◆ Crisis intervention and management ◆ Risk management (e.g. genetic counselling)

Counselling and psychotherapy

Much of the above can also be applied to psychotherapy and parts of this chapter do indeed overlap significantly with psychotherapy. There may well be variation between countries in what is defined as counselling or psychotherapy. Counselling and psychotherapy each have distinct features including different historical roots. *Psychotherapy* arose from the seminal works of Freud in the late nineteenth and twentieth centuries, and in the past, psychotherapists tended to offer a long-term psychodynamic approach. Now, however, psychotherapy also includes interpersonal, humanistic, and cognitive models. In the United States, *counselling* was originally linked with vocational guidance, personnel management, and the workplace,[5] and as such was much more advisory and directive than the analytic processes of psychotherapy. Carl Rogers, the founder of non-directive counselling in the 1940s, initiated the movement away from practical guidance and problem-solving towards collaborative and person-centred models, forming the basis of counselling today.

Differences between psychotherapy and counselling tend to relate more to the individual psychotherapist's or counsellor's training and interests and to the setting in which they work, rather than to any intrinsic difference in the two activities. In medical and mental health settings, psychotherapists are more likely to work with patients with severe psychological disorders, offering long-term therapy, whereas counsellors may concentrate on difficulties amenable to short-term work—mild to moderate psychological disorders, relationship difficulties, or bereavement. Counsellors who work for voluntary agencies or in educational settings such as schools and colleges usually concentrate more on the 'everyday' problems and difficulties of life than on severe psychological disorders, although agencies such as MIND, Alcoholics Anonymous, or

Narcotics Anonymous offer counselling to people with serious mental illnesses. In private practice, however, a counsellor's work will overlap with that of a psychotherapist.

In a pragmatic vein, Feltham states that *practitioners and the public stand to gain much more from the assumption of commonality than from spurious or infinitesimal distinctions . . . little is to be gained practically from further controversy about professional titles and distinctions*.[5] In 2000, the British Association for Counselling lent weight to greater rapprochement by becoming the British Association for Counselling and Psychotherapy. On a practical level, it is interesting that many recipients of 'talking therapies' other than counselling, such as psychotherapies, cognitive-behaviour therapy and problem-focused discussions with GPs, psychiatrists, or nurses, say that they have received *counselling*, reinforcing Feltham's plea that the issue of definition and distinction is academic rather than of practical value.

Counselling skills and counselling practice

Counselling skills are integral to the practice of psychiatry and all the 'helping professions', as basic ingredients of effective interviewing, accurate history-taking, diagnosis, and treatment-planning.[6] The skills of listening, summarizing, reflecting, checking, understanding, gaining rapport, and communicating enable other people to feel understood. They are essential for engagement and eliciting information, especially when the person is afraid, in pain, or mistrustful. The health worker's counselling skills may influence the patient's collaboration with an active participation in treatment, and thereby the outcome of a wide range of medical and even surgical treatments. Many helping and health professionals such as social workers, occupational therapists, probation officers, and speech and communication therapists use counselling skills as an integral part of their work but would not be seen as primarily counsellors.

In contrast, counsellors as *professionals*, who use counselling as a specific intervention, work in many areas of mental health practice alongside psychotherapists, clinical and counselling psychologists, psychiatrists, psychiatric nurses, and social workers.[7] For professional counsellors, counselling skills are central to their work.

Counselling as a specific planned intervention in psychiatry can be differentiated into two broad and overlapping categories, defined by aims into *decision-making* and *treatment*:

- *Decision-making* is an important ingredient in many forms of therapeutic counselling but, conversely, some forms of decision-oriented counselling (e.g. genetic counselling) embody no explicit therapeutic intention.
- Counselling as a primary *treatment* for problems is used in the management of a range of conditions as an adjunct to other interventions including medication, as an integral component of a multimodal treatment method (e.g. crisis intervention), or as a specific treatment in its own right (e.g. for postnatal depression).

Counselling psychology

As well as professional counsellors, counselling psychologists have a particular role to play within counselling provision in mental health. The area of counselling psychology, now developing in the United Kingdom in line with other parts of the world, is a distinctive profession within applied psychology, which aims to foster the psychological development of the individual and help people develop more effective and fulfilled lives. It is based on the fundamental tenets of counselling, but in addition aims to integrate the application of psychological theory and research into its practice.[8] Counselling psychologists use a variety of therapeutic models, including person-centred, psychodynamic, and cognitive. Although the training of counsellors can be varied, as discussed below, counselling psychologists undergo standardized post-graduate doctorate training leading to chartered status within the British Psychological Society, or equivalent in other countries.

Is counselling an effective method of treatment?

Despite the proliferation of counsellors in many areas of medicine and psychiatry, counselling has tended to lag behind medicine and other health care professions in engaging in and promoting research to establish its effectiveness and efficacy.[9] The nature of counselling can mean that standard methods such as RCTs are not appropriate means of evaluation, whereas qualitative research methodology is better able to assess meaningful changes.[10] However, counselling as a profession is now engaging in better quality research, concentrating on outcomes in routine practice as well as qualitative analysis. New practice-based methods of evaluation, such as CORE (Clinical Outcomes in Routine Evaluation), and the aggregation of data across UK NHS counselling services can lead to national benchmarks. Methods of case-study research, and the development of measures of the client's perspective on psychological distress, PSYCHLOPS,[11] enable more client-focused research. Such emphasis on evidence-based practice will lead to more careful targeting of specific counselling approaches to specific problems,[3] clearer information for the public and will improve counselling's parity with other health care professions.

Currently, counselling has an image that it is more appropriate for people with mild to moderate difficulties. The Department of Health[4] recommends that counselling should not be the main intervention for people with severe and complex mental health problems or personality disorders. Patients who are adjusting to life events, illnesses, disabilities, or losses may benefit from brief therapies such as counselling. However, counsellors such as those in primary care and the voluntary sector are already offering an important and valued service. Although people with more serious and enduring mental health difficulties require primarily psychiatric and pharmacological intervention, offering emotional support, advice, and problem-solving can form an important, although under-researched, part of their care.

The core conditions of counselling and the therapeutic relationship

Counselling depends primarily on the interaction between the counsellor and client, what goes on in that interaction and the qualities of both client and counsellor. Carl Rogers'[12] definition of the conditions necessary for therapeutic change was a radical departure from traditional psychotherapeutic practice, in emphasizing the *qualities* and *attitudes* of the counsellor rather than specifying what the counsellor must *do*. His work led to the following as *necessary* and *sufficient* conditions for therapeutic change:

- The client is in a state of *incongruence*, being vulnerable or anxious.
- The therapist is *congruent* and *genuine* in the relationship with the client.

- The therapist experiences *unconditional positive regard* for the client.
- The therapist experiences an *empathic understanding* of the client's frame of reference or way of seeing things.
- The therapist feels *non-possessive warmth* towards the client.
- The client *perceives* the therapist's unconditional positive regard and empathic understanding.

These core conditions have been used and developed in many models of counselling and therapy; even therapies traditionally seen as more technical have always maintained their importance.[13] *Empathy*, for example, a core condition which is central to all good therapeutic relationships, enables clients to know that they are heard and understood. At its simplest, empathy is a simple restatement of someone else's words. At its richest, it involves ... *a fearless exploration of another's inner world, a sensing of meanings unspoken, a compassionate naming of pain ... the fullest empathy does not censor or discriminate. It sees the world as the other person sees it.*[14]

While Rogers took the view that such core conditions are both necessary and sufficient for therapeutic change to occur, other models of counselling have defined such conditions as necessary but not in themselves sufficient for change. However, the core conditions remain the bedrock upon which counselling is practised.

The therapeutic relationship

Across very diverse treatments, including cognitive and psychopharmacological,[15] measures of the strength of the relationship, or alliance, have been the strongest and most consistent process correlates of treatment outcome.[3,16] Clients who have strong alliances with their therapists tend to have better outcomes.

Although recognized as an essential component of change, different models have different conceptual and practical approaches to the relationship. Three examples illustrate the differences:

- Person-centred models take a here-and-now perspective, looking at the immediate interaction between client and counsellor. The client's perception of the therapist's empathy, unconditional positive regard, and congruence enables therapeutic change.
- Cognitive models regard the relationship as necessary but not sufficient for therapeutic change. The relationship is primarily collaborative, with an active, working bond formed between client and counsellor to facilitate the tasks of therapy.
- Psychodynamic models distinguish the real relationship between client and counsellor, and the transference relationship, consisting of both client transference and therapist counter transference. The working alliance therefore is only partly based in reality, also containing aspects of both parties' histories.

Counselling methods and techniques

There have always been many approaches to counselling and psychotherapy, and this diversity grew into a veritable 'multiverse' during which some authors estimated that there were over 400 brand therapies in existence. There are also different settings and agencies which offer counselling—clinics, institutes, health centres, or voluntary bodies, each with its own particular features. Within each model and setting, there are different formats of counselling including self-help materials on CD-rom and the Internet, as well as individual, couple, group, family, and organizational. Such a range can be confusing to potential clients and organizations, and the question of what works, for whom, and in which setting, has to be central in matching client, problem, therapy, and therapist.

Specific models of counselling are usually differentiated by a number of factors:

- Basic assumptions or philosophy
- Formal theory of human personality and development
- Clinical theory defining the goals, principles, and processes of change
- Therapeutic skills and techniques[17]

One useful distinction exists between *schools* of counselling and *theoretical approaches*.[18] A theoretical approach presents a single position regarding the theory and practice of counselling, whereas a school is a grouping of different theoretical approaches with common characteristics(see Table 6.3.1.2).

The three main schools are humanist-existential, psychodynamic, and cognitive behavioural. Humanistic-existential models will be described in detail, with briefer mention of psychodynamic and cognitive behavioural models, which are covered in other chapters. The section on methods also looks at the trend towards integration and eclecticism within counselling, whereby counsellors use a variety of methods and approaches adopted from different models. Although not clearly fitting into any one school, information-giving and *problem-solving* are counselling methods widely used in psychiatry, and are therefore described first.

Information-giving and problem-solving

Giving information is an important part of all medical and psychiatric practice, reflecting an open and collaborative approach to treatment, providing patients and their carers with the material necessary for informed decision-making. For example, for people with schizophrenia or those who misuse alcohol, the provision of information about the diagnosis, causes, and potential consequences of their condition is essential for mobilizing motivation and compliance with treatment. Giving information about the actions and potential side-effects of a prescribed medication enables people to

Table 6.3.1.2 Overview of counselling schools and main approaches (Reproduced from Nelson-Jones, R. *Theory and practice of counselling and therapy* (4th edn.), copyright 2006, Sage Publications.)

Psychodynamic school
Classical psychoanalysis (Sigmund Freud)
Analytical therapy (Carl Jung)
Humanistic-existential school
Person-centred therapy (Carl Rogers)
Gestalt therapy (Fritz Perls)
Transactional analysis (Eric Berne)
Existential therapy (Irvin Yalom and Rollo May)
Cognitive behavioural school
Behaviour therapy (Ivan Pavlov, BF Skinner and Joseph Wolpe)
Rational emotive behaviour therapy (Albert Ellis)
Cognitive therapy (Aaron Beck)
Multimodal therapy (Arnold Lazarus)

play an active role in pharmacological intervention. Information-giving is always crucial when communicating a diagnosis and fundamental to counselling for risk, as in genetic counselling, and to any intervention in which the individual is helped to make decisions.

Psycho-educative methods have a place in most models of counselling and psychotherapy, but have specific importance in problem-solving and cognitive behavioural models. For example, a psychologist or counsellor may describe to the client a psychological model of a specific condition, such as the cognitive model of panic, to help the client understand their particular symptoms.

Information-giving and psycho-education involves more than just giving information to a passive recipient. Wherever possible the individual's curiosity about their condition is promoted, encouraging them to ask questions and, when appropriate, to find their own answers. The Socratic method and guided learning are central to cognitive approaches. Information is not provided in a didactic fashion, but in response to the client's questions, as client and therapist are engaged in collaborative enquiry. Whatever the information given, the practitioner checks whether the client has understood the information and its meaning. Information-giving is rarely the endpoint of an intervention, serving instead as the basis for decision-making or continuing therapeutic work.

Problem-solving has been used and empirically validated as a specific treatment, particularly for depression, and is used by many cognitive behavioural and humanistic counsellors. Problem-solving forms a major part of brief solution-focused therapy.[19] From a problem-solving perspective, depression results from the interaction between negative life events, current problems, and deficient problem-solving abilities, and therefore facilitating solving problems is a means to alleviate depression.[20] Therapist and client work collaboratively to identify and prioritize key problem areas, break them down into specific manageable tasks, solve problems, and develop appropriate coping behaviours. The approach involves several stages:

- Identification and formulation of the client's problem(s)
- Setting clear and achievable goals
- Generation of alternatives for coping
- Selection and operationalization of a preferred solution
- Evaluation of progress, with further problem-solving as necessary

Research in the United Kingdom has shown that problem-solving delivered by general practitioners is as effective as pharmacological treatment for moderate and major depression in primary care.[21,22] The intervention can be extremely useful for clients who do not want or cannot tolerate pharmacological treatment and is recommended in NICE guidelines as a treatment for mild depression. It can be offered by counsellors, general practitioners, and nurses, and may be a means to improve treatment adherence for people with psychotic disorders, as part of a psycho-educational intervention including motivational interviewing.[23]

Brief solution-focused therapy developed from its roots in family therapy to applications in counselling, mental health, group work, education, drug and alcohol work, social work, and business. It is the preferred mode of working for counsellors in the workplace, given its brief and focused approach. The model arose from family therapists' observations that clients made significant changes when focusing on their preferred futures rather than on current problems. By articulating solutions, and building on existing skills and strengths, clients saw their problems in a different light and could effect change.

The 'miracle question' is a classic method of solution-focused therapy which is integrated into other models. The client is asked to think about and describe waking up one day to find that all problems have vanished. The counsellor explores the impact of the miracle on people and situations. The question enables the client to get into a problem-solving cognitive set, enabling identification of what needs to happen for the problems to change. The method has been studied in a range of client groups and settings, including with repeat offenders in the forensic service, and can produce positive outcomes.[19]

Humanistic and existential models

Humanistic and existential approaches include person-centred therapy, gestalt therapy, transactional analysis, and existential approaches. Of these, the person-centred model is the most well known, and the one that comes most readily to mind when describing the philosophy of counselling.

Client-centred, *or* ***person-centred*** as it is more often called, counselling originates from the work of Carl Rogers, whose emphasis on the recognition and empowerment of the help-seeker challenged the perceived authoritarianism of both the medical model and psychoanalysis. The model highlights respect for the person, and adopts the optimistic assumption that each person has an inner potential for healthy development and achievement, or 'self-actualization'. Person-centred approaches often use the analogy of a plant to describe the concept of growth and change. No one can make a plant grow, but if the plant is provided with the right conditions—water, light, soil, nutrients—then it will become the best plant it can be. Person-centred therapy assumes that people have an inbuilt motivation to change, and also have the skills necessary with which to effect changes. Rogers' model of counselling is non-directive. The counsellor's task is to create the core relationship conditions of empathy, warmth, unconditional positive regard, and genuineness, described above, in which the client's inner resources and potential will be unlocked, leading to the spontaneous resolution of problems and developmental growth.

The central features of person-centred counselling form the bedrock of other models of counselling, including cognitive approaches. Carl Rogers, in initiating the person-centred approach, has also had a wide influence in the helping professions—the term 'person-centred' is used frequently in policy documents and guidelines within health care organizations, as one of the standards of service and as a philosophy of health care.

While a non-directive and reflective approach has value, and may be useful for initial data-gathering and supportive work, caution must be applied to the use of Rogerian counselling in psychiatry. Resource constraints require practitioners to impose time limits on counselling, which therefore must be more focused and 'active'. Furthermore, very disturbed people may be unable to access an inner potential for spontaneous change and growth, implicit within the client-centred model. There are some for whom a reflective non-directive approach may be harmful, risking an overwhelming upsurge of avoided or forgotten memories of traumatic experiences without providing methods for coping with them.

Victims of childhood sexual abuse or other destructive experiences may be re-traumatized by unstructured reflective counselling.

It is likely that the person-centred approach will continue to form the basis of good counselling and psychotherapeutic practice regardless of the model used, with increased emphasis on more 'skills-based' approaches such as cognitive behavioural and other models that lend themselves more easily to measurement, structured working, and evidence-based practice.

Gestalt therapy was originated by Fritz Perls, who described his approach as dealing with the total existence of a person, rather than being primarily occupied with symptoms or character structure.[24] Gestalt therapy argues that the past is past and the future unknowable, therefore the focus of counselling should be the present moment—an approach, interestingly, espoused by the development of mindfulness in psychiatry and psychotherapy.[25] The goal of therapy is to put clients in touch with what they are thinking, feeling, and sensing, in the here and now, and how they restrict or limit themselves by continual focus on the past or future. Gestalt therapists regard the therapeutic relationship as a 'working' relationship, with client and counsellor taking responsibility for themselves. Attaining awareness is an essential aim within the relationship.

Gestalt therapy uses many techniques, including dream-work and psychodrama. The classic 'two chair' method of gestalt therapy enables clients to work with 'unfinished business' which may be influencing current problems. For example, a client with memories of a difficult relationship with a parent is encouraged to have a dialogue with the parent in the empty chair, to see both client and parent's point of view. The client may put themselves, metaphorically, in the empty chair, to enable greater understanding and acceptance of the self.

Research has only recently played a role in the development of gestalt theory and practice. Most of the studies concern the effectiveness of the two-chair method, an approach which is being integrated into other models, for example Greenberg's[26] emotion-focused psychotherapy, and cognitive approaches.[13]

Transactional analysis (TA) was founded by Canadian psychiatrist Eric Berne, and provides a theory of personality, child development, and psychopathology as well as a theory of counselling. The method assumes, as for other person-centred approaches, that people are born with a drive for growth and health—the 'I'm OK, you're OK' life position. TA characterizes the personality into three groups of 'ego states'—parent, adult, and child, each with behavioural, social, historical, and phenomenological aspects. Psychopathology arises from the repetition of unhelpful life scripts, or patterns of being, often learned early in life. Counselling enables the individual to identify and modify problematic patterns.

Very little research has been conducted into the effectiveness of TA as a therapy although many theoretical concepts and practical techniques have been assimilated into psychotherapy and counselling.[27] The method has also led to the concept of the 'reflective practitioner', a theme embodied by the BPS Division of Counselling Psychology.

Existential approaches originated in applied philosophy, and focus on helping people to come to terms with life in all its confusing complexity. Rather than curing people of pathology, the aim is to help people deal with the contradictions, dilemmas, and paradoxes of everyday existence.[28] Anxiety and depression, rather than to be avoided, are to be embraced and understood in order to live life to the full. The main method is conversational, enabling clients to confront rather than avoid the reality of situations.

Although existential approaches may sound idealized and unrealistic, much of existential therapy aims to help people build confidence and competence in tackling everyday problems. The methods and approaches have very little outcome research, because of the opposition of existential therapists to what is seen as the reductionist tendencies of research—i.e. what is effective in therapy is not open to evaluation using standard methodology. It may be that the approach offers a number of factors which can be usefully integrated into other, more evidence-based models, such as the focus on validating experience, creation of meaning to enable traumatic events to be processed, and authenticity.

Cognitive behavioural approaches

Cognitive behaviour therapy (CBT) is currently receiving excellent press internationally, and occupies a central place in the move towards evidence-based practice. NICE recommends CBT more often than other therapeutic approaches for many psychological problems. Despite its popularity and evidence-base, cognitive approaches have not been readily embraced by the counselling world.[13] The structured and focused approach, and use of techniques to promote change, rested uncomfortably with counsellors trained in client-centred approaches, and cognitive therapy was felt to pay insufficient attention to the therapeutic relationship and to the influence of past events on current problems. However, the last few years have seen a major change in the way cognitive therapy is being adopted within counselling, and a large proportion of counsellors integrate at least some of the approaches into their work.

The attraction of cognitive therapy to counsellors is increasing, with more overt focus placed on the therapeutic relationship, long-term approaches, and schema-focused work inherent in newer models, which enables counsellors to abandon their prejudices against CBT.[29] There is enormous scope for counsellors to adopt cognitive therapy in a more systematic and rigorous manner, particularly in light of the empirical evidence supporting its effectiveness and increasing demand for briefer interventions.[13] However, counsellors trained in different schools of counselling can be tempted to borrow specific methods from cognitive therapy, such as monitoring negative thoughts, and to use them in an eclectic way. The risk is that the effective components of the approach such as collaboration, structure, focus and homework, may be lost, thus diluting CBT's established effectiveness.

Psychodynamic counselling

Psychodynamic counselling[30] draws from the theoretical traditions of Alderian therapy, Jung's analytical psychology, Freudian psychoanalysis, and Kleinian psychodynamic therapy. Psychodynamic approaches pay particular attention to past experience, particularly adverse relationship experiences during early life, the continuing influence of which may be mediated by unconscious processes. These are seen to influence attachment patterns, psychosocial development, and later psychological functioning. Unconscious processes derived from early experiences contribute to the generation and maintenance of abnormal psychological states. In psychodynamic counselling and psychotherapy, these unconscious processes may be identified through examining

transference and counter-transference in the therapeutic relationship.

The search for the personal meaning of the client's problem or symptoms is central to psychodynamic counselling. The counsellor encourages clients to talk about their difficulties, but also to reflect and gain insight on spontaneous associations and attitudes towards the counsellor as potential sources of information about the presenting problems. Insight alone may be sufficient to enable clients to spontaneously bring about the required changes in their lives. Psychodynamic counselling may also use methods akin to problem-solving and behavioural experiments to facilitate identification and rehearsal of new and more adaptive interpersonal strategies.

Eclectic-integrative approaches

Many practitioners assimilate conceptual and practical way of working that can be attributed to more than one theoretical perspective, formulating the client's difficulties and choosing a mix of methods using more than one theoretical framework. Formally working with a variety of models and methods may be described eclectic or integrative.[31] Such generic therapies often emphasize non-specific factors such as building the therapeutic alliance and engendering hope. Whether this gives the best of each world, or risks the worst of all, is very much open to question, and by nature, eclectic therapy is difficult to standardize for RCTs. The worst kind of *eclecticism* may be an arbitrary pick-and-mix approach, whereby a generically skilled counsellor trains in a variety of approaches and applies these with clients in a way in which he or she deems best. There is little evidence that such an approach is any more effective than the core conditions of counselling allow. Lazarus[32] describes *technical eclecticism*, the drawing of interventions from different sources without necessarily subscribing to their founding discipline. Wherever possible, technically eclectic therapists use treatments based on empirical evidence and client need.

Integration, in contrast, combines identifiable and specific aspects of models in a predetermined way, allowing the evolution of a defined form of therapy such as *cognitive analytic therapy*.[33] *Psychodynamic interpersonal therapy* offers NHS counsellors a point of convergence between predominantly humanistic counselling and more clinically and dynamically orientated approaches often used within psychiatry.[34]

Integration and eclecticism in counselling and therapy will no doubt continue to develop as the nature of clients' problems and the ways of doing therapy evolve. Environmental and technological changes may lead to increasing use of the Internet in counselling and psychotherapy, with face-to-face interactions possible even when client and counsellor are in different locations. It is essential that new counselling approaches are thoroughly supported by empirical evidence so we do not see creeping eclecticism washing out the effectiveness of established methods.

Applications of counselling to specific conditions

Depending on the settings in which they work, counsellors need to be equipped to work with clients with a range of psychological difficulties. For example, the primary care counsellor's caseload is likely to include client difficulties ranging from mild to moderate anxiety or depression to bereavement and relationship problems. Whether or not counselling is an effective intervention for the range of problems seen in these settings has not yet been clearly established. The following looks at counselling for problems where there is good evidence for effectiveness.

Common psychological problems

A Cochrane systematic review[35] compared counselling with normal GP care for people suffering from anxiety, depression, or stress disorders. The authors concluded that overall, significant benefits were seen in mental health improvement from counselling compared with usual GP care, or GP care plus antidepressant treatment, in the short-term (up to 4 months). However these benefits were not maintained in the longer term, over 9 to 12 months. Counselling may be more effective for depression than as a treatment for anxiety. Barrowclough *et al.*[36] compared CBT with supportive counselling (SC) in the treatment of anxiety symptoms in older adults. The CBT group did better than the SC group following treatment, and at follow-up. Overall, cognitively orientated models of counselling and therapy are more effective for depression and anxiety in the long-term than generic counselling. However, counselling may enable people to recover more quickly from depression, and is therefore a valuable and valued intervention.

Counselling for adjustment disorder

Adjustment disorder is defined as a problematic response to a normal stressor, not caused by another mental health problem or bereavement. Such stressors include normal transitions such as leaving home, migration, adverse interpersonal experiences (e.g. relationship breakdown), and unexpected losses such as redundancy. Individual vulnerability can play a part in a person's reaction to life changes, such as previous losses or other adversity, social or cultural isolation, economic deprivation or physical illness.

Counselling is recommended as the first line of treatment for people having difficulty adjusting to life events, illnesses, disabilities or losses, including childbirth and bereavement.[4] The counselling relationship is an important source of security when much has changed in the person's life. The client is helped to identify the stressors, to explore the personal significance of the changes experienced, and to express the emotions generated. It can be necessary to examine unresolved past experiences which may impact on the current adjustment—for example, an individual may not begin to come to terms with redundancy until he recognizes and addresses his unresolved feelings about being abandoned by a parent in childhood. Problem-solving methods are used to identify adaptive goals and ways they may be achieved. The counsellor may encourage the client not to use unhelpful solutions such as denial, excessive use of alcohol, or emotional suppression. The aim is for the client to resolve the crisis themselves.

Relationship problems

Government statistics from both the United States and the United Kingdom show that an ever-growing proportion of marriages fail, with around one in two marriages ending in divorce and an even higher rate in other relationships. Many divorces and relationship problems involve children under the age of 16. A number of specialized services have evolved, including pre-marital counselling, counselling for sexual problems, infertility counselling, bereavement and divorce counselling, and counselling for those involved in

second and subsequent relationships.[37] Telephone counselling and drop in services are frequently used by individuals and couples aiming to clarify the problems and find appropriate help. The kind of issues couples bring to relationship counselling include:

- Communication difficulties
- Conflicts in need between different parties
- Extra-relationship affairs
- Sexual problems
- Conflicts as parents
- Gender role changes
- Violence
- Substance abuse
- Jealousy or possessiveness

Because of the wide range of difficulties, relationship counsellors tend to offer a variety of interventions rather than working within one therapeutic model, and it is unclear whether any one theoretical approach is generally more effective than another.[38] Brief, dynamic work can start the process of internal change, so that the couple is able to work on their own to practise new patterns of relating. Brief, focused work can offer immediate and early solutions to issues which might otherwise threaten the relationship. Longer interventions may be required for couples dealing with major life events and experiences, and offer the opportunity to look at childhood and other roots of persistent or destructive patterns of relating.

The effectiveness of counselling depends a great deal on the willingness of participants to engage in a process which can be painful, challenging often long-established patterns of relating, and one partner in the relationship may be more enthusiastic than the other to promote change. Research shows that many of those who experience relationship counselling understand themselves better, become less emotionally disturbed and understand their partner and relationships better.[39] In some cases, a good outcome is for the relationship to end, in a way which causes least disruption to all parties including children. In the latter case, referral to other agencies may be essential, such as when child protection issues are involved.

Grief counselling

Grief is not a pathological state in itself, and most people emerge from the natural grieving process in a healthy way. Counselling has a role both in facilitating grieving for those who experience difficulties in the process, and in helping those with complex grief reactions.[40] Counselling might involve more than one person in a bereaved family or other grouping, for example the college friends of a student killed in an accident.

In health settings grief counselling is undertaken by trained professionals or volunteers, and in the community by self-help voluntary agencies such as Cruse (in the United Kingdom) or groups attached to hospices.[41] Voluntary agencies are often staffed by people who have themselves experienced bereavement, and group counselling in this context provides a valuable opportunity for acceptance, sharing of experience, and the hope borne out of talking with others who have already come to terms with their loss.

The research on grief counselling[42] shows that professional services and professionally supported voluntary and self-help services can reduce the risk of psychiatric and physical problems following bereavement, and reduce the risk level of 'high-risk' widows to that of a 'low-risk' group. Reid *et al.*[43] found that support by hospice volunteers of high-risk bereaved relatives substantially reduced their levels of anxiety and need for medical care.

Drug and alcohol problems

Drug and alcohol dependence and related problems generate many controversies about their nature and treatment, such as whether people diagnosed as alcoholics can ever return to harm-free, controlled drinking, and the motivation and ability of those addicted to drugs and alcohol to change. Many volunteer agencies offer drug and alcohol treatments, advocating a multifaceted approach, with a combination of methods drawn from motivational interviewing, person-centred and cognitive behavioural therapy known as 'motivational enhancement therapy'. There are several different models of drug and alcohol use, including that of Narcotics Anonymous, similar to Alcoholics Anonymous, who view substance abuse as a pre-existing, biochemical abnormality, necessitating lifelong abstinence. Other views seek to minimize the harm caused by drugs and alcohol, by reducing risks, reducing intake, and possibly changing to another, less harmful substance. A third view sees addiction as a pattern of inappropriate coping: cognitive behavioural principles are used to recognize and deal with situations likely to lead to drug use.

A major trial in the United States randomized 487 patients to one of four, 6-month, treatments. All treatments included group drug counselling following a 12-step model, focusing on achieving abstinence. The group program was offered either alone, or in combination with individual drug counselling, CBT or individual supportive-expressive psychotherapy. Attrition rates in all groups were high, with only 28 per cent completing treatment. All interventions led to a reduction in drug use, but the greatest reduction was for individual plus group counselling.[44]

Counselling for recent and past trauma

Increased public awareness of global trauma arising from natural disasters, war, and terrorism, has led to the development of psychological interventions designed to prevent the onset of post-traumatic stress disorder (PTSD) in those exposed to traumatizing events. However, after many years of early interventions in the form of active single-session 'debriefing' for individuals and groups, there is no evidence for their effectiveness in preventing PTSD. One of the problems in the field is the tendency to 'medicalize' normal distress in traumatic situations, leading to the construction of 'disaster therapists', perhaps with limited understanding of the culture in which the disaster occurred, ready to offer advice and counselling to survivors who may not see themselves as having a mental health problem.[45]

There is also no evidence that non-directive counselling is effective in treating acute stress disorder, which itself constitutes a risk factor for PTSD.[46] Short individualized preventive interventions in the style of 'psychological first aid' may be most effective.

There is no evidence that non-directive or reflective counselling is effective in the treatment of post-traumatic stress disorder. The advocates of counselling for post-traumatic stress disorder describe active-focused methods such as cognitive behaviour therapy,[47]

and using debriefing, to enable people to build a cognitive and emotional account of their experiences. Given that many clients with PTSD are, understandably, mistrustful or avoidant, the methods have to be used within the context of a sound therapeutic relationship, meeting the core counselling conditions.

Postnatal depression

Postnatal depression is often mild and remits spontaneously for many women. However, effective treatment is important because of the potential adverse effect on the child's emotional and cognitive development. Home-based counselling is as effective as antidepressant medication in the treatment of postnatal depression, and is more acceptable to mothers.[48] Counselling can be delivered by health visitors trained in cognitive behavioural counselling methods.[49] The research suggests that depressed mothers benefit from an opportunity to talk about their concerns, not all of which necessarily focus on their baby, with a receptive and non-judgemental professional person. Counselling and other psychological interventions are highly acceptable to mothers with postnatal depression, and preferred over pharmacological treatment. If counselling and other forms of psychotherapy enable women to recover from postnatal depression more rapidly than usual care, this alone may be a valuable service to offer in addition to routine care.

Counselling settings

Counselling takes place in a large number of settings relevant to psychiatry. These include primary care, general medical settings, student counselling services, workplace counselling services, and the voluntary sector. These settings will be described below, aiming to discuss the ways in which counselling may be best adapted to the individual settings, with indications of outcome data on effectiveness.

Counselling in primary care

Primary care is one of the most conspicuous areas of growth in counselling, stimulated by greater demands for alternatives to medication for emotional problems, and by continuing debate about the most effective way of managing mental health problems in primary care. The Layard report on the need for CBT for depression and NICE guidelines stress the importance and effectiveness of psychological interventions including counselling, with a particular focus on CBT. It is likely that the number of counsellors who practice CBT will grow in order to meet demand.

In the United Kingdom, around half of general practices employ a qualified counsellor and the majority meet national criteria for good practice.[50] The development of UK primary care counselling is no doubt part of a wider international trend towards more accessible counselling services at the primary care level.

Counsellors in primary care are a diverse group, in the patients that they see, the counselling models used, and the length of counselling offered. Those identified as primary care counsellors include practice nurses, health visitors, and district nurses trained in counselling skills; clinical and counselling psychologists; community psychiatric nurses, and social workers, and qualified counsellors and psychotherapists. Counselling in primary care is usually provided through one of three main service delivery models[51]:

- Counsellors based in GP practices and provided by a local agency or cooperative
- Managed counselling services provided by the PCT or mental health trust, with counsellors based in GP practices or a central site
- Voluntary agencies (in cases where PCTs have contracts to refer to externally managed services in the voluntary sector)

Counsellors are valued in primary care for a number of reasons,[52] providing time for patients to talk through and reflect on problems, where general practitioners are unable to spend the necessary time on individual patients, as well as a valued alternative or addition to pharmacotherapy. Counselling in primary care also facilitates early identification and intervention for mental health difficulties.

Counsellors in primary care need to be flexible in the way that they work, using different models as appropriate to each client. They also need to be flexible about boundaries and confidentiality, communicating with general practitioners and other health professionals as appropriate. Counselling is generally six to eight 50-min sessions, with a maximum of 20 sessions, and therefore focusing on presenting difficulties rather than long-term issues. The role of the counsellor is varied, and may include offering individual or group counselling, offering advice or training to primary care staff on managing mental health problems, and general consultation.

Despite the growth and popularity of counselling in primary care, it is not clear how effective it is compared to other models such as CBT. Studies have shown mixed findings.[9] Trials comparing counselling for anxiety and depression with usual general practitioner care, CBT, and anti-depressant medication have shown significantly greater clinical effectiveness of counselling compared to usual general practitioner care in the short-term but not in the long-term.[53–55] For people with chronic depression, there were no significant differences between usual care, CBT and short-term psychodynamic counselling,[56] although at 12 months, both psychological therapies were superior to usual care. Counselling in primary care can be cautiously reported as a valuable service, particularly for people with mild to moderate emotional disturbance as well as bereavement and relationship difficulties. Clients improve in the short-term, and the service is appreciated and valued by general practitioners and patients. Primary care services are probably best offered as a range of mental health services, linking closely with community mental health services.

Counselling in general medical settings

Medical patients are understandably at higher risk of psychological difficulties compared to the general population, and many hospital departments and clinics employ counsellors as part of the multidisciplinary team to meet patients' psychosocial needs.[57] Counselling benefits patients in many hospital settings such as gastroenterology, cardiology, obstetrics, and gynaecology, the families of children with medical problems and disabilities and patients with diabetes, renal failure, disfigurement, cancer, head injuries, and chronic conditions such as multiple sclerosis.[58] Counsellors offer what is becoming increasingly limited in medicine—time. A large amount of psychosocial adjustment is needed in most serious illnesses and conditions, and counsellors in health care settings are assisting people cope with potentially life's most challenging moments.

For some conditions, such as HIV/AIDS and genetically transmitted illnesses, counselling forms an important aspect of treatment. From its outset, HIV/AIDS attracted the attention of psychological therapists, since at first there was little else to offer

to help people deal with a strange and potentially fatal illness. A range of professional and voluntary services grew to support those infected. The stages of the illness present different needs, from pre-testing counselling, dealing with the emotional, social, and physical consequences of a positive diagnosis, and managing the adjustment to living with a chronic condition. *Genetic counselling* is specialized branch of counselling practice with increasing applications within psychiatry. In the coming years, the expected identification of susceptibility genes for psychiatric disorders may bring new opportunities and expectations from patients and families for psychiatric genetics.[59]

Counselling is usually offered through referral to a specific hospital department or ward-based counsellor. Many health care workers and hospital chaplains use counselling skills, which are generally regarded as supportive and therapeutic for patients. This does not replace the need for managed counselling services delivered by trained and professional therapists.

Counselling in educational settings

Counselling services in college and university settings cater for students with a wide range of issues. Because of their age and developmental stage, many students have problems adjusting to the new freedoms and demands of college life and also face the developmental challenges of adolescence and young adulthood—conflicts between dependence and independence, psychosexual development, issues to do with self and body image and eating disorders. Other common difficulties include financial, study, and interpersonal problems. Mature students may also contend with the stress of juggling study with children and home-life, and being minorities within a younger peer group. With increasing admissions of overseas students, issues such as identity and loneliness will also arise.

Student counselling services are often arranged so that practical (e.g. financial guidance or careers counselling) and psychological help are offered separately to provide discreet and confidential access. Most clients refer themselves, but may be referred by staff or the student's doctor. Services need to include or work closely with psychiatrists and other mental health professionals in order to meet the needs of students with mental illness. The majority present with less severe emotional or psychological problems, but these may be highly disruptive to their studies and social integration.

Short-term counselling is usually appropriate for students, partly because of their urgency and the structure of the academic year, but also because their natural developmental potential enables most young people quickly to change. This process may be accelerated even more by the intelligence inherent in students, though emotional development can lag behind intellectual development. The task of counselling has been likened to helping the young person back on to the track of normal psychosexual development. More severe derailments, however, may require longer counselling, specialized psychotherapy, or psychiatric treatment.

Counselling in the voluntary sector

Counselling within the voluntary sector has vastly increased, with a growing number of support groups and voluntary organizations offering counselling, mainly on self-referral basis. The most well known in the United Kingdom include Alcoholics Anonymous, Cruse Bereavement Care, Relate for relationship difficulties, and the Samaritans, with equivalent organizations in other countries. Many mental health charities offer support, befriending, and counselling at a 'grassroots' level; some, such as MIND in the United Kingdom, are organized nationally, whereas others operate at a local level. These organizations contribute a great deal to mental health provision, offering the opportunity to talk through and reflect on problems, and in offering support to individuals with more severe mental health problems and their families.

The interface between voluntary and statutory services is varied and at times uncomfortable, with the two sharing different models of care, philosophies, and policies on issues such as confidentiality. There is less research on the effectiveness of counselling provision within the voluntary sector, although one study has shown it to be at least as effective as statutory provision and is often carried out by appropriately trained staff.[60]

Counselling in the workplace

Mental health and employment are known to be significantly related: satisfaction at work is positively correlated with mental health and the unemployed experience higher rates of mental health problems compared to those in employment. The provision of workplace counselling has steadily expanded over the past 20 years, with more than 75 per cent of medium and large organizations in Britain and North America making counselling available to their staff. Counselling may be part of a benefits package, occupational health, human resources or a service brought in to help with specific problems, such as redundancy. Workplace counselling can be viewed as the application of methods of brief psychological interventions that have been shown to be effective in other settings. However, a distinctive strength of seeing a counsellor in the workplace is that the counsellor will be sensitive to the combination of personal and work pressures that the person may present. Workplace counselling is a systemic, as well as individual, intervention in that the organization that pays the counsellor is always present, consciously or unconsciously, influencing the number of sessions and confidentiality boundaries.

Counselling for work-related difficulties is effective in reducing stress-related problems at work and sickness.[61,62] Those who receive counselling are highly satisfied, believing it helps them resolve their problems. Clinically significant improvement in levels of anxiety and depression are reported in 60–75 per cent of clients. Counselling is associated not only with reduction in sickness absence but also improvement in other organizational outcomes such as more positive work attitudes, fewer accidents, and enhanced work performance.

The main provision of counselling at work include employee assistance programmes (EAPs) and specialized staff counselling services.

Employee assistance programmes were first introduced in the United States after the Second World War, to rehabilitate oil-industry employees with alcohol problems. EAPs have become widespread in North America and are increasing in the United Kingdom. They are reported to achieve good results in terms of the percentage of employees who are rehabilitated for work, the reduction in alcohol consumption, improvement of work performance, and cost savings to the company.[63]

EAPs provide a comprehensive confidential counselling service to employees and their families, allowing employee's problems to be identified and resolved at an early stage, and are normally incorporated into the company's benefits package as a form of private emotional health care. EAPs include 24-h access, telephone

counselling, and helplines as well as individual counselling offered at short notice. One of the advantages of counselling organized through EAPs as opposed to in-house staff counselling is improved confidentiality: staff may be reluctant to use counselling services at work if they are not convinced of full confidentiality, and if they fear their career prospects may be adversely affected.

Staff counselling services: many private and public sector organizations now have in-house counselling services.[61] One of the first was set-up by the Post Office in the early 1980s, in recognition of the need to provide emotional and psychological support to employees. Mental health issues, mainly anxiety and depression, formed 46 per cent of the caseload, as well as relationship problems, alcohol problems, bereavement, assault, physical illness or disability, and social problems. Staff counselling schemes are now promoted by many health and education authorities, the Royal College of Nursing, the British Medical Association, and MIND at Work. Problem-solving and cognitive methods of counselling appear to be the most valuable models for workplace settings.[61]

Evaluation of the London Transport Counselling and Trauma Unit showed that the service made huge savings in its first year of operation, in terms of reduced sickness absence and other treatment costs. Research from the United States indicates a return for every dollar invested in Employee Assistance Programmes of between $3 and $7.[61,62] Other benefits may be less quantifiable but nevertheless valuable. For example, a qualitative study[64] of police officers and support staff who had received counselling for work-related difficulties, showed that many described themselves as learning something new and useful about themselves as a result of counselling. For example, an experienced detective stated, *I am 100 per cent better at listening now to a person.*

Telephone and electronically delivered counselling

The telephone is a valuable method of counselling, as shown by organizations such as The Samaritans. Telephone helplines respond to millions of calls each year, and offer a means of talking about feelings and gaining support, information, and advice. The telephone is excellent for crisis intervention and short-term work—one of the key reasons why people phone telephone helplines is because they are in crisis and want to talk anonymously and confidentially.

The telephone and teleconferencing can also be used to conduct some or all of a course of individual or group counselling, or as an adjunct to pharmacological treatment. Telephone advice and structured counselling improved outcome and satisfaction for clients starting antidepressant medication,[65] and also improved adherence to medication.[66]

Electronically delivered counselling: since the mid-1990s, new forms of text technology—the Internet, chatrooms, email, and mobile phone texting—are being developed to deliver counselling and psychotherapy. There is an increasing body of evidence that using text to conduct a therapeutic relationship is not only possible but also in many cases more desirable than face-to-face interaction. Electronically delivered counselling has been addressed by counselling organizations in the United Kingdom, United States and internationally, with published guidelines on issues such as confidentiality and data protection, contracting and informed consent, assessment of suitability of clients, boundaries, and practitioner competence.[67] The International Society for Mental Health Online (http://www.ismho.org/) was formed in 1997 to 'promote the understanding, use and development of online communication, information, and technology for the international mental health community'.

Working electronically has benefits those who cannot access therapy because of, for example, disability or geography. Communicating from a distance can make for a more honest and open relationship, clients diverging information or issues which they would find difficult to discuss face-to-face. Such disinhibition may be empowering—clients having a cathartic experience, so they can then disappear into cyberspace—or potentially hazardous, with traumatic issues remaining unresolved. Although the evaluation of such services is in its infancy, and there are many issues to be addressed, such as practitioner competence and confidentiality, electronic forms of mental health provision are likely to increase. They provide a means of providing help to clients who not only cannot but do not wish to meet with a helper, clients who have been traditionally excluded from mental health provision.

Counselling accreditation, training, and registration

The number of counsellors and psychotherapists in the United Kingdom and other countries is growing rapidly. Professional status and accountability are key concerns about counselling and counsellors, along with other providers of non-medical health care[68]—in the past, anyone could practise as a counsellor, with no standards of training, supervision, or attachment to a regulatory body. As a result, the vulnerable users of such practitioners are at risk. However, there is a strong move within counselling towards clear standards of training and accreditation, to be able to provide an effective and accountable service to the public.

For full *accreditation*, professional counsellors are required to follow, as a minimum, a 3-year full-time training course in the theory and practice of counselling. The standards required are covered by two main organizations in the United Kingdom, with equivalents in other countries:

- The British Association of Counselling and Psychotherapy, which offers accreditation for counsellors from a variety of disciplines, mainly humanistic and psychodynamic.
- The British Psychological Society, which offers accredited chartered status for counselling psychologists, trained to meet standards in applied psychology.

The requirements for accreditation or registration vary between organizations. Common requirements are a minimum of 450 taught contact hours and 450 h of supervised practice, evaluated via case studies and process reports; academic knowledge of counselling theory and research; personal counselling or psychotherapy; and, for the British Psychological Society, research skills and experience. The qualification requires that practitioners follow a code of practice and ethics, stipulating ethical practice, the need for supervision, appropriate confidentiality, and other standards for professional practice.

One of the problems in accreditation is ensuring that counsellors have followed a known and recognized training course. Alongside the expansion of interest in counselling, the number of counselling courses rises each year, ranging from short evening classes in active

listening and short courses in counselling skills for health professionals, to full-time training leading to professional accreditation or chartering. A recent review of UK training found that many psychotherapy and counselling organizations are small, and a significant number of training providers are linked to no external quality assurance systems, 63 per cent having no professional body recognition.[69] There are a large number of titles for both training courses and individual counsellors and psychotherapists, which can only cause confusion to both potential trainees and the public. The review[69] made a number of recommendations to improve the quality of training courses, to lead to greater standardization.

A further change within counselling is the move towards registration. Currently there is no statutory protection for using the terms 'counsellor' or 'psychotherapist', and therefore no means for the prevention of bad practice or abuse. Without registration through a professional body, clients may have no redress for incompetent practice. The National Service Framework for mental health, which sets clear standards for the delivery of effective services, emphasize the importance of 'talking therapies' but stresses the need to register counsellors, psychotherapists, and psychologists. The registering organizations are now moving towards statutory registration and legal protection for the terms counsellor, psychologist, and psychotherapist; and at the time of writing, definitive legislation is likely to be in place by 2008 or so. In the meantime, it is vital that the public and health professionals are aware of the need to seek help only from qualified practitioners of counselling.

Conclusions

For psychiatrists and other mental health care professionals, it can be difficult to make sense of the range of counselling models available, and it is therefore not surprising that potential purchasers and consumers are similarly confused about the nature and advisability of seeking counselling. Counselling has only recently been subject to the rigorous evaluation necessary to meet the standards of evidence-based practice. The most central aspects of counselling, the therapeutic relationship, and qualitative nature of the work, can be difficult to evaluate using established research methodology.

However, counselling is an evolving profession, moving away from, but not forgetting, its roots in Rogerian and psychoanalytic practice, and working hard to meet standards of effectiveness, evaluation, registration, and accountability.

Far more attention is being paid to evaluation, with the development of research paradigms suited to counselling. Systematic evaluation will eventually make it possible to identify, on the basis of clear evidence, the indications for specific models of counselling as well as their limitations. Issues of training, standards, ethics, and accountability are being addressed, to enable counselling to become fully consolidated and integrated within mental health services.

Counselling is a vital part of psychiatry for many reasons. There is a significant gap between the demand for psychological therapy and the available supply. One proposal to overcome this problem is to increase efficiency of provision through the adoption of briefer 'minimal interventions' within stepped care models.[70] Counselling is likely to play an increasingly important part of provision of psychological therapies, particularly for those with mild to moderate mental health problems and social and relationship difficulties. It makes sense for interventions to be offered at an early stage of difficulties: while many problems do resolve on their own, people welcome the support and understanding that counselling can offer, and it is valuable in hastening recovery from emotional distress. Counselling may be appropriate as part of the treatment for people with serious mental illness, offered by psychiatric nurses, social workers, occupational therapists, and workers in the voluntary sector.

Although CBT is being advocated as the treatment of choice for many psychological problems, there is a risk that over-enthusiastic endorsement of the benefits of CBT at the expense of other models may lead to a gap in care: not everyone responds to cognitive approaches, and even when they do, the level of recovery varies. Therefore, the breadth and repertoire of counsellors can add a variety and richness to mental health care provision. *It is this aspect of the human condition, the recognition that we must learn to 'weep for the plague, not just cure it', that is an essential component of meaningful therapy and meaningful relationships. When we experience what seems awful and horrible in our lives, we often take solace in knowing that another person understands, or, at least, is attempting to understand, our pain.*[71]

Acknowledgements

My thanks to Dr Sietske Boeles and Frank Wills for comments on earlier drafts of the chapter.

Further information

Bower, P. and Rowland, N. (2006). Effectiveness and cost effectiveness of counselling in primary care. *Cochrane Database Systematic Review* **3**: Art. No. CD001025.

Feltham, C. and Horton, I. (eds.) (2006). *The Sage handbook of counselling and psychotherapy* (2nd edn). Sage, London.

Nelson-Jones, R. (2006). *Theory and practice of counselling and therapy* (4th edn). Sage, London.

Roth, A. and Fonagy, P. (2005). *What works for whom? A critical review of psychotherapy research* (2nd edn). Guildford Press, New York.

Woolfe, R., Dryden, W., and Strawbridge, S. (eds.) (2003). *Handbook of counselling psychology* (2nd edn). Sage, London.

Resources

The British Association of Counselling and Psychotherapy represents counsellors and psychotherapists in the United Kingdom, with information, training and accreditation: www.bacp.co.uk.
Website for MIND, with information about counselling: www.mind.org.uk.
American Counseling Association: www.counseling.org.
American Mental Health Counselors Association: www.amhca.org.
Canadian Counselling Association: www.ccacc.ca.
Australian Counselling Association: www.theaca.net.au.
European Association for Counselling: www.eacnet.org.

References

1. Chilvers, C., Dewey, M., Fielding, K., *et al.* (2001). Antidepressant drugs and generic counselling for treatment of major depression in primary care: randomised trial with patient preference arms. *British Medical Journal*, **322**, 772–5.

2. Sibbald, B., Addington-Hall, J., Brenneman, D., *et al.* (1996). *The role of counsellors in general practice*. Royal College of General Practitioners, London.
3. Roth, A. and Fonagy, P. (2005). *What works for whom? A critical review of psychotherapy research* (2nd edn). Guilford Press, New York.
4. DOH. (2001). *Treatment choice in psychological therapies and counselling: evidence based clinical practice guidelines*. Department of Health, London.
5. Feltham, C. (2006). What are counselling and psychotherapy? In *The Sage handbook of counselling and psychotherapy* (eds. C. Feltham and I. Horton), pp. 3–10. Sage, London.
6. Burnard, P. (2006). *Counselling skills for health professionals* (4th edn). Nelson Thornes Ltd., Cheltenham.
7. Dryden, W. (2007). *The handbook of individual therapy* (5th edn). Sage, London.
8. Woolfe, R., Dryden, W., and Strawbridge, S. (eds.) (2003). *Handbook of counselling psychology* (2nd edn). Sage, London.
9. Rowland, N. and Goss, S. (eds.) (2000). *Evidence-based counselling and psychological therapies: research and applications*. Routledge, London.
10. McLeod, J. (2002). *Qualitative research in counselling and psychotherapy*. Sage, Thousand Oaks, CA.
11. Ashworth, M., Shepherd, M., Christey, J., *et al.* (2004). A client-generated psychometric instrument: the development of 'PSYCHLOPS'. *Counselling and Psychotherapy Research*, **4**, 27–31.
12. Rogers, C. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting and Clinical Psychology*, **21**, 95–103.
13. Sanders, D. and Wills, F. (2006). *Cognitive therapy: an introduction*. Sage, London.
14. Tolan, J. (2006). *Skills in person-centred counselling and psychotherapy*, p. 18. Sage, London.
15. Krupnick, J., Stosky, S., Simmens, S., *et al.* (1996). The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: findings in the national institute for mental health treatment of depression collaborative research program. *Journal of Consulting and Clinical Psychology*, **64**, 532–9.
16. Hubble, M., Duncan, B., and Miller, S. (eds.) (1999). *The heart and soul of change: what works in therapy*. American Psychological Association, Washington, DC.
17. Horton, I. (2000). Models of counselling and psychotherapy. In *The Sage handbook of counselling and psychotherapy* (eds. C. Feltham and I. Horton), pp. 234–6. Sage, London.
18. Nelson-Jones, R. (2005). *The theory and practice of counselling psychology* (4th edn). Sage, London.
19. O'Connell, B. (2005). *Solution-focused therapy* (2nd edn). Sage, London.
20. D'Zurilla, T. and Nezu, A. (2006). *Problem-solving therapy: a positive approach to clinical intervention* (3rd edn). Springer Publishing Company, Springer.
21. Mynors-Wallis, L., Gath, D., Day, A., *et al.* (2000). Randomised controlled trial of problem solving treatment, antidepressant medication, and combined treatment for major depression in primary care. *British Medical Journal*, **320**, 26–30.
22. Huibers, M., Beurskens, A., Bleijenberg, G., *et al.* (2003). The effectiveness of psychosocial interventions delivered by general practitioners. *Cochrane Database of Systematic Reviews*, **2**, CD003494.
23. Staring, A., Mulder, C., van-der-Gaag, M., *et al.* (2006). Understanding and improving treatment adherence in patients with psychotic disorders: a review and a proposed intervention. *Current Psychiatry Reviews*, **2**, 487–94.
24. Perls, F., Hefferline, R., and Goodman, D. (1994). *Gestalt therapy: excitement and growth in the human personality*. Souvenir Press, London.
25. Segal, Z., Williams, J., and Teasdale, J. (2002). *Mindfulness-based cognitive therapy for depression: a new approach to preventing relapse*. Guilford, New York.
26. Greenberg, L. (2002). *Emotion focused psychotherapy: coaching clients to work through their feelings*. American Psychological Association, Washington, DC.
27. Novey, T. (2002). Measuring the effectiveness of transactional analysis: an international study. *Transactional Analysis Journal*, **32**, 8–24.
28. Spinelli, E. (2007). *Practising existential psychotherapy: the relational world*. Sage, London.
29. Wills, F. (2006). CBT: can counsellors help fill the gap? *Healthcare Counselling and Psychotherapy*, **6**, 6–9.
30. Jacobs, M. (2006). *The presenting past: the core of psychodynamic counselling and therapy* (3rd edn). Open University Press, Milton Keynes.
31. O'Brien, M. and Houston, G. (2007). *Integrative therapy: a practitioner's guide* (2nd edn). Sage, London.
32. Lazarus, A. (2005). Multimodal therapy. In *Current psychotherapies* (eds. R. Corsini and D. Wedding), pp. 337–71. Thomson Brooks/Cole, Belmont, CA.
33. Ryle, A. and Kerr, I. (2002). *Introducing CAT principles and practice*. Wiley, Chichester.
34. Jenkins, P. (2003). Psychodynamic interpersonal therapy: bridging the gap within the medical model. *Healthcare Counselling and Psychotherapy*, **3**, 15–18.
35. Bower, P., Rowland, N., and Hardy, R. (2003). The clinical effectiveness of counselling in primary care: a systematic review and meta-analysis. *Psychological Medicine*, **33**, 203–15.
36. Barrowclough, C., King, P., Colville, J., *et al.* (2001). A randomized trial of the effectiveness of cognitive-behavioural therapy and supportive counselling for anxiety symptoms in older adults. *Journal of Consulting and Clinical Psychology*, **69**, 756–62.
37. Hill, D. (2006). Relationship problems. In *The Sage handbook of counselling and psychotherapy* (eds. C. Feltham and I. Horton), pp. 453–7. Sage, London.
38. Jacobson N. and Gurman, A. (eds.) (2002). *Clinical handbook of couple therapy*. Guildford, New York.
39. McCarthy, P., Walker, J., and Kain, J. (1998). *Telling it as it is: the client experience of Relate counselling*. Newcastle Centre for Family Studies, Newcastle.
40. Worden, J. (2005). *Grief counselling and grief therapy: a handbook for the mental health practitioner* (3rd edn). Bruner-Routledge, Hove.
41. Parkes, C., Relf, M., and Couldrick, A. (2002). *Counselling in terminal care and bereavement*. BPS Blackwell, Oxford.
42. Parkes, C. (2001). *Bereavement: studies of grief in adult life* (3rd edn). Taylor & Francis, Philadelphia.
43. Reid, D., Field, D., Payne, S., *et al.* (2006). Adult bereavement in five English hospices: types of support. *International Journal of Palliative Nursing*, **12**, 430–7.
44. Crits-Christoph, P., Siqueland, L., McCalmont. E., *et al.* (1999). Psychosocial treatments for cocaine dependence: national institute on drug abuse collaborative cocaine treatment study. *Archives of General Psychiatry*, **56**, 493–502.
45. Summerfield, D. (2006). Survivors of the tsunami: dealing with disaster. *Psychiatry*, **5**, 255–6.
46. Wessely, S. and Deahl, M. (2003). Psychological debriefing is a waste of time. *The British Journal of Psychiatry*, **183**, 12–14.
47. Foa, E., Keane, T., and Friedman, M. (eds.) (2004). *Effective treatments of PTSD*. Guildford, New York.
48. Cooper, P., Murray, L., Wilson, A., *et al.* (2003). Controlled trial of the short- and long-term effect of psychological treatment of post-partum depression. *The British Journal of Psychiatry*, **182**, 412–19.
49. Appleby, L., Hirst, E., Marshall, S. *et al.* (2003). The treatment of postnatal depression by health visitors: impact of brief training on skills and clinical practice. *Journal of Affective Disorders*, **77**, 261–6.

50. Mellor-Clark, J., Simms-Ellis, R., and Burton, M. (2001). *National survey of counsellors working in primary care: evidence for growing professionalization?* Royal College of General Practitioners, London.
51. BACP. (2006). *Shaping effective counselling services in health care: case studies of service delivery and outcomes*. British Association of Counselling and Psychotherapy, Rugby.
52. Keithley, J., Bond, T., and Marsh, G. (eds.) (2002). *Counselling in primary care*. Oxford University Press, Oxford.
53. Bower, P. and Rowland, N. (2006). Effectiveness and cost effectiveness of counselling in primary care. *Cochrane Database of Systematic Reviews*, **3**, CD001025.
54. King, M., Sibbald, B., and Ward, E. (2000). Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care. *Health Technology Assessment*, **4**, 1–83.
55. Ward, E., King, M., Lloyd, M., *et al.* (2000). Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy, and usual general practitioner care for patients with depression. I: clinical effectiveness. *British Medical Journal*, **321**, 1383–8.
56. Simpson, S., Corney, R., Fitzgerald, P., *et al.* (2000). A randomised controlled trial to evaluate the effectiveness and cost-effectiveness of counselling patients with chronic depression. *Health Technology Assessment*, **4**, 1–83.
57. Thomas, P., Davison, S., Rance, C. (eds.) (2001). *Clinical counselling in medical settings*. Brunner-Routledge, London.
58. Davis, H. and Fallowfield, L. (1996). *Counselling and communication in health care*. John Wiley & Sons, Chichester.
59. Finn, C. and Smoller, J. (2006). Genetic counseling in psychiatry. *Harvard Review of Psychiatry*, **14**, 109–21.
60. Moore, S. (2006). Voluntary sector counselling: has inadequate research resulted in a misunderstood and underutilised resource? *Counselling and Psychotherapy Research*, **6**, 221–6.
61. Greenwood, A. (2006). *Counselling for staff in health service settings*. Royal College of Nursing, London.
62. McLeod, J. and Henderson, M. (2003). Does workplace counselling work? *The British Journal of Psychiatry*, **182**, 103–4.
63. Carroll, M. and Walton, M. (eds.) (1999). *Handbook of counselling in organisations*. Sage, London.
64. Millar, A. (2002). Beyond resolution of presenting issues: clients' experiences of an in-house police counselling service. *Counselling and Psychotherapy Research*, **2**, 159–66.
65. Simon, G., Ludman, E., Tutty, S., *et al.* (2004). Telephone psychotherapy and telephone care management for primary care patients starting antidepressant treatment: a randomized controlled trial. *The Journal of the American Medical Association*, **292**, 935–42.
66. Tutty, S., Simon, G., and Ludman, E. (2000). Telephone counseling as an adjunct to antidepressant treatment in the primary care system. A pilot study. *Effective Clinical Practice*, **3**, 170–8.
67. Goss, S. and Anthony, K. (eds.) (2003). *Technology in counselling and psychotherapy. A practitioner's guide*. Palgrave Macmillan, Basingstoke.
68. DOH. (2006). *The regulation of the non-medical healthcare professions*. Department of Health, London.
69. Aldridge, S. and Pollard, J. (2005). *Interim report to the Department of Health on initial mapping project for psychotherapy and counselling*. British Association for Counselling and Psychotheraphy, Rugby.
70. Bower, P. and Gilbody, S. (2005). Stepped care in psychological therapies: access, effectiveness and efficiency: narrative literature review. *The British Journal of Psychiatry*, **186**, 11–7.
71. Leahy, R. (2001). *Overcoming resistance in cognitive therapy*. Guildford, New York.

6.3.2 Cognitive behaviour therapy

Contents

6.3.2.1 Cognitive behaviour therapy for anxiety disorders

David M. Clark

Introduction

Cognitive behaviour therapy for anxiety disorders is a brief psychological treatment (1 to 16 sessions), based on the cognitive model of emotional disorders. Within this model, it is assumed that it is not events per se, but rather people's expectations and interpretations of events, which are responsible for the production of negative emotions such as anxiety, anger, guilt, or sadness. In anxiety, the important interpretations, or cognitions, concern perceived physical or psychosocial danger. In everyday life, many situations are objectively dangerous. In such situations, individuals' perceptions are often realistic appraisals of the inherent danger. However, Beck[1] argues that in anxiety disorders, patients systematically overestimate the danger inherent in certain situations, bodily sensations, or mental processes. Overestimates of danger can arise from distorted estimates of the likelihood of a feared event, distorted estimates of the severity of the event, and/or distorted estimates of one's coping resources and the availability of rescue factors. Once a stimulus is interpreted as a source of danger, an 'anxiety programme' is activated. This is a pattern of responses that is probably inherited from our evolutionary past and originally served to protect us from harm in objectively dangerous primitive environments (such as attack from a predator). The programme includes changes in autonomic arousal as preparation for flight/fight/fainting and increased scanning of the environment for possible sources of danger. In modern life, there are also situations in which these responses are adaptive (such as getting out of the path of a speeding car). However, when, as in anxiety disorders, the danger is more imagined than real, these anxiety responses are largely inappropriate. Instead of serving a useful function, they contribute to a series of vicious circles that tend to maintain or exacerbate the anxiety disorder.

Two types of vicious circle are common in anxiety disorders. First, the reflexively elicited somatic and cognitive symptoms of anxiety become further sources of perceived danger. For example, blushing can be taken as an indication that one has made a fool of oneself, and this may lead to further embarrassment and

blushing; or a racing heart may be taken as evidence of an impending heart attack and this may produce further anxiety and cardiac symptoms. Second, patients often engage in behavioural and cognitive strategies that are intended to prevent the feared events from occurring. However, because the fears are unrealistic, the main effect of these strategies is to prevent patients from disconfirming their negative beliefs. For example, patients who fear that the unusual and racing thoughts experienced during panic attacks indicate that they are in danger of going mad and often try to control their thoughts and (erroneously) believe that if they had not done so, they would have gone mad.

Within cognitive models of anxiety disorders, at least two different levels of disturbed thinking are distinguished. First, negative automatic thoughts are those thoughts or images that are present in specific situations when an individual is anxious. For example, someone concerned about social evaluation might have the negative thought, 'They think I'm boring', while talking to a group of acquaintances. Second, dysfunctional assumptions are general beliefs, which individuals hold about the world and themselves which are said to make them prone to interpret specific situations in an excessively negative and dysfunctional fashion. For example, a rule involving an extreme equation of self-worth with social approval ('Unless I am liked by everyone, I am worthless') might make an individual particularly likely to interpret silent spells in conversation as an indication that others think one is boring.

Cognitive behaviour therapy attempts to treat anxiety disorders by (a) helping patients identify their negative danger-related thoughts and beliefs, and (b) modifying these cognitions and the behavioural and cognitive processes that normally maintain them. A wide range of procedures are used to achieve these aims, including education, discussion of evidence for and against the beliefs, imagery modification, attentional manipulations, exposure to feared stimuli, and numerous other behavioural assignments. Within sessions there is a strong emphasis on experiential work and on working with high affect. Between sessions, patients follow extensive homework assignments. As in cognitive behaviour therapy for other disorders, the general approach is one of collaborative empiricism in which patient and therapist view the patient's fearful thoughts as hypotheses to be critically examined and tested.

Background

Historical development of cognitive behaviour therapy

Modern cognitive behaviour therapy for anxiety owes its development to pioneering work since the 1950s and 1960s in which the principles of classical conditioning were applied to the understanding and treatment of phobias.[2] It was argued that (a) phobic stimuli are conditioned stimuli that acquired their aversive properties by being paired on one or more occasions with a traumatic event, and (b) avoidance is the main reason why phobias fail to extinguish. This suggestion led naturally to the development of various forms of exposure therapy, in which patients were systematically exposed to phobic stimuli. Initially therapists were concerned that elicitation of strong anxiety responses would be counter-therapeutic so exposure was very gradual, often starting with brief, imaginal presentations, followed by relaxation. Subsequent research showed that such a gentle approach was unnecessary and relatively rapid, *in vivo* exposure became the norm. By the mid-1970s, it was clear that up to 70 per cent of phobics obtained worthwhile improvements from *in vivo* exposure.[3] However, many were less than fully recovered and it was not clear how exposure therapy could be applied to non-phobic anxiety states (such as panic disorder and generalized anxiety disorder). In an attempt to enhance treatment effectiveness further, researchers attempted to identify additional factors that might maintain anxiety. Several cognitive processes outlined below received empirical support. As a consequence, more comprehensive cognitive behavioural treatments that attempt to modify a range of maintaining factors were developed. This chapter describes these treatments.

Cognitive content of anxiety disorders

Although there is no substitute for a careful assessment of each patient's ideation, research shows that most anxiety disorders are characterized by a specific type of fearful ideation and successful therapy generally focuses on such ideation.[4]

(a) Panic disorder

Panic disorder is characterized by a fear of an immediately impending internal disaster (e.g. heart attack, cessation of breathing, mental derangement) and a sense of loss of control over physical and mental functions. Many of panic patients' negative thoughts can be viewed as misinterpretations of normal bodily sensations (such as palpitations or a slight feeling of breathlessness). Indeed, cognitive theorists[5] argue that panic attacks result from a vicious circle in which catastrophic misinterpretations of body sensations lead to an increase in anxiety and associated sensations, which are in turn interpreted as further evidence of impending, internal disasters (e.g. heart attack, fainting, going mad). Panic disorder with agoraphobia is often also accompanied by fear of the interpersonal consequences of attacks (e.g. 'I'll make a fool of myself').

(b) Social phobia

Social phobia is characterized by exaggerated fears of being evaluated, of having one's weaknesses exposed, and of being judged adversely by other people. While in feared social situations, the social phobic continually monitors his or her performance, fears that this performance will be viewed as evidence that he or she is inept, boring, or stupid, and expects that such judgements will have dire long-lasting implications (loss of status or worth and failure to achieve key goals such as friendship, marriage, promotion). Often social phobics have excessively high standards for social performance (e.g. 'My speech must be perfectly fluent', 'I must always appear intelligent and witty'). Typically, social anxiety is triggered when individuals have a strong desire to convey a particular, favourable impression of them and have marked insecurity about their ability to do so.

(c) Generalized anxiety disorder

Generalized anxiety disorder is characterized by excessive worry about a number of life circumstances (e.g. finance, health, work, children, etc.) and the subjective impression that the worry is difficult to control.[6] Beck *et al.*[7] suggested that generalized anxiety disorder patients are anxious about many topics because their beliefs about themselves and the world make them prone to interpret a wide range of situations and circumstances in a threatening fashion. Although their beliefs are quite varied, Beck suggested that they mainly revolve around issues of acceptance, competence, responsibility, and control, as well as the symptoms of anxiety. Borkovec *et al.*[8] have shown that, compared with non-patients,

the worry of general anxiety disorder patients involves less imagery about specific feared outcomes and more verbal rumination in which problems are cast in a more abstract, more difficult to solve, form. Wells[9] has highlighted the importance of positive and negative beliefs about worry (meta-cognition).

(d) Obsessive–compulsive disorder

Obsessive–compulsive disorder is characterized by intrusive and distressing thoughts, impulses, or images about possible harm coming to oneself or others. Thoughts with a similar content to the intrusions of obsessional patients (e.g. a young mother having an intrusive thought about dropping her baby) are common in the general population.[10] For this reason, it has been suggested that the key cognitive abnormality in obsessive–compulsive disorder is not the content of obsessional thoughts, but rather the way the thoughts are interpreted.[11] In particular, it would appear that obsessional patients interpret recurrent obsessional thoughts and impulses as a sign that something terrible will happen, for which they will be responsible. For example, the young mother mentioned above may think that because she had a thought of dropping her baby, she is very likely to do so, despite finding the idea repugnant. In order to prevent the feared consequences of their obsessional thoughts, patients engage in a wide range of 'putting right' acts including (when relevant) washing and checking.

(e) Post-traumatic stress disorder

Surveys[12] indicate that unwanted, intrusive, and distressing memories and the other symptoms of post-traumatic stress disorder (avoidance of reminders and hyperarousal/numbing) are common immediately after traumatic events. Over the next few months many people recover but in a subgroup post-traumatic stress disorder becomes chronic. It is the latter group that normally present for treatment. Research indicates that chronic post-traumatic stress disorder is associated with appraising the traumatic event and/or its sequelae in a manner that would produce a sense of serious current threat to one's view of oneself and/or the world.[13] Examples are given in Table 6.3.2.1.1. There is also evidence that chronic post-traumatic stress disorder tends to be associated with a fragmented memory for the traumatic event and that recovery is associated with developing a more coherent narrative.[14,15]

Why do negative thoughts and beliefs persist?

If the world is not as dangerous as anxiety disorder patients assume, why do they not notice this and correct their thinking? For many patients with chronic anxiety disorders, the persistence of their fears can seem strangely irrational, at least at first glance. Consider, for example, panic disorder patients who think during their panic attacks that they are having a heart attack. Before they come for treatment they may have had several thousand panic attacks, in each one of which they thought they were dying, but they are not dead. Despite what might appear to an outsider as stunning disconfirmation of their belief that a panic attack can kill, their thinking has not changed.

Several factors that appear to prevent patients from changing their negative thinking are outlined below. Such factors are important because reversing them is likely to be a particularly efficient way of treating anxiety disorders.

(a) Avoidance, escape, and safety-seeking behaviours

Early conditioning theorists identified avoidance of, and escape from, feared stimuli as important factors in the maintenance of anxiety disorders. It is easy to see how avoidance of a feared situation (e.g. a supermarket for an agoraphobic) or escape from the situation before a feared event (e.g. a panic attack) occurs could prevent phobics from disconfirming their fears. However, situational avoidance/escape is not so obviously relevant to non-phobic anxiety and some phobics regularly endure feared situations without marked improvement in their fears. Salkovskis[16] introduced the concept of in-situation safety behaviours to deal with this problem. In particular, Salkovskis suggested that while in feared situations most patients engage in a variety of (often subtle) behaviours that are intended to prevent, or minimize, a feared outcome. For example, cardiac concerned panic disorder patients may sit down, rest, and slow down their breathing during attacks and believe, erroneously, that performing these safety behaviours is the reason why they did not die. Experimental studies have confirmed that (a) anxious patients engage in safety behaviours while in feared situations, and (b) dropping these behaviours facilitates fear reduction.[4]

Recent work[17] has highlighted several other important features of safety behaviours. First, although termed 'behaviours', many are internal mental processes. For example, patients with social phobia who are worried that what they say may not make sense and will sound stupid, often report memorizing what they have said and comparing it with what they are about to say, whilst speaking. If everything goes well, patients are likely to think 'It only went well because I did all the memorizing and checking; if I had just been myself people would have realized how stupid I was'. In this way their basic fear persists. Second, it is common for patients to engage in a large number of different safety behaviours while in a feared situation. Table 6.3.2.1.2 illustrates this point by summarizing the safety behaviours used by a patient who had a fear of blushing, especially while talking to men whom she thought other people would think were attractive. Third, safety behaviours can create some of the symptoms that patients fear. For example, responding to a feeling of breathlessness in panic attacks by breathing more quickly and deeply (hyperventilating) can enhance the feeling of being short of breath. Similarly, post-traumatic stress disorder patients who are concerned that unwanted intrusive recollections of the trauma mean they are going mad and often try hard to suppress such recollections. Unfortunately, active suppression increases the probability that the intrusion will occur. Fourth, some safety

Table 6.3.2.1.1 Some examples of idiosyncratic negative appraisals leading to a sense of current threat in post-traumatic stress disorder

What is appraised?	Negative appraisal
Fact that trauma happened	'Nowhere is safe'
One's behaviour/emotions during trauma	'I cannot cope with stress'; 'It was my fault'
Initial post-traumatic stress disorder symptoms	
Irritability, anger outbursts	'My personality has changed for the worse'
Flashbacks, intrusive recollections, and nightmares	'I'm going mad'; 'I'll lose control of my emotions'
Other people's reactions after trauma	
Positive responses	'They think I am too weak to cope on my own'
Negative responses	'Nobody is there for me'; 'I can't rely on other people'

Table 6.3.2.1.2 Safety behaviours associated with a fear of blushing

Feared outcome	Safety behaviour intended to prevent feared outcome
'My face (and neck) will go red'	Keep cool (open windows, drink cold water, avoid coffee, wear thin clothes) Avoid eye contact. If in a meeting, pretend to be writing notes Keep topic of conversation away from 'difficult' issues Tell myself the man is not really attractive. He's no more than a 2 (out of 10)
'If I do blush, people will notice'	Wear clothes (scarf, high collar) that would hide part of the blush Wear make-up to hide the blush Put hands over face. Hide face with long hair Stand in a dark part of the room
'If people notice, they will think badly of me'	Provide an alternative explanation for the red face, e.g. 'it's hot in here'. 'I'm in a terrible rush today', 'I'm recovering from flu', etc.

behaviours can draw other people's attention to problems that patients wish to hide. For example, a secretary who covered her face with her arms whenever she felt she was blushing discovered that colleagues in her office were much more likely to look at her when she did this than when she simply blushed. Finally, some safety behaviours influence other people in a way that tends to maintain the problem. For example, the tendency of social phobics to monitor continually what they have said, and how they think they come across, often makes them appear distant and preoccupied. Other people can interpret this as a sign that the phobic does not like them and, as a consequence, they respond to the phobic in a less warm and friendly fashion.

(b) Attentional deployment

Selective attention plays an important role in maintaining some anxiety disorders. Patients with panic disorder or hypochondriasis fear certain bodily sensations and symptoms, believing they indicate the presence of a serious physical disorder (heart attack, cardiac disease, cancer, etc.). Such patients have often had several medical investigations that indicate they do not have the physical illness(es) they fear, but they are not convinced. One reason appears to be that their fears lead them to focus attention on relevant parts of their bodies and, as a consequence of this attentional deployment, they become aware of benign bodily sensations that other people do not notice.[5] The presence of such sensations is then taken by the patient as evidence that a serious physical illness has been missed. (Hypochondriasis is classified as a somatoform disorder in DSM-IV[6] and as a somatization disorder in ICD-10.[18] However, it has many features in common with anxiety disorders and can be conceptualized as such for the purposes of psychological treatment).

Social phobia appears to be associated with two attentional biases. First, when in feared social situations, patients with social phobia report becoming highly self-focused, constantly monitoring how they think and feel they are coming across, and paying less attention to other people.[17] Reduced processing of other people means that social phobics have less chance to observe other people's responses in detail and, therefore, are unlikely to collect from other people's reactions information that would help them to see that they generally come across more positively than they think. Second, there is some evidence that when social phobics do focus on other people, they are particularly good at detecting negative reactions[19] and are poor at detecting positive reactions.

(c) Spontaneously occurring images

Spontaneously occurring images are common in anxiety disorders and also appear to play a role in maintenance. Patients with social phobia often report 'observer-perspective' images in which they see themselves as if viewed from outside.[20] Unfortunately, in their images they do not see what a true observer would see, but rather their fears visualized. For example, a teacher who was anxious about talking with colleagues in coffee breaks noticed that before speaking she felt tense around her lips. The tension would trigger an image in which she saw herself with a twisted and contorted mouth, looking like 'the village idiot'. At that moment, she was convinced everyone else thought she was stupid. Negative images are also used as information in other anxiety disorders. For example, obsessional patients who have images of committing a repugnant act (e.g. stabbing one's child) take the occurrence of the image as evidence that they are in danger of performing the act. Similarly, patients with post-traumatic stress disorder report that flashbacks increase the perceived likelihood of a future trauma.

(d) Emotional reasoning

A further source of misleading information that can enhance patients' perception of danger is anxiety itself.[21] For example, social phobics often think they look as anxious as they feel, but in general this is not the case.[22] Similarly, generalized anxiety disorder patients often take feeling on edge as a sign that something bad is about to happen.

(e) Memory processes

Some anxiety disorders are associated with a tendency for the selective recall of information that would appear to confirm the patient's worst fears. For example, high socially anxious individuals selectively recall negative information about the way they think they have appeared to others in the past when anticipating a stressful social interaction.[22] Similarly, patients with hypochondriasis selectively recall illness-related information. In post-traumatic stress disorder, a failure to elaborate memories at the time of the trauma and enhanced associative learning appear to play a key role in maintaining the re-experiencing symptoms.[13]

(f) Rumination

Anxious patients often spend protracted periods of time ruminating about negative things that could happen in the future and about how bad they would be. They may also ruminate about things that they feel have gone wrong in the past. Studies by Davey and Matchett[23] indicate that such rumination can enhance fear. There are several ways in which rumination might operate. First, thinking about an event may directly increase its subjective probability. Second, selectively focusing on past negative events, feelings, and impressions may further enhance the perceived likelihood of future danger. Third, rumination is rarely focused on constructively processing perceived threats, but instead often seems to elaborate the

threats or make them more abstract and hence difficult to deal with. For example, patients with post-traumatic stress disorder often ask themselves 'Could I have done something different?' during their traumatic event without thinking through in detail what their alternative options might have been, and how feasible they would have been at the time.

Treatment

Assessment interview

Table 6.3.2.1.3 summarizes the main topics covered in the assessment interview. The aims of the interview are as follows: (a) to obtain a detailed description of the patient's fears and behaviour; (b) to identify maintaining factors; (c) to normalize the problem; (d) to develop a model of the problem that can be used to guide treatment.

The interview would start by asking the patient to provide a brief description of the main presenting problem(s). For example, intense anxiety attacks, anxious apprehension, and avoidance of places where the attacks seem particularly likely or would be embarrassing. The interviewer then obtains a detailed description of a recent occasion when the problem occurred or was at its most marked. This would include the situation ('Where were you?', 'What were you doing?'), bodily reactions ('What did you notice in your body?', 'What sensations did you experience?'), thoughts ('At the moment you were feeling particularly anxious, what went through your mind? What was the worst that you thought might happen? Did you have an image/mental picture of that? How do you think you looked?'), behaviour ('What did you do?'), and the behaviour of others ('How did X react?', 'What did X say/do?'). Having obtained a detailed description of a recent occasion, the interviewer should check whether the occasion was typical. If not, further descriptions of other recent occasions should be elicited to provide a complete picture.

Table 6.3.2.1.3 Summary of topics to be covered in assessment interview

Brief description of presenting problem(s)
For each problem
Detailed description of a recent occasion when problem occurred/was at its most marked
Situation
Bodily reaction
Cognitions
Behaviour
List of situations when the problem is most likely to occur/be most severe
Modulators (things making it better or worse)
Possible maintaining factors
Avoidance of situations/activities
Safety behaviours
Attentional deployment
Faulty beliefs
Attitudes and behaviour of others
Medication
Beliefs about cause of the problem
Previous treatment (types, whether successful)
Onset and course
Personal strengths and assets
Social and financial circumstances

Next a list of situations in which the problem is most likely to occur or is most severe is elicited ('Are there any situations in which you are particularly likely to have a panic attack?'), together with information about modulators ('Are there any things that you notice make the symptoms stronger/more likely to occur?', 'Are there any things that you've noticed make the symptoms less likely/less severe/more controllable?').

Possible maintaining factors should be identified, including the following:

- avoidance of situations or activities ('What situations/activities do you avoid because of your fears?')
- safety behaviours ('When you are afraid that X might happen, is there anything you do to try to stop it happening?')
- attentional deployment ('What happens to your attention when you are worried about X? Do you focus more on your body? Do you become self-conscious?')
- faulty beliefs (e.g. an obsessive–compulsive disorder patient, believing that thinking something can make it happen)
- attitudes and behaviour of significant others ('What does Y think about the problem?'; 'What does Y do when you are particularly anxious?')
- current medication

There are several ways in which excessive use of both prescribed and non-prescribed medications can maintain anxiety disorders. For example, painkillers and tranquillizers can cause derealization and sleep disturbance respectively, and drinking before social occasions prevents disconfirmation of one's social fears.

It is also important to assess patients' beliefs about the cause of their problems as some beliefs may make it difficult for patients to engage in therapy. For example, patients with post-traumatic stress disorder who think the best way of dealing with a painful memory is to push it out of their mind are unlikely to engage in imaginal reliving of the event until this belief is dealt with.

Finally, a brief description of the onset and subsequent course of the problem should be obtained. This description should particularly focus on factors, which may have been responsible for initial onset and for fluctuations in the course of the symptoms and is primarily used to make the development of the problem seem understandable to the patient.

It is not always possible to obtain all the information needed for a cognitive behavioural formulation in an assessment interview. Sometimes it is necessary to follow-up the interview with homework assignments in which the patient collects more information to clarify the formulation. For example, a hypochondriacal patient who was concerned that palpitations meant that she had cardiac disease was asked to record what she did each hour and how many palpitations she experienced. To her surprise, palpitations were not associated with exercise, as she expected, but rather were most common when she was sitting quietly, reading, watching television, or studying. This realization helped convince her that her problem may be disease preoccupation rather than a faulty heart.

Developing an idiosyncratic model of the patient's problem

Assessment ends with the development of an idiosyncratic version of the cognitive model. In particular, therapists aim to show patients

how the specific triggers for their anxiety produce negative automatic thoughts relating to feared outcomes and how these are maintained by safety behaviours and other maintenance processes. The model is usually drawn on a whiteboard, so that patient and therapist can look at it and discuss it together. Figure 6.3.2.1.1 shows an example for a panic disorder patient. His panic attack started with a twinge in his chest muscles, and he then had the thought, 'There is something wrong with my chest area, maybe I am having a heart attack'. This interpretation made him start to feel anxious, his chest muscles tightened up more, he started to feel dizzy, his heart raced more, and he then thought, 'I'm dying, I'm having a heart attack', and also, interestingly, 'If I don't die, people will notice I'm anxious and think it is odd'. He then engaged in a series of safety behaviours to try to prevent himself from dying. He thought he had read somewhere that paracetamol (aminacetophen) is good for people with heart problems and so he took a paracetamol. This is incorrect information, but the key point is that he believed it. He also sat down and rested, took the strain off his heart, and took deep breaths, trying to slow down his heart rate. He believed that the main reason he had not died was that he had engaged in the safety behaviours. The reader will also notice that some of the safety behaviours (taking deep breaths and monitoring the heart) will also have augmented his feared symptoms.

Figure 6.3.2.1.2 shows a further example with a social phobic patient. The patient's main fear was that other people would think she was stupid and boring. The situation used to develop the model was a recent coffee break at work during which the patient had difficulty joining a conversation with colleagues. When attempting to join the conversation she had the thought, 'I'll sound stupid and everyone will think I am dumb'. In order to prevent herself from sounding stupid, she engaged in an extensive set of safety behaviours which (a) prevented her from discovering that her spontaneous thoughts are interesting to other people, (b) made her appear preoccupied and uninterested in her colleagues, and (c) made her excessively self-conscious. While self-conscious, she became particularly aware of anxiety symptoms (sweaty palms, stiff muscles around her mouth) that she thought other people might see, and indeed, had an image of herself in which she looked very strange, with a twisted and rigid mouth and appeared stupid.

Normally idiosyncratic models of the form illustrated in Figs 6.3.2.1.1 and 6.3.2.1.2 will be developed at the end of the first interview, and certainly not later than the second session. Such models are used as blueprints to help therapist and patient organize and develop the rest of therapy.

Monitoring progress

Once treatment has started, it is important to monitor progress continually in order to decide whether a particular treatment procedure is working or whether the case needs reformulating and new treatment procedures need to be implemented. Usually patients are asked to complete a small number of self-report questionnaires before each therapy session. Typically, these include frequency and severity ratings for the main anxiety problems (often using simple 0–8 Likert-type scales), a measure of negative thoughts, and general measures of anxiety and depression (such as the Beck Anxiety Inventory[24] and the Beck Depression Inventory[25]). Table 6.3.2.1.4 summarizes some of the most commonly used weekly measures. In some instances these are supplemented by more individualized diaries and ratings. More global standardized measures of symptom severity are also often administered at the beginning, middle, and end of therapy in order to provide normative data (see Table 6.3.2.1.4).

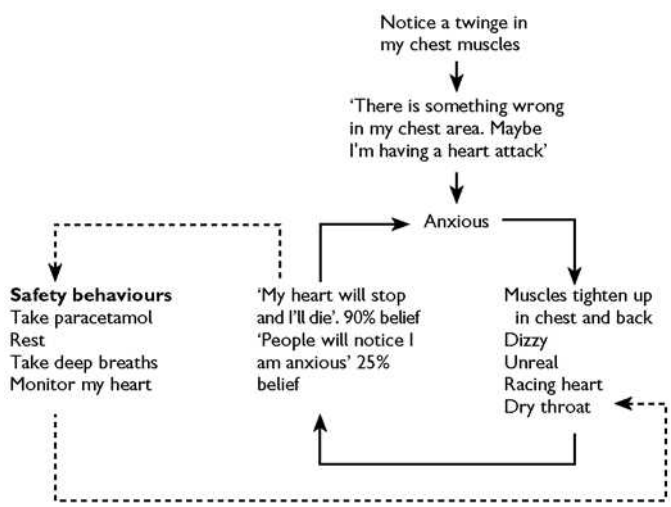


Fig. 6.3.2.1.1 A cognitive model of a patient's panic attacks. (Reproduced from Clark, D.M., Panic disorder: from theory to therapy. In *Frontiers of cognitive therapy* (ed. P.M. Salkovskis), pp. 318–344, Copyright 1996, Guildford Press, New York.)

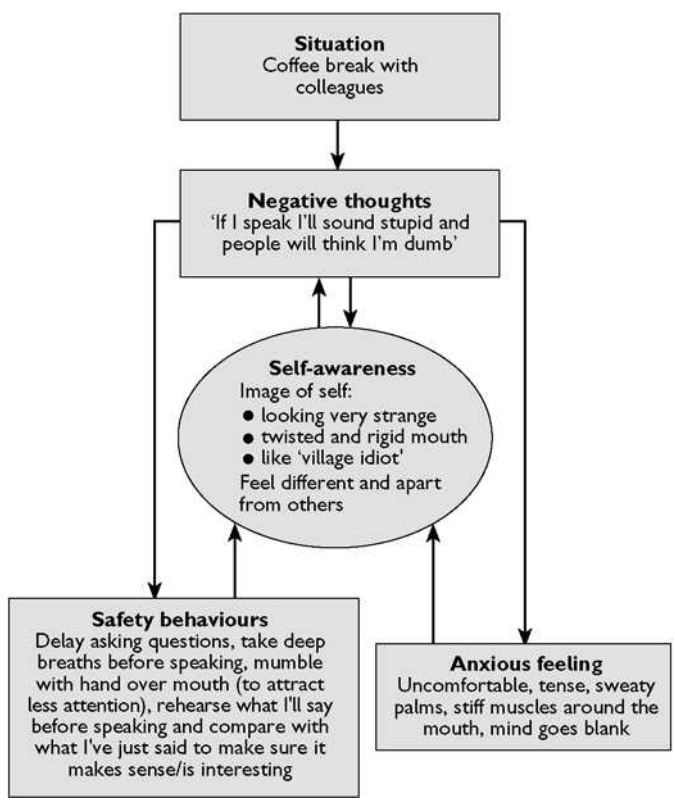


Fig. 6.3.2.1.2 A cognitive model for a patient with social phobia.

Treatment procedures

A wide range of procedures can be used to modify patients' negative beliefs and linked maintenance processes. For clarity the procedures are described separately. However, in practice the techniques are closely interwoven. Within a given session, therapists will usually use a mixture of discussion and experiential techniques to help patients to challenge convincingly their negative beliefs. As with cognitive behaviour therapy for other disorders, patients are given extensive homework assignments and it is assumed that a sizeable amount of therapeutic change is the result of homework assignments.

(a) Identifying patients' evidence for their negative beliefs

Anxiety disorder patients usually have reasons for believing that the things they fear are dangerous, however strange their fears may seem. The therapist, therefore, tries to 'get inside the patient's head' and see what the evidence is. Often the evidence is an event or piece of information that the patient has misinterpreted. Identifying and correcting such misinterpretations can be helpful. For example, a panic disorder patient believed that experiencing high anxiety could kill her. When asked by the therapist what her evidence was, she explained that she had seen it happen. Further enquiry revealed she had entered Dresden the day after the fire bombing of that city by the allies during the Second World War and had helped search for survivors. When opening up cellars below demolished houses, she repeatedly observed that the occupants were either dead or behaved in a dazed confused manner, even though the fire had not entered their cellars. She concluded that fear had killed the occupants or sent them mad. However, further questioning from the therapist revealed that the cellar occupants all had bright cherry-red lips. This allowed the therapist to explain that they were suffering from carbon monoxide poisoning, not the effects of intense fear. This correction considerably reduced the patient's fear of anxiety.

(b) Education

Education about the symptoms of anxiety is often helpful, especially if it directly targets patients' idiosyncratic fears and concerns. For example, post-traumatic stress disorder patients often think their flashbacks and emotional outbursts mean they are going mad or have permanently changed for the worse. In such cases, detailed assessment of the patient's post-traumatic stress disorder symptoms and explanation that each are common reactions to a trauma can greatly help. Similarly, panic disorder patients with cardiac concerns often cite left-sided chest pain as evidence for their belief that they have a cardiac disorder. In such cases discussion of Fig. 6.3.2.1.3 (from a study of chest pain in patients referred to a cardiac clinic[26]) is useful. In particular, the patient discovers that left-sided chest pain is more characteristic of non-cardiac chest pain than of either confirmed angina or myocardial infarction. Further questioning helps patients to see that the association between left-sided pain and attacks is probably a consequence of their fears. That is to say, they can experience pain on either side of the chest but only panic when it is on the left side. Finally, patients with obsessive–compulsive disorder who are perturbed by the apparently repulsive and unusual nature of their intrusive thoughts often benefit from reviewing Rachman and De Silva's classic

Table 6.3.2.1.4 Commonly used measures for monitoring progress

Anxiety disorder	Measure		
	Symptoms	**Thoughts**	**Global severity**
Panic disorder	Panic Rating Scale[63] Panic Diary[63] BAI[33] BDI[2]	Agoraphobic Cognitions Questionnaire[65]	Fear Questionnaire[64] Mobility Inventory[66]
Social phobia	Social Summary Scales (Table 4) BAI[33] BDI[2]	Social Cognitions Questionnaire[53]	Liebowitz Social Anxiety Scale[67] Social Performance Scale[68] Social Interaction Anxiety Scale'[68]' Social Phobia and Anxiety Inventory[69]
Generalized anxiety disorder	BAI[33] BDI[2]	Worry Domains Questionnaire[70] Thought Control Questionnaire[72]	Penn State Worry Questionnaire[71] Spielberger State Trait Inventory[73]
Obsessive- compulsive disorder	BAI[33] BDI[2]	Responsibility Interpretations Questionnaire[74]	Padua Inventory[75] Yale–Brown Obsessive Compulsive Scale[76]
Post-traumatic stress disorder	Post-traumatic Diagnosis Scale[77,78] BAI[33] BDI[2]	Post-traumatic Cognitions Inventory[79] Personal Beliefs and Reactions Scale[81]	Impact of Events Scale[80] Post-traumatic Diagnosis Scale[78]

Owing to their length, the Post-traumatic Cognitions Inventory and the Personal Beliefs and Reactions Scale are not suitable for weekly administration.

BAI, Beck Anxiety Inventory; BDI. Beck Depression Inventory.

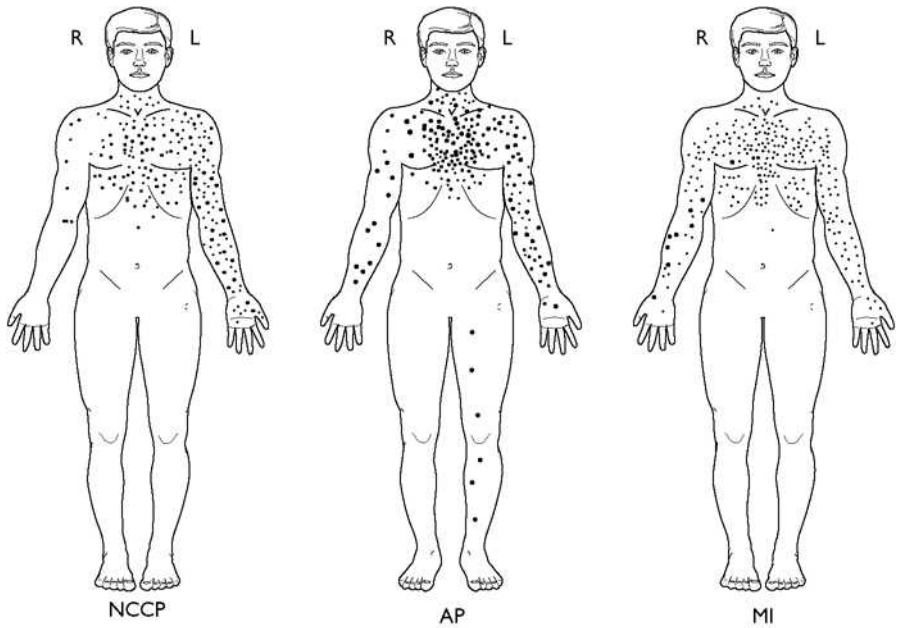


Fig. 6.3.2.1.3 Distribution of chest pain in patients referred to a cardiac clinic and subsequently diagnosed as non-cardiac chest pain (NCCP), angina pectoris (AP), or myocardial infarction (MI). (Reproduced from Beunderman, R. *et al.* (1998), Differentiation in prodromal and acute symptoms of patients with cardiac and non-cardiac chest pain, In *Advances in theory and practice in behaviour therapy* (ed. P.M.G. Emmelkamp, *et al.*), Copyright 1998, Swets and Seitlinger, Taylor and Francis Group, an informa business.)

paper[10] which demonstrated that thoughts with identical content to obsessional intrusions are common in the general population.

(c) Identifying observations that contradict patients' negative beliefs

As anxiety disorder patients' beliefs about the dangerousness of feared stimuli are generally mistaken, patients have often experienced a number of events that contradict their beliefs before they come into therapy. Therapists can make considerable progress, even in an assessment interview, by spotting these events and helping patients understand their significance. For example, panic disorder patients who are worried that their symptoms mean they are about to have a heart attack, often report that in some attacks something unexpected happened to distract them (e.g. a telephone call) and then their symptoms went away. Therapists could then pause and help the patient understand what this means, perhaps asking, 'Would a cardiologist prescribe telephone calls as a treatment for a heart attack?' The patient would probably answer 'No', to which the therapist might reply, 'If telephone calls would not stop a heart attack, how might they work? If the problem was the negative thought, could they help (by distracting one from the thoughts)?'

(d) Imagery modification

Images play an important role in many anxiety disorders. Most images represent feared catastrophes and can be treated as predictions to be tested (see behavioural experiments below). However, when the images are stereotyped and repetitive it is often also necessary to work directly with the images and to restructure them explicitly.

The problem with anxiety-related images is that they seem very realistic at the time they occur and, as a consequence, greatly enhance fear. A common restructuring technique involves discussing with the patient whether the image is realistic. Once it is intellectually agreed that the image is an exaggeration, patients are asked to recreate intentionally the negative image and to hold it in mind until they start to feel anxious. They are then asked to transform it into a more realistic image, or an image, which convincingly indicates that the original image was unrealistic. A common observation is that patients' spontaneous images generally stop at the worst moment. For example, agoraphobic patients who fear fainting in a supermarket might see themselves collapsed on the floor, but not see themselves getting up, recovering, and going home. A useful transformation in such cases is to 'finish out' the image by asking patients to run it on until they see the positive resolution. Of course, sometimes simply running on an image does not produce a positive resolution. For example, a patient who feared she would go mad frequently experienced an image of two men in white coats entering her house to take her away to a locked ward. In the image, the men were extremely powerful and she felt powerless. Transformation, following suggestions from her, involved shrinking the men and then turning them into ridiculous looking (and hence non-threatening) white poodles.

An interesting observation about spontaneous imagery is that it often fails to incorporate positive information that would seriously undermine the impact of the image, even when the patient has such information. For example, a mother whose children died in a house fire, repeatedly experienced intrusive flashbacks in which she saw the house going up in flames and smelled burning flesh, despite having seen her children in the mortuary, knowing that they had not been burnt, but instead were rapidly overcome by fumes.

For imagery restructuring to be effective it is important that it is not done as a cold, intellectual exercise, but instead includes eliciting the affect normally associated with the image. Transformation may have to be done in several steps. It is often best to start with the

most threatening aspect of the image. Possible alternative images should be generated by patients, rather than simply imposed by the therapist.

(e) Cognitive restructuring

All the above techniques are examples of cognitive restructuring in which the therapist provides information and asks a series of questions to help the patients challenge their fearful thoughts and images. A list of some of the questions that can be particularly useful for helping anxiety disorder patients challenge their negative thoughts is given in Table 6.3.2.1.5. Further useful questioning techniques can be found in Chapter 6.3.2.3.

It is sometimes helpful to use graphical methods for discussing alternatives to negative thoughts. In situations where there are several non-threatening alternative explanations for a feared event, pie charts are particularly useful. When constructing a pie chart the therapist draws a circle which is meant to represent all the possible causes of a particular event and asks the patient to list all the possible non-catastrophic causes of the event and allocate a section of the circle to each cause. At the end of the exercise, there is often very little of the circle left for the patient's negative explanations. Figure 6.3.2.1.4 illustrates the use of a pie chart to challenge a generalized anxiety disorder patient's belief that he would be 100 per cent responsible for people not enjoying themselves at his dinner parties. The belief was preventing him from making new social contacts after a painful divorce. Pie charts are particularly helpful for dealing with distorted beliefs about responsibility and hypochondriacal concerns (e.g. 'Headaches mean I have a brain tumour').

When considering the worst that could possibly happen in a feared situation patients frequently ignore the fact that there are many intermediate events, each with a probability of less than 1, which have to occur for the catastrophe to be realized. The inverted pyramid can be a good way of representing this. Figure 6.3.2.1.5 shows an example with a patient who was afraid of blushing. His worst fear was that other people would think they were greatly superior to him if he blushed. Whenever he felt his face becoming hot, he was convinced other people were thinking they are superior to him and gloating. However, careful discussion helped him to see that there were many intermediate steps between him feeling hot and the feared outcome. Once the conditional probabilities were taken into account, there was only a minute chance that his worst fear would be correct.

Table 6.3.2.1.5 Useful questions for challenging anxiety-related thoughts

What is the evidence for this thought?
Is there any alternative way of looking at the situation?
Is there an alternative explanation?
How would someone else think about the situation?
Are you focusing on how you felt, rather than on what actually happened?
Are you setting yourself an unrealistic or unobtainable standard?
Are you forgetting relevant facts or overfocusing on irrelevant facts?
Are you thinking in all-or-nothing terms?
Are you overestimating how responsible you are for the way things work out?
What if the worst happens? What would be so bad about that? How could you cope?
How will things be x months/years afterwards?
Are you overestimating how likely the event is?
Are you underestimating what you can do to deal with the problem/situation?

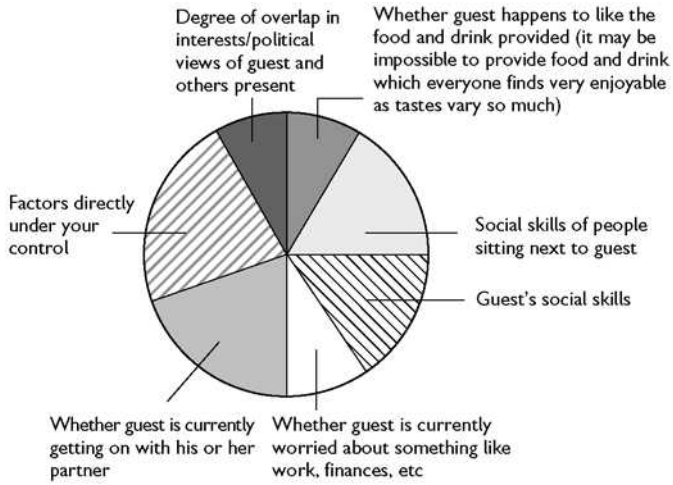


Fig. 6.3.2.1.4 A pie chart representing factors that might contribute to guests enjoying themselves at a dinner parties to challenge a generalized anxiety disorder patient's belief that he would be 100 per cent responsible for people not enjoying themselves. (Reproduced from Clark, D.M., Anxiety states: panic and generalized anxiety, In *Cognitive therapy for psychiatric problems: a practical guide* (ed. K. Hawton, *et al.*), pp. 52–96. Copyright 1989 with permission from Oxford University Press.)

It is important to remember that anxiety results from overestimating the cost of feared events as well as their probability. Discussions aimed at modifying perceived cost are often helpful. This can be true even in cases where it might seem obvious that the feared event is objectively costly. For example, in hypochondriacal patients who are worried about dying, therapists may be tempted to focus exclusively on whether or not the patients are likely to die from the symptoms they are concerned about. Accepting that dying is a bad thing, the therapist may not be inclined to ask, 'What would be so bad about dying?' However, Wells and Hackmann[27] found that many hypochondriacal patients have distorted beliefs and images about death and the process of dying. For example, they think that when they die they will remain conscious and will continue to experience all the pain they had up to that point. Such people can benefit greatly from discussion of their beliefs about the cost of dying.

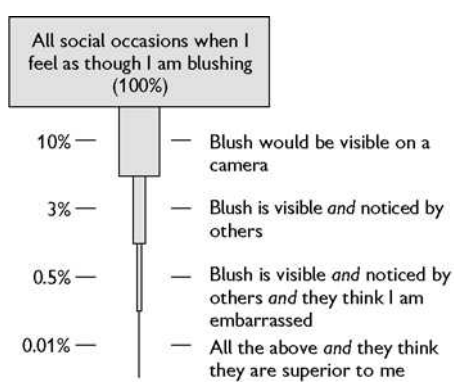


Fig. 6.3.2.1.5 An inverted pyramid representing conditional probabilities between 'feeling hot' and 'others thinking they are superior to you' constructed for a patient who was afraid of blushing.

(f) *In vivo* exposure to feared situations, activities, and sensations

Systematic exposure to feared and avoided situations has a long history in cognitive behaviour therapy and is one of the most effective ways of helping patients to discover that the things they are afraid of will not happen or are more manageable than they anticipate. Initially, exposure was often conducted in imagination but it is now known that *in vivo* exposure is a more effective way of dealing with situational fear.[3] During the 1970s and 1980s the dominant framework for exposure was habituation. It was assumed that repeated prolonged exposure was required to achieve fear reduction. More recent cognitive formulations have suggested that exposure is likely to be optimally effective when set-up in a way that maximizes the extent to which patients are able to disconfirm their fears, and considerable attention is now devoted to setting up exposure assignments in a way that will maximize cognitive change. Before entering a feared situation, patients are asked to specify what is the worst they think could happen, how likely they think it is, and what they would normally do to prevent the feared catastrophe (safety behaviours). They are then asked to enter the feared situation while dropping their safety behaviours and to observe carefully whether the feared outcome occurs. Afterwards, discussion focuses on whether the feared catastrophe occurred. If it did not, how does the patient explain its non-occurrence? Was it because the patient now thinks the feared outcome is unrealistic or does the patient think it was because of 'luck' or the continued use of safety behaviours? In the latter two instances, further exposure assignments with further encouragement to drop safety behaviours are required.

In addition to avoiding feared situations, anxiety disorder patients can also avoid feared sensations. Such avoidance is particularly prominent in panic disorder. For example, because of their fears about the meaning of increases in heart rate, dizziness, sweating, and other autonomic cues, panic disorder patients often avoid exercise. Increasing exercise can be an excellent way of helping them to challenge their negative beliefs, as can other ways of inducing bodily sensations such as ingesting caffeine, and hyperventilating. In each instance, the key point is to help patients discover that they can experience intense physical sensations without dying, losing control, or experiencing some other catastrophe.

Table 6.3.2.1.6 shows a record sheet that can be useful for planning and summarizing the results of exposure assignments, with illustrations from patients with social phobia and agoraphobia. Because of the intensity of patients' fears, and their tendency to attribute good outcomes to 'luck', it is often necessary to move up a hierarchy of feared situations and to consolidate successes by repetition.

In obsessive–compulsive disorder, the compulsive rituals act as safety behaviours and it is necessary to ensure that patients refrain from engaging in rituals (which are often also termed 'putting right' acts) during exposure assignments. This procedure is called 'exposure and response prevention'. For example, obsessional washers would be asked to 'contaminate' themselves by touching feared objects and then not put things right by washing. Similarly, obsessional checkers may be asked to expose themselves to activities that would normally provoke their checking (e.g. turning on the gas cooker) and then refrain from checking more than would be normal. In both instances, patients usually find that although exposure initially provokes considerable distress, the distress systematically declines during prolonged response prevention.[28]

Unlike most phobic fears, the fears of obsessive–compulsive disorder patients (e.g. developing a fatal disease from touching an object that is believed to be contaminated) often cannot be disconfirmed during a single or indeed multiple, exposure assignments. Discussing this issue, Salkovskis[11] has suggested that exposure and response prevention may work by providing patients with a different understanding of their problems. In particular, the decline

Table 6.3.2.1.6 Record sheet for noting behavioural experiments

	Situation	Prediction	Experiment	Outcome	What I learned
		(What exactly did you think would happen? How would you know?) (Rate belief 0–100%)	(What did you do to test the prediction?)	(What actually happened? Was the prediction correct?)	(1) Balanced view (rate belief 0–100%)? (2) How likely is what you predicted to happen in future (Rate belief 0–100%)
Social phobic patient	Coffee break; sitting with other teachers; trying to join in the conversation	If I just say things as they come into my mind, they'll think I'm stupid (50%)	Say whatever comes into my mind *and* watch them like a hawk; don't focus on myself; this only gives me misleading information (such as images of myself as the 'village idiot'), and means I can't see them	I did it and I watched the others; one of them showed interest and we talked; she seemed to quite enjoy it	I am probably more acceptable than I think (70%)
Agoraphobic patient	Shopping in a supermarket	I will feel dizzy and have a panic attack (90%). Unless I grip the trolley tightly or sit down at that moment, I collapse (80%)	Go into the supermarket. When I start to feel dizzy, remind myself it is just anxiety, my heart rate is up and I can't faint. Then move away from the trolley and stand unsupported	I felt dizzy but didn't faint, even though I didn't sit down or hold on to the trolley	Feeling dizzy in anxiety attacks will not make me faint (60%)

in distress during exposure and response prevention helps the patient to discover that they are suffering from a worry problem, rather than being in objective danger.

(g) Imaginal exposure in post-traumatic stress disorder

Although imaginal exposure is rarely used in most anxiety disorders, it plays an important role in the treatment of post-traumatic stress disorder. It is known[29] that avoidance of thinking about the traumatic event is an important predictor of persistent post-traumatic stress disorder. In the light of this finding, clinicians have attempted to treat post-traumatic stress disorder by repeated, imaginal reliving of the traumatic event, and controlled trials[30] have shown that this technique is effective. At this stage it is not known why reliving works. One suggestion is that the intrusive symptoms of post-traumatic stress disorder are the result of a fragmented and disorganized memory for the trauma that is poorly integrated with other autobiographical information. Reliving might, therefore, facilitate the production of a more organized narrative account of the event that can be placed in the broader context of the individual's life.[13,31] Two types of imaginal reliving have been used in controlled trials: writing out details of the event and reliving the event in imagery. In either case, it seems important to focus not only on what happened, but also on patients' feelings and thoughts, both at the time and now, looking back at the event. Problematic idiosyncratic meanings that can be addressed with cognitive restructuring are often identified during reliving exercises.

(h) Behavioural experiments

Behavioural experiments play a central role in the treatment of anxiety disorders. In a behavioural experiment, therapist and patient plan and implement a behavioural assignment that will provide a test of a key belief. The *in vivo* exposure assignments outlined above are examples of behavioural experiments. Several further examples are given to illustrate the technique.

Patients with post-traumatic stress disorder often think their intrusive recollections mean they are going mad or losing control in some way, and as a consequence, try to push the intrusions out of their mind. If this problem is identified during the first session of therapy, therapists often conduct an experiment to illustrate the undesired consequences of thought suppression. For example, the therapist might say to the patient, 'It doesn't matter what you think about in the next few minutes as long as you don't think about one particular thing. The thing is a fluorescent green rabbit eating my hair!' Most patients find they immediately get an image of the rabbit and have difficulty getting rid of it. Discussion then helps them to see that an increase in the frequency of target thoughts is a normal consequence of thought suppression. This result can then be used to set-up a homework assignment in which the patient is asked to collect data to test the idea that thought suppression may be enhancing intrusions. The experiment involves not trying to push the intrusions out of one's mind, but instead just letting them come and go, watching them as though they were a train passing through a station. Often patients report this simple experiment produces a marked decline in both the frequency of intrusions and the belief that they are a sign of impending insanity or loss of control.

Patients with social phobia often overestimate the significance of their anxiety symptoms for other people. A useful behavioural experiment to illustrate this point involves having either the patient or the therapist conduct a survey in which other people are asked for their views about the feared symptom. For example, in the case of fear of blushing, other people might be asked:

Why do you think people blush?
Do you notice other people blushing?
Do you remember it?
Do you think badly about people who blush?
If you do, what do you think about them?'

A further helpful experiment can involve intentionally displaying a feared symptom (e.g. handshaking or forgetting what one is talking about) and closely observing other people's responses. A particularly effective behaviour experiment for modifying social phobics distorted self-images involves the use of video feedback. Patients are asked to engage in a difficult social task while being videotaped. Afterwards they are asked to describe in detail how they think they appeared. They are then asked to view the video, watching themselves as though they are watching a stranger, ignoring memories of how they felt and simply focusing on how they would look to other people. In this way they often discover that they come across better than they would expect on the basis of their self-imagery. This experiment is often a powerful way of correcting distorted self-images.

Patients with panic disorder or hypochondriasis persistently think that normal bodily signs and/or symptoms are caused by a serious physical disorder. Numerous behavioural experiments can be used to demonstrate the correct, innocuous causes of their symptoms. For example, reading pairs of words which represent patient's illness interpretations (e.g. palpitations-dying, breathlessness-suffocate) has been shown to induce feared sensations.[5] Similarly, reproducing patients' fear-driven behaviours can produce the very symptoms the patients take as evidence for a serious physical illness. For example, patients who feel short of breath in a panic attack often respond by breathing quickly and deeply (hyperventilation), which paradoxically produces more breathlessness. Similarly, patients who are concerned about cancer may palpate body parts and then take the resulting soreness or discomfort as evidence of the presence of cancer.

(i) Therapy notes

Over a series of sessions therapist and patient will generate a substantial number of arguments against the patient's fearful beliefs. In order to maximize the impact of this accumulation, patients are asked to keep a running record of evidence against their beliefs in a notebook that can easily be consulted at times of doubt. Table 6.3.2.1.7 shows an illustrative example from a panic disorder patient's notebook. At the start of therapy, the patient had been concerned that there was something seriously wrong with his heart.

(j) Anger management

Although anxiety is the predominant problematic emotion in anxiety disorders, some patients also report significant problems with other emotions such as depression and anger. Techniques for dealing with depression can be found in Chapter 6.3.2.3. Some empirically validated techniques for dealing with anger are described here. Although presented in the context of anger accompanying anxiety disorders, these techniques are also relevant to anger in other disorders and to people without an Axis I disorder.

Table 6.3.2.1.7 A panic disorder patient's notebook: evidence for the two alternative explanations for chest pains

'There is something seriously wrong with my heart'	'My problem is my belief that there is something wrong with my heart'
1 *I hear my heart thumping sometimes, even in my ear.* But because of my fears I focus on my body and that makes me notice it. When I notice it I get anxious and that makes it louder because my heart beats are bigger	1 I think I am dying in a panic attack and that thought makes me anxious, producing many more sensations and setting up a vicious circle
2 *I have chest and rib tightness throughout the day.* But cardiac patients don't. They get chest pain (often crushing and more localized) during heart attacks. It is muscle tension due to work stress. It is mild after a good night's sleep and easier at weekends. It is worst after a stressful day at work	2 Distraction sometimes helps. That makes sense if the problem is my thoughts. It does not make sense if the problem is a heart attack. The same argument applies to leaving the situation. That would not stop a heart attack but it makes me feel more comfortable and undermines the negative thoughts
3 *I occasionally get tingling in my fingertips.* But this is a common symptom of anxiety. Also deep breathing—which I do when I think there is something wrong—causes tingling	3 I get symptoms most often at the end of the day, when I have come to expect them and have time to dwell on them
	4 I have proved to myself that there is nothing wrong with my heart with vigorous exercise. All that happens is that my heart beats faster and pumps harder, as it should do in order to supply my muscles with the energy they need

(k) Cognitive content and other assessment issues

Anger is triggered when other people are seen to have broken one's personal rules about what is right and fair.[1] Angry individuals invariably think that they have been badly treated and ascribe their perceived ill treatment to intention or unacceptable neglect on the part of others. A key first step in assessment is to help patients become aware of their automatic thoughts during periods of anger. It is also helpful to keep a record of the situations and behaviours of other people that routinely trigger anger. Review of such triggers often reveals a particular theme and an implicit rule that the patient thinks other people should abide by. A detailed description of how the person behaves when angry and what effect the behaviour has on others is also essential.

(l) Intervention

As patients' rules about the way that others should behave are often highly idiosyncratic, a useful tactic involves asking patients to consider whether the problem is assuming that others hold the same rule as them when they do not. This can help reduce the conviction that others' actions are actively malicious. Other useful questions include the following.

- Is there any other explanation for what happened?
- Did the other people know that their actions would harm me?
- Am I mind-reading?
- Am I over-applying the 'shoulds'?
- What are the advantages and disadvantages of responding with anger?
- Are there other ways I could behave which will be more likely to put things right/help me to get over it?

Although identifying and changing anger-related thoughts is a useful tactic, it is important to remember that anger is an action-orientated emotion. When angry, patients have a strong compulsion to hit out verbally or physically, and have great difficulty in thinking rationally. For patients with recurrent anger problems, it is often useful to teach them first to pause and relax or remove themselves from the anger-provoking situation before trying to challenge their thoughts and to delay-taking action (such as writing angry letters to others) until they have calmed down and had time to consider the appropriateness/usefulness of the action. To enhance further the generalizability of thought-challenging work, it is often useful to summarize the answers to typical anger-related thoughts on a flash card that patients can carry around and consult whenever they become angry.

Anger can sometimes be the result of chronic under-assertiveness, with patients' fears preventing them from making their point of view known until they feel overwhelmed and irritated by the demands placed on them. In such cases, discussion of the fears that prevent earlier and more appropriate assertion and role-playing in which the patients try out and evaluate ways of communicating their views to others in a prompt and constructive fashion can be helpful.

Indications and contraindications

Cognitive behaviour therapy is suitable for most patients with anxiety disorders and the low dropout rates reported in many controlled trials[4,32] suggest that it is well tolerated. In cases with additional severe comorbid problems (e.g. alcohol dependence, depression) it is sometimes necessary to bring these problems under control before starting cognitive behaviour therapy for anxiety. Concurrent use of prescription anxiolytic medication (benzodiazepines, tricyclics, selective serotonin reuptake inhibitors) is not a contraindication. At one time it was thought that anxiolytics may facilitate treatment by helping patients to confront their fears more quickly. However, there is little evidence that concurrent medication enhances initial response.[32] In addition, combining medication (alprazolam or imipramine) with cognitive behaviour therapy has been shown to produce poorer long-term outcome than cognitive behaviour therapy alone in panic disorder.[33] The latter result suggests that if a patient is not already taking anxiolytic medication, it is probably best to start treatment with cognitive behaviour therapy alone. Medication might then be added at a later stage, if response to cognitive behaviour therapy alone is poor.

Efficacy

Controlled trials involving comparisons with other psychological interventions and waiting-list control groups indicate that cognitive

behaviour therapy is an effective and specific treatment for panic disorder, social phobia, specific phobia, generalized anxiety disorder, hypochondriasis, obsessive–compulsive disorder, and post-traumatic stress disorder.[4,32] Results comparing immediate response to cognitive behaviour therapy alone and pharmacotherapy alone have been mixed, with superiority for cognitive behaviour therapy, equivalence for cognitive behaviour therapy, and superiority for pharmacotherapy all being reported. In contrast to the immediate response data, the follow-up analyses after medication discontinuation that are currently available favour cognitive behaviour therapy.[34] However, the database is modest and further research is required. For anger problems, controlled trials have shown that the cognitive behavioural procedures described here are effective.[35]

Training and supervision

Most controlled trials have used therapists who have received specialized training in cognitive behaviour therapy and there is some evidence that deviation from therapy protocols and/or poor implementation is associated with less good outcome.[36] For these reasons, clinicians are likely to benefit from specialized training and supervision. Where local training institutes exist, it is wise to take advantage of their expertise. Even when no local institute is available, expert cognitive behaviour therapists from established centres often travel internationally to deliver workshops and supervision. Several professional organizations run regular training workshops and can be contacted through the Internet. The organizations include the British Association of Behavioural and Cognitive Psychotherapies (http://www.babcp.org.uk), the Association of Behaviour and Cognitive Therapies (www.abct.org), the International Association of Cognitive Psychotherapy (http://www.cognitivetherapyassociation.org), the American Psychological Association (http://www.apa.org), and the American Psychiatric Association (www.psych.org). A comprehensive list of the competencies required for the main cognitive behaviour therapies for anxiety disorders can be found at: http://www.ucl.ac.uk/clinical-health-psychology/CORE/CBT_Framework.htm

Further information

A number of texts describe the theory and practice of cognitive behaviour therapy for specific anxiety disorders[37–39] and for anger problems[40] in considerable detail. Texts are frequently updated. Readers interested in the latest therapy guides are recommended to visit the following websites: www.oup.com, www.oup.com/us/ttw, www.guilford.com, www.wiley.com. Video illustrations of therapy sessions are also available for some anxiety disorders (see ABCT, American Psychological Association, and Guilford Press websites).

References

1. Beck, A.T. (1976). *Cognitive therapy and the emotional disorders*. International Universities Press, New York.
2. Rachman, S. (1996). The evolution of cognitive behaviour therapy. In *Science and practice of cognitive behaviour therapy* (eds. D.M. Clark and C.G. Fairburn), pp. 3–26. Oxford University Press, Oxford.
3. Marks, I.M. (1975). Behavioural treatment of phobic and obsessive-compulsive disorders. In *Progress in behavior modification* (eds. M. Hersen, R.M. Fisher, and P.M. Miller), **1**, pp. 66–158. Academic Press, New York.
4. Clark, D.M. (2004). Developing new treatments: on the interplay between theories, experimental science and clinical innovation. *Behaviour Research and Therapy*, **42**, 1089.
5. Clark, D.M. (1996). Panic disorder: from theory to therapy. In *Frontiers of cognitive therapy* (ed. P.M. Salkovskis), pp. 318–44. Guilford, New York.
6. APA. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington, D.C.
7. Beck, A.T., Emery, G., and Greenberg, R.L. (1985). *Anxiety disorders and phobias: a cognitive perspective*. Basic Books, New York.
8. Borkovec, T.D., Ray, W.J., and Stober, J.W. (1998). A cognitive phenomenon intimately linked to affective, physiological, and interpersonal behavioral processes. *Cognitive Therapy and Research*, **22**, 561–76.
9. Wells, A. (2000). *Emotional disorders and metacognition: innovative cognitive therapy*. Wiley, Chichester.
10. Rachman, S.J. and De Silva, P. (1978). Abnormal and normal obsessions. *Behaviour Research and Therapy*, **16**, 233–48.
11. Salkovskis, P.M. (1985). Obsessional–compulsive problems: a cognitive-behavioural analysis. *Behaviour Research and Therapy*, **23**, 571–83.
12. Kessler, R.C., Sonnega, A., Bromet, E., *et al.* (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry*, **52**, 1048–60.
13. Ehlers, A. and Clark, D.M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy*, **38**, 319–45.
14. Brewin, C.R. and Holmes, E.A. (2003). Psychological theories of posttraumatic stress disorder. *Clinical Psychology Review*, **23**, 339–76.
15. Foa, E.B., Molnar, C., and Cashman, L. (1995). Change in rape narratives during exposure therapy for posttraumatic stress disorder. *Journal of Traumatic Stress*, **8**, 675–90.
16. Salkovskis, P.M. (1996). The cognitive approach to anxiety: threat beliefs, safety-seeking behaviour, and the special case of health anxiety and obsessions. In *Frontiers of cognitive therapy* (ed. P.M. Salkovskis), pp. 48–74. Guilford Press, New York.
17. Clark, D.M. and Wells, A. (1995). A cognitive model of social phobia. In *Social phobia: diagnosis, assessment and treatment* (eds. R. Heimberg, M. Liebowitz, D.A. Hope, and F.R. Schneier), pp. 69–93. Guilford Press, New York.
18. WHO. (1992). *International statistical classification of diseases and related health problems* (10th edn). WHO, Geneva
19. Veljaca, K.A. and Rapee, R.M. (1998). Detection of negative and positive audience behaviours by socially anxious subjects. *Behaviour Research and Therapy*, **36**, 311–21.
20. Hackmann, A., Surawy, C., and Clark, D.M. (1998). Seeing yourself through others' eyes: a study of spontaneously occurring images in social phobia. *Behavioural and Cognitive Psychotherapy*, **26**, 3–12.
21. Arntz, A., Rauner, M., and Van den Hout, M. (1995). If I feel anxious, there must be danger: ex-consequential reasoning in inferring danger in anxiety disorders. *Behaviour Research and Therapy*, **33**, 917–25.
22. Mansell, W. and Clark, D.M. (1999). How do I appear to others? Social anxiety and processing of the observable self. *Behaviour Research and Therapy*, **37**, 419–34.
23. Davey, G.C.L. and Matchett, G. (1994). Unconditioned stimulus rehearsal and the retention and enhancement of differential 'fear' conditioning: effects of trait and state anxiety. *Journal of Abnormal Psychology*, **103**, 708–18.
24. Beck, A.T. and Steer, R.A. (1993). *Beck anxiety inventory manual*. Psychological Corporation, San Antonio, TX.
25. Beck, A.T. and Steer, R.A. (1993). *Beck depression inventory*. The Psychological Corporation, San Antonio, TX.
26. Beunderman, R., Van Dis, H., Koster, R.W., *et al.* (1988). Differentiation in prodromal and acute symptoms of patients with cardiac and

non-cardiac chest pain. In *Advances in theory and practice in behaviour therapy* (eds. P.M.G. Emmelkamp, W.T.A.M. Everaerd, F. Kraaimaat, and M.J.M. van Son). Swets & Zeitlinger, Amsterdam.
27. Wells, A. and Hackmann, A. (1993). Imagery and core beliefs in health anxiety: content and origins. *Behavioural and Cognitive Psychotherapy*, **21**, 265–73.
28. Rachman, S.J., De Silva, P., and Roger, G. (1976). The spontaneous decay of compulsive urges. *Behaviour Research and Therapy*, **14**, 445–53.
29. Ehlers, A., Mayou, R.A., and Bryant, B. (1998). Psychological predictors of chronic posttraumatic stress disorder after motor vehicle accidents. *Journal of Abnormal Psychology*, **107**, 508–19.
30. NICE. (2005). *Post-traumatic stress disorder (PTSD): the management of PTSD in adults and children in primary and secondary care (Clinical Guideline 26)*. National Institute for Clinical Excellence, London, UK (www.nice.og).
31. Foa, E.B. and Riggs, D.S. (1993). Post-traumatic stress disorder in rape victims. In *Annual review of psychiatry* (eds. J.M. Oldham, M.B. Riba, and A. Tasman), pp. 273–303. American Psychiatric Association, Washington, DC.
32. Nathan, P.E. and Gorman, J.S. (eds.) (2002). *A guide to treatments that work*. Oxford University Press, New York.
33. Barlow, D.H., Gorman, J.M., Shear, M.K., *et al.* (2000). Cognitive-behavioral therapy, imipramine, or their combination for panic disorder. A randomized controlled trial. *Journal of the American Medical Association*, **283**, 2529–36.
34. Clark, D.M. and Wells, A. (1997). Cognitive therapy for anxiety disorders. In *Review of psychiatry* (eds. L.J. Dickstein, M.B. Riba, and J.M. Oldham), pp. 9–44. American Psychiatric Press, Washington, DC.
35. Del Vecchio, L. and O'Leary, K.D. (2004). Effectiveness of anger treatments for specific anger problems: a meta-analytic review. *Clinical Psychology Review*, **24**, 15–34.
36. Schulte, D., Kunzel, R., Pepping, G., *et al.* (1992). Tailor-made versus standardized therapy of phobic patients. *Advances in Behaviour Research and Therapy*, **14**, 67–92.
37. Hawton, K.E., Salkovskis, P.M., Kirk, J., *et al.* (1989). *Cognitive behaviour therapy for psychiatric problems*. Oxford University Press, Oxford.
38. Wells, A. (1997). *Cognitive therapy of anxiety disorders: a practice manual and conceptual guide*. Wiley, Chichester, UK.
39. Barlow, D.H. (ed.). (2007). *Clinical handbook of psychological disorders* (4th edn). Guilford Press, New York.
40. Howells, K. (1998). Cognitive behavioural interventions for anger, aggression and violence. In *Treating complex cases* (eds. N. Tarrier, A. Wells, and G. Haddock), pp. 295–318. Wiley, Chichester.
41. Spielberger, C.D., Gorsuch, R.L., and Lushene, R.E. (1970). *Manual for the state—Trait anxiety inventory*. Consulting Psychologists Press, Palo Alton, CA.
42. Clark, D.M., Salkovskis, P.M., Hackmann, A., *et al.* (1994). A comparison of cognitive therapy, applied relaxation and imipramine in the treatment of panic disorder. *British Journal of Psychiatry*, **164**, 759–69.
43. Marks, I. and Mathews, A.M. (1979). Brief standard self-rating for phobic patients. *Behaviour Research and Therapy*, **17**, 263–7.
44. Chambless, D.L., Caputo, G.C., Bright, P., *et al.* (1984). Assessment for fear of fear in agoraphobics: the Body Sensations Questionnaire and the Agoraphobia Cognitions Questionnaire. *Journal of Consulting and Clinical Psychology*, **52**, 1090–7.
45. Chambless, D.L., Caputo, G.C., Jasin, S.E., *et al.* (1985). The mobility inventory for agoraphobia. *Behaviour Research and Therapy*, **23**, 35–44.
46. Fresco, D.M., Coles, M.E., Heimberg, R.G., *et al.* (2001). The Liebowitz social anxiety scale: a comparison of the psychometric properties of self-report and clinical-administered formats. *Psychological Medicine*, **31**, 1025–35.
47. Mattick, R.P. and Clarke, J.C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, **36**, 455–70.
48. Turner, S.M., Beidel, D.C., Dancu, C.V., *et al.* (1989). An empirically derived inventory to measure social fears and anxiety: the social phobia and anxiety inventory. *Psychological Assessment*, **1**, 35–40.
49. Tallis, F., Davey, G.C.L., and Bond, A. (1994). The Worry Domains Questionnaire. In *Worrying: perspectives on theory, assessment and treatment* (eds. G. Davey and F. Tallis). Wiley, Chichester.
50. Molina, S. and Borkovec, T.D. (1994). The Penn State Worry questionnaire: psychometric properties and associated characteristics. In *Worrying: perspectives on theory, assessment and treatment* (eds. G. Davey and F. Tallis). Wiley, Chichester.
51. Wells, A. and Davies, M. (1994). The Thought Control Questionnaire: a measure of individual differences in the control of unwanted thoughts. *Behaviour Research and Therapy*, **32**, 871–8.
52. Salkovskis, P.M., Wroe, A.L., Gledhill, A., *et al.* (1999). Responsibility attitudes and interpretations are characteristic of obsessive compulsive disorder. *Behaviour Research and Therapy*, **38**, 347–72.
53. Van Oppen, P., Hoekstra, R.J., and Emmelkamp, P.M.G. (1995). The structure of obsessive-compulsive symptoms. *Behaviour Research and Therapy*, **33**, 15–23.
54. Goodman, W.K., Price, L.H., Rasmussen, S.A., *et al.* (1989). The Yale-Brown Obsessive-Compulsive Scale (YBOCS) Part 1: development, use and reliability. *Archives of General Psychiatry*, **46**, 1006–11.
55. Foa, E., Riggs, D.S., Dancu, C.V., *et al.* (1993). Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress*, **6**, 459–73.
56. Foa, E., Ehlers, A., Clark, D.M., *et al.* (1999). The post-traumatic cognitions inventory (PTCI): development and validation. *Psychological Assessment*, **11**, 303–14.
57. Horowitz, M.J., Wilner, N., and Alvarez, W. (1979). The Impact of Event Scale: a measure of subjective stress. *Psychosomatic Medicine*, **41**, 209–18.
58. Mechanic, M.B. and Resick, P.A. (1993). The personal beliefs and reactions scale: assessing rape-related cognitive schemata. *Paper presented at the 9th annual meeting of the International Society for Traumatic Stress Studies*, San Antonio, TX.
59. Clark, D.M., Ehlers, A., Hackmann, A., *et al.* (2006). Cognitive therapy and exposure plus applied relaxation in social phobia: a randomised controlled trial. *Journal of Consulting and Clinical Psychology*, **74**, 568–78.

6.3.2.2 Cognitive behaviour therapy for eating disorders

Zafra Cooper, Rebecca Murphy, and Christopher G. Fairburn

Introduction

The eating disorders provide one of the strongest indications for cognitive behaviour therapy. This bold claim arises from the demonstrated effectiveness of cognitive behaviour therapy in the treatment of bulimia nervosa and the widespread acceptance that cognitive behaviour therapy is the treatment of choice.[1] Cognitive behaviour therapy is also widely used to treat anorexia nervosa although this application has not been adequately evaluated. Recently its use has been extended to 'eating disorder not

otherwise specified' (eating disorder NOS),[2] a diagnosis that applies to over 50 per cent of cases,[3] and emerging evidence suggests that it is just as effective with these cases as it is with cases of bulimia nervosa.

In this chapter the cognitive behavioural approach to the understanding and treatment of eating disorders will be described. The data on the efficacy and effectiveness of the treatment are considered in the chapters on anorexia nervosa and bulimia nervosa (see Chapters 4.10.1 and 4.10.2 respectively), as is their general management.

The cognitive behavioural account of the maintenance of eating disorders

Although both the DSM and ICD schemes for classifying eating disorders encourage the view that anorexia nervosa and bulimia nervosa are distinct clinical states, consideration of their clinical features and course over time does not support this.[4] Patients with anorexia nervosa, bulimia nervosa, and eating disorder NOS have many features in common, most of which are not seen in other psychiatric disorders, and studies of their course indicate that most patients migrate between these diagnoses over time. This temporal movement, together with the fact that the disorders share the same distinctive psychopathology, has led to the suggestion that common 'transdiagnostic' mechanisms are involved in the persistence of eating disorder psychopathology.[5]

Anorexia nervosa, bulimia nervosa, and most cases of eating disorder NOS are united by a distinctive core psychopathology: patients over-evaluate the importance of their shape and weight and their ability to control them. According to the cognitive behavioural view it is this dysfunctional scheme of self-evaluation that is of central importance in maintaining these disorders. Whereas most people evaluate themselves on the basis of their perceived performance in a variety of domains of life, people with eating disorders judge themselves primarily in terms of their shape and weight and their ability to control them. Most of their other clinical features can be understood as stemming directly from this 'core psychopathology', including the extreme weight-control behaviour (i.e. the dieting, self-induced vomiting, laxative misuse, and over-exercising), the various forms of body checking and avoidance, and the preoccupation with thoughts about eating, shape, and weight. Fig. 6.3.2.2.1 provides a 'transdiagnostic' representation (or 'formulation') of the main processes involved in the maintenance of eating disorders.

The only feature that is not obviously a direct expression of the core psychopathology is binge eating, present in all cases of bulimia nervosa, many cases of eating disorder NOS and some cases of anorexia nervosa. The cognitive behavioural theory proposes that binge eating is largely a product of attempts to adhere to multiple extreme, and highly specific, dietary rules. These patients' tendency to react in a negative and extreme fashion to the (almost inevitable) breaking of these rules results in even minor dietary slips being interpreted as evidence of poor self-control. Patients respond to this perceived lack of self-control by temporarily abandoning their efforts to restrict their eating. This produces a highly distinctive pattern of eating in which attempts to restrict eating are repeatedly interrupted by episodes of binge eating. The binge eating maintains the core psychopathology by intensifying patients' concerns about their ability to control their eating, shape, and weight. It also encourages further dietary restraint, thereby increasing the risk of further binge eating.

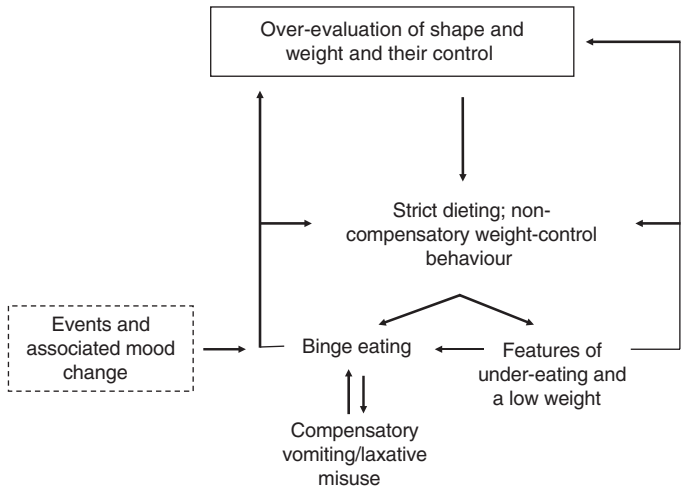


Fig. 6.3.2.2.1 The transdiagnostic 'template' formulation of the maintenance of eating disorders. (Reproduced from Fairburn, C.G., *Cognitive behavior therapy and eating disorders*, Copyright 2008, Guilford Press, NY).

Three further processes also maintain binge eating. First, difficulties in the patient's life and associated mood changes increase the likelihood that they will break their dietary rules. Second, binge eating temporarily ameliorates such mood states and distracts patients from thinking about their difficulties. Third, if the binge eating is followed by compensatory vomiting or laxative misuse, it is also maintained because patients' mistaken belief in the effectiveness of such 'purging' undermines a major deterrent to their binge eating. They do not realize that purging has little effect on energy absorption.[6]

Patients with anorexia nervosa share the distinctive core psychopathology of those with bulimia nervosa and eating disorder NOS. The major difference between patients with anorexia nervosa and those with other eating disorders lies in the fact that in anorexia nervosa under-eating predominates and therefore patients become extremely underweight. This has certain physiological and psychological consequences (see Chapter 4.10.1) that contribute to the persistence of the eating disorder.[7] For example, delayed gastric emptying results in a sense of fullness even after eating modest amounts of food and secondary social withdrawal magnifies patients' isolation from the influence of others.

The composite 'transdiagnostic' formulation shown in Fig. 6.3.2.2.1 represents the core processes that maintain any eating disorder. The specific maintaining processes operating in any patient depend upon the nature of the eating disorder psychopathology present. In some cases only certain of the processes are active (for example, in most cases of binge-eating disorder), but in others (for example, cases of the binge eating/purging subtype of anorexia nervosa) most are operating. The formulation highlights the maintaining processes that need to be addressed in treatment, thereby allowing the clinician to design a bespoke treatment to fit the individual patient's psychopathology.

Evidence for the cognitive behavioural account

There is a sizeable body of research that supports the cognitive view of the maintenance of eating disorders.[8] This includes descriptive and experimental studies of the clinical characteristics of these

patients and the research on dietary restraint and 'counter-regulation' (a possible analogue for binge eating).[9] There is also strong indirect support from a large body of research indicating that cognitive behaviour therapy based on this model has a major and lasting impact on bulimia nervosa (see Chapter 4.10.2) and emerging evidence that this is also true of eating disorder NOS. Further support comes from the finding that 'dismantling' the cognitive behavioural treatment for bulimia nervosa by removing those procedures designed to produce cognitive change attenuates its effects and results in patients being markedly prone to relapse.[10] Direct support comes from studies that have shown that dietary restraint appears to mediate this treatment's effect on binge eating[11] and that continuing over-evaluation of weight and shape in those who have recovered in behavioural terms is predictive of subsequent relapse.[12,13]

Transdiagnostic cognitive behavioural treatment of eating disorders

There follows a brief description of cognitive behaviour therapy for eating disorders. Full details are provided elsewhere.[14] The treatment is outpatient-based and, as applied in research settings, involves 20 individual treatment sessions over 20 weeks for the 80 per cent or more of patients who are not significantly underweight (body mass index over 17.5). The remaining patients with a body mass index of 17.5 or below receive 40 sessions over 40 weeks. The indications for the treatment are the presence of an eating disorder of clinical severity. It is not appropriate for those whose psychiatric state, general physical health, or degree of weight loss is such that they cannot safely be treated on an outpatient basis.

The treatment has four stages.

(a) Stage one

The aims of the first stage are as follows: to educate patients about treatment and the disorder; to engage the patient in treatment and change, and to introduce and establish a pattern of regular eating and weekly weighing. This stage comprises approximately eight sessions which are held twice weekly over 4 weeks.

(i) Jointly creating the formulation

This is usually done in the first treatment session. The therapist draws out the relevant sections of Fig. 6.3.2.2.1 incorporating the patient's own experiences and terms. This helps patients to realize both that their behaviour is comprehensible and that it is maintained by a variety of self-perpetuating mechanisms which are open to change. The formulation provides a guide to what needs to be targeted in treatment if patients are to achieve a full and lasting recovery.

(ii) Establishing real-time self-monitoring

This is the ongoing 'in-the-moment' recording of eating and other relevant behaviour, thoughts, feelings, and events (see Chapter 4.10.2 for an example monitoring record). Self-monitoring is initiated in the first session and continues throughout treatment. It serves two purposes: it assists in the identification of the patient's problems and progress and, more importantly, it facilitates change by helping patients address problems as they occur.

(iii) Establishing 'weekly weighing'

The patient and therapist check the patient's weight once a week and plot it on an individualized weight graph. Patients are strongly encouraged not to weigh themselves at other times. Weekly in-session weighing has several purposes: first, it provides patients with accurate data about their weight at a time when their eating habits are changing; second, it provides an opportunity for the therapist to help patients interpret the numbers on the scale, which otherwise they are prone to misinterpret and, third, it addresses the important maintaining processes of excessive body weight checking or its avoidance.

(iv) Providing education

From the second session onwards, an important element of treatment is education about weight and eating since many patients have misconceptions that maintain their eating disorder. The following topics need to be covered:

- Body weight and its regulation: the body mass index and its interpretation; natural weight fluctuations; and the effects of treatment on weight.
- Physical complications of binge eating, self-induced vomiting, the misuse of laxatives and diuretics, and the effect of the eating disorder on hunger and fullness.
- Ineffectiveness of vomiting, laxatives, and diuretics as a means of weight control.
- Adverse effects of dieting: the types of dieting that promote binge eating; dietary rules versus dietary guidelines.

To provide reliable information on these topics, patients are asked to read relevant sections from one of the authoritative books on eating disorders[6,15,16] and their reading is discussed in subsequent treatment sessions.

(v) Establishing 'regular eating'

The establishment of a pattern of regular eating is fundamental to successful treatment whatever the form of the eating disorder. It addresses an important type of dieting ('delayed eating'); it displaces episodes of binge eating and, for underweight patients, it introduces regular meals and snacks that can be subsequently increased in size. Early in treatment (usually by the third session) patients are asked to eat three planned meals each day, plus two (or if underweight three) planned snacks and they are asked not to eat between them. Patients may choose what they eat at these times with the only conditions being that the meals and snacks are not followed by any compensatory behaviour and that there should rarely be more than a 4-hour interval between these occasions of eating. The new eating pattern should take precedence over other activities but should not be so inflexible as to preclude the possibility of adjusting timings to suit the patients' commitments each day.

Patients should be helped to adhere to their regular eating plan and to resist eating between the planned meals and snacks. Two rather different strategies may be used to achieve this: the first involves helping patients to identify activities that are incompatible with eating or make it less likely, and the second is to help patients to recognize that the urge to eat is a temporary phenomenon. Through using these strategies patients learn to distance themselves from the urge to eat which they find gradually fades with time.

(vi) Involving significant others

The treatment is primarily an individual treatment for adults and hence it does not actively involve others. Despite this, it is our

practice to see 'significant others' with the patient if this is likely to facilitate treatment and the patient is willing for this to happen. There are two specific indications for involving others: if others could help the patient in making changes or if others are making it difficult for the patient to change by, for example, commenting adversely on eating or appearance.

(b) Stage two

Stage two is a transitional stage which generally comprises two appointments, a week apart. Whilst continuing with the procedures introduced in stage one the therapist and patient conduct a joint review of progress to date, identify problems still to be addressed, revise the formulation if necessary, and design stage three.

(c) Stage three

The aim of this stage is to address the key mechanisms that are maintaining the patient's eating disorder. The order in which these mechanisms are addressed depends upon their relative importance in maintaining the particular patient's psychopathology. There are generally 8-weekly appointments.

(i) Addressing the over-evaluation of shape and weight

The first step involves explaining the concept of self-evaluation and helping patients identify the life domains which contribute to their judgement of themselves. The relative importance of these domains can be visually represented on a pie chart, which for most patients is dominated by a large slice representing shape and weight and controlling eating.

The patient and therapist then identify the problems inherent in this scheme for self-evaluation. Briefly there are three related problems: first, the over-evaluation of shape and weight tends to marginalize other domains and thus self-evaluation is overly dependent on performance in one area of life; second, the area of controlling shape and weight is one in which success is elusive, thus undermining self-esteem; and third, the over-evaluation leads to behaviour which is unhelpful and which itself maintains the disorder.

The final step in educating about self-evaluation involves identifying its three main expressions which occur to varying degrees in different patients. These are body checking, body avoidance, and feeling fat. The therapist explains how these behaviours and experiences serve to maintain and magnify the patient's concerns about shape and weight and it is agreed therefore that they need to be addressed in treatment.

(ii) Addressing body checking and avoidance

Patients are often not aware that they are engaging in body checking and that it is maintaining their body dissatisfaction. The first step in addressing body checking involves obtaining a detailed account of the behaviour by asking patients to record it. An example monitoring record is shown in Fig. 6.3.2.2.2. Patients are then helped to realize that body checking is not a helpful way of assessing their

Day ……Friday…….. Date …..6th Sept ……..

Time	Food and drink consumed	Place	*	V/L	Checking (what done, time taken)	Place	Context and comments
7.45	1 piece of toast with butter and marmite. 1 cup of tea	Kitchen					Not hungry but know I should have breakfast
8.15					Looking at stomach and thighs in bedroom mirror while getting dressed (2 mins)	Bedroom	Depressing. Can't see any muscles – only fat.
10.30	1 apple & diet coke	Office					
11.15					Scrutinisng stomach in mirror standing sideways (1 min)	Office toilet	Ok
			*				How can it be so fat?? Have hardly eaten anything today!
12.30	Cheese and tomato sandwich & banana and kit kat	Canteen					Chocolate was too much. Feel too full.
1.30					Feeling and pinching stomach while sitting at desk (10mins)	Office	myself sick.
							Fat disgusting flesh. Feel massive.
3.10	Cup of coffee and 1 yoghurt	Office					Had planned to have other 1/ 2 of chocolate bar. Can't do it, am already too fat!
3.30	Water				Assessing shape of woman on street (10secs)	Office – through window	Feel unhappy. Wish I was as thin as her. Can't accurately compare myself to others – never really sure what my shape is. Frustrating.
6.45	Salad with tuna 1 glass of red wine	Living room					
7.30					Looking at reflection in window while doing dishes (5mins)	Kitchen	Still want more food but won't let myself since might lose control.
						Bathroom	
8.10			*	V	Touching and pinching stomach and thighs while having a bath (15mins)		I'm so huge! Wish I was tall and elegant. How depressing.
9.30	1 kit kat and half a crunchie 1 glass of red wine	Living room					Feel disgusted. Am I ever going to get rid of this fat?
							Feel fat! Too much chocolate today. Have to get rid of the food. Go to bed to stop thinking about it.

Fig. 6.3.2.2.2 An adapted monitoring record illustrating body checking.

shape or weight as it provides unreliable and biased information. Certain forms of body checking are best stopped altogether. In the case of more normative checking such as mirror use, education should stress that, as with other forms of body checking, what one finds depends to an important extent upon how one looks (e.g. scrutiny of perceived flaws tends to magnify them). For patients who avoid seeing their bodies, the therapist needs to explain that this too maintains dissatisfaction. Patients need to be encouraged to get used to the sight and feel of their body. Participation in activities that involve a degree of body exposure can be helpful, for example, swimming.

(iii) Addressing 'feeling fat'

'Feeling fat' is an experience reported by many women but the intensity and frequency of this feeling appears to be far greater among people with eating disorders. Feeling fat is a target for treatment since it tends to be equated with being fat (irrespective of actual shape and weight) and hence maintains body dissatisfaction. Although this topic has received little research attention, clinical observation suggests that in many patients feeling fat is a result of mislabelling certain emotions and bodily experiences. It may be addressed by helping patients appreciate that feeling fat tends to be triggered by the occurrence of certain negative mood states (e.g. feeling bored or depressed) or by physical sensations that heighten body awareness (e.g. feeling full, bloated, or sweaty). Patients can then be encouraged to question the feeling when it occurs and correctly label and address the underlying triggering state using a problem-solving approach.

(iv) Developing marginalized domains for self-evaluation

Tackling the expressions of the over-evaluation of shape and weight will gradually reduce it. At the same time, it is also important to encourage the patient to increase the number and significance of other domains for self-evaluation. Although this is an indirect means of diminishing the over-evaluation of shape and weight, it is nevertheless a powerful one.

(v) Exploring the origins of the over-evaluation

Towards the end of stage three it is often helpful to explore the origins of the patient's sensitivity to shape, weight, and eating. An historical review can help to make sense of how the problem developed and evolved, highlight how it might have served a useful function in its early stages and help patients distance themselves from the past. If a specific event appears to have played a critical role in the development of the eating problem, the patient should be helped to reappraise this from the vantage point of the present.

(vi) Addressing dietary restraint

A major goal of treatment is to reduce, if not eliminate altogether, strict dieting. This dieting has two aspects: an attempt to limit eating termed 'dietary restraint' and actual under-eating in physiological terms termed 'dietary restriction'. 'Regular eating' will already have addressed one form of dietary restraint (delayed eating). Patients need to recognize that their multiple extreme and rigid dietary rules lead to preoccupation with food and eating, encourage binge eating and impose practical and social restrictions. It should therefore be agreed that dietary restraint needs to be addressed. To do this, the patient's various dietary rules should be identified together with the beliefs which underlie them. The patient should be helped to break the rules in order to test the beliefs in question and to learn that the feared consequences that maintain the dietary rule (typically sudden weight gain or binge eating) are not an inevitable result of breaking it. With patients who binge eat it is important to pay particular attention to food avoidance and to help them systematically reintroduce such foods into their diet.

(vii) Addressing event-triggered changes in eating

Among patients with eating disorders, eating habits may change in response to outside events. The change may involve eating less, stop eating altogether, overeating or binge eating. If these changes persist into stage three, they should be addressed by helping patients to tackle the triggering events using a problem-solving approach and by helping patients to accept the occurrence of intense-mood states and identify ways (that are not harmful) of modulating their moods.

(d) Stage four

The aims in stage four are to ensure that the changes made in treatment are maintained over the following months and that the risk of relapse is minimized in the long term. There are three appointments, each 2 weeks apart. During this stage, as part of their preparation for the future, patients discontinue self-monitoring and transfer from in-session weighing to weighing themselves at home.

To maximize the chances that progress is maintained the therapist and patient jointly devise a specific plan for the patient to follow over the following few months until a post-treatment review appointment. Typically this includes further work on body checking, food avoidance, and perhaps further practice at problem-solving. In addition, the therapist encourages patients to continue their efforts to develop new interests and activities.

There are two elements to 'relapse prevention'. First, patients must have realistic expectations regarding the future. A common problem is that many hope never to experience any eating difficulties again. It needs to be explained that this makes them vulnerable to relapse since it encourages a negative reaction to even minor setbacks. Patients should be told to expect lapses with the eating problem continuing to be their Achilles' heel. The goal is for patients to identify setbacks as early as possible, view them as a 'lapse' rather than a 'relapse', and use a well-developed plan to deal with them. Thus, the second element of relapse prevention is the construction of such a plan. The therapist and patient should review the components of treatment with the aim of identifying the principles and procedures that were most relevant and helpful and devise a plan for the future incorporating this information.

Underweight patients

When treating patients who are underweight (most are cases of anorexia nervosa but some are cases of eating disorder NOS) three main modifications to the treatment are required: the motivation of these patients needs to be enhanced, their state of starvation needs to be corrected, and significant others are more likely to be involved. As a result treatment needs to be considerably longer.

(a) Enhancing motivation to change

The poor motivation of these patients needs to be addressed from the outset of treatment. There are various ways of enhancing motivation[17] including focusing on establishing a sound therapeutic relationship, ensuring that the patient feels understood, making it clear that one is working on behalf of the patient and not their

relatives or concerned others, accepting the patient's beliefs and values as genuine and comprehensible, and adopting an experimental approach in which the therapist and patient together explore the advantages and disadvantages of making changes. This includes educating the patient about the physiological and psychological effects of starvation; for example, impaired concentration, preoccupation with food and eating, sleep disturbance, sensitivity to cold, ritualistic eating, social withdrawal, and enhanced fullness secondary to delayed gastric emptying.[7] It is best to focus particularly on those features that the patient views as a problem and explain how they tend to perpetuate the eating disorder. In addition, an exploration of the broader impact of the eating problem on the patient's life is important. When exploring the advantages and disadvantages of change, it is important to draw a distinction between the short-term and long-term consequences of change since patients tend to focus on the immediate present rather than the future.

(b) Restoring a healthy weight

Unless the weight loss is rapid or extreme, or the patient's health is endangered by physical complications, weight restoration can usually be accomplished on an outpatient basis. Before focusing on weight gain, however, it is best to devote several sessions to establishing a collaborative working relationship and to developing a joint formulation and treatment plan. Thereafter, weight gain and the subsequent maintenance of a healthy weight must be an integral part of treatment. A target weight range should be identified in excess of a body mass index of 19.0.

The weight gain should be gradual and steady (at an average rate of about 0.5 kg/week). This requires an energy surplus to be established. This can be achieved by providing patients with energy-rich drinks to supplement their food intake (which should be increased such that it is sufficient to maintain their current weight). The energy-rich drinks may be viewed as weight restorative 'medicine' designed to produce the energy surplus. The drinks should be phased out once the target weight range has been reached. Whilst regaining weight patients should be helped to address their shape and weight concerns and dieting in much the same way as described above. Once a satisfactory weight has been reached patients need time to learn how to maintain their weight in a normal manner.

(c) Involving the significant others

Weight regain is a protracted process that takes considerable time and effort on the part of the patient. If the patient lives with others it can be helpful to involve them if doing so is consistent with the nature of their relationship. Significant others can help the patient choose what to eat and provide guidance concerning portion sizes. This is likely to be especially important with younger patients (many underweight patients are adolescents) who are still living at home with their parents.

Acknowledgements

We are grateful to the Wellcome Trust for its support. CGF holds a Principal Research Fellowship (046386). ZC and RM are supported by a programme grant (046386).

Further information

Fairburn, C.G. (2008). *Cognitive behavior therapy and eating disorders*. Guilford Press, New York.

The International Journal of Eating Disorders.

References

1. National Institute for Clinical Excellence. (2004). *Eating disorders—core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders*. NICE Clinical Guideline No. 9. NICE, London. www.nice.org.uk
2. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders*. American Psychiatric Association, Washington, DC.
3. Fairburn, C.G. and Bohn, K. (2005). Eating disorder NOS (EDNOS): an example of the troublesome 'not otherwise specified' (NOS) category in DSM-IV. *Behaviour Research and Therapy*, **43**, 691–701.
4. Fairburn, C.G. and Harrison, P.J. (2003). Eating disorders. *Lancet*, **361**, 407–16.
5. Fairburn, C.G., Cooper, Z., and Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: a 'transdiagnostic' theory and treatment. *Behaviour Research and Therapy*, **41**, 509–28.
6. Fairburn, C.G. (1995). *Overcoming binge eating*, pp. 48–54. Guilford Press, New York.
7. Garner, D.M. (1997). Psychoeducational principles in treatment. In *Handbook of treatment for eating disorders* (eds. D.M. Garner and P.E. Garfinkel), pp. 145–77. Guilford Press, New York.
8. Vitousek, K.M. (1996). The current status of cognitive–behavioral models of anorexia nervosa and bulimia nervosa. In *Frontiers of cognitive therapy* (ed. P. Salkovskis), pp. 383–418. Guilford Press, New York.
9. Polivy, J. and Herman, C.P. (1993). Etiology of binge eating: psychological mechanisms. In *Binge eating: nature, assessment and treatment* (eds. C.G. Fairburn and G.T. Wilson), pp. 173–205. Guilford Press, New York.
10. Fairburn, C.G., Jones, R., Peveler, R.C., *et al.* (1993). Psychotherapy and bulimia nervosa: the longer–term effects of interpersonal psychotherapy, behaviour therapy and cognitive behaviour therapy. *Archives of General Psychiatry*, **50**, 419–28.
11. Wilson, G.T., Fairburn, C.G., Agras, W.S., *et al.* (2002). Cognitive-behavioral therapy for bulimia nervosa: time course and mechanisms of change. *Journal of Consulting and Clinical Psychology*, **70**, 267–74.
12. Fairburn, C.G., Peveler, R.C., Jones, R., *et al.* (1993). Predictors of 12–month outcome in bulimia nervosa and the influence of attitudes to shape and weight. *Journal of Consulting and Clinical Psychology*, **61**, 696–8.
13. Fairburn, C.G., Stice, E., Cooper, Z., *et al.* (2003). Understanding persistence in bulimia nervosa: a 5-year naturalistic study. *Journal of Consulting and Clinical Psychology*, **71**, 103–9.
14. Fairburn, C.G. (2008). *Cognitive behavior therapy and eating disorders*. Guilford Press, New York.
15. Cooper, P.J. (1995). *Bulimia nervosa and binge eating: a guide to recovery*. Robinson, London.
16. Schmidt, U. and Treasure, J. (1993). *Getting better bit(e) by bit(e)*. Erlbaum, Hove.
17. Vitousek, K., Watson, S., and Wilson, G.T. (1998). Enhancing motivation for change in treatment–resistant eating disorders. *Clinical Psychology Review*, **18**, 391–420.

6.3.2.3 Cognitive behaviour therapy for depressive disorders

Melanie J. V. Fennell

Introduction

This chapter describes A.T. Beck's cognitive behaviour therapy (CBT) for depression.[1] Beck's is probably the most fully developed, comprehensively evaluated, and widely disseminated cognitive behavioural approach to depression. Additionally, CBT is an effective treatment for a range of acute psychiatric disorders, shows promise for severe mental illness and personality disorder, and is thus helpful not only with primary depression, but also with a range of comorbid conditions.

Central characteristics of CBT

The general principles and nature of CBT are described elsewhere. Two specific points relate to depression:

(a) Demands of CBT

Given the nature of depression, CBT challenges both therapist and patient. It requires **active engagement** (e.g. willingness to carry out self-help assignments), yet depressed patients often lack motivation and energy. It is based on a **friendly collaboration**, yet depressed people often find it hard to talk and clinicians may find their negativity aversive. It is an **educational approach**, using written materials and record sheets, yet depressed patients often have concentration and memory difficulties. Its stance is **optimistic**, yet depressed patients are often afraid (or convinced) that change is impossible. Therapists should be alert to these difficulties, understand them as aspects of depression rather than blaming the patient ('She must want to be depressed') or themselves ('I'm no good at this'), and maintain a persistent, problem-solving stance.

(b) Advantages of CBT

Nonetheless, CBT has real advantages for depressed patients. Its **structure** discourages rumination, and helps patients to focus systematically on their difficulties. Its emphasis on a warm **therapeutic relationship** encourages empathy, while its **goal orientation** implies that change *is* possible. The **coherent model** of human functioning, on which it is based, allows it to address many issues, including depression itself, comorbid conditions, problems in living, long-standing difficulties (such as low self-esteem), patients' responses to therapy and therapist, and therapists' responses to patients. Its **emphasis on collaboration and on transfer of knowledge and skill** empowers patients to become their own therapists and to take control of their lives.

Background

Beck's interest in the role of cognition grew out of his practice as a psychoanalytical therapist. Dissatisfied with analytical understandings of depression, he became curious about the role of the negative thinking he observed in depressed patients.[2–4] Beck's clinical observation and research consistently showed the thinking of depressed people to be dominated by self-derogation, negative expectations, overwhelming problems and responsibilities, deprivation and loss, and escapist and suicidal wishes, themes fuelled by systematic biases in information processing. He suggested that patients could recover from depression by learning to re-evaluate everyday cognitions, and to understand the long-standing idiosyncratic schemas under lying them.

This early scientist–practitioner stance has remained central CBT, stimulating an integrated flow of experimental investigation, practice development, and research into treatment efficacy, which has continued to the present day. The first successful outcome trial of CBT for depression appeared in 1977.[5] A detailed treatment manual emerged shortly afterwards.[1] Thirty years of randomized controlled trials now support the treatment's effectiveness.[6] Like other short-term focused psychological treatments, it has consistently proved as effective as antidepressant medication post-treatment. It reduces the likelihood of relapse by about 50 per cent, and this effect endures,[7] and can be enhanced by booster sessions in the months following treatment.[8] Thus CBT emerges as surprisingly cost-effective.[9] With adequate training and supervision equivalent results can be achieved even with severe depression and high comorbidity.[10,11]

Technique

Cognitive case conceptualization

(a) Enduring cognitive vulnerability to depression

The **cognitive model of depression** proposes that enduring cognitive structures and processes shape how everyday experience is interpreted, and are in turn reinforced by these interpretations. This model (Fig. 6.3.2.3.1) forms the basis for an **individualized conceptualization**, developed and shared with the patient, which informs and guides therapy. It suggests that **experience** (loss, events with lasting implications for self-worth)[12,13] leads people to reach **fundamental conclusions** about themselves, others, and the world ('basic' or 'core' beliefs, or schemas). They devise **guidelines for living** ('conditional assumptions'), which allow them to operate in the world, assuming the truth of those conclusions. Using schemas and rules to organize experience and guide behaviour is a normal part of human functioning. However, where schemas are globally negative (e.g. 'I am inferior') and assumptions extreme and resistant to change (e.g. stringent perfectionism), they become counterproductive. Evidence for **cognitive vulnerability** to depression prior to first onset is now emerging,[14] as is evidence that recurrent episodes leave an tendency to re-experience depressogenic processing patterns in the presence of mild, normal depressed mood ('cognitive reactivity').[15,16]

(i) Relationship between thinking and other aspects of depression

Dysfunctional beliefs and assumptions are **activated by events** that match the person's particular sensitivities. So a person with negative beliefs about the self whose psychological well-being depends on love and approval might become depressed after experiencing rejection. Activation of the system results in an upsurge of '**negative automatic thoughts**'—'negative' in that they are associated with painful emotions, and 'automatic' in that they pop into the person's mind rather than being a product of reasoned reflection. Such thoughts reflect **processing biases** such as overgeneralization. Depression is characterized by biased negative thoughts about the self (e.g. 'I'm useless'), the world (e.g. 'My situation is intolerable'), and the future (e.g. 'Nothing will ever change'). The latter (hopelessness) is central to suicidality.

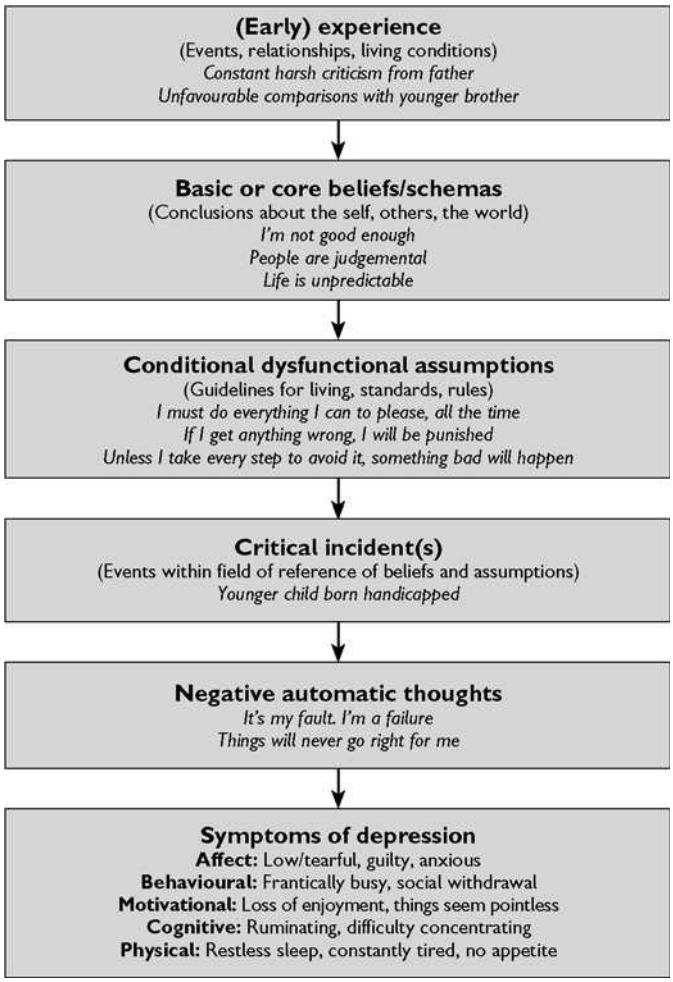


Fig. 6.3.2.3.1 Cognitive model of depression.

The more depressed people become, the more negative thoughts they think, and the more they believe them. The more negative thoughts they think and the more they believe them, the more depressed they become. Thus depression is maintained by **vicious circles** in which negative thinking and other symptoms reinforce one another. Experimental and clinical research reflects this **reciprocal relationship between affect and cognition**: modifying depressive thinking modifies depressed mood, while modifying depressed mood modifies depressive thinking.

(ii) Objectives of CBT for depression

The first goal is to **break the vicious circle** described above by teaching patients to work effectively with negative automatic thoughts. Attention then turns to cognitive predisposing factors (beliefs, assumptions) in order to **reduce vulnerability** and decrease the likelihood of future episodes. The aim is not to teach realistic thinking *per se*, but rather to help patients to resolve their problems by teaching them cognitive behavioural skills.

(iii) Overview of CBT for depression

Treatment usually proceeds through the following **stages**:

- diagnosis, assessment, problem identification
- cognitive interventions designed to reduce the frequency of negative thoughts
- behavioural assignments intended to tackle behavioural and motivational deficits
- monitoring, questioning, and testing negative automatic thoughts (the main body of therapy)
- relapse prevention.

Patients usually move from one stage to the next as each is mastered. The **starting point varies**: severely depressed patients often begin with simple behavioural interventions, whereas relatively mild depressions may immediately be amenable to cognitive work. At each stage, **cognitive and behavioural interventions are closely integrated**.[17] Thus behavioural interventions such as activity scheduling present opportunities to identify, question, and test negative thoughts, while work on thoughts and assumptions is closely linked to changes in behaviour.

Traditionally, **up to 20 sessions** of therapy are offered, twice weekly for 4 weeks and once weekly thereafter. Most patients respond within about 15 sessions. Some do well with 4–6 sessions, but severe chronic depressions may require more than 20 sessions, as well as shorter, more frequent sessions early on. Post-treatment boosters help to increase confidence and consolidate and extend skill. Sessions start with agenda-setting (prioritizing what to work on), homework review, and feedback on the previous session. After the day's main topics have been discussed, more homework practice is agreed to ensure generalization and reinforce new learning. Key points are summarized, and the therapist asks for reactions to the session, including anything that has been uncomfortable or unclear.

Indications and contraindications

CBT was developed as a treatment for moderate-to-severe unipolar depression,[1] and has been consistently effective with this population. However, good average results in outcome trials conceal wide variations in responsiveness.[18] Consistent predictors of treatment response remain elusive.

Patients presenting with endogenous symptoms are as likely to respond well as non-endogenous patients. However, results for severe depression[19] and bipolar disorder[20] remain somewhat conflicting.

Some factors may facilitate a positive response; well-developed pre-existing cognitive and behavioural coping skills, acceptance of the cognitive model, willingness to engage in self-help assignments, an early focus on teaching specific cognitive-therapy skills, ease of access to thoughts and feelings, problem specificity, and ability to form a collaborative alliance. Such criteria can be explored at assessment, especially if it is divided into two sessions with a simple intervening homework assignment.

Patients who are unsuitable for short-term CBT[21] may nonetheless respond to a modified approach with:

- greater emphasis on socialization and on cultivating a solid working relationship
- more extended behavioural work
- more work on enduring depressogenic schemas in the later stages of treatment
- more emphasis on integration with antidepressant medication
- more careful attention to environmental factors (including not only life stresses and family relationships, but also the ward and health-care team)

Finally, given that CBT and antidepressant medication generally produce similar results, the patients' wish for psychological treatment should also be taken into account.

Selection procedure

Diagnosis: recognizing and labelling depression

CBT was designed for moderate-to-severe major unipolar depression. The diagnostic criteria for major depressive episode,[22] remarkably consistent for over 30 years, describe a symptom pattern (including cognitive features) that has been recognized throughout history, and appears basically consistent across age, gender, race, and culture. However, the relative emphasis on different symptoms and the manner in which distress is expressed varies, and assessment procedures should be adapted to explore cultural context. Additionally, sociocultural factors necessarily influence belief systems, and therapists should be sensitive to such differences, rather than assuming that their own assumptions are shared by their patients.

Severity

Severity should be taken into account when deciding whether concurrent (or alternative) physical treatments or hospitalization are necessary, and in determining where to begin within CBT. Severity can be assessed through clinical interview (e.g. intensity, pervasiveness, and reactivity of depressed mood; extent of behavioural and interpersonal deficits). The **Beck Depression Inventory** (BDI-II)[23] a well-established self-report measure of depression, provides a rapid overview of symptoms. Weekly completion allows clinicians to observe overall progress, as well as tracking scores on particular items, e.g. hopelessness. Hopelessness and suicidality should be routinely assessed[24] and any sign of suicide risk investigated in-depth.[25]

Clinical interview

Interview assessment for CBT includes the following:

(a) Symptoms and associated cognitions

Negative automatic thoughts both trigger and enhance symptoms of depression (Table 6.3.2.3.1). Identifying meanings attached to symptoms prepares the ground for more helpful perspectives (e.g. 'These are symptoms of depression, not a reflection of my worth as a person').

(b) Impact on functioning

It is important to establish how depression affects relationships, work performance, and leisure time. It may be necessary to take practical steps to improve the patient's situation (e.g. gradual reintroduction to work).

(c) Coping strategies

The more depressed the patient, the more likely that she/he has adopted coping strategies which help in the short-term, but are in the longer-term self-defeating (e.g. alcohol or drugs, social withdrawal, bed). Therapist and patient can discuss the pros and cons of these, and how cognitive behavioural strategies might be more beneficial. The aim is for patients to reach the point of trying more adaptive coping strategies for themselves.

Table 6.3.2.3.1 Negative automatic thoughts and symptoms of depression

Symptoms	Negative automatic thoughts
Behavioural	
Lowered activity level	I can't do it. It won't work.
Procrastination	I'll never get it done. It's too much for me.
Motivational	
Loss of energy	It's too much effort. I'll wait till I fell better.
Loss of pleasure, interest	I won't enjoy it. What's the point? I can't be bothered.
Affective	
Sadness	I've lost everything important to me.
Guilt	I'm letting everybody down.
Anger	Why can't people just leave me alone?
Shame	What must everyone think?
Anxiety	I'm not going to be able to cope.
Cognitive	
Indecisiveness	Whatever I do will go wrong.
Poor concentration	I must be going senile.
Physical	
Loss of sleep	If this goes on, I won't be able to function.
Loss of appetite	I'm going to make myself ill.
Loss of sexual appetite	Our marriage is at an end.
Other	
Suicide	This is unbearable. There is no other way out.
Problems in living	There's nothing to be done.

(d) Onset of current episode

Information about the onset of episodes may provide valuable clues about beliefs and assumptions. For example, a young woman became depressed when her husband took a job abroad, and was away for long periods. She believed that he would not have done so if he truly loved her. In fact, he had taken the job because it paid exceptionally well and the savings they could make would allow them to start the family they had been planning. The therapist noted the patient's interpretation of her husband's behaviour, and later in therapy used this clue as a starting point to identify long-standing doubts about her attractiveness, and a linked dysfunctional assumption: 'If someone is not there for me all the time, it means they don't care about me'.

(e) Background

(i) Previous treatment

Many depressed patients presenting for CBT have already received other treatment (most commonly antidepressant medication and counselling). The therapist should inquire about the **outcome** of such treatment, and **what the patient makes of it**. Depressed patients often conclude that the incomplete success of prior treatment means that they will also be unable to benefit from this new approach. Such thoughts can be worked with using straightforward cognitive behavioural methods (identify, question, test). When the patient has received psychological therapy, it is often helpful to ask **what they learned** from it. If they feel they learned nothing, this may predispose them to approach CBT with pessimism.

Alternatively, if they learned something of value, CBT can build on this positive experience.

(ii) Expectations of CBT

These may reflect general pessimism about change, especially if patients know nothing about the approach. Those who have heard or read about CBT, and are aware of outcome data, may be more optimistic, although still anxious lest it fail to help *them*. Others may have heard negative reports (e.g. that it is mechanistic and fails to address deep issues). **Non-defensive discussion** of expectations allows doubts and misconceptions to be addressed, as well as encouraging open-mindedness and continued frank feedback from the patient.

(iii) Early experience and resultant beliefs and assumptions

CBT is **conceptualization driven**,[26] **closely based on the cognitive model of depression**. In order for patients to understand how their problems developed, and that current beliefs are learned opinions rather than reflections of fact, it is helpful to know about relevant formative experiences. This is particularly true when working with severe, chronic problems, and personality disorders. However, **obtaining historical information is often not an immediate priority**. It is more important initially to convey the hopeful message that something can be done to change things for the better. Therefore details about history, beliefs, and assumptions may not be explored until work on behavioural deficits and negative automatic thoughts has produced a reduction in hopelessness and depressed mood. That said, where patients are relatively mildly depressed, psychologically minded, and able to articulate their difficulties with ease, a draft conceptualization incorporating historical information, beliefs, and assumptions sometimes emerges even from the initial assessment.

Managing treatment

(a) Starting treatment

The **first treatment interview has four main objectives**:

- to establish a warm, collaborative therapeutic **alliance**
- to list specific **problems and** associated **goals**, and select a first problem to tackle
- to educate the patient about the **cognitive model**, especially the vicious circle that maintains depression
- to give the patient first-hand experience of the focused, workman-like, empirical **style** of CBT.

These convey two important messages: (1) it is possible to make sense of depression; (2) there is something the patient can do about it. These messages directly address hopelessness and helplessness.

(i) Identifying problems and goals

The **problem list** usually includes symptoms of depression. It may also contain aspects of other disorders (e.g. panic attacks), problems in living (e.g. family conflicts), and, in some cases, long-standing psychological problems (e.g. fear of intimacy). Developing the list provides the therapist with a 'map of the territory', which suggests possible targets for intervention, as well as an opportunity to foster the therapeutic relationship by demonstrating empathy. It suggests that apparently chaotic experience can be broken down into manageable problem areas. **Goal identification** then implies that progress is possible.

(ii) Introducing the cognitive model of depression

The therapist's next task is to demonstrate **how negative thinking influences emotion and behaviour**, relating this to the patient's experience using material derived from the session. The therapist explains that the patient will learn to notice negative automatic thoughts, to stand back, question them, and develop more realistic and helpful perspectives. Patients are often doubtful about their ability to do this. It is important therefore to present **CBT as a learning opportunity** during which skills can be acquired, step by step, with the therapist's guidance. The therapist is not obliged to convince patients that CBT will work for them, but a willingness to try it in practice is essential.

(iii) Where to start?

A **first treatment target** is chosen towards the end of the first session, and an appropriate **homework** task agreed upon. Suitable homework tasks include: observing a recording of the session and noting important points (this assignment follows every session), reading assignments,[27] and self-monitoring assignments. The initial target varies. Where patients are only relatively mildly depressed, and remain active and capable of experiencing interest and pleasure, monitoring negative automatic thoughts can begin right away. Where the depression is more severe, with significant behavioural and motivational deficits, it is best to begin with behavioural interventions.

(b) Behavioural interventions

(i) Reducing rumination

In severe depression, access to more positive perspectives may be blocked by depressed mood, making modifying negative automatic thoughts difficult or impossible. At the same time, depressed patients spend a great deal of time ruminating about their difficulties and shortcomings.[28] Learning to **direct attention elsewhere** reduces the frequency of negative thoughts and hence improves mood. This palliative measure will not resolve the patient's problems, but its impact reinforces the model, and feeling somewhat better can facilitate more constructive thinking.

(ii) Monitoring activities

Lowered activity levels and loss of interest and pleasure are often central to depression, and interventions designed to address them are known to be powerful in their own right.[19] Early behavioural interventions serve to **maximize engagement in activities providing a sense of pleasure and mastery**. This has a direct impact on mood, and provides opportunities to test negative thoughts that block engagement and, in a more global sense, prevent recovery (e.g. 'I can't do anything to change how I feel').

Patients record what they do, hour by hour, on a Weekly Activity Schedule (Fig. 6.3.2.3.2). Each activity is rated **0–10 for Pleasure (P) and Mastery (M)**. P ratings indicate how enjoyable the activity was, and M ratings how much of an achievement it was. 'P' is usually easily understood, but M can present difficulties. Depressed people often feel that nothing they do is an achievement, perhaps because most of their activities are routine ('What's so special about that?') or do not meet their standards ('I should have done more'). M should therefore be explained as 'an achievement, *given how you felt at the time*'. Thus even simple activities (e.g. making a cup of tea) are real achievements when patients are hampered by low mood and loss of energy. Ratings should be made immediately

Name _____ Week beginning _____

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
9–10			Bed (P5, M0)	Take kids to school (P0, M9)	Bed (P4, M0)		
10–11			Bed (P4, M0)	Shopping with friends	Bed (P0, M0)		
11–12			Bed (P2, M0)	↓ (P6, M7)	Bed (P0, M0)		
12–1			Bed (P1, M0)	Lunch with friend (P7, M3)	Up, shower, dress (P2, M7)		
1–2			Bed (P0, M0)	Went for a walk (P7, M5)	Lunch, read paper (P3, M4)		
2–3			Up, shower, dress (P0, M5)	Gardening (P4, M7)	Therapy Session		
3–4			Pick up kids from school (P3, M7)	Pick up kids from school (P1, M3)			
4–5			Tea with friends (P4, M9)	Tea & TV with kids (P5, M2)			
5–6			Tea with friends (P6, M7)	Ironing (P0, M5)			
6–7			Get supper (P2, M5)	Get supper (P1, M5)			
7–8			Read with kids (P8, M3)	Read with kids (P7, M3)			
8–12			TV (P3, M1) Bed (P8, M0)	TV (P4, M0) Bed (P9, M0)			

Fig. 6.3.2.3.2 Weekly activity schedule. P, pleasure; M, mastery.

after each activity, since retrospective ratings may be distorted by negatively biased recall. In addition, it is helpful for patients to **review each day**, asking questions like: 'What worked for me?' 'What did not work?' 'What do I need more of? Less of? Different?'

(iii) Planning activities

Once self-monitoring is mastered, **each day is planned in advance** on an hour-by-hour basis. This:

- provides a structure and helps with setting priorities
- averts the need to keep making decisions about what to do next
- reduces what may seem like chaos to a manageable list
- increases the chances that activities will be carried out
- enhances patients' sense of control.

A pattern of activities is sought in which **mastery and pleasure are balanced and maximized**. The plan is likely to contain a blend of obligations (e.g. the ironing) and pleasures (e.g. listening to music). Avoided tasks can be included, broken down into manageable steps ('graded task assignment'). Again, it is helpful for patients to review each day in detail, identifying unhelpful thoughts to be worked on in the next therapy session (e.g. 'If I can't complete the task, I might as well not bother at all'). Thus the cognitive element is present even when behaviour change is the primary target.

(c) Working with negative automatic thoughts

Once behavioural methods have been mastered, patients learn to **identify, question, and test negative automatic thoughts**. The main tool here is the Dysfunctional Thoughts Record (Fig. 6.3.2.3.3). The example summarizes a lengthy discussion that took place when a patient experienced a serious setback midway through treatment.

(i) Identifying negative automatic thoughts

Patients learn to **record upsetting incidents** as soon as possible after they occur (delay makes it difficult to recall thoughts and feelings accurately). They learn:

1 **To identify unpleasant emotions** (e.g. despair, anger, guilt), signs that negative thinking is present. Emotions are rated for intensity on a 0 to 100 scale. These ratings (though the patient may initially find them difficult) help to make small changes in emotional state obvious when the search for alternatives to negative thoughts begins. This is important, since change is rarely all-or-nothing and small improvements may otherwise be missed.

2 **To identify the problem situation**. What was the patient doing or thinking about when the painful emotion occurred (e.g. 'waiting at the supermarket checkout', 'worrying about my husband being late home')?

3 **To identify negative automatic thoughts associated with the unpleasant emotions**. Sessions direct the therapist towards asking: 'And what went through your mind at that moment?' Patients become aware of thoughts, images, or implicit meanings that are present when emotional shifts occur, and record them word-for-word. Belief in each thought is also rated on a 0 to 100 per cent

Date	Emotion(s) *What did you feel? How bad was it (0–100%)?*	Situation *What were you doing or thinking about?*	Automatic thoughts *What exactly was running through your mind? Write your thoughts down, word for word. How far did you believe each of them (0–100%)?*	Alternative views *What alternatives are there to the automatic thoughts? How far do you believe each of them (0–100%)?*	Outcome *1. How far do you now believe the original thoughts (0–100%)? 2. How do you now feel (0–100)? 3. What can you do now (action plan, experiment)?*
Thurs	Depressed (95%) Hopeless (90%)	Three terrible days	I'm back to square one—I've lost everything I learned (100%)	Not true — even now, I'm not as bad as when I came into hospital (100%)	1. (60%)
				I am doing my housework, looking after the children, doing my job. I am getting some satisfaction out of it — it's not a total failure (75%)	2. Depressed (70%) Hopeless (35%)
			There's no point in doing anything. Nothing will work (100%)	I've been feeling very bad, but setbacks are to be expected — disappointment at the contrast with last week makes it worse (100%)	3. Accept setbacks as part of recovery, not the end of the world.
			I've tried everything now, and nothing has changed (100%)	I've been getting this therapy for 7 weeks — I've been depressed for 3 years. It's not surprising I haven't got over it completely. Already I can manage 75% of my depression, as opposed to 25% (100%)	Make a detailed plan to help me deal with these feelings if they come again.
			I've failed again (100%)	This setback is not my failure —it is part of the problem (80%)	Keep using what I have learned to deal with my depression
			I will always be like this (100%) The only solution is to kill myself (95%)	Suicide is not the answer. Keep working on your thoughts. The past 7 weeks, and today, show it can work (100%)	

Fig. 6.3.2.3.3 Dysfunctional thoughts record.

scale (100 per cent represents complete belief, 50 per cent a moderate degree of belief, and so on). Again, this helps to make small changes in conviction evident at the next stage.

The skill of identifying painful emotions and associated thoughts is best learned if therapist and patient **work through examples on the sheet** before the patient self-monitors independently. Therapists can make sure that patients understand what is required, and are prepared for possible difficulties. For example, patients sometimes avoid recording thoughts because doing so is upsetting. Therapists can reassure them that this phase will pass once they learn to answer their thoughts, and suggest that they follow recording by engaging in an absorbing and pleasurable activity. Sometimes thoughts recorded do not seem to 'fit' the emotion experienced; in this case, therapists may need to help patients to **'unpack' the meaning of the thought** (for example, 'That didn't go too well' may turn out to mean 'I'm a total failure'). Time taken to learn accurate self-monitoring varies; many patients acquire the skill within a few days, but others take much time and coaching.

(ii) Questioning negative automatic thoughts

Once patients can record thoughts and feelings, they learn to **search for alternative views**, writing these in the fifth column of the Dysfunctional Thoughts Record. There is no such thing as a 'right' answer to a negative thought; the 'right' answer is the one that helps the patient to feel better and handle the situation more constructively. Accordingly, the therapist's task is not to suggest alternatives, but rather to elicit them through '**guided discovery**', a process of sensitive questioning, which allows patients to reach new interpretations independently. It is helpful for therapists to develop a personal 'library' of questions, through discussion with colleagues, observation of other therapists, attendance at workshops, and reading. Productive areas of inquiry include:

1 **What is the evidence?** Processing biases in depression mean that patients give weight to information consistent with prevailing perspectives at the expense of information, which suggests that they may not be wholly true. The therapist thus needs to examine 'evidence' believed to support the thought, and also to seek information that might contradict it.

2 **What alternative views are there?** Questions such as the following can prompt alternative perspectives: 'How would you have reacted to this before you became depressed?' 'What is your perspective on this when you feel relatively well?' 'What might someone whose views you trusted make of this?' 'If someone you cared about came to you with this problem, what would you say?'

3 **What are the advantages and disadvantages of this way of thinking?** This approach is particularly helpful with self-critical thinking. Patients often believe that self-criticism is an effective way of bringing about change; in fact, it only intensifies depression. Patients who habitually self-criticize can be helped to draw

up an analysis of pros and cons. Apparent advantages (e.g. 'It keeps me on my toes') may in fact be outweighed by disadvantages (e.g. 'It paralyses me').

4 **What are the biases in my thinking?** The tendency to make inferential errors such as overgeneralization has already been mentioned. Learning to recognize these can be helpful, especially when patients regularly make the same mistake.

Alternatives reached by questioning negative automatic thoughts are recorded on the Dysfunctional Thoughts Record. The patient rates them for degree of belief, to ensure that they are *sufficiently* convincing (they do not require belief ratings of 100 per cent). If alternatives are not at all convincing, they will have no impact on the strength of the original automatic thoughts or associated emotions. These are now re-rated in the final column as a check that plausible alternatives have been found.

As with self-monitoring, these skills are **best learned by working through examples in session** before the patient attempts to answer thoughts independently. Even then, patients are sometimes unable to find alternatives, especially if emotion is high. This is quite normal, given that questioning one's thoughts is a complex skill. It may be helpful to leave searching for alternatives until the storm is past. Sometimes alternatives make no difference to the original thoughts or emotions. This may be because the patient has reservations about their validity ('Yes, but . . . '), which can dealt with like other negative thoughts. Alternatively, it may be that non-verbal methods (e.g. imagery work, experiential learning) are necessary to facilitate emotional change, or that the resistant thought is a more or less direct statement of an underlying belief of much longer duration, which will take longer to change.

(iii) Testing negative automatic thoughts: what can I do now?

It is important that cognitive changes brought about by questioning are consolidated through **behavioural experiments**.[17] These are often designed to test out the validity of the new perspective by seeking further information or acting differently and observing the results. They may also include practical plans to solve genuine life problems and to deal with the trigger situation differently should it occur again.

(d) Ending treatment

Although most episodes of depression are time-limited, **relapse and recurrence are common**—the more so, the more episodes a person has experienced. CBT therefore emphasizes working on cognitive vulnerability factors, summarizing and consolidating learning, and preparing for possible setbacks. A new approach, *Mindfulness-Based Cognitive Therapy*,[29] which integrates elements of CBT with intensive meditation practice, has been developed specifically to tackle this problem. Its effectiveness with patients who have experienced three or more episodes of depression has been demonstrated in two clinical outcome trials.[30,31]

(i) Re-evaluating dysfunctional assumptions

Once patients are skilled at answering negative automatic thoughts, attention turns to dysfunctional assumptions that make them vulnerable to depression. Often these emerge from information gathered earlier, for example repeating themes in Dysfunctional Thoughts Records. They may also be identified using a 'downward arrow' technique, which involves identifying situations that typically distress the patient, and associated thoughts. Instead of responding directly to these, the therapist asks: 'If that was true, what would it mean to you?' This question (or variants) is repeated until a general assumption or rule, relevant to a range of situations, emerges. The validity of the rule is then questioned and tested. This process normally takes several sessions to complete. A helpful sequence of questions is given below (these are not exhaustive:

1 **Where did this rule come from?** Identifying the source of a dysfunctional assumption (e.g. parental criticism) often helps to encourage distance by suggesting that its development is understandable, though it may no longer be relevant or useful.

2 **In what ways is the rule unrealistic?** Dysfunctional assumptions do not fit the way the world works. They operate by extremes, which are reflected in their language (always/never rather than some of the time; must/should/ought rather than want/prefer/would like).

3 **In what ways is the rule helpful?** Dysfunctional assumptions are not usually wholly negative in their effects. For example, perfectionism may lead to genuine high-quality performance. If such advantages are not recognized and taken into account when new assumptions are formulated, the patient may be reluctant to move forward.

4 **In what ways is the rule unhelpful?** The advantages of dysfunctional assumptions are normally outweighed by their costs. Perfectionism leads to rewards, but it also undermines satisfaction with achievements and stops people learning from constructive criticism.

5 **What alternative rule might be more realistic and helpful?** Once the old assumption has been undermined, it is helpful to formulate an explicit alternative (e.g. 'It is good to do things well, but I am only human-sometimes I make mistakes'). This provides a new guideline for living, rather than simply undermining the old system.

6 **What needs to be done to consolidate the new rule?** As with negative automatic thoughts, re-evaluation is best made real through experience: behavioural experiments. These encourage patients to challenge specific examples of old rules, as well as testing out the validity of new ones by acting as if they were true and observing the results This systematic work may need to continue for weeks or indeed months, given that assumptions have often been in place for many years.

(ii) Re-evaluating negative core beliefs

Negative thoughts often disappear as patients recover, whether the depression is treated by psychological means or not. Sometimes, however, they reflect enduring beliefs about the self, the world (including other people), or the future, which if left untouched may predispose the patient to become depressed again. Methods for dealing with these have primarily been developed in the context of CBT for personality disorder,[32,33] but can often be used within short-term CBT. This is important, given limited resources.

The cognitive model suggests that negative beliefs contributing to vulnerability to depression are (like dysfunctional assumptions) based on early learning, and maintained by a consistent bias in favour of information, which confirms them, and against information, which contradicts them. Therapists help patients to become aware of this bias, to question the 'evidence' that upholds

the negative beliefs (much as the 'evidence' in favour of negative automatic thoughts is questioned), and to search actively for information which contradicts it. Once a relevant belief has been identified (e.g. 'I am no good') and rated for degree of belief (0–100 per cent), the suggestion is introduced that this may be more of an opinion than a fact (work at the level of automatic thoughts should have prepared the ground for this idea). If possible, the patient is asked to suggest a more positive alternative (e.g. 'If you were not 'no good', how would you like to be?'), and belief in the alternative (which is likely to be low) is also rated. The alternative provides a new 'address' at which to store information inconsistent with the old belief. However, it is not always possible to find one at this stage (e.g. when the patient has predicated his or her life on the belief and accumulated a large body of supporting evidence). In this case, an alternative may only become available once the old belief has been systematically weakened.

Supporting 'evidence' may include events from the distant past, which have been interpreted in a self-derogatory way (e.g. childhood abuse), as well as later experiences (e.g. a broken marriage) and everyday events of the kind already recorded on the Dysfunctional Thoughts Record. Each item is questioned, and new and more adaptive interpretations arrived at. In addition, patients are asked to record evidence that would support a more positive alternative to the old belief (e.g. examples of their strengths and skills). The success of these interventions is assessed by repeatedly rating the degree of belief in the old system, as well as in the new alternatives. This work too may take a considerable time, especially if negative beliefs have had a sizeable impact on the person's life. Where treatment time is limited for practical reasons, clinicians may find it helpful to space out later sessions, ensuring that intervening weeks are used to consolidate and extend within-session work.

(iii) Consolidating learning: 'blueprints'

Preparation for ending treatment begins with the treatment contract. The implication of offering a limited number of sessions is that treatment will end, and that patients will acquire the skills necessary to deal with depression independently. Throughout therapy, they are encouraged to take increasing responsibility for determining session content, making practical suggestions, devising homework assignments, summarizing learning, and applying new skills in fresh areas. Written session summaries and therapy tapes encourage reflection and consolidation.

At the end of treatment, gains are summarized in a personal action plan or 'blueprint for the future'. The blueprint is confined to one or two sheets of paper, guides continued learning, and helps deal with relapse or recurrence. It draws on the case conceptualization, session notes, homework records, reading materials, and the like. The therapist should examine sessions, records, etc. independently, so that the plan is drawn up jointly, nothing important is forgotten, and the patient goes away with as full a summary as possible. The following questions provide a useful framework:

1 **How did my problems develop?** (unhelpful beliefs and assumptions, the experiences that led to their formation, events precipitating onset)

2 **What kept them going?** (maintenance factors)

3 **What did I learn from therapy that helped?** (techniques (e.g. activity scheduling) and ideas (e.g. 'I can do something to influence my mood'). Techniques should be detailed so that patients know exactly what to do should depression recur. Examples of handouts and record sheets can be included.)

4 **What were my most unhelpful negative thoughts and assumptions? What alternatives did I find to them?** (summarized in two columns)

5 **How can I build on what I have learned?** (a solid, practical, clearly specified action plan)

(iv) Preparation for setbacks

The blueprint should also be used to plan for relapse. It is helpful right at the beginning of treatment to tell patients that, however well they do, they may well experience a setback at some point, not least because periods of low mood are a normal part of human experience. CBT will not prevent the patient from ever having another moment's distress; it will provide tools for dealing with distress more effectively. This information can help patients to respond with less fear and despair when they do encounter setbacks.

Preparation for setbacks can be framed by the following questions:

1 **What might lead to a setback for me?** For example, future losses (e.g. children leaving home) and stresses (e.g. financial difficulties), i.e. events which impinge on patients' vulnerabilities and are thus liable to be interpreted negatively. For people who have experienced recurrent depression, mild normal low mood (without any major environmental stimulus) can act as a trigger for negative thinking which, if unchecked, can spiral down into clinical depression.

2 **What early warning signs do I need to be alert for?** Feelings, behaviours, and symptoms that might indicate the beginning of another depression are identified and listed, using careful analysis of this and previous episodes and of fluctuations in mood occurring during treatment.

3 **If I notice that I am becoming depressed again, what should I do?** Clear simple instructions, which will make sense despite low mood, are needed here. Specific ideas and techniques summarized earlier in the blueprint should be referred to. General encouragement can also be included (e.g. 'Don't panic'), as well as a specific plan for what to do if cognitive behavioural methods do not lift mood within a specified period (e.g. contact the general practitioner about medication, contact the therapist for telephone discussion or booster sessions, contact emergency services or telephone helplines in the event of serious suicidal thoughts). Recontacting therapists is often difficult, as patients may feel that they have failed or them down. Therapists should make it clear that they consider it a sign of courage to ask for further help, not a sign of weakness.

(e) Training

Cognitive behaviour therapy for depression is a sophisticated treatment, requiring theoretical knowledge, research familiarity, and clinical expertise. The latter is best developed through practical training and close supervision. Core skills have been operationalized in measures of therapist competency, such as the *Cognitive Therapy Scale*,[34] which allow practitioners to judge for themselves whether they are indeed practising cognitive behaviour therapy and to monitor skills development.

Treating depressed patients with cognitive behaviour therapy is a challenge, especially with severe, chronic, and relapsing depressions

and depression comorbid with other conditions. Therapist competency (the ability to carry out the treatment as intended and to an adequate standard) has a direct impact on outcome. Thus the need for experienced therapists, and for adequate training and ongoing supervision, especially with more difficult groups, cannot be overemphasized.[10,11]

Where established training institutes exist, it is wise to take advantage of their expertise. Even where no local institute is available, expert therapists from established centres often travel internationally to deliver conferences, workshops, seminars, and supervision—notably the triennial World Congress of Cognitive and Behaviour Therapies.

However, novice therapists will sometimes have little option but to supervise themselves, using clinical texts as a knowledge base. Therapists at all skill levels will benefit from regularly monitoring audio or video recordings of their treatment sessions, using the *Cognitive Therapy Scale* to identify strengths and areas in need of improvement. Even if no more experienced practitioner is available, peer supervision with interested colleagues is helpful. It provides external feedback on clinical practice, and makes other forms of learning possible (role play, study groups, etc.).

(f) Self-care

Depression is an infectious disease. It is easy for psychological therapists (especially if relatively inexperienced) to become contaminated by patients' hopelessness. Therefore supervision should include a focus on the therapist's own thoughts and feelings. Therapists with a substantial proportion of depressed clients should also ensure that they leaven their day by planning life-enhancing and pleasurable experiences, and should be prepared to use cognitive behavioural methods to address their own negative thoughts.

Training and supervision

The Beck Institute for Cognitive Therapy and Research in Philadelphia, Pennsylvania, runs extramural courses and training programmes (www.beckinstitute.org), as does the Oxford Cognitive Therapy Centre (www.octc.co.uk), who also offer a selection of CBT oriented booklets for patients and clinicians. The Academy of Cognitive Therapy offers information about CBT, training in CBT, certification as a cognitive therapist, and (once a member) access to a ListServe on which issues relating to theory, research, and clinical practice can be discussed with experienced cognitive therapists (www.academyofcognitivetherapy.org) The International Association of Cognitive Psychotherapy (www.cognitivetherapyassociation.org) also has a ListServe and hosts regular conferences in member countries. Different countries also have their own national organizations promoting CBT, for example, the American Association for Behavioural and Cognitive Therapies (www.aabt.org), the Australian Association for Cognitive and Behavioural Therapies (www.aabct.org), and the British Association for Behavioural and Cognitive Psychotherapies (www.babcp.com). These will be able to offer information about training opportunities.

Further information

A number of texts describe the theory and practice of CBT for depression in some detail.[1,17,21] Practical ideas can also be found in self-help texts for patients.

For the clinician:

Beck, J.S. (1995). *Cognitive therapy: basics and beyond*. Guilford Press, New York.

Fennell, M.J.V. (1989). Depression. In *Cognitive behaviour therapy for psychiatric problems: a practical guide* (eds. K. Hawton, P.M. Salkovskis, J. Kirk, and D.M. Clark). Oxford Medical Publications, Oxford.

Fennell, M. J. V., Westbrook, D., and Benneth-Levy, J. (2004). Depression. In *The Oxford guide to behavioural experiments in cognitive therapy* (eds. J. Benneth-Levy, G. Butler, M.J.V. Fennell, *et al.*). Oxford University Press, Oxford.

For the patient:

Beck, A.T. and Greenberg, R.L. (1974). *Coping with depression*. Available from The Beck Institute of CBT and Research, GSB Building, City Line and Belmont Avenues, Suite 700, Bala Cynwyd, Philadelphia, PA 19004–1610, USA.

Burns, D.D. (1980). *Feeling good*. New American Library, New York.

Butler, G. and Hope, A. (2007). *Manage your mind* (2nd edn). Oxford University Press, Oxford.

Fennell, M.J.V. (1999). *Overcoming low self-esteem*. Constable Robinson, London.

Gilbert, P. (1996). *Overcoming depression*. Constable Robinson, London.

Greenberger, D. and Padesky, C.A. (1995). *Mind over mood*, Chap. 10. Guilford, New York.

Westbrook, D. (1999). *Coping with depression*. Available from: Oxford Cognitive Therapy Centre, Dept of Clinical Psychology, Warneford Hospital, Oxford OX3 7JX, UK.

Subjective experience:

Solomon, A. (2002). *The noonday demon*. Simon & Schuster, New York.

Styron, W. (1991). *Darkness visible*. Jonathan Cape, London.

Wolpert, L. (1999). *Malignant sadness*. Faber, London.

Current research and theoretical developments are best followed through professional journals (primarily psychiatry and clinical psychology—see the reference list for examples), available in libraries and through the Internet.

References

1. Beck, A.T., Rush, A.J., Shaw, B.F., *et al.* (1979). *Cognitive therapy of depression*. Guilford Press, New York.
2. Beck, A.T. (1963). Thinking and depression: 1. Idiosyncratic content and cognitive distortions. *Archives of General Psychiatry*, **9**, 324–33.
3. Beck, A.T. (1964). Thinking and depression: 2. Theory and therapy. *Archives of General Psychiatry*, **10**, 561–71.
4. Beck, A.T. (1967). *Depression: clinical, experimental and theoretical aspects*. Harper and Row, New York.
5. Rush, A.J., Beck, A.T., Kovacs, M., *et al.* (1977). Comparative efficacy of cognitive therapy and pharmacotherapy in the treatment of depressed outpatients. *Cognitive Therapy and Research*, **1**, 17–37.
6. Hollon, S.D., Thase, M.E., and Markowitz, J.C. (2002). Treatment & prevention of depression. *Psychological Science in the Public Interest*, **3**, 39–77.
7. Paykel, E.S., Scott, J., Cornwall, P.L., *et al.* (2005). Duration of relapse prevention after cognitive therapy in residual depression: follow-up of controlled trial. *Psychological Medicine*, **35**, 59–68.
8. Jarrett, R.B., Kraft, D., Doyle, J., *et al.* (2001). Preventing recurrent depression using cognitive therapy with and without a continuation phase. *Archives of General Psychiatry*, **58**, 381–8.
9. Vos, T., Corry, J., Haby, M.M., *et al.* (2005). Cost-effectiveness of cognitive-behavioural therapy and drug interventions for major depression. *The Australian and New Zealand Journal of Psychiatry*, **39**, 683–92.

10. DeRubeis, R.J., Hollon, S.D., Amsterdam, J.D., *et al.* (2005). Cognitive therapy vs. medications in the treatment of moderate to severe depression. *Archives of General Psychiatry*, **62**, 409–16.
11. Hollon, S.D., DeRubeis, R.J., Shelton, R.C., *et al.* (2005). Prevention of relapse following cognitive therapy vs. medications in moderate to severe depression. *Archives of General Psychiatry*, **62**, 417–22.
12. Ingram, R.E. (2003). Origins of cognitive vulnerability to depression. *Cognitive Therapy & Research*, **27**, 77–88.
13. McGinn, L.K., Cukor, D., and Sanderson, W. (2005). The relationship between parenting style, cognitive style and anxiety and depression: does increased early adversity influence symptom severity through the mediating role of cognitive style? *Cognitive Therapy & Research*, **29**, 219–42.
14. Alloy, L.B., Abramson, L.Y., Whitehouse, W.G., *et al.* (2006). Prospective incidence of first onsets and recurrences of depression in individuals at high and low cognitive risk for depression. *Journal of Abnormal Psychology*, 115, 145–56.
15. Segal, Z.V., Gemar, M., and Williams, S. (1999). Differential cognitive response to a mood challenge following successful cognitive therapy or pharmacotherapy for unipolar depression. *Journal of Abnormal Psychology*, **108**, 3–10.
16. Lau, M.A., Segal, Z.V., and Williams, J.M.G. (2004). Teasdale's differential activation hypothesis: implications for mechanisms of depressive relapse and suicidal behaviour. *Behaviour Research and Therapy*, **42**, 1001–17.
17. Fennell, M.J.V., Westbrook, D., and Bennett-Levy, J. (2004). Depression. In *The Oxford guide to behavioural experiments in cognitive therapy* (eds. J. Bennett-Levy, G. Butler, M.J.V. Fennell, *et al.*). Oxford University Press, Oxford.
18. Tang, T.Z., DeRubeis, R.J., Beberman, R., *et al.* (2005). Cognitive changes, critical sessions, and sudden gains in cognitive-behavioural therapy for depression. *Journal of Consulting and Clinical Psychology*, **73**, 168–72.
19. Dimidjian, S., Hollon, S.D., Dobson, K.S., *et al.* (2006). Randomized trial of behavioural activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology*, **74**, 658–70.
20. Scott, J., Payken, E., Morris, R., *et al.* (2006). Cognitive behavioural therapy for severe and recurrent bipolar disorders. Randomised controlled trial. *The British Journal of Psychiatry*, **188**, 313–20.
21. Moore, R.G. and Garland, A. (2003). *Cognitive therapy for chronic and persistent depression*. Wiley, Chichester, UK.
22. American Psychiatric Association. (1994). *Diagnostic and statistical classification of diseases and related health problems* (4th edn). American Psychiatric Association, Washington, DC.
23. Beck, A.T., Ward, C.H., Mendelson, M., *et al.* (1961). An inventory for measuring depression. *Archives of General Psychiatry*, **4**, 561–71.
24. Beck, A.T., Weissman, A., Lester, D., *et al.* (1974). The measurement of pessimism: the Hopelessness Scale. *Journal of Consulting and Clinical Psychology*, **42**, 861–5.
25. Beck, A.T. (1993). *Beck scale for suicide ideation* (*manual*). Psychological Corporation, San Antonio.
26. Persons, J.B. (1989). *Cognitive therapy in practice: a case formulation approach*. Norton, New York.
27. Beck, A.T. and Greenberg, R.L. (1974). *Coping with depression*. Available from The Beck Institute of Cognitive Therapy and Research, GSB Building, City Line and Belmont Avenues, Suite 700, Bala Cynwyd, Philadelphia, PA 19004–1610, USA.
28. Papageorgiou, C. and Wells, A. (2003). *Depressive rumination: nature, theory and treatment*. Wiley, Chichester.
29. Segal, Z.V., Williams, J.M.G., and Teasdale, J.D. (2002). *Mindfulness-based cognitive therapy for depression: a new approach to preventing relapse*. Guilford, New York.
30. Teasdale, J.D., Segal, Z.V., Williams, J.M.G., *et al.* (2000). Reducing risk of recurrence of major depression using Mindfulness-Based Cognitive Therapy. *Journal of Consulting and Clinical Psychology*, **68**, 615–23.
31. Ma, J. and Teasdale, J.D. (2004). Mindfulness-based cognitive therapy for depression: replication and exploration of differential relapse prevention effects. *Journal of Consulting and Clinical Psychology*, **72**, 31–40.
32. Padesky, C.A. (1994). Schema change processes in cognitive therapy. *Clinical Psychology and Psychotherapy*, **1**, 267–78.
33. Beck, A.T., Freeman, A., Davis, D.D., *et al.* (2004). *Cognitive therapy of personality disorders*. Guilford Press, New York.
34. Young, J.E. and Beck, A.T. (1980). *Development of an instrument for rating cognitive therapy: the cognitive therapy scale*. University of Pennsylvania, Philadelphia, PA.

6.3.2.4 Cognitive behaviour therapy for schizophrenia

Max Birchwood and Elizabeth Spencer

Introduction

Cognitive behaviour therapy (**CBT**) for schizophrenia focuses on the core psychotic symptoms of hallucinations and delusions. Other psychosocial approaches to psychosis (e.g. intervention with families and to promote medication compliance) also frequently use CBT techniques. In this chapter, however, we focus on CBT for delusional beliefs and other psychotic phenomena and review evidence for its efficacy.

Background: assumptions and common components

The CBT approach to psychotic symptoms comprises two different strands each with their own theoretical basis, although of late these two approaches have become conjoined in practice.

Coping strategy enhancement

The first approach is inspired by the stress-vulnerability model of schizophrenia. Vulnerability here is viewed as a 'black box', drawing mainly on the biomedical tradition. It is assumed that stressors capable of triggering or exacerbating symptoms may be generated or modulated by the individual. For example, stressors emanating from the social environment are modulated by the patient's own appraisal of their stressfulness and his or her coping strategies.

Another class of stressors consists of the symptoms themselves. It is assumed that certain strategies used to cope with symptoms are unhelpful and generate stress in the individual, in turn, exacerbating symptoms. These strategies are conventionally divided into affective strategies (e.g. relaxation, sleep, etc.), behavioural strategies (being active, drinking alcohol, etc.), and cognitive strategies (distraction, challenging voices, switching attention away from voices, etc.). This underpins the approach known as Coping Strategy Enhancement[1] whereby patients are offered a range of strategies which are implemented in an empirical fashion to determine their effectiveness in symptom control. For example, Falloon and Talbot[2] documenting the coping strategies used by voice-hearers, concluded that those who had multiple strategies available

to them were more able to cope with their voices. Tarrier[3] on the other hand, focusing on a wider range of psychotic symptoms, concluded that those who applied strategies consistently tended to fare the best. This approach views the individual as an active agent who attempts to reduce the threat or distress posed by psychotic symptoms, but does not concern itself with the content or meaning that psychotic symptoms may have to the individual. There is also, in this approach, assumed to be a fundamental discontinuity between normal and abnormal functioning that comes about once the biological vulnerability is 'online'.

CBT for delusions and hallucinations

The second CBT strand draws its theoretical strength from the cognitive therapy approach.[4] Early work in this area focused on the similarity between normal (but strongly held) beliefs and delusions, in terms of the psychological processes at play in their maintenance. For example, Brett-Jones *et al.*[5] showed that delusions, like everyday beliefs, lead the individual to recruit evidence to support them and to de-emphasize or dismiss contradictory evidence. Continuing the exploration of the continuity between normal and delusional beliefs, Birchwood and Chadwick[6] argued that certain beliefs about voices' power may be considered as a quasi-rational response to anomalous experience, with the meaning attributed to them in terms of identity, power, and the consequence of disobedience determining distress and behaviour in relation to the voice. Other work has drawn on the cognitive therapy approach in depression, which emphasizes the importance of evaluative beliefs about the self (e.g. self-worth) in the genesis and maintenance of depressed mood.[7] The application of this to psychosis also emphasizes evaluative beliefs about the self. The precise relationship between self-evaluative beliefs and delusional thinking is a much debated issue of the present time. It has been argued, for example, that delusions may serve the function of defending the individual from the full impact of low self-worth through blaming others for negative events rather than the self. This is the so-called 'paranoid defence'.[8] The content of psychotic thinking often reflects such personal issues. For example, for the patient who has been sexually abused, this theme tends to crop up in the content of voice activity or in the supposed identity of the voice.[6]

This early work has been elegantly drawn together in a cognitive model of psychosis by Phillipa Garety and her colleagues. In a seminal paper[9] they propose a model of the cognitive processes leading to the positive symptoms of psychosis. In brief, positive symptoms in psychosis are hypothesized to begin with basic cognitive disturbances with lead to ambiguous sensory input, the intrusion into consciousness of unintended material from memory, or to difficulties with the self-monitoring of intentions and actions, such that they are experienced as alien. This result in anomalous conscious experiences such as actions being experienced as unintended, racing thoughts, thoughts appearing to be broadcast, and thoughts experienced as voices.

However, the authors argue that such anomalous experiences alone do not develop into full-blown psychotic experiences unless an individual appraises them as externally caused and personally significant. Such appraisals are the results of particular reasoning processes (e.g. data gathering bias or externalizing attributional style), dysfunctional personal schemas (e.g. low self-esteem born of adverse social experience), emotional states (e.g. anxiety and depression), and appraisal of the experience of illness.

This model integrates the two strands of cognitive therapy. It suggests, for example, that the reduction in dysfunctional emotional states through, for example, coping strategy enhancement, will contribute to alterations in the attributions, which are important in the formation and maintenance of the positive symptoms. Similarly, it provides a theoretical basis for the basic techniques traditionally used in the second strand of CBT. Such techniques encourage the individual to weigh evidence that contradicts a delusion as a strategy to compensate for the basic information-processing abnormality, challenge negative self-schemata, and combat depression.

Evaluation

In recent years the volume of trials evaluating CBT for psychosis has greatly expanded, with approximately 20 such studies now reported. Most of these have been conducted in the United Kingdom among patients with chronic schizophrenia. The strength of the data is now sufficient for The UK National Institute for Clinical excellence to state that cognitive behavioural therapy should be available as a treatment option for people with schizophrenia.[10]

An up-to-date review[11] reporting the analysis of 19 studies of CBT for positive symptoms in schizophrenia, found a mean effect size of 0.37. The authors concluded that 74 per cent of these studies achieved small effect sizes, 32 per cent moderate effect sizes, and 16 per cent large effect sizes in improving positive psychotic symptoms, relative to standard psychiatric care. Furthermore, they argued that this is unlikely to be due to publication bias. The effect sizes for CBT versus standard treatment among patients with chronic illness were greater than those among acutely ill patients. This may have been due to a ceiling effect caused by the effectiveness of medication in reducing symptoms in floridly ill inpatients. Similarly, the better the design of the trial, the smaller the treatment effect size, suggesting that CBT is not the panacea for all psychotic ills that it may have originally appeared.

With regard to relapse prevention, CBT appears to be more successful when the intervention is focused on relapse prevention, rather than relapse prevention being one of a series of components.[11] For example, Gumley and his colleagues[12] were able to demonstrate that a group of patients with psychosis receiving targeted CBT for relapse prevention had almost half the rate of relapse over a 12-month period compared with a similar group receiving treatment as usual (18.1 versus 34.7 per cent). In this study, the CBT treatment consisted of an engagement phase, early signs of relapse monitoring with a personalized questionnaire, and targeted CBT at the first sign of impending relapse.

While large, pragmatic trials of CBT treatment packages have yielded the above favourable results, investigating the active elements of the interventions is difficult because CBT for psychosis now refers to a wide range of treatments.

Furthermore, although the conceptual basis of CBT emphasizes the link between emotion, cognition, and behaviour, modifying emotion in psychosis has been relatively neglected in CBT trials, in favour of outcomes based on modification of delusions and hallucinations themselves.

It has been proposed that CBT should be focused into more targeted interventions aimed at emotional dysfunction or distress and/or behavioural anomaly in psychosis that is directly or indirectly linked to psychosis symptoms.[13] This approach recognizes

that while changing the psychosis symptoms might not always be possible, it may well be feasible to change the affective consequences of the symptoms or the diagnosis. This affective change may have further benefits in reducing the severity of the psychosis experience *per se*.

For example:

1 CBT can be used to reduce distress, depression, and problem behaviour associated with commanding voices, without changing the frequency or content of the voices themselves.[14]

2 CBT can focus on anxiety, depression, and interpersonal difficulty in individuals at high risk of developing psychosis.[15]

3 CBT can focus on the relapse prodrome to prevent relapse in psychosis.[12]

4 CBT can focus on 'comorbid' depression and social anxiety, including the patient's appraisal of the diagnosis and its stigmatizing consequences.[16]

5 CBT can be used to reduce stress reactivity, thereby increasing resilience to life stress and preventing psychotic relapse.[17]

6 CBT can be used to increase self-esteem and social confidence in people with psychosis.[18]

Management

Coping strategy enhancement

Coping strategy enhancement involves developing a coping repertoire and over-rehearsing it to facilitate an automatic coping response.[19] It can either be used to improve an individual's attempts to cope with his or her voices by developing an understanding of factors that trigger or improve the voices, or to test the reality of thoughts about the voices.

For example, therapy with a patient who hears threatening and frightening voices exacerbated of being alone might involve: An explicit congratulation on the strength involved in withstanding the voices' incessant activity; strengthening of coping strategies involving seeking company and social support; and the use of a personal stereo for distraction.

Cognitive therapy to challenge delusions and dysfunctional assumptions

The application of cognitive therapy in challenging of delusions and dysfunctional beliefs draws upon the approach described by Chadwick *et al.*[20] and builds upon the pioneering work of Chadwick and Lowe.[21]

Engaging patients is perhaps the greatest challenge facing a therapist. It is noticeable that many individuals either never attend or do so for a few sessions and then stop. Once individuals get past the opening strategies of cognitive therapy they usually see therapy through. Careful attention to appropriate therapeutic technique can maximize client engagement. Similarly, Kuipers *et al.*[22] report that a response to CBT in their study was associated with greater cognitive flexibility concerning delusions at baseline. This suggests that the sufferer may need some small degree of insight into the fact that he might be mistaken, to benefit from cognitive therapy for delusions.

The process involves six basic steps as summarized in Table 6.3.2.4.1.

Table 6.3.2.4.1 A summary of the steps in cognitive therapy for delusions

1 Viewing delusions as beliefs, not facts
2 Developing a rationale for questioning the delusion
3 Weakening delusions
4 Utilizing inconsistency and irrationality
5 Reformulating delusions as reactions to, and attempts to make sense of, specific experience
6 Assessing the delusion and alternative, and empirical testing

(a) Viewing delusions as beliefs, not facts

The first technical difficulty to be encountered is the necessary move to aid the client in conceptualizing a delusion as a belief and not a fact. This move is an essential part of CBT for all emotional problems, but is difficult at the best of times. With depression, for example, patients often struggle to appreciate that their sense of worthlessness, which is so concrete to them, is actually a belief they hold and is different from a knowledge of events and facts. With delusions there is the added complication that the therapist might be perceived as being just another person who disbelieves the patient.

There are two central points to bear in mind when seeking to reconceptualize delusions as beliefs, not facts—why it is being done and how it is done.[20]

The purpose of clarifying that delusions are beliefs, not facts, is to empower the patient and offer a way of easing his or her distress. If the patient really is being persecuted by a powerful organization, or has a radio transmitter and receiver in his head, neither he nor the therapist can actually change this. The patient feels that he knows this as a fact, with the consequence that he feels frustrated and helpless as well as distressed. However, if the patient only believes this to be true, then he gains the freedom to examine his beliefs and perhaps change his distressing feelings and behaviour and experiences himself. In this sense it is in his best interest for the delusion to be false.

How this process takes place is critical. The process of Socratic questioning is not one of persuading a patient that he is wrong and that you, the therapist, are right. This mistake is made all too often. Rather, in Socratic dialogue the therapist helps the patient to draw on his own doubt and experience in order to realize that there are other ways in which he is able to make sense of his experience. So, when the therapist pursues the conceptual step of clarifying that a delusion is only a belief, the patient's own doubt, past or present, his own contradictory experience and behaviour, and concern about the possibility that the delusion is wrong are accessed. Many patients have 'double awareness' of delusions—on the one hand they believe them firmly and are distressed and disturbed by them, yet on the other hand they behave in ways that contradict the delusion, and they believe that working with a therapist might ease the problem. Finally, the therapist must accept that it is acceptable if the patient does not alter his belief. The process is 'collaborative empiricism, not indoctrination'.[17]

(b) Developing a rationale for questioning the delusion

Patients are usually used to being told by family and carers that their beliefs are wrong, that they are deluded. It is easy for a therapist to

prepare the intervention well and embark on it before the patient is clear of the purpose and possible benefit, thus causing early loss of engagement. It is revealing to turn the engagement question on its head and to consider why a patient should ever wish to engage in therapy. With emotional problems patients identify their problems as depression, anger, anxiety, guilt, etc.; with delusions and hallucinations this is not so—patients predominantly present problems which they believe are actual events (persecution, voices, passivity). This means that they have no clear objective and therefore have no particular motivation to engage. The key reason for a patient to reconsider delusional beliefs is that it will help him feel less distress and it will free him to behave differently and to pursue the things he wants more directly. What the therapist does gradually through the unfolding cognitive assessment is to clarify with the patient that he is experiencing emotional and behavioural problems, and that these are tied to his beliefs (delusional and evaluative). The therapist then needs to explore with the patient how the delusion affects his life and how his life would be different (i.e. better or worse) if the delusion were false.[20] In this way, the therapist slowly encourages the patient to view the delusion not as an important discovery but as a belief that results in distress (e.g. fear, anxiety) and causes him to behave in ways which he would rather not (e.g. avoid things he would otherwise like to do).

(c) Weakening delusions

Disputing comprises four elements:[20]

- The evidence for the belief is challenged, in inverse order of its importance to the delusion.
- The internal consistency and plausibility of the delusional system is questioned.
- Following Maher,[23] the delusion is reformulated as being an understandable response to, and way of making sense of, specific experience, and a personally meaningful alternative is then constructed.
- The individual's delusion and the alternative are assessed in the light of the available information.
- Challenging the evidence for the belief.

Watts *et al.*[24] argued that a danger when trying to modify delusions, indeed, all strongly held beliefs, was psychological reactance, whereby too direct an approach served only to reinforce the belief. They offered two principles to minimize this possibility: begin with the least important belief, and also work with the evidence for the belief rather than the belief itself.

Accordingly a 'verbal challenge' of delusions begins by questioning the evidence for the belief, and this process begins with the least significant item of evidence and works up to the most significant one. Our preferred approach is that with each item of evidence the therapist questions the patient's delusional interpretation and puts forward a more reasonable and probable one. The customary approach in CBT is for the patient to be asked to generate the alternative interpretation(s), rather than the therapist supply one, but we have found that for certain patients this conventional tactic is a weak intervention.

When the therapist questions the evidence for a delusion there are two distinct but related objectives. One is to encourage the patient to question and perhaps even to reject the evidence for his or her belief, and in this way perhaps to undermine the patient's conviction in the delusion itself. For some individuals challenging the evidence is a very powerful intervention and one that produces a substantial reduction in delusional conviction. However, more commonly this does not happen, but challenging evidence is still valuable in that it does impart insight into the connection between events, beliefs, affect, and behaviour. This is the second objective of challenging evidence, namely to convey the essentials of the ABC approach, i.e. that strongly held beliefs influence affect, behaviour, and cognition (i.e. interpretation) for all people. Core beliefs recruit or bias everyday inferences and automatic thoughts. However, this means that people often impose an interpretation on to events, which is unwarranted, and because we are prone towards selectively processing information that confirms our beliefs, this goes undetected. In other words, it is understandable that a patient should interpret a particular event in line with his delusion because this is merely one occurrence of a general tendency, and confirmation bias, common to all of us. In therapy, it is helpful to convey the ordinariness and normality of this process with everyday examples.

Having considered the alternatives, the patient is then asked to rate his conviction about each; regardless of how convinced he remains that the delusional interpretation is correct, it is usual to move on to the next piece of evidence. The therapist does not have to change what the patient thinks, but only to offer a fresh insight into the way he is thinking.

(d) Utilizing inconsistency and irrationality

Although delusions contain differing degrees of inconsistency and irrationality, they all seem to contain some. For example, Margaret believed that she could not act or make a decision without reference to her voice; however, she described periods when she was relaxed and the voices quiescent where she would be making decisions. Such inconsistencies can be therapeutically useful.

(e) Reformulating delusions as reactions to, and attempts to make sense of, specific experience

We always construe a delusion as both a reaction to and an attempt to make sense of certain puzzling and often threatening experience. It is an understandable and reasonable attempt to find meaning at a time when the individual is bewildered, anxious, and frightened. But, the delusion carries a cost in terms of distress and disturbance, which the individual might not otherwise experience. This is how delusions may be explained to patients.[23]

At this stage the therapist has commenced the process of challenging the delusional belief, re-formed the delusion as an attempt to make sense of certain experiences (e.g. primary symptoms, trauma) and raised the idea that the delusion is psychologically functional (i.e. it eases puzzlement) and perhaps linked to evaluative beliefs.

(f) Assessing the delusion and alternative, and empirical testing

Finally, the patient and therapist need to assess the delusion and alternative in the light of the available evidence and previous discussion. The therapist may spell out the advantages of the alternative interpretative framework, which can also be discussed by relating it to the patient's experiences.

It is an integral part of CBT that the belief or assumption under consideration be tested empirically. Such reality testing involves planning and performing an activity that validates or invalidates a belief, or part of a belief.

When working with delusions, we set up a clear alternative belief in opposition to the delusion, clarifying with the patient in advance precisely what has to happen for each to be supported and refuted. For example,[20] Alison believed that by repeating her voice's command (e.g. 'The price of milk *will* rise a billion times') it would actually happen by transmitting the thought to a member of the government who would act upon it. The empirical test involved repeating the voice and purchasing milk before and after doing so, predicting that within 2 weeks the price of milk would at least double. If it did not, the alternative (that the power to change events was very weak) would be strengthened.

Cognitive therapy for beliefs about auditory hallucinations

The above techniques can be applied to challenge beliefs about 'voices' using the following three steps:

(a) Assessment

Assessing the personal meaning a voice has for a person is the defining feature of the cognitive approach to assessing auditory hallucinations. The delusional beliefs found to be most significant are those relating to a voice's identity, purpose, power, knowledge, and the consequences of compliance and resistance. The semi-structured interview schedule developed by Chadwick and Birchwood[6,25] is recommended.

(b) Disputing beliefs about voices

The thrust of the therapist's challenge is that the beliefs are reasonable and understandable reactions to, and attempts to make sense of, the auditory hallucinations. The therapist reviews evidence and inconsistency, and plans tests, with the aim always of evaluating two possible meanings: that the beliefs are true, a discovery, or that they are reasonable and understandable, but mistaken. As ever, it is vital that the therapist really practises Socratic questioning and works collaboratively. This involves drawing out patients' own doubts, puzzlement, double awareness, critical faculty, etc., rather than forcing a contradiction on them.

The major piece of evidence for the delusional beliefs is always the actual voices, especially their content—these are, after all, the activating events that the delusions are invoked to explain. The role of beliefs is critical because individuals usually attribute voices with a power and knowledge that goes well beyond what they have actually said. Several examples of challenging follow.

It is really quite common for beliefs about compliance (e.g. 'If I don't do what my voice says I will be punished') not to fit patients' experiences, and perhaps it is only their emotional impact which prevents patients from abandoning them. Kate, for example, believes 'If I drop my guard the voices will kill me', but in fact she has dropped her guard on many occasions without consequence. This might be pointed out as follows. 'Kate, you say that the voices have the power to kill you and you must be on your guard constantly. I certainly appreciate the fear that this must create. What puzzles me though is that your guard is often down, like when you are asleep. How is it that they have not succeeded in all these years?'

The appearance of being all knowing (omniscience) is a vital aspect of many voices[22] and often features as a key piece of evidence that the identity of the voice is superhuman. It leaves individuals feeling exposed and vulnerable and very prone to guilt and shame. Alice believes that her voice is a prophet endowed with the ability to foretell the future. In particular, the voice anticipates exactly the arrival of her husband home from work each day. To begin the process of questioning she was asked: 'Let's suppose for a moment that the voice cannot foretell the future; can you think of any other possible explanations for last night's prediction?' One such possibility was that the voice was making a very safe guess.

(c) Testing beliefs about control

A useful strategy is to use a procedure whereby the patient and therapist learn to engineer situations to start or increase the probability of hearing voices, and then to stop or reduce them. In this way the patient gains a surprising degree of control over the voice. The initial assessment provides information about cues that provoke voices for a particular individual; concurrent verbalization is known to stop or diminish voices temporarily. This information is combined in the following five steps.

- Identify cues that increase and decrease voices.
- Practise the use of 'increasing' and 'decreasing' strategies within a session.
- Propose the notion that 'control' requires the demonstration that voice activity can be turned up/on or down/off.
- In sessions encourage the patient to initiate or increase voice activity for short periods then reduce or stop it.
- Elicit changes in the patient's belief about his control over the voices.

The above process has been applied in a targeted way to the special case of beliefs about command hallucinations. These are high-risk, distressing, and relatively common symptoms of schizophrenia.

For example, Byrne and her colleagues[26] have developed a specific cognitive therapy for command hallucinations, which draws on the above techniques. Using the methods of collaborative empiricism and Socratic dialogue, the therapist seeks to engage the client to question, challenge, and undermine the power beliefs, then to use behavioural tests to help the client gain disconfirming evidence against the beliefs. These strategies are also used to build clients' alternative beliefs in their own power and status, and finally, where appropriate, to explore the origins of the schema so clients have an explanation for why they developed those beliefs about the voice in the first place. They were able to show that this process produce significant reductions in compliance behaviours and favourable changes in beliefs about the power, superiority and need to comply with the voices, despite the frequency, loudness, and content of the voices staying the same.[14]

Conclusion

CBT for psychosis is a rapidly developing field, and one that has borne considerable fruit in terms of providing effective treatments and a basis upon which a dialogue between the patient and the professional can take place about matters of great concern and a source of much distress.

Further information

Nelson, H. (2005). *Cognitive behavioural therapy with delusions and hallucinations: a practice manual* (2nd edn). Nelson Thornes, Cheltenham.

Morrison, A.P. (2002). *A casebook of cognitive therapy for psychosis*. Brunner/Routledge, London.

Byrne, S., Meaden, A., Trower, P., *et al.* (2006). *A casebook of cognitive behaviour therapy for command hallucinations: a social rank theory*. Routledge, London.

References

1. Tarrier, N., Yusupoff, L., Kinney, C., *et al.* (1998). Randomised controlled trial of intensive cognitive behavioural therapy for patients with chronic schizophrenia. *British Medical Journal*, **317**, 303–7.
2. Falloon, I.R.H. and Talbot, R. (1981). Persistent auditory hallucinations: coping mechanisms and implications for management. *Psychological Medicine*, **11**, 329–39.
3. Tarrier, N. (1987). An investigation of residual positive symptoms in discharged schizophrenic patients. *British Journal of Clinical Psychology*, **26**, 141–3.
4. Beck, A.T., Rush, A., Shaw, B., *et al.* (1979). *Cognitive therapy of depression*. Guilford, New York.
5. Brett-Jones, J., Garety, P., and Hemsley, D.R. (1987). Measuring delusional experiences: a method and its application. *British Journal of Clinical Psychology*, **26**, 257–65.
6. Birchwood, M. and Chadwick, P.D. (1997). The omnipotence of voices. II. Testing the validity of the cognitive model. *Psychological Medicine*, **27**, 1345–53.
7. Garety, P.A., Kuipers, L., Fowler, D., *et al.* (1994). Cognitive behavioural therapy for drug resistant psychosis. *The British Journal of Medical Psychology*, **67**, 259–71.
8. Bentall, R.P. and Kinderman, P. (1998). Psychological processes and delusional beliefs: implications for the treatment of paranoid states. In *Outcome and innovation in psychological treatment of schizophrenia* (eds. T. Wykes, N. Tarrier, and S. Lewis), pp. 119–44. Wiley, Chichester.
9. Garety, P.A., Kuipers, A.E., Fowler, D., *et al.* (2001). A cognitive model of the positive symptoms of psychosis. *Psychological Medicine*, **31**, 189–95.
10. National Collaborating Centre for Mental Health. (2003). *Schizophrenia: full national clinical guideline on core interventions in primary and secondary care*. Royal College of Psychiatrists and the British Psychological Society, London.
11. Tarrier, N. and Wykes, T. (2004). Is there evidence that cognitive behavioural therapy is an effective treatment for schizophrenia? A cautious or cautionary tale. *Behaviour Research and Therapy*, **42**, 1377–401.
12. Gumley, A., O'Grady, M., McNay, L., *et al.* (2003). Early intervention for relapse in schizophrenia: results of a 12-month randomized controlled trial of behavioural therapy. *Psychological Medicine*, **33**, 419–31.
13 Birchwood, M. and Trower, P. (2006). The future of cognitive behavioural therapy for psychosis: not a quasi-neuroleptic. *British Journal of Psychiatry*, **188**, 107–8.
14. Trower, P., Birchwood, M., and Meaden, A. (2004). Cognitive therapy for command hallucinations: randomised controlled trial. *British Journal of Psychiatry*, **184**, 312–20.
15. Morrison, A.P., French, P., Walford, L., *et al.* (2004). Cognitive therapy for the prevention of psychosis in people at ultra-high risk: randomised controlled trial. *British Journal of Psychiatry*, **185**, 291–7.
16. Iqbal, Z., Birchwood, M., and Chadwick, P. (2000). Cognitive approach to depression and suicidal thinking in psychosis. 2. Testing the validity of a social ranking model. *British Journal of Psychiatry*, **177**, 522–8.
17. Myin-Germeys, I., Delespaul, P., and van Os, J. (2005). Behavioural sensitization to daily life stress in psychosis. *Psychological Medicine*, **35**, 733–41.
18. Hall, P.L. and Tarrier, N. (2003). The cognitive behavioural treatment of low self esteem in psychotic patients: a pilot study. *Behaviour Research Therapy*, **41**, 317–32.
19. Tarrier, N. (1992). Management and modification of residual psychotic symptoms. In *Innovations in the psychological management of schizophrenia* (eds. M. Birchwood and N. Tarrier), pp. 38–72. Wiley, Chichester.
20. Chadwick, P.D., Birchwood, M., and Trower, P. (1996). *Cognitive therapy for delusions, voices and paranoia*. Wiley, Chichester.
21. Chadwick, P.D.J. and Lowe, C.F. (1990). Measurement and modification of delusional beliefs. *Journal of Consulting and Clinical Psychology*, **58**, 225–32.
22. Kuipers, E., Garety, P., Fowler, D., *et al.* (1997). London-East Anglia randomised controlled trial of cognitive behavioural therapy for psychosis. I. Effects of the treatment phase. *The British Journal of Psychiatry*, **171**, 319–27.
23. Maher, B.A. (1988). Anomalous experience and delusional thinking: the logic of explanation. In *Delusional beliefs* (eds. F. Oltmans and B.A. Maher), pp. 5–22. Wiley, New York.
24. Watts, F., Powell, E., and Austin, S.V. (1973). The modification of abnormal beliefs. *The British Journal of Medical Psychology*, **46**, 359–63.
25. Chadwick, D. and Birchwood, M. (1994). The omnipotence of voices. A cognitive approach to auditory hallucinations. *British Journal of Psychiatry*, **164**, 190–201.
26. Byrne, S., Meaden, A., Trower, P., *et al.* (2005). *A casebook of cognitive behaviour therapy for command hallucinations: a social rank theory approach*. Routledge, London.

6.3.3 Interpersonal psychotherapy for depression and other disorders

Carlos Blanco, John C. Markowitz, and Myrna M. Weissman

Introduction

Interpersonal psychotherapy (IPT) is a time-limited, diagnosis-focused therapy. IPT was defined in a manual. Research has established its efficacy as an acute and chronic treatment for patients with major depressive disorder (MDD) of all ages, as an acute treatment for bulimia nervosa, and as adjunct maintenance treatment for bipolar disorder.[1–9] The research findings have led to its inclusion in treatment guidelines and increasing dissemination into clinical practice.

Demonstration of efficacy in research trials for patients with major depressive episodes (MDEs) has led to its adaptation and testing for other mood and non-mood disorders. This has included modification for adolescent and geriatric depressed patients[10,11] patients with bipolar[12] and dysthymic disorders;[13,14] depressed HIV-positive[15] and depressed pregnant and postpartum patients;[16,17] depressed primary care patients;[18] and as a maintenance treatment to prevent relapse of the depression.[5] Most of the modifications have been relatively minor and have retained the general principles and techniques of IPT for major depression.

Non-mood targets have included anorexia, bulimia, substance abuse, borderline personality disorder, and several anxiety disorders. In general, outcome studies of IPT have suggested its promise for most psychiatric diagnoses in which it has been studied, with the exceptions of anorexia, dysthymic disorder, and substance use disorders.[14,19,20]

IPT has two complementary basic premises. First, depression is a medical illness, which is treatable and not the patient's fault. Second, depression does not occur in a vacuum, but rather is influenced by and itself affects the patient's psychosocial environment. Changes in relationships or other life events may precipitate depressive episodes; conversely, depressive episodes strain relationships and may lead to negative life events. The goal of treatment is to help the patient solve a crisis in his or her role functioning or social environment. Achieving this helps the patient to gain a sense of mastery over his or her functioning and relieves depressive symptoms.

Begun as a research intervention, IPT has only lately started to be disseminated among clinicians and in residency training programmes. The publication of efficacy data, the promulgation of practice guidelines that embrace IPT among antidepressant treatments, and economic pressures on length of treatment have led to increasing interest in IPT. This chapter describes the concepts and techniques of IPT and its current status of adaptation, efficacy data, and training. The chapter provides a guide to developments and a reference list, but not a comprehensive review.

Background

IPT traces its theoretical and clinical origins to the interpersonal psychoanalytic theory of Meyer and Sullivan and builds on work of other relational theory including object relations, particularly with regard to attachment. However, it applies this theory within a conceptual and clinical framework that differs significantly from that of Sullivan and much of relational theory. In contrast with psychoanalytically inspired schools of thought, IPT sees its goals in treating depression and other medical disorders, rather than trying to change overall personality. Pragmatically, IPT opts to narrow its focus to address the area of interpersonal life that seems to require the most immediate attention.

Acknowledging the importance of personality and early experience, IPT emphasizes the role of recent stressful events in triggering depression and other psychiatric disorders in vulnerable individuals, while it also recognizes the protective role social supports play against psychopathology. Nevertheless, IPT is less interested in discerning the *cause* of a depressive episode (since it assumes the aetiology of depression to be multifactorial) than in using the connection between current life events and the onset of depressive symptoms to help the patient understand and combat the episode of illness.

Compared to other psychotherapies, such as psychoanalytic psychotherapy or even cognitive behavioural therapy, IPT is relatively young. It is less concerned about maintaining an established orthodoxy than about adapting itself to the needs of the patient. Although IPT theorists have taken into account theoretical developments in psychiatry and related fields, much of IPT's evolution has been based on the results of clinical trials. As investigations continue into IPT as a treatment for different disorders and populations, further modification of its theoretical aspects as well as techniques are likely.

Indications

IPT research has demonstrated its efficacy for major depressive disorder across a range of patient ages and contexts, and for bulimia nervosa. One large trial indicates its efficacy (modified as interpersonal social rhythms therapy, or IPSRT) as an adjunctive treatment for bipolar disorder. [6] Lesser evidence suggests the potential benefits of IPT for several anxiety disorders. [21,22] IPT has shown no advantages over control psychotherapies for dysthymic disorder or substance abuse disorders. [14,20] For depressed adolescents, IPT has shown not only efficacy but effectiveness in a school-based programme. [10,11]

Both the physician and patient guides in primary care guidelines for depression list IPT, cognitive behavioural therapy (CBT), behavioural, brief dynamic, and marital therapy as treatments for depression. IPT is spreading from its initial research base in the United States. The IPT manual has been translated into Italian, German, Japanese, Spanish, and French, and is being used ever more widely around the world. Descriptions of IPT have appeared in Spanish, Norwegian, Finnish, and Dutch journals. An International Society for Interpersonal Psychotherapy, established at the American Psychiatric Association Annual meeting in May 2000 in Chicago, has a growing membership and biennial international meetings in 2004 and 2006, and maintains a bibliography of studies.

Because IPT focuses clinically on the social context of the depressive episode, researchers have sometimes adapted IPT when applying it to different treatment populations, developing manuals for different age groups or subpopulations, and occasionally adding focal problem areas. IPT has also been used at different lengths, in different formats, in one pilot couple's adaptation, and as a telephone intervention. Nonetheless, all these adaptations involve the basic principles that constitute IPT: a no-fault definition of the patient's problem as a medical illness, excusing the patient from blame for his or her symptoms; and a continual focus on the relationship between the patient's moods and life situation. The continuing growth of IPT research precludes an exhaustive description of studies. This chapter presents a selection of key research trials of IPT for mood and other disorders (see **Efficacy**) and offers selected references for further reading.

Contraindications

Although formal contraindications (i.e. situations in which IPT would worsen the patients' situation) are not known, IPT was never intended to function as a monotherapy for patients with psychotic depression or bipolar disorder. In addition, three controlled trials have found no benefit of IPT as a treatment for substance use disorders.

Conducting IPT

Each of the four IPT interpersonal problem areas has discrete goals for therapist and patient to pursue. The therapist helps the patient relate life events to mood and other symptoms. In this section we outline the phases of IPT, as well as common strategies and techniques used in IPT treatment. We also outline some differences with cognitive behavioural therapy, to which it is often compared.

Phases of treatment

As an acute treatment, IPT has three phases. The *first phase*, generally covering sessions 1–3, includes diagnostic evaluation,

psychiatric history, and setting the treatment framework. The therapist reviews symptoms, diagnoses the patient as depressed according to DSM-IV (or ICD-10) criteria, and gives the patient the sick role. The psychiatric history includes the 'interpersonal inventory', which is not a structured instrument but a careful review of the patient's past and current social functioning and close relationships, their patterns, and mutual expectations. The relationships are examined to see to what extent they are satisfactory, whether there have been recent changes in those relationships, or whether the patient desires to change them. As part of this review, the therapist commonly links the main social and interpersonal situations of the patient's life to the onset of depressive symptoms.

During the opening phase the therapist also sets a time limit for the acute treatment, generally between 12 and 16 sessions. The optimal number of sessions for IPT requires further research. One study suggests that as few as eight sessions may be effective for some patients, but similar to pharmacological treatment, different doses (i.e. number of IPT sessions) might be necessary for different patients. Sessions are generally scheduled weekly. This allows sufficient time to pass that things will happen in the patient's outside life, on which the treatment focuses. Yet it is frequent enough to maintain momentum and thematic continuity. However, in certain cases logistical difficulties (e.g. due to a general medical illness) might require less frequent sessions.

At the end of the first phase, the therapist links the depressive syndrome to the patient's interpersonal situation focusing on one of the four interpersonal problem areas: (1) *grief*; (2) interpersonal *role disputes*; (3) *role transitions*; or (4) *interpersonal deficits*. Once the patient explicitly accepts this formulation as the focus for treatment, IPT enters its middle phase.

It is important to keep treatment focused on a simple theme. Any formulation necessarily simplifies a patient's life narrative. Although some patients may present with multiple interpersonal problems, the goal of the formulation is to isolate one or, at most, two salient problems related to the patient's mood disorder (whether as precipitant or consequence). More than two foci would risk diffusing the treatment and diluting its efficacy. Sometimes a number of interpersonal problems contribute to the depressive episode, making it apparently difficult to choose a focus. However, research has shown that IPT therapists agree in choosing foci, and patients find those foci credible. Moreover, resolution of the interpersonal treatment focus appears to correlate with symptomatic improvement.

An important task of the initial phase requires deciding whether or not to use medication. A growing literature suggests that combined treatment with antidepressants and IPT works at least as well as, but is not always superior to IPT alone. Thus, except for very severe cases or possibly the elderly, the choice between IPT alone or combined with medication relies more on cost, availability of resources, and patients' preference than on existing empirical evidence.

The *middle phase* involves approaches specific to the chosen interpersonal problem area. For **Grief**—complicated bereavement following the death of a loved one—the therapist facilitates mourning and helps the patient find new activities and relationships to compensate for the loss. **Role disputes** are conflicts with significant others: a spouse, a child, other family members, co-workers, or a close friend. The therapist helps the patient explore the relationship, the nature of the dispute, and available options to negotiate its resolution, including ending the relationship. **Role transition** includes change in life status: for example, beginning or ending a relationship or career, moving, promotion, demotion, retirement, graduation, having a baby, or diagnosis of another medical illness. The patient learns to manage the change by mourning the loss of the old role, recognizing positive and negative aspects of the new role, and taking steps to master this new role.

Interpersonal deficits are used as a focus for patients who lack any of the first three focal life situations. Such patients are isolated or lack social skills, have problems in initiating or sustaining relationships. The goal is to help the patient to develop new relationships and skills. Some patients who fall into this category may in fact suffer from dysthymic disorder or social anxiety disorder, for which separate strategies have been developed.

The *final phase* of IPT, occupying the last 2–4 sessions of acute treatment, builds the patient's newly acquired sense of independence and competence by recognizing and consolidating therapeutic gains. Compared to psychodynamic psychotherapy, IPT de-emphasizes termination: it is a bitter–sweet graduation from successful treatment. The sadness of separating from the therapist is contrasted with depressive feelings.

If the patient has not improved, the therapist emphasizes that the *treatment* has failed, not the patient, and that alternative effective treatments, such as medication or other psychotherapies exist. If the treatment has succeeded, the therapist underscores the patient's competence to function without further therapy by emphasizing that the depressive episode has improved because of the patient's actions in changing a life situation. The therapist also helps the patient to anticipate triggers for and responses to depressive symptoms that might arise in the future.

Patients with multiple prior MDE's or significant residual symptoms, who successfully complete acute treatment but remain at high risk for relapse or recurrence, may contract for maintenance therapy as acute treatment draws to a close. At the end of the treatment (acute or maintenance, depending on the case) the patient is also explicitly told that, should depression recur, the patient should immediately seek treatment, just as the patient would do if any other medical illness recurred.

Techniques

Readers new to IPT will find that much of what we describe below sounds familiar and overlaps with other psychotherapies. Thus, on one level, IPT demands few novel skills from therapists and is relatively easy to learn.

The challenges of IPT lie not in the use of any individual technique, but in organizing these approaches to establish and maintain a coherent primary treatment focus and to resist the temptations of digressing into clinical material outside that focus. Additional challenges may arise from 'unlearning' reflexive responses from prior training experiences such as making transference-focused interventions (for psychodynamic therapists) or identifying automatic cognitions and schemas (for cognitive therapists). In our exposition of strategies and techniques, we focus on major depressive disorder, the first and still best tested indication for IPT, although the same principles may apply to other disorders.

(a) General strategies

IPT is organized around four important concepts.

(i) Psychoeducation

The therapist helps the patient to recognize that the problem is a common medical illness, a mood disorder, with a predictable set of symptoms, not the personal failure or weakness of the patient. IPT therapists define depression as a treatable condition that is not the patient's fault. This definition displaces guilt from the patient to the illness, decreases the patient sense of isolation by feeling part of a larger group (those with depression), and provides hope for a response to treatment.

Underscoring this approach, IPT therapists give depressed patients the 'sick role'. This role temporarily excuses them from what their illness prevents them from doing while assigning them the task of working as patients in order to recover their previous healthy role. The resolution of the sick role is to regain the healthy, euthymic role by the end of treatment. The time-limited structure of IPT also energizes patients and protects against regression during treatment.

(ii) Focusing on the positive

IPT therapists take an empathic, supportive, and encouraging stance. They emphasize their patients' successes, although they also commiserate on their difficulties. 'Focusing on the positive' means underscoring positive events; it does *not* mean ignoring negative affect. By doing this, IPT therapists may facilitate the therapeutic alliance that is crucial to good outcome. By solving an interpersonal crisis—a complicated bereavement, a role dispute or transition, or an interpersonal deficit—the IPT patient has the dual opportunity to improve his or her life situation and simultaneously relieve the symptoms of the depressive episode.

This coupled formula, validated by randomized controlled trials in which IPT has been tested, can be offered with confidence and optimism. Symptomatic relief may correlate with the degree to which the patient solves his or her interpersonal crisis. This therapeutic optimism, while not specific to IPT, very likely provides part of its power in remoralizing the patient.

(iii) Focus on the present, not the past

IPT deals with current rather than past interpersonal relationships, focusing on the patient's immediate social context. The IPT therapist attempts to intervene in depressive symptom formation and social dysfunction rather than addressing enduring aspects of personality, which are difficult to assess accurately during an episode of an Axis I disorder. However, IPT does build new social skills such as self-assertion and increased ability to understand interpersonal exchanges, which may be as valuable as changing personality traits.

(iv) Link mood to life events

A core strategy of IPT is constant attention to the link between the patient's current mood state and recent interpersonal experiences. Stressful life events and negative interpersonal encounters trigger lower mood and can lead to depressive episodes in vulnerable individuals. Conversely, depressed mood impairs social functioning, which can lead to further negative life events. IPT is postulated to work by helping the patients manage interpersonal relationships more effectively, which leads to improved mood. Improved mood then allows patients to more effectively manage interpersonal experiences in an iterative fashion.

Unlike psychodynamic psychotherapy, IPT *does not* focus on early childhood experiences and long-standing familial dynamics. Thus, the patient's current mood state is linked to recent experiences rather than those rooted in the distant past. Nor does IPT focus on transferential material, except in the relatively rare instance when problems arise in the therapeutic alliance. Thus the treatment highlights recent experiences outside the office.

(b) Specific techniques

To achieve the general goals of IPT, the following techniques are frequently used:

1 *An opening question*: 'how have things been since we last met?', which leads the patient to provide an interval history of mood and events. The therapist begins each session after the first one with this tactic. It is common, particularly at the beginning of the treatment that the patient will focus exclusively either on the mood or on a recent event. When that occurs, the therapist gently asks about the other aspect and helps the patient connect mood and recent events.

2 *Communication analysis*, a detailed recreation of recent, affectively charged circumstances. This detailed analysis often helps the patient uncover nuances of the interpersonal exchanges that had been missed prior to the session.

3 An exploration of the patient's wishes and options, to help the patient realize and voice the desired outcomes.

4 *Decision analysis*, to help the patient integrate communication analysis, the wishes and options and the constraints of the situation and decide on a specific course of action.

5 *Role-playing*, to help the patient rehearse that course of action before implementing in real life.

Similarities and differences with other psychotherapies

Because IPT and CBT are the two best empirically supported psychotherapies, they are often compared. [23] IPT shares with CBT an orientation towards making the patient feel understood, a 'here and now' focus, a general feeling of hope and optimism, psychoeducation, and the use of role-playing to favour the acquisition of new skills. It addresses interpersonal issues in a manner familiar to marital therapists. Although like CBT a time-limited treatment targeting a syndromal constellation (e.g. major depression), IPT is considerably less structured, and focuses on interpersonal problem areas rather than automatic thoughts. IPT overlaps to some degree with psychodynamic psychotherapies, yet IPT also meaningfully differs from them: in its focus on the present, not the past; its focus on practical, real-life change rather than self-understanding; its medical model; and its avoidance of the transference and of genetic and dream interpretations.

Efficacy

Research findings: IPT for mood disorders

IPT outcome research is ongoing, with new studies published every year. What follows is a selection of key research trials of IPT for mood and other disorders. For some of these trials, IPT was adapted in a separate treatment manual, but in all cases the general principles of the treatment remained the same.

(a) Acute treatment of major depression

IPT was first tested as an acute antidepressant treatment in a four-cell, 16-week randomized trial. This compared weekly IPT, amitriptyline (AMI), their combination, and a monthly supportive psychotherapy treatment for 81 outpatients with major depression.[24,25] The outcome of patients receiving amitriptyline and IPT was similar and superior to that of supportive psychotherapy. Patients who received both amitriptyline and IPT had better depression outcomes and better scores on a range of social adjustment measures including overall adjustment, work performance, and communication than those on amitriptyline alone, suggesting an additive effect of IPT on medication treatment. At 1-year follow-up, many patients sustained improvement from the brief IPT intervention, and IPT patients had developed significantly better psychosocial functioning whether or not they received medication. This effect on social function was not found for AMI alone and had not been evident for IPT at the end of the 16-week trial.

Still probably the most important study to date involving IPT is the National Institute of Mental Health Treatment of Depression[24,25] Collaborative Research Program (TDCRP), investigators randomly assigned 250 outpatients with major depression to 16 weeks of IPT, CBT, or either imipramine (IMI) or placebo with clinical management.[2] Most subjects completed at least 15 weeks or 12 treatment sessions. Patients with milder depression (defined as a 17-item Hamilton Depression Rating Scale [HDRS] score <20) improved equally in all four treatments. For more severely depressed patients (HDRS>20), IMI worked fastest and most consistently better than placebo. IPT and IMI were comparable on several outcome measures, including HDRS, and superior to placebo for more severely depressed patients. In some analyses, IPT appeared to be slightly superior to CBT. CBT was not superior to placebo among the more depressed patients.

A follow-up study of TDCRP subjects 18 months later found no significant difference in recovery among remitters (who had minimal or no symptoms after the end of treatment, sustained during follow-up) among the four treatments.[26] Thirty per cent of CBT, 26 per cent of IPT, 19 per cent of imipramine, and 20 per cent of placebo subjects initially randomized to those treatments remitted and remained in remission during that time span. Among acute remitters, relapse over the 18-month follow-up was 36 per cent for CBT, 33 per cent for IPT, 50 per cent for imipramine (medication having been stopped at 16 weeks), and 33 per cent for placebo. The authors concluded that, for many patients, 16 weeks of specific treatments were insufficient to achieve full and lasting recovery.

Special populations and settings

(a) Depressed primary care patients

There has been a study comparing IPT and nortriptyline with usual care for depressed patients in a primary care setting. If patients were hospitalized for a general medical condition, IPT was continued in the hospital when possible. Depressive symptom severity declined more rapidly with either nortriptyline or IPT than in usual care. Approximately 70 per cent of treatment completers receiving nortriptyline or IPT, but only 20 per cent in usual care, had recovered after 8 months. Subjects with a lifetime history of comorbid panic disorder had a poorer response across treatments, compared to those with major depression alone.

(b) Depressed HIV-positive patients (IPT-HIV)

IPT has also been investigated for depressed HIV patients (IPT-HIV), emphasizing common issues among this population including concerns about illness and death, grief and role transitions. A randomized study echoing the TDCRP of 101 HIV-positive patients with depressive symptoms found IPT and imipramine each superior to CBT and supportive therapy. Many patients reported improvement in depressive physical symptoms that they had mistakenly attributed to HIV infection. IPT may have been a better fit than CBT for these patients due to the extreme life events they faced at the height of the HIV epidemic.

(c) Peripartum depression

Pregnancy and the postpartum also provide natural role transitions as an IPT focus. Exploring these role transitions addresses the depressed pregnant woman's self-evaluation as a parent, physiologic changes of pregnancy and postpartum, and altered relationships with the spouse or significant other and with other children. Timing and duration of sessions are adjusted in response to bedrest, delivery, obstetrical complications, and childcare. Postpartum mothers may bring children to sessions. As with depressed HIV-positive patients, telephone sessions, and hospital visits are sometimes necessary. A controlled clinical trial comparing IPT to a didactic parent education group in depressed pregnant women showed advantages for IPT. A study of depressed postpartum women found superiority of IPT over a wait-list control group. A small randomized trial has also suggested the possibility that group IPT may serve to prevent MDD relapse during the postpartum period.

(d) Conjoint IPT for depressed patients with marital disputes (IPT-CM)

Marital conflict can precipitate or complicate depressive episodes. Some clinicians believe that individual psychotherapy for patients in marital disputes may lead to premature rupture of marriages. Researchers at Yale University developed a manual for conjoint therapy of depressed patients with marital disputes (IPT-CM). IPT-CM includes the spouse in all sessions and focuses on the current marital dispute. Eighteen patients with major depression linked to the onset or exacerbation of marital disputes were randomly assigned to 16 weeks of either individual IPT or IPT-CM. Patients in both treatments showed similar reductions in depressive symptoms, but patients receiving IPT-CM reported significantly better marital adjustment, marital affection, and sexual relations. These pilot findings require replication with a larger sample and other control groups.

(e) Depressed adolescents (IPT-A)

IPT has also been modified to incorporate adolescent developmental issues. Three randomized trials, one of them conducted in Puerto Rico, have shown the efficacy of IPT-A. It is important to note that in the Puerto Rico study, the only one that also included CBT in the design, IPT appeared superior to CBT in certain measures (e.g. self-esteem and social adaptation), consistent with the findings of the TDCRP and the study of depressed HIV-positive patients.

(f) Maintenance treatment

Based on the success of IPT as acute treatment for MDD, the recurrent nature of mood disorders, and the efficacy of medication in preventing relapse and recurrence, IPT was adapted as a once

monthly maintenance treatment for MDD (IPT-M). This was a novel development and allowed the first real testing of psychotherapy as a maintenance treatment for patients who had remitted from acute depression. Since IPT-M begins with patients who have remitted, its goal is to maintain the remitted state. Both patient and therapist are vigilant for early signs of interpersonal problems similar to those of which the patient and therapist previously identified as associated with the onset of the patient's most recent depressive episode. At the same time, the therapist works to enhance strengths that appear to have been present prior to the patient's illness or began to emerge as the most recent depressive episode remitted. In contrast with the acute phase application of IPT, which usually focuses on one or at most two interpersonal problem areas, IPT-M may shift problem areas over time. Three studies have compared medication and IPT as maintenance treatment for MDD.

In the first study, 128 outpatients with recurrent depression initially treated with combined high dose (>200 mg/day) imipramine and weekly sessions of IPT. [4,5] Responders remained on high-dosage medication while IPT was tapered to a monthly frequency during a 4-month continuation phase. Patients who remained remitted were then randomly assigned to 3 years of either: (1) imipramine plus clinical management; (2) imipramine plus monthly IPT; (3) monthly IPT alone; (4) monthly IPT plus placebo; or (5) placebo plus clinical management.

Both IPT and imipramine were significantly superior to the placebo group in delaying MDD relapse. Imipramine was superior to IPT-M in the ability to prevent relapse. The group that received IPT with imipramine had a numerically lower rate of recurrence at 1 year (16 per cent) than the group on imipramine alone (40 per cent), but those results were not statistically significant. Two different studies have had very similar findings in comparisons of IPT and nortriptyline for geriatric patients with recurrent major depression. [7, 27]

Further study is required to determine the efficacy of IPT relative to newer medications (e.g. selective serotonin-reuptake inhibitors), and the efficacy of dosages other than once monthly maintenance IPT. A study of differing doses of maintenance IPT for depressed patients in Pittsburgh has not found differences in outcome based on frequency of sessions.[28] Perhaps optimal dosing of maintenance IPT depends on individual patients' needs.

The success of IPT in treating MDD has led researchers to investigate its efficacy in bipolar and dysthymic disorder.

(i) Bipolar disorder

The modification of IPT used as an adjunct to medication in the treatment of bipolar disorder is called interpersonal and social rhythm therapy (IPSRT). Its use rests on the hypothesis that disruptions of social rhythms are destabilizing for bipolar patients and contribute to trigger their relapse. By decreasing the number and intensity of those disruptions, IPSRT should improve the course of bipolar disorder. The behavioural component helps to protect sleep patterns and limit the disruptions that may provoke mania; the IPT approach to depression remains largely the same.

After stabilizing bipolar I patients with appropriate pharmacotherapy and either IPSRT or intensive clinical management, patients were randomized again to either IPSRT or clinical management for preventive treatment.[6] They found that participants assigned to IPSRT acutely had longer survival times to a new affective episode, irrespective of maintenance treatment assignment. Participants in the IPSRT group had higher regularity of social rhythms at the end of the acute treatment, and this increased regularity of social rhythms during the acute treatment mediated the reduced likelihood of recurrence during the maintenance treatment. Further research appears necessary to more firmly establish the optimal timing and treatment duration of IPSRT.

(ii) Dysthymic disorder

A modification of IPT for dysthymic disorder[29] encourages patients to reconceptualize what they have considered their lifelong character flaws as ego-dystonic, chronic mood-dependent symptoms: as chronic but treatable 'state' rather than immutable 'trait'.

Three randomized trials have examined the efficacy of IPT in dysthymic disorder. In the first study, 35 patients with an ICD-10 diagnosis of dysthymia with or without comorbid MDD were randomized to moclobemide alone ($n = 19$) or moclobemide plus IPT ($n = 16$). Patients were assessed with the 17-item Hamilton Rating Scales for Depression (HAM-D), the Global Assessment Scale (GAS), and the Quality of Life and Satisfaction Questionnaire at baseline, 12, 24, and 48 weeks. Both groups showed statistically significant improvement in all measures across time. There were no differences between the two treatments at week 12. However, patients in the combined group had statistically better scores than the patients in the moclobemide group on all outcome variables at weeks 24 and 48.

In the second study, 707 adults in primary care clinic with DSM-IV dysthymic disorder were randomized, with or without past and/or current MDD (15 per cent of the sample had current MDD), to treatment with sertraline alone (50–200 mg), IPT alone (10 sessions), or sertraline with IPT combined.[30] At the end of treatment, response rates were 60 per cent for sertraline alone, 47 per cent for IPT alone, and 58 per cent for sertraline with IPT. After an additional 18-month naturalistic follow-up phase, there were no statistically significant differences in symptom reduction between sertraline alone and sertraline with IPT. However, both were more effective than IPT alone in reducing depressive symptoms. It is important to note, though, that IPT was given as a brief treatment, while sertraline was generally continued for the full 2 years of the study.

A third study compared IPT adapted for the treatment of dysthymia (IPT-D), brief supportive psychotherapy, sertraline, and combined IPT-D/sertraline for patients with pure dysthymic disorder (i.e. without 'double' depression) in 94 subjects treated over 16 weeks.[14] Patients improved in all conditions, with the cells including sertraline pharmacotherapy showing superiority over psychotherapy alone in response and remission. The results of this study are consistent with an emerging literature suggesting that pharmacotherapy may acutely benefit patients more than psychotherapy. In conjunction with the other studies in patients with dysthymic disorder, it suggests that IPT alone may not be an efficacious treatment for dysthymic disorder.

IPT for non-mood disorders

The efficacy of IPT as an antidepressant treatment has led to its adaptation as a treatment for other psychiatric disorders, based on the premise that life events are ubiquitous.

(a) Bulimia

Fairburn *et al.* modified IPT for the treatment of bulimia, eliminating the use of the sick role and of role-playing in order to contrast

distinct therapeutic strategies in comparing IPT and CBT. Initial trials showed that although CBT worked faster to relieve bulimic symptoms, IPT had longer-term benefits comparable to CBT and superior to a behavioural control condition.[1] A subsequent multisite trial found CBT superior to IPT.[9]

Following a model closer to the original IPT principles, Wilfley *et al*. modified IPT in a group format (IPT-G) and compared it to group CBT and a wait-list control for 56 women with non-purging bulimia.[8] The initial IPT phase was conducted individually. The interpersonal area for almost of all subjects was formulated as 'interpersonal deficits'. At termination, binge eating decreased in the IPT-G and CBT groups, but not in the control condition. Results persisted at 1-year follow-up. A randomized clinical trial of 162 women, comparing group IPT, and CBT for 20 sessions over 20 weeks, yielded similar results.

A research group in Christchurch, New Zealand studied the application of IPT to anorexia nervosa.[19] In their trial, neither IPT nor CBT showed efficacy as an outpatient treatment, consistent with the general anorexia outcome literature.

(b) Anxiety disorders

IPT has not yet been tested in controlled studies for anxiety disorders. Promising results have been found in for social anxiety disorders, PTSD, and panic disorder.[22, 31–33] Several groups are currently conducting controlled trials for these disorders.

(c) Substance abuse

IPT has failed to demonstrate efficacy in three clinical trials for patients with substance dependence.[20, 34, 35] These negative studies suggest limits to the range of utility of IPT as a main treatment for substance use disorders, but do not necessarily preclude its use to treat MDD comorbidity in those patients.

(d) Other applications

Research groups are testing the applicability of IPT to body dysmorphic disorder, chronic somatization in primary care patients, depressed patients postmyocardial infarction, depressed cancer patients, borderline personality disorder, insomnia, and other disorders. The IPT focus on life events suggests its potential applicability to patients with medical illness.

(e) IPT by telephone

Because many patients avoid or have difficulty reaching an office for face-to-face treatment, IPT and IPC are being tested as a treatment delivered over the telephone. Weissman and Miller conducted a successful pilot feasibility trial comparing IPT by telephone to wait-list control in 30 patients with recurrent major depression, and found IPT to be the superior control condition in reducing depressive symptoms and improving psychosocial functioning.[36] Neugebauer and colleagues found telephone IPC a helpful intervention for women with subsyndromal depression following a miscarriage.[37, 38]

(f) Interpersonal counselling (IPC)

Many patients presenting for treatment, particularly outside mental health settings, report psychiatric symptoms but do not meet threshold criteria for a psychiatric disorder. Nonetheless, their symptoms can be debilitating, interfering with their daily functioning, and often result in increased use of general medical services. Interpersonal counselling (IPC), based on IPT, was designed to treat distressed primary care patients who do not meet full syndromal criteria for psychiatric disorders. IPC is administered for a maximum of six sessions by health care usually by professionals who lack formal psychiatric training such as nurse practitioners. The first session can last up to 30 min; subsequent sessions are briefer.

IPC therapists assess the patient's current functioning, recent life events, occupational and familial stressors, and changes in interpersonal relationships. They assume that such events provide the context in which emotional and bodily symptoms occur. Klerman and colleagues studied 128 patients in a primary care clinic who scored 6 or higher on the Goldberg General Health Questionnaire (GHQ), randomizing them to IPC or to usual care without psychological treatment.[39] Over an average of 3 months, often receiving only one or two IPC sessions, IPC subjects showed significantly greater symptom relief on the GHQ than controls, especially mood improvement. IPC subjects were more likely to subsequently make use of mental health services, suggesting a new awareness of the psychological aspect of their symptoms.

Predictors of response to IPT

Five studies have examined predictors of response to IPT. Analyses of the TDCRP data identified general predictors of response to MDD treatment, as well as predictors for specific treatment modalities.[40,41] Seven patient characteristics predicted outcome across treatments: social dysfunction (higher social function predicted better response), cognitive dysfunction (better cognitive function predicted better response, particularly to CBT), expectation of improvement (higher expectation predicted better response), therapeutic alliance (a stronger alliance predicted better response), endogeneity of depression (endogenous depression tended to have better response, a finding supported by another study examining the relationship between EEG patterns and response to IPT in depressed patients[42]), double depression (its presence predicted poorer outcome), personality traits (their presence predicted worse response), and duration of current episode (longer duration was associated to worse response). In addition prior social adjustment, as measured by previous attainment of a marital relationship and higher satisfaction with social relationships in general, differentially predicted good response to IPT. This finding is consistent with reports in the general psychotherapy literature documenting that various indicators higher baseline psychosocial functioning predict good psychotherapy response. Two other studies suggest that comorbidity tends to worsen the prognosis of treatment with IPT alone, but not the prognosis of combined treatment.[43,44]

Summary of research findings

IPT has demonstrated efficacy as an acute and maintenance monotherapy and as a component of combined treatment for major depressive disorder. It also appears to have utility for other mood and non-mood syndromes, although the evidence for these is sparser. It has not shown benefit for substance use disorders or as a monotherapy for dysthymic disorder. Since monotherapy with either IPT or pharmacotherapy is likely to suffice for most patients with major depressive disorder, combined treatment is probably best reserved for severely or chronically ill patients. How best to combine time-limited psychotherapy with pharmacotherapy—for which patients, in what sequence, etc.—is an exciting area for future research.

Training

Until recently, IPT therapists were few, and practiced almost exclusively in research studies. Publications supporting its efficacy have led to clinical demand for this empirically supported treatment. IPT training is now increasingly included in professional workshops and conferences, with training courses conducted at University centres in Canada, the United Kingdom, continental Europe, Asia, New Zealand, and Australia. IPT is taught in a small but growing minority of psychiatric residency training programmes in the United States and as well as some family practice and primary care training programmes.

Although the principles and practice of IPT are relatively straightforward, any psychotherapy requires innate therapeutic ability, comfort with the so-called common factors of psychotherapy: tolerating and exploring affect, helping the patient to feel understood, engendering hope, etc. IPT training requires more than reading the manual: psychotherapy is learned by doing. Most IPT training programmes are designed to help experienced therapists refocus their treatment by learning new techniques, not to teach novices psychotherapy. Candidates should have a graduate clinical degree (MD, Ph.D., MSW, RN), several years of experience conducting psychotherapy, and clinical familiarity with the diagnosis of patients they plan to treat.

The IPT training in the TDCRP became the model for subsequent research studies. It included a brief didactic programme, reading the manual, and a practicum in which the therapist treated 2–3 patients under close supervision monitored by videotapes of the sessions. For research certification, we continue to recommend at least two or three successfully treated cases with hour for hour supervision of taped sessions.

Although many clinicians would like a formal certificate or diploma in IPT, there is no gold standard of IPT proficiency and no accrediting board. When IPT practice was limited to research settings, this posed no problem: one research group taught another, in the manner described above. As IPT spreads in clinical practice, the educational and accreditation process for IPT requires further study. The newly created International Society for Interpersonal Psychotherapy (ISIPT) may provide an appropriate forum in which to discuss these increasingly important issues.

Future directions

The history of IPT has been a succession of outcome trials. These studies have helped to define diagnostic indications for this treatment, but we know far less about the dosage and indications of IPT than about antidepressant medication. Future outcome trials may continue to define the scope of efficacy (response to treatment under ideal conditions) and effectiveness (response to treatment in more general clinical settings) of IPT. These should include both tests for different diagnoses, such as the anxiety disorders, testing of dosage—optimal frequency and duration of IPT sessions—and also studies of the sequencing of IPT with other treatments. Other research may help to determine the cost-effectiveness and potential cost-offset of IPT as a treatment that improves both symptoms and social functioning.

Most of the work on IPT to date has focused on treatment outcome. By contrast, little is known about the process aspects in IPT such as the specific value of many IPT interventions. Although it appears that solving an interpersonal problem area correlates with treatment outcome, it is unclear, for example, whether the choice of a particular treatment focus over other makes a difference for patients, or whether particular sorts of life events are helpful or unhelpful foci. Patient and therapist characteristics may also potentially influence treatment outcome.

Finally, while the initial work on IPT was conducted in the United States, over the last few years, IPT trials have also been conducted in other countries. As those studies continue to be conducted, it will become easier to discern to what extent IPT addresses topics are universal across cultures.

In summary, IPT is a time-limited, forward-looking, pragmatically focused psychotherapy that defines psychiatric disorders as treatable medical illnesses and links them to the patient's current social situation. This strategy has proved efficacious for patients with major depression and bulimia, and shows promise for other mood and non-mood disorders.

Further information

Elkin, I., Shea, M.T., Watkins, J.T., *et al.* (1989). National Institute of Mental Health treatment of depression collaborative research program: general effectiveness of treatments. *Archives of General Psychiatry*, **46**, 971–82.

Weissman, M.M., Markowitz, J.C., and Klerman, G.L. (2000). *Comprehensive guide to interpersonal psychotherapy*. Basic Books, New York.

Weissman, M.M., Markowitz, J.C., and Klerman, G.L. (2007). *Clinician's quick guide to interpersonal psychotherapy*. Oxford University Press, New York.

Website of the International Society of Interpersonal Psychotherapy: http://www.interpersonalpsychotherapy.org/

References

1. Fairburn, C.G., Jones, R., Peveler, R.C., *et al.* (1993). Psychotherapy and bulimia nervosa. Longer-term effects of interpersonal psychotherapy, behavior therapy, and cognitive behavior therapy. *Archives of General Psychiatry*, **50**(6), 419–28.
2. Elkin, I., Shea, M.T., Watkins, J.T., *et al.* (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program. General effectiveness of treatments. *Archives of General Psychiatry*, **46**(11), 971–82; discussion 983.
3. Klerman, G.L. and Weissmann, M.M. (1987). Interpersonal psychotherapy (IPT) and drugs in the treatment of depression. *Pharmacopsychiatry*, **20**(1), 3–7.
4. Frank, E., Kupfer, D.J., Wagner, E.F., *et al.* (1991). Efficacy of interpersonal psychotherapy as a maintenance treatment for recurrent depression: contributing factors. *Archives of General Psychiatry*, **48**, 1053–9.
5. Frank, E., Kupfer, D.J., Perel, J.M., *et al.* (1990). Three-year outcomes for maintenance therapies in recurrent depression. *Archives of General Psychiatry*, **47**(12), 1093–9.
6. Frank, E., Kupfer, D.J., Thase, M.E., *et al.* (2005). Two-year outcomes for interpersonal and social rhythm therapy in individuals with bipolar I disorder. *Archives of General Psychiatry*, **62**(9), 996–1004.
7. Reynolds, C.F. III, Frank, E., Perel, J.M., *et al.* (1999). Nortriptyline and interpersonal psychotherapy as maintenance therapies for recurrent major depression: a randomized controlled trial in patients older than 59 years. *The Journal of the American Medical Association*, **281**(1), 39–45.
8. Wilfley, D.E., Welch, R.R., Stein, R.L., *et al.* (2002). A randomized comparison of group cognitive-behavioral therapy and group interpersonal psychotherapy for the treatment of overweight individuals with binge-eating disorder. *Archives of General Psychiatry*, **59**(8), 713–21.

9. Agras, W.S., Walsh, B.T., Fairburn, C.G., *et al.* (2000). A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa. *Archives of General Psychiatry,* **57**(5), 459–66.
10. Mufson, L., Dorta, K.P., Wickramaratne, P., *et al.* (2004). A randomized effectiveness trial of interpersonal psychotherapy for depressed adolescents. *Archives of General Psychiatry*, **61**, 577–84.
11. Mufson, L., Weissman, M.M., Moreau, D., *et al.* (1999). Efficacy of interpersonal psychotherapy for depressed adolescents. *Archives of General Psychiatry*, **56**(6), 573–9.
12. Frank, E., Cyranowski, J.M., Rucci, P., *et al.* (2002). Clinical significance of lifetime panic spectrum symptoms in the treatment of patients with bipolar I disorder. *Archives of General Psychiatry*, **59**(10), 905–11.
13. Markowitz, J.C. (1994). Psychotherapy of dysthymia. *American Journal of Psychiatry*, **151**, 1114–21.
14. Markowitz, J.C., Kocsis, J.H., Bleiberg, K.L., *et al.* (2005). A comparative trial of psychotherapy and pharmacotherapy for "pure" dysthymic patients. *Journal of Affective Disorders*, **89**(1–3), 167–75.
15. Markowitz, J.C., Svartberg, M., and Swartz, H.A. (1998). Treatment of HIV-positive patients with depressive symptoms. *Archives of General Psychiatry*, **55**, 452–7.
16. Spinelli, M.G. (1997). Interpersonal psychotherapy for depressed antepartum women: a pilot study. *The American Journal of Psychiatry*, **154**(7), 1028–30.
17. Spinelli, M.G. and Endicott, J. (2003). Controlled clinical trial of interpersonal psychotherapy versus parenting education program for depressed pregnant women. *The American Journal of Psychiatry*, **160**(3), 555–62.
18. Schulberg, H.C., Block, M.R., Madonia, M.J., *et al.* (1996). Treating major depression in primary care practice. Eight-month clinical outcomes. *Archives of General Psychiatry*, **53**(10), 913–9.
19. McIntosh, V.V., Jordan, J., Carter, F.A., *et al.* (2005). Three psychotherapies for anorexia nervosa: a randomized, controlled trial. *The American Journal of Psychiatry*, **162**(4), 741–7.
20. Rounsaville, B.J., Glazer, W., Wilber, C.H., *et al.* (1983). Short-term interpersonal psychotherapy in methadone-maintained opiate addicts. *Archives of General Psychiatry*, **40**(6), 629–36.
21. Lipsitz, J.D., Fyer, A.J., Markowitz, J.C., *et al.* (1999). Open trial of interpersonal psychotherapy for the treatment of social phobia. *The American Journal of Psychiatry*, **156**(11), 1814–16.
22. Bleiberg, K.L. and Markowitz, J.C. (2005). A pilot study of interpersonal psychotherapy for posttraumatic stress disorder. *The American Journal of Psychiatry*, **162**(1), 181–3.
23. Markowitz, J.C., Svartberg, M., and Swartz, H.A. (1998). Is IPT time-limited psychodynamic psychotherapy? *Journal of Psychother Practice and Research*, **7**(3), 185–95.
24. DiMascio, A., Weissman, M.M., Prusoff, B.A., *et al.* (1979). Differential symptom reduction by drugs and psychotherapy in acute depression. *Archives of General Psychiatry*, **36**(13), 1450–6.
25. Weissman, M.M. (1979). The psychological treatment of depression. Evidence for the efficacy of psychotherapy alone, in comparison with, and in combination with pharmacotherapy. *Archives of General Psychiatry*, **36**(11), 1261–9.
26. Shea, M.T., Elkin, I., Imber, S.D., *et al.* (1992). Course of depressive symptoms over follow-up. Findings from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Archives of General Psychiatry*, **49**(10), 782–7.
27. Reynolds, C.F. III, Frank. E., Kupfer, D.J., *et al.* (1996). Treatment outcome in recurrent major depression: a post hoc comparison of elderly ("young old") and midlife patients. *The American Journal of Psychiatry*, **153**(10), 1288–92.
28. Frank, E., Kupfer, D.J., Buysse, D.J., *et al.* (2007). Randomized trial of weekly, twice-monthly, and monthly interpersonal psychotherapy as maintenance treatment for women with recurrent depression. *The American Journal of Psychiatry*, **164**(5), 761–7.
29. De Mello, M.F., Myczcowisk, L.M., and Menezes, P.R. (2001). A randomized controlled trial comparing moclobemide and moclobemide plus interpersonal psychotherapy in the treatment of dysthymic disorder. *Journal of Psychother Practice Research*, **10**(2), 117–23.
30. Browne, G., Steiner, M., Roberts, J., *et al.* (2002). Sertraline and/or interpersonal psychotherapy for patients with dysthymic disorder in primary care: 6-month comparison with longitudinal 2-year follow-up of effectiveness and costs. *Journal of Affective Disorders*, **68**(2–3), 317–30.
31. Lipsitz, J.D., Gur, M., Miller, N.L., *et al.* (2006). An open pilot study of interpersonal psychotherapy for panic disorder (IPT-PD). *The Journal of Nervous and Mental Disease*, **194**(6), 440–5.
32. Lipsitz, J.D., Gur, M., Vermes, D., *et al.* (2007). A randomized trial of interpersonal therapy versus supportive therapy for social anxiety disorder. *Depression and Anxiety*, **25**(6), 542–53
33. Lipsitz, J.D., Fyer, A.J., Markowitz, J.C, (1999). Open trial of interpersonal psychotherapy for the treatment of social phobia. *The American Journal of Psychiatry*, **156**(11), 1814–6.
34. Carroll, K.M., Fenton, L.R., Ball, S.A., *et al.* (2004). Efficacy of disulfiram and cognitive behavior therapy in cocaine-dependent outpatients: a randomized placebo-controlled trial. *Archives of General Psychiatry*, **61**(3), 264–72.
35. Carroll, K.M., Rounsaville, B.J., and Gawin, F.H. (1991). A comparative trial of psychotherapies for ambulatory cocaine abusers: relapse prevention and interpersonal psychotherapy. *The American Journal of Drug Alcohol Abuse*, **17**(3), 229–47.
36. Miller, L. and Weissman, M. (2002). Interpersonal psychotherapy delivered over the telephone to recurrent depressives. A pilot study. *Depression and Anxiety*, **16**(3), 114–7.
37. Neugebauer, R., Kilne, J., Bleiberg, K., *et al.* (2007). Preliminary open trial of interpersonal counseling for subsyndromal depression following miscarriage. *Depression and Anxiety*, **24**(3), 219–22.
38. Neugebauer, R., Kline, J., Markowitz, J.C., *et al.* (2006). Pilot randomized controlled trial of interpersonal counseling for subsyndromal depression following miscarriage. *The Journal of Clinical Psychiatry*, **67**(8), 1299–304.
39. Klerman, G.L., Budman, S., Berwisk, D., *et al.* (1987). Efficacy of a brief psychosocial intervention for symptoms of stress and distress among patients in primary care. *Medical Care*, **25**(11), 1078–88.
40. Sotsky, S.M., Glasss, D.J., Shear, M.T., *et al.* (1991). Patient predictors of response to psychotherapy and pharmacotherapy: findings in the NIMH Treatment of Depression Collaborative Research Program. *The American Journal of Psychiatry*, **148**(8), 997–1008.
41. Barber, J.P. and Muenz, L.R. (1996). The role of avoidance and obsessiveness in matching patients to cognitive and interpersonal psychotherapy: empirical findings from the treatment for depression collaborative research program. *Journal of Consulting & Clinical Psychology*, **64**(5), 951–8.
42. Thase, M.E., Buysse, D.J., Frank, E., *et al.* (1997). Which depressed patients will respond to interpersonal psychotherapy? The role of abnormal EEG sleep profiles. *The American Journal of Psychiatry*, **154**(4), 502–9.
43. Frank, E., Shear, M.K., Rucci, P., *et al.* (2000). Influence of panic-agoraphobic spectrum symptoms on treatment response in patients with recurrent major depression. *The American Journal of Psychiatry*, **157**(7), 1101–7.
44. Brown, C., Suhulberg, H.C., Madonia, M.J., *et al.* (1996). Treatment outcomes for primary care patients with major depression and lifetime anxiety disorders. *The American Journal of Psychiatry*, **153**(10), 1293–300.

6.3.4 Brief individual psychodynamic psychotherapy

Amy M. Ursano and Robert J. Ursano

Introduction

Interest in brief dynamic psychotherapy has flourished in recent years. The psychodynamic psychotherapies, including brief psychodynamic psychotherapy, aim to change behaviour through new understanding and the recognition of maladaptive patterns of behaviour enacted since childhood but not previously observed. Through this process, perceptions, expectations, beliefs, and, therefore, behaviours and feelings are altered.[1]

Historically, 'brief psychotherapy' and 'long-term psychotherapy' were used synonymously with 'supportive' and 'explorative' psychotherapy, respectively. However, brief and long-term describe only the duration rather than the technique, focus, or goal of treatment.[2] The time limits of brief dynamic psychotherapy give it a unique character and distinguish it from long-term psychotherapy and psychoanalysis. Because of its limited goals, the brief dynamic psychotherapist must confront his or her ambitiousness and perfectionism as well as any exaggerated ideal of personality structure and function.

Psychotherapy in general, and brief individual psychodynamic psychotherapy in particular, is perhaps the most elegant form of micro-neurosurgery. Psychotherapy strives to alter behaviour (i.e. cognitions, affects, and actions) with verbal interchange—fundamentally to change neurone A that used to connect to neurone B so it will now connect to neurone C. Although the therapist in the individual psychodynamically derived psychotherapies does not 'require' behavioural change, the end result of the therapist's technical expertise is to achieve behavioural change, including changes in well-being, physical health, social supports, and societal productivity as well as symptomatic relief. As in all of medicine, both non-specific and specific curative factors affect the outcome of this work. The non-specific curative factors—abreaction, the provision of new information, and maximizing success experiences—are present in all forms of medical treatment including brief psychotherapy. Brief individual dynamic psychotherapy also has specific technical interventions and procedures above and beyond the non-specific curative factors. As in other medical therapies, there are contraindications and dangers in the use of this treatment.

Background

Evolving from psychoanalysis in the mid-twentieth century, brief individual psychodynamic psychotherapy, like other psychodynamic treatments, is based on the principle that meanings and past experience play an important role in behaviour and illness. Although psychoanalysis is now a lengthy procedure usually requiring a number of years to complete, the early psychoanalytic literature, including Freud's first cases, contain histories of successful short analyses. During the first 30 years of psychoanalysis, it was unusual for treatments to extend beyond 1 year.[3] Ferenczi was the first analyst to advocate shortening psychoanalysis. He advocated 'active therapy' a more directive, focused, and briefer treatment. Rank was the first one to explicitly to set a time limit on treatment. Ferenczi and Rank[4] articulated the advantages of brief dynamic psychotherapy.

Following the Second World War, the interest in psychoanalysis resulted in greater demand for psychotherapy and increased pressure to develop briefer treatments. In the mid-1940s, Alexander and French advocated shortening treatment by decreasing the frequency of sessions in order to minimize regression. They proposed to focus treatment on the present rather than the past, using historical conflicts to inform the therapist in providing the best corrective emotional experience for the patient in the present.

The community-based mental health treatment movement, the increasing cost of mental health care, and the rise of managed care in the United States; have stimulated efforts to find briefer forms of psychotherapy. Contemporary brief individual psychodynamic psychotherapy is heavily influenced by the British School's development of brief focal psychotherapy. Balint sponsored a workshop of experienced psychoanalytic psychotherapists, which focused on clinical evaluation and attempted to understand which patients might be suitable for briefer treatment. After Balint's death, Malan carried on the work of the group. At the Tavistock clinic, Malan developed and applied the principles of psychodynamic treatment to brief treatment, delineating methods for evaluating process and outcome variables. He emphasized the importance of therapeutic planning and the identification of a focal conflict.

Concurrently, Sifneos, at the Massachusetts General Hospital, was studying brief psychotherapy.[5,6] Sifneos developed 'short-term anxiety-provoking psychotherapy' as a technique and theory with strict inclusion and exclusion criteria for choosing patients. Davanloo broadened the focus of the brief psychodynamic psychotherapies to include more than one conflict. He also expanded the inclusion criteria to individuals with character pathology and chronic phobic and obsessional neuroses, and advocated actively confronting resistances. Mann's time-limited psychotherapy identified a central issue related to the meaning of time, as the focus of the treatment. Mann related this to the patient's difficulties in confronting loss and separation and the reality of time and death.

In recent years, brief psychotherapy has become increasingly research based. Strupp, Luborsky, and Horowitz have all introduced manualized focused psychodynamic treatments which substantially contribute to our research understanding of this treatment modality.

Brief dynamic psychotherapy technique

(a) Evaluation and setting

The evaluation is particularly important in brief individual psychodynamic psychotherapy because of the need for rapid and accurate assessment. In contrast to longer term treatments, brief individual psychodynamic psychotherapy does not offer the luxury of time to re-evaluate and correct mistakes. Although at times we think of psychotherapy as beginning as soon as the doctor sees the patient, this is a hyperbole, used to underscore the importance of interpersonal and transferential elements in the initial meeting with the patient. In fact, it is extremely important, particularly in brief individual psychodynamic psychotherapy, to distinguish the diagnostic interviews from the ongoing treatment.

The interventions and technical procedures performed during the evaluation phase, usually one to four sessions, are substantially different from the technical aspects of brief individual psychodynamic psychotherapy itself. The evaluation phase includes the diagnosis, consideration of the interaction among the patient's ego strength, physical health, and selection variables, and the treatment recommendation, including considering the option that no treatment is indicated.

As in all medical treatments, brief individual psychodynamic psychotherapy is given to patients rather than to diseases. The ability to participate in brief individual psychodynamic psychotherapy process requires the patient to be able to access his or her fantasy life in an active and experiencing manner (i.e. psychologically minded) and, importantly, is able to get up and leave this process behind at the end of a session and not be lost in reverie or uncontrolled fantasies or fears. Note that this does not mean the patient requires a 'high IQ'. In fact, a high IQ, when accompanied with rigidity, intellectualization, and rumination, as is often seen, can be a contraindication to a brief psychodynamic treatment since these defences can be quite formidable. The availability of interpersonal support in the patient's real environment and the patient's ability to experience and simultaneously observe highly charged affective states are necessary to a successful treatment. Individuals who are in an emergent crisis (e.g. imminently suicidal, psychotic, recent major life trauma) and therefore are very concerned and focused on the real events in their life cannot enter into a brief psychodynamic psychotherapy without first having a period of supportive treatment. A true life crisis does not allow the patient the opportunity to explore fantasies.

Negotiation with the patient is an important part of reaching a treatment decision in brief individual psychodynamic psychotherapy. The patient must rapidly feel a part of the treatment and committed to the process. The process of setting a time limit at the beginning of the treatment can be an important element in decreasing the dropout rate from this form of treatment,[7] particularly with the patient who is concerned about dependency, 'becoming addicted' to the therapist, or who needs to maintain a substantial sense of control. What is dealt with in treatment can only be what the patient is able to bring into focus, what the patient can tolerate talking about, and what he or she can tolerate the therapist talking about.[8] Although this is not different than other psychodynamic treatments, the limited time of brief individual psychodynamic psychotherapy means that there is limited ability to interpret multiple defences that might open new areas of exploration.

(b) Technique

The rapid establishment of the therapeutic alliance is critical to brief individual psychodynamic psychotherapy.[9] Identifying the patient's initial anxieties related to beginning therapy is an important technique in the early sessions of brief individual psychodynamic psychotherapy in order to assure the alliance and to establish the conditions under which the patient can favourably hear and respond to the interpretations that the therapist will later give. As the therapy unfolds, the therapist operates on the hypothesis that each session is related to the previous one. The therapist strives in each session to identify the continuity of meaning related to the treatment focus that is present but hidden.[10] This continuity is driven by the 'experience bias' of the patient, and his or her tendency to experience the world in a certain way due to unique developmental experiences that have moulded his or her perception, interpersonal beliefs, and expectations.[11]

Brief individual psychodynamic psychotherapy is more focused, and more 'here and now' oriented with fewer attempts to reconstruct the developmental origins of conflicts than the extensive reworking of personality undertaken in longer term psychotherapies. Through the exploration of the patient's metaphors and symbols, both defensive patterns and disturbances in present interpersonal relations are identified in the treatment setting as well as in the patient's life. The importance of being able to hear what the patient has to say and to understand its meaning remains central as in other psychoanalytically oriented treatments.

Free association and inquiry: Free association is part of the technique of brief individual psychodynamic psychotherapy. But what constitutes free association—as in all dynamic therapies—requires thoughtful consideration. In its most basic form, and particularly highlighted in brief individual psychodynamic psychotherapy, free association means that the patient is free to choose what they wish to talk about. This rather direct definition emphasizes that free association is always relative. In addition, in brief dynamic psychotherapy, the patient is always somewhat more task focused that in open-ended treatments or psychoanalysis and this focus should not be discouraged by the therapist. Rather it is the therapist's task to hear the themes in the patient's concerns. The therapist asks questions, directs the patient's attention, and uses benign neglect, i.e. avoids some areas of conflict that cannot be dealt with at this time or in a short period of time. The therapist identifies those spots at which free association breaks down (the presence of a defence) or at which the narrative is carrying a single emotional story out of the patient's awareness. As in all dynamic treatments, often when the patient is able to talk freely and with a coherent narrative about their conflicts, the work of the treatment is completed.

Defence and transference: Brief individual psychodynamic psychotherapy emphasizes understanding (a) the mechanisms of defence used by the patient to decrease anxiety and other uncomfortable feelings associated with areas of conflict which are out of awareness, and (b) the characteristic transference relationships which distort the patients response to their adult world. Typically these two areas, defence and transference, create the world of meaning and expectations in which the patient lives. The techniques of the brief psychodynamic psychotherapy are directed towards clarifying these areas and presenting them to the patient to increase understanding and in this manner change symptoms and behaviour. Often only one defence is concentrated on in a given brief treatment. As the defence is clarified, the transference relationship may become evident. The developmental narrative of how the patient came to see the world in the way he or she does, provides the 'glue' through which the patient can integrate this knowledge into their life experience and behaviours, and recall it for practice and future use.

The brief individual psychodynamic psychotherapy therapist, similar to longer term psychodynamic work, must both enhance the patient's observing capacity in order that the transference can be observed by the patient and therapist, and create the therapeutic situation in which the patient can hear the therapist's interpretations in a useful manner. Dreams, as well as slips of the tongue and symptoms, can provide an avenue to the understanding of unconscious conflict which can be taught and explored with the

patient. The therapist strives to interpret both the triangle of anxiety (wish-defence-anxiety) and the triangle of insight (transference figure in the present—the therapist/patient interaction—transference figure from the past).

Frequently, when the transference is most evident, other elements of the past are simultaneously experienced in the patient's life. In brief individual psychodynamic psychotherapy these can be particularly important to the patient's understanding the feeling elements of the transference in a mutative manner since the depth and intensity of the transference is much less and much briefer than in long-term work. In contrast, however, the presence of a recent precipitant to the patient's problems, as is usually the case in brief psychodynamic psychotherapy, can considerably intensify transference responses and be a central element in developing the psychodynamic understanding for the patient. The transference experience—the transference, the life experiences being relived, and particularly the precipitant—provide the web of meaning that is the focus of interpretation and the mutative force in brief individual psychodynamic psychotherapy.

Often the transference in brief individual psychodynamic psychotherapy is paternal or maternal, but it has also been noted that, perhaps due to the time-limited nature of the work, sibling and transference figures from adolescence may more often be recalled in brief individual psychodynamic psychotherapy. The transference is rarely as deep as that seen in long-term treatment. It requires a skilled eye to note and bring the transference to the attention of the patient in a manner that is neither intrusive nor offensive.[12] Interpretations usually occur over several sessions, in the middle or later third of the treatment, during which past, present, and transference experiences are linked together. In the context of the affective arousal associated with this transference experience and the simultaneous understanding of the experience, behavioural change occurs and the patient's ability to perceive previously hidden feelings and relationships as well as his or her view of the future and the past can change.

Countertransference: Countertransference is also an important element in brief individual psychodynamic psychotherapy as in other psychodynamic treatments.[13] Analysis of countertransference reactions can allow the therapist to recognize subtle aspects of the transference relationship and to understand the patient's experience better. Because of the more active stance, the brief psychodynamic psychotherapist can be particularly prone to countertransferences that show up as over-involvement or aggression. In addition, the brief time available for treatment can make recovery from countertransference errors quite difficult.

(c) Medication

Medication is frequently used in conjunction with brief psychodynamic psychotherapy. This can complicate the treatment and its progress as well as aid in symptom recovery. The therapist must explore the meaning of the medication and its role in the patient's view of himself or herself and interpersonal strengths and vulnerabilities. At times, brief individual psychodynamic psychotherapy can also serve as an alternative to medication treatment for less severe symptoms or when medication is contraindicated. Medication may have also begun during the initial brief psychodynamic psychotherapy and then continued after the psychotherapy has formally stopped and the patient is followed with less frequent meetings to monitor medication. This sequence has many advantages including resolving present stressors and precipitants, encouraging medication compliance, and ongoing medical follow-up after therapy either in maintenance or intermittent frequency. Another course of brief dynamic therapy may be indicated at a later date if the response to combined treatment is ineffective or if new problems appear. Greater education of clinicians and research on this combined and sequential treatment is needed.

Comparison of the brief psychodynamic psychotherapies

The work of Malan, Sifneos, Mann, and Davanloo shows substantial overlap in each author's goals, selection criteria, technique, and duration of treatment.[14] The goals of all of these models of brief psychotherapy include facilitating health-seeking behaviours and mitigating obstacles to normal growth. From this perspective, brief psychotherapy focuses on the patient's continuous development throughout adult life and the context-dependent appearance of conflict, depending on environment, interpersonal relationships, biological health, and developmental stage. This picture of brief psychotherapy supports modest goals that require the therapist to refrain from perfectionism. Malan, Sifneos, Mann, and Davanloo also seem to agree with Stierlin's[15] contrast between brief psychotherapy's use of the 'propitious moment' and long-term treatment's use of 'a shared past' between therapist and patient. Both the propitious moment and the shared past carry psychotherapeutic advantages and disadvantages, emphasizing certain technical possibilities and limiting others.

Selection criteria: Many of the selection criteria emphasized by Malan, Sifneos, Mann, and Davanloo are common to all kinds of psychodynamic psychotherapy. However, unique selection criteria are required due to the brief duration of treatment. Patients in brief psychodynamic psychotherapy must be able to engage quickly with the therapist, terminate in a short period of time, and be able to carry on much of the working through and generalizing of the treatment effects on their own.

The necessity for greater independent action by the patient requires that the patient have high levels of ego strength, motivation, and responsiveness to interpretation. Sifneos's rather unique emphasis on intelligence as a criterion may be related to his anxiety-provoking interpretations, which require a broader educational context in order to be understood. The importance of the rapid establishment of the therapeutic alliance underlies a substantial number of the selection and exclusion criteria.

Focus of brief psychotherapy: All authors mention the central importance of the focus in brief psychotherapy, and therefore the evaluation sessions to determine this focus. Mann formulates the focus to the patient in terms of the patient's fears and pain. However, he would probably agree with Malan, Davanloo, and Sifneos in the importance of constructing the psychodynamic focus at a deeper level in one's own understanding of the work being done. Maintaining the focus is the primary task of the therapist. This enables the therapist to deal with complicated personality structures in a brief period of time. Resistance is limited through benign neglect of potentially troublesome but non-focal areas of the personality. The elaboration of techniques for establishing and maintaining the focus of treatment is critical to all brief individual psychodynamic psychotherapies.

Transference: The manner and rapidity in which transference is dealt with vary considerably among proponents of brief individual psychodynamic psychotherapy. Malan takes a more typical psychoanalytic approach of waiting for transference to become resistance before it is interpreted. Sifneos, in his emphasis on the Oedipal relationship, is more aggressive in handling the deep conflictual areas of transference material. Davanloo is confrontational in developing a transference experience. This confrontational style may at times confuse the patient's experience of the real and the transferential therapist. However, Davanloo often treats severe obsessional disorders. In these cases, the need to increase the patient's affective awareness is high. These may be the patients in which this particular technique is most useful. Aggressive, competitive, and hostile feelings, which might otherwise remain firmly defended, may thus become available to these patients.

Countertransference: The role of countertransference in brief psychotherapy is as complicated as it is in long-term treatment. Countertransference issues related to the aggressive techniques used by Sifneos and Davanloo have been observed. Countertransference experiences related to termination and loss can also be prominent.[16] The goal-directed techniques of brief psychotherapy limit the development of regressive countertransference responses.[13]

Duration of treatment: There is remarkable agreement on the duration of brief psychotherapy. Although the duration ranges from 5 to 40 sessions, authors generally favour 10 to 20 sessions. The duration of treatment is critically related to maintaining the focus within the brief psychotherapy. Shlien *et al.*[17] have found in Rogerian therapy, a correlation between the number of sessions and recovery. In general, they report an increasingly successful outcome (measured by the patient's self-concept) up to about 20 sessions. Howard *et al.*[18] using a meta-analytical technique, found 75 per cent of patients showing some improvement by 26 sessions. However, this study includes a wide range of types of treatment. When treatment extends beyond 20 sessions, the therapist frequently may find himself or herself enmeshed in a broad character analysis without a focal conflict. Change after 20 sessions may be quite slow. Clinical experience generally supports the idea that brief individual psychodynamic psychotherapy should be between 10 and 20 sessions although more complicated cases will require greater length of treatment. Often extending treatment beyond 20 sessions is recognition that treatment will be beyond 40 or 50 sessions.

Brief psychodynamic psychotherapy for depression, narcissistic disturbances, panic disorder, substance abuse, and post-traumatic stress disorder have been described.[14,19] Horowitz *et al.*[20] have described brief psychotherapy focused on the stress responses evidenced by various personality styles. He emphasizes that this psychotherapy is directed towards dealing with the process of the stress response and not character change. However, his outcomes indicate that selected character changes are possible in some areas. The distinction between recovery from a disruption in homoeostatic balance, reconstitution of self-esteem and self-concept, and changes in character structure require further exploration.

Critical points: The identification of critical points during brief psychotherapy, when the 'danger' of becoming a long-term treatment is most acute, clarifies the technical handling of brief psychodynamic psychotherapy. At these points, the therapist often notes an increasing vagueness of the goals of the treatment, decreased activity by the therapist, and the emergence of the transference as the central element. These variables indicate the potential of a short-term psychotherapy becoming a long-term treatment. The fourth to sixth hour of weekly 12-session therapy is often a point at which incipient or potential regression may suddenly appear. The patient at this time is testing the boundaries of the treatment. Action by the therapist is required if a brief psychotherapy is to remain exactly that—brief. The study of technical interventions, which occur at these critical moments, will further elucidate the technical handling of limited regression in brief psychodynamic psychotherapy.

Malan and the Tavistock group: focal psychotherapy: Developed from the workshops of Balint and Malan, focal psychotherapy is an example of applied psychoanalysis.[21] Malan has carried on Balint's earlier work.[22,23] Previous attempts to develop brief forms of psychoanalytic psychotherapy primarily involved the use of 'activity' which was frequently equated with manipulation. On the contrary, Malan emphasized the importance of choosing and maintaining a narrow focal area to be dealt with in a brief period of time. He stresses the importance of finding the appropriate focus in the patient's story and consistently interpreting the focal problem area.[23] Through selective attention and neglect, the therapist maintains the focus and completes a brief psychotherapy. The importance of determining the focus underscores the value of the diagnostic process, including the psychodynamic assessment of the patient prior to the initiation of psychotherapy.[24]

Malan identifies the following factors as leading to the lengthening of treatment: resistance, overdetermination, a need for working through the roots of conflict in early childhood, transference, dependence, negative transference connected with termination, and the transference neurosis. In addition, some therapist characteristics may lengthen treatment. These include a tendency towards passivity, a sense of timelessness conveyed to the patient, therapeutic perfectionism, and a preoccupation with deeper earlier experiences. All of these factors must be dealt with in order to maintain a brief therapy. For Malan, identifying a focal conflict acceptable to the patient is critical to a successful outcome (Table 6.3.4.1). In addition, the patient must have the capacity to think in feeling terms, demonstrate a high motivation, and exhibit a good response to trial interpretations made during the evaluation phase. Patients who have had serious suicidal attempts, drug addiction, long-term

Table 6.3.4.1 Brief psychodynamic psychotherapies

Goal of treatment
Identify the defence, the anxiety, and the impulse
Link the present, the past, and the transference
Focus of treatment
Internal conflict present since childhood
Selection criteria
Patient is able to think in feeling terms
Highly motivated
Good response to trial interpretation
Duration of treatment
Up to 1 year
Mean 20 sessions
Termination
Set definite termination date at beginning of treatment

hospital stays, more than one course of electroconvulsive therapy, chronic alcoholism, incapacitating severe chronic obsessional symptoms, severe chronic phobic symptoms, or gross destructive or self-destructive acting-out are excluded from treatment. The patient is also excluded from focal psychotherapy if the therapist anticipates any of the items in Table 6.3.4.2.

For Malan, the criteria in Table 6.3.4.2 represent specific dangers. If the therapist cannot make contact with the patient, or low motivation or rigid defences are present, it will be difficult to form an effective therapeutic working alliance within a short time. Complex or deep-seated issues, which must be dealt with to resolve a conflict area, require a longer period of treatment. Difficult transference relationships may also prevent timely termination or lead to premature termination. The occurrence of severe depressive or psychotic episodes during treatment can be a danger to the patient and require adjunctive treatments. Thus, Malan takes seriously the time limitation in brief therapy, which requires the rapid establishment of a therapeutic alliance and the ability to terminate therapy without the development of unexpected serious symptoms.

Malan, in contrast with other practitioners, does not automatically exclude patients with serious psychopathology. He sees the balance between motivation and focality as the primary criteria. A patient with only moderate motivation but a highly focal conflict might be accepted into treatment. Similarly, a patient with high motivation but not as focal a conflict might also be accepted into treatment with the hope that clarification of the focus would occur in a short period of time.

Identifying the precipitating factors, early traumatic experiences, or repetitive patterns can indicate the area of internal conflict present since childhood and the possible focus of treatment. The therapist should assess the congruence between the current conflict and the 'nuclear' or childhood conflict during the evaluation phase. The patient's response to interpretations about aspects of this conflict may lead to acceptance into treatment. According to Malan, the greater the probability that the conflict area will manifest itself in the transference, the more positive the outcome will be.

Malan is less concerned with technique than with the importance of choosing the focus. He employs the usual technical procedures of psychoanalytic psychotherapy and emphasizes the importance of making interpretations of the transference and connecting these to current and past relationships. This 'triangle of insight' (the transference, the current relationship, and the past relationship) leads to the patient's cure. Overall, the goal is to clarify the nature of the defence, the anxiety, and the impulse, which the patient is experiencing, and to link these to the present, the past, and the transference. Once the defence and the anxiety are clarified, the link to the past can be made. The interpretation that links to the past may be experienced as reassuring by the patient because of its emphasis on the conflict belonging to the world of fantasy rather than to the world of the present. Malan emphasizes transference interpretations as the most therapeutically effective interpretations because of their 'here and now' character.

In the brief therapy unit at the Tavistock Clinic, a time limit was almost always given at the beginning of treatment. For trainees this was usually 30 sessions. However, in his publications, Malan indicates a mean of 20 sessions for those cases with favourable outcomes. The longer time for trainees gives the opportunity to correct mistakes that might occur. In some published cases, therapy was extended up to 1 year (46 sessions). In general, Malan advocates the importance of a definite date rather than a number of sessions. Practically speaking, this eliminates the need for the patient and therapist to keep count of the number of sessions and eliminates complications related to whether or not to make up sessions that the patient has missed. Such a time limit gives a definite beginning, middle, and end to the therapy. It helps to concentrate the patient's material and the therapist's work, to maintain the focus, and decrease the diffuseness that might lead into long-term work.

Sifneos: short-term anxiety-provoking psychotherapy: Sifneos emphasizes the importance of patient selection because of the anxiety-provoking nature of his brief psychotherapy techniques (Table 6.3.4.1). He distinguishes anxiety-provoking therapy from anxiety-suppressing therapy, commonly referred to as supportive psychotherapy. For short-term anxiety-provoking psychotherapy, the patient must be of above average intelligence and have had at least one meaningful relationship with another person during his or her lifetime. The patient who has had such a relationship will be able to withstand the anxiety produced by the therapy and to develop rapidly a mature collaborative relationship with the therapist. This criterion tends to exclude narcissistic disorders. In addition, the patient must be highly motivated for change, not only for symptom relief. Sifneos also identifies several criteria for the patient selection based on the presentation of the patient during the evaluation. The patient must have a specific chief complaint. If the patient has a number of complaints, Sifneos asks the patient which complaint is of top priority. The patient's ability to identify one conflict area and to postpone work on others is taken as an indication of the patient's ability to tolerate anxiety. Sifneos looks for patients with anxiety, depression, phobias, conversion, and mild obsessive-compulsive features or personality disorders involving clear-cut interpersonal difficulties. During the evaluation, the patient must show an ability to interact with the evaluating psychiatrist, to express feelings, and to show some flexibility.

Sifneos is one of the few authors who clarifies his assessment of motivation. He defines motivation as including the patient's ability to recognize symptoms as psychological, a tendency to be introspective and honest about emotional difficulties, and a willingness to participate in the treatment situation. In addition, motivation includes curiosity, willingness to change as well as a willingness to make reasonable sacrifices, and a realistic expectation of the results of psychotherapy.

Sifneos focuses on the Oedipal conflict and does not expect a good outcome in dealing with other than Oedipal conflict areas.

Table 6.3.4.2 Exclusion criteria for Malan's and the Tavistock group's brief focal psychotherapy

1. Therapist is unable to make affective contact with the patient during the evaluation
2. Therapist anticipates that extended work will be needed
   To generate motivation
   To decrease rigid defences
   To reach complex or deep-seated issues
   To resolve unfavourable, intense transference, or dependence which may develop
3. Depressive or psychotic disturbance may intensify and place the patient at risk

The majority of failures using short-term anxiety-provoking psychotherapy have occurred in patients who complained of reactive depression following the loss of a loved one. He believes that this failure is due to the non-triangular (non-Oedipal) origins of the ambivalent feelings in some patients. In such cases, when the issue of termination arises, the patient regresses and an impasse is reached.

During the initial phase of psychotherapy, the therapist must establish good rapport with the patient in order to create a therapeutic alliance. The therapist uses anxiety-provoking confrontations in order to clarify issues around the patient's early life situation and present-day conflict. The therapist avoids areas such as passivity, dependence, and acting-out, which might lead to extensive regression. The use of anxiety-provoking confrontations in a direct attack on the patient's defences distinguishes short-term anxiety-provoking psychotherapy from other brief psychotherapies. Although it is made clear to patients during their evaluation that the psychotherapy is expected to last only a few months, no specific number of sessions or termination date is given. Interviews are held weekly and last for 45 min. The vast majority of treatments last from 12 to 16 sessions, and none go beyond 20 sessions. The aggressive confrontational style of this treatment underscores the importance of excluding pre-Oedipal problems and the importance of countertransference reactions in the therapist related to being too aggressive.

Mann: time-limited psychotherapy: Mann has focused on the specific limitation of time in brief psychotherapy. Mann sees the variable of time as a specific operative factor in psychotherapy as well as an element in its curative effect.[25,26] The experiences of the timelessness of treatment and of the treatment's termination are significant elements in Mann's view of the psychotherapeutic process.

Usually there are two to four evaluation meetings prior to beginning psychotherapy. Mann limits psychotherapy to a total of 12 treatment hours, distributed according to patient need. This may result in weekly 30-min sessions for 24 weeks or twice weekly hour-long sessions for 6 weeks. In practice, however, nearly all patients are seen in once-weekly 45- or 50-min sessions for 12 weeks. Mann admits having chosen the number 12 somewhat arbitrarily; however, his clinical experience indicates that somewhere between 10 and 14 sessions is a sufficient number. Mann emphasizes the importance of a uniform number of sessions for evaluating the psychotherapeutic process among different therapists. In this way, the relationship between the patient's presenting problems and psychotherapeutic technique can be more easily studied. Also, the provision of a specific number of sessions can be more easily accepted by the patient as a typical medical 'prescription'. Finally, the setting of a specific last session in the initial contract with the patient allows the therapy to have a clear beginning, middle, and end (see Table 6.3.4.1).

Mann indicates a number of exclusionary criteria: serious depression, acute psychosis, borderline personality organization, and the inability to identify a central issue. Mann sees Sifneos' criteria as primarily excluding borderline patients. He does not agree with Sifneos' emphasis on superior academic or work performance.

To some extent, Mann initially minimized selection as a central issue for brief psychotherapy. Later, Mann expanded his selection criteria by emphasizing the importance of the patient's ego strength as measured by prior work performance and past relationships.[26] Patients who may have difficulty engaging and disengaging rapidly from treatment are excluded. This includes schizoid patients, certain obsessional patients, patients with strong dependency needs, some narcissistic patients, some depressive patients who will not be able to form a rapid therapeutic alliance, and some patients with psychosomatic disorders who do not tolerate loss well.

According to Mann, the selection of the central issue for the psychotherapy is the critical event. It is the vehicle through which the patient is engaged in the work of therapy and on which a successful outcome depends. Mann looks for a central issue that is developmentally and adaptively relevant and has been recurrent over time. He describes this issue as the patient's 'present and chronically endured pain' and characterizes it as preconscious. Mann has further described the central issue as including a particular image of the self.[25] The central issue formulated in terms of time, affect, and an image of the self is the 'paradigm of the transference' expected to emerge in treatment. The therapist's statement of the central issue is a clarification, which can be readily recognized, felt, and held onto by the patient. Time-limited psychotherapy is intended to resolve this present and chronically endured pain and the patient's 'negative self-image'. The therapist frames the central issue to the patient in terms of a general statement about feelings.

Mann and Goldman[26] described in detail the phrasing of the central issue to the patient. It is the central issue that specifies the therapeutic contract and the goal of the therapy. In the case of a 41-year-old depressed woman who was preoccupied with her husband and children being even a minute late, Mann suggested the central issue: 'You've encountered extreme life situations and have managed them remarkably well ... yet you fear and have always feared that despite your best efforts you will lose everything'. In a 31-year-old married man attempting to gain a college degree who was consumed with a fear of failing, Mann suggested the central issue: 'Because there have been a number of sudden and very painful events in your life, things always seem uncertain, and you are excessively nervous because you do not expect anything to go along well. Things are always uncertain for you'.[26]

Mann uses the usual psychoanalytic psychotherapy techniques: defence analysis, transference interpretation, and genetic reconstruction. Transference is interpreted from within the central identified conflict area and in terms of the adaptive processes of the patient. However, Mann does not confront the patient. In general, his interventions are very close to the conscious material provided by the patient. Mann identifies specific dynamic events that unfold during the 12 sessions. The opening sessions are understood as filled with the unconscious magical expectation that past pains will now be resolved. During the initial phase, the therapist makes few comments and accepts the positive transference of the patient. Important aspects of the current problem, defence mechanisms, coping styles, and genetic roots of the central issue become clearer during this phase. In the middle four sessions, resistance is likely to appear, as well as the negative transference. The patient experiences the frustration that all of the wished for changes may not occur. In the ending phase of treatment, termination and the patient's resistances to termination in the face of unresolved problems in other areas of life are prominent.

Mann sees the importance of confronting separation and termination issues as critical to the success of brief psychotherapy. Frequently, the patient unconsciously reveals an awareness that the mid-point of treatment has come. The patient experiences

separation from the transference-invested therapist as a separation from an ambivalently experienced person from the past, without having achieved the fantasized magical resolution. The goal is to enable the patient to separate from the transference-invested therapist less ambivalently than he had done from this earlier important figure. Consequently, both the resolution of the central issue and the unfolding of an attachment-separation process in the 12-session treatment contract are intimately related through the development and interpretation of the transference.

Davanloo: broad-focus short-term dynamic psychotherapy: Davanloo writes about broad-focus short-term dynamic psychotherapy.[27] His selection criteria include patients with an Oedipal focus, those with a loss focus, and those with multiple foci. Davanloo is particularly interested in patients suffering from long-standing obsessional and phobic neuroses. His research data indicate that 30 to 35 per cent of the psychiatric outpatient population can benefit from this mode of therapy. Most information about his technique is derived from the publication of cases, presentations, and brief descriptions of his research that accompany case presentations.

The initial evaluation is a specific focused interview in which the patient's defences against 'true' feelings are gently but consistently confronted. Davanloo says that this is not a universal technique for the initial interview and cautions on its use with patients with severe psychopathology. Selection is based on psychological mindedness, the quality of the patient's interpersonal relations, and, in particular, on the presence of at least one meaningful relationship in the patient's past. The patient's ability to tolerate and experience anxiety, guilt, and depression are important (Table 6.3.4.1). The patient must be motivated to complete the treatment process and to resolve neurotic problems. His or her ability to respond to interpretation is an important selection criterion. In particular, response to transference interpretations, which link the transference with the present and the past, is a critical feature in the assessment for broad-focus short-term dynamic psychotherapy. Davanloo finds no value in criteria based on severity and duration of illness. Finally, the presence of flexibility in the ego's defensive pattern and a lack of use of the primitive defences of projection, splitting, and denial are important factors in selecting patients.

The technique Davanloo uses in therapy is a continuation of that used in the initial interview. The emotional experience of the patient in the transference is emphasized. The patient is 'gently but relentlessly' confronted about his defences against feelings in the transference relationship and in the past. All the usual techniques of psychoanalytic psychotherapy are employed: defence analysis, transference interpretations, and genetic reconstruction. Dreams and fantasy materials are also used. Transference interpretations tend to be made early. Because of the confrontive style, a strong therapeutic alliance is necessary. Patients frequently experience hostile, angry feelings towards the therapist because of being confronted. Davanloo actively pursues the patient's defences against recognizing the anger and its transference elements. Davanloo warns therapists that passive dependent and obsessional characters may develop a symbiotic transference relationship. This may be avoided through active confrontation and selection of patients. The active confrontation of defences and early transference interpretations tend to mobilize powerful affects and memories early on in treatment.

Davanloo recommends from 5 to 40 sessions, depending on the patient's conflict area (Oedipal versus multiple foci) and other selection criteria. In general, his treatments fall between 15 and 25 sessions. He does not recommend setting a specific termination date but rather makes clear to the patient that treatment will be short. Shorter time periods (5–15 sessions) are chosen for patients with a predominantly Oedipal focus, longer durations (20–40 sessions) for the more seriously ill group.

Comparison of psychodynamic, cognitive, and interpersonal brief psychotherapies

Interpersonal psychotherapy[28,29] and cognitive behavioural psychotherapy[30] derive from the psychodynamic model and therefore share many common elements with brief psychodynamic psychotherapy but with distinct approaches and interventions. All three modalities, interpersonal psychotherapy, cognitive behavioural therapy, and brief individual psychodynamic psychotherapy, are complex methods of treatment that must be custom-tailored to the individual patient. Brief by definition, they all lack the extended working through and application period of psychoanalysis and intensive (long-term) psychodynamic psychotherapy. All demand a high degree of clinical judgement and considerable experience to acquire competency. The relationship between the therapist and patient and the establishment of a therapeutic alliance are essential (Table 6.3.4.3).

Table 6.3.4.3 Comparison of the brief dynamic psychotherapy with cognitive psychotherapy and interpersonal psychotherapy

	Brief dynamic psychotherapy	Cognitive psychotherapy	Interpersonal psychotherapy
Free association	++	+	+
Directiveness	+	+++	++
Neutrality	+++	+++	+++
Time-limited	+++	+++	+++
Defence analysis	+++	+++ Schema/distortions	+
Transference	+++ Interpersonal	+	+++ patterns
Behavioural interventions	—	+++	+
Published manuals	+	++	++
Concurrent use of medication	++	+++	+++
Empirical research indicates efficacious treatments	+	+++	+++
Training in long-term dynamic psychotherapy helpful	+++	+	+++

While sharing many similarities, it is ultimately in the conception of the problem, the goals, and therapeutic interventions that these treatments differ. It is unclear to what extent behavioural changes may be attributed to the similarities or differences between treatments. All psychotherapies, including brief individual psychodynamic psychotherapy, interpersonal psychotherapy, and cognitive behavioural therapy teach new skills-problem-solving skills directed at how to resolve interpersonal and emotional problems when they arise. Differences among these psychotherapies in their interventions are more striking than the differences in their goals or the problem areas they identify for therapeutic work. In psychodynamic psychotherapy the structure of the session is determined by the flow of the patient's thoughts and their interaction with the therapist's interpretive comments. In contrast, cognitive and interpersonal psychotherapies use more directive, structured, and behavioural interventions. Whereas the brief individual psychodynamic psychotherapy like other psychodynamic psychotherapies relies on the patient to activate and practice new behaviours without direction. The therapist remains an empathic interpreter, a sharer of the patient's experience and perspective. While in other therapies, especially cognitive, the therapist may direct, prescribe, enjoin, educate, or role play.

Practical problems in brief psychodynamic psychotherapy

The choice of focus is perhaps the most important and the most difficult aspect of brief individual psychodynamic psychotherapy. It is helpful to identify several foci during the evaluation process, recognizing that there are inevitably several conflict areas active at any one time in a patient's life. Then the therapist can begin the process of thinking through what the treatment of each focus would entail (Table 6.3.4.4).

The therapist can begin to decide which focal conflict will be more difficult to reach in a brief period of time, which will threaten the therapeutic alliance more and therefore require a deeper working relationship that may take more time, and which focus requires interpreting more primitive defences and therefore may be more complicated.

Choice of a particular focus can also create more family or external disruption or support which can aid or disrupt the treatment.

Use of medication requires carefully explaining to the patient the relationship of the medication to the psychotherapy. Often the medication treatment will continue beyond the psychotherapy.

Table 6.3.4.4 Identifying and selecting the focal conflict in brief dynamic psychotherapy

Identifying the focal conflict
Explore
Precipitant of symptoms
Early life traumas
Repetitive patterns of behaviour
Listen for inhibitions/avoidance
Watch for conflicts about success as well as loss/failure
Selection among several foci
Choose the focus that is presently active
Use trial interpretation to identify active focus
Select focus related to only one transference figure

If repeated complicated medication alterations are needed or if serious side effects of the medication occur, the psychotherapy plan may have to be altered to allow time to understand them from the patient's perspective.

New therapists are often concerned about setting the date of termination at the time of the evaluation, fearing that they may not be able to complete the work by the deadline. Supervision with an experienced colleague can be very helpful to assure confidence and avoid mistakes that may lengthen the treatment. Alternatively, the new therapist may feel too much relief in setting the termination date when treating a very dependent patient and therefore miss the intensity with which the patient is attached and experiencing the therapist as an important, needed, or feared figure from the past.

The management of missed sessions should be made clear at the beginning of treatment. Usually it is best not to 'make up' the sessions, but to keep to the termination date. If the therapist is concerned about this as a potential issue in the treatment, the therapist may wish to plan several additional sessions in the overall treatment to assure this can be discussed and understood therapeutically. Of course if an emergency arises it is always appropriate to schedule appointments as needed for the health and safety of the patient.

The patient who 'divulges' new 'secret' information near the end of the treatment is a challenge to all therapists. Understanding to what extent this represents narcissistic, or sociopathic issues, fear of the therapist or the treatment, or the emergence of hope for the future or a transference enactment will determine how to respond.

Brief individual psychodynamic psychotherapy is best learned in conjunction with the skills of longer term psychodynamic psychotherapy. In the longer work, the therapist will be able to see more easily the possible conflict areas and think about the sequencing of the treatment of these, i.e. which is closer to the patient's awareness or which is more defended. In addition there is more time to correct errors and repair untoward events in the therapeutic relationship. The brief individual psychodynamic psychotherapist will have less time to correct mistakes and must more quickly identify conflict areas and assess their relative importance and potential for resolution through treatment.

Efficacy: research and evaluation

The brief psychodynamic treatments have a small empirical database. Much further research is needed.[31] In general, studies have supported the efficacy of this treatment approach. However, methodological issues are prominent in most research in this area. The development of handbooks for treatment has gone far in improving research in the psychoanalytically oriented brief treatments.[19,32–34]

The effectiveness of psychotherapy in general, is not argued as in the past.[8,35–37] Brief psychodynamic psychotherapy has been shown to have an effect size similar to many other medical treatments. Short-term psychodynamic psychotherapy has shown modest to moderate, often sustained gains for a variety of patients.[38] However, the question of which psychotherapy is suitable for which patient and by which therapist is still unclear. The cost-effectiveness of psychotherapeutic treatment remains hotly debated and is a focus of substantial research.[9,39,40] Individual psychotherapy has been shown to result in fewer days of hospital stay for patients on medical or surgical services of a general hospital. In health clinics

or health maintenance organizations, brief psychotherapy decreases the number of visits to primary health care providers, reduces the number of laboratory and radiographic studies, decreases the number of prescriptions given, and, overall, reduces direct health care costs. Recently summaries of the cost-offset effects of outpatient mental health treatment, the majority of which were short-term are hopeful but not unambiguous. One study found outpatient psychotherapy resulted in a 33 per cent average reduction in medical care utilization. Furthermore, these reductions occurred mostly in the more expensive, inpatient medical services. In another study, 72 patients with significant emotional problems and treated only by internists in a general medical clinic were compared with 62 patients who, in addition to being treated by internists for medical problems, received 10 weekly psychotherapy visits. Both groups had approximately an equal degree of emotional disturbance. At 4-month and 1-year follow-ups, the brief psychotherapy group reported significantly more global improvement than the non-psychotherapy group. Also, more patients in the brief psychotherapy group became employed at 1-year follow-up than in the non-psychotherapy group. This study suggests specific beneficial effects of brief psychotherapy when used in a medical setting by skilled psychotherapists. Combining psychotherapy with antidepressant medication has also been shown to give the best outcome at 1 year when compared to either treatment alone. Whether a therapist keeps to a consistent frame of reference in the treatment may also be a predictor of success if brief individual psychodynamic psychotherapy, regardless of what that perspective is.[41]

Malan's finding of the importance of making the transference-parent link for the successful outcome of treatment is significant and requires further exploration.[2] One reanalysis of Malan's data confirmed his finding and one did not.[42] In addition, one replication of this finding has been published.[43] Importantly, more recently, the overuse of transference interpretations has been shown to lead to poorer outcome. The therapeutic alliance, particularly when measured from the patient's perspective, has a consistent although modest contribution to outcome.[9,44] It has been shown that independent of the type of treatment and early clinical improvement, the therapeutic relationship contributes directly to the positive therapeutic outcome.[45]

The quality of the therapeutic interaction and the handling of the transference and countertransference appears to be critical to success or failure in brief individual psychodynamic psychotherapy.[34] Patients treated by therapists who have not been professionally trained, may, on average, be as improved as patients treated by professional brief dynamic therapists. However, such non-experienced therapists run out of relevant material and are unwilling to continue to treat patients over an extended period of time.[46] One of the important tasks of training in psychotherapy may be the development of the ability to 'endure' with the patient and, over time, with numbers of patients. Technical training and a theoretical framework may allow the therapist to maintain a sense of competence, direction, and interest in the work which the non-professional therapist cannot.

Interpersonal psychotherapy and cognitive-behavioural therapy have been much more extensively studied than dynamic psychotherapy, particularly in combination with medications. Recently, telephone psychotherapy with cognitive behavioural therapy in primary care settings when initiating antidepressant medication has been shown to improve clinical outcome.[47,48] To the extent that these treatments share techniques and outcomes, similar results might be expected with brief dynamic psychotherapy; however, this still needs to be shown. Focal directive psychotherapies generally appear to be more effective than traditional unstructured psychodynamic psychotherapy for a number of types of patients, but a delineation of which psychotherapy for which patient over what time and with which medication remains to be demonstrated. Good clinical sense dictates combined treatments with matching the patient's cognitive and affective style with treatment type and making medication compliance a focus of any psychotherapy. Additionally, further research of brief psychodynamic psychotherapies in specific psychiatric disorders as well as across the life span are needed.[49,50]

Conclusion

Brief dynamic psychotherapy is an important treatment for numerous disorders, primarily the adjustment, anxiety, and mood disorders. Both alone and in combination with medication brief dynamic psychotherapy is an effective part of the treatment armamentarium. Clinicians should be trained in the brief as well as the longer term treatments and their use as brief, intermittent, and maintenance treatments. Skill in the longer term psychotherapies is important to developing skill in the brief dynamic psychotherapy where the needs for rapid establishment of the therapeutic alliance and the accurate assessment of transference and defence patterns are important.

Empirical studies comparing well-defined brief dynamic psychotherapy with cognitive and interpersonal psychotherapies are limited. Future research must address which form of brief psychotherapy may be most helpful for which patient. An individual's preferred learning path-what he or she may see and observe most easily such as thoughts or feelings or interpersonal relations-may be an important variable in determining which brief psychotherapy for which patient. State, trait, and contextual variables will influence this learning modality. The process of change in brief individual psychodynamic psychotherapy, a process of altering neuronal organization through verbal means, is influenced by the patient's diagnosis, medications, past history, cognitive style, developmental stage, and affective availability, as well as the doctor–patient match.

Further information

Ursano, R.J., Sonnenberg, S., and Lazar, S. (2004). *Concise guide to psychodynamic psychotherapy: principles and techniques of brief, intermittent and long term psychodynamic psychotherapy*. American Psychiatric Press, Washington, DC.

Levinson, H., Butler, S.F., Powers, T.A., *et al.* (2002). *Concise guide to brief dynamic and interpersonal psychotherapy* (2nd edn). American Psychiatric Publishing Inc., Washington, DC.

Luborsky, L. and Luborsky, E. (2006). *Research and psychotherapy: the vital link*. Jason Aronson, Lanham MD.

Dewan, M.J., Steenbarger, B.N., and Greenberg, R.P. (eds.) (2004). *The art and science of brief psychotherapies: a practicioner's guide*. American Psychiatric Publishing Inc., Arlington, VA.

References

1. Gabbard, G.O. (1994). *Mind and brain in psychiatric treatment*. Institute of Pennsylvania Hospital Strecker Award Monograph Series 31. Pennsylvania Hospital, Philadelphia, PA.

2. Ursano, R.J. and Silberman, E.K. (1988). Individual psychotherapies. In *Textbook of psychiatry* (eds. J.A. Talbott, R.E. Hales, and S.C. Yudofsky), pp. 855–89. American Psychiatric Press, Washington, DC.
3. Michels, R. (1997). Psychodynamic psychotherapy in modern psychiatry. *Journal of Practical Psychiatry and Behavioral Health*, **3**, 95–8.
4. Ferenczi, S. and Rank, O. (1925). *The development of psychoanalysis*. Nervous and Mental Diseases Publishing Company, New York.
5. Sifneos, P.E. (1972). *Short-term psychotherapy and emotional crisis*. Harvard University Press, Cambridge, MA.
6. Sifneos, P.E. (1984). The current status of individual short-term dynamic psychotherapy and its future: an overview. *American Journal of Psychotherapy*, **37**, 472–83.
7. Sledge, W.H., Moras, K., Hartley, D., *et al.* (1990). Effect of time-limited psychotherapy on patient dropout rates. *The American Journal of Psychiatry*, **147**, 1341–7.
8. Crits–Christoph, P. and Barber, J.P. (eds.) (1991). *Handbook of short-term dynamic psychotherapy*. Spectrum, New York.
9. Ursano, A.M., Sonnenberg, S.M., and Ursano, R.J. (in press). Physician patient relationship in psychiatry (3rd edn) (eds. A. Tasman, J. Kay, and J.A. Lieberman). Wiley & Sons, Ltd., West Sussex, England.
10. Coleman, J.V. (1968). Aims and conduct of psychotherapy. *Archives of General Psychiatry*, **18**, 1–6.
11. McGuire, M. (1965). The process of short-term insight psychotherapy. *The Journal of Nervous and Mental Disease*, **141**, 83–94.
12. Frances, A. and Perry, S. (1983). Transference interpretations in focal therapy. *The American Journal of Psychiatry*, **140**, 405–9.
13. Klan, H. and Frances, A. (1984). Countertransference in focal psychotherapy. *Psychotherapy and Psychosomatics*, **41**, 38–41.
14. Levinson, H., Butler, S.F., Powers, T.A., *et al.* (2002). *Concise guide to brief dynamic and interpersonal psychotherapy* (2nd edn). American Psychiatric Press, Washington, DC.
15. Stierlin, H. (1968). Short-term versus long–term psychotherapy in the light of a general theory of human relationships. *British Journal Medical Psychology*, **41**, 357–67.
16. Hoyt, M. and Farrell, D. (1984). Countertransference difficulties in a time–limited psychotherapy. *International Journal of Psychoanalytic Psychotherapy*, **10**, 191–203.
17. Shlien, J.M., Mosik, H.H., and Dreikurs, R. (1962). Effective time limits: a comparison to psychotherapy. *Journal of Counselling Psychology*, **9**, 31–4.
18. Howard, K., Kopta, S., Krause, M., *et al.* (1986). The dose-effect relationship in psychotherapy. *The American Psychologist*, **41**, 159–64.
19. Milrod, B.L., Busch, F.N., and Cooper, A.M. (eds.) (1996). *Manual of panic-focused psychodynamic psychotherapy*. American Psychiatric Press, Washington, DC.
20. Horowitz, M.J., Marmar, C., Krupnick, J., *et al.* (1984). *Personality styles in brief psychotherapy*. Basic Books, New York.
21. Balint, M., Ornstein, P., and Balint, E. (1972). *Focal psychotherapy*. Lippincott, Philadelphia, PA.
22. Malan, D.H. (1975). *A study of brief psychotherapy*. Plenum Press, New York.
23. Malan, D.H. (1980). *Toward the validation of dynamic psychotherapy*. Plenum Press, New York.
24. Ursano, R.J., Sonnenberg, S., and Lazar, S. (2004). *Concise guide to psychodynamic psychotherapy: principles and techniques of brief, intermittent and long term psychodynamic psychotherapy.* American Psychiatric Press, Washington, DC.
25. Mann, J. (1980). *Time–limited psychotherapy*. Harvard University Press, Cambridge, MA.
26. Mann, J. and Goldman, R. (1995). *A casebook in time-limited psychotherapy*. Jason Aronson, New York.
27. Davanloo, H. (ed.) (1980). *Short-term dynamic psychotherapy*. Jason Aronson, New York.
28. Markowitz, J.C. (ed.) (1998). *Interpersonal psychotherapy*. American Psychiatric Press, Washington, DC.
29. Klerman, G.L., Weissman, M.M., Rounsaville, B.J., *et al.* (1984). *Interpersonal psychotherapy of depression*. Basic Books, New York.
30. Beck, A.T. and Rush, A.J. (1995). Cognitive therapy. In *Comprehensive textbook of psychiatry*, Vol. VI (eds. H.I. Kaplan and B.J. Sadock), pp. 1847–57. Williams and Wilkins, Baltimore, MD.
31. Kay, J. (1997). Brief psychodynamic psychotherapies: past, present and future challenges. *Journal of Psychotherapy Practice and Research*, **6**, 330–7.
32. Luborsky, L. (2000). *Principles of psychoanalytic psychotherapy: a manual for supportive expressive treatment*. Basic Books, New York.
33. Miller, N.E., Luborsky, L., Barber, J.P., *et al.* (eds.) (1993). *Psychodynamic treatment research*. Basic Books, New York.
34. Strupp, H.H. and Binder, J. (1984). *Psychotherapy in a new key: time-limited dynamic psychotherapy*. Basic Books, New York.
35. Gabbard, G.O., Lazar, S.G., Hornberger, J., *et al.* (1997). The economic impact of psychotherapy: a review. *The American Journal of Psychiatry*, **154**, 147–55.
36. Lazar, S.G. (ed.) (1997). *Extended dynamic psychotherapy: making the case in an era of managed care.* Psychoanalytic inquiry supplement. Analytic Press, Hillsdale, NJ.
37. Crits–Christoph, P. (1992). The efficacy of brief dynamic psychotherapy: a meta-analysis. *The American Journal of Psychiatry*, **149**, 151–8.
38. Abass, A.A., Hancock, J.T., Henderson, J., *et al.* (2007). *Short-term psychodynamic psychotherapies for common mental disorders* (*Review*), pp. 1–47. The Cochrane Collaboration, John Wiley & Sons, Ltd, New York.
39. Ursano, R.J. and Silberman, E.K. (1999). Psychoanalysis, psychoanalytic, psychotherapy, and supportive psychotherapy. In *Textbook of psychiatry* (eds. R.E. Hales, S.C. Yudofsky, and J.A. Talbot), pp. 1157–84. American Psychiatric Press, Washington, DC.
40. Wiborg, I.M. and Dahl, A.A. (1996). Does brief dynamic psychotherapy reduce the relapse rate of panic disorder? *Archives of General Psychiatry*, **53**, 689–94.
41. Barbar, J.P., Crits–Christoph, P., and Luborsky, L. (1996). Effects of therapist adherence and competence on patient outcome in brief dynamic therapy. *Journal of Consulting and Clinical Psychology*, **64**, 619–22.
42. Hoglend, P. (1996). Long-term effects of transference interpretations: comparing results from a quasi-experimental and a naturalistic long-term follow-up study of brief psychotherapy. *Acta Psychiatrica Scandinavica*, **93**, 205–11.
43. Marziali, E.A. (1984). Prediction of outcome of brief psychotherapy from therapist interpretive interventions. *Archives of General Psychiatry*, **41**, 301–4.
44. Joyce, A.S., Ogrodniczuk, J.S., Piper, W.E., *et al.* (2003). The alliance as mediator of expectancy effects in short-term individual therapy. *Journal of Consulting and Clinical Psychology*, **71**, 672–9.
45. Zuroff, D.C. and Blatt, S.J. (2006). The therapeutic relationship in the brief treatment of depression: contributions to clinical improvement and enhanced adaptive capacities. *Journal of Clinical Psychology*, **74**, 130–40.
46. Strupp, H.H. (1980). Success and failure in time-limited psychotherapy: with special reference to the performance of lay counselors. *Archives of General Psychiatry*, **37**, 831–41.
47. Simon, G.E., Ludman, E.J., Tutty, S., *et al.* (2004). Telephone psychotherapy and telephone care management for primary care patients starting antidepressant treatment: a randomized controlled trial. *The Journal of the American Medical Association*, **292**, 935–42.
48. Ludman, E.J., Simon, G.E., Tutty, S., *et al.* (2007). A randomized trial of telephone psychotherapy and pharmacotherapy for depression: continuation and durability of effects. *Journal of Consulting and Clinical Psychology*, **75**, 257–66.

49. Leichenring, F, Rabung, S, and Leibing, E. (2004). The efficacy of short-term psychodynamic psychotherapy in specific psychiatric disorders. *Archives of General Psychiatry*, **61**, 1208–16.
50. Shefler, G. (2000). Time-limited psychotherapy with adolescents. *Journal of Psychotherapy Practice and Research*, **9**, 2.

6.3.5 Psychoanalysis and other long-term dynamic psychotherapies

Peter Fonagy and Horst Kächele

Introduction

Basic assumptions

The term psychodynamic psychotherapy has no specific referent. It denotes a very heterogeneous range of psychological treatment approaches which arguably have in common an intellectual heritage of psychoanalytic theory. Psychoanalytic theory itself is no longer based on a unitary body of ideas[1] but a number of ideas appear to be core to most psychodynamic approaches. These notions are:

(a) A shared notion of psychological causation, that mental disorders can be meaningfully conceived of as specific organizations of an individual's conscious or unconscious beliefs, thoughts, and feelings.

(b) Psychological causation extends to the non-conscious part of the mind, and to understand conscious experiences, we need to refer to other mental states of which the individual is unaware.

(c) The mind is organized to avoid unpleasure arising out of conflict[2] in order to maximize a subjective sense of safety.[3]

(d) Defensive strategies are a class of mental operations that seem to distort mental states to reduce their capacity to generate anxiety, distress, or displeasure. Individual differences in the predisposition to specific strategies have often been used as a method for categorizing individuals or mental disorders.[4,5]

(e) Varying assumptions are made concerning normal and abnormal child and adolescent development but therapists are invariably oriented to the developmental aspects of their patients' presenting problems.[6]

(f) Relationship representations linked with childhood experience are assumed to influence interpersonal social expectations including the transference relationship with the therapist[7] and to shape the representations of the self.[8–11]

(g) These relationship representations inevitably re-emerge in the course of psychodynamic treatments.[12]

Brief overview of theories

Psychoanalytic theory has evolved from the work of Freud following two broadly separate paths which converged over the past 25 years only to separate again. In the United States followers of the Vienna school in the 1950s and 1960s evolved a systematic psychology of the ego, a conflict-oriented complex psychological model of the mind and its disturbances.[13] In Europe, only Anna Freud and her followers in London pursued this tradition of psychoanalytic thought.[14] Based on the Berlin school of Karl Abraham, Melanie Klein and her followers established a distinct approach focusing on the understanding of disturbance rooted in infantile destructiveness and sadism.[15] Some psychoanalysts, influenced by Klein and the idea of the pathogenic nature of the experiences of infancy, gradually discarded the mechanistic psychology of drives and psychology of internal structures in favour of theories of intrapsychic interpersonal relationships (object-relations theory).[16]

As these schools developed in the United Kingdom, their influence travelled across the Atlantic. First, Kohut, strongly influenced by Winnicott (albeit without explicit acknowledgement), evolved a psychoanalytic psychology of the self.[17] Shortly after, Kernberg arrived at an imaginative integration of ego-psychological and Kleinian ideas.[18] In the meantime, in the United Kingdom, the Kleinian movement rapidly progressed in their understanding of psychoanalytic clinical experience, moving beyond Klein's original work and integrating some of the key features of the Anna Freudian and the British object-relations traditions.[19] In the United States, disillusionment with the false certainty provided by ego-psychology became intense throughout the late 1970s and early 1980s and a radical change in psychoanalytic thinking took place with the emergence of the interpersonal relational perspective, which is in part rooted in the work of Harry Stack Sullivan.[20,21] The relational psychoanalysis of the 1980s and 1990s consolidated several lines of thought initiated by justified critiques of traditional analytic theory[22]; including feminism, the hermeneutic-constructivist critique of the analyst's authority, infancy research, and, closely related to this, the intersubjectivist-phenomenological philosophy of mind—as well as a general political movement to improve and democratize access to analytic ideas and training.[23]

There are many other new psychoanalytic theoretical approaches, bringing the field increasingly close to total fragmentation.[24] This is because the emergence of new approaches in no way signals the demise of any previous orientations, most of which continue to enjoy considerable popularity among specific groups of psychoanalysts.

Psychoanalytic therapy as treatment

The history of psychoanalysis as a therapeutic approach is rather different. Broadly speaking, it may be argued that psychoanalysis and other long-term psychodynamic therapies are predominantly verbal, interpretive, insight-oriented approaches which aim to modify or re-structure maladaptive relationship representations. It is implicitly assumed that genetic and early environmental factors give rise to partial, unintegrated, and generally troublesome relationship representations (e.g. a helpless 'infant' requiring total care from an adult, a self with exaggerated sense of power, and entitlement requiring constant confirmation from outside) that lie at the root of psychological disturbance. It is believed that the integration of these partial representations into more complex schemata, primarily but not exclusively through the use of insight, leads to improved internal and social adjustment.

Psychoanalysis is the most intensive form of these long-term therapies. The analysand attends treatment three or more times a week over a period of years. The use of the couch and the instruction to the analysand to free associate have been considered hallmarks. The distinction between psychoanalysis and other forms of psychotherapy is normally made in terms of the frequency of

sessions rather than in terms of the therapeutic stance of the analyst. It is difficult to avoid the conclusion that in the absence of plausible, theoretically based criteria for what is or is not psychoanalytic, against the background of an overwhelming diversity of theoretical frameworks, psychoanalysts have attempted to find common ground in readily identifiable treatment parameters. This problem arises as a consequence of an extremely loose relationship between psychoanalytic theory and clinical practice.[24] It is an indisputable fact that, whereas theory has evolved extremely rapidly in the last half of the twentieth century and continues to change, psychoanalytic practice has, until recently, changed surprisingly little and continues to provide the core of the psychoanalytic identity. On the other hand, the follow-along study by Sandell *et al.*[25] found that psychoanalysis and psychoanalytic psychotherapy were 'separate things'. When psychotherapy was performed using mainly psychoanalytic techniques, it was less effective than psychotherapy performed with modified and adjusted techniques (that is, not performed as an 'as-if analysis'). The findings from the Stockholm study suggest that psychoanalysis and psychoanalytic psychotherapy may be separate endeavours, although how exactly they differ is far from clear.

In this chapter we will not consider the theoretical richness of this field but instead will focus on the clinical constructs which run across the diverse intellectual approaches. The intersection of the two is perhaps clearest in one area which we shall consider in some detail—namely, the therapeutic action of long-term psychoanalytically oriented psychotherapeutic treatment.

Background

Historical development of the psychoanalytic approach to treatment

As is well known, Freud's discovery of the talking cure[26] was really that of an intelligent patient (Anna O) and her physician (Breuer). The patient reported that certain symptoms disappeared when she succeeded in linking up fragments of what she said and did in an altered state of consciousness (which we might now call dissociative) with forgotten impressions from her waking life. Breuer's remarkable contribution was that he had faith in the reality of the memories which emerged and did not dismiss the patient's associations as products of a deranged mind. The patient's response to treatment was probably less complete than Breuer and the young Freud had hoped[27] but the 'treatment' defined the basic elements of the 'cathartic' method-linking memory of trauma (the circumstances of her experience of her father's death) to her many symptoms.

At first Freud rigorously pursued the traumatogenic origins of neuroses. Later, when confronted by evidently incorrect statements, he modified his theory, assuming consistency between recollection and childhood psychic reality rather than physical reality.[28] The issue of accuracy of memories of childhood sexual trauma remains controversial, although its relevance to psychoanalytic technique is at best tangential.[29] Freud's technique, however, was dramatically modified by his discoveries. The intense emotional relationship between patient and physician, which had its roots in catharsis following hypnotic suggestion, had gradually subsided into what was principally an intellectual exercise to reconstruct the repressed causes of psychiatric disturbance from the fragments of material derived from the patient's associations. It was a highly mechanistic approach reminiscent of a complex crossword puzzle. In the light of therapeutic failures, however, Freud once more restored the emotional charge into the patient–physician relationship.[30] However, in place of hypnosis and suggestion, he used the patient's emotion, signs of transference of affect and affective resistance which were manifest in the analytic relationship. Instead of seeing the patient's intense emotional reaction to the therapist as an interference, Freud came to recognize the importance of transference as a representation of earlier relationship experiences which could make the reconstruction of those experiences in analysis highly meaningful to that individual.[31]

Freud's early clinical work evidently lacked some of the rigour which came to characterize classical psychoanalysis.[32] His occasional encouragement to his patients to join him on holiday might now be considered a boundary violation.[33] What is perhaps less well known is that Freud remained somewhat sceptical about the effectiveness of psychoanalysis as a method of treatment.[34] Indeed, autobiographies of some of his patients testify to his great flexibility as a clinician and use of non-psychoanalytic techniques, including behavioural methods.[35] Nor was Freud the only clinician to use psychoanalytic ideas flexibly. The Hungarian analyst Sandór Ferenczi should be credited with the discovery of the treatment of phobic disorders by relaxation and exposure[36] although many of his well-intentioned actions were criticized by contemporaries and more recently on arguable ethical grounds.[37]

The technique of psychoanalysis after Freud's death came to be codified. Those (such as Alexander and French and Freda Fromm-Reichmann) who attempted to revive or retain Freud's original clinical flexibility were subjected to powerful intellectual rebuttals.[38] In reality, psychoanalysts probably continued to vary in the extent to which they observed the ideals of therapeutic neutrality, abstinence, and a primarily interpretive stance, but these deviations could no longer be exposed to public scrutiny for fear of colleagues' forceful condemnation. Personal accounts of analyses with leading figures yield fascinating insights into variations in technique, principally in terms of the extent to which the analyst made use of a personal relationship.[39] There has been an ongoing dialectic throughout the history of psychodynamic approaches between those who emphasize interpretation and insight and those who stress the unique emotional relationship between patient and therapist as the primary vehicle of change. The controversy dates back to disputes concerning the work of Ferenczi and Rank[40] but re-emerged with the first papers of Balint and Winnicott in London opposing a Freudian and Kleinian tradition, and somewhat later in the United States with Kohut and more subtly Loewald opposing classical ego psychology.

In the last two decades, the pluralistic approach of modern psychoanalysis has brought out into the open many important dimensions along which psychoanalysts' techniques may vary. In particular, the recent trend to consider analyst and patient as equal partners engaged in a mutual exploration of meaning[41] directly challenged many of the classical constructs. The emphasis on the mutual influence of infant and caregiver shaped the emerging relational model of therapy as a two-person process in which there was little room for a detached analyst with pretensions of 'objectivity'. Drawing on the assumption that humans are predisposed towards two-person co-constructed systems that provide a context for psychic change, the quality of engagement

between therapist and patient became the core of therapeutic action. What changes the mind is not the insights gained but learning from the interactional experience of being with another person. Neither the analyst nor the patient can be considered as forging meaning; rather, meaning is co-constructed.

Technique—principal features

Neutrality and abstinence

Based in the classical framework of libidinal theory, Freud made an explicit injunction against the analyst giving in to the temptation of gratifying the patient's sexual desire.[42] Obviously, this is primarily an ethical issue. However, within the psychoanalytic context it also justifies the analyst's stance of resisting the patient's curiosity or using the therapeutic relationship in any way that consciously or unconsciously could be seen as motivated by the need to gratify their own hidden desires. Within this classical frame of reference, the patient must also agree to forgo significant life changes where these could be seen as relevant to current psychotherapeutic work. In practice, such abstinence on the part of the patient is rare. Yet long-term psychodynamic treatment may founder if the emotional experiences of the therapy are obscured by the upheavals of significant life events.

The primary function of abstinence is to ensure the neutrality of the therapist. The analyst assumes an attitude of open curiosity, empathy, and concern in relation to the patient. The therapist resists the temptation to direct the patient's associations and remains neutral irrespective of the subject matter of the patient's experiences or fantasies. While it is easy to take this issue too lightly, (and it is perhaps this aspect of the psychoanalyst's therapeutic stance which makes them most vulnerable to ridicule), it is probably genuinely critical for the therapist to retain emotional distance from the patient to a degree which enables the latter to bring fantasies and fears of which they feel uncertain. Nevertheless, neutrality at its worst denies the possibility of sensitivity; recent literature on the process and outcome of psychotherapy makes it clear that the therapist's genuine concern for the patient must become manifest if significant therapeutic change is to be achieved.[43] The quality of the alliance is one of the better predictors of outcome[44] and alliance is impacted by the patient's attachment style and quality of object-relations.[45]

Mechanisms of defence

The term 'psychic defences' may risk reification and anthropomorphism (precisely who is defending whom against what?) yet the existence of self-serving distortions of mental states relative to an external or internal reality is generally accepted, and frequently demonstrated experimentally.[46–48] Within classical psychoanalytical theory and its modern equivalent (ego psychology), intra-psychic conflict is seen as the core of mental functioning.[49] Here defences are seen as adaptations to reduce conflict. Within many object-relations theories, defences are seen as helpful to the individual to maintain an authentic or 'true' self-representation or a nuclear self.[17] Models of representations of relationships are of course often defensive. Traumatic experiences may give rise to omnipotent internal working models to address a feeling of helplessness. Within attachment theory, defences are construed as assisting in the maintenance of desirable relationships.[50] The Klein–Bion model makes limited use of the notion of defence mechanisms but uses the term in the context of more complex hypothetical structures called defensive organizations.[19] The term underscores the relative inflexibility of some defensive structures, which are thus best conceived of as personality types. For example, narcissistic personality disorder combines idealization and destructiveness; genuine love and truth are devalued. Such a personality type may have been protective to the individual at an earlier developmental stage, and has now acquired a stability or autonomy which must be rooted in the emotional gratification which such a self-limiting form of adaptation provides.[51]

Irrespective of the theoretical frame of reference, from a therapeutic viewpoint clinicians tend to differentiate between so-called primitive and mature defences based on the cognitive complexity entailed in their functioning.[52] In clinical work, primitive defences are often noted together in the same individual. For example, individuals loosely considered 'borderline' tend to idealize and then derogate the therapist. Thus they maintain their self-esteem by using splitting (clear separation of good from bad self-perception) and then projection. Projective identification[53] is an elaboration of the process of projection. An individual may ascribe an undesirable mental state to the other through projection but when the other can be unconsciously forced to accept the projection and experience its impact, the defence becomes far more powerful and stable. The analyst's experiencing of a fragment of the patient's self-state, has in recent years been considered an essential part of therapeutic understanding.[54]

Whether in fantasy or in actualized form, through projective identification the patient can experience a primitive mode of control over the therapist. Bion argued that when the self is experienced as being within another person (the therapist) the patient frequently attempts to exert total control over the recipient of the projection as part of an attempt to control split-off aspects of the self. Bion[55] also argued that not all such externalizations were of 'bad' parts of the self. Desirable aspects of the self may also be projected, and thus projective identification can be seen as a primitive mode of communication in infancy. There are other aspects of projective identification which we commonly encounter clinically. These include the acquisition of the object's attributes in fantasy, the protection of a valued aspect of the self from internal persecution through its evacuation into the object, and the avoidance or denial of separateness. It is thus a fundamental aspect of interpersonal relationship focused on unconscious fantasy and its appreciation is critical for the adequate practice of long-term psychotherapy.[56]

Classifications of defences have been frequently attempted[52,57–61] and often as a method for categorizing individuals or mental disorders.[4,5] An attachment theory-based classification rooted in the notion of habitual deactivation or hyperactivation of the attachment system ('attachment style') has achieved general acceptance.[62,63] Deactivating ('avoidant' or 'dismissing') strategies include suppression of ideas related to painful attachment experiences, repressing painful memories, minimizing stress and distress, segregated mental systems that result in the defensive exclusion of distressing material from the stream of consciousness.[64,65] Ingenious experimental studies have shown that individuals who habitually use avoidant defences are more efficient, when instructed, at suppressing conscious thoughts and associated feelings about a romantic partner leaving them for someone else[66] and are more likely to attribute their own unwanted traits to others

(projection) which serves to both increase self-other differentiation and enhance self-worth.[67] In a further, remarkable study the same group of researchers demonstrated that the above advantages of the suppression strategy of those using avoidant defence fall away in the laboratory situation if a cognitive load is placed on the participant which then leaves them literally defenceless so that they experience a heightened rebound of previously suppressed thought about painful separation.[68] The cognitive and socio-cognitive strategies associated with reducing anxiety or displeasure and enhancing safety, which both the attachment theory and psychoanalytic literatures tend to refer to as defences, are perhaps better thought of not as independent classes of mental activity or psychological entities but as a pervasive dynamic aspect of complex cognition interfacing with attachment relationships and emotional experience. Some mechanisms of defence are thought to be more characteristic of the less severe psychological disorders (e.g. depression, anxiety, obsessive–compulsive disorders, etc.). It is beyond the scope of this chapter to consider the various defence mechanisms in detail.

Modes of therapeutic action

The primary mode of the therapeutic action of psychoanalytic psychotherapy is generally considered to be insight.[69] Insight may be defined as the conscious recognition of the role of unconscious factors on current experience and behaviour. Unconscious factors encompass unconscious feelings, experiences, and fantasies. The psychodynamic model has been seen as a model of the mind that emphasizes repudiated wishes and ideas which have been warded off, defensively excluded from conscious experience. In our view this is a narrow and somewhat misleading way to define the therapeutic mechanism for approaches that are considered as psychodynamic. The psychodynamic approach is better seen as a stance taken to human subjectivity that is comprehensive, and aimed at understanding all aspects of the individual's relationship with her or his environment, external, and internal. Freud's great discovery ('where id was, there ego shall be', Freud[70] p. 80), often misinterpreted, points to the power of the conscious mind radically to alter its position with respect to aspects of its own functions, including the capacity to end its own existence through killing the body. Psychodynamic, in our view, refers to this extraordinary potential for dynamic self-alteration and self-correction—seemingly totally outside the reach of non-human species. Engaging with this potential to bring change through understanding, is the science and the art of the psychodynamic clinician.

Conscious insight is more than mere intellectual knowledge[71,72] or descriptive insights. Prototypically, psychodynamic therapy achieves demonstrated or ostensive insights which represent a more direct form of knowing, implying emotional contact with an event one has experienced previously. Working with what is non-conscious is at the heart of the dynamic approach to bringing about psychological change because of the force that awareness of unconscious expectations can bring to the interpretation of behaviour. Although specific formulations of the effect of insight depend on the theoretical framework in which explanations are couched, there is general agreement that insight has its therapeutic effect by in some way integrating mental structures.[72] Kleinian analysts[73] tend to see the healing of defensively created splits in the patient's representation of self and others as crucial. Split or part-objects may also be understood as isolated representations of intentional beings whose motivation is insufficiently well understood for these to be seen as coherent beings.[74] In this case insight could be seen as a development of the capacity to understand internal and external objects in mental state terms, thus lending them coherence and consistency.[75] The same phenomenon may be described as an increasing willingness on the part of the patient to see the interpersonal world from a third person's perspective.[76]

A simple demonstration to the patient of such an integrated picture of self or others is not thought to be sufficient.[31] The patient needs to 'work through' a newly arrived integration. Working through is a process of both unlearning and learning: actively discarding prior misconceptions and assimilating learning to work with new constructions. The technique of working through is not well described in the literature, yet it represents the critical advantage of long-term over short-term therapy.[77] Working through should be systematic and much of the advantage of long-term treatment may be lost if the therapist does not follow through insights in a relatively consistent and coherent manner.

In contrast to the emphasis on insight and working through are those clinicians who, as we have seen, emphasize the 'relationship aspect' of psychoanalytic therapy (Balint, Winnicott, Loewald, Mitchell, and many others). This aspect of psychoanalytic therapy was perhaps most eloquently described by Loewald when he wrote about the process of change as: 'set in motion, not simply by the technical skill of the analyst but by the fact that the analyst makes himself available for the development of a new 'object-relationship' between the patient and the analyst . . . '(Loewald, 1960, pp. 224–5).[78] Sandler and Dreher[79] have recently observed 'while insight is aimed for it is no longer regarded as an absolutely necessary requirement without which the analysis cannot proceed'. There is general agreement that the past polarization of interpretation and insight on the one hand, and bringing about change by presenting the patient with a new relationship on the other, was unhelpful. It seems that patients require both, and both may be required for either to be effective.[80]

Controversy remains even if all accept that neutrality is an impossible and undesirable fiction and that patient and therapist affect each other in myriad mutually influencing ways. Projective identification is seen as occurring in a bidirectional interpersonal field between analyst and patient—a model clearly adapted from Kleinian approaches to infant-caregiver interaction.[23] If we take this perspective seriously, we have to concede that all analytic interventions change the situations into which they are introduced, and their content and style always reflect the analyst's countertransference/response to the treatment situation.[81] Relational psychoanalysis advocates making the interactional influence of analyst upon patient explicit. As Levenson[82](p. 9) put it, the key therapeutic question is not 'what does this mean?' but rather 'what is going on around here?' The therapist will 'act' on the patient; this is not a therapeutic disaster but rather a potentially progressive and certainly inevitable part of the process.

It has been suggested that change in analysis will always be individualized according to the characteristics of the patient or the analyst.[83] For example, Blatt[84] suggested that patients who were 'introjective' (preoccupied with establishing and maintaining a viable self-concept rather than establishing intimacy) were more responsive to interpretation and insight. By contrast, anaclitic patients (more concerned with issues of relatedness than of self-development) were more likely to benefit from the quality of

the therapeutic relationship than from interpretation. Taking a second look at large-scale outcome investigations Blatt found strong evidence for the oft made but rarely demonstrated claim of patient personality—therapeutic technique fit.[85]

Indications and contraindications and selection procedures

Medical treatments normally have indications and contraindications. In psychodynamic treatment the term 'suitability' indicates a looser notion of the appropriateness of the approach.[86] Nevertheless, based primarily on clinical experience, some writers have arrived at specific criteria for long-term psychodynamic therapy.[87] Some authors have also suggested relatively systematic methods of assessment yielding both diagnostic and prognostic information.[88] The majority of psychodynamic clinicians, however, rely on clinical judgements based on interpersonal aspects of their first meeting with the patient.[71] The three areas of assessment are personal history, the content of the interview, and the style of the presentation.

A history of one good relationship has been traditionally regarded as a good indicator.[89] By contrast, a history of psychotic breakdown, severe obsessional states, somatization, and lack of frustration tolerance are generally considered contraindications. For example, a challenging set of re-analyses of the Treatment of Depression Collaborative Research Program found that the trait of perfectionism was associated with poor outcome, and could undermine the therapeutic alliance and the patient's satisfaction with social relations, limiting their improvement in the course of brief treatment for depression.[90]

Empirical literature, to the meagre extent that this is available, suggests that many of the presuppositions about suitability are unfounded. It was, for example, assumed that patients who manifested more serious mental illness, especially disturbances in reality testing, were unsuitable for psychoanalysis; however, a recent study showed that some patients with serious disturbances in reality testing were able to benefit from psychoanalysis when their analysts were able to tolerate and analyse this level of psychopathology.[91] What does seem to be consistent is that severity of symptoms, as well as functional levels in work and relationships, are correlated with the outcome of psychotherapy[92]—although no single patient variable is a strong predictor of outcome. This is why the effects of psychotherapy, good and bad, can sometimes be surprising.

Predicition based on the content of assessment interviews is hard. In general, the presence of some kind of 'mutuality' between therapist and patient is a positive indicator. Some clinicians offer 'trial interpretations' which summarize their initial impressions, and a positive thoughtful response to these is regarded a good indication. The capacity to respond emotionally within the assessment session is a further indicator.[93] Motivation for treatment is harder to ascertain. Most patients express enthusiasm for the treatment, which falls away once they are asked to confront unpleasant or unflattering parts of themselves.

More recently, psychodynamic therapists have given increasing consideration to the style of the patient's discourse during assessment rather than its content. Holmes,[94] for example, attempts to identify whether patients' narrative styles are avoidant (sparse and dismissing of interpersonal issues) or enmeshed and entangled (excessive current anger about past hurts and insults). The findings of one study indicate that, in a severely personality disordered population at least, the avoidant type of patient has a better prognosis in psychodynamic therapy.[95] A further relevant capacity is reflective function or mentalization, often reflected in narrative; this has been variously described as seeing oneself from the outside,[96] reflecting on one's inner world[87] or having fluidity of thought.[97]

Managing treatment

Starting treatment

(a) Establishing parameters

Most psychodynamic therapists, explicitly or implicitly, convey objectives and expectations to their patients. The details of this agreement normally include arrangements for a time and a place as well as the length and frequency of sessions. Usually a tentative idea is offered as to the likely duration of therapy: 'It is likely to take years rather than months.' Most therapists also describe the expected behaviour of the patient and the therapist: 'I would like you to be as open and honest with me as possible and say absolutely everything that comes into your mind. This is the fundamental rule.' In fact it is very likely, in view of the variety of such agreements that tend to be made, that its emotional context is more relevant than the specific items agreed upon. Such a 'contract' implies recognition by both patient and therapist that the process of therapy needs protecting and that it is important enough to require a sacrifice from both parties.

In the treatment of severe personality disorders, contracts may have an additional important function—that of protecting the therapy from incessant enactments, self-harming, parasuicidal gestures, and so on. In Kernberg's approach to the treatment of borderline patients, the patient formally undertakes not to seek the therapist's help outside of office hours, not to engage in acts of violence and to deal with self-destructive acts through normal medical channels.[98] Whilst such agreements are commonly made in long-term therapy, it is by no means clear that they are either essential or useful. For example, in an alternative form of psychodynamic therapy, Mentalization-Based Treatment (MBT), contracts are not recommended.[99]

(b) Formulation of patients' problems

An important part of initiating any psychosocial treatment is arriving at least at a preliminary formulation of the patient's problems. In the case of psychodynamic therapies this represents a special challenge because of the diversity of the possible theories to draw on. In principle, psychodynamic formulations would identify key unconscious conflicts, central maladaptive defences, unhelpful unconscious fantasies and expectations, deficits in personal development, and so on. The complexity of such formulations is such that agreements are hard to arrive at even when clinicians follow similar orientations. In the absence of a generally accepted format for formulating the patient's problems, a list of key parameters for the level of maturity of personality organization may be offered:

(a) the maturity of relationship representations (three or more persons versus just a self-other dimension)

(b) the maturity of psychic defences (primarily based on projective versus internalizing processes)

(c) the extent of whole as opposed to part object-relations (e.g. whether a person is represented as performing more than a single function for the patient)

(d) the general mutuality of the relationship patterns described; the quality of attachment to others.

It should be noted that psychodynamic formulations tend to change as treatment progresses. Indeed, Winnicott described psychoanalysis as 'an extended form of history taking'.[100] Within certain psychodynamic approaches formulation is communicated formally to patients (e.g. by letter in cognitive analytic therapy Ryle[101]).

The middle phase

(a) Supportive and directive interventions in psychodynamic therapy

Supportive techniques are used both explicitly and implicitly in psychodynamic treatment. They include offering explicit support and affirmation; offering reassurances concerning, for example, irrational anxieties about the therapeutic arrangements; expressing concern and sympathy to a patient who has suffered a recent loss; and general empathy for the patient's anxieties and struggles with the treatment.[102]

From a psychodynamic point of view, such supportive interventions are by no means straightforward. For example, Feldman[103] illustrated how patients may sometimes experience the therapist's submission to a demand for reassurance as a source of anxiety rather than comfort. They may be unconsciously aware that the therapist's true stance is not compatible with reassurance and therefore face anxieties about the therapist's weakness in allowing themselves to be manipulated. By contrast, Kohut's[17] emphasis on interpersonal empathy was probably a welcome antidote to the somewhat rigid interpretive stance of American ego psychologists, particularly for those whose history of psychosocial deprivation meant that they had experienced little by way of genuine warmth or concern in the past.

The most common use of supportive and directive techniques in psychodynamic psychotherapy are in the service of the therapy itself. Elaborative techniques (e.g. the simple question: 'Could you tell me more?') are undoubtedly directive in specifying a topic of interest, but at the same time may be crucial antecedents to interpretive work. Clarification stands in between supportive and interpretive interventions. It is a restatement in the therapist's words of the patient's communication. It may also be crucial in offering a verbal (symbolic) label for a confused set of internal experiences which the patient is poorly equipped to represent coherently. Confrontation is also in between a directive and an interpretive approach. At its gentlest, confrontation may involve the therapist simply identifying an inconsistency in the patient's communication and bringing this to the patient's attention. For example: 'You seem to express no sadness about this loss, yet in the past you claimed to have cared a great deal for him'.

(b) Regression

An important facet of psychoanalysis and long-term psychodynamic therapy is the activation and exploration of parts of the patient's personality which may be normally hidden behind an over-riding demand to adapt to the demands of every day life. Access to these aspects of personality is achieved through the process of regression. It has been suggested that rather than encouraging regression, the process is best conceived of as inhibiting 'an anti-regressive function' in much the same way that certain intimate interpersonal experiences, large group situations, and alcohol appear to bring out the more infantile aspects of our character.[104] Some psychoanalysts consider regression to be crucial to successful psychoanalytic treatment, but others consider the concept and its clinical application outmoded and counterproductive.[105] The extent to which a particular treatment involves significant regression appears to be a function of the patient's personality as well as the therapist's particular approach. Fear of regression is an important source of resistance to long-term psychotherapy, particularly amongst those with previous experience of psychotic episodes.[104]

(c) Resistance

Resistance is inevitably encountered in any long-term psychodynamic treatment. In fact, the presence of resistance is implied by the term dynamic, which suggests psychic forces both pulling against and pushing towards change. Like regression, resistance fluctuates in the middle stage of treatment. In borderline and narcissistic disorders, the patient's intense resistance signals the patient's desperation to protect extremely fragile self-esteem. In less severe cases, what appears to be at issue is preventing a painful integration of experience, such as the integration of love and hate directed towards the same object.[106]

In clinical practice resistance takes a variety of forms. In repression resistance, the patient may experience a temporary difficulty in gaining access to particular ideas and feelings; for example, failing to remember dreams. In transference resistance the patient may appear to wish to keep their relationship with their therapist at an extremely superficial level. In a negative therapeutic reaction the increase of symptomatology occurs alongside therapeutic progress. In Freud's formulation this may be attributed to unconscious guilt. It is quite likely that in at least some patients this form of resistance against psychotherapy is part of a pervasive so-called 'envious' predisposition to eradicate any aspect of their life that they experience as 'good' but beyond their immediate control.[107]

(d) The experience of the transference

Patients may experience a whole range of feelings about an analyst including love, admiration, excitement or anger, disappointment, and suspicion. The feelings appear to have little to do with the therapist's actual personality as different patients are likely to bring quite disparate feelings about the same analyst at the same time. While clearly not realistic, the actual nature of transference experience and its use in therapy is quite controversial.[108] Object-relations theorists consider the analyst a vehicle onto which an internal object (a person, an aspect of a person, the self, or an aspect of the self) is projected.[109] Clearly internal objects are representations which are heavily distorted by both fantasy and defensive processes.

For John Bowlby[64] transference feelings are based on expectations gathered through past relationship experience with an attachment figure. Patients resist understanding of the past relationship by insisting on repeating it. Bowlby's[110] suggestion that therapists function as secure bases implies that psychodynamic therapists are, in part, conducting attachment therapy as inevitably they serve as attachment figures for their patients. There is accumulating evidence for this claim[111–114] with a number of studies linking specific

transference schemas and attachment.[115–117] Many analysts do not accept such an isomorphism between past and present. Rather, they see it as something which gives coherence to the patient's experience of the analytic relationship—an aspect of narrative rather than a representation of the historical realities of the patient's experience.[118] In contrast, analysts who work in the Klein–Bion frame of reference see transference as providing an inevitably accurate picture of the patient's current internal world.[119] For example, a transference where the analyst is idealized may reflect psychotic anxieties in the patient linked to an intensification of the death instinct. The idealization serves to protect both the patient and the analyst from fantasized destruction which threatens to engulf them both. Marcia Cavell[120] demonstrated that these alternative models of transference have their philosophical roots in the debate between correspondence and coherence models of truth.

There is significant debate regarding from what point and how much psychoanalytic therapists should work 'in the transference'. Some analysts are inclined to see transference as pertinent to every aspect of the psychoanalytic situation. For example, Joseph[119] considers the therapeutic situation in toto as mirroring the internal state of the patient. Thus the therapeutic alliance or the 'real relationship'[121] are regarded as subsumed under the transference relationship. In this context it makes little sense to interpret anything other than the transference from the very beginning of the analysis. By contrast, Strachey[122] understood transference as an attempted externalization of the patient's superego. Unlike other people in the patient's life, the analyst does not accept this externalization, whether it is idealized, denigratory, or judgemental. The analyst conveys his or her understanding of the externalization by a so-called 'mutative interpretation'. While Strachey implied that only interpretation of the transference is therapeutic, his view clearly admits other aspects of the therapeutic relationship. Other therapists, particularly Freudian psychoanalysts, regard transference interpretations as an important but not uniquely therapeutic way of providing the patient with insight and consider the almost exclusive reliance on understanding the patient through their thoughts and feelings about their therapists as unhelpful and even dangerous.[123] The only systematic investigation of this technical controversy, where patients were randomly assigned to a transference and a non-transference-oriented psychological therapy, could not show a significant difference between the overall effectiveness of these two treatments, although there was a tendency for those with more dysfunctional object-relationship representations to do better in therapy which used transference interpretations.[124,125]

The nature of the transference appears to systematically relate to specific clinical groups and hence may have an aetiological significance. For example, specific transference patterns appear to characterize particular groups of narcissistic patients.[17] The 'mirroring' transference is one where patients crave the approbation and admiration of the therapist. This may be a consequence of the failure of the original self-objects (parents) in their mirroring function. If this transference is undermined by premature interpretations, an opportunity for restoring self-esteem is lost. The 'idealizing' transference also enables the patient to address a deficiency in self-esteem by secretly identifying with the object of admiration (the analyst). If the analyst destroys this idealized image, within Kohut's framework, this is equivalent to a direct attack on the patient's self-regard. Other analysts would suspect that behind such an exaggeratedly positive image lies the patient's true image of the analyst as frustrating or inadequate, an image which is simply placed out of harm's way by the idealization. An interesting empirical study of clinicians' experience of the transference with personality disordered patients was reported from Drew Westen's laboratory.[115] The study identified five transference dimensions: angry/entitled, anxious/preoccupied, avoidant/counterdependent, secure/engaged, and sexualized which were associated in predictable ways with Axis II pathology and confirmed that the way patients interact with their therapists can provide important data about their personality, attachment patterns, and interpersonal functioning.

Commonly, transference includes an erotic component, regardless of the age or even the gender of the analyst.[126] Admitting to such feelings may border on the unacceptable for some patients. Attachment theorists may suggest that sexual fantasies are used in the service of obtaining the attention of an unresponsive attachment figure.[127] Eroticized transference, relatively common in severely traumatized patients, represents an expression of a need for sexual gratification which, in the context of the therapy, is not considered by the patient as unrealistic.[71] Some view this phenomenon as an indication of an immature mode of representing internal reality, where only the physically observable outcome is believed to be real.[128]

(e) Experience of the countertransference

Countertransference is a somewhat controversial concept in psychoanalytic clinical work. The therapist during the course of an intensive long-term treatment is likely to have a range of feelings which are related to the patient's current experience but which may serve to either illuminate or obscure this. Some countertransference experiences may be instances of projective identification and thus can be appropriately attributed to the patient,[129] whereas others are likely to be the analyst's neurotic emotional reactions to the patient's behaviour or the material he or she brings. For Freud,[130] countertransference was always of this latter type, a neurotic reaction which was likely to obstruct psychoanalytic treatment. It was not until Paula Heimann[131] pointed out that the analyst's feelings and thoughts could contain important clues about the patient's unconscious mental state that countertransference started to be seriously considered as part of the analyst's therapeutic armamentarium. Those following an interpersonalist tradition saw the recognition of the complementarity of the therapeutic relationship as highly appropriate. From this point of view, the assumption of perfect neutrality on the part of the analyst who is a participant as well as an observer is both an anathema and an anachronism.[132] The psychotherapeutic process is more accurately viewed as a complex mixture of complementary interpersonal processes which establish themselves in 'custom designed' configurations in each treatment.[133]

The therapist's feelings may be either complementary to or concordant with those of the patient.[134] Concordant countertransferences are the product of primitive, empathic processes within the therapist who 'feels' for the patient, who may unconsciously react to experiences implied but not yet verbalized by the patient; for example, inexplicable overwhelming sadness. Complementary countertransferences tend to occur when the patient treats the analyst in a manner consistent with interpersonal interactions within a past relationship. Most commonly this occurs when the patient treats the therapist as he or she experienced being treated as a child. This is known as the 'reverse transference'.[135]

The mechanisms of countertransference are poorly understood. To assert that countertransference functions via projective identification merely brings one poorly understood phenomenon to account for a second even less well understood one. Sandler[136] suggested that an instantaneous process of automatic mirroring of one's partner in an act of communication accounted for concordant countertransference. The process, which he termed primary identification, was non-conscious and could be brought into awareness only upon reflection. Recent work on the mirror neurone system[137,138] suggests that the fundamental mechanism that allows us to understand the actions and emotions of others involves the activation of the mirror neurone system for actions and the activation of visceromotor centres for the understanding of affect. An alternative account suggests that a secondary mode of encoding is available within language whereby the use of a language of pretend gestures at the phonemic, syntactic, or even semantic level enables the communicator to address directly the unconscious of the recipient of the communication.[139] In other words, anything that can be said in gestures may be communicated unconsciously through language, through phonemic distortion, intonation, and other paralinguistic features and picked up impressionistically by the therapist.

When either concordant or complementary countertransferences mobilize defensive processes within the analyst, countertransference is in danger of becoming disruptive to therapeutic understanding. The analyst may react by unconsciously withdrawing from the therapeutic relationship. For example, in the case of a concordant countertransference where the patient's feelings of inadequacy create a similar feeling in the analyst, the analyst's vulnerability in this area may lead him or her to become defensively angry or excessively motivated to demonstrate his or her efficacy. There may be no simple way of regulating such reactions and the only reasonable strategy might be to carefully monitor one's style of relating, noting anything that is unusual. A number of analysts have pointed to the importance of reflectiveness in this context.

Some feelings in relation to the patient are not provoked either by the patient's projections or the neurotic feelings these give rise to in the therapist. It required someone of the stature of Donald Winnicott[140] to make the self-evident observation that the provocative behaviour of certain patients (particularly those in the borderline spectrum) can lead to a normal reaction of 'objective hate'. These reactions are merely indications of the therapist's humanity. Analytic understanding of these sometimes intense reactions to patients helps, but models of countertransference ill-fit such experiences. The objective study of countertransference has had to wait for a recent ingenious methodological development from Westen's laboratory.[141] The Countertransference Questionnaire yielded eight clinically and conceptually coherent factors that were independent of clinicians' theoretical orientation: (i) overwhelmed/disorganized, (ii) helpless/inadequate, (iii) positive, (iv) special/overinvolved, (v) sexualized, (vi) disengaged, (vii) parental/protective, and (viii) criticized/mistreated. Countertransference patterns were systematically related to patients' personality pathology across therapeutic approaches, suggesting that clinicians, regardless of therapeutic orientation, can make diagnostic and therapeutic use of their own responses to the patient.

(f) Interpretation

Interpretive interventions are at the core of psychoanalytic and psychodynamic treatment. However, the importance of interpretation is often exaggerated in relation to other aspects of the therapy. It is a sobering reminder that follow-up studies of long-term psychodynamic therapies invariably demonstrate that patients remember their analyst not for their interpretive interventions, rarely remembering individual interpretations, but rather for their 'emotional presence', regardless of the analyst's therapeutic perspective.[142]

Interpretations may be classified according to the aspect of a conflict they aim to address: the defence, the anxiety, or the underlying wish or feeling. Similarly, the content of the interpretation may be used in classifying interpretations: whether it relates to external reality, the transference relationship, or childhood relationships. In principle, in the earliest phases of treatment interpretations relating to current events are most common and, as the treatment progresses, transference issues and the patient's past may increasingly take over as foci of analytic work. Interpretations should start with the patient's anxiety, by identifying the defence used by the patient to protect himself from repudiated wishes and affects. In reality, these are guidelines that are rarely followed in practice. For example, very long-term treatments tend to end up being principally supportive explorations of the patient's current experience.[143] Furthermore, interpretations of the distant past tend to be least helpful to individuals with severe personality disorders.[144] Working in the so-called 'here and now' is more effective with those patients whose representation of the past is unreliable and distorted.[145]

Steiner[146] distinguished analyst-centred from patient-centred interpretations. The former refers to comments on the patient's reactions in terms of what the patient thinks may be going on in the analyst's mind, while the latter directly addresses the analyst's perception of the patient's non-conscious mental state. In either case the patient is directly learning about how minds interact in the context of social relationships. The distinction is important since when patient-centred interpretations are used exclusively the therapist may appear to be persecutory and not to be cognizant of the patient's genuine difficulties in being in an intimate relationship with another person. Others have argued, that at least in the case of severe personality disorder, interpretations, if they were to have therapeutic value, should focus on the patient's understanding of thoughts and feelings in themselves or in others at the level of what was conscious rather than unconscious, what patients could discover for themselves rather what they received as a communication from a 'mind expert'.[147] This implies that interpretation of the transference is about helping the patient represent their own and their therapist's mental states in the treatment room in all their complexity but with a stance conveying enquiry and playful curiosity about something that is not readily knowable (the mental state of the other is always opaque) with the aim of making thinking about thoughts and feelings safe again rather than communicating powerful insights.

The idealization of the transference has led some therapists to neglect interpretation of the patient's behaviour outside of the therapy. Most clinicians now agree that a balance needs to be struck between these two approaches. Treatment which is over-focused on the transference becomes a claustrophobic enclave.[148] In certain instances, the direct communication of the therapist's experience of frustration (objective hate in Winnicott's terms) may help to break a rigid repetitive pattern in the therapy.[149] Disclosing the therapist's experience is one of the cutting edges of the relational

approach to psychodynamic therapy.[150] In cases where the therapeutic alliance falters, perhaps following an empathic failure on the part of the therapist, it turns out that the recovery of the alliance may have particular therapeutic value both in showing the possibility of repair[151] but also as an opportunity to understand misunderstanding, an ideal opportunity for the recovery of mentalization.[152]

Ending treatment

The ending of psychoanalytic therapy is often idealized in clinical descriptions. As there is little agreement on the goals of psychoanalytic therapy,[79] it is hardly surprising that there is little general agreement about when ending is appropriate. Desirable final outcomes are mostly stated in terms of the process of treatment and are thus mostly specified in theoretical terms (e.g. increased awareness of impulses and fantasies, a reintegration of aspects of the self lost through projective identification, the capacity to engage in self-analysis, etc.). All these, even if observable in the course of treatment, are only loosely related to the aims the patient might have in concluding a lengthy treatment process.

The patient's own goals tend to be outcome rather than process goals and are more easily defined: the decline of symptoms, improved relationships, greater well-being, increased capacity for work, higher self-esteem, a capacity for assertiveness. As such changes are clearly achievable without psychodynamic treatment, many psychodynamic clinicians erroneously regard such criteria for ending as superficial. Independent evidence will be required to show that the achievement of process aims results in a more permanent or general achievement of outcome aims, in order to validate process aims as an appropriate criterion for ending.

Ending itself, of course, is a process. There is significant disagreement between authors, however, as to its nature; it has been labelled among other things as a mourning,[153] a detachment,[71] and a maturation.[154] It is inevitable that there is disappointment and disillusionment at the ending of long-term therapy as what is achieved is never quite the same as what has been hoped for.[155] Also, the patient loses the object who has been available as a receptacle for projections.[146] It is not surprising then, that symptoms sometimes return, even if only briefly, as part of the process of termination and the full benefit is not seen until some months after termination.[25] There is general agreement, however, that with these unconscious issues worked through the ending of therapy requires no special form of intervention on the part of the therapist.

Efficacy

It is often said that there are no studies on the effectiveness of psychoanalysis and long-term psychodynamic psychotherapy. In fact, this is not true. There are a number of comprehensive reviews[156–160] and they tend to come to similar conclusions. There is considerable evidence for the effectiveness of psychoanalytic approaches but definitive randomized controlled trials of its efficacy are still lacking.

The Boston Psychotherapy study[161] compared long-term psychoanalytic therapy (two or more times a week) with supportive therapy for clients with schizophrenia in a randomized controlled design. On the whole clients who received psychoanalytic therapy fared no better than those who received supportive treatment. In a partial-hospital RCT[162,163] the psychoanalytic arm of the treatment included therapy groups three times a week as well as individual therapy once or twice a week over an 18 month period.

The Stockholm Outcome of Psychotherapy and Psychoanalysis Project[164–166] followed 756 persons who received national insurance funded treatment for up to 3 years in psychoanalysis or psychoanalytic psychotherapy. The groups were matched on many clinical variables. Four or five times weekly analysis had similar outcomes at termination when compared with one to two sessions per week psychotherapy. During the follow-up period, psychotherapy patients did not change but those who had had psychoanalysis continued to improve, almost to a point where their scores were indistinguishable from those obtained from a non-clinical Swedish sample.

The German Psychoanalytic Association undertook a major follow-up study ($n = 401$) of psychoanalytic treatments undertaken in that country between 1990 and 1993.[159,167] Between 70 per cent and 80 per cent of the patients achieved (average 6.5 years after the end of treatment) good and stable psychic changes according to the evaluations of the patients, their analysts, independent psychoanalytic and non-psychoanalytic experts, and questionnaires commonly applied in psychotherapy research. The evaluation of mental health costs showed a cost reduction through fewer days of sick leave during the 7 years following the end of long-term psychoanalytic treatments. In the absence of pre-treatment measures it is impossible to estimate the size of the treatment effect.

The Research Committee of the International Psychoanalytic Association recently prepared a comprehensive review of North American and European outcome studies of psychoanalytic treatment.[157] Four case record studies, 13 naturalistic pre-post or quasi-experimental studies, nine follow-up studies, and nine experimental studies were identified. In addition, six process-outcome studies were also reviewed. The committee concluded that existing studies failed to demonstrate unequivocally the efficacy of psychoanalysis relative to either alternative treatment or active placebo. Studies showed a range of methodological and design problems including absence of intent to treat controls, heterogeneous patient groups, lack of random assignments, failure to use independently administered standardized measures of outcome, etc.

Another overview[168] suggested that psychoanalytic treatments may be necessary when other treatments proved to be ineffective. The authors concluded that psychoanalysis appears to be consistently helpful to patients with milder disorders and somewhat helpful to those with more severe disturbances. More controlled studies are necessary to confirm these impressions. A number of studies testing psychoanalysis with 'state of the art' methodology are ongoing and are likely to produce more compelling evidence over the next years. Despite the limitations of the completed studies, evidence across a significant number of pre-post investigations suggests that psychoanalysis appears to be consistently helpful to patients with milder (neurotic) disorders and somewhat less consistently so for other, more severe groups. Across a range of uncontrolled or poorly controlled cohort studies, mostly carried out in Europe, longer intensive treatments tended to have better outcomes than shorter, non-intensive treatments (demonstration of a dose-effect relationship). The impact of psychoanalysis was apparent beyond symptomatology, in measures of work functioning and reductions in health care costs. Studies report results which other psychotherapies have not been able to achieve; some studies show very long-term benefits from psychoanalytic treatment; the

results tend to be highly consistent across studies; some of the populations studied have been larger than most better controlled treatment trials. So whereas it is true to say that little that is definite can be stated about the outcome of psychoanalysis, a number of suggestive conclusions may be drawn and these are listed below.

Across a number of studies and measures psychoanalysis has been shown to benefit the majority of those who are offered this treatment[169] and can bring the functioning of a clinical group to the level of the normal population.[167] Completed treatments tend to be associated with greater benefits.[170] On the whole longer treatments have better outcomes[171] and intensive psychoanalytic treatment is generally more effective than psychoanalytic psychotherapy,[25] but its superiority sometimes only becomes apparent on long-term follow-up.[172] Psychoanalysis can lead to a reduction in health care related use and expenditure[173] and this is maintained for a number of years after therapy ends[174] but it does not invariably achieve this.[166] Psychoanalytic treatment can lead to a reduction in the use of psychotropic medication amongst inpatients.[175] Long-term psychoanalytic therapy can reduce symptomatology in severe personality disorders such as BPD[162,176,177]and these improvements are maintained.[163]

Training

Training in psychoanalytic psychotherapy and psychoanalysis has three components: a personal psychoanalytic psychotherapy, theoretical training, and supervised clinical practice. A variety of trainings are available, although in most countries there is only one training organization that is recognized by the International Psychoanalytic Association. Training is long, chiefly because of the length of supervised treatments. Training standards are carefully monitored by national and international bodies.

Conclusion

Psychoanalysis is hardly a practical treatment alternative for the twenty-first century. The principles derived from this treatment, however, have powerfully influenced other psychotherapeutic approaches, whether long-term or short-term therapy or psychiatric care more generally, particularly in the United States. At the time of its invention, it was the unique effective psychosocial treatment method for psychiatric disorder which offered a genuine alternative to the sometimes barbaric and generally ineffective treatment methods available. Not surprisingly, its proponents adopted an almost religious zeal in defending its value against alternative approaches. While understandable, such an attitude has no place in the sophisticated evidence base underpinning multi-agency service planning. Psychoanalytic clinicians face a challenge in identifying their niche in the complex mental health care delivery systems of the twenty-first century.

Further information

Budd, S. and Rusbridger, R. (eds.) (2005). *Introducing psychoanalysis: essential themes and topics*. Routledge, London.

Fonagy, P. and Target, M. (2003). *Psychoanalytic theories: perspectives from developmental psychopathology*. Whurr, London.

Yeomans, F.E., Clarkin, J.F., and Kernberg, O.F. (eds.) (2002). *A primer of transference-focused psychotherapy for the borderline patient*. Jason Aronson, Northvale, NJ.

References

1. Fonagy, P. and Target, M. (2003). *Psychoanalytic theories: perspectives from developmental psychopathology*. Whurr, London.
2. Smith, H.F. (2003b). Conceptions of conflict in psychoanalytic theory and practice. *The Psychoanalytic Quarterly*, **72**, 49–96.
3. Sandler, J. (2003). On attachment to internal objects. *Psychoanalytic Inquiry*, **23**, 12–26.
4. Bond, M. (2004). Empirical studies of defense style: relationships with psychopathology and change. *Harvard Review of Psychiatry*, **12**(5), 263–78.
5. Lenzenweger, M.F., Clarkin, J.F., Kernberg, O.F., *et al.* (2001). The inventory of personality organization: psychometric properties, factorial composition, and criterion relations with affect, aggressive dyscontrol, psychosis proneness, and self-domains in a nonclinical sample. *Psychological Assessment*, **13**(4), 577–91.
6. Fonagy, P., Target, M., and Gergely, G. (2006). Psychoanalytic perspectives on developmental psychopathology. In *Developmental psychopathology: theory and methods*, Vol. 1 (2nd edn) (eds. D. Cicchetti and D.J. Cohen), pp. 701–49. John Wiley & Sons, Inc., New York.
7. Brumbaugh, C.C. and Fraley, R.C. (2006). Transference and attachment: how do attachment patterns get carried forward from one relationship to the next? *Personality and Social Psychology Bulletin*, **32**(4), 552–60.
8. Adler, G. and Buie, D. (1979). Aloneness and borderline psychopathology: the possible relevance of some child developmental issues. *The International Journal of Psycho-analysis*, **60**, 83–96.
9. Eagle, M. (2003). Clinical implications of attachment theory. *Psychoanalytic Inquiry*, **23**(1), 27–53.
10. Mikulincer, M. and Shaver, P.R. (2004). Security-based self representations in adulthood: contents and processes. In *Adult attachment: theory, research and clinical implications* (eds. W.S. Rholes and J.A. Simpson), pp. 159–95. Guilford, New York.
11. Winnicott, D.W. (1958). The capacity to be alone. In *The maturational processes and the facilitating environment*, pp. 29–36. International Universities Press, New York, 1965.
12. Westen, D. and Gabbard, G.O. (2002). Developments in cognitive neuroscience. II. Implications for theories of transference. *Journal of the American Psychoanalytic Association*, **50**(1), 99–134.
13. Hartmann, H. (1939). *Ego psychology and the problem of adaptation*. International Universities Press, New York, 1958.
14. Freud, A. (1965). *Normality and pathology in childhood: assessments of development*. International Universities Press, Madison, CT.
15. Klein, M. (1948). On the theory of anxiety and guilt. In *Envy and gratitude and other works, 1946–1963*. (eds. M. Masud and R. Khan) Delacorte Press, New York, 1975.
16. Fairbairn, W.R.D. (1952). *An object-relations theory of the personality*. Basic Books, New York, 1954.
17. Kohut, H. (1984). *How does analysis cure*? University of Chicago Press, Chicago.
18. Kernberg, O.F. (1976). *Object relations theory and clinical psychoanalysis*. Aronson, New York.
19. Rosenfeld, H. (1987). *Impasse and interpretation*. Tavistock Publications, London.
20. Aron, L. and Harris, A. (eds.) (2005). *Relational psychoanalysis: innovation and expansion*, Vol. II. Analytic Press, Hillsdale, NJ.
21. Sullivan, H.S. (1953). *The interpersonal theory of psychiatry*. Norton, New York.
22. Mitchell, S.A. and Aron, L. (eds.) (1999). *Relational psychoanalysis: the emergence of a tradition*. Analytic Press, Hillsdale, NJ.
23. Seligman, S. (2003). The developmental perspective in relational psychoanalysis. *Contemporary Psychoanalysis*, **39**, 477–508.
24. Fonagy, P. (2003). Some complexities in the relationship of psychoanalytic theory to technique. *The Psychoanalytic Quarterly*, **72**, 13–48.

25. Sandell, R., Blomberg, J., and Lazar, A. (2002). Time matters. On temporal interactions in psychoanalysis and long-term psychotherapy. *Psychotherapy Research*, **12**, 39–58.
26. Freud, S. and Breuer, J. (1895). Studies on hysteria. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 2 (ed. J. Strachey), pp. 1–305. Hogarth Press, London.
27. Castelnuovo-Tedesco, P. (1994). On rereading the case of Anna O: more about questions that are unanswerable. *Journal of the American Academy of Psychoanalysis*, **22**, 57–71.
28. Freud, S. (1899). Screen memories. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 3 (ed. J. Strachey), pp. 301–22. Hogarth Press, London.
29. Fonagy, P. and Target, M. (1997). Perspectives on the recovered memories debate. In *Recovered memories of abuse: true or false*? (eds. J. Sandler and P. Fonagy), pp. 183–216. Karnac Books, London.
30. Freud, S. (1912a). The dynamics of transference. In *Standard edition of the complete psychological works of Sigmund Freud*, Vol. 12, (ed. J. Strachey) pp. 97–109. Hogarth Press and the Institute of Psycho-Analysis, London.
31. Freud, S. (1914). Remembering, repeating, and working through. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 12 (ed. J. Strachey), pp. 145–56. Hogarth Press, London.
32. Jones, E. (1953). *The life and work of Sigmund Freud*, Vol. I. Basic Books, New York.
33. Celenza, A. and Gabbard, G.O. (2003). Analysts who commit sexual boundary violations: a lost cause? *Journal of the American Psychoanalytic Association*, **51**(2), 617–36.
34. Freud, S. (1937). Analysis terminable and interminable. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 23 (ed. J. Strachey), pp. 209–53. Hogarth Press, London.
35. Walter, B. (1946). *Theme and variations*. Knopf, New York.
36. Ferenczi, S. (1930). The principle of relaxation and neocatharsis. *The International Journal of Psycho-analysis*, **11**, 428–43.
37. Szecsödi, I. (2007). Sándor Ferenczi–the first intersubjectivist. *Scandinavian Psychoanalytic Review*, **30**(1), 31–41.
38. Eissler, K.R. (1953). The effect of the structure of the ego on psychoanalytic technique. *Journal of the American Psychoanalytic Association*, **1**, 104–43.
39. Guntrip, H. (1975). My experience of analysis with Fairbairn and Winnicott. *International Review of Psychoanalysis*, **2**, 145–56.
40. Ferenczi, S. and Rank, O. (1925). *The development of psychoanalysis*. International Universities Press, Madison, CT, 1986.
41. Altman, N., Briggs, R., Frankel, J., *et al.* (2002). *Relational child psychotherapy*. The Other Press, New York.
42. Freud, S. (1915). Observations on transference love. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 12 (ed. J. Strachey), pp. 157–71. Hogarth Press, London.
43. Lambert, M. (ed.) (2004). *Bergin and Garfield's handbook of psychotherapy and behavior change.* Wiley, New York.
44. Orlinksy, D.E., Ronnestad, M.H., and Willutski, U. (2004). Fifty years of psychotherapy process-outcome research: continuity and change. In *Bergin and Garfield's handbook of psychotherapy and behavior change* (ed. M. Lambert), pp. 307–90. Wiley, New York.
45. Pinsker-Aspen, J., Stein, M., and Hilsenroth, M. (2007). Clinical utility of early memories as a predictor of early therapeutic alliance. *Psychotherapy: Theory, Research, Practice, Training*, **44**, 96–109.
46. Blagov, P.S. and Singer, J.A. (2004). Four dimensions of self-defining memories (specificity, meaning, content, and affect) and their relationships to self-restraint, distress, and repressive defensiveness. *Journal of Personality*, **72**(3), 481–511.
47. Lyons-Ruth, K. (2003). Dissociation and the parent-infant dialogue: a longitudinal perspective from attachment research. *Journal of the American Psychoanalytic Association*, **51**(3), 883–911.
48. Shamir-Essakow, G., Ungerer, J.A., Rapee, R.M., *et al.* (2004). Caregiving representations of mothers of behaviorally inhibited and uninhibited preschool children. *Developmental Psychology*, **40**(6), 899–910.
49. Brenner, C. (1982). *The mind in conflict*. International Universities Press, New York.
50. Walins, D.J. (2007). *Attachment and psychotherapy*. Guilford Press, New York.
51. Steiner, J. (2000). Containment, enactment and communication. *The International Journal of Psycho-analysis*, **81**(2), 245–55.
52. Vaillant, G.E. (1992). *Ego mechanisms of defense: a guide for clinicians and researchers*. American Psychiatric Association Press, Washington, DC.
53. Klein, M. (1946). Notes on some schizoid mechanisms. In *Developments in psychoanalysis* (eds. M. Klein, P. Heimann, S. Isaacs, and J. Riviere), pp. 292–320. Hogarth Press, London.
54. Heimann, P. (1956). Dynamics of transference interpretation. *International Journal Psycho-analysis*, **37**, 303–10.
55. Bion, W.R. (1962). *Learning from experience*. Heinemann, London.
56. Greatrex, T. S. (2002). Projective identification: how does it work? *Neuro-Psychoanalysis*, **4**, 187–97.
57. Fraiberg, S. (1982). Pathological defenses in infancy. *The Psychoanalytic Quarterly*, **51**, 612–35.
58. Freud, A. (1936). *The ego and the mechanisms of defence*. International Universities Press, New York, 1946.
59. Horowitz, M.J. (1995). Defensive control states and person schemas. In *Research in psychoanalysis: process, development, outcome* (eds. T. Shapiro and R.N. Emde), pp. 67–89. International Universities Press, Madison, CT.
60. Kaye, A.L. and Shea, M.T. (2000). Personality disorders, personality traits, and defense mechanisms. In *Handbook of psychiatric measures* (ed. Task Force for the Handbook of Psychiatric Measures), pp. 713–49. American Psychiatric Association, Washington, DC.
61. Spitz, R. (1961). Some early prototypes of ego defenses. *Journal of the American Psychoanalytic Association*, **9**, 626–51.
62. Cassidy, J. and Kobak, R.R. (1988). Avoidance and its relation to other defensive processes. In *Clinical implications of attachment* (eds. J. Belsky and T. Nezworski), pp. 300–23. Erlbaum, Hillsdale, NJ.
63. Mikulincer, M. and Shaver, P.R. (2003). The attachment behavior system in adulthood: contents and processes. In *Advances in experimental social psychology*, Vol. 35 (ed. M.P. Zanna), pp. 53–152. Academic Press, San Diego, CA.
64. Bowlby, J. (1980). *Attachment and loss, Vol. 3: loss: sadness and depression*. Hogarth Press and Institute of Psycho-Analysis, London.
65. George, C. and West, M. (2001). The development and preliminary validation of a new measure of adult attachment: the adult attachment projective. *Attachment & Human Development*, **3**(1), 30–61.
66. Fraley, R.C. and Shaver, P.R. (1997). Adult attachment and the suppression of unwanted thoughts. *Journal of Personality and Social Psychology*, **73**(5), 1080–91.
67. Mikulincer, M. and Horesh, N. (1999). Adult attachment style and the perception of others: the role of projective mechanisms. *Journal of Personality and Social Psychology*, **76**(6), 1022–34.
68. Mikulincer, M., Dolev, T., and Shaver, P.R. (2004). Attachment-related strategies during thought suppression: ironic rebounds and vulnerable self-representations. *Journal of Personality and Social Psychology*, **87**(6), 940–56.
69. PDM Task Force. (2006). *Psychodynamic diagnostic manual*. Alliance of Psychoanalytic Organizations, Silver Spring, MD.
70. Freud, S. (1933). New introductory lectures on psychoanalysis. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 22 (ed. J. Strachey), pp. 1–182. Hogarth Press, London.
71. Etchegoyen, H. (1991). *The fundamentals of psychoanalytic technique*. Karnac, London.
72. Thomä, H. and Kächele, H. (1987). *Psychoanalytic practice. I. Principles*. Springer-Verlag, New York.

73. Spillius, E.B. (2001). Freud and Klein on the concept of phantasy. *The International Journal of Psycho-analysis*, **82**(2), 361–73.
74. Gergely, G. (2000). Reapproaching Mahler: new perspectives on normal autism, normal symbiosis, splitting and libidinal object constancy from cognitive developmental theory. *Journal of the American Psychoanalytic Association*, **48**(4), 1197–228.
75. Allen, J.G. (2006). Mentalizing in practice. In *Handbook of mentalization based treatments* (eds. J.G. Allen and P. Fonagy), pp. 3–30. Wiley, Chichester.
76. Britton, R. (1998). *Belief and imagination*. Routledge, London.
77. Lipsius, S.H. (2001). Working through in psychoanalytic psychotherapy: an alternative and complementary path. *Journal of the American Academy of Psychoanalysis*, **29**, 585–600.
78. Loewald, H.W. (1960). On the therapeutic action of psycho-analysis. *The International Journal of Psycho-analysis*, **41**, 16–33.
79. Sandler, J. and Dreher, A.U. (1996). *What do psychoanalysts want? The problem of aims in psychoanalysis*, Vol. 24. Routledge, London and New York.
80. Chodorow, N.J. (2003). The psychoanalytic vision of Hans Loewald. *The International Journal of Psycho-analysis*, **84**, 897–913.
81. Hoffman, I.Z. (2006). The myths of free association and the potentials of the analytic relationship. *The International Journal of Psycho-analysis*, **87**(Pt 1), 43–61.
82. Levenson, E. (1983). *The ambiguity of change*. Basic Books, New York.
83. Pine, F. (1998). *Diversity and direction in psychoanalytic technique*. Yale University Press, New Haven, CT.
84. Blatt, S.J. (2004). *Experiences of depression: theoretical, clinical and research perspectives*. American Psychological Association, Washington, DC.
85. Blatt, S.J., Auerbach, J.S., Zuroff, D.C., *et al.* (2006). Evaluating efficacy, effectiveness, and mutative factors in psychodynamic psychotherapies. In *Psychodynamic diagnostic manual* (ed. PDM Task Force). Alliance of Psychoanalytic Organizations, Silver Spring, MD.
86. Varvin, S. (2003). Which patients should avoid psychoanalysis, and which professionals should avoid psychoanalytic training? A critical evaluation. *Scandinavian Psychoanalytic Review*, **26**, 109–22.
87. Coltart, N. (1988). Diagnosis and assessment for suitability for psycho-analytic psychotherapy. *British Journal of Psychotherapy*, **4**, 127–134.
88. Kernberg, O.F. (1981). Structural interviewing. *The Psychiatric Clinics of North America*, **4**, 169–95.
89. Piper, W.E., Ogrodniczuk, J.S., McCallum, M., *et al.* (2003). Expression of affect as a mediator of the relationship between quality of object relations and group therapy outcome for patients with complicated grief. *Journal of Consulting and Clinical Psychology*, **71**(4), 664–71.
90. Shahar, G., Blatt, S.J., Zuroff, D.C., *et al.* (2003). Role of perfectionism and personality disorder features in response to brief treatment for depression. *Journal of Consulting and Clinical Psychology*, **71**(3), 629–33.
91. Leuzinger-Bohleber, M. (2002). A follow-up study critical inspiration for our clinical practice? In *Outcomes of psychoanalytic treatment. Perspectives for therapists and researchers* (eds. M. Leuzinger-Bohleber and M. Target). Whurr Publishers, London and Philadelphia.
92. Clarkin, J.F. and Levy, K.N. (2004). The influence of client variables on psychotherapy. In *Bergin & Garfield's handbook of psychotherapy and behavior change* (ed. M.J. Lambert), pp. 194–226. Wiley, New York.
93. Piper, W.E., Joyce, A.S., Azim, H.F.A., *et al.* (1994). Patient characteristics and success in day treatment. *Journal of Nervous and Mental Diseases*, **179**, 432–8.
94. Holmes, J. (2003). Borderline personality disorder and the search for meaning: an attachment perspective. *The Australian and New Zealand Journal of Psychiatry*, **37**(5), 524–31.
95. Fonagy, P., Leigh, T., Steele, M., *et al.* (1996). The relation of attachment status, psychiatric classification, and response to psychotherapy. *Journal of Consulting and Clinical Psychology*, **64**, 22–31.
96. Sandler, J., Dare, C. and Holder, A. (1992). *The patient and the analyst* (2nd edn). Karnac, London.
97. Limentani, A. (1972). The assessment of analysability: a major hazard in selection for psychoanalysis. *The International Journal of Psycho-analysis*, **53**, 351–61.
98. Kernberg, O., Clarkin, J.F., and Yeomans, F.E. (2002). *A primer of transference focused psychotherapy for the borderline patient*. Jason Aronson, New York.
99. Bateman, A.W. and Fonagy, P. (2004b). *Psychotherapy for borderline personality disorder: mentalization based treatment.* Oxford University Press, Oxford.
100. Winnicott, D.W. (1965). *The maturational process and the facilitating environment.* Hogarth Press, London.
101. Ryle, A. (2004). The contribution of cognitive analytic therapy to the treatment of borderline personality disorder. *Journal of Personality Disorders*, **18**(1), 3–35.
102. Gorman, H.E. (2002). Growing psychoanalysis. *Canadian Journal of Psychoanalysis*, **10**(1), 45–69.
103. Feldman, M. (1993). The dynamics of reassurance. *The International Journal of Psycho-analysis*, **74**, 275–85.
104. Sandler, J. and Sandler, A.M. (1994). Theoretical and technical comments on regression and anti-regression. *The International Journal of Psycho-analysis*, **75**, 431–9.
105. Inderbitzin, L.B. and Levy, S.T. (2000). Regression and psychoanalytic technique. *The Psychoanalytic Quarterly*, **69**(2), 195–223.
106. Smith, H.F. (1997). Resistance, enactment, and interpretation: a self-analytic study. *Psychoanalytic Inquiry*, **17**(1), 13–30.
107. Cairo-Chiarandini, I. (2001). To have and have not: clinical uses of envy. *Journal of the American Psychoanalytic Association*, **49**, 1391–404.
108. Smith, H.F. (2003a). Analysis of transference: a north American perspective. *The International Journal of Psycho-analysis*, **84**, 1017–41.
109. Kernberg, O.F. (1984). *Severe personality disorders: psychotherapeutic strategies*. Yale University Press, New Haven, CT.
110. Bowlby, J. (1988). *A secure base: clinical applications of attachment theory*. Routledge, London.
111. Diamond, D., Stovall-McClough, C., Clarkin, J.F., *et al.* (2003). Patient-therapist attachment in the treatment of borderline personality disorder. *Bulletin of the Menninger Clinic*, **67**(3), 227–59.
112. Farber, B.A., Lippert, R., and Nevas, D. (1995). The therapist as attachment figure. *Psychotherapy*, **32**, 204–12.
113. Mallincrodt, B., Porter, M., and Kivlighan, M. (2005). Client attachment to therapist, depth of in-session exploration and object relations in brief psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, **42**, 85–100.
114. Parish, M. and Eagle, M.N. (2003). Attachment to the therapist. *Psychoanalytic Psychology*, **20**(2), 271–86.
115. Bradley, R., Heim, A.K., and Westen, D. (2005). Transference patterns in the psychotherapy of personality disorders: empirical investigation. *The British Journal of Psychiatry*, **186**, 342–9.
116. Eames, V. and Roth, A. (2000). Patient attachment orientation and the early working alliance: a study of patient and therapist reports of alliance quality and ruptures. *Journal of Psycho-Therapy Research*, **10**, 421–34.
117. Waldinger, R.J., Seidman, E.L., Gerber, A.J., *et al.* (2003). Attachment and core relationship themes: wishes for autonomy and closeness in the narratives of securely and insecurely attached adults. *Psychotherapy Research*, **13**(1), 77–98.
118. Spence, D.P. (1982). *Narrative truth and historical truth. Meaning and interpretation in psychoanalysis*. Norton, New York/London.
119. Joseph, B. (1985). Transference: the total situation. *The International Journal of Psycho-analysis*, **66**, 447–54.
120. Cavell, M. (1994). *The psychoanalytic mind*. Harvard University Press, Cambridge, MA.

121. Hausner, R.S. (2000). The therapeutic and working alliances. *Journal of American Psychoanalytic Association*, **48**(1), 155–87.
122. Strachey, J. (1934). The nature of the therapeutic action of psychoanalysis. *The International Journal of Psycho-analysis*, **15**, 275–92.
123. Couch, A.S. (2002). Extra-transference interpretation. *Psychoanalytic Study of the Child*, **57**, 63–92.
124. Hoglend, P., Amlo, S., Marble, A., *et al.* (2006). Analysis of the patient-therapist relationship in dynamic psychotherapy: an experimental study of transference interpretations. *The American Journal of Psychiatry*, **163**(10), 1739–46.
125. Hoglend, P., Johansson, P., Marble, A., *et al.* (2007). Moderators of the effect of transference interpretation in brief dynamic psychotherapy. *Psychotherapy Research*, **17**(2), 162–74.
126. Bollas, C. (1994). Aspects of the erotic transference. *Psychoanalytic Inquiry*, **14**(4), 572–90.
127. Bowlby, J. (1977). The making and breaking of affectional bonds. II. Some principles of psychotherapy. *The British Journal of Psychiatry*, **130**, 421–31.
128. Fonagy, P., Gergely, G., Jurist, E., *et al.* (2002). *Affect regulation, mentalization and the development of the self*. The Other Press, New York.
129. Spillius, E.B. (1992). Clinical experiences of projective identification. In *Clinical lectures on Klein and Bion* (ed. R. Anderson), pp. 59–73. Routledge, London.
130. Freud, S. (1912b). Recommendations to physicians practising psychoanalysis. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 12 (ed. J. Strachey), pp. 109–120. Hogarth Press, London.
131. Heimann, P. (1950). On countertransference. *The International Journal of Psycho-analysis*, **31**, 81–4.
132. Renik, O. (1998). The analyst's subjectivity and the analyst's objectivity. *The International Journal of Psycho-analysis*, **79**, 487–97.
133. Mitchell, S.A. (1997). *Influence and autonomy in psychoanalysis*. Analytic Press, Hillsdale, N.J.
134. Racker, H. (1968). *Transference and countertransference*. Hogarth Press, London.
135. King, P. (1978). Affective response of the analyst to the patient's communications. *The International Journal of Psycho-analysis*, **59**, 329–34.
136. Sandler, J. (1993). Communication from patient to analyst: not everything is projective identification. *British Psycho-Analytical Society Bulletin*, **29**, 8–16.
137. Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, **8**(9), 396–403.
138. Rizzolatti, G. and Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, **27**, 169–92.
139. Fonagy, P. and Target, M. (2007). The rooting of the mind in the body: new links between attachment theory and psychoanalytic thought. *Journal of the American Psychoanalytic Association*, **55**(2), 411–56.
140. Winnicott, D.W. (1949). Hate in the countertransference. *The International Journal of Psycho-analysis*, **30**, 69–75.
141. Betan, E., Heim, A.K., Zittel Conklin, C., *et al.* (2005). Countertransference phenomena and personality pathology in clinical practice: an empirical investigation. *The American Journal of Psychiatry*, **162**(5), 890–8.
142. Leuzinger-Bohleber, M., Stuhr, U., Ruger, B., *et al.* (2003a). How to study the 'quality of psychoanalytic treatments' and their long-term effects on patients' well-being: a representative, multi-perspective follow-up study. *The International Journal of Psycho-analysis*, **84**(Pt 2), 263–90.
143. Blum, H.P. (1989). The concept of termination and the evolution of psychoanalytic thought. Annual meeting of the American Psychoanalytic Association (1987, Montreal, Canada). *Journal of the American Psychoanalytic Association*, **37**, 275–95.
144. Bateman, A.W. and Fonagy, P. (2006). *Mentalization based treatment for borderline personality disorder: a practical guide*. Oxford University Press, Oxford.
145. Fonagy, P. (1999). Memory and therapeutic action (guest editorial). *The International Journal of Psycho-analysis*, **80**, 215–23.
146. Steiner, J. (1993). *Psychic retreats: pathological organisations in psychotic, neurotic and borderline patients*. Routledge, London.
147. Fonagy, P. and Bateman, A. (2006). Progress in the treatment of borderline personality disorder. *The British Journal of Psychiatry*, **188**, 1–3.
148. O'Shaughnessy, E. (1992). Enclaves and excursions. *The International Journal of Psycho-analysis*, **73**, 603–11.
149. Symington, N. (1983). The analyst's act of freedom as agent of therapeutic change. *International Review of Psycho-analysis*, **10**, 783–92.
150. Ehrenberg, D. (1993). *The intimate edge*. Norton, New York.
151. Safran, J.D. (2003). The relational turn, the therapeutic alliance, and psychotherapy research: strange bedfellows or postmodern marriage? *Contemporary Psychoanalysis*, **39**, 449–75.
152. Bateman, A.W. and Fonagy, P. (2004a). Mentalization-based treatment of BPD. *Journal of Personality Disorders*, **18**(1), 36–51.
153. Klein, M. (1950). On the criteria for the termination of a psychoanalysis. *The International Journal of Psycho-analysis*, **31**, 78–80.
154. Payne, S. (1950). Short communication on criteria for terminating analysis. *The International Journal of Psycho-analysis*, **31**, 205.
155. Pedder, J. (1988). Termination reconsidered. *The International Journal of Psycho-analysis*, **69**, 495–505.
156. Bachrach, H.M., Galatzer-Levy, R., Skolnikoff, A., *et al.* (1991). On the efficacy of psychoanalysis. *Journal of the American Psychoanalytic Association*, **39**, 871–916.
157. Fonagy, P., Kachele, H., Krause, R., *et al.* (2002). *An open door review of outcome studies in psychoanalysis* (2nd edn). International Psychoanalytical Association, London.
158. Lazar, S.G. (ed.) (1997). *Extended dynamic psychotherapy: making the case in an era of managed care*. Analytic Press, Hillsdale, NJ.
159. Leuzinger-Bohleber, M. and Target, M. (eds.) (2002). *The outcomes of psychoanalytic treatment*. Whurr, London.
160. Richardson, P., Kachele, H., and Renlund, C. (eds.) (2004). *Research on psychoanalytic psychotherapy with adults*. Karnac, London.
161. Stanton, A.H., Gunderson, J.G., Knapp, P.H., *et al.* (1984). Effects of psychotherapy in schizophrenia. I. Design and implementation of a controlled study. *Schizophrenia Bulletin*, **10**, 520–63.
162. Bateman, A.W. and Fonagy, P. (1999). The effectiveness of partial hospitalization in the treatment of borderline personality disorder-a randomised controlled trial. *The American Journal of Psychiatry*, **156**, 1563–9.
163. Bateman, A.W. and Fonagy, P. (2001). Treatment of borderline personality disorder with psychoanalytically oriented partial hospitalization: an 18-month follow-up. *The American Journal of Psychiatry*, **158**(1), 36–42.
164. Blomberg, J., Lazar, A., and Sandell, R. (2001). Outcome of patients in long-term psychoanalytical treatments. First findings of the Stockholm outcome of psychotherapy and psychoanalysis (STOPP) study. *Psychotherapy Research*, **11**, 361–82.
165. Grant, J. and Sandell, R. (2004). Close family or mere neighbours? Some empirical data on the differences between psychoanalysis and psychotherapy. In *Research on psychoanalytic psychotherapy with adults* (eds. P. Richardson, H. Kächele, and C. Renlund), pp. 81–108. Karnac, London.
166. Sandell, R., Blomberg, J., Lazar, A., *et al.* (2000). Varieties of long-term outcome among patients in psychoanalysis and long-term psychotherapy: a review of findings in the Stockholm outcome of

psychoanalysis and psychotherapy project (STOPP). *The International Journal of Psycho-analysis*, **81**(5), 921–43.
167. Leuzinger-Bohleber, M., Stuhr, U., Ruger, B., *et al.* (2003b). How to study the quality of psychoanalytic treatments and their long-term effects on patients' well-being: a representative, multi-perspective follow-up study. *The International Journal of Psycho-analysis*, **84**, 263–90.
168. Gabbard, G.O., Gunderson, J.G., and Fonagy, P. (2002). The place of psychoanalytic treatments within psychiatry. *Archives of General Psychiatry*, **59**(6), 505–10.
169. Fonagy, P. (2006). Evidence-based psychodynamic psychotherapies. In *Psychodynamic diagnostic manual* (ed. PDM Task Force). Alliance of Psychoanalytic Organizations, Silver Spring, MD.
170. Bachrach, H.M., Weber, J.J., and Murray, S. (1985). Factors associated with the outcome of psychoanalysis. Report of the Columbia psychoanalytic research center (IV). *International Review of Psychoanalysis*, **12**, 379–89.
171. Erle, J. and Goldberg, D. (1984). Observations on assessment of analyzability by experienced analysts. *Journal of the American Psychoanalytical Association*, **32**, 715–37.
172. Sandell, R., Blomberg, J., Lazar, A., *et al.* (1997). *Findings of the Stockholm outcome of psychotherapy and psychoanalysis project* (*STOPPP*). Paper presented at the Annual Meeting of the Society for Psychotherapy Research, Geilo, Norway.
173. Dührssen, A. (1962). Katamnestische Ergebnisse bei 1004 Patienten nach analytischer Psychotherapie. *Psychosomatic Medicine*, **8**, 94–113.
174. Breyer, F., Heinzel, R., and Klein, T. (1997). Kosten und Nutzen ambulanter Psychoanalyse in Deutschland (Cost and benefits of outpatient psychoanalytic therapy in Germany). *Gesundheitsökonomie und Qualitätsmanagement*, **2**, 59–73.
175. Bateman, A.W. and Fonagy, P. (2003). Health service utilization costs for borderline personality disorder patients treated with psychoanalytically oriented partial hospitalization versus general psychiatric care. *The American Journal of Psychiatry*, **160**(1), 169–71.
176. Clarkin, J., Levy, K.N., Lenzenweger, M.F., *et al.* (2007). Evaluating three treatments for borderline personality disorder: a multiwave study. *The American Journal of Psychiatry*, **164**, 922–8.
177. Giesen-Bloo, J., van Dyck, R., Spinhoven, P., *et al.* (2006). Outpatient psychotherapy for borderline personality disorder: randomized trial of schema-focused therapy vs transference-focused psychotherapy. *Archives of General Psychiatry*, **63**(6), 649–58.

6.3.6 Group methods in adult psychiatry

John Schlapobersky and Malcolm Pines

Introduction

After a century of development, group therapy is today one of the most widely practised treatment methods in psychiatry with an extensive literature. There are three principles common to its wide range of applications. First, the therapist calls the 'community' into the consulting room where, together with the therapist, it becomes the therapeutic agent. Second, the therapist assembles a group of people who can contribute to a commonly held resource from which its members can each derive benefits. And third, the therapist does nothing for them in the context of the group, that they can do for themselves, and one another.

This chapter starts by providing a conceptual framework that differentiates methods, models, and applications for the practice of group therapy in adult psychiatry. After classification of the different methods and applications we discuss the main theoretical models; explore the dynamic life of therapy groups; consider some of the key clinical issues facing practitioners; their applications to a range of patient populations and settings; their evaluation and justification and their historical evolution this century. In the conclusion we consider the planning of group services and the training of their practitioners. This revision of the chapter has brought it up-to-date with the contemporary literature in a field that has seen a great deal of innovation since the original 2000 edition.

The developing evidence base for group psychotherapy is 'Guardedly optimistic. The literature has become stronger and deeper and is capable of supporting evidence-based treatment recommendations for some patient populations.' The evidence base for the effectiveness of group psychotherapy has been growing with the field. Some 700 studies, spanning the past three decades, have shown that the group format consistently produced positive effects with diverse disorders and treatment models.[1] These show that both individual and group psychotherapy will effect much the same set of results. For group therapy to be effective it has to utilize those therapeutic factors originally laid out by Foulkes[2] and later by Yalom[3]—the group has to be the primary focus of therapy; patients need to be well selected; and therapists need to be adequately trained. The chapter will address these questions of focus, selection, and training.

Although the two authors of this chapter are both group analysts, we have set out to provide a full account of the wide range of group work practice. The United Kingdom is our own working location which lends emphasis to the chapter but it is compiled with sources and references that address the international field and it gives attention to current literature in many countries including North and South America and Continental Europe.

Basic methods

In Fig. 6.3.6.1 we have used two simple factors—therapeutic goals and group leadership—to provide a simple classification of the many different methods.

Therapeutic goals

Groups will be more or less specific in their therapeutic goals. For example, those catering for a homogeneous population with a commonly defined problem whose solution provides the basis for entry to the group—such as overcoming drink or drug dependence—are classified here with specific goals. Groups that provide psychoanalytic psychotherapy, whether run according to Interpersonal, Tavistock, or Group-Analytic models, are classified here with non-specific goals. There is a wide range of variation between these extremes and within each of these main psychodynamic models.

Leadership

The more the leader directs the group, the more prominent he becomes as the group's 'model object'. The less the leader directs the group, the more scope there is for the emergence of unconscious dynamics and for attention to transference and counter-transference. In this case, therapy progresses through the development of relationships. The greater the leadership activity, the more likely it

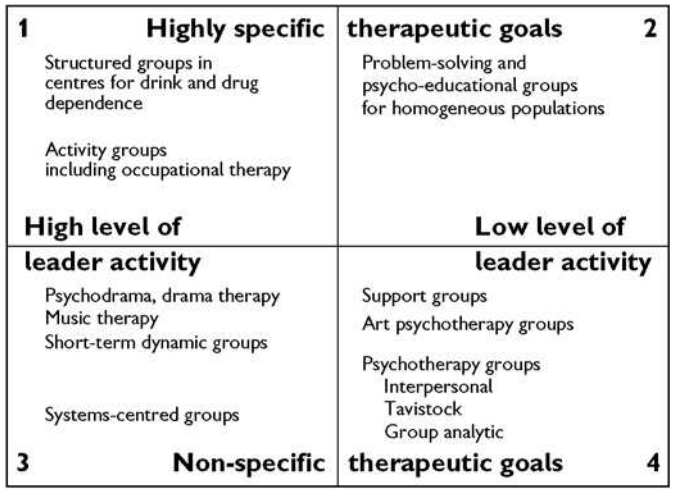


Fig. 6.3.6.1 A simple classification of group methods.

is that group members are being offered a technique or skill in the setting of a group. The lower this activity, the stronger will be the relational content of the therapy and the therapist's skills of fostering relationships will equip the group to work developmentally and in depth on both the obvious and hidden issues that its members bring. The three principal psychodynamic methods discussed in The principal model of psychodynamic group therapy below share a non-directive philosophy with subtle but significant differences between their models of practice.

Using these two basic indicators—specificity of goals and levels of leadership activity—the four quadrants in the diagram provide us with a simple way of 'placing' the different group therapies.

Goal-specific therapy with high level of leader activity: quadrant 1

In many drug and alcohol dependency regimes, participants are required to fulfil obligations tied to each stage of a structured programme. They move forward when stage-specific obligations are fulfilled. As the novice moves up, he/she becomes a trainer to the newcomers with the therapist(s) directing the process in active terms. Cognitive therapy given in a group setting uses the group as an assembly who learn from and discuss with the expert. Dependency on a shared and valued leader, and attention to group dynamics amplify the learning and some group cohesion develops, but this is not the primary focus of the therapy.

Goal-directed with lower level of leader activity: quadrant 2

Problem-solving or psycho-educational groups for homogeneous populations, such as those set up for eating disorders or offenders, which are run along analytic lines, can be placed in this category. Although there are clear and directed goals, the leader's level of activity is confined to a facilitating, linking, or enabling one, followed by analysis and interpretation. Group discussion and cohesion amplify the affective experience and enhance the learning.

Non-specific goals with high level of leader activity: quadrant 3

In **psychodrama groups**, leadership is explicitly vested in the psychodrama director. The needs and goals of group members are diffuse and often diverse and, in psychiatric practice, will have to do with relief from mental suffering. The psychodrama director can draw on many techniques. Affective arousal can be high so the power of sharing through discussion and the sympathy; and empathy of group members towards one another become powerful therapeutic tools. Strong conflict arousal and its subsequent resolution is similarly therapeutic.

Systems-centred therapy (as developed by Agazarian)[4] similarly provides a high level of leadership activity for groups that have non-specific goals. Short-term dynamic groups are frequently constituted with non-specific goals but are run over 10 or 20 sessions by leaders who maintain a high level of involvement and direction, often demarcated according to the different stages of the group's progress.

Non-specific goals with low level of leader activity: quadrant 4

The goals of group-analytic or psychoanalytic group therapy are most frequently diffuse and non-specific involving relief from symptoms and other forms of suffering; personal growth; and psychological change. There are three main schools considered in this chapter—the Interpersonal (Yalom), Tavistock (Bion or group-as-a-whole), and the Group-Analytic. They share non-specific goals and have low levels of leader activity but differ from one another in how the leadership role and function is understood and discharged. They share assumptions about the importance of unconscious individual and group dynamics and look to the group for its transformational potential. Their differences affect the way in which transference and counter-transference is understood and worked with. There is a comparative appraisal of these models below.

The application of basic methods

We now provide a more detailed overview of the current field and offer a brief set of training requirements for practitioners in each of the methods discussed.

The field can be divided into **five basic methods**—activity, supportive, problem-solving, psycho-educational, and psychodynamic. The first three methods are goal-specific as indicated by their descriptions, the fourth is less specific and the fifth is a non-directive analytic psychotherapy. In supportive and problem-solving groups, therapeutic leadership can be highly directed or not, depending on the approach. Activity and psycho-educational groups will inevitably have a high level of directed group leadership, whilst psychodynamic groups have a much lower level of directed group leadership. All five methods rely on the same basic procedures—the selection and grouping of a number of people seeking help who have regular meetings together with one or more well-trained therapist(s).

Activity groups

The most vulnerable and disturbed patients can be placed in therapy groups defined by an activity that provides a convening function such as exercise or cooking. They can then be used to create conditions for a wide range of secondary functions that foster affiliations, develop social skills, address unspoken anxieties, and express troubling emotions. Occupational therapists and nurses using art media or other socially syntonic activity like gardening or

hair-dressing have been developing a wide range of group services in both acute and rehabilitation psychiatry for many years.[5] The approach has been used in a wide range of other settings including medical rehabilitation, rehabilitation with refugees, social work and fostering, and adoption programmes. Groups that keep the original activity as their primary focus, working with art media for example, need to be differentiated from those which use such media to develop an analytic focus on psychological work. The arts psychotherapies belong to this latter group. They have non-specific therapeutic goals and might, as in the case of music therapy, have a high level of leadership activity or, in the case of art therapy, have a low level of leadership activity.

Whilst therapists do not engage in the uncovering and exploration of unconscious dynamics, they will need leadership abilities, capabilities in organizing group activities, and should have a basic understanding of psychopathology and group dynamics.

Supportive groups

These groups function as a form of social support providing containment, the improvement of social skills, and the enhancement of participants' capacities for social adaptation. They aim to reduce the deleterious effects of social isolation, bring people out of withdrawal into a social context, and provide opportunities for problem sharing. They cater to patient populations with long-standing personality disorders not open to uncovering exploration; those with chronic mental and physical illness,[6] physical handicap, mental retardation, and carers for those with any of these problems. They will often allow a certain amount of psycho-education with the group leader influencing members' attitudes as in the case, for example, of a group for young sexually active adults with learning disability who might receive guidance on contraception.

Whilst therapists do not engage in the uncovering and exploration of unconscious dynamics, they will need leadership abilities, capabilities in organizing group activities, and should have a basic understanding of psychopathology and group dynamics.

Problem-solving groups

Group therapy is provided for a set of referral criteria to resolve a defined and sometimes circumscribed problem. Alcoholics Anonymous, Alanon, Gamblers Anonymous, and groups for people with poor impulse control, eating disorders, or other habitual problems such as smoking, are a few of the examples. These groups can take on many of the features of long-term support groups, in that they offer ego-supportive and adaptive resources, providing an extended service for monitoring by the patient or by professionals, without necessarily committing members to the deeper and more radical analytic work entailed by a psychodynamic group. In many cases, the problem-solving focus provides a convening frame by which to engage a population who are soon drawn into psychodynamic work that sees them through profound changes. Many of the groups run by clinicians in primary hospital care—occupational therapists, nurses, doctors, and psychologists—take this form. The Group Work Programme at the Medical Foundation For Victims Of Torture in London is another example (see also section on trauma in specical population, below).

Where they cater for the more severely disturbed, staff will need to be well-trained in one of the core professions. They will need leadership abilities, capabilities in organizing group activities, should have a basic understanding of psychopathology, and be sufficiently well-trained to explore the dynamic group issues that lead from the problem back to the personality structure of their membership. If therapeutic goals involve major changes in personality and social functioning, this will involve the uncovering and exploration of unconscious dynamics.

Psycho-educational groups

The original groups for servicemen with war neurosis at Northfield took this form in which people were given the role of students of their disorders rather than 'sick' people. Patients become open to new information and are better able to unlearn maladaptive attitudes about the nature of their disorder. This more cognitive approach can be applied in homogeneous problem-solving groups. Information can be provided through lectures, discussions, and suitable reading material. There are different ways of lowering anxiety and uncovering maladaptive and inappropriate attitudes towards such problems as anxiety states, phobias and obsessions, and psychosomatic disorders. Many of the groups run for those with serious physical illness (see also section on the medically ill in special populations, below) take this form. And there is often a major psycho-educational component in support groups—for example, those with chronic mental illness who can be helped to understand and cope with delusions, hallucinations, and the stigma of illness.[7] (see also section on the mentally ill in special populations, below)

Staff need leadership abilities; capabilities in organizing group activities, a basic understanding of psychopathology, and need to be sufficiently well-trained in their chosen problem area to be able to relate its educational focus to thematic group issues. If therapeutic goals involve major changes in personality and social functioning, this will involve the uncovering and exploration of unconscious dynamics.

Psychodynamic groups

There are supportive, problem-solving, and psycho-educational components in all psychodynamic groups, but the description 'psychodynamic' is reserved for those in which the declared goal is lasting personal change through a non-directive, free-associative therapy. The range of different group contexts is so varied that—at first sight—they might appear to have little in common. But there will be common principles offering therapy to a group of people on an in-patient unit recovering from psychosis and meeting thrice weekly; those in a secure unit for violent offenders meeting once weekly; and those—including mental health trainees—attending a group in private practice once or twice weekly. These principles can be summarized in the table, Table 6.3.6.1 below.

Within these parameters therapy is part of the cultural domain of all shared, conversational experience in which people struggle with meaning—in congregational life, in the confessional, in theatre, narrative or poetry.[8,9]

Staff need to be trained to the level already described. They need good leadership ability, capabilities in organizing group activities, and a good understanding of psychopathology. Beyond these requirements, therapeutic goals will involve major changes in personality and social functioning involving the uncovering and exploration of unconscious dynamics. So staff will need access to a

Table 6.3.6.1 Organizing principles for group therapy

1	Members will have been chosen by the therapist.
2	They will have chosen to join and participate.
3	They will be expected to justify their place by reliable attendance and participation.
4	The work will be governed by a psychotherapeutic contract.
5	The contract will include definitions of confidentiality and other boundaries.
6	A therapeutic alliance with individual members will be established either prior to their joining the group or during the early stages of their attendance in the group.
7	Agreed parameters will include the duration and time-boundaries of the group as well as its membership and composition.
8	Groups can be homogeneous or heterogeneous.
9	Groups may have a fixed time limit or continue on a slow–open basis for many years.
10	Groups may have a stable and fixed, or a rotating membership with empty places taken by new members.

range of specialized training opportunities which should provide psychodynamic theory, clinical supervision and, ideally, some opportunity for the practitioners' own personal development.

The contemporary field and its history

The paradigm shift that led to a vision of the group as a whole began over a period of time and in a number of locations. Trigant Burrow coined the term group analysis in the USA in the 1920s.[10,11] Further development in the years after World War II period enabled workers to recognize the dynamics of groups and institutions and led to group and family therapy, milieu therapy, and therapeutic community concept and practice.

Second World War

In Britain and the US, during the Second World War, an appreciation of group psychology led to a wide range of innovations, the most important of which included:

- The use of group methods for selection and allocation of work responsibilities
- Studies of group morale
- The integration of psychiatric knowledge to the management of large groups through the role of the command psychiatrist
- The treatment of acute and prolonged battle stress and the rehabilitation of returned prisoners of war.

Clinicians used the opportunities created within army psychiatry to apply methods developed in the pre-war years.[12] Brigadier J.R. Rees, Director of the Tavistock Clinic in the 1930s, largely created this opportunity in the British Army. His Tavistock colleagues formed an 'invisible college' and were responsible for signal achievements in the advances in selection and treatment.[13]

Foremost amongst these were Northfield, the military hospital near Birmingham where S.H. Foulkes was a senior medical officer.[14] A refugee from Nazi Germany, Foulkes brought with him from the Frankfurt Institute the revised understanding of Freudian theory that was also to prove influential in the US. In the New School for Social Research, New York, and in the work of Neo-Freudians like Erich Fromm, Frieda Fromm-Reichman, and social theorists like Adorno, Marcuse, and Norbert Elias, psychoanalysis and Marxist theory were brought into a new, creative relationship.[15] In the United Kingdom Foulkes first developed his approach to group therapy in Exeter before the war and was able to apply it successfully on a large scale to the treatment of war neuroses at Northfield.

Post-war period

Group psychotherapy moved from inspirational and didactic models to psychodynamic and analytic ones in the post-war period.

(a) United Kingdom

The Group-Analytic Tradition: Foulkes gathered around him a small group of clinicians and others who developed his ideas and practises. Drawing on the ideas of Trigant Burrow, they called it group analysis and later established the Group-Analytic Society, and trained generations of clinicians. His first book written in the heat of the Northfield experience outlined the basics of his approach.[16] Other publications followed and, with Malcolm Pines, training courses were established which lead to the founding of the Institutes of Group Analysis and Family Therapy, the Association of Family Therapists, and Association of Therapeutic Communities. There are now training courses in group analysis in many centres in the United Kingdom and continental Europe. The Journal, *Group Analysis*, established by Foulkes, continues to be the major publication in European group psychotherapy. Group-analytic psychotherapy has undergone clinical evaluation by a number of clinicians.[17–19]

The 'Tavistock' Approach: The approach originates in the work of Bion, Ezriel, Sutherland, and their colleagues. It shares with group analysis an interest in the underlying pattern of object relations in groups but, under the influence of Bion—its major exponent—his 'basic assumption theory' is applied to the exclusion of almost everything else. (Bion's monograph was his only publication on groups and marked the end of his interest in the subject.[20]) The approach has been especially influential in staff training and consultancy which, given the slender theoretical foundations on which it rests, suggests a wide responsiveness in the field of basic assumption theory. The approach has undergone further development in the United States where it is often referred to as group-as-a-whole. When employed as a therapy, it can overlook the individuality of a group's members, disturbing some patients whose experience of the group situation can repeat early developmental traumas of neglect and misunderstanding by caretakers. Malan's study of effectiveness[21] raised serious questions about the model's efficacy in its clinical applications but its training applications continue to influence the field.

(b) United States

(i) Early pioneers

Jacob Moreno was the innovator of group psychodrama, a pioneer form of group psychotherapy.[22] He also introduced sociometry, a scientific method for the study of group affiliations and conflicts, widely accepted and used by social psychologists. **Slavson** was an educationalist of psychoanalytic persuasion who became the central figure in the early development of group psychotherapy. His clinical influence, particularly with groups for the parents of children in difficulty, and his focus on the dynamics of projection in groups has been of lasting importance.[23] His organizational efforts lead to the formation of the American Group Psychotherapy

Association. **Emanuel Schwartz** began to apply psychoanalytic ideas to group psychotherapy in the late 1930s and was later joined by **Alexander Wolf**.[24, 25] In their approach, people underwent an individual psychotherapy in the setting of a group, a kind of parallel process alongside their fellow patients, with attention focused on the transferential relationship between each individual and their therapist. The approach has been of lasting importance in creating a clinical framework for combining individual and group therapy. Foulkes' criticism at the time was that the approach overlooked any systematic use of group-specific process. In contrast to their 'psychotherapy in the group' he offered group analysis as a clinical alternative, describing it as 'psychotherapy by the group'.

(ii) Irving Yalom

Yalom's interpersonal approach is influenced by the interpersonal psychotherapy of Sullivan and Frank. His *Theory And Practice of Group Psychotherapy,* now in its fifth edition and written jointly with Leszcz, is the first systematic account of groups informed by research and remains one of the most influential books in the field.[3] Yalom's later text on inpatient group psychotherapy systematized group work in that setting.[26]

(iii) The contemporary field

There are many centres of excellence, a wide range of methods and models and an empirical base grounded in research. The most useful single text is by Rutan and Stone, now in its fourth edition.[27] Collections by Kaplan and Sadock[28] and by Alonso and Swiller[29] cover the field. Psychoanalytic models have a rich diversity of theory with contributions from object relations, self psychology, and social systems theory.[30] The Modern Group movement is amongst the most innovative, beginning with a classic text by Spotnitz[31] and developing through an active training programme, a journal, *The Modern Group,* and publications by Ormont[32] and others.

(iv) South America

There is a vigorous field of development throughout South America that draws on both the Tavistock and Group Analytic traditions but is informed by independent sources based largely on the work of Pichon-Riviere. Tubert-Oklander and Hernandez de Tubert have introduced this approach, referred to as Operative Groups, to the English-speaking world.[33]

(v) Continental Europe

Group methods have played an active part in the reconstruction of mental health services throughout Europe in the post-war period. Distinctive approaches are emerging. Those in Germany include psychosomatic practice[34] and the Gottingen model.[35] The journal of the Heidelberg Institute of Group Analysis gives access to a vigorous field. A major research study is in process, based in Germany, in which therapists throughout Europe are taking part and which aims to provide a detailed evaluation of group therapy, its patients, and its therapists.[36] Other distinctive developments include those in Italy[43] and original training models, for example, those in Greece[37] and Norway.[38]

Principal models of psychodynamic group therapy

The therapist

The therapist is responsible to the group— and to the institution in which it is set—for achieving and maintaining professional competence and should have a level of training appropriate to the task. A formal qualification in psychotherapy is the ideal training. This will have included theory, personal therapy for the therapist, and clinical supervision. Mental health professionals from all disciplines make an active contribution to a rich and diverse service with the training requirements of theory and supervision arranged at their workplace. The opportunity to run a group is provided in most psychiatric and psychology training programmes. Many centres and training institutions offer training in group methods and several are wholly committed to the training of group therapists who have their own professional associations in the UK and internationally. Private once- or twice-weekly analytic groups are now regarded by many mental health professionals as the therapy of choice for their personal development. Other requirements for a good therapist are listed in table 6.3.6.2.

(a) Making a beginning

The establishment of a group begins as a management task in the definition of its goals, recruitment of its members, protection of its setting, venue and timetable, and in the maintenance of its ongoing life. It evolves as a therapeutic task in which the therapist is responsible for maintaining a therapeutic attitude to the individual members and to the group as a whole. Powerful affects and attitudes will be directed towards her which she will monitor and transform into verbal and non-verbal therapeutic responses.

The therapeutic rationale will allow the therapist to be discriminating and consistent about interventions of various kinds during the life of the group and what follows below provides an orientation—based on the dynamic elements of structure, process and content—to the three main models used in the UK.

Structure, process, and content: the dynamic elements of a group

Regardless of the therapist's method, people usually start in groups with a form of serial monologue. Out of this arise the capacities to talk and listen that are often undeveloped or even non-existent at the outset of therapy but which are its core constituents. From talking and listening comes self-disclosure and out of this social exchange identification emerges, which in due course leads to dialogue and differentiation. So the conductor must give a place to monologue whilst, at the same time, cultivating dialogue—the exchange between members or sub-groups—and, ultimately,

Table 6.3.6.2 Requirements for therapeutic competence

Requirements for therapeutic competence include:
1 The ability to follow complex interactions and processes.
2 The ability to discriminate between appropriate activity and a containing form of silence.
3 A reflective attitude and the capacity to consider and reflect upon the processes concerning both the individual members and the group as a whole.
4 An eye for both the visible and invisible group and a curiosity about the unconscious or otherwise hidden aspects of a group's life. This will require the therapist's access to their own internal process and a capacity to make use of it.
5 A therapeutic rationale for action related to the group tasks and leadership requirements including the psychopathology of the individuals, their psychodynamics, and group dynamics.

promoting discourse, defined here as the free interaction of participants in the flexible and complex exchange that distinguishes the communication of a group.

Structure describes the more enduring aspects of any group's makeup, the 'architecture' of its interpersonal relations conceptualized first in terms of the setting and its boundaries and then conceptualized in the bond between each individual, the therapist(s), and the group as a whole. Process describes the fluid and dynamic fluctuations of emotion and experience, the business of relating and communicating, the changes of association and inter-member responses. The content of a group's exchange is in its visible and audible events, in the narrative line and dramatic content of peoples' encounters, the topics raised, their thematic development, and the extent to which they are explored or avoided.

As Fig. 6.3.6.2 illustrates, each of these three dynamic elements has a determining influence on each of the others. For example, a group in which there was a problem caused by the institution's failure to honour its commitment to reliable space for regular meetings, would have a serious structural problem. Intrusion, relocation, or a conflict over space might then emerge in the content of the members' associations as they talked about shared past experience. The therapist would need to decide whether to direct the process towards the connection between past and present anxieties, or reassurance that the therapist would—from now on—be able to protect their space.

(a) Overview of the Interpersonal, Tavistock, and Group-Analytic Models

In the Interpersonal school, intra-group interactions, including those between its members and the leader, are taken in their totality, but differentiating the leader as a different 'sort' of person from the others. In the Tavistock model, a two-body psychology is used to analyse the interchange between the leader and the group taken as a whole. The therapist's principal role is in the analysis and interpretation of defences against primitive anxieties (or basic assumptions). The Group-Analytic Model calls on elements of both foregoing models. Like the Tavistock model it considers the leader as structurally different to other group members but like the Interpersonal model, it encourages the leader to work in the group with individuals. A three-body psychology is used to understand the role of the leader who is referred to here as the conductor. Therapy proceeds through the dynamic interaction between each individual, the conductor, and the group as a whole.

(b) The model of interpersonal group therapy

The focus is on interpersonal learning as a primary mechanism of change. The group provides the antidote to maladaptive interpersonal beliefs and behaviours through feedback from others and encouragement to experiment with healthier behaviours first within the group and then outside. The joint examination of intra-group transference reactions allows members to replace processes that have a historical origin in the 'there and then', the dynamic past, with those more appropriate to the 'here and now', the dynamic present. The approach emphasizes the educational opportunities of working in the 'here and now' of the group. The therapist takes the responsibility for leading the group towards awareness of these interpersonal dynamics and their expressions. There is also greater therapist transparency than in other psychodynamic approaches with the therapist modelling desired behaviours, sharing the reactions to events in the group directly, and being open to feedback from other group members.

As the diagram indicates, interpersonal dynamics are kept at the forefront of members' attention by the therapist. This sets a pattern in which the content of members' discussions and the process of their interactions gives the group its agenda. The interpersonal approach places the therapist amongst other members of the group without giving him a distinctive structural identity and omits any formal demarcation for the boundaries of the group as a whole.

The model provided the early descriptive research on the phases of small groups, on the basis of which Yalom tabulated the **curative factors** in a group's life (see Table 6.3.6.3).

This construction has been very influential. Yalom's use of the term 'curative' poses many problems for those clinicians who see the goals of therapy involving personal growth and change. He did much to address this difficulty by singling out the last of his 'curative' factors—existential issues—for special treatment in a subsequent text.[49]

(c) The Tavistock model

Bion's ideas have an explanatory power and simplicity of application that continues to prove illuminating.[40] In a group at any

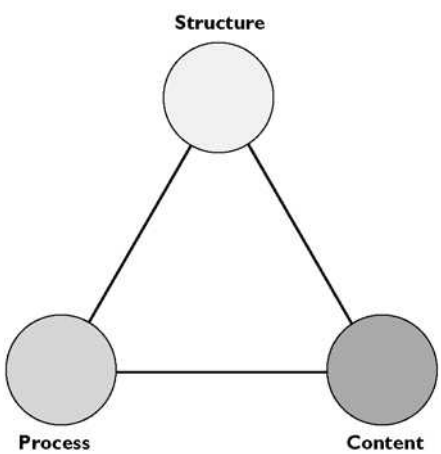


Fig. 6.3.6.2 The dynamic elements of a group.

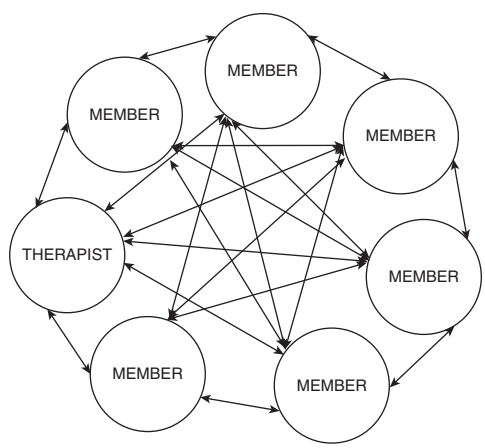


Fig. 6.3.6.3 The elements of an interpersonal group.

Table 6.3.6.3 Yalom's curative factors

1	Instillation of hope
2	Universality
3	Imparting information
4	Altruism
5	Corrective recapitulation of primary family group
6	Development of socializing techniques
7	Imitative behaviour
8	Interpersonal learning
9	Group cohesiveness
10	Catharsis
11	Existential factors

point in time, its culture and climate are governed by primitive, unconscious anxieties that impede its capacities for rational work in which the person or representation of the leader plays a crucial part. The anxieties, organized into one of three categories, referred to as **basic assumptions**, are dependency, fight or flight, and pairing. They affect the group as a whole in which only one basic assumption is believed to be operative at any point in time. Bion saw basic assumptions as interfering with the 'work group', the more rational, higher-level functioning of the group and its members. The therapist's key task lies in understanding and interpreting the operative basic assumption to the whole group. The meaning of individuals' experience is subsumed by this understanding of the whole. This therapist-centred approach sees transference only as directed towards the therapist who represents authority. In **dependency**, the group tries to elicit protection through passive or dependent behaviour. In **fight/flight** they will attack the therapist or some other issue; or retreat and withdraw. And in **pairing** they may create a group illusion that some magical form of rescue may arise from the dilemmas of group life through charged partnerships. Hopper has introduced a fourth basic assumption that he calls **massification/aggregation** in which the defensive structures of groups or societies in crisis is thought to entail either a rigid fusion of identities excluding individuality, or extensive withdrawal preventing mutuality.

The two-body psychology used here enforces a series of clinical constraints that reduce the complexity of group interaction to a bi-personal exchange between therapist and group taken as a whole. As Fig. 6.3.6.4 illustrates, intra-group dynamics are considered only in their entirety for what they reveal about the unconscious state of the group as a whole, and for what they indicate about the nature of the group's relationship with the therapist. Figure 6.3.6.4 illustrates how the therapist stands outside the group in a stance that is not only neutral and dispassionate but also opaque and withholding of self.

Ezriel, basing his work on Bion, developed his theory of **common group tension**.[41] He believed that the group would be caught up at any given time in a commonly shared conflict centred on the unconscious fear of catastrophe, what he called **the dreaded state**. People would avoid a state in the group—say one in which they talked about sad feelings—because of the unconscious fear that talking about sadness would lead to a dreaded state, in this case a depressive collapse. A group would be driven into unconscious, defensive organization—what he called **the required state**—to keep sadness at bay. For example, an extended period of manic humour, the required state, would help prevent **the avoided state**, sadness, and this would in turn protect against the dreaded state. Interpretations would allow members to become increasingly aware of the underlying catastrophic fears and reduce their need for defensive organization.

Horwitz calls Ezriel's approach 'deductive', in that it relates individual's contributions only to the common group tension. He realized that this deductive approach was clinically unproductive and developed, what he called an 'inductive' method which is group-centred.[42] Interventions are first addressed to individual members in the group. Only after working with patients individually does the therapist introduce a common theme that binds them together. Thus the therapist, as in the Group-Analytic model, works on a figure/ground basis in which individual contributions are valued and explored in their own right before they are contextualized in the life of the group as a whole.

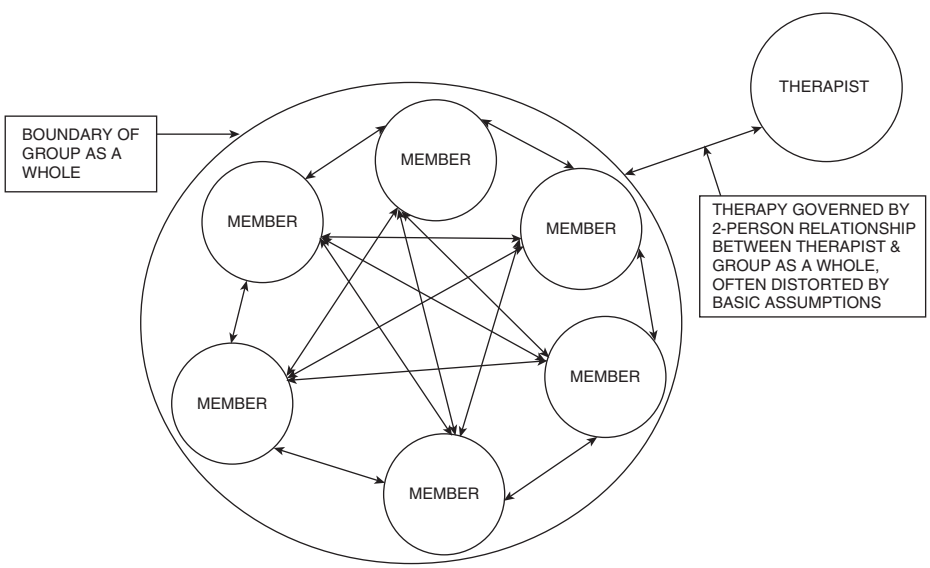


Fig. 6.3.6.4 The elements of a Tavistock group.

Another approach, **focal conflict theory**, as developed by Whitaker and Lieberman[43] is similar to Ezriel's in focussing the therapist on a conflict that becomes his point of emphasis, but it answers Horwitz's criticisms. On their account, underlying **disturbing motives** in group behaviour are acted against by unconscious, **restrictive solutions**. On their account the therapist—by focusing on such key conflicts—helps give members access to the unconscious anxieties and once these have been relieved they can construct more **enabling solutions** to the shared dilemmas of the group. The idea of focal conflicts, conceived of in these broader terms, has become an integral part of the Group-Analytic model.

(d) The Group-Analytic model

This approach integrates important aspects of the two preceding models but introduces a number of new elements. As Fig. 6.3.6.5 suggests, the therapist is encouraged to address the individual as well as the whole group and considers the more conscious and individual dynamics as well as the unconscious and potentially destructive whole-group dynamics. The approach is guided by an integrated set of concepts relating structure, process, and content to one another in which the group conductor works both as therapist and as group member to foster and cultivate the ordinary language of shared conversational experience. He will at times take up the position of the group's manager, and at other times he will speak personally as one of its members. Groups may begin with a relatively high level of leadership activity, referred to as **dynamic administration**, which is flexibly reduced with a decrescendo of responsibility as the group becomes the therapist and the leadership function is devolved upon its membership who becomes active co-therapists in each other's treatment[44]. Figure 6.3.6.6 indicates how, in this approach, at one key moment in the group its theme can focus on structural dynamics that link one member to both the therapist and to the group as a whole. The web of interconnecting dynamics between any one member and all the others, is summarized by 'the group as a whole', and this is called on to represent the inter-connecting latticework of relationships that includes all the members and the therapist.

(c) The Matrix

As Fig. 6.3.6.6 indicates, the conductor is inside and a part of the group, the structural elements of which provide a way of understanding the crucial links between each member, the conductor and the group as a whole. The triangle by which the group's psychological objects are linked to each other illustrates one of the 6 corresponding patterns of connection. When replicated for each of the 6 members, the diagram will produce a **matrix** of relational patterns, a complex relational field that will undergo change in terms of alliances, sub-groups, and polarizations. This concept of the matrix is crucial in group analytic theory. It allows us to accept that all events in a group will become part of an unconscious network that is intrapsychic, interpersonal, and transpersonal. The developing matrix creates the capacity to receive, contain, and eventually transform individuals' contributions, fostering integration at the individual level as it does so in the group as a whole.

Free-floating discussion: Free floating discussion is the group-analytic equivalent of free-association. The term originates in Foulkes' own writing and describes a set of key clinical concepts in therapeutic practice that distinguish the group-analytic approach. The language of the group is discussed in the dynamic life of groups, below.

Group-specific process: These processes, also mapped out originally by Foulkes, have been studied further by Pines[45] and by Agazarian and Peters.[46] The key concept of **resonance** describes the unconscious communication of emotion. The group provides its members with a wide field of meaning which is explored as they **mirror** one another's experience, find their emotions **amplified** by association with one another, and find **condensed**, sometimes highly aroused, cathartic experience, moments charged with significance.

Content analysis: Foulkes described four levels at which the content of the group's discussion can be analysed in the search for meaning.[47] In the Tavistock model this is the therapist's exclusive task, whereas here interpretation is only one amongst a number of others. The therapist's overall stance is to foster communication and educate the group's members about the dynamic links between the group's structure, the content of the discussion, and the form

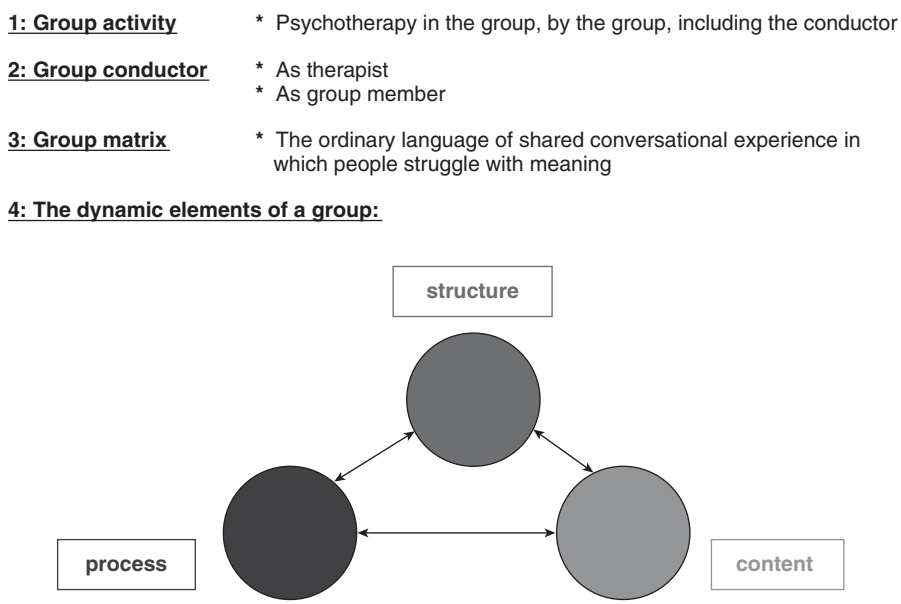


Fig. 6.3.6.5 Group Analytic psychotherapy.

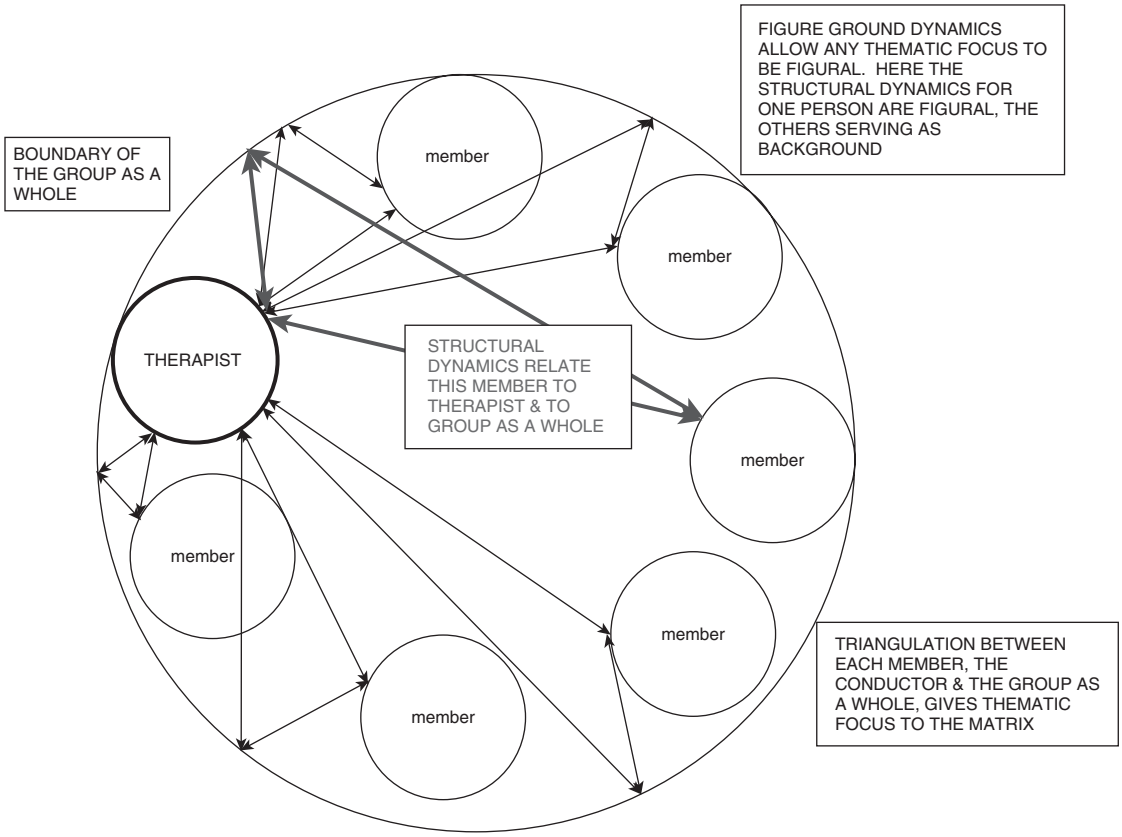


Fig. 6.3.6.6 The elements of a group-analytic theory.

in which it takes place. Figure 6.3.6.7 illustrates the conductor's therapeutic role in relation to each of these dynamic elements of the discussion.

(f) Recent developments in group-analytic psychotherapy

The role of dreams in group psychotherapy has been reconsidered.[48] Friedman described two of the unconscious functions they serve in group therapy—requests for containment and influence on relations with the dream audience.[49] Lipgar and Pines[50] work towards an integration of the ideas and methods of Bion and Foulkes with discussion from an international panel. Nitsun[51] considers sexuality in group psychotherapy: sexual identity, boundary transgression, erotic connection, dissociation of desire, the group as witness, erotic transference and counter-transference, and the effectiveness of psychotherapy.

(g) The conductor as therapist and group member

At times of coherence, when the members are close in the shared experience of a moment, or when there is an issue charged with meaning, the group-analytic approach comes into its own. The symbolic content of the discourse might evolve in the language content, the flux of interactions, or the attention given to an individual's problems. The conductor needs to be able to model this use of imaginative play—with images, associations, or exchanges—and then stand back to allow members to take the enquiry forward. Cox[52] shows how images can safely hold experience too painful or brittle to tolerate much analysis. People discover that images can touch the depths before they stir the surface, giving access to profoundly felt and deeply hidden concerns. When used and played with in this way, **the mutative use of metaphor** provides the whole group with a vehicle for change. Other aspects of the therapist's activity are summarized in Table 6.3.6.4.

The dynamic life of groups

In the sections that follow we draw on our own Group-Analytic model to examine a range of clinical considerations and make them as relevant as possible to the widest range of practitioners, regardless of their own models.

Group development theory

Many schemata have been proposed, usually derived from time-limited experiential and study groups but some of these can provide a useful orientation. Bennis and Shepard,[53] followed by Yalom, write about the initial stages of **orientation**, involving a search for structure and goals, dependence on the leader, and concern with boundaries. Their second stage is characterized by **conflict**, particularly over **norms**, authority, and control. Their third stage is achieved with a high level of group **coherence** that allows for inter- and intra-personal exploration. They acknowledge that the boundaries between phases are not clear and that a group never graduates permanently from any one phase.

In slow open groups, especially, this idea of discrete phases has its limits. The group-analytic approach provides a cyclic model in which focal issues such as affection, intimacy, personal history, and change, are understood and struggled with repeatedly. A group

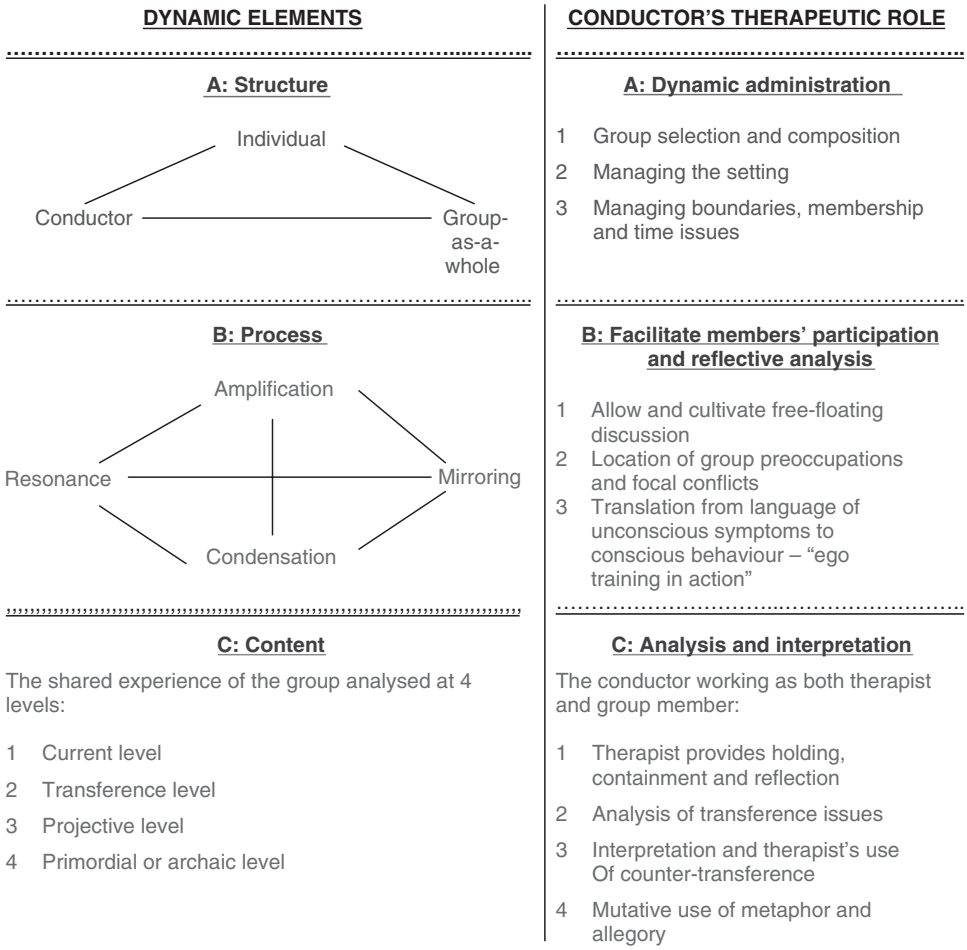


Fig. 6.3.6.7 Structure, process, and content—the conductor's therapeutic map.

Table 6.3.6.4 The therapist's activity

A: Leadership and analysis
1 Model a capacity for open, direct communication
2 Maintain therapeutic neutrality
3 Attend to boundary events
4 Provide holding and containment
5 Withhold personal material
6 Drawn on counter-transference for
7 Reflection on group events
8 Bring events from background to foreground, or vice versa
9 Provide linking communications
10 Clarification and confrontation with individuals
11 Attention to omissions, avoidance, denial
12 Maintain silence
B: Interpretation
1 Locate group preoccupations
2 Translate from the language of unconscious (individual and group) behaviour
3 Interpret or provide metaphorical constructions for
I Defences and resistances
II Transference and projective process
III Archaic and primordial experience

understands in these terms struggles with **developmental tasks** rather than phases, in the course of which it is the individuals who enjoy growth, differentiation, and progressive change.

(a) Developmental stages and thematic focus

Figure 6.3.6.8 provides a map of these developmental tasks conceived of in logical rather than sequential terms. The way in which one person comes into the group will be different to the arrival of another. The terms on which one person joins will have a determining influence on each of the stages they pass through. For example, preoccupations about how someone came into the group—the terms of their engagement—might be resolved only when the person leaves perhaps one, three, or five years later, when they have to assess the outlook for their future as they look back over the years spent in therapy. So the diagram can be used as a thematic map for one person's journey through therapy, or as a way of appraising the stage reached by the group as a whole.

Early stages are dominated by the anxiety of being involved in a new situation and by questions about other group members. Preoccupations are likely about boundary issues, confidentiality and security. In the second stage, those familiar with psychotherapy—or otherwise accustomed to talking about themselves—will be at an advantage. There will be a range of tensions about group norms and disclosures, reasons for joining and discrepant levels of confidence about using the group. In the third stage, members will defend against intimacy with one another and struggle with questions of trust, attachment, and affiliations. In the

fourth stage, as members become increasingly able to trust the group with self-disclosure, observed changes might become manifest as self-exploration yields the beneficial experience of individuation and differentiation. In the concluding stage, people prepare for departure and find themselves comparing points of difference between changes achieved in the group and the state of their lives outside—generalising from the arena of therapy to that of real life. Has therapy made a lasting difference? Will it be maintained outside?

The language of the group

Foulkes suggested that 'symptoms, in themselves unsuitable for sharing, exert, for this very reason, an increasing pressure upon the individual to express them'.[54] The group equips the person to transform the mute and inchoate language of symptoms into a socially understandable form of discourse. Following publication of Schlapobersky's paper, The Language of the Group, there is increasing interest in characterizing group phases in terms of the language that predominates, using theory from discourse analysis and the Foulksian concept of free-floating discussion.[16]

It is possible to differentiate between three primary forms of speech that arise in the matrix of any group. At the most basic level **monologue**—speaking alone (with or without an audience)—is a form of individual self-expression. At the next level **dialogue**—a conversation between two people—is the form of communication that distinguishes a bipersonal exchange. And at the third level **discourse**—the speech pattern of three or more people—allows the free interaction of all its participants in a flexible and complex exchange that distinguishes the communication of a group. These patterns of speech are universal cultural forms arising in all communication and are present in the life of every group, although in no set order. Monologue can be understood as a soliloquy, dialogue as the resolution of opposites or the search for intimacy, and discourse as the work of a chorus. The use of free-floating discussion allows a pattern of exchange to move freely between these different speech forms, each of which constitutes a distinctive type of communication. It is through this movement—from monologue through dialogue to discourse and back again—that the group-analytic method comes into its own, creating an arena in which the dialectic between the psyche and the social world helps to refashion both.[13,14]

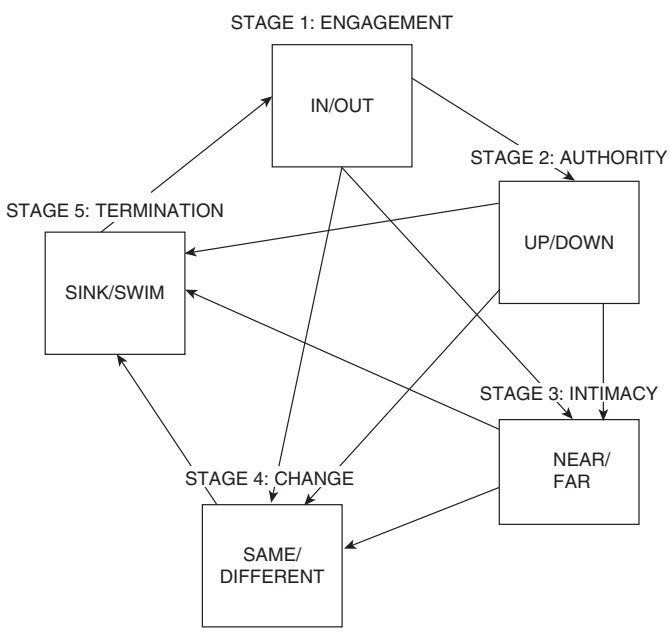


Fig. 6.3.6.8 Developmental stages and thematic focus.

Leadership

The group analyst works as both group member and therapist, beginning with dynamic administration, assuming an active role in a new group and allowing a **decrescendo** of his own role as the group gains authority. He is responsible for helping the group with the **location** of disturbance in its process and for providing a balance between analytic and integrative forces whilst manifest content is translated into language that describes the unconscious. Transference is prominent but the work is undertaken in the dynamic present. Foulkes' account in the following passage,[55] of the conductor at work in a group, stands in dramatic contrast to Freud's account of the psychoanalyst at work behind the couch:

> He treats the group as adults on an equal level to his own and exerts an important influence by his own example . . . representing and promoting reality, reason, tolerance, understanding, insight, catharsis, independence, frankness, and an open mind for new experiences. This happens by way of a living, corrective emotional experience.

Disturbed, narcissistic, or borderline patients bring to the group more primitive psychic structures and processes that put strain on the resources of other group members. Such patients can create turmoil in which the leader's task is to maintain the group at a more mature level of psychic organization. By responding to part-object relationships and processes on the level of whole-object relations, containing responses can be established. Progressively, these help to build for the disturbed patient a more benign world of inner object relationships and processes. More disturbed patients desperately seek attention in ways that are inappropriate and disruptive. This search for **attention** arises because the patient cannot establish a sense of **connection** between herself and the processes of the group. Mirroring and resonance can steadily come to replace these isolated and fragmentary responses allowing the patient to attain—for the first time—a coherent sense of self and a capacity to recognize the identity of others.

(a) The dynamics of change

As noted above, Yalom cited 11 'curative' factors responsible for change in groups. Foulkes believed there were four key group-specific factors: mirroring, exchange, social integration, and activation of the collective unconscious. Tschuschke and Dies[56] have identified five key factors that they could study empirically (Table 6.3.6.5).

(b) Binding forces: alliance, cohesion, and coherence

Group alliance has as its major focus the quality of the relationship that develops between each individual member and the therapist(s). These alliances grow as each individual develops a transference relationship to the therapist and can feel that their own particular individual dynamics are recognized. Group cohesion used to be compared to the concept of therapeutic alliance in individual psychotherapy but current research shows that alliance (defined as the affective bond that develops between each group member and the therapist) can be differentiated from group cohesion.

Group cohesion describes the bonds between group members, their attitudes, and their commitment to therapeutic work, in particular

Table 6.3.6.5 Five factors in the dynamics of change open to empirical evaluation[56]

1 Cohesiveness	Closely related to 2 & 3 as determining factors that influence outcome in 4 and 5
2 Self-disclosure	Closely related to 1 & 3 as determining factors that influence outcome in 4 and 5
3 Feedback	Closely related to 1 & 2 as determining factors that influence outcome in 4 and 5
4 Interpersonal learning output	Evidence of working through process seen in 4
5 Family re-enactment	Findings of change in these patterns provides evidence of enduring psychic change

their feelings of attraction and dependency. These processes are interdependent and in combination they provide optimal conditions for positive group process and outcome.[57]

Group coherence is a more evolved group state requiring but going beyond cohesion. It evolves as a semantic matrix, built on the earlier, relational matrix. When a group moves through support and understanding to be able to recognize and work through conflict, it can achieve a sense of containment.[58] At this stage the group becomes a more complex, self-evolving, and self-defining entity capable of reaching deeper levels of exploration, acceptance, and understanding.

(c) Corrective emotional experience

The concept was introduced by Alexander to describe the patient's recognition of discrepancies between present and past experience. Grotjahn, a colleague of his, applied this to the group, describing it as 'the corrective family experience'.

> There is a built-in correction of the transference phenomena through the peer relationship in groups. An analyst is trained to let the transference neurosis to grow to full bloom. Members of a group are neither trained nor willing to accept such projections . . . and will correct them. This is the basis of a corrective therapeutic family experience.[59]

Oedipal, sibling, and pre-oedipal constellations are activated in group therapy and can be worked through as the members play parts in each other's family scenarios.[60] The task of the therapist interprets transference and helps the group as a whole to develop coherent norms of understanding and responsiveness that equip its members to go beyond the private preoccupations that might have brought them into therapy. Garland calls this **taking the non-problem seriously**.[61]

(d) Resonance, mirroring, and other dynamic processes

Foulkes described the group as 'a hall of mirrors'. Each person, he thought, could see aspects of themselves reflected in the personality and behaviour of others and could often more easily recognize these aspects of the self than by direct introspection. We cannot know ourselves in the absence of reflective mirroring, the feedback and information we obtain from others according to our presence and behaviour in the group brings us to a greater awareness of who we are. Resonance in the group is the unconscious communication of emotion by which the other process dynamics—like mirroring, amplification and condensation—are effected.

Insight and outsight, regressive and progressive forces in groups

The dictum, 'where id was, ego shall be', coined originally by Freud to ground the psychoanalytic enterprise in an easily comprehensible principle, has been superseded for group workers by the idea of **ego training in action**, coined by Foulkes. It emphasizes two related issues. The first is that **insight**—the understanding of the self—is related to **outsight** defined as the understanding of the other(s). And the second is that people find growth and change as much in what they do for others—**progressive behaviour**—as in what others can do for them—**regressive behaviour**. So the group is designed to create an arena in which there is continuous flux between these different elements.

(e) Imitation, identification, internalization, and differentiation

(i) Imitation

From early in infancy, we observe and copy what others do. A therapeutic group is designed to create conditions in which imitative behaviour, or modelling, can be experienced, monitored, and understood. It brings the group members closer together and, as it develops into identification and internalization, it increases the cohesiveness of the group.

(ii) Identification

Peoples' predominant modes of relatedness to one another—such as compliance, avoidance, dominance, receptivity, exploitation, and need—become apparent in the group. The group exposes these patterns of relatedness to reveal how the internal objects with which members have identified can be externalized and encountered in the group through projection and introjection. Members compel each other to play and re-play the dramas of their interior lives and past injuries in the 'here and now' of group experience.

(iii) Internalization

Intimacy with others is developed through the exchange of understanding. As it is given and received, it allows for new forms of intimacy within the self. **Constructive** experience in the group is taken in and becomes part of the self. This can allow the recognition and reappraisal or **deconstruction** of repetitive patterns and fixed, maladaptive characteristics. And this can lead to the **reconstruction** of the self in the concluding phases.

(iv) Differentiation

Over time people become aware of significant differences in their reactions, emotions, and psychological structures. Examples are increased tolerance of affect, understanding and modification of self-inflicted pain, diminution of guilt and shame, retrieval of lost aspects of the self, increased openness with others, increased spontaneity and creativity, and, most critically, feelings of tolerance for or forgiveness of the self, for the constraints into which it has been driven by past injury and present defence.[62]

(f) Personal and group resistances

Some resistances are manifestations in the group of the characteristic defences of the members, evident in their interpersonal behaviour. Others are resistances of the group as a whole, shown in blockage of free-floating discussion, opposition to group interaction, sub-grouping, and opposition to the deepening and broadening of the group's exploration. A constructive function of resistances should be kept in mind when monitoring the pace at which both individuals and the group can progress without experiencing

overwhelming anxiety. Resistances can protect people from fears of loss of self and identity; and from fears of engulfment through excessive intimacy. Resistances then become opportunities for understanding fear of change, for the re-working early developmental patterns, and for discovering the freedom that can follow release from excessive internal control.

(g) The anti-group

Nitsun developed an understanding of the anti-group, bringing together the work of Bion and Foulkes.[63] He introduced the concept to help understand the negative experiences therapists face in periods of stagnation, hostile silence, severe conflict in or premature departure from the group, or negative feelings about the work of the group. These situations can arise when the group recapitulates early experiences of loss, deprivation, anger and envy or in an effort to avoid these emotions. Negative emotions can then be projected towards the group and the conductor who represent early care-givers. These feelings can be worked on when the therapist uses counter-transference to recognize and verbalize projected feelings. The idea of **negative elaboration** (see below) is a useful guide in this process.

Basic clinical issues

Dynamic administration

(a) Selection and composition

Group composition is the therapist's first and most enduring contribution to the group for its membership will determine the outcome of therapy. Preparing patients for treatment with several individual sessions or, if necessary, an extended programme of preparatory work will provide the therapist and patient with a basis for judgement about therapeutic prospects. Preparatory work has been found likely to enhance participation and reduce drop-out rates, although findings are not conclusive.[64]

The criteria for selection are exclusive rather than inclusive since most patients seeking psychotherapy can be accommodated in a group, provided a suitable one is available. The selection process should take into account both the patient needs and the composition of the group. A service should, ideally, provide a selection of groups into which people can be placed both according to their needs and characteristics, and those of the particular group. Selection criteria aim to optimize the 'fit' between the needs and resources of the individual and those of the group.

Table 6.3.6.6 lists inclusion criteria, while Table 6.3.6.7 gives a shortlist of excluding criteria. Lists of this kind should be used with caution. In general, at least four inclusive criteria should be found amongst those to be included in outpatient, dynamic psychotherapy groups with a mixed population. If there are four exclusion criteria, one should be very wary about including the person in a group. However, there are many exceptions.

The criteria in Tables 6.3.6.6 and 6.3.6.7 hold good for mixed groups and outpatient services generally. With homogeneous groups for special populations, the range of potential candidates is much wider, for example, Hearst describes a population of severely deprived mothers drawn entirely from those on the exclusion list.[65]

(b) Homogeneous and heterogeneous groups

Homogeneous groups: Are for people with similar symptomatic or diagnostic pictures, such as phobias, anxiety, or depression. Such groups offer more immediate support to members, are better attended, and provide faster symptomatic relief. However they may remain at a more superficial level with less interpersonal learning.[67]

Homogeneous groups are also used for those with similar personality structure or life-history, particularly those in socially extreme categories. For example, men with histories of sexual violence or women who have suffered rape or torture may be treated.

Heterogeneous groups: Melnick and Woods[67] suggest that group composition should be guided by an optimal balance between conditions ensuring group maintenance or homogeneity, and those maximizing interpersonal learning or heterogeneity. A group which shares one strong characteristic—a diagnosis such as an eating disorder, or a personal attribute like intelligence—can accommodate a diversity of presenting problems or social backgrounds. If, on the other hand, the members are similar in social background, diversity can be incorporated on another basis, such as diagnosis. Thygesen[68] found that diversity enabled group members to recognize and work with differences in mental and emotional attitudes, life histories, and developmental problems. Recognizing and working with difference develops emotional resources, promotes flexibility, and the tolerance of emotional tension. And it encourages the group to move from **cohesion**, in which security is based on identification, to **coherence** in which relationships are based on differentiation.

It is useful also to identify members likely to be isolated from the rest of the group by age, ethnicity, gender, personality, or problems or a history that noone else shares, for they are likely to find

Table 6.3.6.6 Inclusion criteria for psychodynamic groups (at least four of these criteria should be present, for someone to join a group)

1	Motivation to address personal issues, to resolve problems
2	Willingness to try and participate
3	Some experience of successful relationships in childhood or present
4	Some interest in exploring and understanding the self
5	Some capacity to talk, listen, and relate
7	Some interest in others
8	Some sense that being amongst others could be helpful
9	Some ability to sympathize or empathize with others' needs and problems
10	Some indication of future reliability in attendance

Table 6.3.6.7 Exclusion criteria for psychodynamic groups[66] (if four or more of these criteria are present, then questions should be raised about inclusion)

1	Those in acute crisis
2	Prior history of broken attendance in therapy
3	Major problems of self-disclosure
4	Major problems with reality testing, i.e. paranoid projections or psychosis
5	Pathological narcissism
6	Difficulties with intimacy generalized into personal distrust
7	Defences that rely excessively on denial and disassociation
8	Emotional unavailability
9	Tendency to be verbally subdued or withdrawn
10	Tendency to be hostile and aggressive, verbally or otherwise

the group experience threatening. We do not put a patient into a position of being isolated.

Managing the structure, setting, and time boundaries

(a) Optimal and sub-optimal size

The optimal number for small group psychotherapy, ranging from five to nine, is determined by practical considerations. A smaller number than this minimum is likely to have an active attendance of only two or three and will not necessarily generate the corporate energy to produce movement. A group larger than this will exceed the number that can be taken into one person's confidence in a face-to-face exchange.

Sub-optimal groups can provide valuable therapy under good conditions.[69] Low and irregular attendance is often associated with a problematic composition, particularly if there is a high proportion of members with character disorders and borderline features. Their inner sense of deprivation and loss can make the group seem unreliable and threatening which can be reinforced if the therapist is seen to be in difficulty. Therapists who can maintain an understanding attitude and positive commitment to the group's future will usually find the situation settles into a working nucleus of members that can then be built upon.

(b) Setting and time-boundaries

The therapist supplies, creates, and maintains the setting throughout the group's life. This requires attention to such matters as meeting times, punctual beginnings and endings, confidentiality, the predictable frequency of its meetings and breaks, and the general guarantee of a stable background.

Every aspect of the group's life, including absences, departures, late attendance, and extra-group communication in terms of letters, phone calls, and messages, referred to generically as **boundary events**, are open for discussion. Boundary events are interpreted for the meaning they might hold for the life of the group as a whole. This is a task initiated by but not confined to the therapist. Also, the setting itself—to which each member is seen to contribute—comes to acquire a capacity to **hold** the individual members and **contain** their anxieties and insecurities. The issue of containment is of great importance, as is the discovery—by people who may have no belief in themselves as responsible members—that they can take responsibility for themselves and expect responsibility from others. The group is thus responsible not only for the nurture, acceptance, and security of its members, but also for their containment, the setting of limits and the maintenance of consistent authority. The therapist will often have to lead the way in modelling both roles for the group's members.

Fostering therapeutic norms and a culture of enquiry

The therapist encourages the sharing of experience and helps to balance participation, recognition, and translation. The thrust of a group's life is towards greater shared involvement and the expression of emotions. The expression of feeling may arise in the recounting of members' life-situations and their reasons for therapy—**narrative emotion**—or it may arise in the interpersonal encounters engendered by the telling—in the **drama** of 'here and now'. Therapy becomes effective when the problems that brought the patient into treatment become recognizable in their interpersonal encounters. At this point the affect lodged in the narrative will interact with the drama of the group's current emotions, creating opportunities for corrective emotional experience. As the group progresses from **constructive** to **deconstructive**, and ultimately to **reconstructive**, experience, it will encompass gesture, behaviour, body-language, and other non-verbal communication, and actions that convey feelings when emotions have no words.

Guidelines for intervention

There are four modalities of time and place in any therapy group. The content of the exchange might be located in the past outside the group, the past inside the group, the present outside the group, and the present in the group.

Free-floating discussion will carry the focus of the exchange between these different modalities. The therapist's task is to follow the interaction, to use interventions sparingly and strategically, to cultivate a reflective curiosity for which Table 6.3.6.8 offers some pointers, and to work towards a progressive shift in the focus of attention, from no. 1, in Fig. 6.3.6.9 towards 4, via 2 and 3. A group governed by narrative in its free-floating discussion is likely to be dominated either by the the group's own past or by the past of its members outside the group or by their current lives outside. A group governed by members' intense experience of one another in the present, is likely to be governed by the drama of immediate encounter and understanding. The corrective recapitulation of early family life is likely to arise when the issues that predominate in no's 1, 2, and 3, are translated into issues that prevail in 4, where they can be addressed in the present which allows new resolutions to be forged.

Groups for special populations

Nine distinctive clinical populations or approaches merit special mention. Some pose distinctive problems for group workers and, as we indicate in the section on borderline patients, they sometimes need co-therapists working together in the group.

Table 6.3.6.8 Interventions in time and place

1	Address process rather than content
2	Help members recognize aspects of themselves in others and accept the viewpoints of others on themselves
3	Monitor the intensity of participation to allow an enabling pace so members can develop resources to deal with the intimacies of each others' lives
4	Establish a sense of enquiry about the fluctuations of mood and outlook, and a sense of curiosity about thematic movement between the different modalities of time and place
5	Help the group recognize role configurations taken up by individuals (therapist's assistant or rival, joker, complainer)
6	Hold in mind the gestalt of figure and ground. If someone stands out, what is the background against which they do so? If the ground changes the underlying pattern of the group, how might the figural person be affected?
7	Help the group recognize sequences and patterns
8	Help the group decode and find meaning in the constructive, deconstructive, and reconstructive elements of its exchange
9	Work towards the coherency of the group
10	Foster the integration and integrity of its members

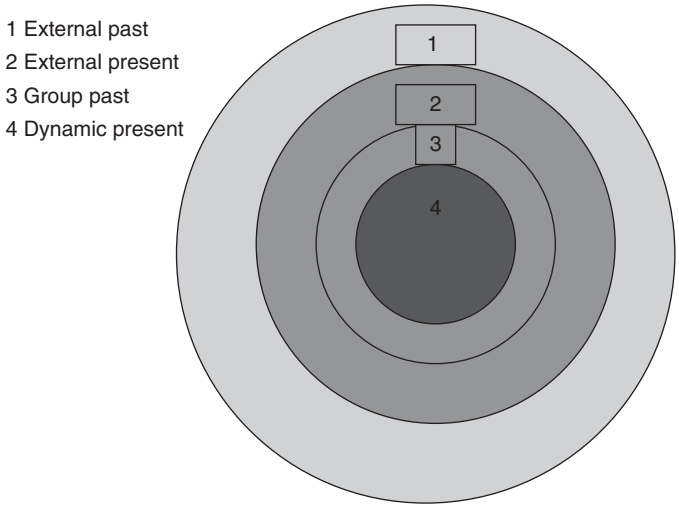


Fig. 6.3.6.9 Location of group discourse in time and place.

Intensive outpatient psychotherapy heterogeneous groups for mixed populations

Group analytic psychotherapy has come into its own as a treatment modality in out-patient settings in the context of clinical agencies in public and private Health Services. Table 6.3.6.9 gives a picture of patients' clinical needs in the private practice of John Schlapobersky over a twenty-year period. Some attempt has been made to indicate how these categories are mapped against those in ICD 10 and DSM IV.

The medically ill

Groups have been used successfully for people with diseases of the cardiovascular, endocrine, gastro-intestinal, pulmonary, neurological, and renal systems. Amongst the applications are groups for those with renal dialysis and organ transplantation; chronic conditions such as diabetes, irritable-bowel syndrome, and rheumatic disease; and life-threatening conditions such as metastatic cancers.

Groups usually have members with a single disorder, last for 20 sessions or fewer, and are psycho-educational in purpose. They emphasize compliance with treatment regimes and opportunities to share and explore the affective consequences of the conditions. The focus is on helping patients come to terms with the complaint; make changes in lifestyle, including diet, exercise, work, and recreation; deal with the inevitable anxieties about death, stigma, demoralisation, feared dependency, and loss of control; and resolve issues in relationships with loved ones and care staff. Clinicians write of how, despite their brevity, these groups provide a forum for sometimes profound exchange about major existential issues including pain, fear, and impending death.[70]

Evaluation indicates such groups are effective in improving quality of life and, to some extent, in prolonging its duration even with metastatic cancer.[71] Improved survival correlates significantly with enhanced, active coping. Spiegel reported a significant, near-doubling of survival effects in 86 women with metastatic breast cancer treated for one year with weekly, supportive-expressive group therapy.[72,73]

The chronic mentally ill

In North American literature there are consistent findings that schizophrenic patients do best in support groups that help them manage their symptoms and devise coping strategies for day-to-day problems.[74] Group discussion is used to help overcome problems of social isolation, develop social skills and coping resources, and learn about maladaptive interactions evident in the 'here and now' of the group. The leader creates a sense of safety and security with supportive feedback. Areas that people find helpful include the management of symptoms, expression of emotion, and relationships with others. Groups are less valued for acquiring insight or receiving guidance about the illness, medication, or economic problems. Interpretations that reveal or explore unconscious conflicts, particularly transferential issues between members or with the therapist, are more likely to be harmful than helpful.[75] Patients respond positively to an interactive, open, and safe group environment, with gradually increased cohesion, decreased avoidance, and decreased conflict.[10]

Recently, Resnik[76] has described the use of insight-orientated psychotherapy with delusional problems using psychoanalytic principles in long-term group psychotherapy for psychotics.

Borderline personality disorders

Working with these patients can pose difficult clinical problems based on their 'stable instability', including their inability to form a stable therapeutic alliance, mood oscillations, poor impulse control, limited reflective ability, painful states of psychic emptiness, followed by self destructive acting-out including self-destructive gestures and self-mutilation, violence; substance abuse, sexual acting out, and eating disorders. Some attempts at short-term group therapy combining cognitive behavioural approaches with the psychodynamic have shown promise. Longer therapies in out-patient groups have been encouraging but have not been evaluated. Good results are reported with in-patient applications at, for example, The Cassel and Henderson Hospitals, and in other forensic settings where day hospital group treatments are followed by outpatient group therapy in a carefully monitored programme.[77]

Homogeneous outpatient groups do not do well because the members use similar 'primitive' defences of splitting, projection and denial, and have not acquired a stable sense of identity. They are suspicious of close interpersonal contact and have little capacity to care for or be cared for by others. Such groups lack cohesion, are pervaded by a sense of pessimism, hostility, rivalry, and easily fragment. Self-destructiveness is turned onto the group and its therapists who may experience intense counter-transference feelings of frustration, despair, and anger.

These problems can be managed in the containing setting of closed or specialized institutions. However, it is generally better to place one or two borderline patients in an otherwise well-functioning group where other members will not always respond to borderline pathology at the same primitive level. They may—like the adult carers of children—respond with understanding, find ways of setting limits, and expect co-operation with the task at hand. After sometime, lengthy periods in which borderline patients maintain a frustrating presence on the margins of the group, or make themselves felt in aggravating terms at its very centre, they can acquire resources to take part in the group's work in more mature terms. This combination of borderline and neurotic members in a carefully composed group can benefit both parties. Borderline members will often have an unerring accuracy of perception about others, they can shake the group into more active interactions, and may not collude with others' neurotic defences.

Table 6.3.6.9 Categories of clinical need catered for by intensive, outpatient group therapy with mixed populations (many of those who join groups are found in more than one of these categories of need)

1. High dependency need including:	2. Problematic reactions to traumatic life events including:	3. Character problems including:	4. Serious relational problems including:	5. Selected people with borderline personality disorders including	6. Training of mental health professionals including:	7. Those seeking personal growth or understanding including:
a Recovery from serious or enduring mental and physical illness b Help with the resolution of moderate psychiatric symptoms c Resolving addiction and substance abuse	a Loss, injury, illness, disorder, infertility b Massive psychic and physical trauma c Adult sequelae of child sexual abuse	a Immaturity and developmental problems b Chaotic life situations c Problems of identity and meaning d Problems of gender, sexuality, and orientation e Occupational problems	a Intimacy avoidance b Recurrent broken relationships c Intractable conflicts	Those with insight who both need and can tolerate containment	Psychotherapists, psychiatrists, psychologists, social workers, clergy	People who would have previously sought this through individual psychoanalysis

(John Schlapobersky: Group-Analytic Practice & Southwood Practice: 1987–2007)

Category 1: Corresponds to Axis I in DSM IV and Groups F1–4 in ICD 10.

Category 2: Corresponds to Axes III, IV, & V in DSM IV and Group F5 in ICD 10.

Category 3: Corresponds to Axes IV and V in DSM IV and Group F8 in ICD 10.

Category 4: Corresponds to Axis V in DSM IV and Group F8 in ICD 10.

Category 5: Corresponds to Axis II in DSM IV and Group F6 in ICD 10.

The containing resources of the other members that set the norms and values of the group can slowly be internalized by the borderlines. As with many of these specialist areas, therapists working with these populations will need to acquaint themselves with the literature, have ongoing supervision and may—to begin with—work more fruitfully in co-therapy.

For a review of group psychotherapy for personality disorders see [78, 79].

Forensic groups

Group therapy has been used for forensic populations, previously thought untreatable. Pioneering work has been done in the UK in special hospitals,[80,81] in prisons such as Grendon,[82] and great strides have been made in outpatient services at the Portman Clinic in London.[83] Therapeutic community treatment for mentally disordered offenders in North America and Western Europe has been reviewed by Lees, Manning, and Rawlings.[84]

The aim of forensic group therapy is to help patients find words, rather than actions, to express impulses and compulsions. A major part of the group's work is to provide psychic space for perspective, negotiation, recognition, acceptance, and verbalization of hurt. Sharing the past and present can make it accessible as internalized, persecutory, and vengeful monologues are brought into dialogue.

In Britain, the Probation Service is the agency most actively involved in the groupwork with offenders.[85] A report for the NHS Centre for Reviews and Dissemination (CRD Report 17) reviews studies of therapeutic communities and small groups and indicates that these show the most promising results of any form of treatment for anti-social personality disorders.[86]

Trauma

Group work with trauma victims is a comparatively new field but one in which there is a vigorous range of applications. The most comprehensive overview is Van der Kolk's text.[87] There are groups for survivors of sexual abuse,[88] war trauma,[89] and torture and other forms of organized violence. Group work applications to traumatized children and adolescents is discussed in reference 90 and more generally in reference 91.

At the Medical Foundation for The Care of Victims of Torture in London, a group work programme caters for refugees and asylum-seekers. The groups provide psychotherapy for massive psychic and physical trauma, for problems of displacement and exile, and for trans-cultural problems.[92] The diversity of different kinds of groups, including a range of activity, problem-solving, and psychodynamic groups, ensures that people can be provided with an environment in which they can each realize their own potential for self-healing. The programme is grounded in commitment to human rights, to team-work that addresses counter-transference issues, and to principles of positive intervention to counter emotions of hopelessness, and despair.[93,94]

Couples groups

These groups cater for people in stable but troubled relationships in which there is some form of pernicious collusion. The approach can provide symptom relief and personality change in even severe difficulties with relationships that last but do not work. Family therapy is concerned with the systemic function served by symptoms. Psychoanalysis is concerned with the origins of the

symptoms in object relationships. Group analysis provides a bridge between these paradigms, helping therapists find a point of intervention between marital and object relationships.

In a group, the process moves between the psychology of the individuals and the dynamics of their marriages. This interplay between the pair and the person—the interactive and historical dynamics—is part of the group's free-floating discussion. The therapist uses this interplay, following the patterns of the group's progress and making interpretations. Work is at times transferential, other times, systemic techniques are used to address immediate issues.[95] Themes to which the group resonates are amplified and members are helped to condense from this discourse, the kind of personal knowledge that promotes growth and change. There is insufficient evaluation of the method's efficacy, but experience in hospital and private practice is very encouraging.[96]

The elderly

This is an area of relatively new therapeutic exploration. A range of groups has been found useful in helping the aged to face their problems, improve their functioning, and feel happier about themselves. The groups include those offering support, activity, psycho-education, problem-solving, and insight, and takes place in many settings. Common themes include the ubiquity of loss, the acceptance of death, and the value of humour in lightening mood.[97]

Brief therapy in groups

Spurred by economic and managerial pressures a distinctive modality has evolved. Sessions are held weekly and number between 6 and 30. Characteristics that differentiate brief from long-term therapy include: clearly defined therapeutic goals agreed at the outset, the early establishment of a therapeutic alliance, active and flexible therapeutic style, a focus on 'here and now' group process, the maintenance of time awareness, monitored in stage-specific terms, and the vigorous, directed exploration of thematic content.

Short-term groups may be homo- or heterogeneous. Homogenous groups have proved effective in helping patients deal with loss and grief, the consequences of trauma and abuse, and common problems coping with physical illness and disability. Heteregenous groups require more psychodynamic commonalities such as shared problems in inter-personal relationships, the ability to recognize and work on psychological issues, and the ability to cope with the speed and intensity of the process. Those who lack psychological sophistication or are not motivated for self-exploration, are not suited.

Research and evaluation

The general trends in literature show ample empirical confirmation that group treatments represent a powerful therapeutic intervention. A comprehensive overview of literature described positive outcomes with alcoholism, anxiety disorders, bereavement, bulimia, depression, schizophrenia, and sexual abuse.[98] There is also evidence of the adverse outcomes in group psychotherapy which can guide the clinicians training for the work.[99]

(a) Outcome research

A meta-analysis of 58 controlled studies of psychotherapy for the treatment of depression showed that in comparison to a waiting list control the average treated patient was better off than 80 per cent of the controls.[100] The efficacy of group and individual therapy was almost identical. Tyllitski[101] also reported no appreciable difference between individual and group therapy effectiveness, both doing better than the control condition. Budman et al. found significant improvement in time-limited individual and group therapies. It seems that most patients who are suitable for psychotherapy will benefit in either modality.

(b) Process research

(i) Preparation of patients for group therapy

In a review of 20 controlled or comparative studies, Piper and Perrault[102] found that preparation has a positive effect upon attendance but that it could not be shown to have a direct effect on outcome.

(ii) Therapist activity

Development of constructive group norms will depend on factors such as careful group composition and leadership style. Foulkes' idea that the conductor has, at first, a relatively high rate of activity which decreases as the group develops its own resources for psychological work has been supported by later research findings.[103]

(iii) Group process variables

Research on therapeutic factors (e.g. self-disclosure and feedback) and leadership technique indicate that members in well-established groups are engaged in many different types of psychological work. They are less group-centred and more likely to be confronting the personal distress and maladaptive interpersonal styles that brought them to treatment in the first place.[104]

(iv) Therapeutic factors

In a study of long-term in-patient groups, Tschuschke and Dies[105] investigated five therapeutic factors: cohesiveness, self-disclosure, feedback, interpersonal learning-output, and family re-enactment. All five therapeutic factors were associated with clinical improvement with group cohesiveness, an important ingredient. They suggested that **affective integration** into the group, that is the high and positive emotional relatedness to co-members, promotes the capacity to disclose and leads to more frequent and intense feedback from fellow patients. It appeared that feedback given earlier in the group had a stronger relationship to treatment outcome. This may suggest that **interpersonal feedback** needs time to be assimilated and worked through before it can be utilized effectively. There are significant differences between successful and unsuccessful patients in terms of level of group cohesion and amount of self-disclosure. Patients who disclose little and do not feel drawn to the group receive relatively little meaningful interpersonal feedback and become neglected. They concluded that **cohesiveness, self-disclosure**, and **feedback** and together promote **interpersonal learning** within the group.

These findings were confirmed by the author's later studies, the most recent of which was published in 2007[106] and by recent independent studies using different parameters; see, for example, references [107–109].

Conclusion: planning a service

For group psychotherapy to be effective the group has to be the primary focus of therapy; patients need to be well selected; and

therapists need to be adequately trained. Therapeutic competence is not a function of mastering the literature so much as it is the outcome of experience in the group situation itself. Courses introducing different group methods are now widely available throughout the UK, Continental Europe, and the USA. Group training is also provided in many general training programmes and, along with clinical supervision, is offered in many health and other service agencies.

Long-term **outpatient** group therapy of 100 sessions or more is effective and economic in producing lasting benefits for patients with a wide range of medical and psychiatric symptoms, interpersonal problems, traumatic life experiences, character and personality disorders. **In-patient** group therapy is an effective resource in the context of acute units working with crisis, and in secure units working with long-term problems. **Short-term** group therapy for selected conditions requires careful composition of the group and an active, flexible therapeutic approach.

We have not tried to cover the range of group services and approaches for children. Chapters by Schamess[110] and by Kymissis[111] cover group work with children and adolescents, respectively. Further reading is available in Evan's text[112] and Melzack.[113]

There are economic arguments for group therapy. In one study, the quality of improvement between individual and group therapy of psychiatric patients was not significantly different but the cost of the service was different. Cost savings were calculated as the reduction in medical consultations and hospital attendance, and lost workdays. For those treated with psychotherapy of any kind, the cost of treatment was 25 per cent per patient less than it was for those who did not receive psychotherapy. The cost of group psychotherapy per patient was about a third less than for individual psychotherapy.[114]

Further information

A. Bateman, D. Brown and J. Pedder's (2000) *Introduction To Psychotherapy* (Routledge) sets group therapy in the context of other psychotherapies. D. Stock Whittaker's (1995) introduction, *Using Groups To Help People* (Routledge) and M. Aveline and W. Dryden's (1988) *Group Psychotherapy In Britain Today* (Open University Press) give a general overview of UK practice in the past. There are three good texts that introduce the group-analytic approach. The most recent by H, Behr and L. Hearst (2005) is *Group Analysis: A Meeting of Minds* (John Wiley). W. Barnes, S. Ernst and K. Hyde have written a recent overview, (1999) *An introduction to groupwork: a group-analytic perspective* (Palgrave Macmillan). And an earlier text by D. Kennard (1993) *The Workbook of Group Analysis* (Routledge) provides a clinically grounded study of practitioners at work. The range of other books in the International Library of Group Analysis and the journals, *Group Analysis, Group, and The International Journal of Group Psychotherapy* give access to the many specialist applications discussed here.

References

1. Burlingame, G.M., Mackenzie, K.R., Strauss, B. (2005). Small-group treatment. Evidence for effectiveness and mechanisms of change. Ch.14. In *Bergin and Garfield's Handbook of psychotherapy and behaviour change*. (ed. M.J.Lambert), New York. Wiley.
2. Foulkes, S.H. and Anthony, J. (1957*). Group psychotherapy: the psychoanalytic approach.* Penguin, Harmondsworth. (Maresfield Reprint, Karnac Books London 1989.
3. Yalom, I. and Leszcz, M. (2005). *The theory and practice of group psychotherapy.* Basic Books, New York.
4. Agazarian, Y. (1997). *Systems-centred therapy for groups*. Guilford Press, New York.
5. Creek, J. (ed.) (1997). *Occupational therapy and mental health*. Churchill Livingstone, Edinburgh.
6. Stone, W. (1996). *Group psychotherapy for people with chronic mental illness*. Guilford, New York.
7. Kanas, N. (1999). Group therapy with schizophrenic and bipolar patients. In G*roup psychotherapy of the psychoses* (eds. V. Schermer and M. Pines), pp. 129–47. Jessica Kingsley, London.
8. Schlapobersky, J. (1993). The language of the group: monologue, dialogue and discourse in group analysis. In *The psyche and the social world: developments in group- analytic theory* (eds. D. Brown and L. Zinkin), pp. 211–31. Routledge, London.
9. Schlapobersky, J. (1996). A group-analytic approach to forensic psychotherapy: from the speech of hands to the language of words. In *Forensic psychotherapy: crime, psychodynamics and the offender patient. Vol. 1: Mainly theory* (ed. C. Cordess and M. Cox), pp. 227–43. Jessica Kingsley, London.
10. Abse, W. (1979). Trigant Burrow and the inauguration of group analysis in the USA. *Group Analysis,* **3**, 218–29.
11. Lewin, K. (1951). *Field theory and the social sciences.* Harper and Rowe, New York.
12. Trist, E., Murray, H. (1990). T*he social engagement of social science Vol. 1, section 1.* University of Pennsylvania Press, Philadelphia, PA.
13. Pines, M. (1991). A history of psychodynamic psychiatry in Britain. In *Textbook of psychotherapy in clinical practice* (ed. J. Holmes), pp. 75–86. *C*hurchill Livingstone, Edinburgh,
14. Harrison, T. (2000). *Bion, Rickman, Foulkes and the Northfield Experiment: Advancing on a different front.* Jessica Kingsley, London.
15. Elliott, A. (1999). *Social theory and psychoanalysis in transition. Chapter 2, Free Association Books,* London. pp. 46–76.
16. Foulkes, S.H. (1948). *Introduction to group analytic psychotherapy.* Heinemann, London. (Maresfield Reprint, Karnac Books, London 1991.)
17. Dick, B. M. (1975). A ten year study of out-patients analytic group therapy. *British Journal of Psychiat*ry **127**, 365–75. Lorentzen, S. (2000) An assessment of change after long-term psychoanalytic group treatment. *Group Analysis* (in press).
18. Sigrell, B. (1992). The long-term effects of group psychotherapy. A thirteen year follow up study. *Group Analysis,* **25**, 333–52.
19. Lorentzen, S. (2000). An assessment of change after long-term psychoanalytic group treatment. *Group Analysis*, **33**.
20. Bion, W.R. (1961). E*xperiences in groups*. London: Tavistock.
21. Malan, D. (1976). A follow up study of group psychotherapy. *Archives of General Psychiatry* **33**, 1303–15.
22. Moreno, J. L. (1953). *Who shall survive? Foundations of sociometry, group psychotherapy and psychodrama*. Beacon House, New York.
23. Slavson, S. (1940). Group psychotherapy. *Mental Hygiene,* **24**, 36–49.
24. Wolf, A. (1949). The psychoanalysis of groups—1. *American Journal.of Psychotherapy,* **3**, 525–58.
25. Wolf, A. (1950). The psychoanalysis of groups—2. *American Journal of Psychotherapy,* **3**, 16–50.
26. Yalom, I. (1983). *In-patient group psychotherapy*. Basic Books, New York.
27. Rutan, J.S. and Stone, W. (2007). P*sychodynamic group therapy* (Fourth Edition). Guilford Press, New York.
28. Kaplan, H.I. and Sadock, B.J. (eds) (1993). C*omprehensive group psychotherapy*. Williams and Wilkins, Baltimore, MD.
29. Alonso, A. and Swiller, H.I. (eds) (1993). *Group therapy in clinical practice.* American Psychiatric Press, Washington, DC.
30. Ashbach, C. and Schermer, V. (1987). *Object relations, the self and the group.* Routledge, London.
31. Spotnitz, H. (1961). T*he couch and the circle*. Knopf, New York.
32. Ormont, L.R. (1992). *The group therapy experience: from theory to practice.* St. Martin's Press, New York.

33. Tubert-Olander, J. and de Tubert, R.H. (2004). *Operative Groups.* Routledge, London.
34. Janunsen, P. (1994). *Psychoanalytic therapy in the hospital setting.* Routledge, London.
35. Koenig, K. and Lindner, W.V. (1994). *Psycho-analytic group therapy.* Jason Aaronson, New York.
36. Tschuschke, V. (2000). The P.A.G.E. study: early treatment effected of long-term outpatient group therapies-first preliminary results. *Group Analysis*, **33**, 3, 397–411.
37. Tsegos, I.K. (1995). Further thoughts on group-analytic training. *Group Analysis,* **28**, 313–26.
38. S. Lorentzen (1990). Block training in Oslo: the experience of being both organiser and participant in the Norwegian Psychiatric Association group psychotherapy training programme. *Group Analysis*, **23**, 361–82.
39. May, R. and Yalom, I.D. (1989). Existential Psychotherapy. In *current psychotherapies* (eds. R. Corsini and D. Wedding.) pp. 363–402. F.E. Peacock, Itasca, Ill.
40. Bion, W. (2000). *Experiences in Groups*. Routledge, London (First published by Tavistock, 1961).
41. Ezriel, H. (1950). A psycho-analytic approach to group treatment. *British Journal of Medical Psychology*, **23**, 56–74.
42. Horwitz, L. (1977). A group-centred approach to group psychotherapy. *International Journal Group Psychothererapy*, **27**, 423–39.
43. Whitaker, D.S. and Lieberman, M.A. (1964). P*sychotherapy through the group process*. Tavistock, London.
44. Pines, M., Hearst, L. and Behr, H. (1982). Group analysis (group analytic psychotherapy). In *basic approaches to group psychotherapy and group counselling.* (ed. G. Gazda), pp. 132–178. Charles C. Thomas, Springfield Ill.
45. Pines, M. (1982). Reflections On Mirroring: 5th. Foulkes Lecture. *Group Analysis,* **15** (Supplement). Reprinted in Pines, M. (1998) *Circular reflections: selected papers on group analysis and psychoanalysis*. Jessica Kingsley, London, pp. 17–39.
46. Agazarian, Y., Peters, R. (1981). *The visible and invisible group: two perspectives on group psychotherapy*. Routledge, London.
47. Foulkes, S.H. (1964). *Therapeutic group analysis*. Allen and Unwin, London. (Maresfield Reprint, Karnac, London1984).
48. Pines, M. and Hearst, L. (1993). Group Analysis in *Comprehensive Group Psychotherapy.* (eds Kaplan I. and Sadock, J.) Williams & Wilkins, Baltimore.
49. Friedman, R. (2008). Dreamtelling as a request for containment: three uses of dreams in groups. *International Journal of Group Psychotherapy*, **58**, 3, 327–44.
50. Lipgar, R. and Pines, M. (2003). *Building On Bion: Vol. 1: Roots; Vol. 2*: Branches. Jessica Kingsley, London.
51. Nitsun, M. (2006). *The group as an object of desire: exploring sexuality in group therapy.* Routledge, London.
52. Cox, M. and Theilgaard, A. (1987). *Mutative metaphors in psychotherapy: the Aeolian mode.* Tavistock Press, London.
53. Bennis, W.G. and Shepard, H.A. (1956). A theory of group development. *Human Relations,* **9**, 415–37. Reprinted in *Sensitivity training and the laboratory approach (eds. R. Golembiewski and A. Blumberg)* pp. 91–115. *F.E. Peacock, Itasca, Ill.1970.*
54. Foulkes, S.H. (1964). *Therapeutic group analysis.* George Allen and Unwin, London. pp. 51–2, 176–77.
55. Foulkes, S.H. (1964). *Therapeutic group analysis*. George Allen and Unwin, London. p. 57.
56. Tschuschke, V. and Dies, R.R. (1994). Intensive analysis of therapeutic factors and outcome in long-term inpatient groups. *International Journal of Group Psychotherapy,* **44**, 185–208.
57. Marziali, C. et al. (1997). The contributions of group cohesion and group alliance to the outcome of group psychotherapy. *International Journal Group Psychotherapy*, **47**, **4**, 475–97.
58. Pines, M. (1986). Coherence and its disruption in the development of self. *British Journal of Psychotherapy*, **2**, **3**, 180–85. Reprinted in Pines, M. (1998) *Circular Reflections: selected papers on group analysis and psychoanalysis,* Chapter 12. Jessica Kingsley, London. pp. 211–23.
59. Grotjahn, M. (1977). *The art and technique of analytic group therapy.* Jason Aaronson, New York. p. 14.
60. Pines, M. (1990). Group analysis and the corrective emotional experience: is it relevant? *Psychoanalytic Inquiry*, **10**, **3**, 389–408.
61. Garland, C. (1982). Group analysis: taking the non-problem seriously. *Group Analysis,* **15**, 4–14.
62. Pines, M. (1995). The universality of shame: a psychoanalytic approach. *British Journal of Psychotherapy*, **11**, 346–57.
63. Nitsun, M. (1996). *The anti-group: destructive forces in the group and their creative potential.* Routledge, London.
64. Salvendy, J.T. (1993). Selection and preparation of patients and organisation of the group. In *Comprehensive group psychotherapy* (ed. H.I. Kaplan and B.J. Sadock). Williams and Wilkins, Baltimore, MD.
65. Hearst, L. (1998). The restoration of the impaired self in psychoanalytic treatment. In *Borderline and narcissistic patients in treatment* (ed. N. Slovinska-Holy) International University Press, New York. Chapter 7.
66. Roback, H.B. and Smith, M (1987). Patient attrition in dynamically oriented treatments in groups. *American Journal of Psychotherapy*, **144**, 426–31.
67. Melnick, J. and Woods, M. (1976). Analysis of group composition: research and theory for psychotherapeutic and growth-oriented groups. *Journal of Applied Behavioural Science*, **12**, 493–512.
68. Thygesen, B. (1992). *Diversity as a group-specific factor. Group Analysis* **25**, 175–86.
69. Zelakowski, P. (1998). The sub-optimal group. *Group Analysis*, **31**, 491–504.
70. Stern, M.J. (1993). Group therapy with medically ill patients. In *Group therapy in clinical practice* (ed. A. Alonso and H.I. Swiller), pp. 185–200. American Psychiatric Press, Washington, DC.
71. Leszcz, M. and Goodwin, P. (1998). The rationale and foundation of group psychotherapy with metastatic breast cancer. *International Journal of Group Psychotherapy,* **48**, 245–73.
72. Spiegel, D., Bloom, J.R. and Yalom, I.D. (1981). Group support for patients with metastatic cancer. *Archives of General Psychiatry*, **38**, 527–33.
73. Spiegel, D. (1999). *Supportive group therapy with cancer patients.* Basic Books, New York.
74. Kanas, N. (1999). Group therapy with schizophrenic and bipolar patients. In *Group psychotherapy of the psychoses* (eds. V. Schermer and M. Pines) Jessica Kingsley, London.
75. Kapur, R. (1993). The effects of group interpretations with the severely mentally ill. *Group Analysis,* **26**, 411–32.
76. Resnik, S. (1987). *The theatre of the dream.* The New Library of Psychoanalysis, London; and (1995) *Mental Space*. Karnac. London.
77. Chiesa, M. and Fonagy, P. (2000). The Cassel Hospital personality disorder study. *British Journal of Psychiatry (forthcoming).*
78. Wilborg,T. and Karterud, S. (2001). The place of group psychotherapy in the treatment of personality disorders. *In Current Opinion in Psychiatry,* **14/2**, 125–30.
79. Karterud, K., Oyvind, U. (2004). Short-term day programmes for patients with personality disorders. What is the optimal composition? *Nordic Journal Psychiatry,* **58**(3) 243–49.
80. Cox, M. (1986). The 'holding function' of dynamic psychotherapy in a custodial setting: a review. *Journal of the Royal Society of Medicine*, **79**, 162–64.
81. Kennard, D. (1993). Group therapy at Rampton. Group Work Co-ordinating Committee, Rampton Hospital.
82. Genders, E. and Player, E. (1995). *Grendon: A study of a therapeutic prison*. Clarendon, Oxford University Press.

83. Welldon, E. (1994). Forensic Psychotherapy. In *Handbook of psychotherapy*, (eds. P. Clarkson and M. Pokorny), pp. 470–93. Routledge, London.
84. Lees, J., Manning, N. and Rawlings, B. (2000). Therapeutic community effectiveness: A systematic international review of therapeutic community treatment for people with personality disorders and mentally disordered offenders. *CRD Report 17: University of York NHS Centre for Reviews and Disemmination.*
85. Brown, A., Caddick, B. (eds) (1993). *Groupwork with offenders*. Whiting and Birch, London. p. 2.
86. Welldon, E. and Wilson, P. (2006). Special edition of Group Analysis, Forensic Psychotherapy. *Group Analysis*, 39.1.
87. Van der Kolk, B. (1993). Groups for patients with histories of catastrophic trauma. In *Group therapy in clinical practice* (ed. A. Alonso and H.I. Swiller). American Psychiatric Press, Washington, DC. pp. 289–305.
88. Hall, Z., Mullee, M. and Thompson, R. (1995). A clinical and service evaluation of group therapy for women survivors of childhood sexual abuse. In *Research Foundations for Psychotherapy Practice* (eds. M. Aveline and D.A. Shapiro). Wiley, Chichester.
89. Lifton, R. (1983). *The broken connection*. Basic Books, New York.
90. Greene, L.R. (ed.) (2005). Special Edition of the International Journal of Group Psychotherapy: Children & Adolescents in the Aftermath of 9/11: Group Approaches Towards Healing, Trauma and Building Resilience, July 2005.
91. Weinberg, H. and Nuttman-Shwartz, O. (eds) (2005). Special Edition of Group Analysis on Trauma, Vol. 38 No 2, June 2005.
92. Schlapobersky, J. and Bamber, H. (1988). Rehabilitation with victims of torture. In *Refugees—The trauma of exile*. pp. 206–22. Nijhoff, The Hague.
93. Woodcock, J. (1997). Groupwork with refugees and asylum seekers. In *Race and groupwork*, (eds. T. Mistry and A. Brown), pp. 254–77. Whiting and Birch, London.
94. Callaghan, K. (1996). Torture—the body in conflict: The role of movement psychotherapy. In *Arts Approaches to Conflict*, (ed. M. Liebmann). Jessica Kingsley, London.
95. Benun, I. (1986). Group Marital Therapy: A Review. *Sexual and Marital Therapy*, **1**, 61–74.
96. Skynner, A.C.R. (1986). Recent developments in marital therapy. In *Explorations with families: group analysis and family therapy: selected clinical papers of Robin Skynner*. Routledge, London.
97. Lesecz, M. (1997). Integrated group psychotherapy for the treatment of depression in the elderly. *Group* **21**, 89–113.
98. Dies, R.R. (1993). Research on group psychotherapy: overview and clinical applications. In *Group therapy in clinical practice*, (eds. A. Alonso and H.I. Swiller), pp. 475–6. American Psychiatric Press, Washington DC.
99. Roback, H. (2000). Adverse outcomes in group therapy. *Journal of Psychotherapy Practice and Research*, **9**, 113–22.
100. Robinson, L.A., Berman, J.S. and Niemeyer, R. (1990). Psychotherapy for the treatment of depression: a comprehensive review of controlled outcome research. *Psychological Bulletin,* **108**, 30.
101. Tyllitski, C.J. (1990). A meta-analysis of estimated effect sizes for group versus control treatments. *International Journal of Group Psychotherapy,* **40**, 215–24.
102. Piper., W.E., Perrault, E.L. (1989). Pre-therapy preparation for group members. *International Journal of Group Psychotherapy* **39**, 17–34.
103. Lorentzen, S. and Heglund, P. (2004). Predicting change in long-term group psychotherapy. *Psychotherapy and Psychosomatic Medicine,* **73**, 1. 125–135; and Lorentzen, S. (2006) Contemporary Challenges For Research. In *Group Analysis*, 3, 321–340; and Mace, C., (2006) Setting the world on wheels: some clinical challenges of evidence-based practice. In *Group Analysis*, 39, 3, 304–20.
104. Dies, R. R. (1993). Research on group psychotherapy: overview and clinical applications. In *Group therapy in clinical practice*, (eds. A. Alonso and H.I. Swiller), p. 502. American Psychiatric Press, Washington DC.
105. Tschuschke, V. and Dies, R.R. (1994). Intensive analysis of therapeutic factors and outcome in long-term inpatient groups. *International Journal of Group Psychotherapy*, **44**, 183–214.
106. Tschuschke, V., Anbeh, T., Kiencke, P. (2007). Evaluation of long-term analytic outpatient group therapies. In *Group Analysis,* **40**, **1**, 140–59.
107. Lorentzen, S., Bogwald, K. and Hogland, P. (2002). Change during and after long-term analytic psychotherapy. *International Journal of Group Psychotherapy*, **52**(3) 419–29.
108. Terlidou, C., Moschonas, D., Kakitsis, P., Manthouli, M., Moschona, T., Tsegos, I. (2004). Personality changes after the completion of long-term group psychotherapy. *Group Analysis,* **37**, **3**, 401–18.
109. Conway, S., Audin, K., Barkham, M., Mellor-Clark, J. and Russell, S. (2003). Practice based evidence for a brief time-intensive multi-modal therapy guided by group-analytic principles and methods. *Group Analysis* **36**, **3**, 413–35.
110. Schamess, G. (1994). Group psychotherapy with children. In *Comprehensive group psycho*therapy (ed. H.I. Kaplan and B.J. Sadock). Williams and Wilkins, Baltimore, MD. pp. 560–77.
111. Kymissis, P. Group psychotherapy with adolescents. In *Comprehensive group psychotherapy* (ed. H.I. Kaplan and B.J. Sadock), pp. 577–84. Williams and Wilkins, Baltimore, MD.
112. Evans, J. (1998). *Active analytic group therapy for adolescents*. Jessica Kingsley, London.
113. Melzak, S. (ed.) (2000). *Children in exile: therapeutic and psychotherapeutic work in the clinic and the community.* Jessica Kingsley, London.
114. Heinzel, R. (2000). Outpatient psychoanalytic individual and group psychotherapy in a nationwide follow-up study in Germany. *Group Analysis*, **33**, 353–72.

6.3.7 Psychotherapy with couples

Michael Crowe

Introduction and background

There is an ongoing crisis in the institution of marriage, at least in Western cultures. There has for some time been a tendency to idealize marriage, and at the same time social forces are operating which tend to undermine it.(1) These influences have probably made a contribution to the increasing divorce rate, as well as the tendency for fewer couples to marry, and have probably also led to an increase in the number of couples seeking help with their relationships.

In the United Kingdom, for example, the number of marriages taking place each year has fallen for the first time in living memory, and the number of divorces is still steadily increasing, reaching 40 per cent of marriages in 1996.(2) There are also a large number of 'common-law' marriages, often with children, as well as more transient cohabiting or non-cohabiting sexual relationships, both heterosexual and homosexual. The stability of these relationships is, of course, not recorded in the marriage or divorce statistics, and the rate of breakup can only be guessed at; however, it is very probable from clinical experience that in these non-marital relationships there is a higher than 40 per cent incidence of breakup. In the wake of these changes, there are a large number of

single-parent families and 'reconstituted' or blended families, as reviewed by Robinson,[3] and there is a decreasing proportion of children who are being brought up in the traditional nuclear family with two biological parents.

In addition to these new factors affecting marriage in the early 21st century we should also be aware of the fact that many countries, especially those in the developed world, have a multicultural society, and that immigrant cultures have different attitudes to marriage and family life. For example, families from the Indian subcontinent often prefer to arrange marriages for their children, and in some cases insist that the couple live in the husband's parents' house. On the other hand, West African couples often leave their children in Africa to be looked after by family members for long periods of time, while the parents work or study in the West.

In the last few years, Gay marriage or Civil Partnership has been recognized in many western countries. Couples in Gay relationships have many of the same problems and satisfactions as heterosexual couples, and in addition, must live with fairly widespread negative attitudes and homophobia from neighbours, family and society generally. Their relationships have to be, if anything, stronger than heterosexual ones to survive these pressures, and may be more in need of therapy.

Couple therapy must be able to take account of these factors, and whilst much of what is contained in this chapter will relate to heterosexual married British couples living with their biological children, it should be understood that there are many other types of relationship which can be helped using a similar approach, with appropriate changes of emphasis. In a later section, there will be some additional discussion of the specific problems relating to couples from other cultures, and ways of managing these.

Couple counselling and couple therapy

The concept of couple counselling dates from the 1920s when in the United States the American Association for Marital Counselling was formed; in the United Kingdom the Marriage Guidance Council (now called Relate) was founded in 1938. Counselling mainly took the form of giving advice on practical issues, but in more recent years, Relate counselling has been orientated more towards psychodynamic approaches, and favours a longer-term involvement with the couple. Couple counselling continues in both countries, and the great majority of couples seeking help with their relationships are seen by couple counsellors, rather than any other types of therapist.

The distinction between couple counselling and couple therapy is not an easy one, because many of the interventions are similar. In a simplistic sense therapy attempts to make a more radical difference to the couple's functioning than counselling, which has the general aim of improving adjustment to the situation as it is. However, many forms of both couple therapy and couple counselling are based on a theoretical formulation which is derived from a related school of individual psychotherapy (for example, cognitive behavioural or psychodynamic). Thus, theoretical formulations in the resultant couple work are so different between therapies (e.g. psychodynamic as against behavioural) that a particular form of counselling may have more in common with a related form of couple therapy than that therapy itself has with another type of couple therapy.

Psychoanalytic/psychodynamic couple therapy

Couple therapy using a psychodynamic model began in the United Kingdom in 1948, when Dicks and his colleagues founded the Institute of Marital Studies. The theories and techniques involved have been ably reviewed by Daniell[4] and Clulow.[5] The central concept used is that the inner (unconscious) world of the two partners determines their interaction and their response to changing circumstances. It is as though each partner has an internal blueprint, both of themselves and each other, formed partly by observation but also partly by the influence of earlier intimate attachment experiences with parents, siblings, or friends. These influences may actually determine the choice of partner, and the nature of each partner's patterns of attachment (secure or insecure) will affect the ways in which they cope with the stresses of the new relationship. There may then be projections which lead one partner to attribute motives such as hostility or sadism to the other, whereas in fact this is a split-off and denied characteristic of the first partner. Other consequences of this unconscious process may include the system of shared fantasies and defences which builds up as the relationship continues.

In therapy, four premises are used, which inform a relatively long-term and open-ended series of sessions. The first is that a person's emotional health is related to his or her capacity to manage both internal conflict and external stress: it is important to be able to experience fear as well as trust, pain as well as pleasure, doubt as well as certainty, frustration as well as satisfaction. Secondly, significant relationships can be used to resurrect, but also change, inflexible patterns of behaviour established in the past. Thirdly, unconscious processes need to be taken into account when attempting to understand problems in relationships. Fourthly, change takes time because it requires a reordering of perceptions of self and others, perhaps with the help of transference interpretations by the therapist involving both partners.

Therapy in this mode may be carried out by one therapist seeing both partners, but is more often done by two therapists either seeing the couple together or in parallel individual sessions, using one partner with one therapist, with joint supervision of the two therapists. An intriguing aspect of this therapeutic format is that sometimes the two cotherapists find themselves interacting in unfamiliar ways, in sessions and between sessions, which are thought to represent the projection of fantasies and feelings by the couple on to the therapists; the therapists' understanding of these projections in their joint supervision may play a role in advancing the therapy itself. If these insights are used to inform the therapists' interaction with the couple, the individual partners may then be made aware of their own conflicts, fantasies, and projections, and thus be able to give up some of their repetitive patterns of behaviour and withdraw damaging projections.

The psychoanalytic approach has been an important source of theoretical ideas in couple therapy, especially the concepts of attachment and loss developed by Bowlby.[6] It has also the distinction of being the first theory to be adapted to this area of work. There are, however, some drawbacks to working in this way, as enumerated by Wile.[7] He sees the emphasis on negative impulses and emotions (e.g. dependence, narcissism, sadism, manipulation, and exploitation) as painting a rather unflattering and negative picture of the couple in therapy, and perhaps therefore reducing

their motivation to continue. A more serious problem with the approach is that the psychodynamic concepts, whether of defence mechanisms, projections, or shared fantasies, are treated as if they were as real as observed behaviour, whereas in fact they must remain assumptions based on hypothetical constructs, and are really only valuable in so far as the therapy based on them is effective.[1]

The question of efficacy is raised later in the chapter, but it must be stated here that the psychodynamic therapies for couple problems have only seldom been submitted to controlled trial, and then usually in a relatively short-term form. The therapy may be quite long term, and the improvements seen are usually not dramatic, so that in the last analysis the approach has to remain of uncertain value.

Behavioural couple therapy

The behavioural approach, in contrast, makes no assumptions about internal conflicts or underlying mechanisms in the individuals. The approach was initiated in 1969 by Stuart[8] and Liberman[9] as behavioural marital therapy. They worked from the principles of operant conditioning and made the assumption that couples who were having difficulties were either giving each other very low levels of positive reinforcement or were using punishment or negative reinforcement to coerce each other into behaving differently. The remedy that they proposed for this situation was to help the partners to learn how to persuade each other to conform to the desired pattern of behaviour by the use of prompting and positive reinforcement. Thus, complaints would be transformed into requests and requests into tasks agreed by both partners.

Behavioural marital therapy relies on the therapist's observation of the couple's behaviour in the session and on the problems they report from the previous week or equivalent timespan. There are two types of therapeutic activity in behavioural marital therapy. The first is reciprocity negotiation, in which the partners request changes in behaviour on each side and negotiate how this can be achieved through mutually agreed tasks. The second is communication training, in which the partners are encouraged to speak directly and unambiguously to each other about feelings, plans, or perceptions, and to feed back what they have heard and understood. In both these approaches, the deeper meanings behind a particular piece of behaviour are ignored, the emphasis being on change in the interaction both in the here and now and in the immediate future. The approach has been the subject of many controlled trials (see below), and is of proven efficacy.

Cognitive behavioural and rational–emotive couple therapy

Aaron Beck,[10] in his cognitive behavioural approach to couple therapy, identifies in the communication of disturbed couples many of the problems found in the thinking of depressed patients, and attempts to correct these. Thus, he tackles misunderstandings, generalizations, untested assumptions, and automatic negative thoughts by challenging assumptions, reducing unrealistic expectations, relaxing absolute rules, improving the clarity of the communication and focusing on the positive rather than the negative.

Similarly, Albert Ellis (reviewed by Dryden[11]) uses a rational-emotive approach to couple problems. Here, the main focus is on the use of words; terms such as 'intolerable' are replaced by (for example) 'difficult to accept', and the couple are encouraged to express desires rather than demands. There is an analysis of the repetitive cycles of cognitive and behavioural disturbance, in which each partner may attribute the other's behaviour to a negative motive and assume that nothing can be done about it. The general thrust of this therapy is similar to that of Beck, but with a more lively and less formalized approach in the session.

Systems therapy for couple problems

The systems approach to couple therapy derives partly from concepts developed by Minuchin[12] and Haley,[13] and partly from the work of Selvini Palazzoli *et al.*[14] All these pioneers worked predominantly with families rather than couples, but many of their ideas and techniques are relevant to the treatment of couples. Although the systems approach to therapy has broadened and deepened since the 1980s, many of the early concepts are still very useful.

A central concept in thinking about couple relationships is 'enmeshment', by which is meant an excessive involvement in what is essentially the private business of another person. It is quite common to find an enmeshed relationship between parents and their teenage children, in which both sides find it very hard to 'let go'. It can also be found in couple relationships where one partner wants to be closer than the other, and a conflict arises as to what is the best distance to maintain. Systems therapy aims to help them to find a compromise 'distance' which suits them both, and thereby to reinforce the necessary 'boundaries' which people need in maintaining their individuality within a relationship.

The concept of circular causality is also central to systems work. This enables the couple to get away from the idea that one person is necessarily to blame for a particular situation by considering the continuous cycle of cause and effect in which A's actions may be caused by B's actions and also B's may equally be caused by A's. Thus, systems therapists, when approaching a couple problem, do not focus on one partner's behaviour, but rather on the pattern of interaction obtaining in the relationship. They will then try to effect a change in which both partners contribute actively to the solution of the problem.

Systems therapists have many techniques at their disposal, including those which increase the couple's understanding of the system they are participating in. These include family genograms (a form of family tree construction which leads to discussion of transgenerational influences or 'systems over time'), family 'sculpting' (in which the members position themselves and each other wordlessly to represent their current relationships), and the discussion of 'family myths' and stories. More active techniques, designed to play a part in changing the family interaction, include creating conflict in the session, giving homework tasks, and the use of 'paradoxical injunctions' in which the therapist tells the family to continue with the current interaction because, even though it is problematic, it seems to be protecting them from worse consequences. These more active techniques will be dealt with in more detail in the main part of this chapter on behavioural–systems therapy.

Mixed or eclectic approaches

Most couple therapists use a mixture of techniques, and it seems that this is probably an inevitable consequence of the difficulties

involved in applying one therapeutic method rigorously in a clinical setting. A number of specific combinations have been advocated, and will be briefly mentioned here.

The first is the psychodynamic-behavioural approach of Segraves.[15] In this, the basic underlying cause of marital disturbance is assumed to be the partners' conflicting internal and unconscious projections, and their interactions. The therapy, however, is not only directed at helping them to understand these (as in psychodynamic therapy) but also to increase their negotiating and communicating skills (as in behavioural marital therapy).

The second is a more comprehensive mixture of theory and technique, known as the intersystem model, and advocated by Weeks.[16] This tries to take account of the individual, interactional, and intergenerational aspects of couple relationships, and combines them in what is probably closest to a systems model, but with more emphasis on the psyche of the individual. Interventions are on both a conjoint and individual basis, and the techniques of decentring (see below) and paradoxical injunctions are often used.

The third eclectic approach is that of Spinks and Birchler.[17] This is called behavioural–systems marital therapy, and makes use of behavioural marital therapy as the main form of intervention, moving into the systems mode when 'resistance' emerges. There are many similarities between this form of treatment and the one described in the main part of this chapter, but our 'behavioural–systems approach' is more integrated as between the two components of the method.

The fourth eclectic approach which should be mentioned is that of Berg-Cross.[18] She uses rational-emotive, sociocognitive, systemic, psychodynamic, humanistic, and theological concepts to understand and modify couple relationships. Like that of Weeks, her approach gives the therapist a very wide canvas to work on, but may lose some of the focus by being very general and all-embracing.

The behavioural–systems approach to couple therapy

Behavioural–systems couple therapy is the approach that will be described in detail in the present chapter, and although it is only one of several approaches to couple problems, it has the advantage of spanning two of them, and issues such as indications for therapy and assessment are shared with both the pure behavioural and systemic approaches. It is the method developed at the Maudsley Hospital Couple Therapy Clinic in the 1980s. It has been expounded at greater length by Crowe and Ridley,[1] and like some of the other eclectic models mentioned above it combines two different approaches, behavioural marital therapy and systems family therapy. The behavioural dimension, similarly to that described by Stuart (8) and Jacobson and Margolin,[19] consists of the relatively straightforward methods of reciprocity negotiation and communication training. The systems dimension is more complicated, and involves systems thinking, structural moves during the session, tasks and timetables for the couple between sessions, and the use of paradox. The method was developed in a predominantly psychiatric setting, and has been found to be particularly suitable for those couples where one or both partners has psychiatric problems in addition to their relationship difficulties. It is also useful as an adjunct to psychosexual therapy where a sexual dysfunction or a sexual motivation problem seems to be connected with relationship issues.

The method should be thought of as a series of menus from which the practitioner can choose techniques rather than as a set course of therapy beginning at one point and ending at another. Thus, the various components of behavioural–systems couple therapy can be incorporated at any time in the therapy session, although in practice, negotiation, communication training, and structural moves are usually employed in the earlier part of the session, while tasks, timetables, and paradox are usually reserved for the 'message' at the end, and are linked to homework assignments to be carried out between sessions.

The different techniques of behavioural–systems couple therapy can be thought of as belonging to a kind of hierarchy. The so-called 'hierarchy of alternative levels of intervention (ALI)'[1] links each type of intervention with a particular set of clinical problems and makes recommendations as to the type of intervention that is appropriate. The ALI hierarchy is shown in diagrammatic form in Fig. 6.3.7.1. As may be seen, where the couple appear to have greater rigidity in their behaviour, where they show more symptoms, and where they show more reluctance to accept the relationship as the focus of work, the therapist needs to move to the systems end of the hierarchy, and use more ingenuity in the development of interventions. If, however, the couple accept the interactional focus and show willingness to recognize the part that the relationship is playing in maintaining the problems, the therapist may be quite comfortable and effective working behaviourally. By and large, the preference is to work behaviourally, since this implies collaborating with the couple and accepting their stated goals, whereas the systems approach puts the therapist into a more managing role, deciding what is best for the couple and suggesting tasks that may not be what they would expect. It should be emphasized that the therapist may at any stage move up or down the hierarchy, according to the couple's response: an increase in flexibility shown by the couple could be the trigger for the therapist to begin working in a more behavioural way, whereas an increase in rigidity or a failure to respond to behavioural work could be met by a more systemic approach.

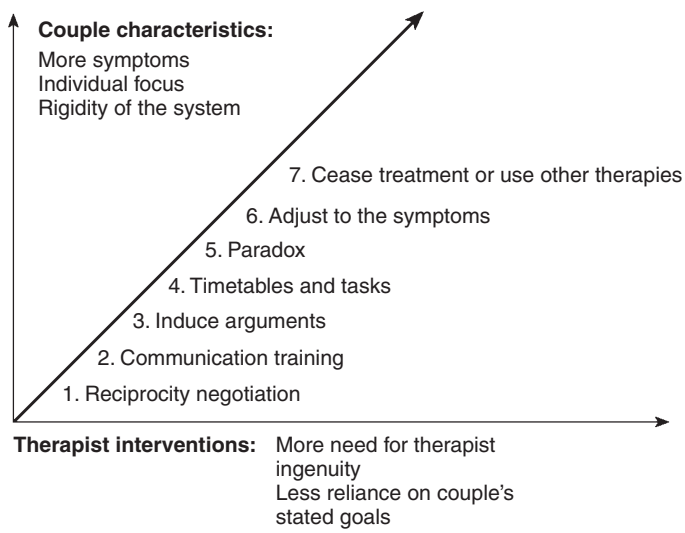


Fig. 6.3.7.1 The alternative levels of intervention hierarchy.

Indications and contraindications

If there is a relationship problem identified by either the couple or their advisers, even if there are also individual psychiatric or behavioural problems, and if the couple are willing to attend together, then in most cases they are suitable for behavioural–systems couple therapy. The breadth of the therapeutic approach, the fact that the behavioural techniques are of proven efficacy (see below), and the fact that the systemic interventions are suitable for those with more psychiatric symptoms or similar problem behaviours, all give the therapy a wide range of positive indications.

(a) The nature of the problem

Clearly those with relationship problems such as arguments and tensions are highly suitable for couple therapy. Another related indication is those relationships in which one partner (who might be attending a counsellor or psychiatrist alone) spends much time complaining about the absent partner's behaviour. A third indication is where the health of one partner suffers following the other partner's individual therapy.

Many problems with sexual function would be suitable for couple therapy, including those couples where there is a disparity in sexual desire, or those where one partner has a specific phobia for sex. In some such cases there is also a need for individual therapy, especially where one partner is the survivor of earlier childhood sexual abuse.

Many people with depression or anxiety, especially those where there is also poor self-esteem, may be suitable for couple therapy. There are often aspects of the illness that are exacerbated by problems in the relationship. Indeed, in a study by Leff *et al.* in 2000[20] it was shown that couple therapy was an effective and acceptable form of treatment for couples in whom one partner was depressed.

Where jealousy is present the problem usually affects the non-jealous partner to a greater or lesser extent, and here it would almost always be useful to have at least a few conjoint sessions with the couple, as suggested by De Silva.[21]

Some problems are perhaps less amenable to couple work, and among these are, for example, phobias which seem unconnected with home life in any way, and post-traumatic stress reactions where the event happened away from the partner. Some alcoholic and drug-addicted patients have so much of their existence involved with the addiction that they are not available emotionally to do couple work, and the work would at that stage be wasted on them. Similarly, those with an acute psychosis would, at the time they are acutely ill, be unavailable to this kind of therapy, and should not be offered it. However, in both cases, when the acute crisis is over and the addiction or psychosis is under control, it would be very appropriate to offer them some kind of couple therapy, even if this had limited aims and expectations. Some of the most useful psychological interventions in schizophrenia, after the acute illness has resolved, involve the nearest relative, as shown by McFarlane.[22]

(b) Degree of connection with the relationship

Some problems in individuals have been in existence long before they entered the present relationship. If this is the case, the therapist should consider whether it is best to embark on couple therapy or whether individual therapy would be better. However, even when there seems no causal connection with the relationship, the effect of the problem on the partner may be such as to warrant at least one or two couple sessions.[23]

(c) Availability and willingness for joint therapy

If the partner of a patient is unavailable or unwilling to attend for therapy, it may be appropriate to let the situation be, and not offer treatment. In some immigrant couples, for example those from the Indian subcontinent, there are cultural reasons given for a wife not attending therapy, and we usually have to respect these. However, in both situations it is sometimes right to put some pressure on the absent partner to attend, because the reasons for non-attendance may be relevant to the couple problems we are trying to treat.

(d) Is the relationship continuing?

If the person who is asking for therapy is going through a divorce or equivalent breakup of a relationship, it may not be appropriate or possible to treat both partners. However, in some cases there is good work to be done in arranging a more satisfactory breakup, in terms of domicile and care of any children. This 'mediation' work is increasingly being done, and many of the processes are similar to those of couple therapy.

Assessment and selection

This is not an easy process, because there is a dearth of research on the types of couple problem presenting for treatment and on the outcome of treatment itself. However, on the basis of the referral letter (usually from a GP or psychiatrist) the therapist will usually arrange a preliminary session with the couple. The only reason not to see the couple in the first instance is that they are not willing to attend as a couple, or the problem is seen clearly as an individual one (see above).

Another, more subtle, form of selection occurs during the first therapy session, when the couple are in contact with the process of therapy. In this session, the therapist considers the ability of the partners to empathize with each other, their pattern of communication, their stage in the 'family life cycle' (see below) with its associated stresses, and their flexibility in response to simple therapeutic interventions. He or she will attempt to use reciprocity negotiation or other straightforward techniques of therapy, partly as a treatment trial, and partly to see whether the couple is ready for this kind of intervention. If not, the therapist can move to a more systemic approach, or decide after a few visits that there is no future at this point in further therapy of this kind, and that something else (e.g. individual therapy) is needed (Fig. 6.3.7.1).

The process of therapy: beginning and continuing

Behavioural–systems couple therapy is essentially a short-term therapy involving perhaps 5 to 10 sessions of 60 minutes each, over a period of 3 to 6 months. Although the therapy was developed in a room with a one-way screen and live supervision, it is quite possible to use behavioural–systems couple therapy in any conventional consulting room without a team or live supervision. If the one-way screen is used, the therapist and the team (which may include students as well as experienced colleagues) begin by reading the referral letter and the biographical questionnaires which the couple will have completed, and discuss the case with a view to formulating the problem from an interactional point of view. This may involve, for example, thinking about the couple's stage in the 'family life cycle' (e.g. birth of the first child or the 'empty nest'),

any recent event such as a bereavement or a new relationship, or the diagnosis of a serious illness. Hypotheses about the possible causation of the recent problems are not necessarily thought of as 'true' explanations, but have the function of informing the therapist's thinking and suggesting what level of the ALI hierarchy to choose at the initial meeting and what strategy to employ in the session.

There are several important issues to address in the first session, which functions both as a kind of assessment and as the beginning of the therapeutic work. It is necessary:

- to remain in control;
- to develop rapport with both partners, without favouring either;
- to maintain the momentum of the session and the interactional focus;
- to maximize the opportunities for the couple to experience a change in the nature of their interaction.

A particularly useful move at some stage in the first session, and one which we almost always use, is to ask the partners to talk directly to each other rather than the therapist; this is the so-called 'decentring' technique originated by Minuchin.[12] Keeping this configuration for as much as possible of the session enables the therapist to observe the couple's typical pattern of interaction, to intervene as a 'theatrical producer' rather than a diplomatic negotiator, to avoid as far as possible taking sides, and to encourage the kind of negotiation which hopefully the couple will be able to carry on at home without the presence of a third party. It is not necessary to remain decentred for the whole session, but if it does not happen at all in the first session, an opportunity for effective work will have been lost.

It may be difficult to remain decentred in the face of pressures from one or both partners to talk to the therapist directly. One way to participate while still remaining decentred is to request the partners to ask each other questions to which the therapist would like to know the answer. For example, one might say: 'Could you ask your partner what she thinks about your coolness on this matter?' or 'Perhaps your partner disagrees with you; could you check with him?'. In this way the couple can continue to talk to each other in the decentred position without undue 'triangling in' of the therapist.

One situation which causes particular problems for the behavioural–systems therapist is where one partner persists with monologues, either about his or her own symptoms and problems or about the partner's unreasonable behaviour. This is a perfectly acceptable way to present a problem in one-to-one therapy, but in couple therapy it slows down the interaction and prevents the therapist focusing on the relationship. One way to overcome this is by decentring, but in some couples this is too difficult, and another possibility is to ask the non-verbal partner to comment on the spokesperson's problems. This can provoke a minor crisis in the couple, and lead to the spokesperson realizing that the problems are not all one-sided. It is a technique related to circular questioning, and is used fairly extensively in family therapy see Chapter 6.3.8

Another obstacle to progress is the situation in which the partners have intractable arguments, perhaps about other family members. If such battles continue to dominate the sessions the therapist may have to devise a method for putting them 'on ice for the time being' and concentrating instead on negotiating everyday problems to do with the house or the children.

Momentum may also be slowed by the therapist's own style of working. Students of the behavioural systems approach may need to 'unlearn' some of their otherwise good therapeutic habits such as being a good and empathetic listener. They may be able to achieve this by decentring or by asking circular questions in order to refocus on the interaction and increase momentum.

On the other hand, it is still important for the therapist to be able to feel and show empathy to the two individuals in therapy. The difficulty may be that to show empathy to one partner may be interpreted by the other as side-taking. A possible remedy lies in a particular skill, which, however, is not easy to acquire, of saying something to show that one has understood one partner without antagonizing the other. One way to develop this skill is for the therapist not to become emotionally involved in the issues, but to concentrate on the process of interaction, thinking all the time in terms of balance and communication rather than worrying about the rights and wrongs of what is being discussed. This must be done while still showing respect to both the partners, taking their problems seriously, and at the same time conveying hope that they can be solved.

It is also very helpful for the partners themselves to be in touch with each other emotionally and take each other seriously. Some individuals are very good at communicating, but only at an intellectual level, and cannot express empathy with each other. Their interaction is like that of fellow committee members, and they tend to suppress any expression of feelings such as sadness or anger which are 'not on the agenda'. In other couples, there is an imbalance, with one partner expressing feelings openly and the other being exclusively logical and self-controlled. In both situations it is necessary to help them to communicate both intellectually and emotionally, encouraging the self-controlled individual to be more open and the person who is emotionally open to try to be more restrained at times.

Ending the session

Every session must have an ending, and the purpose in the behavioural–systems approach is to send the couple away with something to work at in the weeks before the next session. About two-thirds of the way through a supervized session the therapist will usually go behind the one-way screen, turning off the camera and closing the shutters (although when working alone this luxury is not available and the therapist must work independently to end the session on a positive note). The team and the therapist then spend about 15 minutes in discussion, planning the 'message' to be given to the couple at the end. Part of this discussion will centre round the team's thinking about the significance of the problems from a systems point of view, but part will also be concerned with how the therapist can help the couple to change their interaction.

The introduction to the message will give the date of the next appointment, and usually also contains some positive and sympathetic comments for both partners. It is important as far as possible to keep them both 'on side' at this stage, so that it would be unwise to say something which could be seen by either partner as favouring the other one. A good example of an introductory comment would be: 'The team and I are aware of the great difficulties you are experiencing, but we think you have what is basically a good relationship, and you are both working hard to improve things'.

The message itself will vary according to the level of the hierarchy at which the therapy is being pitched. If it is mainly a session of reciprocity negotiation the main theme may be simply to reiterate the negotiated plans for both partners which have emerged from the earlier discussion. If, however, the therapist is working more systemically, the message may contain a task, a timetable, or a paradoxical injunction which is designed to alter both the behaviour of the couple and the way they conceptualize their relationship. In some cases, it is appropriate to use a 'split-team' message , in which one part of the team is said to favour a more behavioural task while the other part believes that that will be impossible to achieve and therefore prefers to 'prescribe the symptom'.

The final part of the message is again likely to reiterate the positive sentiments of the introduction. There are good therapeutic reasons for this, in that people tend to remember the positive things that they hear about themselves, and may then link these to the more specific tasks or injunctions that are given with them. In many cases we also send a written copy of the message to the couple, so that they can think it over between sessions and not forget what has been discussed.

Where therapists are working alone, they will be unable to have the 15 minute break in the session, but it is always useful to spend a little time thinking about the message to be given at the end of the session, and to think what other team members might have said or suggested for a final message.

Specific techniques in behavioural–systems couple therapy

(a) Reciprocity negotiation

Within the alternative levels of intervention hierarchy, reciprocity negotiation is at the lowest level, relating closely to the goals that the couple themselves has set, and depending on a fairly co-operative attitude on both sides. The partners state their complaints in everyday terms, and the task of the therapist is then to help them to achieve a compromise by each doing what the other partner wants in a reciprocal way.

Reciprocity negotiation is partly based on operant conditioning and partly on the social exchange theory of Thibault and Kelley.[24] The assumption is that satisfaction in marriage and other intimate relationships is based on a relatively equal and high level of input by each partner of positive (i.e. rewarding) behaviour and a relatively low input of negative or unacceptable behaviour. Problematic marriages have a low level of these mutually rewarding behaviours on both sides, or may have a gross imbalance in the input from the two different partners. Instead of exchanging positive behaviour, the partners may use coercive methods to try and force the other to stop doing those things of which they disapprove.

The remedy proposed by behavioural marital therapy is that each partner should state their **complaints**, but that these complaints should then be translated into **wishes** for an alternative way of behaving which is more acceptable, and, as a second stage, into **tasks**. It is very useful to concentrate on practical, domestic issues for these tasks, as these are easily grasped, frequently repeated, and more likely to be remembered than more abstract tasks. In principle the tasks for each partner should be linked and reciprocal, but if this is not possible a 'bank account' approach can be used in which each partner builds up a fund of good behaviour and they work out at the end of a period of time whether it has been mutually acceptable. In moving from complaints to tasks one also moves from past to future, and this is one of the most characteristic features of reciprocity negotiation. The therapist is thus more interested in what will happen next week than in what happened last week or last year.

The way that reciprocity negotiation is used in behavioural–systems couple therapy is a little different from its use in behavioural marital therapy. We will usually have the couple in a decentred position while negotiating, and feel that this helps the process both to be effective, and to translate more successfully to their home setting. We also use it quite briefly at different stages of therapy, rather than as the mainstay of therapy throughout.

The tasks developed for each partner in reciprocity negotiation should be:

1 specific,

2 positive,

3 repeatable,

4 practicable, and

5 acceptable to both partners.

They should also be concerned with everyday activities, rather than once-only events such as arranging an overseas holiday. Sometimes sexual problems can be brought in to the negotiation.

Reciprocity negotiation is a well-tried and effective method of couple therapy in those who accept that they have marital problems. It is also an advantage that the therapist here works in a way which is straightforward and takes an adult-to-adult approach. But it is also, in our setting, a way of assessing whether the couple are ready for this sort of intervention; if not, they can be offered a more systemic input until they are more ready to negotiate.

(b) Communication training

The second strategy in the alternative levels of intervention hierarchy is training in communication. This too is part of the behavioural marital therapy spectrum, but not so exclusively, because work on communication is part of most types of couple therapy. The characteristic feature of the form of communication training used in our setting, however, is that it aims for efficient and clear communication, with positive and constructive requests rather than complaints. Other forms of communication training[25] emphasize other skills such as empathy, reflective listening, and supportive comments. In the present form of communication training these are also issues to be considered, but the main emphasis is on issues such as reducing misunderstandings, ensuring that both partners have an equal say, and helping them both to speak from the 'I' position.

Problems encountered in couple therapy amenable to communication training include:

- lack of empathy
- inability to express emotion
- failing to listen
- monologues with no break for feedback
- one partner the spokesperson and the other silent
- mind-reading (i.e. A knowing better than B what is in B's mind)

- sting in the tail (a positive comment followed by a criticism)
- wandering off the topic
- continual criticism.

In carrying out communication training, the therapist first decentres him- or herself, and asks the couple to converse about a relevant topic. When a problem of communication arises the therapist acts as a 'director' and asks them to discuss the topic in another way. If the problem observed is one of lack of empathy, this may include asking one partner to attend to the emotional state of the other, and perhaps to feed back his or her understanding. If it is of inability to express emotion, the therapist may try to intensify the interaction, pointing out the way in which they are holding back their emotions, and encouraging more expressiveness.

The next three problems are connected: failing to listen, talking in monologues, and the 'spokesperson' problem. Remedies can be decentring, encouraging each partner to speak for him- or herself, stopping the talkative partner (perhaps by asking them to listen to what the other partner has to say), and cutting any monologues short by asking for feedback from the other partner. In dealing with mind-reading one may have to be quite diplomatic, because the process is rather similar to psychotherapeutic interpretation, and some partners may feel that this is a legitimate way of giving insight; however, it should be tactfully blocked, usually by asking the partner whose mind is being 'read' to say whether that is what he or she really thinks.

The 'sting in the tail' is dealt with usually by simply pointing it out, but in some cases it can be neutralized by asking the speaker to restate the idea the opposite way round with the 'sting' first. An example of this is given by a man who said 'I realize you were hurt by what I did, but I had no intention to harm you (i.e. you are being oversensitive)'. He was asked to rephrase it as 'I had no intention to harm you, but I realize that you must have been hurt', and his wife found this much more acceptable, because she could respond to the more positive part of the comment.

The problems of wandering off the topic and continuous criticism are often rather intractable. One way, however, of keeping them to task is to bring them back frequently to the problem first presented, and ask whether they can concentrate on solving it. In the case of mutual criticism, one way of coping is to slow down the interaction so that each partner speaks only after the therapist has intervened to reframe what has just been said.

As with reciprocity negotiation, communication training is used not as a self-contained therapy in itself, but rather as part of a menu of techniques to be chosen according to the problem presented or observed at the time.

(c) Structural moves in session

The main interventions under this heading are raising arguments (or heated discussions) in the session, reversed role play, and 'sculpting'.

There are many couples in which there is a reluctance to enter any sort of conflict. They avoid differences of opinion, and pretend that there is agreement on almost every issue. The more dominant partner, usually more at ease verbally, effortlessly takes the spokesperson role. The other partner is either silent much of the time or spends much effort placating the other in order to reduce conflict.

One strategy with such couples is to ask them to argue (or, to put it more acceptably, to have a **heated discussion**) about a fairly trivial topic. An example of this might be whether the toilet seat should be left up or down after it has been used. It must be a genuine difference of opinion, and not simply one manufactured for the purpose, but it is important that it should be of a trivial nature, as otherwise the couple may feel inhibited about discussing it.

They are then asked to discuss the issue with the therapist observing, and the therapist particularly encourages the more submissive partner to participate with enthusiasm. It may be necessary to ask him or her to speak louder, or to ask the other partner to listen more carefully to what the quiet partner has said, but the therapist should not take sides as such. What is being dealt with is not the issue itself, but the process of arguing. The outcome does not matter, except that the submissive partner should not be allowed to 'get away with' their usual tactic of giving in for the sake of peace. The couple may 'agree to differ' or the submissive partner may have a better than usual hearing, and even win the argument.

This intervention is particularly useful for those couples where there is a degree of depression in the quieter partner, or where the quieter partner is very reluctant to be involved sexually, and is blaming him- or herself.

Another intervention in session which can have an impact on the interaction is the '**reversed role play**'. Here the couple is asked to discuss a particular issue, but they are asked to act as if they were the other partner, even perhaps changing chairs for the purpose. The exercise is useful for some couples who have difficulty understanding each other's point of view, and may promote better mutual understanding.

A third intervention in session is the use of '**sculpting**', in which the partners position themselves and each other wordlessly in a kind of tableau to express some aspects of the relationship. For example, a wife who feels herself excluded from her husband's life may place him looking away from her, while her husband might place the two arm-in-arm and facing the same way. Neither position would represent the objective truth, but each would gain some understanding of the views of the other. The different views could also be the subject for discussion in session or during 'homework'. As with reversed role play, sculpting, with the accompanying 'experiential' insight, can be useful in those couples where there is little understanding of the other's point of view.

(d) Timetables and tasks

These are perhaps the most frequently used of our interventions. They are always given as part of the 'homework' at the end of the session, and may be of a behavioural nature or more systemic. Systems tasks are usually used for behaviour which is thought of as being out of control. Thus, they may be used in a couple where there is a jealous partner: this partner would be asked to raise his or her doubts about the other's fidelity, but only at a specified time each day and for a limited period (e.g. half an hour). If the topic comes up at any other time, they are asked to postpone any discussion till the appointed time. This can be frustrating for the jealous partner, although he or she will perhaps be reassured that the other will give the topic his or her full attention at the set time: but for the other partner it can come as a great relief that the issue of jealousy is at last under some sort of control, even in this simple form.

A timetabled task may be used in other situations, for example, when one partner has a series of complaints which the other is rejecting. The couple can again be asked to discuss the issue only

at certain times and for a limited duration. The advantage of a timetable under these circumstances is that the therapist does not have to adjudicate as to who is right or wrong in the content of the argument, but simply deals with the process of arguing by asking the couple to raise their legitimate complaints at home at an appropriate but limited time.

Another frequently used timetable is the 'talk' timetable, in which a couple who do not communicate very often are asked to set a time each day or evening when they can get together for a discussion about the day's events. In cases where there are difficulties with empathy, it may be useful in addition to ask each partner at the daily talk session to repeat back what the other one has said to reassure the other that they have understood what is meant.

One situation which responds particularly well to timetabling is where the male partner is very keen on sex and the female (while not having a sexual dysfunction as such) is much less enthusiastic. Here the partners are encouraged to reach a compromise on the agreed frequency at which sexual relations might occur, and then they agree on a suitable timetable. The day of the week has to be fixed in advance, since if this is not done the usual arguments will ensue as to whether sex should take place that night, and they are also asked to make the chosen night something special, with perhaps a dinner and the telephone disconnected. If, however, the enthusiastic partner suggests sex on another night, the other can simply remind him that it has been arranged and that they should stick to the arrangement. This remedy may seem somewhat crude, and it is often simply a temporary measure. However, it can be said to have virtually saved some relationships, because it takes the heat out of the sexual conflict which could otherwise lead to divorce, and its use can open up the discussions in subsequent sessions to include non-sexual topics which would otherwise be pushed out by the sexual issue.

It should be mentioned here that many relationship problems have a sexual and a general dimension. When the sexual difficulty is motivational rather than dysfunctional, it is often most productive to deal with it in couple relationship therapy either alone or in combination with psychosexual therapy. In such a case it is quite appropriate to suggest techniques such as the Masters and Johnson 'sensate focusing'[26] in addition to the couple therapy approaches already mentioned.

(e) Paradoxical interventions

Paradox is a relatively infrequently used option in couple therapy (Fig. 6.3.7.1), and is brought in when other methods are ineffective or where the couple relationship seems so rigid that no other intervention can be used. The rationale for paradox depends on a systemic hypothesis which states that the homeostatic forces in the system may be so strong that no straightforward intervention will alter it. All systems tend towards a resistance to change, but in some the resistance is maintained by powerful forces which themselves seem to be informed by extreme anxiety.[14] In these couples or families the only intervention likely to succeed in changing the system is one which prohibits change, but for unacceptable reasons. Although the above explanation is somewhat unsatisfactory, paradox remains in practice a technique which can unlock an otherwise stuck relationship and get the couple back on course for continued therapy.

Paradox is always applied with care and in a sympathetic manner. A common form is to 'prescribe the symptom', that is to advise the couple that it is best 'for the time being' to persist with both the behaviour complained of and the reciprocal behaviour in the other partner. The reason given for this conclusion is a plausible, but challenging and perhaps unacceptable, explanation based on systemic understanding of the relationship.

In using paradox the therapist should think in four stages. First, there should be a positive connotation of the 'symptom' and the reciprocal behaviour. Secondly, there should be a rehearsal of why they are at present helpful for the couple. Thirdly, a statement should be made of the hypothesized feared consequences if the behaviours were to stop. Fourthly, the symptom and the reciprocal behaviour should be prescribed. A case example may make this process a little clearer.

Case Study: A couple who presented with depression in the wife (Edna) and a rather overprotective attitude on the husband, George's, side were in therapy for some weeks without much progress. Following a session in which the therapist asked many questions of both of them about the circumstances and consequences of the depressive episodes, the paradox was presented as follows. 'This depression seems in some ways to be quite good for you as a couple, because it enables Edna to help George by giving him a role in life as her protector. If the depression were to disappear it might be difficult for you both to continue your peaceful relationship, because the differences between your views and ideals would become very clear and you might argue all the time. So for the time being it is better for Edna to remain depressed and for George to be her spokesman and protector'.

This intervention led to quite an outburst from the wife, who up to that time had always been very quiet, and she began to talk of some of the differences of opinion that they had actually had. The husband looked rather disconcerted, and questioned the therapist's reasoning. In the next two sessions the couple reversed their imbalance to some extent, the wife became more assertive than the husband, and her depression became less severe.

The paradox can thus be a powerful mechanism for change, but it must be used with some caution, since an instruction given paradoxically may be taken literally. So it would be inappropriate to include in a paradox any instructions to break the law, to harm oneself or others, or to act irresponsibly. If given as recommended, however, the paradox can unlock a 'stuck' system and put the couple on the road to change and improvement.

In a team setting it is probably best to use a **'split team'** message rather than a paradox as such. This presents the paradox as above, but in the form of an alternative, for example in the form of a disagreement between the therapist and the supervising team. 'I feel that you can carry on with the tasks that I have been giving you, but my team think I am being naïve, and that you really need the depressive symptoms and the overprotection to keep your marriage from falling apart'. The effect is similar, but the impact is softened somewhat by this technique.

Couple therapy with couples from other cultures

As mentioned above, most western countries are now multicultural, and a significant minority of those seeking therapy, especially in urban centres, are from immigrant backgrounds. Probably the most frequently seen in Britain are those of South Asian origin, but Eastern European, African and African- Caribbean

couples are also seen quite often. In the USA there are many people of Latin American and Oriental backgrounds, and here again these will be more commonly encountered in urban settings. The author's experience is mainly with South Asians, and the examples will be mainly from this group.

Cultural factors in counselling have been highlighted by d'Ardenne and Mahtani,[27] and include the need for awareness in the counsellor that he or she also comes from a specific culture which may be just as difficult for the client to comprehend as the client's is to the counsellor. They emphasize the need for humility in the face of difference, and the responsibility of the counsellor to check with the clients before making assumptions about their lifestyles and beliefs. Their advice on the use of interpreters is that unofficial interpreters, including members of the clients' own families, should be discouraged because they are inclined to act as therapists themselves, may translate inaccurately, may ignore cultural differences and may even exploit the clients. It is better to use official interpreters, though this may become expensive to the treatment unit, and the interpreters themselves may translate inaccurately in accord with what they think the therapist wants to hear. Ideally a unit would have multilingual counsellors, but this is not always practicable. In practice, it is often possible with a modicum of understanding of the language for a couple relationship therapist to carry out therapy in English without an interpreter, but with a little bit of necessary translation by the partner whose English is better.

Little has been written specifically on cultural factors in relationship therapy, but Ahmed and Bhugra[28] have reviewed the role of culture in sexual dysfunctions, and their observations are also relevant to couple therapy. They emphasize the need to adapt the techniques of 'western' sex therapy to accommodate the cultural backgrounds of the patients, in particular the gender roles in the culture, and they highlight the risks of ignoring these in therapy.

Couple therapists need to be flexible in regard to the aims of therapy in these couples, which may be rather different from the typical white British couple, and there may also be limits to the kind of therapeutic change possible. For example, a couple from South Asia may be orientated towards a male dominated marital pattern, and both partners may be reluctant to accept the kind of equal relationship that typical couple therapy would expect. In other cases, the more traditional male partner may be concerned to retain the traditional dominance, while the (more westernized) woman may be demanding equality. Similar problems may arise in couples who come from different cultures, and there are clearly more interracial relationships developing as the cultures become more integrated.

Religious considerations may bring difficulties to therapy. Strict Muslim couples may be reluctant to attend therapy together, particularly with a male therapist, because of the difficulty of the wife talking to a male stranger. Masters and Johnson[26] found that one of the most reliable prognostic factors in their therapy was the negative effect of any strong religious belief on the outcome. It has also been observed that in a sexual dysfunction clinic Asian couples were more likely than white couples to default from therapy, a finding put down to their pursuit of organic explanations for the problems and educational and language barriers.[29]

Similarly African-Caribbean couples will have different aims and limitations from white British couples. The father in these families may traditionally be more of an absentee, and leave the upbringing of the children to his partner, who becomes the main authority figure in the family. Again, the therapist must remain aware of possible differences from his/her own culture, and remember that the key consideration is the wishes and wellbeing of the couple rather than any imposed set of rules derived from theory. In particular it is usually impossible to persuade the man to take a more active role in child care, even when the woman wants this, and their relationship will usually remain 'semi-detached'.

One particular 'problem constellation' which the author has seen many times is that of an Englishman married to a North American woman. This should theoretically present no problems, as the language and cultures of both are similar. However, they are divided by the tendency for the British man to be reserved and sometimes resentful and the American woman to be outspoken and critical. Such difficulties have also been found with couples from other disparate backgrounds, such as North American and Latin American partners, and Southern European and British couples. These differences in outlook can lead to repetitive quarrels, in which neither partner can understand where the other is coming from, and often there is also a lack of sexual contact between them. In therapy it is usually necessary to explain each partner to the other, using positive terms and helping them to appreciate the cultural differences without condemning the partner. Then they can usually cooperate with reciprocity negotiation and communication training.

Although the examples given are mainly of South Asian and Western couples, the principles for dealing with cross-cultural relationships of all sorts are basically similar. The therapist needs to respect the differences between their culture and his/her own, trying not to impose solutions which are alien to the couple's own culture. In those cases where they each come from a different background, the general approach is to try to build bridges between them, and use the techniques of behavioural systems therapy to solve their difficulties.

Efficacy of couple therapy

The efficacy of couple therapy is not an easy topic to discuss. Problems arise as to how one should assess efficacy, and while most authors would agree that a measurable improvement in marital adjustment is a valid measure of improvement, some authors dismiss that as being too subjective or too superficial. On the other hand, to use an objective criterion such as divorce as an outcome variable might be seen as being too strict on the therapy, since divorce happens for many reasons, and it might not actually be a bad outcome in some relationships.

A review of efficacy in couple therapy has been carried out by Baucom *et al.*[30] They did a very thorough search of the literature, and made some far-reaching and challenging observations. They comment that the untreated improvement rate is very low in couple problems, and that many of the non-behavioural approaches are of unproven efficacy. They conclude however that behavioural marital therapy (comparing mean effect size over a series of 17 independent controlled outcome studies) is an efficacious and specific intervention for marital distress. The improvement is likely to last for up to a year after treatment but there is less certainty over longer follow-up periods. The addition of cognitive restructuring to behavioural couple therapy did not add anything to the efficacy, but the numbers were rather small.

Snyder and Wills[31] evaluated the outcomes of behavioural versus insight-orientated marital therapy, and found that there was equal improvement in the two conditions, with both being superior to waiting list controls. There was, however, a difference at follow-up, with more of those who had had behavioural therapy divorcing than those who had had insight-orientated therapy.

In a controlled study of behavioural versus interpretative couple therapy, Crowe[32] found that both approaches were effective, but that the behavioural approach produced results more quickly. The follow-up at 18 months showed both methods to be of lasting efficacy, with no differences between them at that point.

Another treatment approach, emotion-focused therapy,[33] has produced good results with couples in therapy. As with behavioural couple therapy the couples improved significantly more than those on a waiting list, but this therapy seems less effective in couples with higher levels of distress.

In one of the relatively few studies on the efficacy of systems-orientated couple therapy. Emmelkamp *et al.*[34] evaluated the effects of behavioural versus systems couple therapy, and concluded that the two approaches had very similar results, but both did better than waiting-list controls.

The London Depression Study[20] (see above) found that systemic couple therapy produced good improvement not only in terms of the couple satisfaction but also on the depression in the depressed partner. The therapy was also associated with a lower drop-out rate than the antidepressant condition, and was thus more acceptable to the patients and their partners.

Thus, the two components of behavioural–systems couple therapy have both been validated by outcome research, although at a much higher level for the behavioural than the systemic. It would be desirable to carry out research on the combined therapy, but this has not yet been done, and the best that can be said is that it is a combination of two probably effective treatment approaches, and therefore likely to be effective.

Training

Training for work with couples using a behavioural–systems approach has been thoroughly reviewed by Crowe and Ridley.[1] It requires an ability in the therapist to understand and use different approaches, and an ability to adapt ones activity to the needs of the couple. Before beginning to work as a couple therapist the trainee should have a basic understanding of the dynamics of couple and family interaction, the phases of human development, the impact on the individual of life events, sexual function and interaction and the impact of physical illness on couple and family relationships. These can be dealt with in the traditional seminar format, in which the trainee can also learn about theoretical and technical aspects of the approach to couple therapy itself. It is also important in selecting candidates for training to ensure that they have some experience of counselling individuals or of being in a therapeutic role, for example as a nurse or doctor.

In addition to the seminars there are also more active training sessions in which the trainee is given the technical skills to carry out therapy. The latter take three main forms: role play, observation, and supervized practice. In role play the trainees are encouraged to use either an existing couple or a fictional one and role play a couple therapy session. Ideally they should each, in different exercises, have the opportunity to play the husband, wife, therapist, and observer in therapy.[35] This helps both in the development of technical skills and in learning to be empathetic to clients through having experienced the client role. In role play the trainee can practise any of the techniques required in therapy, but it is perhaps especially useful in the area of communication training, in which the therapist needs to be alert to the problems shown by the partners and able to apply the appropriate technique smoothly and effectively.

In observation and supervision other aspects of the therapy can be taught, especially the more systemic methods such as arguments, reversed role play, and the framing of messages. The trainees move quite quickly from live observation of therapy in the clinic to being firstly a co-therapist and then the sole therapist in the session, supported and supervised by the trainer and the observation team on the other side of the screen. It is in this activity that trainees begin to display their skills or deficits as therapists, and the trainer must be able to assess progress at this stage and take remedial action if a trainee does not seem to be working as well as expected.

Conclusions

The field of couple therapy is a wide and varied one, and there are almost as many different approaches to treatment as in individual psychotherapy. The relatively brief therapeutic method presented here, behavioural systems couple therapy, is an eclectic one, taking techniques from two approaches of proven efficacy and combining them into a flexible and versatile therapy capable of being used in a wide variety of presenting problems. These include simple relationship problems, psychosexual problems, and such psychiatric conditions as anxiety, depression, and morbid jealousy. It is relatively easy to teach, and although it has not yet been subjected to controlled trials it can be assumed to be no less effective than its component therapies which are both effective. It has recently also been recommended in a package for self-help[36] with homework exercises and theoretical explanations to be used without the intervention of a therapist. There are few contraindications for the therapy, and it can be used both as a therapy in its own right or as an adjunctive therapy in, for example, the treatment of depression, psychosis or sexual dysfunctions. It can thus be a useful addition to the various methods available for the reduction of distress, whether in couples or individuals.

Further information

Crowe, M. and Ridley, J. (2000). *Therapy with Couples: a behaviouralsystems approach to marital and sexual problems* (2nd edn.). Blackwells Science, Oxford.

Clulow, C. (ed.) (2001). *Adult Attachment and Couple Therapy*. Brunner Routledge, London.

Organizations which provide information about how to obtain couple therapy

British Association for Sexual and Relationship Therapy www.basrt.org.uk
British Association for Counselling and Psychotherapy www.bacp.co.uk
United Kingdom Council for Psychotherapy www.psychotherapy.org.uk
Institute of Family Therapy www.instituteoffamilytherapy.org.uk
American Association for Marriage and Family Therapy www.aamft.org

Organizations which provide information on training in Sexual and Relationship Therapy

British Association for Sexual and Relationship Therapy www.basrt.org.uk

Porterbrook Clinic, Sheffield (Sheffield Hallam University) www.porterbrookclinic.org.uk
London South Bank University ww.lsbu.ac.uk/psychology
University of Central Lancashire: Lancashire School of Health and Postgraduate Medicine www.uclan.ac.uk
Relate Institute, Doncaster www.relate.org.uk
American Association for Marriage and Family Therapy www.aamft.org

References

1. Crowe, M. and Ridley, J. (2000). *Therapy with couples: a behavioural–systems approach to marital and sexual problems* (2nd edn). Blackwell Science, Oxford.
2. National Statistics Office (1998). *Population trends.* HMSO, London.
3. Robinson, M. (1991). *Family transformation through divorce and remarriage*. Tavistock Press, London.
4. Daniell, D. (1985). Marital therapy: the psychodynamic approach. In Marital therapy in Britain, Vol. 1 (ed. W. Dryden), pp. 169–94. Harper and Row, London.
5. Clulow, C. (ed.) (2001). *Adult Attachment and Couple Psychotherapy*. Brunner Routledge, London.
6. Bowlby, J. (1969). Attachment and loss, Vols 1 and 2. Hogarth Press, London.
7. Wile, D.B. (1993). *Couples therapy, a nontraditional approach* (2nd edn). Wiley, Chichester.
8. Stuart, R.B. (1980). *Helping couples change*. Guilford Press, New York.
9. Liberman, R.P. (1970). Behavioural approaches in family and couple therapy. *American Journal of Orthopsychiatry*, **40**, 106–18.
10. Beck, A. (1988). *Love is never enough.* Harper and Row, New York.
11. Dryden, W. (1985). Marital therapy, a rational emotive approach. In Marital therapy in Britain, Vol. 1 (ed. W. Dryden), pp. 195–221. Harper and Row, London.
12. Minuchin, S. (1974). *Families and family therapy*. Tavistock Press, London.
13. Haley, J. (1980). *Leaving home.* McGraw Hill, New York.
14. Selvini Palazzoli, M., Boscolo, L., Cecchin, G., *et al.* (1978). *Paradox and counter–paradox.* Aronson, New York.
15. Segraves, R.T. (1982). *Marital therapy: a combined psychodynamic–behavioural approach.* Plenum Medical, New York.
16. Weeks, G.R. (1989). *Treating couples, the intersystem model of the Marriage Council of Philadelphia.* Brunner–Mazel, New York.
17. Spinks, S.H. and Birchler, G.R. (1982). Behavioural–systems marita l therapy: dealing with resistance. *Family Process*, **21**, 169–85.
18. Berg–Cross, L. (1997). *Couples therapy*. Sage, Thousand Oaks, CA.
19. Jacobson, N.S. and Margolin, G. (1979). *Marital therapy: strategies based on social learning and behavioral exchange principles.* Brunner–Mazel, New York.
20. Leff, J., Vearnalls, S., Brewin, C.R., *et al.* (2000) The London depression intervention trial. Randomized controlled trial of antidepressants v. couple therapy in the treatment and maintenance of people with depression living with a partner: clinical outcome and costs. *British Journal of Psychiatry*, **177,** 95–100.
21. De Silva, P. (1997) Jealousy in couple relationships: nature, assessment and therapy. *Behaviour Research and Therapy*, **35,** 937–85.
22. McFarlane, W. (2000) Psychoeducational multi-family groups. Adaptations and outcomes. In *Psychosis: Psychological Approaches and their Effectiveness.* (ed. B. Martindale, A. Bateman, M. Crowe and F. Margison) Gaskell, London.
23. Crowe, M. (2005) Couples and mental illness. *Sexual and Relationship Therapy*, **19**, 309–10.
24. Thibault, J.W. and Kelley, H.H. (1959). *The social psychology of groups.* Wiley, New York.
25. Olson, D.H., McCubbin, H.I., Barnes, H., *et al.* (1983). *Families: what makes them work.* Sage, Los Angeles, CA.
26. Masters, W.M. and Johnson, V.E. (1970). *Human sexual inadequacy.* Little, Brown, Boston, MA.
27. d'Ardenne, P. and Mahtani, A. (1989). *Transcultural Counselling in Action*. Sage, London.
28. Ahmed, K. and Bhugra, D. (2007). The role of culture in sexual dysfunction. In *Psychiatry: Sexual Disorders and Psychosexual Therapy*. (ed M.Crowe). Medicine Publishing (Elsevier), London.
29. Bhui, K. (1998). Psychosexual care in a multi-ethnic society. *Journal of Social Medicine*, **91,** 141–3.
30. Baucom, D.H., Shoham, V., Mueser, K.T., *et al.* (1998). Empirically supported couple and family interventions for marital distress and adult mental health problems. *Journal of Consulting and Clinical Psychology*, **66**, 53–88.
31. Snyder, D.K. and Wills, R.M. (1989). Behavioural versus insight orientated marital therapy: effects on individual and interpersonal functioning. *Journal of Consulting and Clinical Psychology*, **57**, 39–46.
32. Crowe, M.J. (1978). Conjoint marital therapy: a controlled outcome study. *Psychological Medicine*, **8**, 623–36.
33. Johnson, S.M. and Greenberg, L.S. (1985). Differential effects of experiential and problem–solving interventions in resolving marital conflict. *Journal of Consulting and Clinical Psychology*, **53**, 175–84.
34. Emmelkamp, P.M.G., van der Helm, M., MacGillavry, D., *et al.* (1984). Marital therapy with clinically distressed couples: a comparative evaluation of system–theoretic, contingency contracting and communication skill approaches. In *Marital interaction: analysis and modification* (ed. K. Hahlweg and N. Jacobson), pp. 36–52.Guilford Press, New York.
35. van Ments, M. (1983). *The effective use of role–play. A handbook for teachers and trainers.* Kogan Page, London.
36. Crowe, M. (2005) *Overcoming Relationship Problems.* Constable Robinson, London.

6.3.8 Family therapy in the adult psychiatric setting

Sidney Bloch and Edwin Harari

The term 'family therapy' covers a range of approaches. At one extreme, it is a method which seeks to help an individual patient. At the other extreme, the focus is on the relationships between people; according to this view psychopathology reflects recurring, problematic interactive patterns among family members. Midway between the two positions is one that views the family as acting potentially either as a resource or a liability for an identified patient. In this chapter, we cover the spectrum but confine ourselves to the adult psychiatric setting.

A historical and theoretical context

The family has long been recognized as a core aspect of social organization. The folklore of all cultures emphasize the family's role to mould the character of its members. In the past 150 years academic disciplines, such as anthropology and sociology, have studied the various forms of family structure found in different cultures, and at different times. Since the 1960s, psychiatry has also developed a clinical and research interest in the family beyond that of genetics.

Scattered through Freud's writings are interesting comments about marital and family relationships and their possible roles in

both individual normal and abnormal development.[1] His description of unconscious processes like introjection, projection, and identification illuminate how individual experiences may be transmitted across generations. In 1921, J.C. Flugel published the first comprehensive psychoanalytic account of family relationships.[2] Influenced by Anna Freud, Melanie Klein, and Donald Winnicott, the child guidance movement in Britain, mainly consisting of social workers, devised a model of one therapist working with the disturbed child and another with the mother. The two clinicians then collaborated in order to appreciate how the mother's anxieties distorted her perception and handling of her child, leading to developmental difficulties.

Proliferation of theoretical schools

Psychoanalytic and related approaches

Things took a different turn in the United States where Nathan Ackerman[3] began in the 1950s to treat families with a disturbed child, using psychodynamic principles. An interest in working with two or more generations arose concurrently with 'transgenerational'-oriented family analysts using object–relations concepts. Thus, Murray Bowen[4] noted that the capacity of psychotic children to differentiate from their families, while still retaining a sense of age-appropriate belonging, was impaired by the effects of unresolved losses and other trauma in parental and grandparental generations. He also devised the genogram, a schematic depiction of family structure, with a notation for notable events; this remains a standard part of family assessment (see below).

Boszormenyi-Nagy and Spark[5] similarly addressed the transgenerational theme, describing how relationships were organized around a ledger of entitlements and obligations, which conferred on each family member a sense of justice or injustice about their situation. This, in turn, reflected childhood experiences of neglect or sacrifices made on another relative's behalf for which redress was sought in adult life.

Systems-oriented (see later)

Bowen[4] also introduced the principles of 'systems theory' into family therapy. A system is defined as a set of interrelated elements that function as a unity within a particular environment and where the whole is larger than the sum of the parts. 'General systems theory', propounded in the 1940s by a German biologist,[6] contains among its key concepts the place of hierarchy and the emergence of new features in the system as it transforms itself, necessarily, from one level of organization to another. A family is an example of a partially open system that interacts with both its biological and socio-cultural environments and changes over time to accommodate developments such as the advent of a first child or the death of a grandparent.

Working with delinquent youth, Salvador Minuchin recognized the relevance of systems thinking. The youngsters often came from poor, emotionally deprived families, headed by a demoralized single parent (usually the mother) who alternated between excessive discipline and helpless delegation of responsibilities to a child or to her own critical mother. Since these families were beyond the reach of conventional 'talking' therapies, Minuchin applied action-oriented techniques which enabled him to 'join' the family and to re-establish an adaptive hierarchy and effective boundaries between subsystems (marital, parent–child, siblings).

Later, treating 'psychosomatic families' where the problem was a child or adolescent suffering from anorexia nervosa, unstable diabetes or asthma, Minuchin and his colleagues noted that these families, while intact and articulate, were often enmeshed. Members avoided challenging the apparent sense of family unity. Typically, marital conflict was detoured through the symptomatic child, resulting in maladaptive coalitions between parent and child (sometimes between grandparent and child) and the involvement of third parties (e.g. helping agencies) in family life; loss of hierarchy and boundaries ensued. Because words were used to avoid change in these well-educated families, non-verbal strategies were devised to face unspoken fears of conflict and change.[7]

Jay Haley's 'strategic therapy'[8] combined features of Minuchin's model with ideas of Milton Erickson whose techniques had skilfully exploited the notion that a covert message lurks behind explicit communication, which defines the power relationship between family members. Related theoretical developments took place in Palo Alto, California in the 1950s, where a group of clinicians, together with the anthropologist Gregory Bateson,[9] observed that implicit in communication were tacit, non-verbal 'meta-communications' which defined the ties between participants. A contradictory quality between these two levels of communication—in which messages carried persuasive, moral, or coercive force for the recipient—formed part of what they called a 'double-bind'; this form of entrapment was proposed, albeit erroneously, as a possible basis for the formal thought disorder found in schizophrenia.[10,11]

(a) Systems-oriented models: further developments

All the above system-oriented views assume that family functioning can be objectively studied. However, therapists are not value-free and may actively orchestrate changes in accordance with their preferred theoretical model; neglected in these circumstances are therapists' biases and their influence.

This tendency probably reflected the determination of family therapists to distance themselves from psychoanalytical theory; but it also led them to neglect the family's past history and changes through the lifecycle, including the relevance of traumatic events.

In response to this criticism there was a shift away from a problem-focused approach, which had typified most communication-based views of psychopathology. The so-called Milan school[12] (see course of therapy below), whose founders were psychoanalysts, launched profound conceptual changes in how to approach the family, particularly in interviewing them. Another innovation was the participation of observers behind a one-way screen whose task was to offer hypotheses about the family-plus-therapist system to the protagonists.

A Norwegian group[13] took the idea one step further by developing the 'reflecting team dialogue'. Here, following a session, the family could observe the therapeutic team discussing their problems and possible causes, and what factors might have prompted them to seek certain remedies—especially those they had persevered with despite the clear lack of effectiveness.

(b) Post-modern developments

Family therapists also began to ask whether families might be hampered from trying out new ways to solve their difficulties because of the ways they themselves had interpreted their past experiences or unwittingly absorbed the explanatory narratives of external 'experts' or society at large.

This led to a shift from considering the family as a system defined by its organizational structure to a linguistic-based one. According to this view the narrative a family relates about themselves is a means to integrate in specific ways their past experience and its significance. Other 'stories' are excluded from consideration. For instance, when a family with an ill member talk to health professionals, the conversations inevitably revolve around problems (a problem-saturated description). The family ignore times when problems were absent or minimal, or when they were confined to manageable proportions. A different story might be told if they were to examine the factors that could have led, or still lead, to better outcomes than those currently deemed pathological.

Several narrative-based approaches apply these concepts.[14–16] Philosophically, they align themselves with post-modernism, a movement which challenges the idea that there is a fundamental truth or grand theory known only by the expert.

(c) Criticism of systems approaches

Many criticisms have been levelled at systems-based approaches, these include:

- disregard of the subjective experiences of family members
- neglect of the family's history
- inattention to unconscious motives in interpersonal behaviour
- not addressing the issue of unequal power in a family, particularly violence against women and child abuse, and
- ignoring various forms of injustice based on societal attitudes regarding gender, ethnicity, and class.

This critique has led to integrating systems-oriented and psychoanalytic concepts, particularly those derived from object–relations theory.[17–20] Specific disorders such as schizophrenia[21] and anorexia nervosa[22] have been targeted. Another noteworthy variant of integration is Byng-Hall's[23] synthesis of attachment theory, systems-thinking, and a narrative approach.

Another criticism of systems-oriented approaches is minimizing the impact of material reality, such as physical handicap, or biological factors, in the causation of mental illness, as well as sociopolitical phenomena like unemployment, racism, and poverty. These are obviously not merely the result of social constructions or linguistic games and the distress they may inflict on people are potentially considerable.

The 'psycho-educational' approach and 'family crisis intervention' have arisen in the context of the burden that severe mental illness, particularly schizophrenia, places on the family and the potential for members to influence dramatically the course of the condition. This has led to a series of family interventions:

- educating the family about the nature, cause, course, and treatment of schizophrenia
- providing the family with opportunities to discuss their difficulties in caring for the patient, and to devise pertinent strategies
- clarifying the role of conflict, not only about the illness but also about other relational issues
- regularly evaluating the impact of the illness on the family, both individually and collectively
- helping to resolve other conflicts possibly aggravated by the demands of caring for a enduringly ill person.

This type of work may be done with a single family or with several families meeting together, known as Multiple-Family Group Treatment (MFGT). The latter has emerged as a powerful adjunct to conventional individual-based treatment of schizophrenia, bipolar disorder, major depression, obsessive-compulsive disorder, somatization disorder, and an array of chronic medical conditions. Good results have been achieved in reducing the relapse rate, duration, and frequency of hospitalization and in boosting compliance with medication.[24] Family crisis intervention, initially devised for families with a schizophrenic relative but since applied to other clinical states, operates on the premise that deterioration or a request by the family to hospitalize a member may reflect change in a previously stable pattern of family functioning. Convening an emergency meeting with the patient, spouse, and other key family members may help to avoid admission. Social and institutional forces outside the family often contribute to a crisis, and may precipitate a psychotic episode in a vulnerable member. The 'open dialogue' model of family crisis interviewing, developed in Finland, fosters discussion about such forces, using concepts and techniques derived from, *inter alia*, the Milan school, narrative approaches, and psychodynamic thinking; this integrated perspective has much potential.[25]

Indications

A measure of controversy has dogged the issue of what constitutes the indications for family therapy. Pioneering practitioners claimed, somewhat overzealously, that their methods were suited to most conditions. A more balanced view since the mid-1990s encompasses a consensus that considering the systemic context is advantageous in assessing and treating any psychiatric problem. However, it does not follow that family therapy is the treatment of choice (or even indicated).

Family therapy, it should be stressed, does not constitute a unitary approach, with one principal purpose. The diversity of theoretical models we have alluded to above, with their corresponding techniques, should make this obvious. Regrettably, attempts to link indications to specific models have contributed little to the field.

It has also become clear that DSM or ICD diagnoses do not serve well as a basis for determining indications for family interventions. DSM has a minuscule section, the V diagnoses, covering 'relational problems'; these are limited in scope and not elaborated upon.[26] We are only informed that the problem in relating can involve a couple, a parent, and child, siblings, or 'not otherwise specified'. ICD neglects this relational area entirely.

In mapping out indications, we need to avoid blurring family assessment and family therapy. A patient's family may be recruited in order to gain more knowledge about diagnosis and treatment. This does not necessarily lead to family therapy. Indeed, it may point to marital therapy or to long-term supportive therapy. Thus, we need to distinguish carefully between an assessment family interview and family therapy *per se*.

A typology of family psychopathology, which might allow us to differentiate one pattern of dysfunction from another and so map out corresponding interventions, remains elusive. Empirical evidence is inconclusive and clinical consensus lacking. An inherent difficulty is in selecting dimensions of family functioning central to creating a typology.[27] Communication, cohesiveness, adaptability,

boundaries between family members and subgroups, and level of conflict are a few of the contenders offered (see our own classification below).

There are no clear correlates between conventional diagnoses and family type. Efforts to establish links, such as an anorexia nervosa family[28] or a psychosomatic family[29] have not been fruitful. Similarly, investigations into the family and schizophrenia have yielded no durable results.[4,10] Clinicians and researchers have reluctantly accepted that models of effective family-based treatment for mental illness may not necessarily follow an understanding of the apparent causes of a condition in terms of observed disturbances in family relating. This complex matter is helpfully reviewed by Eisler regarding studies of the treatment of anorexia nervosa, but has implications for the entire field.[30]

What follows is our attempt to distil clinical and theoretical contributions.[31] Given the considerable overlap in clinical practice, categories are not mutually exclusive; and a family may require family therapy based on more than one indication. We should stress that family dysfunction is obvious in certain clinical situations and covert in others, often being concealed by a specific member's clinical presentation. Six categories emerge:

1 The problem manifests in explicit family terms and the therapist readily notes the family's dysfunction. For example, a marital conflict dominates, with repercussions for the children; or tension between parents and an adolescent child dislocates family life with everyone ensnared in conflict. In these situations the family is the target of intervention by dint of its clear dysfunctional pattern, and family therapy undoubtedly is the treatment of choice.

2 The family has experienced a disruptive life event which has led to its dysfunction. These events are either predictable or accidental and include, for instance, suicidal death, financial embarrassment, diagnosis of a serious physical illness, and the unexpected departure of a child from home. Any family stability that prevailed previously has been disturbed; the ensuing disequilibrium becomes associated with family dysfunction and/or the development of symptoms in one or more members. Family efforts to rectify the situation may inadvertently aggravate it.

3 Continuing, demanding circumstances in a family are of such a magnitude as to lead to ineffective adjustment. The family's resources may be stretched to the hilt; external sources of support may be scanty or unavailable. Typical situations are chronic physical illness, persistent or recurrent psychiatric illness, and the presence of a frail elderly member.

4 An identified patient may have become symptomatic in the context of a dysfunctional family; symptoms are in fact an expression of that dysfunction. Depression in a mother, an eating problem in a daughter, alcohol misuse in a father, through family assessment, are adjudged to reflect underlying family difficulties.

5 A family member is diagnosed with a conventional condition such as schizophrenia, agoraphobia, obsessive-compulsive disorder, or depression; the complications are the adverse reverberations within the family stemming from that diagnosis. For example, the son with schizophrenia taxes his parents in ways that exceed their 'problem-solving' capacity; an agoraphobic woman insists on the constant company of her husband in activities of daily living; a recurrently depressed mother comes to rely on the support of her eldest daughter. In these circumstances, members begin to respond maladaptively to the diagnosed relative, which paves the way for a deterioration of her condition, manifest as an enduring or relapsing course.

6 Thoroughly disorganized families, buffeted by many problems, are viewed as the principal target of help. This is apposite, even though, for instance, one member abuses drugs, another is prone to violence, and a third manifests antisocial behaviour. Regarding the family as the core dysfunctional unit is the rationale rather than a focus on each member's individual problems.

To reiterate, family therapy may not be the only treatment indicated. Thus, in helping a disturbed family struggling to deal with a schizophrenic member, supportive therapy and medication for the patient are usually as pertinent as any family treatment. Similarly, an indication for family therapy does not negate the possible use of another psychological approach for one or more family members. For instance, an adolescent striving to separate and individuate may benefit from individual therapy following family treatment (or in parallel with it), while his parents may require a separate programme to focus on their sexual relationship.

Contraindications

These are self-evident and therefore mentioned only briefly.

1 The family is unavailable because of geographical dispersal or death.

2 Shared motivation for change is lacking. One or more members may wish to participate, but their chances of benefiting from a family approach are likely to be less than if committing themselves to individual therapy. We need to distinguish here between poor motivation and ambivalence; in the latter, the assessor teases out factors that underlie it and may encourage the family to engage.

3 The level of family disturbance is so severe or long-standing, or both, that a family approach seems futile, according to the best possible clinical judgement. For example, a family that has fought bitterly and incessantly for years is unlikely to engage in the constructive purpose of exploring their patterns of functioning.

4 Family equilibrium is so precarious that the inevitable turbulence[32] arising from family therapy is likely to lead to decompensation of one or more members; for example, a sexually abused adult may do better in individual therapy than by confronting the abusing relative.

5 The patient is too incapacitated to withstand the demands of family therapy. Someone in the midst of a psychotic episode or buffeted by severe melancholia is too affected by the illness to engage in family work.

6 An identified patient acknowledges family factors in the evolution of his problem, but seeks the privacy of individual therapy to explore it, at least initially. For example, a university student struggling to achieve a coherent sense of identity may benefit more from her individual pursuit of self-understanding. Such an approach does not negate an attempt to understand the contribution of family factors to the problem.

Assessment

Family assessment, an extension of individual psychiatric assessment, adds a broader context to the formulation. The range and pace of the enquiry depends on the specifics of the case. Its phases are history from the patient, a provisional formulation concerning the relevance of the family, an interview with one or more members, and a revised formulation. In some cases, it is clear from the outset that the problem resides in the family group, thus rendering the phases below superfluous.

History from the patient

The most effective way to obtain a family history is by constructing a family tree. Apart from showing the structure, it allows relevant information about noteworthy life events and a range of family features to be added. Scrutiny of the tree also provides a source of issues warranting exploration and, eventually, the potential for formulating hypotheses.

Personal details such as age, date of birth and death, occupation, education, and illness are recorded for each member, as well as critical family events (for example, migration, crucial relationship changes, notable losses, and achievements), and the quality of relationships. For an excellent discussion of the family tree—its construction, interpretation, and clinical uses—see McGoldrick and Gerson.[33] (See Fig 6.3.8.1 for genogram conventions.)

Useful principles are to work from the presenting clinical problem to the broader context, from the current situation to its historical origins and evolution, from 'facts' to inferences, and from non-threatening to more sensitive themes.

Questions are best preceded by a statement such as: 'In order to understand your problems better I need to know something of your background and your current situation'. This can be enriched by questions that allude to interactive patterns: 'Who knows about the problem? How does each of them see it? Has anyone else in the family faced similar problems? Who have you found most helpful and least helpful so far? What do they think needs to be done'. Attitudes of family members can be thus explored and light shed on the clinical picture.

The presenting problem and changes in the family

Questions to understand the current context include: 'What has been happening recently in the family? Have there been any changes (e.g. births, deaths, illness, losses). Has your relationship with family members changed? Have relationships in the family altered?'

The wider family context

A broader enquiry flows logically in terms of other family members to be considered, and in the time span of the family's history. Other significant figures, which may include caregivers and professionals, should not be forgotten.

Apart from information about the extended family's structure, questions about their response to major events can be posed: for example, 'How did the family react when your grandmother died? Who took it the hardest? How did migration affect your parents?'

Relationships are explored at all levels, covering those between the patient and other members and between these other members. Conflicted ties are particularly illuminating. Understanding who takes what 'roles' is also useful: 'Who tends to take care of others? Who needs most care? Who tends to be the most sensitive to what is going on in the family?' Asking direct questions about members is informative, but a better strategy is to seek the patient's views about their beliefs and feelings and to look for differences between members: 'What worries your mother most about your problem? What worries your father most?' Several lines of enquiry may reveal differences.

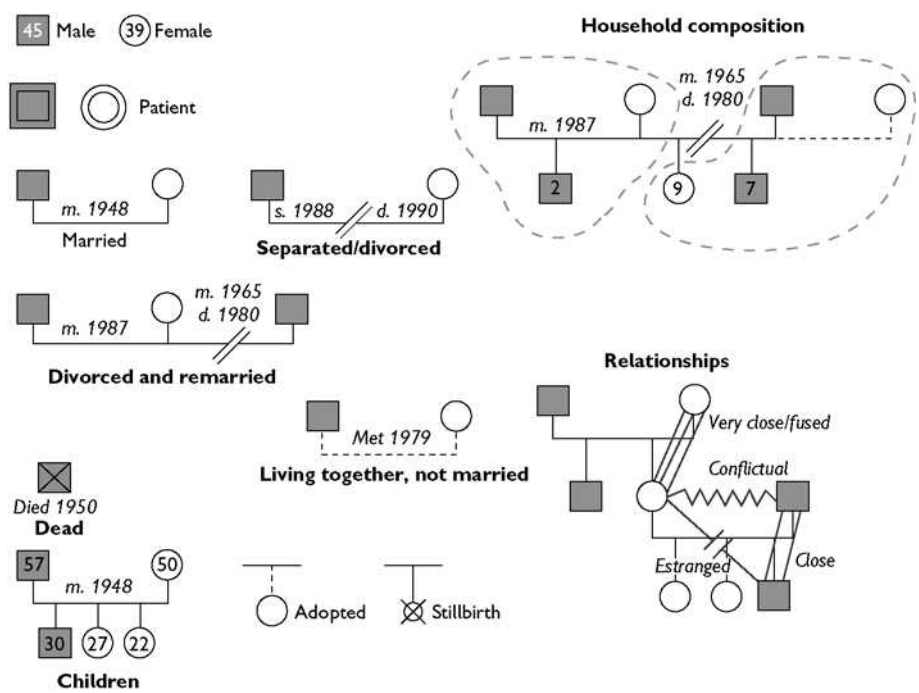


Fig. 6.3.8.1 Genogram conventions.

- Pursuing sequential interactions: 'What does your father do when you say your depressions are dreadful? How does your mother respond when your father advises you to pull up your socks? How do you react when she contradicts him?
- 'Ranking' responses: 'Everyone is worried that you may harm yourself.Who worries most? Who is most likely to do something when you talk about suicide?'
- Looking for relational changes since the problem: 'Does your husband spend more or less time with you since your difficulties began? Has he become closer or more distant from your daughter?'
- Hypothetical questions dealing with imagined situations: 'How do you think your relationship with your wife will change if you don't improve? Who would be most likely to notice that you were getting better?'

Triadic questions help to gain information about relationships which go beyond pairs; for example: 'How do you see your relationship with your mother? How does your father see that relationship? How would your mother react to what you have told me if she were here today?'

Making a provisional formulation

Two questions arise from the above interview—how does the family typically function and are there any family features relevant to the patient's problems?

(a) How does the family typically function?

A schema to organize ideas about family functioning builds from simple to complex observations: structure, changes, relationships, interaction, and the way in which the family works as a whole.

- The family tree will reveal the many family structures possible—single-parented, divorced, remarried, siblings with large age gaps, adoptees. Unusual configurations invite conjecture about inherent difficulties.
- Data will be obtained about notable family changes and events; the timing of predictable transitions is pertinent. Have external events coincided with these transitions (times at which the family may be more vulnerable)? How has the family met such changes?
- Relationships refer to how members interact with one another. What is the degree of closeness and emotional quality (e.g. warm, tense, rivalrous, hostile)? Major conflicts may be noted, as may be overly intense relationships.
- Particular interactive patterns may become apparent which go beyond pairs. Triadic relationships are more revealing about how a family functions overall. A third person is often integral to defining the relationship between another pair. A conflict, for instance, may be re-routed through the third person, preventing direct resolution. A child may act in coalition with one parent against the other or with a grandparent against a parent.
- At a higher level of abstraction, the clinician notes how the family works as a whole. Particular patterns (possibly a series of triads) may emerge that may have recurred across generations. For example, mothers and eldest sons have fused relationships, with fathers excluded, while daughters and mothers-in-law are in conflict.

Idiosyncratic shared beliefs may be discerned, explaining much of the way the family does things. 'Rules' governing members' behaviour towards one another or to the outside world may flow from these beliefs. For example, a family may hold that 'You can only trust your own family; the outside world is always dangerous'; they may therefore avoid conflict at any cost, and prohibit seeking external support.

Evidence of family difficulties may be found at each of these five levels. If they are, the question arises whether these do or do not relate to the patient's problems.

(b) Are family factors involved in the patient's problems?

Links between family functioning and the patient's problems take various forms, but the following categories cover most situations:

- the family as reactive
- the family as a resource, and
- the family in problem maintenance

Often, more than one will apply.

The family as reactive

The patient's illness, or its exacerbation, may have occurred at a time of family upheaval. While the precipitant for the upheaval may have been inherent in the illness itself, an escalating combination of the two may pertain. The illness may have occurred in the face of family stress; it pressurizes the family all the more, and this in turn exacerbates the illness.

The family as a resource

The family may be well placed to assist in treatment. This may be as straightforward as supervising medication, ensuring clinic attendance, and detecting early signs of relapse, or providing a home environment that promotes and maintains recovery. The family may also call on friends and agencies, professional or voluntary, to offer support.

The family in problem maintenance

Interactions revolving around the patient's illness may act to maintain it.

1 First, the illness becomes a way of 'solving' a family problem, the best that can be achieved. For example, anorexia nervosa in a teenager due to attend a distant university may lead to her abandoning this plan since she feels unable to care for herself. Were she to leave, parental conflict would become more exposed and her mother, with whom the patient is in coalition against her father, would find herself unsupported. The illness therefore keeps the patient at home and enmeshed in the parental relationship, and also provides a focus for shared concerns and an ostensible sense of unity.

2 Maintenance of the illness does not solve a family problem but may have done so in the past. An interactive pattern persists even though it lacks utility. In the previous example, the father's mother died 9 months later. His wife subsequently expressed feelings of closeness which he had not experienced for years; their relationship gradually improved. Both parents, however, continued to treat their daughter as incapable of achieving

autonomy, reinforcing her own uncertainty about coping independently if she were to recover.

3 Persistence of illness reflects a perception by the family of themselves and their problems, to which they are bound by the persuasive power of the narrative they have shaped for themselves. This often stems from the health care professional's explanatory schema.

Interview with key informants

The clinician will by now have made an initial assessment of the patient's problems and of the family context. The next step is an interview with one or more informants, usually family members, to corroborate the story, to fill in gaps, to determine influences impinging on the patient, and to recruit others to help. A family meeting is most effective to accomplish these goals.

Implementing the session may prove difficult since the patient may oppose it for all sorts of reasons: symptoms have been kept secret, he regards it as unfair to burden others, he is ashamed of seeing a psychiatrist, he is fearful the family will be blamed, he is suspicious of them, and so forth. These concerns need ventilating, particularly if the family context is pivotal and it is likely that treatment will be enhanced by their involvement. The patient will agree in most cases. Where the safety of the patient or others is threatened, refusal may be overridden on ethical grounds. Otherwise, refusal must be respected. A family session can be suggested again after a more trusting relationship has been established.

Who should be seen depends on the purpose of the interview; generally, all those living in the household are likely to be affected by the patient's illness. The more family factors pertain, the more desirable the attendance by all. The patient's views are sought since he will provide insight into the members he deems crucial to his 'story'.

The family interview

Much information will have been garnered by the time the family is seen. The clinician should consider any biases that may have infiltrated her thinking about the family, and how best to avoid being drawn into alliances. A non-judgemental stance is paramount.

Introductions are made in the initial phase. Names and preferred modes of address are clarified. The clinician then explains the meeting's purpose, details of which may well influence future participation. She invites everyone to share views about the nature and effects of problems they face.

The clinician has an idea about how the patient's problems relate to family function, and can test it out by asking probing questions and observing interactions. This is kept to herself since it is unhelpful for a hypothesis to be offered prematurely. Instead, details about everyday events are sought and inferences drawn later. For example, rather than focusing on 'closeness', questions can be asked about time spent together, whether intimate experiences are shared, who helps with family tasks, and so on.

Triadic relationships can be scrutinized both through questioning (what does A do when B says this to C?) and observation (what does A do when B and C reveal tensions?). The scope for such 'circular' questioning (a method ushered in by the Milan school) is enhanced if several members participate. A third person may be asked to comment on what two others convey to each other when a particular event occurs. This strategy of not posing questions to which the family may have stereotypical responses challenges them to think about their relationships in a fresh way.

Information is elicited that elaborates the family tree. Observations may be made concerning family structure and functioning; for example, who makes decisions, who controls others and in what areas, the quality of specific dyadic relationships, conflict, alliances, how clearly people communicate and how they solve problems. The discussion then extends to all spheres of family life: beliefs, traditions, rules, and values.

Throughout the interview the clinician affirms the experiences of all family members. Concerns are attended to and the members' strengths and efforts acknowledged.

The interview concludes with a summary of what has emerged. The clinician may wish to continue the assessment or recommend family therapy. If the latter, an explanation of its aim and rationale is then given.

Arrangements are made for a follow-up session, purportedly the launch of family therapy *per se*, but in essence a continuation of 'work' in progress.

Revised formulation

Since new information becomes available at each point, the initial formulation is revised as necessary.

Five observational levels —of structure, transitions, relationships, patterns of interaction, and global functioning—are re-examined in terms of the family as reactive, resourceful, or problem-maintaining. We now turn to the course of family therapy.

The course of therapy

With a family approach agreed upon, therapy begins. However, when a family is referred as a group on the premise that the problem is inherently a family one, it is made explicit that the initial stage incorporates assessment.

Given the plethora of 'schools' of family therapy it would be laborious to chart the course of treatment based on each. We shall focus on the Milan approach,[12] but stress that it has undergone refinements. Our account highlights core features but first we comment briefly on the different roles the therapist may assume.

Role of the family therapist

Beels and Ferber,[34] early observers of possible roles for family therapists, divide them into 'conductors' and 'reactors'; this differentiation remains useful since it transcends schools. Virginia Satir[35] is a good illustration of the conductor given that she espoused the notion of family therapist as a teacher who shares her expertise in how to communicate well by setting goals and the direction of treatment. She guided the family to adopt a new form of language to resolve communication problems, which she saw as the root of their troubles. Additionally, the therapist instils confidence, promotes hope for change, and makes them feel comfortable. Conductor-type therapists are explicit authorities, who intervene actively.

The therapist as reactor resonates with, and responds to, what the family exhibits to her. Psychoanalytically oriented and 'pure' systems therapists are representative since they typically share observations about patterns of relating that emerge. We will illustrate this in our account of the Milan school.[12]

The Milan approach

With assessment complete, the therapist (sometimes a pair) meets the family for about an hour. With her preparatory knowledge, she develops a hypothesis about the nature of the family's dysfunction. She has the opportunity on observing patterns *in vivo* to confirm her notions. Patterns usually emerge from the start, making the therapist's job correspondingly easier. Apart from hypothesis-testing, another key task is to engage the family so they will be motivated to reattend. We could interpolate a dictum here: a primary aim of the first session is to facilitate a second session. A vital element in encouraging engagement is for the therapist to promote a sense of curiosity in family members so that they raise questions about themselves and the family as a group.[36]

The chief strategy is circular questioning.[37] Its main purpose is to address the family's issues indirectly; this avoids applying pressure on members and possibly inviting their resistance. For example, the therapist questions an adolescent about how his parents get on with each other, a mother about how her husband relates to the eldest son, a grandmother about which grandchild is closest to the parents, and so forth. This generates illuminating data about individual members and about the family as a group. In this phase, it helps to clarify the hypothesis and to engage participants. It also affords the therapist a greater facility to remain neutral. Because the system and not a patient is the target of change, the therapist is wary of showing any bias.

Various options are then available. If the therapist works as part of a team, her colleagues have busily observed the proceedings through a one-way screen. The family's consent for this will of course have been obtained. During a break, the team—observers and therapist(s)—pool impressions.[38] This is always enlightening since team members note something that others have missed. A consensus evolves, conclusions are drawn and converted into 'messages'. The therapist returns to the family to convey them. The actual messages and their oracular quality comprise a potent intervention, but not necessarily more cogent than circular questions made during the working session. We should mention here that the narrative school has brought with it a de-emphasis on the 'message' on the grounds that 'truth' is a shared construction.

One to three messages are usually given, with maximal clarity. These have a range of purposes including promotion of intersessional 'work'. Homework may be assigned and another session planned (unless termination was set for this point). Meetings commonly occur 3 to 4 weeks apart, and for good reason. During this time, the family, armed with new ideas, tackle them in their day-to-day lives. It is not critical how they go about it but that they do so. As Cecchin has posited,[36] the family's interest in their functioning should have been so aroused that they will be motivated to continue looking at themselves between sessions.

The varied nature of the message makes them classifiable.[39] Messages are supportive, hypothesis-related, or prescriptive. First, the message has a reassuring and encouraging quality and is not related to the hypothesis. A *complimentary message* might be that 'The team were impressed by how open you all were in the session', and a *reassuring message* that 'This is tantamount to a new start for the family and uncertainties are likely'.

Hypothesis-related messages refer to the hypothesis worked out by the team, and may assume diverse forms. It may be stated directly; for example, 'Susan has assumed the role of therapist for her parents and sister to save the family from breaking up'. There may be reference to change such as, 'The team sees John taking responsibility; John and his father's improved relationship has enabled this to occur'. The family may be offered choices related to the hypothesis; for instance, 'The family could risk openness or remain self-preoccupied'. *Paradoxical messages* are a means to communicate a hypothesis which invites the family to revisit a feature of their functioning so that the family's difficulties are positively promoted and explicitly encouraged; for example, 'The team sense that your problem is working for the good of your marriage; sticking with your illness can save the marriage'. More creatively, the paradox may be split, in that the family are told about a divergence of opinion in the team[40]; for instance, some members believe it is too risky for them to communicate openly, others suggest the family can begin to do so. Through a *prescriptive message* the family is given a task. This may or may not be related to the hypothesis. For example, the family is urged to meet on their own before the next session to explore what inhibits a member from relating closely to the others.

Whatever the form of message, the therapist de-emphasizes the pathological status of the patient and applies what the Milan school calls *positive connotation*. This brilliant innovation rests on the premise that all behaviour is purposeful, and that the purpose can be construed positively. An adolescent's 'open grieving' is reframed as sparing the family the anguish of grief. This quality of message calls for creative thinking and flies in the face of the customary view of symptoms as evidence of psychopathology. Again, curiosity enters the picture as the family hears a positive communication concerning an issue they hitherto regarded as abnormal.

The process described continues in succeeding meetings, with attention also paid to family life between sessions. Duration of therapy depends on how entrenched the family dysfunction is rather than on the status of the patient's problems. Thus, systemic change is aimed for, with the family invited to consider a substitute mode of functioning that is feasible and safe. In practice, the number of sessions ranges from 1 to 12. If progress has not been made by about the seventh session, it is likely that alternate ways of helping family and/or patient are needed.

Ending therapy

Termination issues are less profound than in individual or group therapy. The reason is obvious. The family has come as a living unit and will continue as such. Even when the therapist is a pure conductor, the family's intrinsic resources are highlighted so that they can be drawn on after the therapist's departure. Determining the end-point is not usually problematic. There is a shared sense that the work has been accomplished.

A hypothesis (or set of hypotheses) has been introduced, tested, and confirmed. The family system has been examined so that impediments are recognized and understood and better modes of functioning devised and implemented. The family is not required to leave functioning optimally. Instead, termination occurs when there is agreement that the family feels confident to try out newly discovered options.

As alluded to earlier, this may be determined alongside a judgement that the identified patient (or another family member) requires another form of therapy in their own right. An adolescent who has felt unable to separate and individuate is a good example. While family work has explored the system that hindered his

'graduation', the sense prevails that he could benefit further from individual or group therapy. In another example, the parents may conclude, with the therapist's support that they have an agenda which is not pertinent to their children and is therefore best undertaken in couple therapy.

Problems encountered in therapy

Where assessment has been carried out diligently and motivation for change sustained, treatment proceeds smoothly. A crisis may still buffet the group but, rather than being derailed, the family regard it as a challenge with which to grapple.

Family treatment does not always succeed. Indeed, deterioration may occur, albeit in a small percentage of cases. What common difficulties are encountered? The non-engaging family is problematic in that while evidence points to the need for family intervention, members cannot participate in the task, usually because they resist letting go of 'the devil they know'. In another variation, engagement of some members may fail. This is particularly so in the case of fathers who, in the wake of their denial, tend to see the target of therapy as the identified patient rather than the family as a group. This belief may apply to any member.

Missed appointments may punctuate therapy, often linked to turbulent experiences between sessions or apprehension about what a forthcoming session may engender. Like any psychotherapy, drop-out is possible. On occasion, this is appropriate in that the indication for family therapy was misconstrued. In other circumstances, drop-out is tantamount to failure and may stem from such factors as therapist ineptitude, unearthing of family conflict which they cannot tolerate, and inappropriate selection of a family based on faulty assessment.

Given that the family continues as a living group during treatment, they are exposed to all manners of vicissitudes, and these may disrupt the therapeutic work. For example, an overdose by the patient, abrupt marital separation, or admission to a psychiatric hospital may take its toll and undermine treatment.

In discussing termination, we commented on outcome. Not all families will benefit. The family's dysfunction may be so intractable that it proves impervious to change, hypotheses may be 'off the mark', the family may lack sufficient psychological-mindedness, members may retreat in the face of change because of insecurity, and so forth.

Occasionally, dependency becomes a problem as the family discards any vestige of autonomy and only feels secure in the authoritative hands of the therapist. The latter may inadvertently foster such dependency by assuming undue authority, so precluding any sense of a growing partnership. The family's inherent resources are then not permitted expression.

Finally, part of the family may harbour a secret that threatens the principle of open communication. The therapist may be inveigled into a subgroup, although knowing that secrets are not conducive to the therapeutic process. For example, a wife calling the therapist to say she has been having an affair but cannot divulge this to her husband or children lest she hurt them imposes a burden on the therapist and the process (see Bloch *et al.*[27] for an overview of confidentiality in family therapy).

Sound clinical judgement is required in all these situations. Since no ready-made prescriptions are available, the therapist must be aware that difficulties may occur even in a highly motivated, well-selected family. The general principle, however, is to prevent their evolution if at all possible or to recognize them early and 'nip them in the bud'.

Research in family therapy

Selective reviews of the vast research literature in family therapy have been provided by Carr.[41,42] Larner[43] has examined political and conceptual issues raised by family therapists' attempts to conform to the criteria of evidence-based practice while Stratton and his colleagues[44] have looked at the impact of such research on clinicians. An argument for the continuing relevance of single-case studies has been mounted by Datillio[45] and the possibilities of conducting qualitative research in a systems model by Burck.[46]

The long-standing debate between the role of 'common basic factors' versus 'model-specific factors', which has bedevilled research in individual therapy, also pertains to family therapy. Simon[47] has proposed a testable hypothesis, rich in its implications, that the most effective model is one whose concepts and values most closely resemble the world view of the therapist. He offers this as a conceptual bridge between the two approaches to therapeutic research, although he curiously says nothing about the family's world view.

Modifying the psychoeducational model of family-based treatment to incorporate clinically relevant socio-cultural factors has been proposed for the treatment of adolescents and young adults suffering from schizophrenia[48] and depression.[49]

In appraising the contemporary state of family therapy research in adult clinical psychiatry, we may be cautiously optimistic. Immense strides have been made in developing theoretical concepts. As can be seen in the first section of the chapter, we have a rich array of therapeutic approaches.[50] On the other hand, the growth occurred at a dizzy pace, with the inevitable consequence of overload. How can we make sense of so many offerings? Is integration needed to forestall fragmentation? Have we reached the point to pause and reflect? Are we in a position to evaluate the effectiveness of diverse approaches, and for various types of clinical problems?

Observers of the research[51,52] have pointed to a complicating factor in contemplating future work, namely the therapist assembling a natural group, of varying composition, in which the principal goal is to improve its functioning. We face the conundrum of what constitutes optimal outcome and how it is best measured. We can best illustrate this by citing Asen and his colleagues.[53] In their trial of family therapy they had agreed to apply multidimensional measures to assess outcome—at individual, dyadic, and family system levels. At follow-up they noticed changes at the first two levels, but not in the family as a whole. The latter involved ratings of such aspects as communication, boundaries, adaptability, and competence. The researchers were candid in sharing their doubts about how to deal with the divergent findings. Several interpretations were offered: for example, no change was achieved in family functioning, the instrument of family functioning was non-reactive to treatment since it was a trait measure, and an inappropriate model of therapy was applied in the first place. The group concluded that the 'assumptive worlds' of therapists and researchers were under scrutiny rather than the families themselves.

A group in Oxford[54] encountered similar difficulties in their study of consecutive families treated in an adult family therapy clinic. Whereas two-thirds of the identified patients were judged improved at termination, only half the families were rated as

functioning at a better level. Again, the investigators were left with questions of how to determine what had actually been achieved.

A methodologically simpler method is to focus only on the identified patient. The work of Hafner *et al.*[55] exemplifies this choice—a case-controlled evaluation of family therapy in an inpatient setting with subsequent hospital admission data as the chief outcome criterion. Satisfactory as this study is in design, the omission of a family-system outcome measure leaves us ignorant about the level of family functioning following the intervention.

With these tricky matters in mind, what does research need to sort out? The diffuse question of whether family therapy works is of little utility, and is reminiscent of the sterile debate that typified individual psychotherapy outcome research for decades.[56] While subsequent meta-analyses demonstrated that psychological interventions overall exerted useful effects across a range of conditions, the field was still open to the criticism that efficacy of a specific approach for a particular clinical state remained unanswered. Family therapy should not repeat the same mistake; instead of posing the futile question of whether family therapy is effective in adult psychiatry, we should ascertain whether a specific family therapeutic approach, whose character is well identified and measurable, is useful for both the identified patient, with a specific clinical presentation, and his family's functioning, again well defined.

We have good examples of such research. Many intervention studies of families with a schizophrenic member have carefully described the principles of treatment, its rationale, process aspects, and outcome measures in the patient and (in some cases) the family.[57–60]

As mentioned earlier, multiple family group therapy (MFGT) is emerging as the preferred psychosocial intervention for adolescents and young adults with schizophrenia. A comparison with the 'open dialogue' model we mentioned earlier has not been conducted, and may not be feasible, since the latter is part of a comprehensive treatment paradigm.[61] A Danish study[62] which combined MFGT with assertive community-based treatment of young people with schizophrenia reported improvement not only of psychotic features (both positive and negative), but also a decline in use of alcohol and illicit drugs. Given the major problem of co-morbidity, this is a welcome development.

MFGT has also been directed to helping the caregivers of patients with schizophrenia. In a randomized controlled trial, Hazel *et al.*[63] found that those in the treatment programme for over a 2-year period experienced greater relief from distress compared to controls, but they were not able to determine the mechanism for this effect.

Work in the area of affective disorders and family therapy has been innovative, with MFGT, again, gaining popularity.[64] As in schizophrenia, the family treatment aims to reduce members' hostility and criticism expressed toward the depressed patient.

An excellently designed and executed study on anorexia and bulimia nervosa illustrates how outcome research can contribute to the clinical sphere.[65] In a well-controlled study, patients were randomized to either family therapy or 'routine individual supportive therapy', following their discharge from a weight-restoration programme. The family intervention focused on providing the members with information about the eating disorder and the effects of starvation. Parental anxiety was acknowledged and efforts made to help them take control of their daughter's diet. In parallel with improved physical status, therapy turned progressively to typical adolescent issues of autonomy and how these might be achieved. Overall, a structural approach was applied, with systemic and strategic methods added as necessary. Applying these principles to groups of families with adolescents with anorexia nervosa appears to be useful.[66]

While the above studies concerning particular diagnoses, and involving an identified patient, is necessary for progress,[67] this does not preclude outcome studies where the family system is the main target of change. We illustrate this with a particular form of family grief therapy.[68] The model was derived from earlier empirical research on the outcome of family grieving in an oncology setting. A 13-month follow-up yielded five family clusters, two of which were distinctly dysfunctional, two functional, and an intermediate group vulnerable to maladaptive grieving. Three dimensions of family relational functioning were critical: cohesiveness, conflict, and expressiveness. The researchers then developed a treatment model highlighting the goals of promoting cohesiveness, expressiveness, and optimal management of conflict. A screening instrument was found which could readily identify dysfunctional families. A randomized controlled trial was then carried out which showed certain family types benefiting but others remaining unchanged, and a small group even being made worse.[69]

This necessarily schematic account of research on family therapy in adult psychiatry points to action needed in future. We can best summarize what investigators should strive for as: 'Specificity is of the essence'.

Training

From charismatic figures devising innovative methods of family therapy, the field has developed into a worldwide enterprise, with dozens of books, scores of training courses, several journals, and a busy programme of national and international conferences and workshops on offer.[70] Formal training may be given as follows[71]:

1 University-based programmes that regard family therapy as a distinct professional pursuit, with a corresponding corpus of knowledge, and offer degree courses at various academic levels.

2 Free-standing institutes that also see family therapy as a distinct discipline and provide training, generally part-time and of briefer duration than university-based programmes.

3 Within university-affiliated hospitals and clinics that arrange professional training in psychiatry, psychiatric nursing, psychology, social work, and occupational therapy. Although there is a tremendous diversity in the above programmes, most include:

- Supervision of clinical work with the experienced practitioner (and perhaps other students) observing the trainee and family from behind a one-way screen. Some clinicians however consider the one-way screen as dehumanizing. They advocate instead a model of co-therapy between trainee and supervisor, often with other students sitting in the same room as the family.
- Video recording the trainee's work, which she then reviews with the supervisor and fellow students, is widely used. Tapes conducted by eminent therapists are also popular.

Whether training requires familiarity with concepts and techniques of diverse schools or whether it is preferable to develop expertise in only one school remains an open question. The free-standing institutes tend to be run by practitioners of a particular school so

that, after a mostly cursory overview of the field, training concentrates on a specific model. Wendel and his colleagues[72] have proposed a model of training for multidisciplinary mental health settings which places most emphasis on integrating empirically derived knowledge; flexibility is a crucial feature for facilitating its optimal application.

Conclusion

Family therapy has the potential to play a major role in the adult psychiatric setting. As we have commented above, research reveals several promising developments. Clinicians should seriously consider applying this mode of treatment; they will be much rewarded in doing so.

Further information

Luepnitz, D.A. (1988). *The family interpreted: feminist theory in clinical practice*. Basic Books, New York.

Byng-Hall, J. (1995). *Rewriting family scripts. Improvisation and systems change*. Guilford, London.

Kissane, D. and Bloch, S. (2002). *Family focused grief therapy*. Open University Press, Buckingham, UK.

References

1. Sander, F. (1978). Marriage and family in Freud's writings. *Journal of the American Academy of Psychoanalysis*, **6**, 157–74.
2. Flugel, J.C. (1921). *The psychoanalytic study of the family*. Hogarth, London.
3. Ackerman, N.W. (1958). *The psychodynamics of family life*. Basic Books, New York.
4. Bowen, M. (1971, 1981). *Family therapy in clinical practice*. Aronson, New York. (For an incisive critique of Bowen's theoretical contributions, see Miller, R., Anderson, S., and Keala, D. (2004). Is Bowen theory valid? A review of basic research. *Journal of Marital and Family Therapy*, **30**, 453–66.)
5. Boszormenyi-Nagy, I. and Spark, G.M. (1984). *Invisible loyalties: reciprocity in intergenerational family therapy*. Brunner-Mazel, New York.
6. von Bertalanffy, L. (1968). *General systems theory: foundation, development, applications.* Braziller, New York.
7. Minuchin, S. and Fishman, H.C. (1981). *Family therapy techniques*. Harvard University Press, Cambridge, MA.
8. Haley, J. (1976). *Problem-solving therapy*. Jossey-Bass, San Francisco, CA.
9. Bateson, G. (1972). *Steps to an ecology of mind*. Ballantine, New York.
10. Bateson, G., Jackson, D.D., Haley, J., *et al.* (1956). Toward a theory of schizophrenia. *Behavioural Science*, **1**, 251–64.
11. Bateson, G., Jackson, D.D., Haley, J., *et al.* (1962). A note on the double-bind. *Family Process*, **2**, 154–61.
12. Selvini-Palazzoli, M., Boscolo, L., Cecchin, G., *et al.* (1980). Hypothesising-circularity-neutrality: three guidelines for the conductor of the session. *Family Process*, **19**, 3–12.
13. Andersen, T. (1991). *The reflecting team: dialogues and dialogues about dialogues*. Norton, New York.
14. Anderson, H. and Goolishian, H.A. (1988). Human systems as linguistic systems: preliminary and evolving ideas about the implications for clinical theory. *Family Process*, **27**, 371–93.
15. De Shazer, S. (1985). *Keys to solution in brief therapy*. Norton, New York.
16. White, M. and Epston, D. (1990). *Narrative means to therapeutic ends*. Norton, New York.
17. Flaskas, C. and Perlesz, A. (eds.) (1996). *The therapeutic relationship in systemic therapy*. Karnac Books, London.
18. Braverman, S. (1995). The integration of individual and family therapy. *Contemporary Family Therapy*, **17**, 291–305.
19. Cooklin, A. (1979). A psychoanalytic framework for a systemic approach to family therapy. *Journal of Family Therapy*, **1**, 153–65.
20. Luepnitz, D.A. (1988). *The family interpreted: feminist theory in clinical practice*. Basic Books, New York.
21. Ciompi, L. (1988). *The psyche and schizophrenia. The bond between affect and logic*. Harvard University Press, Cambridge, MA.
22. Dare, C. (1997). Chronic eating disorders in therapy: clinical stories using family systems and psychoanalytic approaches. *Journal of Family Therapy*, **19**, 319–51.
23. Byng-Hall, J. (1995). *Rewriting family scripts. Improvisation and systems change*. Guilford, London.
24. McFarlane, W.R. (ed.) (2002). *Multiple family groups in the treatment of severe psychiatric disorders*. Guilford, New York.
25. Seikkula, J. and Olsen, M. (2003). The open dialogue approach to acute psychosis: its poetics and micropolitics. *Family Process*, **42**, 403–18.
26. American Psychiatric Association. (1994). *Diagnostic and statistical classification of diseases and related health problems* (4th edn). American Psychiatric Association, Washington, DC.
27. Bloch, S., Hafner, J., Harari, E., *et al.* (1994). *The family in clinical psychiatry*. Oxford University Press, Oxford.
28. Minuchin, S., Rosman, A., and Baker, L. (1978). *Psychosomatic families: anorexia nervosa in context*. Harvard University Press, Cambridge, MA.
29. Stierlin, H. (1989). The psychosomatic dimension: relational aspects. *Family Systems Medicine*, **7**, 254–63.
30. Eisler, I. (2005). The empirical and theoretical base of family therapy and multiple family day therapy for adolescent anorexia nervosa. *Journal of Family Therapy*, **27**, 104–31.
31. Clarkin, J., Frances, A., and Moodie, J. (1979). Selection criteria for family therapy. *Family Process*, **18**, 391–403.
32. Jenkins, H. (1989). Precipitating crises in families: patterns which connect. *Journal of Family Therapy*, **11**, 99–109.
33. McGoldrick, M. and Gerson, R. (1985). *Genograms in family assessment*. Norton, New York.
34. Beels, C. and Ferber, A. (1969). Family therapy: a view. *Family Process*, **8**, 280–332.
35. Satir, V. (1967). *Conjoint family therapy*. Science and Behaviour Books, Palo Alto, CA.
36. Cecchin, G. (1987). Hypothesizing, circularity, and neutrality revisited: an invitation to curiosity. *Family Process*, **26**, 405–13.
37. Tomm, K. (1987). Interventive questioning: part II. Reflexive questioning as a means to enable self-healing. *Family Process*, **26**, 167–83.
38. Selvini, M. and Selvini Palazzoli, M. (1991). Team consultation: an indispensable tool for the progress of knowledge. Ways of fostering and promoting its creative potential. *Journal of Family Therapy*, **13**, 31–52.
39. Allman, P., Bloch, S., and Sharpe, M. (1992). The end-of-session message in systemic family therapy: a descriptive study. *Journal of Family Therapy*, **14**, 69–85.
40. Papp, P. (1980). The Greek chorus and other techniques of paradoxical therapy. *Family Process*, **19**, 45–58.
41. Carr, A. (2005). Thematic review of family therapy journals in 2004. *Journal of Family Therapy*, **27**, 399–421.
42. Carr, A. (2006). Thematic review of family therapy journals in 2005. *Journal of Family Therapy*, **28**, 420–39.
43. Larner, E. (2004). Family therapy and the politics of evidence. *Journal of Family Therapy*, **26**, 17–39.
44. Stratton, P., McGovern, M., Wetherall, A., *et al.* (2006). Family therapy practitioners researching the reactions of practitioners to an outcome measure. *Australian and New Zealand Journal of Family Therapy*, **27**, 199–207.
45. Datillio, F. (2006). Case-based research in family therapy. *Australian and New Zealand Journal of Family Therapy*, **27**, 208–13.

46. Burck, C. (2005). Comparing qualitative research methodologies for systemic research: the use of grounded theory, discourse analysis and narrative analysis. *Journal of Family Therapy*, **27**, 237–62.
47. Simon, G. (2006). The heart of the matter: a proposal for placing the self of the therapist at the centre of family therapy, research and training. *Family Process*, **45**, 331–44.
48. Weisman, A., Duarte, E., Koneru, V., *et al.* (2006). The development of a culturally-informed, family-focused treatment for schizophrenia. *Family Process*, **45**, 171–86.
49. Breland-Noble, A.M., Bell, C., and Nicolas, G. (2006). Family first: the development of an evidence-based family intervention for increasing participation in psychiatric clinical care and research in depressed African-American adolescents. *Family Process*, **45**, 153–69.
50. Gurman, A. and Kniskern, D. (eds.) (1991). *Handbook of family therapy*, Vol. II (2nd edn). Brunner-Mazel, New York.
51. Gurman, A., Kniskern, D., and Pinsof, W. (1986). Research on marital and family therapy. In *Handbook of psychotherapy and behaviour change* (3rd edn) (eds. S. Garfield and A. Bergin), pp. 565–624. Wiley, New York.
52. Bednar, R., Burlingame, G., and Masters, K. (1988). Systems of family treatment: substance or semantics? *Annual Review of Psychology*, **39**, 401–34.
53. Asen, K., Berkowitz, R., Cooklin, A., *et al.* (1991). Family therapy outcome research: a trial for families, therapists, and researchers. *Family Process*, **30**, 3–20.
54. Bloch, S., Sharpe, M., and Allman, P. (1991). Systemic family therapy in adult psychiatry: a review of 50 families. *The British Journal of Psychiatry*, **159**, 357–64.
55. Hafner, J., MacKenzie, L., and Costain, W. (1990). Family therapy in a psychiatric hospital: a case-controlled evaluation. *Australian and New Zealand Journal of Family Therapy*, **11**, 21–5.
56. Alexander, J., Holtzworth-Munroe, A., and Jameson, P. (1994). The process and outcome of marital and family therapy: research review and evaluation. In *Handbook of psychotherapy and behaviour change* (4th edn) (eds. A. Bergin and S. Garfield), pp. 595–630. Wiley, New York.
57. Falloon, I., Boyd, J., and McGill, C. (1986). *Family care of schizophrenia: a problem-solving approach to the treatment of mental illness*. Guilford, New York.
58. Dixon, L. and Lehman, A. (1995). Family interventions for schizophrenia. *Schizophrenia Bulletin*, **21**, 631–43.
59. Mueser, K. and Bellack, A. (1995). Psychotherapy and schizophrenia. In *Schizophrenia* (eds. S. Hirsch and D. Weinberger), pp. 626–48. Blackwell Science, Oxford.
60. McFarlane, W., Dixon, L., Lukens, E., *et al.* (2003). Family psychoeducation and schizophrenia. A review of the literature. *Journal of Marital and Family Therapy*, **29**, 223–45.
61. Alanen, Y., Lehtinen, V., Lehtinen, K., *et al.* (2000). The Finnish integrated model for early treatment of schizophrenia and related psychosis. In *Psychosis: psychological approaches and their effectiveness* (eds. B. Martindale, A. Bateman, M. Crowe, and F. Marginson), pp. 235–65. Gaskell, London.
62. Peterson, L., Jeppesen, P., Thorup, A., *et al.* (2005). A randomised multicentre trial of integrated versus standard treatment for patients with first-episode psychotic illness. *British Medical Journal*, **331**, 1065–69.
63. Hazel, N., McDonell, M., Short, R., *et al.* (2004). Impact of multiple-family groups for outpatients with schizophrenia on caregivers' distress and resources. *Psychiatric Services*, **55**, 35–41.
64. Keitner, G.I., Drury, L.M., Ryan, C.E., *et al.* (2003). Multiple family group therapy for major depressive disorder. In *Multiple family groups in the treatment of severe psychiatric disorders* (ed. W. Mcfarlane), pp. 244–67. Guilford, New York.
65. Russell, G.F., Szmukler, G., Dare, C., *et al.* (1987). An evaluation of family therapy in anorexia nervosa and bulimia nervosa. *Archives of General Psychiatry*, **44**, 1047–56.
66. Sholz, M., Rix, M., Sholz, K., *et al.* (2005). Multiple family therapy for anorexia nervosa: concepts, experiences and results. *Journal of Family Therapy*, **27**, 132–46.
67. Rowe, C. and Liddle, H. (2003). Substance abuse. *Journal of Marital and Family Therapy*, **29**, 97–120.
68. Kissane, D. and Bloch, S. (2002). *Family focused grief therapy*. Open University Press, Buckingham.
69. Kissane, D., McKenzie, M., Bloch, S., *et al.* (2006). Family focused grief therapy: a randomized controlled trial in palliative care and bereavement. *The American Journal of Psychiatry*, **163**, 1208–18.
70. Liddle, H. (1991). Training and supervision in family therapy: a comprehensive and critical analysis. In *Handbook of family therapy*, Vol. II (2nd edn) (eds. A. Gurman and D. Kniskern), pp. 638–97. Brunner-Mazel, New York.
71. Goldenberg, I. and Goldenberg, H. (1996). *Family therapy: An overview*. Brooks-Cole, Pacific Grove, CA.
72. Wendel, R., Gouze, K., and Lake, M. (2005). Integrative module-based family therapy: a model for training and treatment in a multidisciplinary mental health setting. *Journal of Marital and Family Therapy*, **31**, 357–70.

6.3.9 Therapeutic communities

David Kennard and Rex Haigh

Introduction

Two of the best-known pioneers of therapeutic communities, Tom Main and Maxwell Jones, defined them as follows:

> An attempt to use a hospital not as an organization run by doctors in the interests of their own greater technical efficiency, but as a community with the immediate aim of full participation of all its members in its daily life and the eventual aim of the resocialization of the neurotic individual for life in ordinary society.[1]

> What distinguishes a therapeutic community from other comparable treatment centres is the way in which the institution's total resources, staff, patients, and their relatives, are self-consciously pooled in furthering treatment. That implies, above all, a change in the usual status of patients.[2]

Today therapeutic communities can be defined by a number of common features, but a word of warning. For reasons of historical coincidence, the term is used in the fields of mental health and addictions to refer to two somewhat different treatment models. In the addiction field they are also known as hierarchical, drug-free or concept-based therapeutic communities, or simply addiction therapeutic communities,[3] in contrast to the more democratized programmes in mental health. The two models have similar goals but their methods differ, although there are signs of increasing rapprochement between them. This chapter deals mainly with therapeutic communities in mental health, but reference will also be made to addiction therapeutic communities and those in long-term care settings. It is worth noting that those admitted to a therapeutic community for treatment are usually referred to as residents, clients, or members, rather than as patients.

Defining beliefs

Certain beliefs about human relationships and the nature of therapy are central to therapeutic communities.

1 Staff are not completely 'well' and residents are not completely 'sick'. There is a basic equality as human beings between staff and residents, who share many of the same psychological processes and experiences.

2 Whatever the symptoms or behaviour problems, the individual's difficulties are primarily in his or her relationships with other people.

3 Therapy is essentially a learning process, both in the sense of learning new skills—how to relate to others or deal more appropriately with distress—and learning to understand oneself and others.

Defining principles

A study of one of the best-known therapeutic communities, Henderson Hospital,[4] identified four principles or 'themes' that have come to be widely associated with therapeutic community treatment.

Four principles of therapeutic community treatment

- **Democratization** Every member of the community should share equally in the exercise of power in decision-making about community affairs.
- **Permissiveness** All members should tolerate from one another a wide degree of behaviour that might be distressing or seem deviant by ordinary standards.
- **Communalism** There should be tight-knit intimate sets of relationships, with sharing of amenities (dining room etc.), use of first names, and free communication.
- **Reality confrontation** Residents should be continuously presented with interpretations of their behaviour as it is seen by others in order to counteract their tendency to distort, deny, or withdraw from their difficulties in getting on with others.

Defining aspects of current practice

The generalizability of these principles to newer therapeutic communities is now being questioned and others are developing theoretical frameworks for different therapeutic communities.[5] In 2002, a quality network including most British therapeutic communities started, the 'Community of Communities', with the explicit aim of defining good practice and improving it. In 2006 the first version of 'Core Standards' was published.[6] This comprised 16 standards which were derived from consensus and consultation exercises to determine what practitioners and service users thought reflected the underlying values of therapeutic communities.

Box 6.3.9.1 illustrates a sample of eight of the standards. Note that 'all community members' should be taken to include both resident or client members, and staff.

Background

Evolution of different types of therapeutic community

Communities providing sanctuary for mentally ill people have been known as far back as the fourteenth century at Geel in Belgium. In 1796 the Retreat was opened by the Quakers in York, England, where personal relationships and social expectations in a family-like atmosphere enabled previously dangerous and unpredictable individuals to control and modify their behaviour.[7] This model, known as 'moral treatment', strongly influenced the creation of asylums in Britain and the United States in the first half of the nineteenth century. In the early twentieth century, pioneers in therapeutic education, inspired by a Christian belief in the therapeutic power of love and by Freud's new method of psychoanalysis (see Chapter 3.1), created residential schools for maladjusted children that demonstrated most of the practices and attitudes outlined above.[8] The modern equivalent of communities such as Geel can be found in the intentional communities run by third sector (voluntary) organizations such as l'Arche and the Camphill communities for people with learning disabilities. (The term 'intentional community' avoids language that implies clinical responsibility or a focus on therapy or change, and has been defined as 'a relatively small group of people who have created a whole way of life for the attainment of a certain set of goals'.) A number of therapeutic communities for children and young people now exist as voluntary organizations in the educational sector, as progressive schools, and as long-term treatment units for very disturbed children.

The history of mental health therapeutic communities for adults began during the Second World War, when the psychoanalyst Wilfred Bion was put in charge of the training wing at Northfield Military Hospital in Birmingham, England. His brief attempt in 1943 to establish a therapeutic community failed, but was soon followed by others who were more successful: Tom Main, S. H. Foulkes, and Harold Bridger at Northfield, and Maxwell Jones at Mill Hill Hospital, London. In dealing with psychiatric casualties among soldiers they developed a radical new approach, which was first described in a series of papers in 1946. One of these coined the term 'therapeutic community'.[1] Main and Jones continued to develop different versions of this new method after the war, Main as director of the Cassel Hospital and Jones at Belmont Hospital Industrial Neurosis Unit, which was renamed the Henderson

Box 6.3.9.1 Core standards for therapeutic communities

1 The whole community meets regularly

2 All community members work alongside each other on day-to-day tasks

3 All community members share meals together

4 All community members can discuss any aspects of life within the community

5 All community members create an emotionally safe environment for the work of the community

6 All community members participate in the process of a new client member joining the community

7 There is an understanding and tolerance of disturbed behaviour and emotional expression

8 Positive risk taking is seen as an essential part of the process of change

Hospital in 1958. The Cassel Hospital continues as an inpatient psychotherapy hospital, and Henderson Hospital replicated itself in 2000 to serve national needs for 'severe personality disorder' provision by founding Main House in Birmingham and Webb House in Crewe.

The creation of the National Health Service in 1948 provided the stimulus to address the major problems of institutionalization revealed in a number of studies of large mental hospitals in the United Kingdom and United States.[9,10] In the 1950s and 1960s social psychiatry was in the ascendancy and a number of these hospitals developed what Clark called the 'therapeutic community approach".[11] In the 1970s and 1980s concepts of collective responsibility fell from favour and individualism prevailed, with a decline in the fortunes of therapeutic communities. The 1990s and 2000s have seen a revival of interest in therapeutic communities within more specific mental health contexts, including prisons, personality disorder services, and for the management of people with enduring mental illness in the community. The problem of degraded and poorly functioning inpatient units is now being addressed by attention to establishing and maintaining 'therapeutic environments' in acute settings, in a direct parallel to the 'therapeutic community approach' 40 years earlier.[12,13]

Alongside these developments two other types of therapeutic community have emerged. In 1958 a self-help organization in the United States called Synanon became the prototype for concept-based therapeutic communities for ex-addicts. Phoenix House and Daytop were two major programmes that grew from this, and today therapeutic communities modelled on them can be found in more than 50 countries worldwide.[3] A development that grew out of the antipsychiatry movement in the 1960s is known at Soteria. These are small low-stress family-like environments where psychosis is responded to with intensive therapeutic support rather than medication. These communities are mainly found in Europe.[14]

Scientific background

Therapeutic communities have drawn on the concepts of psychoanalysis, group analysis (see Chapter 6.3.6), humanistic and integrative psychotherapies, and on sociological studies of mental hospitals which identified phenomena such as the total institution[10] and patterns of behaviour associated with psychiatric treatment in institutions.[15] They are also underpinned by studies of the impact of unconscious processes in organizations,[16,17] and by anthropological studies such as that of Rapoport[4] which found a typical pattern of oscillation in the therapeutic community.

A developmental model based on the 'required emotional experiences' of attachment, containment, communication, inclusion, and agency has been proposed by Haigh.[18] This identifies ways in which a range of psychological theories and approaches are relevant to therapeutic community practice, and illustrates how they are replicated in the structures and culture of a therapeutic community. It also proposes that disturbance of 'primary emotional development' (which all humans undergo early in life) can to some extent be made good by a satisfactory experience of 'secondary emotional development' in a therapeutic community.

Technique—how change is brought about

Since the therapeutic community *is* the treatment, managing treatment involves attention to two parallel processes: the progress of each resident through the community, and the effective functioning of the therapeutic community as a whole. Responsibility for managing these two processes ultimately belongs to the staff, though it is shared with the residents when the community is functioning well.

Most if not all the treatment in a therapeutic community takes place in groups and in the everyday life of the community, although some also use individual psychotherapy. The essence of the therapeutic community technique has been encapsulated in two phrases.

1 A *living–learning* situation: this refers to the fact that everything that happens between members of a therapeutic community in the course of living together, and in particular when a crisis occurs, is used as a learning opportunity.[19]

2 *Culture of enquiry*: this refers to the creation not just of certain structures but of a basic culture among the staff of 'honest enquiry into difficulty'. There is a conscious effort to identify and challenge dogmatic assertions or accepted wisdoms.[20]

The basic mechanism of change is not difficult to explain. The therapeutic community provides a wide range of lifelike situations in which the difficulties a member has experienced in their relationships with others outside are re-experienced, with regular opportunities in small group and community meetings to examine and learn from these difficulties. If the therapeutic community is to work as a therapeutic method, all its constituent parts described in this chapter must be in good-working order. This requires a process by which new members adopt the values of the community, emphasizing openness, responsibility, and active participation, and in turn pass these on when the next new members arrive. To operate this mechanism requires both staff and residents to fulfil a number of roles (Box 6.3.9.2).

Staff training

The need for specialized training for leading or working as a staff member in a therapeutic community has been a matter of some debate.[21] There is an argument that the emphasis on egalitarianism and democratization means that this form of treatment is best delivered by people without special training who can just 'be themselves'. Unqualified social therapists often form a key part of the staff complement. However, the other argument is now increasingly accepted, that while being oneself is an important part of the staff role (see above), therapeutic community work requires a high level of skill and knowledge in a number of areas, together with a well-developed capacity for open honest discussion and reflection.

Training courses exist in the United Kingdom, Finland, Norway, the Netherlands, and Greece, but there is no set standard or curriculum. Some relevant theoretical training is obtainable as part of group therapy and systemic therapy courses, and the case has also been made for a placement in a therapeutic community to be part of professional training in psychiatry and psychotherapy. The benefits include learning at first hand about the treatment of personality disorders, experience of group-based treatment, and working as a member of a multi-disciplinary team, thus gaining first-hand experience of institutional dynamics and the way individuals react to their wider social networks. One of the most popular short courses in the United Kingdom and the Netherlands

Box 6.3.9.2 Staff and client/resident roles in a therapeutic community

Role	Role activity of client/resident members	Role activity of staff members
Participation and involvement in the daily life of the community('Living–learning')	To explain how the community works to those referred, to visitors, and new members	To spend informal time with client/residents (those who appear aloof may be challenged about this)
	To take responsibility for various tasks which contribute to the running of the community	To work alongside clients/residents in day-to-day tasks
	To contribute to various 'extras' such as being involved in teaching and research	To monitor other staff members' emotional involvement and consider in supervision
	To notice and include those who isolate themselves	
Contributing to and managing therapeutic processes ('Culture of enquiry')	To use identification to challenge or support peers	To use therapeutic interventions in groups of various modalities (e.g. group analytic, psychodrama, art therapy, CBT)
	To be open to the challenge and support of one's peers	To encourage client/resident members to take a therapeutic role, in some cases delegating responsibility for out-of- hours support to them
	To support those in crisis, often including out-of-hours	
	To maintain community structures such as rules and timekeeping	
Responsibility for decision-making and 'Democratization'	To make decisions about day-to-day and domestic matters, including rotas and elections	To decide the level of decision-making which would be optimal for the therapeutic benefit of the client group, according to their capabilities and needs
	To participate in decisions about therapeutic matters (e.g. deciding consequences of breaking rules)	To monitor the functioning of the therapeutic environment and titrate the level of staff input and leadership required
	To be involved in planning local events and service developments (e.g. open days, starting new groups)	To ensure good communication with managers, commissioners, referrers, clinical colleagues, and other relevant organizations
	Senior and ex-members can offer invaluable support in maintaining effective external relations (e.g. with organizational executives)	

is a brief residential simulated therapeutic community.[22] Here, 20 to 30 health professionals live together for a few days in the roles of residents with a 'staff' group working with them. This provides a valuable opportunity to experience at first hand the workings and impact of this form of treatment.

Indications and contraindications

Universal indications for therapeutic community treatment are difficult to give. Modified therapeutic communities have been developed for people with different types and levels of psychiatric disorder, and even the same therapeutic community may fluctuate in its capacity to absorb difficult members. An individual's suitability will need to be judged in relation to a particular therapeutic community at a particular time. Having made this caveat, the general indications and contraindications will usually apply (Box 6.3.9.3).

Pathways and process: phases in the therapeutic journey

There are usually four distinct but overlapping phases in a member's journey to and through a therapeutic community:

Engagement phase: Referral, preparation, and selection procedures are an integral part of therapeutic community practice, involving both the prospective member and existing residents as active participants in the process, which start with referral or self-referral. Many prospective members of therapeutic communities are wary or fearful of the forthcoming therapy, and need support and encouragement to persist. This is often effectively delivered by current or ex-members, arranged in partnership with voluntary agencies, or with internet support groups. During this time, any regular support from mental health teams or other agencies should continue.

Assessment and preparation phase: When a decision to proceed towards formal treatment has been made, several arrangements may need to be made. These include formal assessment processes, practical planning, and agreeing a treatment contract. This work is frequently arranged through the use of an 'assessment and preparation group', which is also designed to be a time-limited foretaste of what the treatment phase entails. The assessment process can be undertaken in this group itself, in smaller groups, or with individual appointments. The practical planning involves matters such as arranging childcare, securing stable accommodation, and agreeing plans for medication and risk management. It also includes an

Box 6.3.9.3 Indications and contraindications for therapeutic community treatment

Indications	Indications for specialized TC	Contraindications
Diagnosis of:		
◆ Personality disorder	First episode psychosis	Physically dependent addictions
◆ Self-harm	Serious and enduring mental illness	Current mania
◆ Adjustment disorders	Less than 18 years old	Depression with severe retardation
◆ Recurrent depressive disorders	Learning disabilities	Dementia
◆ Bipolar disorder	Perpetrators of sexual abuse	Dangerously low weight
◆ Intractable anxiety disorders		Antisocial PD with history of intimidation and deception
◆ Eating disorder		No capacity for social involvement
◆ Addictions		Inability to see problems in terms of relationships
		Unwillingness to engage in informal, intimate, and open style of relating with professionals
Age:		
No age limits: young children to elderly members can show benefit		Belief that only experts can help

explicit treatment contract, which may be a verbal agreement about understanding the community rules, or a formal written and signed agreement.

Treatment phase: This usually begins with a formal 'case conference' or 'selection panel' including current community members with a decision made by voting. Subsequent therapy programmes vary considerably: from 1 day per week to whole-time residential; from predominantly sociotherapy to a range of psychoanalytic, cognitive behavioural, humanistic and interpersonal and systemic groups; from a few weeks duration to several years, either time-limited or open-ended; and from group size of less than 6 to more than 50. Some communities include individual therapy, but others consider this inimical to the group dynamic process. A modal or typical programme would be for between 3 and 5 days per week, with all interventions in groups comprising a mixture of community meetings, small therapy groups, shared lunch, and informal time together. A typical community would have between 12 and 24 members divided into three small groups, who would stay for 12 to 18 months. Suitable arrangements would be in place for crisis meetings to be called at short notice, as would a system for members to support each other out-of-hours.

During the first few weeks of treatment the new member will be feeling his or her way, forming attachments to one or two others, but still wary of the groups. After the first month or two he or she will begin participating more actively in the groups, taking part in the full life of the community with certain role responsibilities, helping and supporting other members. This will probably include experience of situations similar to those triggering referral, such as having to deal with authority, fear of failure, feeling rejected or abandoned, situations evoking rivalry and competition, or many others. As before, these may trigger destructive or violent impulses towards the self or another person, or the experience of other symptoms of distress. Through the group meetings the member is confronted with the effects of their behaviour on fellow members and the meaning of the behaviour or symptoms is explored, making full use of the insights and understanding of fellow members. Through this repeated process the member gradually comes to experience himself and others differently. As one member wrote: *Bit by bit, almost grudgingly, the fact dawned on me that I wasn't surrounded by forty sticks of furniture but by Jim, Gary, Jane . . .* [23]

Re-entry phase: Until recently, many therapeutic communities had a 'cliff-edge' ending, where one day members are able to have the community's full emotional and practical support, and the next are not allowed to contact any other members. Although this has some theoretical justification in terms of 'coming to term with endings', and has strong advocates amongst ex-members of therapeutic communities, it is now generally considered better practice to support members over the leaving process, and then into re-establishing mainstream social networks. This can be done with a specific 'leavers group', that members join while in the full treatment phase and continue afterwards, for either a fixed or indefinite period. These groups normally include a practical focus, and are social and supportive rather than exploratory and therapeutic. For those who are ready and able, they can have objectives of securing employment or education for members. In the case of prison-based therapeutic communities for ex-addicts the provision of drug-free housing and vocational training have been found to improve success rates.[24]

As well as planned endings there are various other types of ending. Some members may leave prematurely, unable to cope with therapy; some may be 'voted out' by the community for a serious or repeated transgression of community rules. Such endings do not necessarily indicate a treatment failure, although the longer members remain the more likely they are to benefit.

Research evidence

The effectiveness of therapeutic communities has been investigated in relation to different clinical problems, which are discussed separately.

Personality disorders

Until recently there has been little systematic evidence of the efficacy of therapeutic communities for treating personality disorders, and disagreement over whether those who did benefit were really suffering from psychopathic or personality disorder. While efficacy in this area is still questioned (see Chapter 4.12.7), the picture has recently become clearer with the publication of the first systematic

review of therapeutic community treatment for people with personality disorders.[25] The authors carried out a full search of therapeutic community publications and grey literature, collecting over 8000 references from 38 countries. These were reduced to 29 research studies that met the criteria of randomized controlled trial design (eight studies) or comparative or controlled studies that reported raw data and used conservative outcome criteria (e.g. reconviction rates rather than psychological improvement). A meta-analysis found that 19 studies showed a positive effect within the 95 per cent level of confidence while the remaining 10 straddled the neutral score. The overall summary log odds ratio was −0.567, with a 95 per cent confidence interval, −0.524 to −0.614. The authors concluded that there is strong evidence for the effectiveness of therapeutic communities. A more recent systematic review assessed evidence for interventions for people with PD in general and for dangerous and severe personality disordered offenders and made clear recommendations about the most promising treatment interventions for PD in use or currently in development. The reviewers covered therapeutic community programmes; cognitive, behavioural, cognitive behavioural, and psychodynamic psychotherapies; pharmacological and physical treatments. They concluded that 'the TC model currently has the most promising evidence base in this poor field'.[26]

Offending behaviour

Therapeutic communities have been established in prisons to deal with disruptive, violent inmates, and also with the underlying problems of antisocial personality disorder. Results at Barlinnie Special Unit in Scotland demonstrated substantial reductions in violent incidents within prison. Reconviction studies carried out at Grendon, a prison run entirely on therapeutic community lines, found that prisoners had lower rates of reconviction, fewer custodial sentences, and fewer reconvictions for violent offences than prisoners on the Grendon waiting list who never went there. Those who stayed at Grendon longer than 18 months showed the greatest reductions in reconviction rates. Re-offending rates have also been found to be lower in the former Federal Republic of Germany for prisoners in Social Therapeutic Institutions than those receiving standard prison sentences.[27] In 2004, the 'Democratic Therapeutic Community Core Model' was accredited as a treatment programme for use in prisons in England and Wales. Such programmes must show evidence of their capacity to impact on dynamic risk factors known to be associated with re-offending. Risk factors that therapeutic communities have been found to have a positive influence on include negative attitudes towards authority, identification with antisocial role models, and acceptance of responsibility for offending behaviour.

Drug dependence

Therapeutic communities for drug dependence use the hierarchical or concept-based model. These form one part of the range of treatments for drug abuse, which includes other residential models such as Christian communities and the Minnesota model, as well as methadone maintenance programmes and psychotherapy. Therapeutic communities for drug dependence, often modified to suit different local cultures, can now be found in many European and international countries as well as in the United States where they originated. Although the model began, and continues, as a residential peer-support programme in the community, it has adapted well to secure environments, and concept-based therapeutic communities for drug dependence have been established in American prisons since the mid-1980s, accompanied by a growing number of aftercare programmes providing employment and drug-free accommodation. Several national studies have evaluated the outcome of these programmes. Randomized controlled trials show that no-treatment groups have a higher level of recidivism than those who complete treatment in a prison therapeutic community, and that recidivism is further reduced by participation in a community aftercare therapeutic community.[24] Since 1995 a number of these therapeutic communities have also been established in English prisons. The general conclusion, for both secure and non-secure therapeutic communities, is that residents who stay in programmes for longer periods have lower rates of drug use and criminal behaviour and higher rates of employment than those who stay for shorter times. However, there are no direct comparisons between therapeutic communities and other treatment models, and it is likely that therapeutic communities are successful for those who are well motivated. Although this may be only a relatively small proportion of all drug abusers, studies, and admission policies, suggest those who enter these high intervention therapeutic communities tend to be severely addicted and damaged, often dually diagnosed with personality disorder, and less likely to respond to low intervention treatments.[28]

First episode psychosis

A small number of therapeutic communities have been developed in the United Kingdom, United States, Switzerland, and Germany on the principle that first episodes of psychosis can be effectively treated in low-stress family-like settings providing round the clock personal support, with no or minimal use of neuroleptics. This has become known as the Soteria model. Two Soteria houses, in California and Berne, have been subjected to randomized or matched control trials comparing them with usual hospital treatment. One study found that completing subjects with schizophrenia exhibited a large effect size benefit with Soteria treatment, especially in the areas of psychopathology, work, and social functioning. Length of stay in Soteria Berne was initially longer but this was subsequently reduced to less than the admission ward. In both studies the 2-year outcomes were at least as good in the Soteria group and less antipsychotics were prescribed for the Soteria group.[14] A 20-year study of an acute psychiatric ward in Finland found that people with acute psychotic episodes and borderline conditions seemed to benefit from the therapeutic community model with a high level of support, negotiation, order, and organization.

Severe and enduring mental illness

The therapeutic community approach[11] has been widely used in large mental hospitals to counter the effects of institutionalization and to mobilize the residual capacity of those suffering from chronic mental illness for social relationships, purposeful employment, and personal responsibility. The method was as much about improving the sense of purpose and morale of the staff and the general quality of life in the institutions as it was about clinical improvement, and its success was demonstrated in the way some large old mental hospitals were turned into centres of excellence.

With the re-provision of services for people with enduring mental illness in the community, therapeutic community principles have been found to be an effective way of structuring staffed hostels and homes. In one version of this, the 'ward in a house', the model is close to the original practice of the York Retreat, an antecedent of therapeutic communities (see above).[29] The challenges presented by the severely mentally ill chemical user have also been addressed using a modified therapeutic community with some evidence of success. Three main modifications required to the TC structure were increased flexibility, decreased intensity, and greater individualization.[30]

Children and adolescents

Therapeutic communities for children and adolescents were first developed in the field of therapeutic education almost a century ago, and now exist for a variety of needs: learning disability, delinquency, and emotional disturbance. Little systematic evaluation has been carried out. A survey of 186 children in nine therapeutic communities for emotionally disturbed children found evidence of increased stability and hopeful outcomes for those who stayed. A 20-year follow-up of 28 children in one community reported evidence of long-term improvement.[31]

Wright and Richardson, reviewing the current state of research for therapeutic communities for children and young people, conclude that, 'when rigorous quality controls are introduced there are as yet too few studies to draw any aggregated conclusions. Perhaps the clearest qualified statement would be that there is low-level evidence that some residential therapeutic placements produce changes in the mental and social functioning of some young people who have been unable to cope with family life'.[32]

Learning disabilities

Long-term residential communities for adults with learning disabilities exist in the charitable sector in many countries around the world. These are value based rather than evidence based. There is some evidence that many families express a strong preference for village-style communities such as the Camphill for their mentally handicapped relatives.[33]

Further information

Campling, P. and Haigh, R. (eds.) (1999). *Therapeutic communities: past, present and future*. Jessica Kingsley, London.

Campling, P., Davies, S., and Farquharson, G. (2004). *From toxic institutions to therapeutic environments*. Gaskell, London.

Community of Communities. (2006). *Service standards for therapeutic communities*. Royal College of Psychiatrists, London.

Kennard, D. (1998). *An introduction to therapeutic communities* (2nd edn). Jessica Kingsley, London. www.therapeuticcommunities.org—website of the Association of Therapeutic Communities.

References

1. Main, T. (1946). The hospital as a therapeutic institution. *Bulletin of the Menninger Clinic*, **10**, 66–70.
2. Jones, M. (1968). *Social psychiatry in practice*, pp. 85–6, Penguin, Harmondsworth.
3. Rawlings, B. and Yates, R. (2001). *Therapeutic communities for the treatment of drug users*. Jessica Kingsley, London.
4. Rapoport, R.N. (1960). *Community as doctor*. Tavistock, London.
5. Campling, P. and Haigh, R. (eds.) (1999). *Therapeutic communities: past, present and future*. Jessica Kingsley, London.
6. Community of Communities. (2006). *Service standards for therapeutic communities*. Royal College of Psychiatrists, London.
7. Tuke, S. (1813). *Description of the retreat*. Reprinted by Process Press, London, 1996.
8. Bridgeland, M. (1971). *Pioneer work with maladjusted children*. Staples Press, London.
9. Barton, R. (1959). *Institutional neurosis*. John Wright, Bristol.
10. Goffman, I. (1961). *Asylums*. Doubleday, New York; Penguin, Harmondsworth, 1968.
11. Clark, D.H. (1965). The therapeutic community—concept, practice and future. *The British Journal of Psychiatry*, **131**, 553–64.
12. Campling, P., Davies, S., and Farquharson, G. (2004). *From toxic institutions to therapeutic environments*. Gaskell, London.
13. Hardcastle, M., Kennard, D., Grandison, S., *et al.* (2007). *Experiences of mental health in-patient care: narratives from service users, carers and professionals*. Routledge, London.
14. Ciompi, L. and Hoffman, H. (2004). Soteria Berne: an innovative milieu therapeutic approach to acute schizophrenia based on the concept of affect-logic. *World Psychiatry*, **3**, 140–6.
15. Stanton, A. and Schwartz, H. (1954). *The mental hospital*. Basic Books, New York.
16. Menzies, I. (1960). A case-study in the functioning of social systems as a defence against anxiety. *Human Relations*, **13**, 95–121.
17. Main, T. (1957). The ailment. *British Journal of Medical Psychology*, **30**, 129–45.
18. Haigh, R. (1999). The quintessence of a therapeutic environment. In *Therapeutic communities: past, present and future* (eds. P. Campling and R. Haigh), Chap. 20. Jessica Kingsley, London.
19. Jones, M. (1968). *Social psychiatry in practice*, pp. 105–12. Penguin, Harmondsworth.
20. Main, T. (1983). The concept of the therapeutic community: variations and vicissitudes. In *The evolution of group analysis* (ed. M. Pines), pp. 197–217. Routledge and Kegan Paul, London.
21. Roberts, J. (1998). Questions of training. In *An introduction to therapeutic communities* (ed. D. Kennard). Jessica Kingsley, London.
22. Rawlings, B. (2005). The temporary therapeutic community—a qualitative evaluation of an ATC training weekend. *Therapeutic Communities*, **26**, 6–18.
23. Mahoney, N. (1979). My stay at the Henderson Therapeutic Community. In *Therapeutic communities: reflections and progress* (eds. R.D. Hinshelwood and N. Manning), pp. 76–87. Routledge and Kegan Paul, London.
24. Wexler, H. (1997). Therapeutic communities in American prisons. In *Therapeutic communities for offenders* (eds. E. Cullen, L. Jones, and R. Woodward), pp. 161–79. Wiley, Chichester.
25. Lees, J., Manning, N., and Rawling, B. (1999). *Therapeutic community effectiveness. A systematic international review of therapeutic community treatment for people with personality disorders and mentally disordered offenders*. CRD Report 17, NHS Centre for Reviews and Dissemination, University of York, York.
26. Warren, F., Preedy-Fayers, K., McGauley, G., *et al.* (2003). *Review of treatments for severe personality disorder*. Home Office Online Report 30/03, Home Office, London.
27. Rawlings, B. (1999). Therapeutic communities in prisons: a research review. *Therapeutic Communities*, **20**, 177–93.
28. Yates, R. and Wilson, J. (2001). The modern therapeutic community: dual diagnosis and the problem of change. In *Therapeutic communities for the treatment of drug users* (eds. B. Rawlings and R. Yates). Jessica Kingsley Publishers, London and Philadelphia.
29. Leff, J. and Trieman, N. (1997). Providing a comprehensive community psychiatric service. In *Care in the community: illusion or reality* (ed. J. Leff), pp. 189–201. Wiley, Chichester.

30. Sacks, S. (2000). Co-occurring mental and substance use disorders: promising approaches and research issues. *Substance Use and Misuse*, **35**, 2061–93.
31. Rose, M. (1997). *Transforming hate to love: an outcome study of the Peper Harow treatment process for adolescents*. Routledge, London.
32. Wright, J.C. and Richardson, P. (2003). The challenge of research. In *Therapeutic communities for children and young people* (eds. A. Ward, K. Kasinski, J. Pooley, and A. Worthington), pp.244–53. Jessica Kingsley, London.
33. Cox, C. (1995). The case for village communities for people with learning disabilities. *British Journal of Nursing*, **4**, 1130–4.

6.4

Treatment by other professions

Contents

6.4.1 Rehabilitation techniques

W. Rössler

The goal of psychiatric rehabilitation is to help disabled individuals to establish the emotional, social, and intellectual skills needed **to live, learn, and work in the community** with the least amount of professional support.(1)

Rehabilitation practice has changed the perception of mental illness. Enabling disabled people to live a normal life in the community causes a shift away from a focus on an illness model towards **a model of functional disability**. As such, other outcome measures aside from clinical conditions become relevant. Social role functioning including social relationship, work, and leisure as well as quality of life and family burden are of major interest for the people affected living in the community.(25)

The relevance of psychosocial and environmental problems is reflected in the DSM-IV and ICD-10. Axis IV of DSM-IV and codes Z55–Z65 and Z73 of ICD-10 are assigned for reporting psychosocial and environmental problems that may affect the diagnosis, treatment, and prognosis of mental disorders.

The International Classification of Functioning, Disability and Health

Long-term consequences of major mental disorders might be described using different dimensions. A useful tool was provided by the International Classification of Impairment, Disability and Handicaps (ICIDH), first published by the World Health Organization in 1980. The ICIDH has been recently revised. The revised '**International Classification of Functioning, Disability and Health'** (ICF) includes a change from negative descriptions of impairments, disabilities and handicaps to neutral descriptions of body structure and function, activities and participation. A further change has been the inclusion of a section on environmental factors as part of the classification. This is in recognition of the importance of the role of environmental factors in either facilitating functioning or creating barriers for people with disabilities. Environmental factors interact with a given health condition to create a disability or restore functioning, depending on whether the environmental factor is a facilitator or a barrier.

ICF is a useful tool to comprehend chronically mentally ill in all their dimensions including impairments at the structural or functional level of the body, at the person level concerning activity limitations and at the societal level with respect to restrictions of participation. Each level encompasses a theoretical foundation on which a respective rehabilitative intervention can be formulated.

Target population

During the course of psychiatric reforms the predominant objective of psychiatric rehabilitation was to resettle patients from large custodial institutions to community settings. Today all patients suffering from severe mental illness require rehabilitation. The core group is drawn from patients with the following:

- persistent psychopathology
- marked instability characterized by frequent relapse
- social maladaptation.(28)

There are other definitions currently used to characterize the chronically mentally ill. But they all share some common elements, that is a diagnosis of mental illness, prolonged duration, and **role incapacity**.

Although the majority of the chronically mentally ill have the diagnosis of schizophrenic disorders, other patient groups with psychotic and non-psychotic disorders are targeted by psychiatric rehabilitation. Up to 50 per cent of people with severe mental illness carry **dual diagnoses** especially in combination with substance abuse.

The role of the psychiatrist in rehabilitation

Psychiatric rehabilitation is by its very nature, **multidisciplinary;** because of the many different competencies required. Monitoring medication is a key task of the psychiatrist.

Pharmacotherapy in psychiatric rehabilitation needs some special consideration. Symptom control does not necessarily have the highest priority as some side effects of pharmacological treatment can weaken a person's ability to perform his social roles, and impair vocational rehabilitation. Many patients living in the community **want to take responsibility for their medication** themselves. This also includes the varying of medication without consultation within certain limits.

The starting point for an adequate understanding of rehabilitation is that it is concerned with the individual person in the context of his or her specific environment. Psychiatric rehabilitation is regularly **carried out under real life conditions**. Thus, rehabilitation practitioners have to take into consideration the realistic life circumstances that the affected persons are likely to encounter in their day-to-day living.

A necessary second step is helping disabled persons to **identify their personal goals**. Motivational interviews provide a sophisticated approach to identify the individuals' personal costs and benefits associated with the needs listed.[8] This makes it also necessary to assess the individuals' readiness for change.[15] Functional assessment and individual goal setting are prerequisites of a differentiated rehabilitation intervention plan and should be repeated in different stages of the rehabilitation process.

The rehabilitative planning process **focuses on the patient's strengths**. Irrespective of the degree of psychopathology of a given patient, the rehabilitation practitioner must **work with the 'well part of the ego'** as 'there is always an intact portion of the ego to which treatment and rehabilitation efforts can be directed' (Lamb 1984[2]). This leads to a closely related concept: the aim of restoring hope to people who suffered major setbacks in self-esteem because of their illness. As Bachrach[2] states 'it is the kind of hope that comes with learning to accept the fact of one's illness and one's limitations and, proceeding from there'.

Psychiatric rehabilitation **concentrates on peoples' rights as a respected partner** and **endorses their involvement and self-determination** concerning all aspects of the treatment and rehabilitation process. These rehabilitation values are also incorporated in the **concept of recovery.**[9] Within the concept of recovery, the **therapeutic alliance** plays a crucial role in engaging the patient in his or her own care planning. It is essential that the patient can rely on his or her therapist's understanding and trust as most of the chronically mentally ill and disabled persons lose close, intimate, and stable relationships in the course of the disease. Recent research has suggested that social support is associated with recovery from chronic diseases, greater life satisfaction, and enhanced ability to cope with life stressors.[24] Therefore, psychiatric rehabilitation is also an exercise in network building.

Current approaches

Psychiatric rehabilitation aims at **changing the natural course** of the disease. Yet, there is no consensus among rehabilitation researchers on what rehabilitation actually does accomplish. Some understand rehabilitation as an approach to help disabled people to compensate for impairments and to function optimally with the deficiencies they have. Other researchers assume that rehabilitation helps the patient to recover from the disorder itself, while the contributing factors to the healing process are not clear.

The overall philosophy of psychiatric rehabilitation comprises two intervention strategies. The first strategy is individual-centered and aims at **developing the patient's skill** to interact with a stressful environment. The second strategy is ecological and is directed towards **developing environmental** resources to reduce potential stressors. Most disabled people need a combination of both approaches.

As a general rule people with psychiatric disabilities tend to have the **same life aspirations** as people without disabilities in their society or culture. They want to be respected as autonomous individuals and lead a life as normal as possible. As such they mostly desire (1) their own housing, (2) an adequate education and a meaningful work career, (3) satisfying social and intimate relationships, and (4) participation in community life with full rights.

Housing

The objective of psychiatric reforms since the mid-50s of the 20th century has been to resettle chronically mentally ill persons from large custodial institutions to community settings. Providing sheltered housing in the community for the long-term patients of the old asylums was one of the first steps in the process of deinstitutionalization. Most long-stay patients can successfully **leave psychiatric hospitals and live in community settings**.

Ideally, a **residential continuum** (RC) with different housing options should be provided. RC ranges from round-the-clock staffed sheltered homes to more independent and less staffed sheltered apartments, which eventually allow individuals moving to independent housing in the community. Critics of RC contended that (1) up to date RC is rarely available in communities, (2) that RC does not meet the varying and fluctuating needs of persons with serious mental illnesses, and (3) that RC does not account for individuals' preferences and choices. **Supported housing**, i.e. independent housing coupled with the provision of support services emerged in the 1980s as an alternative to RC. Supported housing offers flexible and individualized services depending on the individual's demands. In the meantime, rehabilitation research could demonstrate that supported housing is a realistic goal for the majority of people with psychiatric disabilities.[23] Once in supported housing, the majority stay in housing and are less likely to become hospitalized. Other outcomes do not yield consistent results.

Work

The beneficial effects of work on mental health have been known for centuries.[11] Therefore, **vocational rehabilitation** has been

a core element of psychiatric rehabilitation since its beginning. Vocational rehabilitation is based on the assumption that work not only improves activity, social contacts, etc., but may also promote gains in related areas such as self-esteem and quality of life, as work and employment are a step away from dependency and a step closer to integration into society. **Enhanced self-esteem** in turn improves adherence to rehabilitation of individuals with impaired insight.

Vocational rehabilitation originated in psychiatric institutions where the lack of activity and stimulation led to apathy and withdrawal of their inpatients. Long before the introduction of medication, occupational and work therapy contributed to sustainable improvements in long-stay inpatients. Today occupational and work therapy are not any longer hospital-based but represent the starting point for a wide variety of rehabilitative techniques teaching vocational skills.

Vocational rehabilitation programs in the community provide a series of **graded steps to promote job entry or re-entry**. For less disabled persons, brief and focused techniques are used to teach how they can find a job, fill out applications, and conduct employment interviews. In transitional employment, a temporary work environment is provided to teach vocational skills, which should enable the affected person to move on to competitive employment. But all too often, the gap between transitional and competitive employment is so wide that the mentally disabled individuals remain in a temporary work environment. Sheltered workshops providing pre-vocational training also quite often prove a dead end for the disabled persons.

One consequence of the difficulties in integrating mentally disabled individuals into the common labour market has been the steady growth of cooperatives, which operate commercially with disabled and non-disabled staff working together on equal terms and sharing in management. The mental health professionals work in the background providing support and expertise.

Today, the most promising vocational rehabilitation model is **supported employment** (SE). In SE, disabled persons are placed in competitive employment according to their choices as soon as possible, and receive all support needed to maintain their position.[4] The support provided is continued indefinitely. Participation in SE programs is followed by an increase in the ability to find and keep employment.[7] Links were also found between job tenure and non-vocational outcomes, such as improved self-esteem, social integration, relationships, and control of substance abuse.[4,29] It was also demonstrated that those who had found long-term employment through SE had improved cognition, quality of life, and better symptom control.[17]

Although findings regarding SE are encouraging, some critical issues remain to be answered. Many individuals in SE obtain unskilled part-time jobs. Since most studies only evaluated short (12–18 months) follow-up periods, the long-term impact remains unclear. Currently, we do not know which individuals benefit from supported SE and which do not.[20] After all, we have to realise that the integration into the labour market does by no means only depend on the ability of the persons affected to fulfill a work role and on the provision of sophisticated vocational training and support techniques but also on the **willingness of society to integrate** its most disabled members.

Building relationships

In recent years, social skills training packaged in the form of modules with different topics has become very popular and has been widely promulgated. The modules focus on medication management, symptom management, substance abuse management, basic conversational skills, interpersonal problem solving, friendship and intimacy, recreation and leisure, workplace fundamentals, community (re-)entry, and family involvement. Each module is composed of skill areas. The skills areas are taught in exercises with demonstration videos, role-play and problem solving exercises, and *in-vivo* and homework assignments.[14]

The results of several control studies suggest that disabled individuals **can be taught a wide range of social skills**. Social and community functioning improve when the trained skills are relevant for the patient's daily life, and the environment perceives and reinforces the changed behaviour. Unlike medication effects, benefits from skills training occur more slowly. Furthermore, long-term training has to be provided for positive effects.[3] Overall, social skills training have been shown to be effective in the acquisition and maintenance of skills and their transfer to community life.[13]

Keeping relationships

As a consequence of deinstitutionalization the **burden of care** has increasingly fallen on the relatives of the mentally ill. Informal caregiving significantly contributes to health care and rehabilitation.[31] Fifty to ninety per cent of disabled persons live with their relatives following acute psychiatric treatment. This is a task many families do not choose voluntarily. Caregiving imposes a significant burden on families. Those providing informal care face considerable adverse health effects, including higher levels of stress and depression, and lower levels of subjective well being, physical health and self-efficacy. Additionally, not all families are equally capable of giving full support to their disabled member and are not willing to replace an insufficient health care system. Caregivers regularly experience higher levels of burden when they have poor coping resources and reduced social support. But **families also represent support systems**, which provide natural settings for context-dependent learning important for recovery of functioning. As such, there has been a growing interest in helping affected families since the beginning of care reforms.

One area of interest deals with the expectations of relatives concerning the provision of care. Relatives quite **often feel ignored**, not taken seriously, and also feel insufficiently informed by health professionals. They also may feel that their contribution to care is not appreciated or that they will be blamed for any patient problems. It certainly is no surprise that there is a lot of frustration and resentment among relatives considering the physical, financial and emotional family burden.

Family intervention programs have produced promising results. Family intervention is effective in lowering relapse rate, and also in improving outcome e.g., psychosocial functioning. Possibly, family intervention can reduce family burden. Furthermore, the treatment gains are fairly stable.[21] But we also have to appreciate, that it is not clear what the effective components of the different models are. Additionally, family interventions differ in frequency and length of treatment. There are also no criteria for the minimum amount of treatment necessary.

Finally, we have to be aware that most family interventions were developed in the context of western societies during deinstitutionalization. Family caregiving might be quite different in a **different cultural context**. This refers to other cultures in total as well as to minority groups in western societies.[31]

Participation in community life with full rights

Practitioners often are confronted with the **deleterious effects of stigma and discrimination** in the lives of people with serious mental illnesses. Numerous studies have examined stigmatizing attitudes toward people with mental illness.[12] In recent years, the scientific interest in the perspective of the labelled individual has increased too. There is extensive empirical evidence of the negative consequences of labelling and perceived stigmatization. These include demoralization, low quality of life, unemployment, and reduced social networks.[10,19] Once assigned the label 'mental illness' and having become aware of the related negative stereotypes, the affected individuals expect to be rejected, devaluated or discriminated against. This vicious cycle decreases the chance of recovery and normal life.

On the other hand, well-integrated people with mental illness exhibit better outcomes regarding psychopathology and quality of life. The importance of social integration is underlined even more when considering the subjective availability of support: perceived social support predicts outcome in terms of recovery from acute episodes of mental illness, community integration, and quality of life.[27,30]

On the basis of comprehensive research in this area during the last decade, several strategies have been developed to fight the stigma and discrimination suffered by those who have mental illnesses.[30] Different research centres developed interventions directed to specific target groups relevant for de-stigmatization, e.g., students[18] or police officers.[22] Persons in contact with mentally ill individuals quite often have a more positive attitude. Contact with the mentally ill persons also reduces social distance,[12] which is a strong argument in favour of community psychiatry. Other initiatives have targeted stigma by means of more comprehensive programs. The World Psychiatric Association launched one of the internationally best-known programs in 1996 (www.openthedoors.com). All these initiatives make clear that efforts in re-integrating persons with serious mental illness into community life must be accompanied by measures on the societal level.

The core elements of modern psychiatric rehabilitation are summarized in Box 6.4.1.1

Developing environmental resources

Effective psychiatric rehabilitation requires **individualized and specialized treatment**, which has to be embedded in a comprehensive and coordinated system of rehabilitative services. But even when a variety of services are available, they are poorly linked in many cases, and costly duplication may occur.

While developing community support systems it became obvious that there is a **need to coordinate and integrate the services** provided as each involved professional concentrates on different aspects of the same patient. Therefore, as a key coordinating and integrating mechanism, the concept of **case management** (CM) originated. CM focuses on all aspects of the physical and social environment. The core elements of CM are the assessment of patients' needs, the development of comprehensive service plans for the patients, and arrangement of service delivery.[26]

Over the past two decades, a variety of different models of CM have been developed which exceed the original idea that CM mainly intends to link the patient to needed services and to coordinate those services. Today most clinical case managers also provide direct services in the patient's natural environment. This model is called **Intensive Case Management** (ICM). ICM on its part is difficult to distinguish from **Assertive Community Treatment** (ACT).

Stein and Test have developed the basic compounds of ACT in the 1970's.[33] The original program was designed as a community based alternative to hospital treatment for persons with severe mental illnesses. A comprehensive range of treatment, rehabilitation, and support services in the community is provided through a multidisciplinary team. ACT is characterized by an assertive outreach approach i.e., interventions are mainly provided in the natural environment of the disabled individuals.[32]

Research on CM and ACT yielded 'mixed' results.[6] While the traditional office-based CM approach obviously is less successful, the ACT model was found to be more beneficial when compared with standard care.[16] ACT can reduce time in hospital,[20] but has moderate or only little effects on improving symptomatology and social functioning. The differing features of the respective services might explain the international variation. Six regularly occurring features of successful services were identified: smaller case loads, regularly visits at home, a high percentage of contacts at home, responsibility for health and social care, multidisciplinary teams, and a psychiatrist integrated in the team.[5]

Box 6.4.1.1 Core elements of psychiatric rehabilitation

- multidisciplinary
- carried out under real life conditions
- identify patient's personal goals
- focus on the patient's strengths
- work with the 'well part of the ego,
- concentrate on people's rights as a respected partner
- endorse their involvement and self-determination
- build a therapeutic alliance

References

1. Anthony, W. (1979). *The principles of psychiatric rehabilitation*. University Park Press, Baltimore, MD.
2. Bachrach, L.L. (2000). Psychosocial rehabilitation and psychiatry in the treatment of schizophrenia – what are the boundaries? *Acta Psychiatr Scand Suppl* **407**, 6–10.
3. Bellack, A.S. (2004). Skills training for people with severe mental illness. *Psychiatr Rehabil J*, **27**(4): 375–91.
4. Bond, G.R. (2004). Supported employment: evidence for an evidence-based practice. *Psychiatr Rehabil J*, **27**(4): 345–59.
5. Burns, T, Catty, J, Wright, C. (2006). De-constructing home-based care for mental illness: can one identify the effective ingredients? *Acta Psychiatr Scand Suppl*, **429,** 33–5.
6. Burns, T, Fioritti, A, Holloway, F. *et al*. (2001). Case management and assertive community treatment in Europe. *Psychiatr Serv*, **52**(5), 631–6.
7. Cook, J.A., Leff, H.S., Blyler, C.R. *et al*. (2005). Results of a multisite randomized trial of supported employment interventions for individuals with severe mental illness. *Arch Gen Psychiatry,* **62**(5), 505–12.
8. Corrigan, P.W., McCracken, S.G., Holmes, E.P. (2001). Motivational interviews as goal assessment for persons with psychiatric disability. *Community Ment Health J*, **37**(2), 113–22.

9. Farkas, M., Gagne, C., Anthony, W. *et al.* (2005). Implementing recovery oriented evidence based programs: identifying the critical dimensions. *Community Ment Health J,* **41**(2), 141–58.
10. Graf, J., Lauber, C., Nordt, C. *et al.* (2004). Perceived stigmatization of mentally ill people and its consequences for the quality of life in a Swiss population. *J Nerv Ment Dis*, **192**(8), 542–7.
11. Harding, C., Strauss, J., Hafez, H. *et al.* (1987). Work and mental illness. I. Toward an integration of the rehabilitation process. *Journal of Nervous and Mental Disease*, **175**, 317–26.
12. Lauber, C., Nordt, C., Falcato, L. *et al.* (2004). Factors influencing social distance toward people with mental illness. *Community Ment Health J*, **40**(3), 265–74.
13. Liberman, R.P., Glynn, S., Blair, K.E. *et al.* (2002). In vivo amplified skills training: promoting generalization of independent living skills for clients with schizophrenia. *Psychiatry,* **65**(2), 137–55.
14. Liberman, R.P., Kopelowicz, A. (2002). Teaching persons with severe mental disabilities to be their own case managers. *Psychiatr Serv* **53**(11), 1377–9.
15. Liberman, R.P., Wallace, C.J., Hassell, J. (2004). Rehab rounds: Predicting readiness and responsiveness to skills training: the Micro-Module Learning Test. *Psychiatr Serv*, **55**(7), 764–6.
16. Marshall, M. (1996). Case management: a dubious practice. BMJ, **312**(7030), 523–4.
17. McGurk, SR., Mueser, K.T. (2003). Cognitive functioning and employment in severe mental illness. *J Nerv Ment Dis*, **191**(12), 789–98.
18. Meise, U., Sulzenbacher, H., Kemmler, G. *et al.* (2000). ["…not dangerous, but nevertheless frightening". A program against stigmatization of schizophrenia in schools]. *Psychiatr Prax*, **27**(7), 340–6.
19. Mueller, B., Nordt, C., Lauber, C. *et al.* (2006). Social support modifies perceived stigmatization in the first years of mental illness: a longitudinal approach. *Soc Sci Med* 2006, **62**(1), 39–49.
20. Mueser, K.T., Bond, G.R., Drake, R.E. *et al.* (1998). Models of community care for severe mental illness: a review of research on case management. *Schizophrenia Bulletin*, **24**, 37–74.
21. Pilling, S., Bebbington, P., Kuipers, E. *et al.* (2002). Psychological treatments in schizophrenia: I. Meta-analysis of family intervention and cognitive behaviour therapy. *Psychol Med*, **32**(5), 763–82.
22. Pinfold, V., Huxley, P., Thornicroft, G. *et al.* (2003). Reducing psychiatric stigma and discrimination–evaluating an educational intervention with the police force in England. *Soc Psychiatry Psychiatr Epidemiol*, **38**(6), 337–44.
23. Rog, DJ. (2004). The evidence on supported housing. *Psychiatr Rehabil J,* **27**(4), 334–44.
24. Rogers, E.S., Anthony, W., Lyass, A. (2004). The nature and dimensions of social support among individuals with severe mental illnesses. *Community Ment Health J*, **40**(5), 437–50.
25. Rössler, W. (2006). Psychiatric rehabilitation today: an overview. World Psychiatry **5**(3), 151–7.
26. Rössler, W., Fätkenheuer, B., Löffler, W. *et al.* (1992). Does case management reduce the rehospitalization rate? *Acta Psychiatrica Scandinavica*, **86**, 445–9.
27. Rössler, W., Salize, H.J., Cucchiaro, G. *et al.* (1999). Does the place of treatment influence the quality of life of schizophrenics? *Acta Psychiatr Scand*, **100**(2), 142–8.
28. Royal College of Psychiatrists (1996). *Psychiatric rehabilitation* (revised edition). Gaskell, London.
29. Ruesch, P., Graf, J., Meyer, P.C. *et al.* 2004). Occupation, social support and quality of life in persons with schizophrenic or affective disorders. *Soc Psychiatry Psychiatr Epidemiol*, **39**(9), 686–94.
30. Rusch, N., Angermeyer, M.C., Corrigan, P.W. (2005). Mental illness stigma: concepts, consequences, and initiatives to reduce stigma. *Eur Psychiatry,* **20**(8), 529–39.
31. Schulze, B., Rössler, W. (2005). Caregiver burden in mental illness: review of measurement, findings and interventions in 2004–2005. *Current Opinion in Psychiatry*, **18**, 684–91.
32. Scott, J.E. and Dixon, L.B. (1995). Assertive community treatment and case management for schizophrenia. *Schizophrenia Bulletin*, **21**, 657–68.
33. Stein, L.I. and Test, M.A. (1980). Alternative to mental hospital treatment. I. Conceptual model, treatment program, and clinical evaluation. *Archives of General Psychiatry*, **37**, 392–7.

6.4.2 Psychiatric nursing techniques

Kevin Gournay

Background

Psychiatric nursing as an entity has really only evolved since the Second World War. Psychiatric nurses (now often referred to as mental health nurses in the United Kingdom and Australasia) can now be found in most countries of the developed world, although in the developing world, psychiatric nursing is still not defined as a specific discipline. In many countries, psychiatric hospitals are still staffed by untrained 'Attendants' who may have some supervision from general trained nurses. Nevertheless, a number of initiatives, notably those of the Geneva Initiative in Psychiatry[(1)] in Eastern Europe and the former Soviet Union and the World Health Organization in African countries, have provided specific training in psychiatric nursing techniques.

The development of psychiatric nursing across the world needs to be seen in the context of changing and evolving patterns of mental health care. De-institutionalization, with the attendant setting up of community mental health teams, has prompted a range of innovations in psychiatric nursing and the psychiatric nurse of today, who in the United States and Europe is likely to be a university graduate, is a very different person to that of the nurse working in the post-Second World War asylums of 40 years ago.

In this chapter, we examine the development of psychiatric nursing in some detail and particularly emphasize the role of psychiatric nurses working in the community. Community psychiatric nursing first developed in the United Kingdom nearly 50 years ago and this model has been followed in countries such as Australia and New Zealand. However, this community role has not developed to any great extent in the United States, where the main presence of psychiatric nursing remains in hospital-based care. Furthermore, in the United Kingdom and Australasia, the development of community initiatives has seen the role of the psychiatric nurse blurring with that of other mental health professionals. Chapters such as this cannot really do justice to the whole range of techniques used by psychiatric nurses; neither can it examine in any detail the differences between psychiatric nursing practices across the world. However, a description of psychiatric nursing in six important areas will provide the reader with an appreciation of the range and diversity of psychiatric nursing skills:

- Inpatient care
- Psychosocial interventions in the community
- Prescribing and medication management

- Cognitive behaviour therapy
- Primary care
- Psychiatric nursing in the developing world.

Psychiatric nursing in inpatient settings

In the past three decades the population in psychiatric hospitals across the developed world has fallen dramatically in England from 160 000 to 30 000 beds over a period of 25 years and the duration of inpatient care in the United Kingdom in 2007 is approximately 36 days. However, today's inpatients are a population with much greater levels of illness than was previously the case; they tend to be more treatment-resistant, have complex problems, and display high levels of substance abuse and violence.[2] As a corollary of this, a greater proportion of patients are now detained under mental health legislation. Inpatient facilities consist of acute psychiatric units, local secure units, and high secure psychiatric hospitals for those patients who pose high levels of danger to themselves and others. In the United Kingdom, four high secure hospitals contain approximately 1600 patients. It should also be noted that, due to the large numbers of people with psychiatric problems in prisons, there are now several hundred psychiatric nurses employed in prison settings to carry out a range of assessment and treatment procedures. In addition, the NHS also has a number of 'in reach' schemes, which include sending NHS staff into prisons on a sessional basis.

Given that community care in the western World is now the norm, inpatient care is now seen as a short-term measure with the dual purposes of stabilizing the patient's condition and keeping the patient safe. Psychiatric nurses have a role to play in the overall assessment of the patient and, given that the nurse is—literally—with the patient 24 h a day, the observation of the patient's mental state and behaviour is of considerable importance. Unfortunately, this is an area where, outside the United States, a number of problems exist and suicide rates by inpatients are unacceptably high.[3] In the United States, inpatient wards tend to be much more secure than wards in countries such as the United Kingdom and Australia and, therefore, the incidence of inpatient suicide is much lower. The *UK National Confidential Inquiry into suicides and homicides*[3] demonstrates that nearly 200 suicides by inpatients occur every year, with hanging on the ward itself being still prevalent at unacceptably high levels. Recently the National Institute for Health and Clinical Excellence (NICE)[4] has published guidelines, which include the observation of patients at risk. This guidance sets out very careful protocols for the observation of patients at risk and includes recommendations regarding the prevention of absconding. In the United Kingdom and Australia, open-door policies still operate in acute psychiatric units and it is being increasingly recognized that balancing the rights of the patients against safety is a difficult issue. Nurses also have a major role to play in providing patients and their families with information about condition and treatment. We also know that there are interventions that can be applied by nurses, which would lead to improved outcomes. For example, Drury *et al.*[5] showed that a cognitive behavioural therapy package improved longer-term outcomes. Similarly, Kemp *et al.*[6] showed that motivational interviewing and psychoeducation methods produced clear, clinical, and economic benefits in patients who have compliance problems with medication.

With regard to the containment of violent behaviour, which is now so common in inpatient settings, nurses in the United Kingdom have been assisted by very comprehensive evidence-based guidance from the National Institute for Health and Clinical Excellence (NICE),[4] which sets out clear guidance on the use of de-escalation techniques and control and restraint, as well as providing a comprehensive algorithm for the use of rapid tranquillization. In respect of rapid tranquillization, nurses are now provided with the necessary skills to observe and monitor patients following rapid tranquillization, including the use of pulse oximetry and blood pressure. Whilst nurses in Australasia use the same methods of managing violent behaviour as nurses in the United Kingdom, psychiatric nurses in most European countries and in the United States use various forms of mechanical restraint and a very wide range of devices, including belts, straps, nets, and jackets. Whilst it needs to be recognized that there are a range of social and cultural influences that determine how violence in mentally ill people is managed, it is important to note that the evidence base for all forms of violence management, including rapid tranquillization, is very poor and a Cochrane review found that there is no evidence base for the use of seclusion and restraint.[7]

Psychosocial interventions in the community

In order to appreciate the current practice of psychiatric nurses working in the community, it is important to say something about the historical context. Until the early 1980s, community psychiatric nurses (CPNs) in the United Kingdom were generally based in large, Victorian psychiatric hospitals and worked mostly within a consultant psychiatrist team responsible for the follow-up of patients after discharge from hospital. Their main responsibilities were the administration of medication and the provision of general, supportive care, mostly to people with schizophrenia, the elderly with functional and organic illnesses, and to people with other serious and enduring mental illnesses. Initial research on the effectiveness of community psychiatric nurses produced very positive results. In a randomized trial conducted by Paykel *et al.*[8] CPNs were compared with psychiatric registrars in the provision of aftercare for patients who had suffered an acute episode requiring hospitalization. In general terms, this study showed that there was an equivalent outcome on clinical, social, and economic measures. Some 20 years ago, CPNs in the United Kingdom began to diversify their practice and separated themselves from consultant psychiatrists, attaching themselves to primary care settings and taking referrals directly from GPs. By 1990, a national survey showed that 40 per cent of CPNs worked in primary care.[9] The vast majority of this work involved treating people with depression, anxiety, and adjustment disorder, using counselling-based approaches. Whilst this work by CPNs became very popular with GPs and mental health professionals in general, research into the effectiveness of their work demonstrated that they were largely ineffective. Gournay and Brooking[10] carried out a randomized controlled trial involving 11 CPNs, working in six primary care settings in North London. In this study, 177 patients were randomized to either routine continuing care from their GP or to CPN intervention. The majority of patients had adjustment disorders and various states of general depression and anxiety. Patients, in both the CPN and

continuing GP care groups, showed significant improvement on a range of measures, clinical status, and social functioning but, at post-treatment and follow-up, there was no difference in outcomes demonstrated. Patients allocated to CPNs showed high levels of dropout (50 per cent) and patient satisfaction rating did not correlate with outcome measures. An economic analysis[11] showed that, per unit of health gain, CPN intervention was very expensive compared with interventions for people with schizophrenia. The Paykel and Gournay and Brooking studies still represent the only research evidence regarding the efficacy of CPNs working with common mental disorders.

During the early 1990s a National Review of Psychiatric Nursing in the United Kingdom led to CPNs refocusing their efforts on the seriously mentally ill and this trend has been followed in Australasia. In the last decade there has been a wide range of psychiatric nursing developments in respect of psychosocial interventions. The initial impetus for this development came from the Thorn Programme, this initiative taking its name from the Sir Jules Thorn Trust, a charitable foundation that provided the funds to inaugurate the first 3 years of the training programme for nurses, commencing in 1992. The initiative was originally led by Dr Jim Birley who, with a group of colleagues from other professions, became impressed by the work of nurses working in cancer care. Birley's initial aim was to train a substantial number of nurses specifically dedicated to the care of people with schizophrenia and their relatives. Indeed, Birley, who was one of the pioneers of Social Psychiatry in the United Kingdom, noted that the families of people with schizophrenia were often in great need of intervention. Previous work in Manchester[12] had confirmed that nurses could be trained in family intervention skills which in turn led to positive outcomes for the patient and family. This training in family work formed the basis of what has now become a more general initiative to train nurses in various evidence-based psychosocial interventions for schizophrenia. The Thorn Initiative has now become the national model of training in psychosocial interventions in the United Kingdom and similar programmes to Thorn have been set-up in Australasia and some European countries. The psychosocial interventions used by nurses are as follows:

- Assertive community treatment
- Family interventions for schizophrenia
- Cognitive behavioural techniques for managing hallucinations and delusions
- Approaches with dual diagnosis
- Medication management.

In addition to psychosocial interventions, training programmes for psychiatric nurses working in the community, now also include approaches to improve the physical health of people with serious mental health problems and nurses are now taking a more active lead in ensuring that this very vulnerable population obtains appropriate medical services including physical screening and health promotion activities. At the time of writing, there are also, in several parts of the United Kingdom, specific training programmes for nurses aimed at helping patients with chronic mental illness to deal with obesity, lack of exercise, and smoking.

On a cautionary note, there are now several studies[13] which demonstrate that more intensive case management may not be effective, the possible reason being that one needs to provide nurses with suitable levels of training in community approaches. In the UK700 study[13] mentioned above the nurses involved only received a few hours training in case management approaches, whist nurses undertaking basic psychosocial interventions training such as a Thorn diploma will receive 250 h of classroom instruction in addition to supervised practice. Whilst there have been considerable numbers of nurses trained in the above-mentioned evidence-based psychosocial interventions, unfortunately there are still many nurses working in the community without such training. Whilst their general psychiatric nursing skills will be reasonably sound, their impact on patient care will be somewhat limited.

In the early part of the twenty-first century, psychiatric nurses working in the United Kingdom, Europe, and Australasia are increasingly working in specialist community teams, for example, assertive outreach services, early intervention teams and crisis intervention, and home treatment teams. Whilst these approaches are commonly used across the United States, such teams are likely to be staffed by case managers who have backgrounds in social work and social care and psychiatric nurses are unlikely to be employed in large numbers. In the United States, psychiatric nurses are often specifically employed to run medication clinics, probably for reasons of cost, whilst in the United Kingdom, approximately 50 per cent of community mental health teams carrying out a very wide range of psychosocial functions are likely to be CPNs.

Prescribing and medication management

The work of Kemp *et al.*[6] who showed that motivational interviewing and psychoeducation methods produced good outcomes for patients who were non-compliant with their medication, led to the development of medication management training for nurses in the United Kingdom. Gray *et al.*[14] using a cluster randomized controlled trial, where 60 CPNs were randomly assigned to medication management training or carrying on with their usual treatment as usual, demonstrated that the nurses who had received medication management training produced very clear benefits in patients with schizophrenia. The study demonstrated a significantly greater reduction in patients' overall psychopathology for the trained group, compared with treatment as usual. At the end of the 6-month study period, the improvement in positive and negative symptoms for the trained nurses over the control was statistically and clinically significant. This training is now used by, literally, thousands of nurses in the United Kingdom, Australasia, and some non-English speaking countries in Europe. Whilst this training, which comprised a number of components, including improving the pharmacological knowledge of the nurses, the use of side effect monitoring and motivational interviewing for non-compliant patients, a European multi-centre trial, which tested adherence therapy as a treatment package over and above routine clinical care and delivered mostly by psychologists and psychiatrists, showed that the treatment package was no more effective than health education in improving quality of care.[15] Both studies raised a number of questions concerning the very complex issue of treatment compliance with medication and the specific difficulties associated with measurement of compliance itself and of patient insight.

Arguably, the most important recent development in psychiatric nursing has been the advent of nurse prescribing. This began more than a decade ago in the United States, where nurses have

prescriptive authority in virtually all states. The situation in the United States is, however, complex with a considerable variation in the level of prescriptive authority across the United States from complete independence to being able to prescribe under a physician protocol. In turn, the educational requirement for nurse prescribers also varies considerably. However, most states have fairly comprehensive regulations concerning not only course content, but also hours of instruction and supervision. The Website of the American Psychiatric Nurses,[16] provided at the end of this chapter, provides very detailed state-by-state information. In 2006, United Kingdom legislation was passed that means nurses may prescribe almost independently,[17] although—as in the United States—there is a variation across the country in terms of training and practice. Across nursing more generally, the law in the United Kingdom means that provided it is within their area of specialist work, nurses may independently prescribe any drug (including controlled substances such as opiates). By contrast to the different legislative frameworks that exist in the United States, the legislative framework for the United Kingdom is unitary. However, interpretation of that framework in the United Kingdom seems to vary between NHS services. There are now similar nurse prescribing initiatives in Australasia, where because of the nature of rural and remote populations, the development of nurse prescribing seems very logical.

At present, there are no randomized controlled trial data to compare nurse prescribing with more conventional doctor prescribing, neither is there any substantial data on patient efficacy. The advantages of nurse prescribing have been clearly set out in a *Maudsley discussion paper* by Gournay and Gray[18] and these include the delegation of routine prescribing tasks to nurses, so that psychiatrists may concentrate their prescribing efforts on difficult-to-manage patients who may be treatment-resistant and/or non-compliant and those patients who have substantial physical health co-morbidity. Another advantage might be that CPNs, who are case managers, may be able to spend more time than their psychiatrist colleagues in the detailed evaluation of effectiveness and side effect monitoring and management.

Cognitive behaviour therapy

For more than 35 years, nurses in the United Kingdom have been trained to provide psychological treatment to patients with various mental health problems. These developments began in 1972, when Isaac Marks, a psychiatrist working at the Maudsley hospital, began a 3-year experiment to determine whether nurses could be trained to deliver behavioural interventions for neurotic disorder. Isaac Marks was one of the first to recognize that the workforce of psychologists would be insufficient to deliver evidence-based treatment. Subsequently, Marks[19] published data which demonstrated both the clinical and economic effectiveness of nurses working with neurotic disorders in primary care. Over the years, training programmes for nurses have developed and now nurses are trained in a variety of university and clinical settings alongside their psychology colleagues in the practice of evidence-based psychological treatments, i.e. cognitive behaviour therapy for a very wide range of disorders. Whilst the original efforts to train nurses were centred on techniques for the treatment of phobias and obsessive–compulsive disorder, nurses are now trained more comprehensively in cognitive behavioural methods, which encompass treatment techniques used in the treatment of depression, schizophrenia, and personality disorders. In recent years there has been a small, but significant, growth in the United Kingdom of psychiatric nurses employed as psychological therapists. However, this trend has not been replicated in Australasia or Europe, where legislative frameworks prevent nurses from obtaining full accreditation as psychological therapists. In the United States, the situation is variable. Nevertheless, the American Psychiatric Nurses' Association membership comprises a significant number of nurses who have full accreditation as therapists in their respective states. Such nurses are now often prepared at a post-doctoral level and their expertise is arguably equivalent to that of their clinical psychologist colleagues.

Primary care

Following the refocus of CPNs efforts on people with schizophrenia and other serious and enduring illnesses, there have recently been a number of United Kingdom policy developments that will lead to more psychiatric nurses being employed in primary care settings. This trend follows the recognition that many people with common mental health problems do not receive evidence-based treatment or, if they do, they are subject to long periods on waiting lists. Psychiatric nurses are now being trained to provide brief evidence-based interventions in primary care and also to provide important assessment and screening functions, so as to ensure that patients who need the services of the community mental health teams are suitably referred and those who can be managed at primary care level are provided with appropriate treatments. Nurses are now also increasingly involved in the delivery of computerized cognitive behaviour therapy, which, as a recent NICE review demonstrated,[20] is effective in the treatment of a wide range of mental health problems. This method of treatment is particularly important given the scarcity of skilled therapist resources. Psychiatric nurses are therefore increasingly involved in the running of 'Computer clinics', which now use a very wide variety of treatment packages for a whole range of disorders. Whist most of these packages come at a cost, there are now free to access programmes on the Internet. These include 'Moodgym',[21] a programme for the self-help treatment of depression, which is based on a cognitive behavioural approach and developed in Australia at the Australian National University in Canberra. Moodgym has been evaluated and its use is supported by positive randomized trial data.[22]

Psychiatric nursing in the developing world

As noted above, there have been a number of training initiatives in the former communist countries of Eastern Europe and the former Soviet Union.[1] Nevertheless, the numbers of trained psychiatric nurses across this region still remains fairly small. It is also clear that the skills of psychiatric nurses, in these countries, are compromised by the relatively poor general standards of mental health care.

With the exception of South Africa, psychiatric nursing is very poorly developed in the African continent. However, the World Health Organization has a number of projects that aim to integrate mental health into primary care provision. For example, the Ministry of Health in Kenya, the Kenyan Psychiatric Association, and the Kenyan Nursing Council are working with the UK Department for International Development on a 5-year programme

that began in 2005, which has the ambitious aim of training 3000 primary care workers with some skills in very basic mental health care.

In Asia, psychiatric nursing is becoming increasingly recognized and there are now substantial training initiatives in the Indian subcontinent and China. Nevertheless, the numbers of psychiatric nurses who are needed in those vast countries are very great and, at present, the best way of describing the psychiatric nursing presence would be to say that it is sparse and patchy, with most nurses working in the large cities. An additional problem is that, where reasonable standards of education and training exist (and this applies particularly to India and China) there is a considerable loss to immigration to the developed world—to countries such as the United Kingdom and the United States, where the recruitment and retention of psychiatric nurses is an ongoing problem, and is financially very attractive.

There is obviously enormous potential for the development of psychiatric nursing across the developing world and, although current initiatives have focused on providing nurses with basic skills, there is obviously enormous potential for the provision of evidence-based psychosocial interventions and, given the tremendous shortage of psychiatrists in many countries, nurse prescribing (providing that the nurses have received adequate education and training) could obviously potentially provide widespread benefits for a substantial proportion of the literally millions of people whose illnesses are currently untreated.

Conclusion

Psychiatric nursing is a relatively new profession, which has evolved over the past 50 years, from a branch of general nursing, where the main role focused on the custodial care of people with chronic serious and enduring mental illnesses—such as schizophrenia—to the present day situation, where nurses are, in the developing world at least, more likely to be university graduates who may be employed in a number of very diverse roles. The setting for these roles could be in inpatient settings, where custodial care is challenging, to say the least, to the community with roles involving psychosocial interventions within primary health care teams, to inpatient and community settings where nurses are increasingly autonomous prescribers of medication or psychological therapists. It is pleasing to note that the education and training of psychiatric nurses is gradually becoming more evidence based and policy makers are apparently much more aware of the need to provide focused skill sets on populations of need.

Further information

From values to action: The Chief Nursing Officer's review of mental health nursing (2006)—England—http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4133839

American Psychiatric Nurses Association—http://www.apna.org/i4a/pages/index.cfm?pageid=1

Newell, R. and Gournay, K. (2008). *Mental health nursing: an evidence based approach* (2nd edn). Elsevier, London.

Australian and New Zealand College of Mental Health Nursing—http://www.acmhn.org/index.html

References

1. Geneva Initiative Training. http://www.geneva-initiative.org/pages/projects/projects.asp
2. Healthcare Commission. (2005). The National Audit of Violence 2003–2005. www.healthcarecommission.org.uk
3. National Confidential Inquiry. (2006). *Avoidable deaths: five year report of the National Confidential Inquiry into suicide and homicide by people with a mental illness*. University of Manchester. www.medicine.manchester.ac.uk/suicideprevention/nci/useful/avoidable_deaths.pdf
4. National Institute for Health and Clinical Excellence (NICE) (2005). Violence. The short-term management of disturbed/violent behaviour in psychiatric inpatient settings and emergency departments. *Clinical Guidelines 25*. www.nice.org.uk
5. Drury, V., Birchwood, M., Cochrane, R., *et al.* (1996). Cognitive therapy in recovery from acute psychosis, a controlled trial. *The British Journal of Psychiatry*, **169**, 593–607.
6. Kemp, R., David, A., Hayward, P., *et al.* (1998). Compliance therapy, an 18 month follow up. *The British Journal of Psychiatry*, **5**, 228–35.
7. Sailas, E. and Fenton, M. (2002). Seclusion and restraint for people with serious mental illnesses. *The Cochrane Library*, (1). Update Software, Oxford.
8. Paykel, E., Mangen, S., Griffith, J., *et al.* (1982). Community psychiatric nursing for neurotic patients: a controlled trial. *The British Journal of Psychiatry*, **140**, 573–81.
9. White, E. (1990). *The third quinnenial survey of CPNs*. Department of Nursing Studies, University of Manchester.
10. Gournay, K. and Brooking, J. (1994). The CPN in primary care: an outcome study. *The British Journal of Psychiatry*, **165**, 231–8.
11. Gournay, K. and Brooking, J. (1995). The CPN in primary care: an economic analysis. *Journal of Advanced Nursing*, **22**, 769–78.
12. Brooker, C., Fallon, I., Butterworth, A., *et al.* (1994). The outcome of training community psychiatric nurses to deliver psychosocial intervention. *The British Journal of Psychiatry*, **165**, 222–30.
13. Burns, T., Fiander, M., Kent, A., *et al.* (2000). Effects of caseload size on the process of care of patients with severe psychotic illness: report from the UK700 trial. *The British Journal of Psychiatry*, **177**, 427–33.
14. Gray, R., Wykes, T., Edmonds, M., *et al.* (2004). Effect of a medication management training package for users on clinical outcomes for patients with schizophrenia: cluster randomised controlled trial. *The British Journal of Psychiatry*, **185**, 157–82.
15. Gray, R., Leese, M., Bindman, J., *et al.* (2006). Adherence therapy for people with schizophrenia. *The British Journal of Psychiatry*, **189**, 508–14.
16. American Psychiatric Nurses' Association. www.apna.org
17. Extended Nurse Prescribing. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH 4006775
18. Gournay, K. and Gray, R. (2001). *Should mental health nurses prescribe? Maudsley discussion paper*. Institute of Psychiatry, London.
19. Marks, I. (1985). *Nurse therapists in primary care*. RCN Publications, London.
20. National Institute for Health and Clinical Excellence (NICE) (2006). The clinical and cost effectiveness of computerised cognitive behaviour therapy for depression and anxiety. *Technology Appraisal 51*. NICE, London. http//www.nice.org.uk/
21. Moodgym website: http://moodgym.anu.edu.au/
22. Christensen, H., Griffiths, K.M., and Jorm, K. (2004). Delivering interventions for depression by using the internet: randomised controlled trial. *British Medical Journal*, **328**, 265.

6.4.3 Social work approaches to mental health work: international trends

Shulamit Ramon

The historical development of mental health social work

Social work was formally established in most European countries and North America at the end of the nineteenth century, before it took off gradually in other countries. It usually developed out of charitable work, which focused on financial support for poor families. The second main strand in social work was represented by the Settlement Movement, which concentrated on improving the communal life of poor people by living with them, using community work methods to support and empower.

The major impetus to developing mental health social work at the beginning of the twentieth century was related to the work of leading psychiatrists and psychologists with shell-shocked (PTSD) soldiers during the First World War.[1] This approach lead to the establishment of the psychodynamically oriented Tavistock Clinic in London, where the first British psychiatric social worker was appointed in 1920.[2] Stuart[3] argues that American mental health social work in the pre-1920 period concentrated more on care in the community in its social, rather than its administrative or psychological meaning, than after 1920, when it shifted further to the psychological dimension.

Social workers in both the children and the adults outpatient services provided comprehensive psychosocial history of the child/adult and their family, enabled parents, teachers, and partners of adult clients to understand the underlying psychological reasons for the index client's mental ill health, and guided them as to how they could actively support that family members.[2] It was only in the 1950s that qualified psychiatric social workers began to work in hospitals.

Since the 1970s all English-speaking countries have also opted for deinstitutionalization as their core mental health policy, leading to the closure of many of their psychiatric hospitals, replacing them by community-based services and by small psychiatric wards in general hospitals.[4,5] With the notable exception of Italy,[6,7] most continental European countries have opted for bed reduction coupled with less extensive community services.

The trend towards deinstitutionalization is formally adhered to also in Latin America, but thus far is only practised in some small-scale projects.[8,9] This applies also to Asia. MHSWs (mental health social workers) there focus mainly on sorting out benefits, though some are based in rehabilitation focused facilities where they work on connecting users to educational and employment opportunities.[10]

This fundamental change has led to the relocation of MHSWs away from institutions into community services,[5] to a renewed interest in rehabilitation, and more recently also in the newly defined recovery.[11]

The paralleled development of private, for-profit mental health services especially—but not exclusively—in the United States, has led to a further shift in the location of MHSWs and their work focus. Most United States MHSWs are to be found today working as psychotherapists in private practice or in managed care residential units.[12] In the latter they work with users who have long-term mental illness to a per capita budget.

Although the not-for-profit sector has grown considerably with the focus on care in the community, MHSWs work there only in certain countries in which the public sector has either been reduced or never played the major part it does in the United Kingdom (e.g. the Netherlands, Hong Kong).

Underpinning values

Social work, including its mental health branch, is ethically governed by a set of values, which are expected to be universal and adhered to in everyday practice,[13] even though its implementation may prove at times to be problematic in terms of balancing care and control.

The values are derived from the liberal collectivist, humanistic, tradition of the twentieth century in which social work has developed.

The core values are social justice, respect for people who social workers meet at their most vulnerable state, readiness to help in a way which will enable the client to retain dignity, self-determination, and enhance their problem-solving abilities. Social workers are expected to take an active stance against any type of discrimination. Furthermore, social workers are committed to pursing a psychosocial approach in any type of their practice, and believe that most clients have the potential to grow and positively change.

Several elements stand out as central to MHSW:

1 The right to fail—this comes as part of the right to self-determination, in that social workers are aware that risk needs to be taken at times to enable people to grow and develop, or as a basic human right of making a mistake. When social workers take this right seriously, they are able to have a genuine discussion with clients as to the pros and cons of risk-taking, of learning from success as much as of learning from failure.[14,15]

2 The wish to take an active stance against discrimination applies to working well with clients who come from ethnic minorities, from sexual orientation minorities, and to combating stigma against mental illness in one's practice.

3 The adherence to a psychosocial approach entails ensuring that both the psychological and the social aspects of users' lives are attended to, an issue of importance in mental health where often biological aspects are attended to, but the psychosocial ones are not getting the same priority.[16]

Conceptual developments

Psychodynamic approaches

As outlined above, mental health social work originated within the psychodynamic fold, though social workers did not practice psychoanalysis as a work method.

Social workers have tended to select from the range of psychodynamic perspectives those theories, which were more focused on the ego, rather than on the id or the unconscious. The impact of ego psychology was/is in evidence in terms of understanding how people come to develop and maintain mental distress and mental

illness, the importance of family dynamics, and of attachment to significant others.[17, 18, 19, 20]

American social workers developed the crisis approach in its application to all areas of social work.[21] Based on Erickson's notion of the normal crisis every person goes through when moving from one stage of life to another, major life events may lead initially to adverse reactions. However, with professional support people can reorganize their reactions more constructively, reduce the duration of these reactions, be more ready for change at the point of crisis, and learn how to improve their coping strategies and emotional responses. The problem-solving approach also originated from the United States, developed by Perlman.[22] Although the psychodynamic understanding of relationships is in evidence in her work, she focused on the process of social work with individuals and families (casework) and the client–worker relationships, beginning with the presenting problem.

The identification of child abuse, especially child sexual abuse, and its implications for the mental health of children and adults in the 1970s and the 1980s led to refocusing on the psychodynamic approach among social workers in this area[23] at a time in which all other approaches have paid less attention to the impact of such abuse on mental health.

Learning theory applications in social work

Behavioural social work

Behavioural social work developed in the United States in the 1950s, and is a leading approach in relation to people with milder forms of mental illness and problems of living.[24] Its application within social work does not differ in any significant way from its application within psychiatry or psychology. In this sense it is not a social work approach. A number of influential texts appeared in the United Kingdom which demonstrated the research evidence pertaining to the effectiveness of the approach in a number of social work areas.[25]

Task-centred social work

This orientation takes further the crisis perspective and the lessons from learning theories and behaviour modification.

Reid and Epstein[26, 27], as well as Marsh and Doel,[28] proposed that people work better on their problems if focused on specific targets and if the problem-solving effort leads to success, however small. Research evidence demonstrated the usefulness of this approach to different aspects of social work, such as direct work with children and their parents, as well as with people suffering from mild mental distress symptoms.

The social dimension in mental health social work

Social workers and theorists interested in the social dimension began usually from the assumption that inequality in opportunities and in civic participation due to poverty may increase the rate of mental illness among poor people. This assumption follows Merton's classical matrix of the reactions to the gap between social goals and means, in which mental illness is a reaction of people who accept socially desirable goals, but withdraw from obtaining them after being frustrated in doing so, whilst at the same time not adopting antisocial means (as in criminal behaviour) or developing an alternative model of society (social rebels).

This strand of thinking was reinforced in the 1960s and 1970s by the application of Marxist thinking and the combined impact of the deviancy and anti-psychiatry orientations.[29] Discrimination on the basis of age, ethnicity, gender, or sexual orientation was added in the 1980s to the likely social factors which foster inequality.

Social workers accepted the logic presented by sociologists such as Goffman and Scheff[30, 31, 32] that the stigma attached to mental illness is largely irreversible, as it is accepted both by others and by the individual concerned who in turn internalizes his or her poor social status.

Interestingly, although accepting the enormity of the labelling process, social workers did not count themselves among the labellers.

The appeal of the anti-psychiatry approach for social workers related to acknowledging the price of labelling for the individual concerned, and the considerable shortcomings of a system focused on the psychiatric hospital and medication in which psychological and social factors were largely ignored.

Today the social perspective implies a greater focus on social inclusion, supporting users and carers-led initiatives, ensuring financial support side-by-side with the critique of the medicalized approach to mental health and illness, of modernity and post-modernity.[16]

A more recent strand of this approach is outlined in the critical social work, which applies a post-modern perspective to the analysis of where social work is, as well as the issues and dilemmas related to mental health social work. Bainbridge[33] highlights the need to focus on the social dimension and sociological understanding of mental ill health in social work, as well as on issues of power and empowerment.

(a) The social role valourization (SRV) and the strength approach

This approach was initially developed by psychologists in the field of learning difficulties.[34–37]

SRV accepts the deviancy approach up to the point at which the impact of labelling and segregation is said to be irreversible. Conversely, SRV is focused on reversing the devaluation of the disabled person and the group, while accepting that a disability exists. The devalued existence can be reversed by the combined impact of the following:

- enabling those who have been segregated to live in the community by providing them with the opportunities to do so and the support they require for this purpose
- enhancing the competencies of the disabled person
- changing their public image, in part by their positive presence in ordinary settings in the community
- upgrading the state of the physical settings in which a disabled group is treated, lives, and works
- changing the derogatory language used in both professional and lay circles in describing people with disabilities.

Its protagonists are critical of professional attitudes, knowledge, and skills, including those of social workers (see Wolfensberger).[35] Yet as a group social workers have within their repertoire more of the attitudes, knowledge, and skills required by this approach than any other mental health profession. Furthermore, SRV offers an

interesting and comprehensive combination of psychological and social dimensions; for an application to how it can work with the Nearest Relative in mental health.[38]

In the United States and Canada, but much less so in the United Kingdom, SRV came into prominence within social work through the **strengths** model of social work.[39] The model is unique in concentrating on the strengths the person and his or her environment possess, and how these could be harnessed to solve the specific problem and lead to an improvement in the person's quality of life. Coming together with a focus on following people's ambitions (as long as these are within socially acceptable norms), this orientation has led to useful and positive outcomes in care management.[40] A further development of the strengths model is the growing interest in focusing on enhancing resilience in mental health social work.[41]

Interestingly, the approach has been adopted by other mental health professionals in the field of employment without acknowledging the debt to social work.

Legally anchored MHSW

The social mandate of social work is anchored within legal and policy frameworks; this applies to MHSW too.

Securing benefits

In all countries social workers are gatekeepers to and advocates for securing benefits either in cash (e.g. disability allowance, Direct Payment agreement) or in kind (e.g. housing, clothing). They often have to make the claim in addition to the client, verify the claim as against eligibility criteria, secure supporting documents from other professionals, at times negotiate with other agencies (e.g. social security, health, education), and in a number of countries they are indeed located in social security services (e.g. Portugal, Israel).

While this work is considered to be routine, its importance in the life of poor people cannot be underestimated. The evidence highlights that most people with enduring mental illness are poor,[42] and that remaining poor is a strong counter-indication to becoming mentally healthy. Furthermore, the evidence related to Direct Payment in mental health,[43] which enables users to take the driving seat as to how they spend their budget in agreement with the local authority and mental health trust, illustrates the social inclusion and recovery value of such a scheme which clearly comes out of the strengths model.

Knowledge of the available resources and eligibility is needed for this type of work, as well as the ability to inform users and enable them to participate as deserving partners.

In a large number of countries this is the main social work task in mental health (e.g. Brazil, Greece, Italy, Ireland, Portugal, and Poland).

The approved social worker

This role illustrates an enhanced legal position for MHSWs; its fullest form is practised in the United Kingdom.

The approved social worker was developed in Britain to provide a complementary measure to the psychiatric perspective within the 1982 amendments to the 1959 Mental Health Act.

Social workers were seen as suitable professional figures who would represent the psychosocial angle in parallel to the psychiatric view in the following instances:

- assessing people when an application has been made for a compulsory admission to a psychiatric unit
- the follow-up to such an admission
- mental health review tribunals (established within the 1959 Mental Health Act)
- work with the Nearest Relative[38]
- coordinate the multi-disciplinary assessment, which needs to be carried out by a psychiatrist and a GP in addition to the social worker. This assessment has to be carried out within a specified limited period of time.

Each of these tasks calls for somewhat different knowledge and skills, as well as emphasis and use of a range of more generic skills.[44] In each task social workers are asked not to replicate the psychiatric assessment but to compliment it. For example, they have to look for the least restrictive alternative to the hospitalization before they can recommend a hospital admission, rather than diagnose mental illness. Social workers have an autonomous position as they can disagree with the views of the other professions. The role requires exercising more social control than care, a contested issue within social work.

Training to become an approved social worker requires 60 days of academic input and supervised practice initially, followed by 5 days refresher training annually. Individual social workers can take it up after 2 years of post-qualification work experience. This compares with 2 days training for general practitioners and 1 day for psychiatrists.

Most of the activities undertaken by MHSWs with adults since 1983 are related to meeting the requirements of this role. Existing evidence[45] highlights that in most cases ASWs are working to a good standard. The role also offered MHSW a higher status and pay, but came with the price tag of giving up most of their previous activities, such as family work, group and community work, and work with users who have minor mental illness. However, the proposed new English Mental Health Act[46] includes the introduction of AMHPs (Approved Mental Health Practitioners) who can come from any mental health discipline, but likely to be nursing because of the numerical dominance of nurses in the English mental health service. ASWs are unhappy at being dethroned of their unique legal role, arguing that nurses do not have the same background training for psychosocial understanding and intervention.

Workforce research into ASWs[47] has highlighted a recent steady decrease in the number of ASWs, a high number of workers approaching retirement age and low morale, factors likely to have played a part in the government's wish to introduce other professions to this role. There are currently 4500 ASWs, a minority in the total workforce of social workers in England which stands at 46 000.

While current ASWs and MHSWs will be able to be part of the AMHP workforce, this change may also enable them to reclaim some of their previous roles and activities.

Care management

In all English-speaking countries and a number of European countries (e.g. the Netherlands, Slovenia), MHSWs are also often engaged in one form or another of care management, which follow the specific laws and regulations of each country (e.g. the Care in the Community 1990 Act in England).

Care management (not be confused with managed care) is a form of coordinating the assessment, planning, and interventions with people who require long-term care, including in mental health. It is aimed at preventing fragmentation and duplication of professional input and services, as well as ensuring that services follow the user's needs, and not vice versa as was—and still is—the case all too often. While in the United Kingdom, psychiatrists are formally the nominated care manager since 1995, and CPNs are the professionals in more frequent contact with the users, in other countries such as Australia, New Zealand, and Canada, social workers are often the nominated care coordinator.

Care management can be practised in a variety of ways, ranging from a purely administrative orientation, through clinical care management, to one anchored within the strengths and recovery orientations.[40, 48, 49] The choice is often dictated by the managers of a local authority or a mental health trust.

Community treatment orders (CTOs)

CTOs constitute a third legally defined area in which MHSWs are engaged. They originated in the United States, and exist in a variety of formats in Canada, Australia, and the United Kingdom.[50–52] They represent a response to the escalation of concerns about risk avoidance in the field of mental health, closely related to the growing fears of risk raised by modernity and post-modernity,[54, 55] and reinforced more recently by fears of terrorism as a particularly threatening type of risk. It is indicative that this preoccupation is not shared between North and South Europe; it is much more prominent in the North.

Following a legal process, CTOs enable the nominated mental health professional to require a service user to adhere to specific restrictions, such as to live in a certain facility or to present themselves for interventions at specific locations. This measure has been introduced to ensure that users with long-term mental illness who lead a disorganized life (e.g. often do not comply with medication, live rough, misuse substances, mishandle money, misagreed appointments, and get into trouble with the law) will have a safety net, which structures their lives. There is some evidence of the effectiveness of CTOs,[48, 49] but it seems restricted to specific subgroups within the broader category of people with severe and enduring mental illness.

Social workers are often the nominated professionals responsible for the agreed plan and enforcement of the CTO. This puts them in a somewhat conflictual position regarding the desired focus on care in social work, as it is tilted more towards the control element.[56]

Non-legally anchored MHSW

The constraints on this type of work come not only from the primacy of legally sanctioned work, but also from working within welfare bureaucracies, for a private employer interested primarily in profit, or for an impoverished not-for-profit service.

Despite these constraints, we have examples of good and innovative MHSW practice in most countries, which include:

- Successful attachment to primary care was established as early as 1965 in London, with the social workers, their clients, and the general practitioners expressing satisfaction with this way of working. Nevertheless, this form was largely abandoned owing to the focus on statutory responsibilities.
- Social workers pioneered collective user involvement approach during the early 1980s[57] some years before it became fashionable in wider circles.
- Applying self-directed group and community work approach to working with mothers of abused children, empowering them to take control over their lives.[58]
- Initiating de-institutionalization in social care institutions in Slovenia.[59]
- Creating family support teams[60] where most staff members are social workers providing brief assessment and consultation to families of children with minor mental distress, and are often successful in preventing the need for referral to more expensive services.
- Establishing the Building Bridges project in which parents with mental health difficulties and their children are supported together as well as separately.[61]
- Creating The Faith Links project, a multi-faith project within an inpatient service in which users, volunteers, and social workers plan joint activities which follow users' spiritual wishes.[62]
- Supported education and employment schemes in mental health initiated by social workers.[63]

Conclusion

Mental health social work is a broad, rather than a rigorous, church. Since the 1980s social workers have gained in professional status by the introduction of the roles of the approved social worker (or licensed to carry out civil commitment in the American context), care co-ordinators, managers of managed care facilities, or psychotherapists. These gains have come at a price outlined in the text above.

Often the cost of closer collaboration within the multi-disciplinary framework has led to the risk of giving up the attempt to hold on to, and further develop, an alternative and complimentary perspective from that of psychiatrists, nurses, or psychologists, as well as raising doubts as to the uniqueness of MHSW.

The increased narrowness of the role is not simply the byproduct of the legal framework. It is also due to increased specialization within mental health on the one hand, and the effects of neo-liberal policies globally on public sector funding on the other hand.

The move to privately contracted work, either in managed care or in psychotherapy so apparent in the United States, is yet another outcome of neo-liberal policies which fragments MHSW. As a trend we are likely to see growing beyond the United States, the increased concentration of mental health social workers within the private sector does not bode well for a profession whose value base focuses on the need to protect the more vulnerable and stigmatized populations, and to provide the dual perspectives of psychosocial input.

Mainly due to governmental pressure related to fear of risk and its potential political fallout, the focus on working exclusively with people experiencing long-term severe mental illness has contributed to the increasing narrowness of the role of social workers in most First World countries. The paralleled withdrawal of social work involvement with people who have milder forms of mental distress within public sector and not-for-profit services, and its

increased availability only to those who can afford it, is a reflection of this situation.

The core qualities of belief, optimism, and caring of MHSWs identified in a cross-national research[64] coupled with the ability of MHSW to innovate as highlighted in this chapter, illustrate the optimistic scenario for positive change within this branch of social work. However, unless theory building and research aspects are given the importance they deserve within MHSW globally, including an inevitable critical dimension of the existing system, mental health social work is likely to be no more than a reflection of the developments in other professions. This will not only mean curtailing its autonomous potential, but also the impoverishment of the multi-disciplinary framework as a whole of a crucial dimension necessary for its comprehensive work, as exemplified in some recent work on the social aspects of MHSW.[16]

In addition, mental health social work will have to develop a much stronger policy making function, if it is to provide a more responsive, effective, and comprehensive service to users, relatives, and the communities in which these people live.

Further information

Norman, E. (ed.) (2000). *Resiliency enhancement: putting the strength perspective into social work practice*. Columbia University Press, New York.

Ramon, S. and Williams, J.E. (eds.) (2005). *Mental health at the crossroads: the promise of the psychosocial approach*. Ashgate Publishing, Aldershot.

Tew, J. (ed.) (2005). *Social perspectives of mental health*. Jessica Kingsley, London.

References

1. Dicks, H. (1970). *Fifty years of the Tavistock*. Tavistock Publications, London.
2. Timms, N. (1964). *Social casework*. Routledge and Kegan Paul, London.
3. Stuart, P.H. (1997). Community care and the origins of psychiatric social work. In *Social work in mental health: trends and issues* (ed. U. Aviram), pp. 25–37. Haworth Press, New York.
4. Goodwin, S. (1997). *Comparative mental health policy: from institutional to community care*. Sage, London.
5. Shera, W., Aviram, U., Healy, B., *et al.* (2002). Mental health systems reform: a multi country comparison. *Social Work in Health Care*, **35**, 547–75.
6. De Leonardis, O., Mauri, D., and Rotelli, F. (1986). *Deinstitutionalisation: a different path: the Italian mental health reform, health promotion*, Vol. **2**, pp. 151–65. WHO Cambridge University Press.
7. Ramon, S. (ed.) (1990). *Psychiatry in transition: British and Italian experiences*. Pluto Press, London.
8. Vasconcelos, E.M. (2005). Structural issues underpinning mental health care and psychosocial approaches in developing countries: the Brazilian case. In *Mental health at the crossroads: the promise of the psychosocial approach* (eds. S. Ramon and J.E. Williams), pp. 95–108. Ashgate Publishing, Aldershot.
9. Ramon, S. and Williams, J.E. (2005). *Mental health at the crossroads: the promise of the psychosocial approach*. Ashgate Publishing, Aldershot.
10. Fung Sheung Chee, B., Law Ka Sin, J., and Lee Yuk Yee, K. (2006). *Clubhouse model in an Asian culture, 5th International Conference on health and mental health social work*, December, Hong Kong.
11. Ramon, S., Healy, B., and Renouf, N., (2007). Recovery from mental illness as an emergent concept and practice in Australia and the UK. *International Journal of Social Psychiatry*, **53**, 108–22.
12. Cohen, G.A. (2003). Managed care and the evolving role of the clinical social worker in mental health. *Social Work*, **48**, 34–43.
13. Beckett, C. and Maynard, A. (2005). *Values and ethics in social work*. Sage, London.
14. McDermott, R. (ed.) (1975). *Self determination in social work*. Routledge and Kegan Paul, London.
15. Ramon, S. (2006). Risk avoidance and risk taking in mental health social work. In *Knowledge in mental health: reclaiming the social* (eds. L. Sapouna and P. Hermann), pp. 39–56. Nova Publications, New York.
16. Tew, J. (ed.) (2005). *Social perspectives of mental health*. Jessica Kingsley, London.
17. Parad, H.J. (ed.) (1958). *Ego psychology and dynamic casework*. American Family Services Association, New York.
18. Hutton, J.M. (ed.) (1977). *Short-term contracts in social work*. Routledge and Kegan Paul, London.
19. Yellowly, M. (1980). *Psychoanalysis and social work*. Van Nostrand, New York.
20. Howe, D. (1995). *Attachment theory for social work practice*. Macmillan, London.
21. Golan, N. (1978). *Treatment in crisis situations*. Free Press, New York.
22. Perlman, H. (1957). *Social casework—a problem-solving process*. Chigaco University Press, Chigaco, IL.
23. Perlberg, R. and Miller, A. (eds.) (1992). *Gender and power in families*. Routledge, London.
24. Gambrill, E.D. (1977). *Behaviour modification*. Jossey-Bass, San Francisco, CA.
25. Hudson, B. and Macdonald, E. (1986). *Behavioural social work: an introduction*. Macmillan, London.
26. Reid, W.J. and Epstein, L. (1972). *Task centered casework*. Columbia University Press, New York.
27. Jackson, V.H. (1996). Behavioral managed care: a social work perspective. *Behavioral Health Management*, 22–3.
28. Marsh, P. and Doel, M. (1993). *Task-centred social work*. Ashgate, Aldershot.
29. Leonard, P. and Corrigan, P. (1978). *The Marxist approach to social work*. Macmillan, London.
30. Goffman, I. (1961). *Asylums*. Penguin, Harmondsworth.
31. Laing, R.D. (1965). *The divided self: an existential study in sanity and madness*. Penguin, Harmondsworth.
32. Scheff, T. (ed.) (1975). *Labelling madness*. Prentice-Hall, Englewood Cliffs, NJ.
33. Bainbridge, L. (1999). Competing paradigms in mental health practice and education. In *Transforming social work practice: postmodern critical perspective* (eds. B. Pease and J. Fook), pp. 179–94. Routledge, London.
34. Nirje, B. (1969). The normalisation principle and its human management implications. In *Changing patterns in residential services for the mentally retarded* (eds. R. Kugel and W. Wolfensberger), pp. 255–87. President's Committee on Mental Retardation, Washington, DC.
35. Wolfensberger, W. (1983). Social role valorisation: a proposed new term for the principle of normalisation. *Journal of Mental Retardation*, **21**, 234–9.
36. Ramon, S. (ed.) (1991). *Beyond community care: normalisation and integration work*. Mind/Macmillan, Basingstoke.
37. Brandon, D. (1991). Implications of normalisation work for professional skills. In *Beyond community care: normalisation and integration work* (ed. S. Ramon), pp. 35–55. Mind/Macmillan, London.
38. Rapaport, J. (2005). The informal caring experience: issues and dilemmas. In *Mental health at the crossroads: the promise of the psychosocial approach* (eds. S. Ramon and J. Williams), pp. 155–70. Ashgate Publishing, Aldershot.
39. Saleebey, D. (ed.) (1992). *The strengths perspective in social work practice*. Longman, New York.

40. Rapp, C. (1998). *The strengths perspective of case management with persons suffering from severe mental illness*. Oxford University Press, Oxford.
41. Norman, E. (ed.) (2000). *Resiliency enhancement: putting the strengths approach into social work practice*. Columbia University Press, New York.
42. Pilgrim, D. and Rogers, A. (2003). *Mental health and inequality*. Palgrave, Macmillan, Basingstoke.
43. Glasby, J. and Lester, H. (2002). *Social work and direct payment*. Policy Press, Bristol.
44. Barnes, M., Bowl, R., and Fisher, M. (1990). *Sectioned: social services and the 1983 Mental Health Act*. Routledge, London.
45. Hatfield, B. and Robinshaw, P. (1994). The use of compulsory powers by approved social workers in five local authorities: some trends over two years. *Journal of Mental Health*, **3**, 339–50.
46. Rapaport, J. (2006). New roles in mental health: the creation of the approved mental health practitioner. *Journal of Integrated Care*, **14**, 37–46.
47. Huxley, P., Evans, S., Webber, M., *et al.* (2005). Staff shortages in the mental health workforce: the case of the disappearing social worker. *Health and Social Care in the Community*, **13**, 504–13.
48. Kanter, J. (1989). Clinical case management: definition, principles, components. *Hospital and Community Psychiatry*, **40**, 361–8.
49. Brandon, D., Atherton, K., and Brandon, A. (1996). *Handbook of care planning*. Positive Publications, London and ref. 26.
50. Hiday, V.A. and Scheid-cook, T.L. (1991). Outpatient commitment for "revolving door" patients compliance and treatment. *Journal of Nervous and Mental Disease*, **179**, 83–8.
51. Campbell, J., Brohpy, L., Healy, B., *et al.* (2006). International perspectives on the use of community treatment orders: implications of mental health social workers. *British Journal of Social Work*, **36**, 1101–18.
52 Canvin, K., Bartlett, A., and Pinfold, V. (2002). A "bittersweet pill to swallow": learning from mental health service users' responses to compulsory community care in England. *Health and Social Care in the Community*, **10**, 361–9.
53. Stanley, N. and Manthorpe, J. (2004). *The age of the inquiry*. Routledge, London.
54. Beck, A. (1992). *The risk society*. Sage, London.
55. Rose, N. (2002). *Powers of freedom: reframing political thought*. Cambridge University Press, Cambridge.
56. Thompson, P. (2003). Devils and deep blue seas: the social worker in-between. *Journal of Social Work Practice*, **17**, 35–47.
57. Hennelly, R. (1990). Mental health resource centres. In *Psychiatry in transition: British and Italian experiences* (ed. S. Ramon), pp. 208–18. Pluto Press, London.
58. Mullender, A. and Ward, D. (1991). *Self-directed groupwork*. Whiting and Birch, London.
59. Flaker, V., Cizely, M., Ferle, Z., *et al.* (2004). Special care homes: the vision: a project of the community of social institutions in Slovenia. *Journal of Social Work, Faculty of Social work*, University of Ljubljana, Ljubljana, Slovenia.
60. Debell, D. and Walker, S. (2003). *Norfolk family support teams: an evaluation of the first two years*. Anglia Ruskin University, Cambridge.
61. Diggins, M. (2000). Innovation as a professional way of life—the building bridges project for parents-users of mental health services and their children. In *A stakeholder approach to innovation in mental health services* (ed. S. Ramon), pp. 77–93. Pavilion, Brighton.
62. Jones, J. (2006). *The faith links project—equality and diversity in Brent, central and north west London*. Mental Health NHS Trust, London.
63. Mowbray, C.T., Collins, M.E., and Bellamy, C.D. (2005). Supported education for adults with psychiatric disabilities: an innovation for social work and psychosocial rehabilitation practice. *Social Work*, **50**, 7–20.
64. Ryan, M., Merighi, J.R., Healy, B., *et al.* (2004). Belief, optimism and caring: findings from across-national study of expertise in mental health social work. *Qualitative Social Work*, **3**, 411–29.

6.4.4 Art therapy

Diane Waller

The fundamental principles of art therapy/art psychotherapy

Definitions

Descriptions of art therapy from two of the oldest and largest professional associations, the British Association of Art Therapists and the American Art Therapy Association refer to: the use of art materials for self-expression and reflection in the presence of a trained art therapist. Art therapy uses the flexible, creative problem-solving potential of art-making to improve and enhance the physical, mental, and emotional well-being of individuals of all ages. The relationship between the therapist, client, and their artwork is of central importance. Art therapy can be used on a one-to-one and group basis.

Art therapy (or art psychotherapy, both titles are protected by law) in the United Kingdom is firmly rooted in psychodynamic and humanistic concepts and practices appropriate to public sector settings, and adapted to the social and mental health of the client. It is a broad-based discipline, involving substantial knowledge of the visual arts, individual and group psychotherapy, social and communication sciences, and the impact of culture on health.[1]

Main premises

- That visual image-making is an important aspect of the human learning process;
- That art made in the presence of an art therapist may enable a person to get in touch with feelings that cannot easily be expressed in words;
- That the creative process helps people to resolve conflicts and problems;
- That art can act as a 'container' for powerful emotions and be a means of communication between client(s) and therapist;
- That the image can serve to illuminate the transference in the case of a psychodynamic approach.

Engagement in image-making is of central importance although clients do not need any prior experience of or skill in art, as the aim is not to produce a 'good' piece of art that can be exhibited. The images made in art therapy may embody thoughts and feelings, be a bridge between the 'inner world' and outer reality, be a mediator between unconscious and conscious, hold and symbolize past, present, and future aspects of a client's life. Ambivalence and conflict can be stated and contained within an image. In art therapy the client tries to give form to what seem to be inexpressible or unspeakable feelings, which they can then share with the art therapist.

The focus of the transference (bringing feelings from the past into the present), can be onto the art object rather than to the therapist directly, adding a 'third dimension' to the therapeutic process.[2–4]

An important aim, as with all psychotherapy, is to bring about change. Positive change may occur when a client can direct their strong feelings into making art and when the therapist helps the client to tell their story through the art. How, when, and if change occurs obviously depends on their capacity to engage with this process and needs much time and patience while the client builds confidence. For verbally inarticulate clients, or those who use words defensively, engagement with the art materials gives the opportunity to understand self and environment, communicate emotions to the therapist, receive feedback, and encouragement.

The historical development of art therapy in the United Kingdom, United States, and Europe

There are parallels in the development of art therapy in the United Kingdom and United States, early history being shared with that of group analytic psychotherapy as a phenomenon of the Second World War rehabilitation movement.[5–7] In the 1940s and 1950s art therapists were simply artists working in hospitals who emphasized the healing role of art. In 1963, the British Association of Art Therapists was formed from this small group of artists and art educators, who set themselves the task of defining and extending the activity, preparing standards for training in the higher education sector, informing the public and other professionals of the potential of art therapy, and working towards a career and salary structure in the National Health Service (NHS). The first *postgraduate trainings* began in the late 1960s. The positive response of the NHS and other organizations to art therapy's beneficial impact led to a petition being made for statutory regulation under the old Council for Professions Supplementary to Medicine in 1991, approved in 1997, after which art therapists, along with music and dramatherapists had their own federal Board at the Council. They were transferred to the Health Professions Council in 2001. Training in the United Kingdom is now at Master's level, in four universities in England, one in Scotland, one in Northern Ireland, and usually follows a degree in art and design. Study of psychotherapeutic principles, visual art, and practical placement are important elements in the training. Elsewhere in Europe the picture is very different with some countries sharing the UK standards, others having no training or a great variety of trainings in both the public and private sector. The United States, Australia, and New Zealand have the same requirements of a Master's level qualification in order to practice. (See website references for more information.)

The development of art therapy with specific client groups

One of art therapy's main advantages as a treatment is its flexibility. It can be used with many different client groups and some of these are discussed as follows:

Children

Many founder art therapists in the United Kingdom and United States were art teachers and were influenced by the 'child-centred' approach to art education that developed in the 1930s. American pioneer Kramer considered that it was art activity itself that had inherent healing properties; and that within a secure relationship with the therapist, a child could sublimate their destructive and aggressive feelings by producing an object, which would symbolize those feelings, prevent them being acted out and lead to more insight and control. This often led to change in behaviour.[8]

Others pioneers from the United States gave examples of how group work could enable angry and shameful feelings to be shared among the group members as well as the therapist, to the relief of the child as well as his peers.[9] Many art therapists specializing in work with children attest to the importance of play and to the role of art materials in allowing regression in the form of mess-making. This seems to be particularly beneficial for children who have suffered sexual abuse[10–12] due to the loosening of control that happens when a child becomes deeply immersed in the physical process of painting and is able to lower defences as a result. Materials may be smeared, spilled, and wasted and it is important that the therapist maintains control of the boundaries and is able to tolerate a high level of anxiety as the child attacks the therapeutic space.[13]

Art therapy is helpful for children suffering from chronic constipation, faecal overflow soiling, and also 'antisocial behaviour' and Aldridge[14] pointed out the relationship she observed between food, painting, and faeces while working with neglected and abused children in the context of a social services Unit and how mess-making was important in their creative development. Ambridge[15] discussed how images may be used to reflect mother–child relationships with children who have been sexually abused and are often so traumatized that they cannot speak about their experiences.

The physical involvement in the art materials in enabling regression and essentially in receiving containment and acceptance from the therapist is very important to all the children mentioned above.

Dubowski[16,17] used a Developmental Art Therapy approach in research with children with learning difficulties, aiming to help the child to achieve his or her maximum potential. Understanding creativity and mark-making in early childhood is as important in this model as understanding psychodynamics. Studies made by Kellog[18] of over 100 000 children's scribbles inform our understanding of the developmental process leading to production of meaningful marks. Visual problem-solving through picture-making is developed between the age of about 18 months (when hand-eye co-ordination has developed to the extent that they can grasp an implement and direct it to a picture surface while attending to the activity) and 4 years, by which time most have developed the capacity to make recognizable pictures endowed with symbolic meaning.[19] This model draws on insights from art educationists, most recently Matthews.[20,21] Art therapists have also contributed to the emotional and educational development of children with Autism.[22,23]

Art therapists also occasionally work with families and this is an emerging area of interest.

People with learning difficulties

Stott and Males[24] were among the first British art therapists to write about their work in a large hospital with people who had

lived in institutions most of their lives. They suggested that art therapy offered a means of communication and of self-expression through which difficulties of life in an institution, such as loss of identity, could be eased. Their goals were: to find the art medium of most use to each resident, bearing in mind any physical handicaps; to set-up the art therapy sessions at regular times and in the same place; record and report the results of sessions; to enable maximum communication to take place. For some long-term residents, the art therapy studio became an 'oasis' in the desert of the hospital and they gained an identity as 'artist'.

Their work was continued by a generation of art therapists concerned about the effects of institutionalization on long-stay residents. Drawing on Gardner's work concerning the categories of personal and spatial intelligence[25] Rees set-up a qualitative research project using a detailed observational schedule with a group of women with severe learning difficulties in a single-sex locked ward, conducted over 3 years to investigate clients' use of physical space and its potential symbolic significance. Rees found that by relating to physical and spatial aspects of their environment, some clients discovered an effective way of maintaining some level of psychological and emotional integration. They were helped to manage their often overwhelming feelings and to develop a stronger identity in their, albeit, very restricted environment.[26] Strand used a group interactive art therapy model with a group of learning disabled clients whom she observed to be suffering from loss and despair as a result of their emotional needs being neglected.[27]

Now that people with learning difficulties mainly live in the community, art therapists are now able to assist clients in developing resources to manage their day-to-day activities, and to improve their quality of living—particularly social interaction—through engagement in creative activity, often in groups. Insights from earlier work on institutionalization and exclusion now inform art therapy with older residents in care homes.

Offenders

Liebmann used art therapy within probation services to address 'offending behaviour' directly[28] summarizing the benefits to offenders as follows: as a means of non-verbal communication, important for the high percentage of offenders who have poor verbal skills, conversely with those who use words defensively; to release angry and aggressive as well as shameful and embarrassed feelings and provide an acceptable way of looking at and dealing with difficult emotions; client and therapist together could look back at the images over a series of sessions, see patterns and note developments; active participation is required, helping to mobilize those who may not be voluntary clients and bring about favourable behavioural changes that outlast the session itself. Liebmann devised strategies to help offenders gain insight into their behaviour, for example, the comic strip where the client is asked to draw an important life event within frames, the sequences of which can then be discussed and alternative options suggested.[29] This approach combines some elements of cognitive behavioural therapy with a psychodynamic approach, providing a structure, which is empathetic but on the other hand does not collude with offending behaviour, nor accept it as inevitable.

Teasdale produced a set of Guidelines while working in the prison service. These focus on inmates and on the prison environment, reinforce many of the points above, such as using art therapy to support prisoners in coping with their imprisonment, address feelings of separation, isolation, loss, low self-esteem. Specific advice is offered for conducting art therapy within a prison context.[30]

People with psychotic illness

From the 1940s art therapists have worked with long-stay clients in psychiatric hospitals modifying the effects of institutionalization through detailed attention to communication when even the most psychotic patients could communicate through images and have this acknowledged. Now that people with psychosis increasingly live in the community and cope with the challenges of everyday life the focus is on issues of independence, isolation, building relationships, managing the illness, and its impact on self and family.[31]

Work with acutely psychotic clients, on the other hand, takes place on wards, is normally brief and aimed at helping the client to interact positively with others usually in 'open' studio groups with a rapidly changing population, or for those in a serious acute state, on a one-to-one basis. Drawing on process-oriented psychology McClelland[32] has devised a new model using art therapy to work directly on the acute state itself, requiring an active and assertive therapist style and meeting the client in their own 'language' however bizarre this may seem.

Older people

With older people's mental health and well-being coming under increased government scrutiny, this is an area where art therapy has much potential to be beneficial. Art therapists in this field have noted that attention to loss, fear of illness and dying, feelings of helplessness and dependency is necessary in alleviating depression. Recent art therapy research with older people with dementia showed some improvements in mental acuity, calmness, sociability, and physical competence following 40 weekly sessions of group work, compared to no change in the control.[33]

As with other client groups, engaging in creative activity can also provide an outlet for frustrations to do with ageing, as well as possibly leading to a challenging and rewarding hobby.

Physical illness

Art therapy is used with people suffering from cancer, including terminal cancer[34,35] as well as with other long-term and progressive illnesses or conditions such as multiple sclerosis, ME, kidney disease, effects of stroke, Parkinson's disease, and in palliative care to manage anxiety and fear about death and dying. The aims are to assist clients in coping with the emotional impact of their illness, enabling expression of feeling, relieving depression and stress, and improving their quality of living.

Other areas where promising work is going on: with eating disorders, drug and alcohol addiction, and with refugees and asylum seekers, many of whom have been the victims of war, of torture and are suffering extreme stress; also with the moderately to severely depressed and those who have work and relationship difficulties.

Contextual issues

In the United Kingdom, over 50 per cent of art therapists work within the NHS where they may form part of the psychological therapies or occupational therapy department, or be autonomous. Others practise in Social Services, Education, Home Office,

non-Statutory services, and a fairly small percentage work privately or are self-employed.

All are capable of assessing the suitability of the client for art therapy. Exclusion criteria are minimal. Initial interviews usually explain that the client does not have to be 'good at art' and that art therapy is not a 'painting class' but that it may arouse strong and sometimes difficult feelings. All therapists are trained to work with individuals and groups, to function as members of multi-disciplinary teams, to manage health and safety issues concerning preparation and maintenance of the art therapy space, liaising regularly with medical, nursing, teaching, or other appropriate staff in the interests of the client. Ethical standards are laid down by the Health Professions Council as for the other arts therapists (drama and music and also dance) with whom there is regular contact.

Research

There is a substantial body of qualitative research in art therapy, however much of this would not meet the criteria for evidence-based practice required by today's public sector. Currently there is very little about art therapy within National Institute of Clinical Excellence guidelines, which have tended to emphasize quantitative studies. This is unfortunate as it gives the impression of a profession without a strong evidence base, which is not so as the flourishing Art Therapy Practice Research Network and a growing body of literature demonstrates.[36,37]

A few books and papers emerging from control group studies feature evaluations of art therapy groups with older people with moderate to severe dementia[38] and with schizophrenia. In 2006, the UK Health Technology Assessment supported a consortium headed by Crawford, Killaspy, and Waller (University of London) for a multi-centre random control group study of art therapy and schizophrenia. The Master's level training in art therapy requires a substantial dissertation, and a significant percentage of art therapists continue to doctorate research designed to use and test the hypotheses emerging from over 60 years of detailed casework.

Further information

British Association of Art Therapists: www.baat.org.

Health Professions Council: www.hpc-uk.org (Art, Drama and Music Therapy).

International Society for the Study of the Psychopathology of Expression and Art Therapy: http//www.online-art-therapy.com/

American Association of Art Therapists: www.arttherapy.org.

References

1. Quality Assurance Agency. (2004). *Benchmark statements for arts therapies*, www.qaa.ac.uk.
2. Dalley, T. (ed.) (1987). *Art as therapy*, pp. 6–19. Tavistock, London.
3. Schaverien, J. (1987). The scapegoat and the talisman: transference in art therapy. In *Images of art therapy* (eds. T. Dalley, *et al.*), pp. 74–108. Tavistock, London.
4. Schaverien, J. (1992). *The revealing image: analytical art therapy in theory and practice*. Routledge, London.
5. Waller, D. (1991). *Becoming a profession: the history of art therapy in Britain*. Routledge, London.
6. Waller, D. (2004). *Art therapists: pragmatic rebels*. Goldsmiths College, London (Inaugural Lecture May 2001).
7. Hogan, S. (2001). *The healing arts: the history of art therapy*. Jessica Kingsley, London.
8. Kramer, E. (1971). *Art therapy with children*. Schocken Books, New York.
9. Rubin, J. (1978). *Child art therapy*. Van Nostrand Reinhold, New York.
10. Lee Drucker, K. (2001). Why can't she control herself? Case study. In *Art therapy with young survivors of sexual abuse* (ed. J. Murphy), pp. 101–25. Brunner-Routledge, Hove and New York.
11. Lillitos, A. (1990). Control, uncontrol, order and chaos: working with children with intestinal motility problems. In *Working with children in art therapy* (ed. C. Case), pp. 72–88. Routledge, London.
12. Sagar, C. (1990). Working with cases of child sexual abuse. In *Working with children in art therapy* (ed. C. Case), pp. 89–114. Routledge, London.
13. Waller, D. (2006). Art therapy and Children: how it leads to change. *Clinical Child Psychology and Psychiatry*, **11**, 271–82.
14. Aldridge, F. (1998). Chocolate or shit: aesthetics and cultural poverty in art therapy with children. *Inscape*, **3**, 2–9.
15. Ambridge, M. (2001). Using the reflective image within the mother-child relationship. In *Art therapy with young survivors of sexual abuse* (ed. J. Murphy), pp. 69–85. Brunner-Routledge, Hove.
16. Dubowski, J. (1984). Alternative models for describing the development from scribble to representation in children's graphic work. In *Art as therapy* (ed. T. Dalley), pp. 45–61. Tavistock, London.
17. Dubowski, J. (1989). Art versus language (separate development during childhood). In *Working with children in art therapy* (ed. C. Case), pp. 7–22. Tavistock/Routledge, London.
18. Kellog, R. (1970). *Analysing children's art*. National Press Books, California.
19. Dubowski, J. and James, J. (1998). Arts therapies with children with learning difficulties. In *Development and diversity: new applications in art therapy* (ed. D. Sandle), pp. 41–56. Free Association Books, London and New York.
20. Matthews, J. (1999). *The art of childhood and adolescence: the construction of meaning*. Falmer Press, London.
21. Matthews, J. (2003). *Drawing and painting: children and visual representation*. Paul Chapman, London.
22. Evans, K. and Rutten-Saris, M. (1998). Shaping vitality affects: enriching communication: art therapy for children with autism. In *Development and diversity: new applications in art therapy* (ed. D. Sandle), pp. 57–77. Free Association Books, London and New York.
23. Tipple, R. (2003). The interpretation of children's art work in a paediatric disability setting. *Inscape*, **8**, 48–59.
24. Stott, J. and Males, B. (1984). Art therapy for people who are mentally handicapped. In *Art as therapy* (ed. T. Dalley), pp. 111–26. Tavistock, London.
25. Gardner, H. (1984). *Frames of mind: the theory of multiple intelligences*. Paladin, London.
26. Rees, M. (1995). Making sense of marking space: researching art therapy with people who have severe learning difficulties. In *Art and music therapy and research* (eds. A. Gilroy and C. Lee), pp. 117–37. Routledge, London and New York.
27. Strand, S. (1990). Counteracting isolation: group art therapy for people with learning difficulties. *Group Analysis*, **23**, 255–63.
28. Liebmann, M. (ed.) (1994). *Art therapy with offenders*. Jessica Kingsley, London.
29. Liebmann, M. (1998). Art therapy with offenders on probation. In *Development and diversity: new applications in art therapy* (ed. D. Sandle), pp. 104–20. Free Association Book, London.
30. Teasdale, C. (2002). *Guidelines for arts therapists working in prisons*. Department for Education and Skills/HM Prison Service, Prisoners' Learning and Skills Unit.
31. Killick, K. and Schaverien, J. (1997). *Art psychotherapy and psychosis*. Routledge, London and New York.

32. McClelland, S. (1992). Brief art therapy in acute stages: a process-oriented approach. In *Art therapy: a handbook* (eds. D. Waller and A. Gilroy), pp. 189–208. Open University Press, Buckingham.
33. Rusted, J., Sheppard, L., and Waller, D. (2006). A multi-centre randomized control group trial on the use of art therapy for older people with dementia. *Group Analysis*, **39**, 517–36.
34. Pratt, M. and Wood, M. (eds.) (1998). *Art therapy in palliative care*. Routledge, London and New York.
35. Waller, D. and Sibbett, C. (eds.) (2005). *Art therapy and cancer care*. McGraw-Hill, Maidenhead.
36. Karkou, V. and Sanderson, P. (2006). *Arts therapies: a research based map of the field*. Elsever, London.
37. Gilroy, A. (2006). *Art therapy, research and evidence based practice*. Sage, London.
38. Waller, D. (ed.) (2002). *Arts therapies and progressive illness*. Brunner-Routledge, Hove and New York.

6.5

Indigenous, folk healing practices

Wen-Shing Tseng

What are indigenous, folk healing practices?

Indigenous, folk healing practices are nonorthodox therapeutic practices based on indigenous cultural traditions, operating outside of official (modern) healthcare systems.[1] These practices are often validated by experience, but are not founded on scientific principles. Indigenous healing practices are observed in 'primitive' or 'pre-industrialized' societies as well as in modern or developed societies. All healing practices or psychotherapies are more or less culturally influenced, including modern and orthodox psychotherapies, but indigenous healing practices are described as 'culturally embedded' because they are often intensely embedded in the cultural systems in which they were invented and in which they are practised. They are, therefore, usually very difficult to transplant to entirely different cultural settings, where they do not have the same meaning or legitimacy.[2]

While indigenous healing practices function in general as healing methods for problems, they are not usually considered by either the healer or the clients to be psychological therapy for the clients' emotional or psychological problems. Rather, they are recognized as religious ceremonies or healing exercises related to supernatural or natural powers. However, from a mental health point of view, the indigenous healing practices often provide psychotherapeutic effects for the clients, and can be considered as folk psychotherapy.

Anthropologists have studied folk healing practices as a part of cultural behaviour. Recently, cultural psychiatrists have become interested in examining indigenous healing practices from clinical perspectives to explore the similarities and differences that exist between folk healing practices and modern psychotherapy, and to disclose the therapeutic mechanisms that are operating in and being utilized by indigenous healing practices. Many people in developed societies utilize folk healing practices as adjunctive to their primary (modern) therapy or as their main way to get help. Therefore, it is relevant for the modern psychiatrists to know what they are and the possible therapeutic mechanism they offer, or the possible negative effects they may receive by utilizing such indigenous healing practices.

Various practices are covered by the loosely defined terms, indigenous or folk healing. Religious healing practices and ceremonies are closely related to a specific religion. Shamanism involves a spirit medium. Divination, or various kinds of fortune-telling, including astrology or physiognomy, may be used by people to solve their psychological problems or to seek answers for life problems, and, therefore, can be viewed as folk counselling practices as well. Furthermore, the practice of meditation, a self-training exercise used to obtain tranquility, growth of mind, and prevention of emotional problems, can be considered a folk healing practice if one defines psychotherapy very broadly, as not only treating a suffering person but also providing a means for preventing problems and improving the quality of a person's mental life.[3]

No matter what terms are used, indigenous healing practices share some common features. They are invented and utilized by local people for the purpose of solving problems or treating suffering—therefore, they are called indigenous in contrast to universal. They are distinctly different from the modern (Western or orthodox) professional medical approaches—thus, they are called folk practices. Most of them are supernaturally oriented and remote from any scientific orientation. Such indigenous practices are usually rooted in traditional beliefs and folk interpretations of problems, and, thus, are closely related to cultural beliefs.

Subdivision of various healing practices

Based on their core nature and their basic therapeutic orientation, healing practices observed in different societies can be subdivided into different categories namely, supernatural orientation (such as spirit mediumship, religious healing ceremony, and divination); nature orientation (such as fortune-telling, astrology, and meditation); medical-physiological orientation (such as mesmerism, acupuncture, and herb medicine); and socio-psychological orientation (such as Zen training, Alcoholics Anonymous, est, and most modern psychotherapy).[4,5] It is recognized that such subdivisions are arbitrary, and often overlap. Yet these subcategories will help us to understand various healing practices that exist on a spectrum which includes the supernatural, natural, physiological, and psychological.

Spirit mediumship (trance-based healing system)

Spirit mediumship broadly refers to a situation in which the healer or the client, or both, experiences alternate states of consciousness in the form of dissociation or a possessed state at the time of the healing ritual. From a psychotherapeutic point of view, it is

important to distinguish which person is in an alternative state of consciousness, as the mechanism of therapy differs depending on whether it is the healer or the client who is dissociating.

(a) Shamanism

It is speculated that the geographic heartland of shamanism is Central and North Eurasia, with widespread diffusion to Southeast Asia and the Americas.[3] Through a religious ceremony, a shaman can work himself into a trance state in which he is possessed by a god. The rhythmic singing, dancing, or praying (quiet meditation) seems to assist the self-induction of the trance state. Among native healers in North and South America, a psychedelic substance (such as may be found in cactus) is frequently used to induce an altered state of consciousness and a special psychic experience for the healing performance. Whether the altered state of consciousness is substance- or self-induced, the healer is considered to be possessed by a supernatural power. The client can then consult the supernatural through the shaman for instructions on dealing with his or her problems.

The causes of problems are usually interpreted according to the folk concepts held by the culture—involving such things as loss of the soul, sorcery, spirit intrusion, or violation of taboos. Disharmony with nature may also be interpreted as the cause of problems. Coping methods are usually magical in nature, such as: prayer, the use of charms, or the performance of a ritual ceremony for extraction or exorcism. Utilizing supernatural powers, acting as an authority figure, making suggestions, and providing hope are some of the main mechanisms for healing provided by the shaman. The goal of the healing practice is to resolve the problems that a client is encountering.

(b) Zar ceremonies

The term *zar* refers to a ceremony as well as a class of spirits. *Zar* ritual is observed primarily in Muslim societies in the Mideast, including Ethiopia, Egypt, Iraq, Kuwait, Sudan, and Somaliland. *Zar* ritual is different from shamanism in that, in addition to the healer, the client also experiences the dissociated or possessed state.

The *zar* ceremony is primarily a female activity. All of those attending the ceremony wear new or clean clothing to please the spirits. The main patient usually wears a white gown, as much gold jewelry as possible, and is heavily perfumed. The ceremony master begins the ceremony with a song and drumming. When a spirit associated with some person in the audience is called, that person begins to shake in her seat, dancing, and trembling until she falls, exhausted, to the floor. Before the spirit consents to leave, it usually demands special favours, such as jewellery, new clothing, or expensive foods. It is the duty of the relatives and friends to gather around the prostrate woman and pacify the spirit. The whole tone of the ceremony is one of propitiation and persuasion, rather than coercion. The ceremony ends with an animal sacrifice and a feast.

The *zar* ceremony is primarily an adult female activity reflecting social conditions of sex-separation, low female status, restriction of women from religious participation, an unbalanced sex ratio, marital insecurity, and relative isolation. The *zar* ceremony provides women an ideal situation for relief of persistent and regular anxieties and tensions arising from their life conditions. The goods demanded during the ceremony are all things that their husbands should provide. This fulfills a woman's wish for attention and care. Emotional catharsis, fulfillment of unsatisfied desire, and compensation for the suppressed female role are some of the therapeutic mechanisms working in this kind of therapeutic ritual. Restoring balance in real life is the implicit goal of this culture-embedded healing practice.[6]

Religious healing ceremonies

A distinction needs to be made between religion and a religious healing ceremony. Religion refers to a system of belief in a divine or superhuman power or spiritual practice. As a part of a religion, some people may perform special ceremonies for the purpose of healing certain problems or disorders. There are various kinds of religious healing ceremonies observed in different societies that are considered by mental health workers to serve a therapeutic function for their participants.

(a) Sprit dancing ceremony

As observed among Salish-speaking Indians of the Pacific Coast of North America, healing ceremonies utilize psychological mechanisms and processes similar to those in brainwashing. The initiate has to go through three major therapeutic approaches: depatterning through shock treatment (such as physical restraint, blindfolding, hitting, kinetic stimulation, or intensive acoustic stimulation, followed by lying still, being forbidden to talk, and starvation); physical training (such as daily running, jumping into ice-cold waters, or frequent rounds of dancing); and, finally, indoctrination.[7]

(b) Sacrificial ritual

An example is found in Yoruba, Africa. A person's problems were identified by palm nuts tossed by the diviner. It was usually interpreted that the person or a member of his family had offended the family *orisa* (the lineage deity) or some other spirit. Then, the sacrifice of a certain animal was prescribed for resolution. In the sacrifice, the supplicant passed his bad luck or illness to the animal, and the animal was killed in the supplicant's stead. The healing power of the ritual lies in its reassurance and generation of conviction. The ritual demonstrates that proper curative steps are being taken.[8]

(c) The religious ceremony of mourning

This is practised among members of the Spiritual Baptist Church in the West Indies. In a desire for spiritual strength and other benefits, church members volunteer to participate in the practice. After ceremonial washing and anointing, the mourners are isolated in a small chamber at the back of the church, where they remain for a period of 7 days. During that time, each individual prays, fasts, and experiences dreams and visions. The mourners claim they obtained beneficial psychological relief on their moods; attainment of the ability to foresee and avoid danger; improvement in their decision-making abilities; cures for physical illnesses; and heightened facility to communicate with God.[9]

(d) Snake-handling cult

An extremely different form of religious ritual was a snake-handling cult in the southern United States. As part of cult activities, members, in trance states, handled poisonous snakes as a sign of being blessed by God. Occasionally, some of the members died when they were bitten by the snakes. Although forbidden by the government to perform such cult rituals, these activities still exists. The gratification of emotional excitement was interpreted

as one of the effects sought by cult members—even at the risk of their lives.[10]

(e) Christian religious healing

It is important to know that religious healing ceremonies are not only observed in primitive societies or among uncivilized populations, but are quite common in many industrialized societies, as well. In Christian religious healing, there is a broad spectrum of beliefs and activities, ranging from Christian Science to the fundamentalism of healers such as Oral Roberts to the Roman Catholic rite of anointing of the sick. The participating client is provided with the hope for supernatural resources against disease, thus increasing his or her security and sense of well-being.[11]

Thus, in various forms of religious healing ceremonies, the therapeutic operation is carried out through the ritual of prayer, testimony, sacrifice, reliving experience, or even spirit possession. Assurance, suggestions, and generation of conviction are some of the healing mechanisms utilized in the practices. The aims of therapy are to heal the problems and give a certain perspective to the client's life.

Divination

Divination refers to the act or practice of trying to foretell the future or the unknown by occult means. It relies on mysterious, magic, or religious methods. Since the interpretation of divine instruction is usually provided by the diviner himself, or an interpreter, the interaction between the diviner/interpreter and the client becomes an important variable.

There is a range of methods of divination. Some methods are very simple, while others are more complicated. For example, in Nigeria, Africa, the divination practised by the Nsukka Ibo (called *Afa*) is carried out by casting four strings containing half-shells of the seeds of the bush mango;[12] and by the people in Yoruba (known as *Ifa*) by tossing palm nuts.[8] In the divination practised in some parts of Africa, the diviner simply offers a certain sign himself. For example, divination may occur while his hand is shaking, with the belief that he is guided by a supernatural power to give instructions.

In ancient China, turtle shells or the bones of big animals were burned during divination ceremonies and divine instruction was interpreted through the cracks made from the heat. An elaborate divination system called *chien* has been developed in China, and a modified version is used in Japan. To obtain answers to questions about their lives, some Chinese or Japanese will visit temples for divination. After a sincere prayer to the god of the temple, the person will ask for divine instruction, which is provided through a fortune stick that the person selects. Corresponding to the number on the stick, there is a fortune paper with an answer written on it. This practice is called *chien* drawing in Chinese,[13] or *kujibiki* in Japanese.

No matter what method of divination is practised , the basic therapeutic operation is performed to provide a clear-cut answer for the problems presented. Thus, it is helpful psychologically for a client to find a definite way to address his problems. Naming effects are among the healing mechanisms operating in divination. It is assumed that human life is under the influence of supernatural regulation. It is the basic goal of the person seeking help to find the proper way to comply with the universe through divine instruction.

(a) Fortune-telling

The system of reference shifts from the supernatural to the natural in the practice of fortune-telling. Based on the concepts of microcosm and macrocosm, fortune-telling is oriented to the basic belief that human life and behaviour are parts of the universe. The nature of the problems is usually explained in terms of an imbalance of vital forces or disharmony with the natural principles that rule the universe. The objective of the practice is to help the client find out how to live compatibly with nature and adjust to the environment more harmoniously.

Based on the sources of information used, fortune-telling can be divided into several groups. In astrology, there is a basic belief that a person's life is correlated to and influenced by the movement of the stars, thus, their movement becomes the essential source of information for predicting one's life course. For the Chinese, an ancient record of universal change, the Oracle of Change (*Yi-Jing*), is used for fortune-telling. A person's date and time of birth, and the number of strokes in the Chinese character for his or her name, is the information needed to calculate an individual's fortune.

Physiognomy is based on the assumption that there is a close correlation between the mind and the body and that one's character, life, and fortune can be read by examining one's physical features. It is assumed that a person is born with a certain predisposition, which is shown in his physical appearance and will lead him to manifest certain behaviour patterns. A physiognomist tries to help a client understand his own character and behaviour patterns, learning how to make good use of his talents and, at the same time, make up for his shortcomings.

Although the basic assumption underlying fortune-telling is that every person has a predetermined course of life, such fate is not absolutely unchangeable—it may be subject to modification. Thus, it is not a completely passive acceptance of fate, but allows room for adjustment. Finding a way to adjust your own fortune is the purpose of fortune-telling.

Even though the basic orientation shifts from a supernatural to a natural one, and the sources of practice rely on the rules of nature, the therapeutic operation, like divination, is still characterized by offering folk-natured interpretation and providing concrete guidance for a client in making choices. Based on the concepts of microcosm and macrocosm, complying with the fundamental rules of nature is the basic goal of the practice.

Common therapeutic factors

Reviewing various forms of folk therapy, it has been pointed out that the core of the effectiveness of different methods of religious and magical healing seems to lie in their ability to arouse hope by capitalizing on the patient's dependency on others.[14] Comparing the healing practices carried out by witch doctors and psychiatrists, it has been pointed out that they share a common root. Both kinds of therapists are able to decrease the client's anxiety by identifying what is wrong with him—that is, to name the cause of the problems, providing the effect of the Rumpelstiltskin principle; the therapist presents certain personal qualities that are admired by the culture and contribute to the therapy; the client's expectations of therapy, and the emotional arousal that is usually enhanced by the therapeutic setting, the therapist's belief in himself, and his reputation; the emerging sense of learning and mastery that the

client obtains through therapy; and finally, the techniques of therapy that enhance the basic components of psychotherapy. It is clear that folk healing practices and modern psychotherapy share a number of nonspecific therapeutic mechanisms.[15] It has been indicated that traditional healing practices have several advantages over cosmopolitan modern medicine, namely: cultural congeniality, maximal use of the personality of the healer, a holistic approach, accessibility and availability (particularly for developing areas), effective use of affect and altered states of consciousness, collective therapy management, and cost-effectiveness.[1] It is important to recognize that both folk healing and modern therapy utilize symbols and metaphors for interpretation and suggestions.[16] In contrast to modern psychotherapists, some folk healers make use of symbolic interpretations and suggestions to enhance the effects of healing.

Attitudes towards indigenous healing practices

Different points of view exist among scholars, clinicians, and public health workers regarding whether or not to encourage or discourage indigenous folk healing practices in various societies. Some people (particularly modern clinicians) see folk healing as merely superstitious and primitive, insisting that such out-of-date practices should be discouraged or prohibited. Others (such as cultural anthropologists and cultural psychiatrists) consider these folk practices to be interesting subjects for academic study—examining the therapeutic elements that are utilized in these primitive healing practices, and why such supernaturally oriented therapeutic exercises are still popular among some groups. Still other people (such as some community health workers) believe that, due to the shortage of professional personnel available in the community, the existence of folk therapies should be supported. The position was taken that any folk healing practice that is proven (or at least considered) to be helpful to the client and useful to the community deserves the support and encouragement of clinicians as well as administrators.

Final comments: clinical implications

The comparative study of indigenous healing practices and modern psychotherapy has revealed the existence of certain universal elements of the healing process that operate as important factors for therapy, whether the therapy is carried out in a primitive or modern form. The universal and nonspecific healing factors identified are: the cultivation of hope, the activation of surrounding support, and the enhancement of culturally sanctioned coping. The study of indigenous healing practices has also pointed out the existence of supernatural dimensions of healing power, which are less intentionally utilized in modern therapy.

Despite the general usefulness of folk therapies, the ill effects of some have not been widely studied and reported. Yet, clinical observation has disclosed that some folk therapists cause harm to the clients who seek their services. Under the guise of treatment, tricking a client out of his money by deceit or fraud, or sexual involvement with a client, are examples of disreputable behaviour that are occasionally reported. Harming a client by prescribing dangerous substances, and physically injuring or even killing a client by accident during the performance of an exorcism, are other examples of serious complications that have occurred.

No matter what position is taken, there is one simple fact that deserves attention, namely, that there exists a wide range of professional quality among so-called folk healers, and different motivations for practice. Some are benign healers motivated by a desire to serve, while others are not. Some are well-trained in their particular professions and know how to practice within its limitations, while others are not—and are liable for malpractice. The major problem is that, from a public health point of view, in most societies, there still are no formal guidelines for regulating folk therapy, as there are for modern therapy. Folk therapy, whether it is shamanistic practice or faith healing, should be subject to periodic surveys and reevaluation by the public health administration, as is modern clinical work, so that its benefits to clients can be protected and any potential malpractice can be prevented. If any folk therapist refuses to be examined and regulated, he or she should be discouraged or prevented from practicing.

Further information

Tseng, W.S. (1999). Culture and psychotherapy: review and practical guidance. *Transcultural Psychiatry*, **36**(2), 131–79.

Tseng, W.S. (2003). 7: Culturally competent psychotherapy. In *Clinician's guide to cultural psychiatry* (ed. W.S. Tseng), pp. 291–342. Academic Press, San Diego.

References

1. Jilek, W.G. (1994). Traditional healing in the prevention and treatment of alcohol and drug abuse. *Transcultural Psychiatric Research Review*, **31**(3), 219–58.
2. Tseng, W.S. (2001). *Handbook of cultural psychiatry*, pp. 515–37. Academic Press, San Diego.
3. Prince, R. (1980). Variations in psychotherapy procedures. In *Handbook of cross-cultural psychology: psychopathology*, Vol. 6 (eds. H.C. Triandis and J.G. Draguns). Allyn and Bacon, Boston.
4. Tseng, W.S. and Hsu, J. (1979). Culture and psychotherapy. In *Perspectives on cross-cultural psychology* (eds. A.J. Marsella, R.G. Tharp, and T.J. Ciborowski). Academic Press, New York.
5. Tseng, W.S. (1999). Culture and psychotherapy: review and practical guidance. *Transcultural Psychiatry*, **36**(2), 131–79.
6. Kennedy, J.G. (1967). Nubian *zar* ceremonies as psychotherapy. *Human Organization*, **26**(4), 185–94.
7. Jilek, W.G. (1976). Brainwashing as therapeutic technique in contemporary Canadian Indian sprit dancing: a case in theory building. In *Anthropology and mental health: setting a new course* (ed. J. Westermeyer). Mounton Publishers, Paris.
8. Prince, R. (1975). Symbols and psychotherapy: the examples of Yoruba sacrificial ritual. *Journal of American Academy of Psychoanalysis*, **3**(3), 321–38.
9. Griffith, E.E.H. and Mahy, G.E. (1984). Psychological benefits of spiritual Baptist mourning. *The American Journal of Psychiatry*, **141**(6), 769–73.
10. La Barre, E.H. (1962). *They shall take up serpents: psychology of the southern snake-handling cult*. University of Minnesota Press, Minneapolis.
11. Hufford, D. (1977). Christian religious healing. *Journal of Operational Psychiatry*, **8**(2), 22–7.
12. Shelton, A.J. (1965). The meaning and method of *Afa* divination among the northern Nsukka Ibo. *American Anthropologist*, **67**, 1441–5.

13. Hsu, J. (1976). Counseling in the Chinese temple: a psychological study of divination by Chien drawing. In *Culture-bound syndromes, ethnopsychiatry, and alternate therapies* (ed. W.P. Lebra). University Press of Hawaii, Honolulu.
14. Frank, J.D. (1961). *Persuasion and healing: a comparative study of psychotherapy*. Schocken Books, New York.
15. Torrey, E.F. (1986). *Witchdoctors and psychiatrists: the common roots of psychotherapy and its future*. Harper & Row Publishers, New York.
16. Kirmayer, L.J. (1993). Healing and the invention of metaphor: the effectiveness of symbols revisited. *Culture, Medicine and Psychiatry*, **17**(2), 161–95.

SECTION 7

Social Psychiatry and Service Provision

This page intentionally left blank

7.1

Public policy and mental health

Matt Muijen and Andrew McCulloch

Introduction

Public policy, and specifically national public policy, is one of the key factors that affects the practice of psychiatry, the shape of mental health services, and the environment within which mental health services work. The specific content of public policy varies greatly across the world and often even across neighbouring countries. It is therefore impossible within the space of this chapter to undertake systematic international comparisons. This chapter gives an overview of:

(a) what policy is and why it might be important;

(b) types of policy and policy development internationally;

(c) international structures and organizations that are relevant to the scope and content of policy, especially in the field of human rights which is often the starting point for policy;

(d) the breadth of policy activity that is relevant to mental health—stretching beyond the health ministry and health policy—and the partnerships that are necessary to tackle the mental health of individuals and populations.

What is public policy?

Koontz and Weihrich[1] define policy or policies as 'General statements or understandings which guide thinking on decision making.' This definition implies that:

1 policy may be stated in writing (statute, guidance, or statement) or may be unwritten and disseminated through management or political chains of command only;

2 policies may exist at various organizational levels albeit here we are mainly concerned with national policies;

3 policy is only one factor in any final management or clinical decision. Other factors will often be more important, and clinical behaviour is notoriously resistant to policy influence except where certain behaviour is directly prohibited.[2]

Therefore, the policy needs to be understood as just one of the factors which determine the nature and state of mental health services and mental health promotion or public mental health within a society.

In many developed countries, the management of the public sector has seen major changes over the last two decades, shifting from a hierarchical and bureaucratic administration to a model of 'managerialism'.[3] In the old model policy making and administration was separated, and governments were responsible for the provision of services often via related public sector agencies. The new model, inspired by business,[4] is characterized by governments introducing competition and market principles and involving the private sector in providing services, setting measurable objectives, and introducing performance management with the aim of delivering higher efficiency. This has affected professionals such as psychiatrists by shifting accountability for good medical practice as judged by peers to reporting on outputs to managers as specified by contracts and practice guidelines. Nevertheless, in most cases strong professional accountability remains, creating a dual accountability that may be appropriate but is always challenging and sometimes conflicting.

Mental health public policy

Public policy in mental health is driven by a range of influences and interests which policy makers will attempt to balance. Positive drivers, that is factors that encourage the development of progressive mental health policies, include:

- epidemiological evidence of the prevalence and the contribution of mental disorders to the burden of disease;
- evidence for effective and efficient interventions;
- public interest as expressed by politicians and the media;
- campaigns by NGOs and professional groups;
- human rights considerations;
- pressure from patient and family organizations;
- growing public expenditure.[5]

The strong interface between mental health issues and human rights explains the importance of human rights conventions and declarations in the shaping of policies and legislation, especially in countries where mental health policy and practice are less developed historically. Negative drivers, that is, factors that limit the commitment of policy makers, include budget limitations,

competition from other priority areas, discrimination, and the perception that mental health is not a fruitful area for investment.

The function of mental health policy

Mental health policy may serve a variety of functions ranging from the purely political and presentational through to delivering on a moral or social imperative. Often policy is mixed in function, enabling politicians to make a statement on an important human rights or social issue whilst also achieving real improvements for people with mental illness and delivering on economic targets. Mental health policy will also affect delivery of other aspects of policy in relation to crime, housing, education, and health. Thus, a coherent set of social policies cannot be developed without addressing mental health in some way. Mental health has been called a 'wicked' social policy issue (see Bogdanor for a discussion of such issues)[6] because it is hard to define, to address, and to deliver on tangibles, but unless we do, many of our social aspirations will not be fully realizable. Finally, the stated mission or desired outcome of policy may vary—for example some countries may have goals in terms of reduced prevalence of an outcome such as suicide, others may wish to achieve de-institutionalization.[7]

Broader public policy and mental health

Whilst policy on mental health services *per se* is likely to be the concern of the health ministry, the impact of other public policy on both mentally ill people, and the practical application of psychiatry may be even bigger. People with severe or persistent common mental health problems often face social exclusion and face difficulties with issues including housing/shelter, employment, education, and welfare or welfare benefits. All mental health professionals in all countries of the world will be aware of how much these issues can affect the lives of people with mental illness and indeed clinical outcomes. Box 7.1.1 summarizes some of the main relevant policy areas which will impact on mental health services.

It should be clear from only brief examination of this table that achieving a totally coherent mental health policy presents huge challenges:

- The views of different parties and the priorities of many different government departments must be reconciled;
- Policy must integrate horizontally—for example, across health, education, and local government, and vertically from small organizations up to government;
- The non-measurable side of what is offered is very important to the stakeholders—services must be responsive, caring, joined up and visible to users, carers, and the general public—yet these features of services are hard to influence from central government;
- The needs and demands are so large and diverse it is difficult to construct and resource a coherent system for delivery even in the richest countries;
- Achieving equity and balancing resource allocation across groups of people suffering from a range of mental disorders with different prevalence and causing a different burden to self, families, and communities. Different groups of disorders require different investment producing different potential benefits yet such decisions must often be made on the basis of ambiguous information whilst subject to conflicting pressures from advocacy groups, communities, media, colleagues, government ministries, and parliament.

Box 7.1.1 Policy links for mental health across government

Government department or ministry (generic description may not accord with structures in all countries)	Mental health relevant policy responsibility (Examples)
Department of health/health and social services	Primary mental health care Human resources for health and social care services Finance for health and social care Specialist mental health care policy Public mental health or mental health promotion
Department of employment/ social welfare	Welfare benefits for people with mental health problems Employment rights and protection Occupational health
Department of education	Education on mental health issues as part of wider curricula in schools and higher education Healthy schools Higher education for relevant vocational qualifications
Department of criminal justice/internal security	Diversion of mentally ill offenders from the criminal justice system Public security
Department of housing/ communities and local government/regions	Housing for mentally ill people Community development Urban policy
Department for constitutional affairs/human rights	Protection of those who lack capacity
Treasury or finance ministry	Financing of the above at a macro level Strategic value for money

(The authors after Jenkins *et al.* (2002))

Comparative policy

It can be argued that mental health policies in most countries fall within four broad groupings:

1 Provision for the basic protection of human rights of people with mental illness or those lacking mental capacity;

2 As one together with a simple health care policy to provide for the health care needs of people with serious mental illnesses;

3 The above but with a more comprehensive health and social care policy which addresses primary and specialist care needs for various groups of disorders;

4 A comprehensive cross-sectoral mental health policy that addresses care, public mental health, and broader issues such as housing.

No country in the world has a fully comprehensive policy although developed countries such as New Zealand, Australia, several

European countries, and states of Canada and the United States have advanced policies. This reflects the nature and resources of these societies, and does not necessarily result in differences in prevalence or outcomes.[8]

Human right conventions and legislation

Human rights have historically been a fundamental driver for the development of mental health policy in most countries of the world. The current Human Rights conventions derive from the United Nations, and are binding once they are ratified by member states following adoption by the United Nations General Assembly. The Universal Declaration of Human Rights was adopted in 1948. This was followed in 1966 by the International Covenant on Political and Civil Rights (ICCPR) and the International Covenant on Economic, Social and Cultural Rights (ICESCR), in combination known as the International Bill of Rights. None of the conventions directly addresses issues related to mental health care, although the right to the highest attainable standard of physical and mental health is included in Article 12 of the ICESCR. In non-binding general comments, the committee on Economic, Social and Cultural Rights specified that right to health covers availability, accessibility with a strong emphasis on equality and non-discrimination, and acceptability including cultural sensitivity and quality. The obligation of signatories to provide information on detentions and measures taken to prevent abuse for people admitted to mental hospitals is important.[9]

In 1991, the United Nations General Assembly adopted 'the principles for the protection of persons with mental illness and the improvement of mental healthcare', better known as the MI principles, which remains a key reference document for national mental health legislation. The document outlines 25 principles, stressing the importance of non-discriminatory practice within the health care system. The rights to which patients with mental health problems are entitled, according to the MI principles, include the right to live and work as far as possible in the community, and the right to the best available treatment and care in the least restrictive settings in conditions suited to the cultural background of the patient. The principles also set out expectations for confidentiality, the protection of autonomy, and the conditions for involuntary detention. In combination the principles offer a good foundation for modern mental health policies and legislation. Although the document is not legally binding, it is clearly stated that member states are expected to implement these principles fully.

An essential principle, confirmed by the World Conference on Human Rights in 1993, is that all human rights are universal, equally applying to people with disabilities and mental disorders. This is a strong driver for national governments to deliver anti-discrimination legislation, guidance, and practice.

International agencies

There are several relevant regional agencies that represent countries, including political organizations such as the European Union, the Council of Europe and the European Court of Human Rights, the African Commission on Human and People's Rights, and the Inter-American Court of Human Rights. All have produced conventions and declarations pertinent to people with mental health problems, some legally binding.[10]

The World Health Organization (WHO), the specialist health agency of the United Nations, is mandated by its constitution to support member states in all health areas, including mental health. Its core functions include to:

- propose conventions, agreements and regulations, and make recommendations with respect to international health matters;
- assist governments, upon request, in strengthening health services;
- promote improved standards of teaching and training in the health, medical, and related professions;
- study and report on, in cooperation with other specialized agencies where necessary, administrative and social techniques affecting public health and medical care from preventive and curative points of view, including hospital services and social welfare.

In 2001, WHO dedicated the World Health Report to mental health. The report entitled 'New Understanding, New Hope' urges member states 'to seek solutions for mental health that are already available and affordable'.[11] The report formulates a set of 10 recommendations for member states to address the challenges faced in various areas of mental health. By adopting the report (World Health Assembly Resolution WHA 55.10), governments have committed themselves to implement its recommendations to strengthen primary care and develop accessible and affordable community-based mental health services.

At WHO regional level there have been several declarations and action plans that elaborated the principles of the World Health Report 2001, and reinforced the commitment by governments to delivery. Examples include the Caracas Declaration for the Americas, anticipating the World Health Report, and the European Declaration for Mental Health signed in Helsinki 2005.

Why public mental health policy?

Although the governments of most countries accept a role for public policy in the regulation, commissioning, funding, resourcing, and provision of mental health care, particularly for persons with severe and enduring mental health problems, a case still should be made why mental health care cannot be left to a combination of the private market and civil society to deliver without central intervention. As we have argued, when national systems are compared, considerable variation emerges, even within developed countries. In some, the role of government is limited to regulation of independent health insurance groups that purchase from independent providers, and the public funding of a safety net for expensive acute care or long-term conditions. In other countries all parts of the health system are state owned and tax funded.

The World Bank produced a report on this question,[12] scrutinizing the arguments for public financing of mental disorders as compared to other conditions that compete for public money. In their opinion, disease burden, cost-effectiveness of interventions, and externalities are not sufficient arguments since these are not unique to mental health. However, the potential of catastrophic cost, the risk of insurance market failure, and the place of involuntary treatment and consequent issues of human rights are all disproportionately present in mental health care and plead in favour of a public role. One could add to this that stigma and

discrimination may negatively affect the availability and access to mental health services if left to market forces together with the argument advanced above that comprehensive social policies cannot logically leave out mental health.

Economic impact of disease burden

An important argument for the view that mental health should be central to health care and public policy in general is the burden of disease evidence, addressed elsewhere in this book (cross ref). The realization that mental disorders contribute so much to the overall burden of disease puts mental health care at least on an equal footing with other groups of disorders that have long been recognized as posing a major public health challenge such as infectious disease, cancer, and heart disease. Even more startling is the very high proportion of years lived with disability attributable to mental health problems and the days lost to employment due to mental health conditions. The majority of mental health problems contributing to disability are anxiety and depression and the burden of substance misuse and organic disorders in developed countries is also very high, and rising.

From a public policy perspective, this means that the consequences of mental health problems are no longer only an issue of health care, but are central to macro-economic national interests, particularly at times of skills shortages. This has become connected with the interest in the mental well-being of the population, based on the observation that level and growth in Gross National Product (GNP) is only very marginally associated with status of mental well-being of the population, and in some cases even inversely correlated.[13] This has resulted in the conclusion by economists that government's macro-economic policies should not simply be driven by the aim to maximize the Gross National Product and the income of its citizens. Rather, the 'happiness' of the population should become the prime driver of policy making.[14] Governments have consequently seen a need to stimulate the employment of people with mental disorders, both as part of the social inclusion agenda and to reverse the very high and increasing cost of mental disability. This could be considered as a somewhat symptomatic approach ignoring the more systemic problem of job stress and insecurity in very competitive market economies. However, there is also no doubt that loss of employment is a disastrous event for individuals and families in its own right, whether considered from an emotional, social, or financial perspective. The strong association of unemployment with divorce, mental illness, suicide, all factors associated with social exclusion[15] in their own right, supports the case for increasing employment rates, if not exclusively so. A challenge mental health services are facing is to identify the most appropriate and effective response to such broader societal needs as many of the levers for change lie outside the health care sector.

The scope of mental health public policy development

The scope of mental health policy making has broadened, as reflected in the content of international declarations. During the second half of the twentieth century mental health policy and expenditure was focused on people with severe and enduring conditions cared for in institutional settings. Increasingly the focus is broadening to incorporate mental health promotion and prevention and treatment of common mental disorders, including a key role for primary care. The reasons for this include:

- an increased demand for the treatment of stress and depression;
- the development of community-based services, lowering the threshold to access for, and acceptability of care;
- the growing evidence base for a range of interventions such as cognitive behavioural therapy (CBT);
- increased media interest in common mental disorders and their treatment;
- the reduction of stigma of common mental disorders;
- growing affluence in many countries making privately purchased therapies more affordable;
- an awareness that mental health problems are associated with many social determinants of health, including inequality, and macro-economic conditions;
- the co-morbidity between mental health problems and many physical diseases including diabetes and CHD, affecting mortality, and recovery rates;
- an awareness of the cost of mental illness to society, and the evidence of cost-effective promotion and prevention activities.

All these have major consequences for public policy. The complex interface between public mental health activities and mental illness services demands a multi-agency approach, crossing the responsibilities of government departments and local services and requires 'joined up government'. This requires clear policies with explicit objectives and targets, specifying responsibilities for delivery and funding. It requires the active ownership of mental health issues beyond the health care sector and its advocacy by leading practitioners across a range of sectors and disciplines.

Policy implementation

The growing range and complexity of mental health-related activities requires explicit strategies outlining the vision and values of reform, the planned service organization, capital and human resource implications, financing, quality monitoring, and human rights protection. The number of countries in the world with such detailed mental health policies as defined above is still low but growing. There are also many examples of good strategies, whether at national, state, regional, or local level, that have produced impressive change, and guidance for the drafting of policy and legislation is available.[16]

The common factors in countries with a successful track record of change include:

- full commitment by government;
- national consultation exercises, gaining the views and support of key stakeholders;
- clear and consistent communication of objectives;
- realistic and sustained resource allocation;
- workforce strategies including education and training.

However, there are also examples of countries that have drafted comprehensive strategies that have not been implemented.

Failure of implementation will create cynicism that may obviate future attempts. Main factors that lead to failure include:

1 Poor connection between the strategy and reality, and lack of focus. In some instances strategies are written by foreign experts ignorant of local circumstances. Other examples are of strategies that are based on ideology and are overambitious, repeating the content of declarations without taking into account local needs and resources.

2 Lack of financial commitment or poor grasp of the financial costs of a new model of care, causing rejection by the finance ministry or inability to deliver locally.

3 Neglect of the human resource implications. New models tend to require more staff with new skills, and implementation can create high levels of anxiety that need to be confronted.

4 Lack of consultation with professional groups, users, carers, and communities, who therefore reject or resist the strategy.

5 Insufficient joint planning either across government or with partner organizations that will carry responsibilities for the funding and delivery of essential components such as social care, housing, and employment. Incentives will need to be created for various groups such as professionals and the private sector to support the strategy. Added to this is the need to align incentives across different areas, and to prevent the development of perverse incentives, leading to conflicts of interest.[17]

6 Lack of a long-term perspective. It is often possible to develop local model services, but generalization to national and sustainable programmes requires different approaches.

7 Lack of all party political support. It is unlikely reform will be initiated and embedded within the lifetime of a single parliament/administration or under the governance of the same minister of health. Unless political support is strong across the political spectrum, the risk of reversal due to weak commitment or even hostility to previously agreed plans is high.

Human resources

Any strategy must include plans for sufficient staff with the right attitudes, skills, and competencies. The growing demand for mental health care has increased the need for staff. Many countries around the world are looking for ways to remedy staff shortages, particularly of doctors and nurses. Three approaches can be distinguished:

- increase of national recruitment;
- creation of different roles and responsibilities for staff;
- international recruitment.

Although the increase of training places is on the face of it an obvious approach, it does not solve the problem immediately since the lag time between the creation, for example, of a medical training place and the production of a psychiatrist will be at the least 10 years. The availability of training places assumes that places will be filled with able candidates, which is also not always the case. In some poorer countries students reject the option to become mental health workers due to stigma, poor working conditions, and poor pay.

An important alternative strategy is to take a more comprehensive competency-based approach, and analyse the roles and responsibilities of staff groups that are available and could be equally effective, sometimes even at lower cost. A potential example of this is nurse prescribing.[18] A step further is the creation of new workforce roles, compatible with the direction of reform. Community-based services may require a multi-disciplinary team involving a larger proportion of social workers and psychologists instead of only doctors and nurses. One could also consider new types of community workers focusing on prevention or early intervention, or support workers for patients with severe and enduring disorders. A primary care-based approach may suggest the development of mental health expertise for primary care workers. In practice, however, such models tend to create additional demands for staff and change roles of existing staff, rather than reducing the need for the specialists who remain in short supply.[19]

A short cut to solving staff shortages is international recruitment, setting up a three-way dilemma between ethical recruiting practice, the interest of both receiving and sending countries, and the benefit of individuals. In some trade zones the free migration of employees, including health care staff, is a basic right. It is also hard to argue that health care staff living in poverty and insecurity should be deprived of the opportunity to improve their lives by working in rich and secure countries. Developed countries argue that it is their responsibility to ensure access to health care for their population, although some consideration for the consequences of the country of origin should be shown. Ethical recruitment guidelines have been developed, distinguishing between active recruitment, that is deliberately setting out to attract staff, and passive recruitment, offering open opportunities, for example, by advertising in an international journal. The presence of a bilateral memorandum of agreement is also a indication of good practice.[20] It has to be recognized that international recruitment is not a phenomenon limited to western host countries. It is a global process, with a stream of staff following the trail of higher salaries.

Funding

If it is accepted that the public sector is responsible for the funding of mental health care, the total amount and by implication the proportion of the health budget to be allocated to mental health has to be determined, whether at a national, regional, or local level. In theory this could be calculated on the basis of a population needs assessment, matching the level of need with the cost of interventions. Many countries have performed epidemiological surveys with a sufficient degree of precision to allow an estimate of the prevalence of diagnostic groups.[21] It is more challenging to take account of the cost of treatment. Attempts to determine the costs of specific interventions are possible in research settings. However, in psychiatry, the cost variation even within homogeneous diagnostic groups attributable to factors such as severity, age, co-morbidity, ethnicity, geography, and social context, as well as supply factors such as the range of treatments and delivery options, make any prediction imprecise, as demonstrated by the challenge of developing Diagnosis Related Groups (DRGs) for mental health care.[22]

The variation of spending on mental health across countries is striking. About 20 per cent of countries in the world dedicate less than 1 per cent of the total health budget on mental health, and a

few more than 10 per cent as reported in the WHO Atlas project (2005),[23] with a strong positive correlation between the mental health budget and GDP. Thus, the poorest countries have a very low public spending on mental health, although this ignores the importance of out of pocket payments in many of these countries, sometimes with catastrophic impact on families.

Variations over time of the proportion of health budget allocated to mental health tend to be minor. This suggests that proportion of budget allocated to mental health can be attributed to static factors such as historical spending, often based on running costs of hospitals, adjusted by political priorities and influenced by factors such as advocacy and level of stigma. We do not know of any example of a national formula of rational health budget allocation at treasury/finance ministry level, although a number of countries use a psychiatric needs index to inform sub-national spending decisions.

Equitable resource allocation

A decision on basis of equity needs to balance the burden of disease for patients, families, and the community (see Chapter 7.5), intensity of human suffering, and the cost-effectiveness of treatment relative to other health conditions. In developed countries budgets can be expected to cover basic needs and simple evidence-based practices across all disease groups. It would be hard to accept in Western Europe or North America that patients have no access to treatments for conditions such as tuberculosis or schizophrenia. Some rationing in a variety of implicit or explicit forms is usually in place for expensive new technologies or costly treatments for common and mild conditions. Whilst mental health has few hi-tech interventions, interventions like CBT would be highly expensive if applied to more than defined subgroups of patients.

In poor countries with very low mental health budgets, many lives will depend on the decision how best to invest the scarce resources in order to gain the greatest cost-effectiveness as measured by DALYs averted. Cost-effectiveness studies have found that episodic treatment with older antidepressant drugs is the most cost-effective treatment in poor countries, whereas the treatment of schizophrenia is relatively more expensive.[24] Even more challenging is to decide on the merit of allocation across disease groups, for example, tuberculosis versus mental health.[25] However, one has to take into account that figures are based on averages, and ignore the very great personal and community benefits that some individual interventions can offer. In mental health, it also has to be considered that interventions are not only about treatment of diseases, but also involve in many instances action against major human rights abuses and the consequences of neglect, such as patients with schizophrenia locked up for years or children ignored in institutions. It is hard to put monetary values on such decisions, and this yet again places mental health activities beyond a narrow health treatment perspective, but shows it as one component of broad public policy making. The challenge is to keep mental health high on the agenda of all public policy makers, rather than it being allowed to slip down every government department's list of priorities.

Further information

Moniz, C. and Gorin, S.H. (2006). *Health and mental health care policy: a biopsychosocial perspective*. Allyn and Bacon, Boston.

Knapp, M., McDaid, D., Mossialos, E., *et al.* (2006). *Mental health policy and practice across Europe*. Open University Press, Maidenhead.

Brody, E. and Kemp, D.R. (eds.) (1993). *International handbook of mental health policy*. Greenwood Press, Westport.

Jenkins, R., McCulloch, A., Friedli, L., *et al.* (2002). *Developing a national mental health policy*. Maudsley Monograph. The Psychology Press, Hove.

References

1. Koontz, H. and Weihrich, H. (1988). *Management*. McGraw Hill, New York.
2. McCulloch, A. and Cohen, A. (2007). *Mental health policy and primary mental health care: present and future. Chapter in Mental health care policy and primary mental health care: present and future*. Royal College of General Practitioners, London.
3. Hughes, O.E. (1994). *Public management and administration*. St. Martin's Press, New York.
4. Peters, T. and Waterman, R. (1982). *In search of excellence: lessons from America's best Run Companies*. Harper and Row, New York.
5. Friedman, B. (2005). *The moral consequences of economic growth*. Oxford University Press, Oxford.
6. Bogdanor, V. (2005). *Joined up government*. Oxford University Press, Oxford.
7. Jenkins, R., McCulloch, A., Friedli, L., *et al.* (2002). *Developing a national mental health policy*. Maudsley Monograph 43. The Psychology Press, Hove.
8. Jablensky, A., Sartorius, N., Ensberg, G., *et al.* (1992). Schizophrenia: manifestations, incidence and cause in different cultures—a World Health Organization ten country study. *Psychological Medicine Monograph*, (Suppl. 20), 1–97.
9. WHO. (2005). *WHO resource book on mental health, human rights and legislation*. World Health Organization, Geneva.
10. WHO. (2005). *WHO resource book on mental health, human rights and legislation*. World Health Organization, Geneva.
11. WHO. (2001). *World Health Report 2001: mental health: new understanding, new hope*. World Health Organization, Geneva.
12. Beeharry, G., Whiteford, H., Chambers, D., *et al.* (2002). *Outlining the scope for public health sector involvement in mental health*. The World Bank, Washington, DC.
13. Offer, A. (2006). *The challenge of affluence*. Oxford University Press, Oxford.
14. Layard, R. (2006). *Happiness. Lessons from a new science*. Penguin Books, London.
15. Warr, P. (1987). *Work, unemployment and mental health*. Oxford University Press, Oxford.
16. WHO. (2003). *WHO mental health policy and service guidance package: organization of services for mental health*. World Health Organization, Geneva.
17. Muijen, M. and Ford, R. (1996). The market and mental health: intentional and unintentional incentives. *Journal of Interprofessional Care*, **10**, 13–22.
18. Department of Health. (2006). *From values to action: the chief officer's review of mental health nursing*. Department of Health, London.
19. Muijen, M. (2006). Challenges for psychiatry: delivering the mental health declaration for Europe. *World Psychiatry*, **5**, 113–17.
20. McIntosh, T., Torgerson, R., and Klassen, N. (2007). The ethical recruitment of internationally educated heath professionals: lessons from abroad and options for Canada. Canadian Policy Research Networks, Ottawa, Ontario.
21. WHO World Mental Health Survey Consortium. (2004). Prevalence, severity and unmet need for treatment of mental disorders in the

World Health Organization World Mental Health surveys. *The Journal of the American Medical Association*, **291**, 2581–9.
22. Burgess, P., Pirkis, J., Buckingham, W., *et al.* (1999). Developing a casemix classification for specialist mental health services. *Casemix*, **4**.
23. WHO. (2005). *Atlas: mental health resources in the world 2005*. World Health Organization, Geneva.
24. WHO. (2006). *Dollars, DALYs and decisions: economic aspects of the mental health system*. World Health Organization, Geneva.
25. Jamison, D.T., Breman, J.G., Measham, A.R., *et al.* (2006). *Disease control priorities in developing countries* (2nd edn). World Bank/Oxford University Press, New York.

7.2

Service needs of individuals and populations

Mike Slade, Michele Tansella, and Graham Thornicroft

Introduction

The importance of needs assessment has been one of the most consistent themes to emerge from the evolution of community mental health services. However, the concept of 'need' is used in different, and sometimes contradictory, ways. The aim of this chapter is to

- define needs assessment
- consider different approaches to assessing needs, both at the individual and at the population levels
- discuss how needs assessments can be applied in real-world settings in planning and delivering clinical care.

Defining needs

At its simplest, a need involves a lack of something. But of what? In operational terms the concept of need is usually applied to a difficulty (in this case in relation to a person with mental illness) for which a possibly effective intervention exists. By implication an experienced difficulty for which there is no known effective intervention is therefore not defined as a need.[1] The clearest categorization of such needs was identified by Brewin, who grouped definitions of need within mental health care into three categories: lack of health, lack of access to services or institutions, and lack of action by mental health workers.[2] Approaches to need within each of these three categories will be reviewed.

(a) Needs for improved health

The psychologist Maslow established probably the best-known hierarchy of need, when he formulated a theory of human motivation.[3] In his model, fundamental physiological needs (such as the need for food) underpin the higher needs of safety, love, self-esteem, and self-actualization. He proposed that people are motivated by the requirement to meet these needs, and that higher level needs could only be met once the lower and more fundamental needs were met. The clinical relevance of this theory is that it implies a hierarchy of clinical priorities—interventions to meet basic physiological need (e.g. to ensure adequate food supply) should take priority over interventions to foster, for example, self-esteem.

In practice health-related needs are often considered in a widely defined way. In England, for example, the requirement to base the provision of services on level of need was first made explicit in the National Health Service and Community Care Act,[4] which defined need as *the requirements of individuals to enable them to achieve, maintain, or restore an acceptable level of social independence or quality of life*. This requirement was retained when national standards for mental health services were set.[5] This involves **needs-led** care planning—basing care for an individual patient on an assessment of their health and social needs. The needs-led approach offers many benefits:

1 The overall level of need gives guidance about which part of the mental health system should treat the patient, for example that people with less disabling mental disorders should be seen in primary care settings.[6]

2 Needs assessment can improve the comprehensiveness of case formulations and care plans by incorporating a broad range of health determinants, such as poor housing or lack of social support.

3 Explicit identification of need can support clinician–patient discussions about care priorities, which is associated with improved treatment satisfaction[7,8] and compliance.[9,10]

4 Identification of needs helps to identify the contribution of services outside the psychiatric sector.

5 Needs-led care can facilitate more individualized treatment planning than diagnosis-driven approaches, by more closely matching the help offered to patient's needs and by explicitly identifying problems which require the involvement of both health and other agencies.

Needs-led care planning focussing on health can be differentiated from the assessment of care needs. Assessing care needs involves identifying whether the patient will benefit from a predefined menu of interventions, and by definition will not identify all unmet needs for individual patients. Assessment of need at the patient level should therefore be a separate process from decisions about what care or treatment to provide. There are, however, other reasons to assess needs for services, which we now review.

(b) Needs for services

The second category of need is a requirement for a particular type of service. At the population level, it is possible to use epidemiological methods to develop prevalence for different disorders, which can be translated into estimates of the need for services. A recent epidemiological survey in the United States, for example, found very considerable unmet need of the population level nationwide.[11] This study identified that between 1990–92 and 2001–03 the overall annual period prevalence of mental illnesses remained constant at between 29.4 and 30.5 per cent. Among these cases, however, there was an increase in the proportion who received any treatment at all, rising from 20.3 to 32.9 per cent between the two time periods. The inverse is however very revealing, namely that the most recent data show that 67 per cent of people with mental disorders in the United States receive no treatment. The situation is worse in other countries. A recent comparative international study of depression found that 0 per cent of patients in St Petersburg received evidence-based treatment in primary care, and only 3 per cent were referred on to specialist mental health care.[12] The inability of patients to afford out-of-pocket costs was the primary barrier to care for 75 per cent of the depressed Russian patients studied.

International comparisons of population-level needs have been conducted in recent years. The ESEMed Study, for example, carried out cross-sectional surveys in Belgium, France, Germany, Italy, the Netherlands, and Spain among 8796 representative members of the general population. Individuals with a 12-month mental disorder that was disabling or that had led to use of services in the previous 12 months were considered in need of care. The study found that about 6 per cent of the sample was defined as being in need of mental health care. Nearly half (48 per cent) of these people reported no formal health care use, so that 3.1 per cent of the adult population had an unmet need for mental health care. In contrast, only 8 per cent of the people with diabetes had reported no use of services for their physical condition.[13]

(c) Needs for action

In health care, the concept of need has been taken to mean the ability to benefit in some way from health care, and thus distinguished from demand (what the person asks for) and supply (services given).[14] For example, the MRC Needs for Care Assessment Schedule is premised on the assumption that need is 'a normative concept which is to be defined by experts'.[15]

Using this approach, an Australian study compared current and optimal treatment for 10 high-burden mental disorders in Australia.[16] This found that current levels of treatment at current coverage avert 13 per cent of the overall burden attributable to these disorders. Providing optimal treatment at current coverage would avert 20 per cent of the burden, and optimal treatment at optimal coverage would avert 28 per cent. The development of a more robust treatment evidence base makes this innovative approach to informing public policy more possible, and the approach can be recommended for evidence-based policy initiatives.

Patient and staff perceptions of need

There has been a long-standing recognition that differences in perceptions of need can exist, in particular between staff and patient. In the 1990s the emphasis was put on acknowledging these differences, but then prioritizing the staff perspective. For example, UK policy stated that *all users … should be encouraged to participate to the limit of their capacity. …Where it is impossible to reconcile different perceptions, these differences should be acknowledged and recorded*.[17] Several societal and scientific developments challenge this prioritization of staff over patient perspectives.

First, general societal changes towards consumerism and an emphasis on rights have produced more assertive mental health service users. Easier access by patients to internet-based information reduces the knowledge disparity. Reduced societal trust in the authoritative expert has eroded the position power of mental health staff. The emphasis put on choice and empowerment raise patient expectations of being more than passive recipients of care.[18,19]

Second, the prioritization of staff perspectives has been actively challenged by an increasingly vociferous and organized user movement. This opposition has found its voice in the 'recovery' movement, which emphasizes the meaning and values of the patient, and the need for services to foster self-management rather than dependency. There has been widespread international policy support for recovery-focussed services[20] although there can be tensions between what professionals construe as their duty of care and being led by the patient perspective on need, which can create ethical dilemmas. Care planning which emphasizes agreement between staff and patients may have additional advantages. A recent study in Verona showed staff–patient agreement on needs was significantly associated with better treatment outcomes both rated by the patient and by staff (psychopathology, social disability, global functioning, subjective quality of life, and satisfaction with care).[21] Similarly, there is emerging evidence that crisis plans (advanced statements) which are jointly agreed between staff and patient can be cost-effective in reducing compulsory admission to hospital.[22,23] Such emerging findings indicate that needs assessment and care planning, which are based on negotiation and jointly agreed analyses of problems and interventions, are likely to become increasingly important in future.

Finally, emerging empirical evidence strongly supports the positioning of the patient perspective at the heart of needs assessment and care planning. Evidence from several studies consistently shows differences between staff and patient perspectives on need,[24,25] so the two perspectives are not interchangeable. Empirical research suggests two reasons for basing care on the patient rather than staff assessment of need. First the patient rating is more stable than the staff rating.[26] Second, longitudinal studies indicate a causal relationship between patient-rated (but not staff-rated) unmet need and quality of life.[27–29] If the goal of mental health services is to improve quality of life, then best available evidence indicates that the patient's perspective on their unmet needs should drive care planning.

Assessing needs

In this section we identify specific approaches to assessing needs.

(a) Individual-level needs assessment measures

Several standardized approaches to the assessment of patient-level need have been developed, primarily in the United Kingdom. These have shown a transition along a continuum, from an initial focus on assessment of need as an objective state to be defined by experts following careful assessment, towards those which emphasize the subjective nature of needs assessment.

The earliest standardized needs assessment measure was the **Medical Research Council Needs for Care Assessment (NFCAS)**.[30]

The NFCAS assesses the need for further action by health care professionals, and links identification of a need with a predefined list of actions. This raises two problems. First, the emphasis on identifying available interventions which would be at least partly effective is problematic, given the complexities of deciding that a treatment has not worked. Second, updating the list of actions has proved problematic. However, as Bebbington notes, 'the inevitable value judgements inherent in the procedure have the virtue of being public and consequently accessible to argument'.[31] An important variation of the NFCAS is the **Cardinal Needs Schedule (CNS)**,[32] which also considers patient willingness to accept help and level of carer concern. Training is needed for using both the NFCAS and the CNS, and they are primarily used for research purposes.

At the other end of this continuum are needs assessment measures which emphasize individual difference and the subjective nature of need. The **AVON Mental Health Measure** was developed by service users, and assesses physical, social, behaviour, access, and mental health domains.[33] It can take up to 20 min for completion by the patient and 5 min by the staff, and its development has emphasized external validity over other psychometric properties. The **Carers and Users Experience of Services (CUES)** was developed by service users and staff, and assesses 16 domains: the place you live, money situation, the help you get, the way you spend time, your relationships, social life, information/advice, access to services, choice of mental health services, relationship with mental health workers, consultation and contact, advocacy, stigma, any treatment, access to physical health services, and relationship with physical health workers.[34] Completion can take up to 30 min. Neither AVON nor CUES have become widely used in mental health services.

The **Camberwell Assessment of Need (CAN)**[35] spans both ends of the continuum. It assesses 22 domains of health and social need, and a key development is that it records staff and patient views separately, without giving primacy to either perspective. Research (CAN-R), clinical (CAN-C), and brief versions (CANSAS) of the CAN have been developed for adults of working age with severe mental health problems,[36] and it has been translated in 22 languages. Variants have been developed for people with learning disabilities and mental health problems (CANDID),[37] mentally disordered offenders (CANFOR),[38] older adults (CANE),[39] and mothers with mental health problems (CAN-M).[40] An updated web resource for the CAN is available at www.iop.kcl.ac.uk/prism/can.

The CAN has become the most widely used needs assessment measure internationally,[41] and is the standardized needs assessment measure which is most relevant to routine clinical practice. The short version, CANSAS, can be recommended for routine use in community services. Two specific approaches have been empirically shown to produce patient-level benefit. First, the patient-rated two-way communication (2-COM) measure is an amended version of the CAN which gives the patient the opportunity to identify unmet needs and also prioritize those which they wish to discuss with their clinician.[42] Asking patients to complete 2-COM before an outpatient appointment and then using that information in the appointment was associated with greater patient satisfaction and more likelihood of treatment change.[7] Second, a structured approach to collating and feeding back staff and patient ratings for CANSAS and other assessments led to a reduction in psychiatric admissions, probably because of earlier intervention during relapse,[43] and improvements in patient-rated unmet need and quality of life for higher premorbid IQ patients.[44] Routine use of CANSAS brings patient-level benefits, and empirical evidence indicates the clinical focus should be on assessment and interventions for patient-rated, rather than staff-rated, unmet needs.

(b) Population-level needs assessment measure

Measures to assess population-based needs can be classified by the data and by the analytic approaches they use. Three types of data are commonly used. The most readily accessible documents the use of current mental health services. While this can be criticized as reflecting only current service provision, its ready availability and nationwide coverage means that is extensively used.

Simple population-based samples are very inefficient in estimating the prevalence of relatively rare conditions such as schizophrenia. Studies thus tend to use two-stage procedures, with a relatively brief initial screening process applied to a large number of people followed in-depth interviews for a selected few. 'Booster' samples, perhaps including all the known psychiatric patients for the areas surveyed, may be sought through mental health services.[45] Population surveys depend on the identification of randomly sampled individuals. Some types of mental health problems, notably substance misuse, are commonly associated with socially marginal lifestyles, making it likely that sufferers will be systematically underrepresented by traditional population sampling approaches. More sophisticated sampling approaches, such as capture–recapture methods, have been used for these situations.[46]

The third type of data relates to the views of local people. Local needs assessment studies entail a structured approach to eliciting the views of service users, their carers, interested voluntary sector organizations, and all statutory agencies with responsibilities in the area. Smith[47] has described how this type of study can be integrated into the overall planning process.

Government initiatives in England have tried to base the allocation of money between areas on the morbidity as well as the size of their populations. This has led to studies modelling this variation. The first widely used index[48] was developed on the basis of consensus between GPs about patient characteristics associated with high use of primary care services. While developed for wider purposes, this was shown to relate reasonably closely to variations in psychiatric admission. Later indices have been established by statistical modelling exercises seeking to quantify the relationship between social variables measured in censuses and either service use,[49] or population-based epidemiological findings. The variation between places in the prevalence of the less severe types of mental illness commonly dealt with in primary care is less than that for problems usually managed by specialist mental health services, which again is much less than that observed for forensic services. Thus models developed for one level of care should not be used to estimate patterns of need for other levels.

In practice, no single approach to assessing the needs of a population will suffice. Needs assessment at this level requires the integration of many perspectives. The Kings Fund review of London's mental health Services[50] illustrates how a detailed perspective can be assembled from many fragments of evidence, each of which would be inadequate in isolation. Recent examples of population-level needs assessment from the United States, Canada, and

New Zealand also reveal that epidemiological studies may not produce data that corresponds directly to needs, and that some sub-populations, for example particular ethnic groups, may be less well represented in such approaches, unless considerable methodological care is taken.[51–53]

If a mental health practitioner is asked to join a committee to plan services, for example, for a local catchment area, what approach is helpful to identify and use population-level needs? Table 7.2.1 indicates a series of steps to find the best available information on population prevalence rates.[54]

As Table 7.2.1 shows, we consider that the best possible information would be local epidemiological data on the occurrence of mental disorders, using a standard system of classification, alongside a measure of the needs for treatment among the prevalent cases identified.[55] Since these assessments are expensive and time-consuming, most sites will not have access to such recent local data. If the data in step (1) are not available then we suggest that country/regional epidemiological data (2) are used instead, and are then weighed for local socio-demographic characteristics. But if such larger scale prevalence data are not available, then a third option is to use international rates from 'comparison' countries or regions, again weighted for local socio-demographic characteristics (3). The results in this case will be less accurate because they are based on the additional assumption that the data can be transferred between countries. A newer set of techniques that offer considerable promise are rapid appraisal/rapid assessment techniques. These are methods to undertake brief assessments of population needs which are focussed upon key focussed questions, for example on how primary care services should be augmented to treat people with depression, and example of these approaches have been used to positive effect in South Africa.[56–58]

In some cases, none of the data described in steps 1–3 will be available, and then the next option (4) is to use a number of experts, some of whom may be from the local area, to produce a consensus statement on the local rates and characteristics of people with mental illness. Such a data synthesis can be based on the best available views, taking into account local factors (e.g. levels of non-health service provision, family support, traditions, degree of affluence, or migration). This pragmatic approach will yield data which are accurate enough to use for local service planning purposes.

Table 7.2.1 Ways to measure or estimate local population mental health prevalence

(1) Actual local epidemiological data on psychiatric morbidity and disability for the particular area by age, sex, ethnicity, social status, and degree of urbanicity
(if not available) ↓
(2) Country/regional epidemiological data weighed for local socio-demographic characteristics
(if not available) ↓
(3) International data from 'comparable' countries or regions, adjusted for local socio-demographic characteristics
(If 1, 2, 3 not sufficient) ↓
(4) Best estimates and expert synthesis and interpretation based on other sources of local information and opinions (e.g. extent of non-health service provision, family support, local traditions, or migration)

This raises the important topics of coverage and focussing. *Coverage* means the proportion of people who receive treatment who could benefit from it.[59] *Focussing* refers to how far those people who actually receive treatment in fact need it: do they have any form of mental illness?[60] Even in the most well-resourced countries one can find both low coverage and poor focussing.[61,62] From the public health perspective, therefore, the key issue is the appropriate use of resources, whatever the level of resources actually available, namely to increase both coverage and focus.

(c) The relationship between individual and population-level needs

We have argued in this chapter that the provision of care for individual patients should be based on assessment of their health and social needs. Can these individual assessments be aggregated to inform service? For population-level service planning, the key question is what types of interventions to provide, and with what capacity. Therefore data from individual needs assessments cannot simply be aggregated to inform service development decisions. While it is theoretically feasible to undertake routine standardized needs assessments on all patients within a service, this approach alone has three drawbacks. Firstly, despite the developing evidence reviewed earlier about the benefits of routine use of standardized outcome assessments, this remains the exception rather than the norm.[63,64] Secondly, there is not yet a sufficiently developed information infrastructure to support the national collection, management, and analysis of such data, despite this being a priority identified two decades ago.[65] Thirdly, even if individual needs assessments were nationally aggregated, the diversity of views (patient, staff, carer, taxpayer) would make a shared interpretation of the data problematic.

Conclusion

In this chapter we have emphasized that it is of central importance when planning mental health service for populations, to do so on the basis of (i) the occurrence of mental disorders in that particular population, (ii) the impairments caused by these disorders that require interventions, (iii) the nature and level of needs among these people, (iv) identifying from among these needs those which are unmet, and then (v) prioritizing new service development on the basis of these unmet needs, including a range of social supports and services (such as housing or employment opportunities, outside the mental health system), the requirements for enhanced physical/general health care, as well as improvements in the provision of specific mental health services. For all of these sectors there is an increasingly clear call from service user/consumer groups for involvement in these priority-setting planning exercises.[18,19]

At the level of individuals with mental illness, there is a similar trend to increasingly involve service users/consumers in assessing needs, with emerging evidence that this produces a more comprehensive basis for care planning. Indeed in the last decade there has been an important conceptual shift away from the view that professionals defined 'needs' while consumers stated 'demands', to a better appreciation of the many advantages to be gained from identifying, as far as possible, unmet needs in a joint and consensual way as a basis for action.

Further information

Andrews, G. and Henderson, S. (eds.) (2000). *Unmet need in psychiatry*. Cambridge University Press, Cambridge.

Lasalvia, A. and Ruggeri, M. (2007). Multidimensional outcomes in 'real world' mental health services: follow-up findings from the South Verona Project. *Acta Psychiatrica Scandinavica Supplementum*, **437**(S116), 3–77.

Thornicroft, G. (2001). *Measuring mental health needs* (2nd edn). Gaskell, Royal College of Psychiatrist, London.

Thornicroft, G., Becker, T., Knapp, M., *et al.* (2006). *International outcome measures in mental health. Quality of life, needs, service satisfaction, costs and impact on carers*. Gaskell, Royal College of Psychiatrists, London.

Thornicroft, G. and Tansella, M. (2008). *Better mental health care*. Cambridge University Press, Cambridge.

Regularly updated web site for the Camberwell Assessment of Need is available at: http://www.iop.kcl.ac.uk/prism/can

References

1. Brewin, C.R. (2001). Measuring individual needs for care and services. In *Measuring mental health needs* (2nd edn) (ed. G. Thornicroft), pp. 273–90. Gaskell, London.
2. Brewin, C. (1992). Measuring individual needs for care and services. In *Measuring mental health needs* (eds. G. Thornicroft, C. Brewin, and J. Wing). Gaskell, Royal College of Psychiatrists, London.
3. Maslow, A. (1954). *Motivation and personality*. Harper & Row, New York.
4. House of Commons (1990). *National health service and community care act*. HMSO, London.
5. Department of Health. (1999). *Mental health national service framework*. HMSO, London.
6. Tansella, M. and Thornicroft, G. (1999). *Common mental disorders in primary care*. Routledge, London.
7. van Os, J., Altamura, A.C., Bobes, J., *et al.* (2004). Evaluation of the two-way communication checklist as a clinical intervention. *The British Journal of Psychiatry*, **184**, 79–83.
8. Lasalvia, A., Bonetto, C., Malchiodi, F., *et al.* (2005). Listening to patients' needs to improve their subjective quality of life. *Psychological Medicine*, **35**, 1655–65.
9. Haynes, R.B., McDonald, H.P., and Garg, A.X. (2002). *Interventions for helping patients to follow prescriptions for medications*. Cochrane Library Update Software, Oxford.
10. Gray, R., Leese, M., Bindman, J., *et al.* (2006). Adherence therapy for people with schizophrenia. European multicentre randomised controlled trial. *The British Journal of Psychiatry*, **189**, 508–14.
11. Kessler, R.C., Demler, O., Frank, R.G., *et al.* (2005). Prevalence and treatment of mental disorders, 1990 to 2003. *The New England Journal of Medicine*, **352**, 2515–23.
12. Simon, G.E., Fleck, M., Lucas, R., *et al.* (2004). Prevalence and predictors of depression treatment in an international primary care study. *The American Journal of Psychiatry*, **161**, 1626–34.
13. Alonso, J., Codony, M., Kovess, V., *et al.* (2007). Population level of unmet need for mental healthcare in Europe. *The British Journal of Psychiatry*, **190**, 299–306.
14. Stevens, A. and Gabbay, J. (1991). Needs assessment needs assessment. *Health Trends*, **23**, 20–3.
15. Bebbington, P. and Rees, S. (2001). Assessing the need for psychiatric services at the district level: using the results of community surveys. In *Measuring mental health needs* (2nd edn) (ed. G. Thornicroft). Gaskell, Royal College of Psychiatrist, London.
16. Andrews, G., Issakidis, C., Sanderson, *et al.* (2004). Utilising survey data to inform public policy: comparison of the cost-effectiveness of treatment of ten mental disorders. *The British Journal of Psychiatry*, **184**, 526–33.
17. Department of Health Social Services Inspectorate. (1991). *Care management and assessment: practitioners' guide*. HMSO, London.
18. Chamberlin, J. (2005). User/consumer involvement in mental health service delivery. *Epidemiologia e Psichiatria Sociale*, **14**, 10–4.
19. Rose, D., Thornicroft, G., and Slade, M. (2006). Who decides what evidence is? Developing a multiple perspectives paradigm in mental health. *Acta Psychiatrica Scandinavica Supplementum*, **429**(S113) 109–14.
20. New Freedom Commission on Mental Health. (2005). *Achieving the promise: transforming mental health care in America*. Department of Health and Human Services, Rockville, MD.
21. Lasalvia, A., Bonetto, C., Tansella, M., *et al.* (2007). Does staff-patient agreement on needs for care predict a better mental health outcome? A 4 year follow-up in a community service. *Psychological Medicine*, doi 10.1017/S0033291707000785.
22. Henderson, C., Flood, C., Leese, M., *et al.* (2004). Effect of joint crisis plans on use of compulsory treatment in psychiatry: single blind randomised controlled trial. *British Medical Journal*, **329**, 136.
23. Flood, C., Byford, S., Henderson, C., *et al.* (2006). Joint crisis plans for people with psychosis: economic evaluation of a randomised controlled trial. *British Medical Journal*, **333**, 729.
24. Lasalvia, A., Ruggeri, M., Mazzi, M.A., *et al.* (2000). The perception of needs for care in staff and patients in community-based mental health services. The South-Verona Outcome Project 3. *Acta Psychiatrica Scandinavica*, **102**, 366–75.
25. Hansson, L., Vinding, H.R., Mackeprang, T., *et al.* (2001). Comparison of key worker and patient assessment of needs in schizophrenic patients living in the community: a Nordic multicentre study. *Acta Psychiatrica Scandinavica*, **103**, 45–51.
26. Slade, M., Leese, M., Taylor, R., *et al.* (1999). The association between needs and quality of life in an epidemiologically representative sample of people with psychosis. *Acta Psychiatrica Scandinavica*, **100**, 149–57.
27. Slade, M., Leese, M., Ruggeri, M., *et al.* (2004). Does meeting needs improve quality of life? *Psychotherapy and Psychosomatics*, **73**, 183–9.
28. Lasalvia, A., Bonetto, C., Malchiodi, F., *et al.* (2005). Listening to patients' needs to improve their subjective quality of life. *Psychological Medicine*, **35**, 1655–65.
29. Slade, M., Leese, M., Cahill, S., *et al.* (2005). Patient-rated mental health needs and quality of life improvement, *The British Journal of Psychiatry*, **187**, 256–61.
30. Brewin, C., Wing, J., Mangen, S., *et al.* (1987). Principles and practice of measuring needs in the long-term mentally ill: the MRC needs for care assessment. *Psychological Medicine*, **17**, 971–81.
31. Bebbington, P. (1992). Assessing the need for psychiatric treatment at the district level: the role of surveys. In *Measuring mental health needs* (eds. G. Thornicroft, C. Brewin, and J. Wing). Gaskell, Royal College of Psychiatrists, London.
32. Marshall, M. (1994). How should we measure need? *Philosophy, Psychiatry and Psychology*, **1**, 27–36.
33. Lelliott, P. (2000). What do people want from specialist mental health services and can this be routinely measured in routine service settings? *Behavioural and Cognitive Psychotherapy*, **28**, 361–8.
34. Lelliott, P., Beevor, A., Hogman, J., *et al.* (2001). Carers' and users' expectations of services—user version (CUES-U): a new instrument to measure the experience of users of mental health services. *The British Journal of Psychiatry*, **179**, 67–72.
35. Phelan, M., Slade, M., Thornicroft, G., *et al.* (1995). The Camberwell assessment of need: the validity and reliability of an instrument to assess the needs of people with severe mental illness. *The British Journal of Psychiatry*, **167**, 589–95.
36. Slade, M., Loftus, L., Phelan, M., *et al.* (1999). *The Camberwell assessment of need*. Gaskell, London.

37. Xenitidis, K., Slade, M., Bouras, N., *et al.* (2003). *CANDID: Camberwell assessment of need for adults with developmental and intellectual disabilities*. Gaskell, London.
38. Thomas, S., Harty, M., Parott, J., *et al.* (2003). *The forensic CAN: Camberwell assessment of need forensic version (CANFOR)*. Gaskell, London.
39. Reynolds, T., Thornicroft, G., Abas, M., *et al.* (2000). Camberwell assessment of need for the elderly (CANE). Development, validity and reliability. *The British Journal of Psychiatry*, **176**, 444–52.
40. Howard, L., Slade, M., O'Keane, V., *et al.* (eds.) (2008). *The Camberwell assessment of need for pregnant women and mothers with severe mental illness*. Gaskell, London.
41. Evans, S., Greenhalgh, J., and Connelly, J. (2000). Selecting a mental health needs assessment scale: guidance on the critical appraisal of standardized measures. *Journal of Evaluation in Clinical Practice*, **6**, 379–93.
42. van Os, J., Altamura, A.C., Bobes, J., *et al.* (2002). 2-COM: an instrument to facilitate patient-professional communication in routine clinical practice, *Acta Psychiatrica Scandinavica*, **106**, 446–52.
43. Slade, M., McCrone, P., Kuipers, E., *et al.* (2006). Use of standardised outcome measures in adult mental health services: randomised controlled trial. *The British Journal of Psychiatry*, **189**, 330–6.
44. Slade, M., Leese, M., Gillard, M., *et al.* (2006). Premorbid IQ and response to routine outcome assessment. *Psychological Medicine*, **36**, 1183–92.
45. Jenkins, R., Bebbington, P., Brugha, T., *et al.* (1997). The National Psychiatric Morbidity surveys of Great Britain-strategy and methods. *Psychological Medicine*, **27**, 765–74.
46. Hay, G. and McKeganey, N. (1996). Estimating the prevalence of drug misuse in Dundee, Scotland: an application of capture-recapture methods. *Journal of Epidemiology and Community Health*, **50**, 469–72.
47. Smith, H. (1998). Needs assessment in mental health services: the DISC framework. *Journal of Public Health Medicine*, **20**, 154–60.
48. Jarman, B. (1983). Identification of underprivileged areas. *British Medical Journal (Clinical research ed.)*, **286**, 1705–9.
49. McCrone, P., Thornicroft, G., Boyle, S., *et al.* (2006). The development of a local index of need (LIN) and its use to explain variations in social services expenditure on mental health care in England. *Health & Social Care in the Community*, **14**, 254–63.
50. Johnson, S., Ramsay, R., Thornicroft, G., *et al.* (1997). *London's mental health. The report of the Kings Fund Commission*. Kings Fund Publishing, London.
51. Messias, E., Eaton, W., Nestadt, G., *et al.* (2007). Psychiatrists' ascertained treatment needs for mental disorders in a population-based sample. *Psychiatric Services*, **58**, 373–7.
52. Kumar, S., Tse, S., Fernando, A., *et al.* (2006). Epidemiological studies on mental health needs of Asian population in New Zealand. *The International Journal of Social Psychiatry*, **52**, 408–12.
53. Hanson, L., Houde, D., McDowell, M., *et al.* (2006). A population-based needs assessment for mental health services. *Administration and Policy in Mental Health*, **34,** 233–42.
54. Thornicroft, G. and Tansella, M. (1999). *The mental health matrix: a manual to improve services*. Cambridge University Press, Cambridge.
55. Thornicroft, G. (2001). *Measuring mental health needs* (2nd edn). Gaskell, Royal College of Psychiatrists, London.
56. Flisher, A.J., Lund, C., Funk, M., *et al.* (2007). Mental health policy development and implementation in four African countries. *Journal of Health Psychology*, **12**, 505–16.
57. Lund, C. and Flisher, A.J. (2006). Norms for mental health services in South Africa. *Social Psychiatry and Psychiatric Epidemiology*, **41**, 587–94.
58. Lund, C. (2002). *Mental health service norms in South Africa*. PhD thesis. University of Cape Town, Cape Town.
59. Habicht, J.P., Mason, J.P., and Tabatabai, H. (1984). Basic concepts for the design of evaluations during programme implementation. In *Methods for the evaluation of the impact of food and nutrition programmes. Food and nutrition bulletin* (eds. D.R. Sahn, R. Lockwood, and N.S. Scrimshaw), pp. 1–25. The United Nations University, New York.
60. Tansella, M. (2006). Recent advances in depression. Where are we going? *Epidemiologia e Psichiatria Sociale*, **15**, 1–3.
61. World Health Organisation (2005). *Mental health action plan for Europe.* World Health Organisation, Copenhagen.
62. World Health Organisation (2005). *Mental health declaration for Europe.* World Health Organisation, Copenhagen.
63. Gilbody, S.M., House, A.O., and Sheldon, T.A. (2002). Psychiatrists in the UK do not use outcome measures. *The British Journal of Psychiatry*, **180**, 101–3.
64. Valenstein, M., Mitchinson, A., Ronis, D.L., *et al.* (2004). Quality indicators and monitoring of mental health services: what do frontline providers think? *The American Journal of Psychiatry*, **161**, 146–53.
65. Ellwood, P. (1988). Outcomes management–a technology of patient experience. *The New England Journal of Medicine*, **318**, 1549–56.

7.3

Cultural differences care pathways, service use, and outcome

Jim van Os and Kwame McKenzie

This chapter discusses the influence of culture on the route an individual takes to access treatment for psychological distress and the treatment received. Culture is difficult to measure. All categories of cultural variables have different meanings and measure different things. Research into them leads to different hypotheses. Given this reality, there is no need to join the cul-de-sac argument of whether one or the other is the most important. In the discussion below, the roles of ethnicity and of socio-economic, political, community, national, and other factors that help to define the culture of an individual or a group are acknowledged. Associations with pathways to care, service use, and outcome will be presented.

Care pathways

Cultural variation in pathways through care

At the beginning of the pathway to care, the individual displays cognitive, physical, or behavioural changes. They or their family, friends, or wider community interpret these as in need of some remedy. The individual's personal resources and then informal resources of family and friends are often triggered to help deal with the problem. These may lead to resolution but if they do not they may lead to presentation through an ever more distant and 'professional' array of caregivers, help agencies, and formal medical services. Most help for psychological problems is not given by mental health services. Interventions and their perceived success or failure move an individual along a pathway.

Pathways through care have differing directions and durations. These depend on where the pathway starts, the presenting symptoms, and psychosocial and cultural factors in the individual, their community, and the services used. Pathways are not random, they are structured and set by a dialogue between the individual, the community, and the code of the statutory services and the law set within that country.[1]

At each level the aim of care is to help the individual move down to less professional interventions until they are either back in the community or at the lowest intensity of care that meets their needs. Traditionally, care pathways have considered routes to getting treatment but with de-institutionalization, social inclusion and the recovery model, pathways out of care need to be considered more widely.

International comparisons

Cultural variation in pathways to care for a mental health problem is readily, though crudely, demonstrated through international comparisons. For example, in an international study of the pathways to care of 1554 patients newly referred to the mental health services in 11 centres in different countries, the majority of patients (63–80 per cent) were referred by their general practitioner in United Kingdom, Spain, Portugal, Czechoslovakia, Cuba, Mexico, and Aden Democratic Republic of Yemen. Only between 0–15 per cent of patients in these centres had referred themselves, with the exception of Mexico (24 per cent). In Kenya, only 7 per cent were referred by general practitioners but 72 per cent were referred by a hospital doctor. In Pakistan and India, a quarter of patients were referred by general practitioners, a quarter by hospital doctors, a third were self-referred, and 11–17 per cent were referred by religious healers. In Indonesia, both primary care and native healer referrals constituted each around a third of all referrals. The differences between these 11 centres largely reflected the people's choice of first port of call for a psychiatric problem.[2]

International differences in pathways are also reflected in primary health care studies. In a 14 country World Health Organization (WHO) investigation, the proportion of attendees with anxiety and depression as defined by the Composite International Diagnostic Interview (CIDI) varied five-fold across centers. Asian sites reported the lowest rates and European and South American sites the highest. The differences in prevalence may reflect a combination of demographic differences between attendees, true differences in population prevalence, the differential availability of other culture-specific pathways to care for the psychologically distressed, and differential sensitivity of the CIDI in picking up psychiatric disorder in different cultures.[3]

Geographically less dispersed countries also demonstrate significant differences in care pathways. In a recent study of 6 Eastern European countries, the percentage of new patients with schizophrenia who first sought care from psychiatric services ranged

from 69 per cent in Bucharest (Romania) to 47 per cent in Zagreb (Croatia). Thirteen percent of patients first sought care from general practitioners in Strumica, Macedonia but 47 per cent in Zagreb. The police were the first port of call in 8 per cent of cases in Bucharest but for none in Zagreb.[4]

This study used the 'encounter form' developed for the WHO which can be used to map and quantify pathways to care (Fig. 7.3.1).

Marked differences in pathways to care also exist within countries and groups. The increase in the use of involuntary admission of African-Caribbeans in the UK and African-Americans in the US is well documented. The reasons for this are unclear but some argue that it reflects service configuration.[5] More recently, increased involuntary admission rates have been reported in the Maori population of Auckland, New Zealand.[6]

Geographic and ecological factors may also contribute to variation in pathways to care within a country. In a Canadian province, involuntary admissions were shown to be related to the size of a community and its proximity to the hospital. Thus, involuntary admission rates were increased if a community was close to the hospital. Rates were also higher in both densely populated inner city areas as well as in cities of less than 500 people. The higher rates of involuntary admission in small towns may be related to a decreased likelihood of being tolerated or remaining anonymous.[7] Despite a similar mental health act, involuntary admissions in Greenland were found to be twice as high as they were in Denmark or the Faroe Islands. The excess risk was associated with the higher homicide rate, lower psychiatric bed availability, lower access to psychiatric care, small settlements, and increased alcohol consumption and violence in Greenland.[8]

Factors associated with cultural variation in pathways

How interpersonal or cultural factors are translated into differences in pathways has rarely been assessed. However, a number of factors have been identified that contribute to cross-national and cross-ethnic differences.

The family may play an important role. An American study reported that Chinese patients were kept for extended periods of time within their families at the beginning of pathways, while Anglo-Saxons and Central Europeans were referred by their relatives or themselves to a range of mental health and social agencies. Native Americans tended to be referred by people other than relatives or themselves.[9] In another study, both Asians and African-Americans showed more extended family involvement, and the involvement of key family members tended to be persistent and intensive in Asians. Ethnicity was also associated with the length of delay, Asians showing the longest delay and white people, the shortest.[10]

The history of and the way that institutions promote themselves can affect the attitude minority patients have to them and so their likelihood of using them. For example, the view of American hospital services by ethnic minority patients has been tarnished by the American Medical Association's support for segregated wards until the 1960s.[11]

The experience of illness is a culturally shaped phenomenon. The monitoring of change, the understanding of symptoms, the language used to present symptoms, and the fears that accompany symptoms are all suffused with cultural interpretations.[12] Cultural differences in displaying distress are most obviously seen in culture-bound syndromes but are present in the content of delusions and somatic presentation of distress. Somatic symptoms are located in multiple systems of meaning that serve diverse psychological and social functions. In one study, the experience of neurotic patients in India was labeled depressive by clinicians using DSM-IIIR, whereas the patients emphasized their somatic experience.[13] Such discrepancies between professional theory and patients' experience may have an impact on the recognition and treatment of psychiatric disorder. For example, ethnic difference between the doctor and the patient or linguistic/communication problems had previously been offered as reasons for British general practitioners missing depression in South Asian women. However, South Asian doctors are also more likely to miss

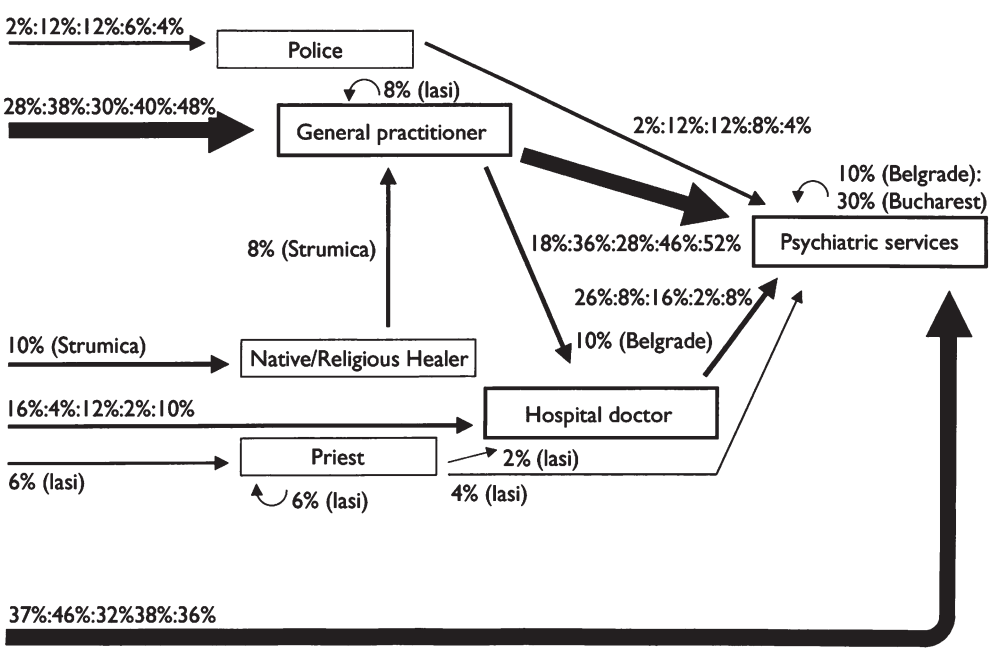


Fig. 7.3.1 Pathways to psychiatric care in Belgrade, Bucharest, Iasi, Strumica and Zagreb. Percentage of those taking each step for each centre respectively, for carers involved in more than 5 per cent of pathways. Steps occurring in only one or two centres are indicated as a single figure followed by the centre name in brackets. (Curved arrows above carer boxes indicate recursive pathways, where patients have gone from one to another of the same type of carer). Reproduced from Gater, R., Jordanova, V., Maric, N. *et al.* (2003). Pathways to psychiatric care in Eastern Europe. *British Journal of Psychiatry*, **186**, 529–35, copyright 2003, The Royal College of Psychiatrists.

depression in South Asian women.[14] Thus, rather than 'ethnic match', the culture of medicine may be the important determinant of recognition of depression in this group and its treatment. Longer delays in first contact with a mental health professional,[10] reduce likelihood of successful drug treatment,[15] and poorer outcomes[16] have been reported in groups of Asian patients. The mismatch between patient experience and culture of medicine appears ubiquitous. In Turkey, psychiatric patients with somatic presentations were shown to have longer delays because they went to see hospital doctors first before being referred to a psychiatrist.[17] A study in Nigeria found that a large proportion of depressed patients initially received another diagnosis because of somatic presentation.[18] However, emphasis on somatic experience on the part of the patient does not preclude the patient's recognition of psychological factors, and addressing culturally shaped experience of illness can help clinicians determine underlying aetiology and understand patients with challenging somatic symptoms.[19]

Beliefs about why the problem has arisen, shape the pathway into care. How important this is depends on how different a culture's models of illness and treatment are from that of the service providers. For example, in a study of help-seeking behaviour of families of patients with schizophrenia in India, most of those who believed in supernatural causation consulted indigenous healers first and those who identified schizophrenia as a medical problem consulted practitioners of modern medicine.[20] Similarly, a study in Ghana found that a perceived supernatural cause of mental health problems was associated with a marked reduction in the likelihood of consulting a mental hospital facility.[21] Surveys among Chinese Americans and Mexican Americans show that[22,23] the association between help-seeking, service and acculturation is mediated, amongst others, by beliefs and explanatory models of psychiatric symptoms. Patient satisfaction is highest if the explanatory model of the patient is matched with that of the service provider.[24] Mental health-care professionals are also prone to differences in explanatory models. For example, a comparison between mental health-care professionals in Saudi Arabia and the UK revealed that the staff in the UK believed in a greater range of possible causes and diagnoses for auditory hallucinations than staff in Saudi Arabia. Differences in belief were associated with different expectations regarding efficacy of possible treatments.[25]

Pathways to care are shaped by public opinion and stigma. Such factors can also contribute to the resources that a society will give, and where and by whom treatment is given. A population survey in Germany showed that the lay public generally held psychotherapy in high esteem and the vast majority of respondents rejected pharmacotherapy for psychological problems. Psychoanalysis was the most popular approach in the western part of Germany but in eastern Germany the preference was for group therapy.[26]

Similarly, because psychiatric practice remains to a large degree opinion-based within Europe, differences exist between national samples of psychiatrists with regard to the diagnosis, aetiology, treatment, and outcome of psychiatric disorders such as schizophrenia. Cultural divergence is especially evident with regard to differential emphasis on psychodynamic and biological approaches,[27] and the level of availability of psychotherapy within the health service in European countries appears to be associated with the dominant therapeutic culture of psychiatrists.[28]

The quality of mental health legislation, and, perhaps more importantly, the degree to which correct implementation is enforced have an effect on care pathways. Increased stigma of psychotic illness and fear of public safety have led to compulsory treatment in the community becoming law in the UK. This could change pathways to and through care.[29] In the Netherlands, a weak and impractical act has been criticized for denying patients with severe mental illness the treatment they need. In Spain, to date no specific mental health act exists. In countries like the United Kingdom and France, involuntary admission is in practice, a largely clinical decision, and therefore more easily to put into effect compared with, for example, The Netherlands and most of Germany where the decision is ultimately made by the judiciary. In the United Kingdom and France, the police may bypass medical referral and take people behaving strangely in a public place directly to a psychiatric hospital. This has been shown to be an important pathway for certain ethnic minority groups, such as African-Caribbeans in the United Kingdom.[30]

Filters on the pathway to care and service use

Filter permeability and service use

The majority of people in psychological distress never present to formal health services. If formal help is sought, the likelihood of receiving such help is influenced by the permeability of a number of filters (see Chapter 7.8). For example, once the decision to seek formal help is taken, actual receipt of help depends on the ability of the professional at the first port of call (usually primary care services) to recognize the presence of mental disorder. The permeability of subsequent filters determines the rate with which patients move from primary care service to specialist mental health outpatient services, and from specialist outpatient to hospital-based psychiatric services and back down the hierarchy.

(a) Cross-ethnic differences

There are important cross-ethnic and cross-national differences in the permeability of filters along the pathway to care. Recognition of symptoms by mental health professionals is an important factor. Recognition of distress is dependent on the way symptoms are elicited. For instance, the recognition by primary care doctors of psychiatric disorder in African-Caribbeans and South Asians in the United Kingdom has been shown to be poor.[31] There are a number of reasons for this including differences between the symptoms expected by doctors and those presented.[12] Despite the fact that they visit their general practitioner more often, women of South Asian origin with depression are less likely to be diagnosed and treated than white British women. Detection of depression depends on the doctor's skill but also whether the patient tells the general practitioner about her worries. Those who believed that a doctor was the right person to deal with depression were found to be more likely to disclose information and more likely to be diagnosed and treated.[12] South Asian women are less likely than white British women to think that a doctor was the right person to deal with depression.[32] A one-year follow-up of the sample of the Epidemiologic Catchment Area Program revealed that African-Americans, Hispanics, and other minorities were much less likely to have consulted with a professional in the specialized mental health care sector than white people. The odds of consultation in African-Americans was less than one-quarter of that in white people even after adjustment for confounders.[33] Similarly,

African-American children and adolescents may also remain under-treated although they may have higher levels of symptomatology.[34] Differences between American ethnic groups are also apparent in populations with identical insurance coverage.[35] Hence, these findings[36] suggest low permeability of filters on the pathway to mental health care. Reasons for this may include that African-Americans are less inclined to seek professional help because of increased tolerance to depressive symptoms, but also because of fear of hospital admission.[37] In comparison to all other ethnic groups, African-Americans make more use of emergency rooms for routine psychiatric care.[38]

Despite the low permeability of the filters on the pathway to care there is an over-representation of African-Caribbeans and African-Americans at the level of hospital-based psychiatric services. Possible mechanisms for this include failure of community services to engage mentally ill African-Caribbean men[39] and bypass of the usual filters by, for example, compulsory admission to hospital with or without police involvement.[40] It has been shown that police involvement and compulsory admission to hospital is strongly associated with the absence of general practitioner involvement.[41] Levels of perceived violence and rates of involuntary admission may be due to stereotyped attitudes of the police and mental health professionals or may be in part due to a higher rate of presentation of psychosis that is superimposed on intact premorbid personalities. It has been suggested that reactive forms of psychotic illness in African-Caribbeans are wrongly labelled as schizophrenia.[42] Higher functioning, less withdrawn patients may be perceived as constituting a higher risk by police and mental health professionals. Another factor is that, despite low rates of recognition by general practitioners, African-Caribbeans are most likely to be referred on to a specialist, followed by white people and then people from south Asia even when socioeconomic class and diagnosis are taken into account.[31]

(b) Cross-national differences

Cross-national differences in filter permeability and service use are difficult to examine. Service organization plays an important role. For example, compared with south Manchester, in the United Kingdom, closer integration between community and inpatient psychiatric services in south Verona, Italy, resulted in a greater permeability of the filter between inpatient and community care, as evidenced by higher hospital admission rates and shorter lengths of stay. Conversely, greater permeability of the general practitioner referral filter in south Manchester resulted in more referrals, and therefore higher treated incidence and prevalence rates of psychiatric disorder.[43] Similarly, compared with the more institution-based system in Groningen, in The Netherlands, the south Verona community-based system provided for a higher degree of continuity of care across services for patients with schizophrenia.[44]

Because of convergence in methodology, fascinating material on cross-national differences in the dynamic balance between psychological distress, need for care, and actual treatment received is now available from large prevalence surveys of representative samples in different countries. Comparative data are available on three large population studies in three countries. In the United States and Ontario (Canada), samples representative of all non-institutionalized individuals aged 18 to 54 years were examined using the Composite International Diagnostic Interview (CIDI) in 1990 and samples of people aged 18 to 64 years in The Netherlands were examined in 1996 (Nemesis Study).[45–47]

The three studies show that help-seeking rates among individuals with a diagnosable disorder vary widely. The contact rate with any type of formal or informal service was lowest in the United States (33.9 per cent) and not far from twice as high in The Netherlands (56.7 per cent) (Table 7.3.1). Among individuals with a diagnosable disorder, ambulatory service use in the general medical sector was much higher in Ontario and The Netherlands than in the United States, especially among those with more severe comorbid

Table 7.3.1 Prevalence of CIDI disorder and service use in three countries

		Service use			
	12-month prevalence	General medical[a]	Ambulatory mental health[b]	Any service	Perceived need for care in non-users of professional services
National Comorbidity Survey					
No disorder[c]	54.0	2.6	2.6	7.6	
One disorder	1S.7	5.6	8.3	18.8	8.4
Two disorders	12.0	10.2	18.6	33.9	
Ontario					
No disorder'	65.6	1.5	1.2	3.3	
One disorder	127	9.7	6.3	17.8	5.4
Two disorders	5.9	24.3	23.5	39.4	
Nemesis					
No disorder[c]	58.8	3.8	1.8	6.1	
One disorder	15.3	15.5	7.8	22.5	3.8
Two disorders	8.1	42.8	30.3	56.7	

[a] Non-psychiatrist physician in any setting or allied health professional in a general medical setting.

[b] Psychiatrist or psychologist in any setting or allied health professional (e.g. nurse, social worker) in a psychiatric or addiction treatment setting.

[c] No lifetime history of any disorder at all.

Weighted data from Kessler *et al.*[45] and Bijl *et al.*[47]

disorders. Individuals in the United States with more severe disorders were less likely to use services in the health-care sector as a whole, but if treatment from self-help and other sources is included, the difference between Ontario and Nemesis on the one hand, and the United States on the other, is attenuated. This suggests that low permeability of the primary care access and primary care referral filters in the United States may lead to increased use of non-professional services to fill the gap.

Table 7.3.1 shows that the higher the contact rates with professional services of individuals with a CIDI diagnosis, the lower the number of individuals who were not using professional services but felt they were in need of such help (level of unmet need) (Table 7.3.1). Such differences in the population level of unmet need are important from the point of view of public health. For example, if 90 per cent in a population of 200 million are non-users of professional services for mental health problems, then the difference between 8.4 per cent (United States) and 3.8 per cent (The Netherlands) in perceived need is a difference of 9.2 million individuals.

The substantial differences in mental health care provided by the general practitioner are likely to have an equally substantive impact on the likelihood of receiving appropriate management. Thus, the proportion of patients with major depression (as defined in DSM-IIIR) in the previous 12 months who received appropriate medication management, defined as a combination of antidepressant medication use and four or more visits to any health-care provider within the previous 12 months, was much higher in Ontario (14.9 per cent) than in the United States (7.3 per cent). This difference was especially marked for the lowest income groups in the two countries. Individuals in the lowest income groups in the United States were found to be 7.5 times less likely to make contact with either general or specialty health-care providers than their peers in Ontario. For the highest income groups, however, contact rates differed only by a factor 2.1.[48] These data suggest that economic barriers play an important role in determining the permeability of the filters on the pathway to care. In a United Kingdom national survey, 16 per cent of patients with a depressive episode in the past week according to the Revised Clinical Interview Schedule were current users of antidepressant medication.[49]

Because the majority of the population does not have a mental disorder, even a small degree of service use by this large segment of the population will take up a considerable part of the total capacity of mental health services. Thus, Katz *et al.*[50] noted that because of the relatively high rates of perceived need for care and help-seeking among individuals without a CIDI diagnosis in the United States, total mental health outpatient service use was higher in that country than in Ontario. Although diagnosis is only an imperfect indicator of need for care, the results nevertheless suggest that the mismatch between need and care in the population is greater in the United States than in Ontario.

A long-standing debate exists whether and how financing of mental health care can be used to maximize the fit between need and care in the population. A frequently expressed concern is that universal coverage will lead to an increase of people with little need using services of unproven value. The opposite argument, however, is that limitations in coverage will result in service use that is poorly matched to need. Although it is thought that differences in type of insurance system have an impact on demand and utilization of mental health services,[51] systematic comparisons between countries have been lacking. The systems of coverage in the United States, Canada, and The Netherlands are different in many respects. In Ontario, universal and relatively comprehensive coverage for mental health services exists, with no or minimal limits on inpatient stays or outpatient visits for mental health services, and minimal patient cost sharing. In The Netherlands, almost all mental health care is covered under the Exceptional Medical Expenses Act, and is available to the entire population. A comprehensive range of public services exists, with few supply-side controls. In the United States, at least 16 per cent of the population is uninsured, and even for the insured mental health coverage is increasingly limited. Although the public health system provides mental health care at little or no cost to the poor and the uninsured, supply-side controls severely and increasingly limit access. Therefore, the results of the comparisons between the three countries do not support the frequently expressed reservation that expansion of insurance coverage for mental health disorders results in an increase in unnecessary use of services. Of the three countries considered, those with broad mental health coverage actually treated a similar number or more people with severe mental illness, but less people who never had a history of mental illness.

Treatment and response

Culture has an important influence on the type of service received. There are reports that African-Caribbean and African-American patients receive antipsychotic medications in a higher prescribed dose, and more frequent use of injectable preparations is made.[30,52–57] Asian patients have been reported to receive lower doses.[58,59] Some ethnic minority groups may receive less information about side-effects,[60] which may result in less vigilance with regard to onset of problems such as tardive dyskinesia. In the United States and the United Kingdom, ethnic minority patients are less likely to receive psychological treatment.[61,62] In the United States, African-American children have substantially lower rates of receiving methylphenidate.[63,64] Such differences may be related to differences in explanatory models of African-American parents, and differences in the rate with which African-American parents receive appropriate information about attention deficit–hyperactivity disorder from the doctor.[65]

There are several important considerations with regard to outcome in relation to cultural variables.[66] The first is that services offered may not be equally efficacious for different groups of people. For example, Chinese, Japanese, Filipino, Korean, and Southeast Asian Americans who were treated in the same setting in Los Angeles County showed different outcomes. Filipinos were under-represented in the system, whereas Southeast Asians were over-represented and had higher rates of service utilization. Despite this, Southeast Asians showed less improvement than the other groups, even after controlling for diagnosis and initial level of functioning.[67] The second is that treatment uptake may differ between groups.[33] In the United Kingdom and the United States, uptake of treatment with antipsychotic medication may be higher in white patients, though the overall influence of ethnicity remains small.[68] The third consideration is that the expectations about the desired endpoint of treatment may not be the same for different groups of people. For example, the expectations of British Asian and white people relating to the process and outcome of a psychological intervention were shown to be different in one study.[69]

Outcome may improve if therapists receive information about their clients' cultural background and expectations before treatment.[70] Perhaps the most important consideration is that outcome is a multidimensional concept defying summary statements. For example, clinical outcome in terms of usual symptom severity and risk of self-harm may be better in African-Caribbean patients with psychosis as compared with white people, yet risk of imprisonment and compulsory admission may be greater,[71] as may be the frequency of relapse[72] and the rate of dissatisfaction with services.[73]

However, it is interesting to note that a recent pan-European study has concluded that despite the fact that differences in health-care systems may affect service provision and cost, the impact of such differences on outcome may be less marked. More work needs to be undertaken on cultural differences in care pathways and treatment to see if they change outcomes.[74] Moreover, at an individual level, recent work has failed to show differences in the duration of untreated psychosis between African Caribbeans and whites in the UK,[75] even though they have different pathways to care at first admission indicating that the impact of cultural or ethnic differences in pathways can be difficult to predict.

Pathways out of care

Pathways out of care are complex and have multiple influences. An important influence is the pathway into care because the relationship set up with services at first contact determines, in part, the trajectory of the clinical career of a patient. Moreover, in specific circumstances, such as where the criminal justice system is involved, certain responsibilities may be placed on services which constrain the ability to discharge patients. There has been little research which aims to understand ethnic differences in pathways out of care.[76] This reflects the fact that the focus of studies to date has been to document any inequalities in access to care.

Pathways out of statutory sector care are influenced by the effectiveness of treatment strategies, the illness models and treatment preferences of patients, the structure and funding of clinical services, the availability of and access to nonstatutory sector support and also the socio-cultural context in which people live.

Effectiveness is a combination of the efficacy of a treatment and the real world context which influences outcomes. Though there may be differences in the efficacy of drugs by racial group, differences in compliance are more likely to be important,[77] and there are well documented differences in adherence to treatment programs which may lead to differences in rates of patients leaving care.[78]

Adherence has been linked to differences in illness models but in psychosis the level of discord between patient and treatment service illness models are not consistently associated with differences in outcome.[79] This may reflect the fact that preferred alternative treatment outside statutory mental health services are not available or that services are more coercive with patients who do not fully subscribe to their model.

The structure and funding of clinical services is important. Filters out of care can be porous—an example of this is Ontario, where rehabilitation is mainly offered by local community based organizations which may have better links with the community and so make transition out of care easier. They also have limited responsibility for their clients compared to the UK and may be less risk averse. Risk averse service cultures can lead to delay in discharge from care.[80] Moreover, putting the onus on risk can mean that ethnic groups who are considered more risky have particular difficulty in getting out of care.

Prevailing law also shapes pathways out of care. Such laws may be applied differently to some ethnic groups. For instance, in New Zealand, the Maori population is more likely to be detained on community treatment orders thus delaying movement out of statutory service care.[81]

The availability of nonprofessional support and the structure of some cultural groups may affect pathways. Those of South Asian origin with severe mental illness spend less time in hospital and outpatient care. Family support is cited as the reason for this. Socio-cultural mores are important but also geographic concentration; dispersed refugee and asylum groups with limited access to community support may stay longer in services than would be expected from their symptoms. But socially cohesive groups can be a double-edged sword. One study has demonstrated higher rates of re-admission in areas with high social capital which arguably could be a result of high levels of health norm policing and low levels of tolerance for deviance.[82]

Socio-economic factors also need to be kept in mind when considering the capacity of communities to offer care. The decreased resources available to low income groups will be important in determining the level of burden they will be able to accept.

Further information

World Health Organization: up to date information on the health with a good archive, research tools and data as well as a section which summarizes health systems organization and funding for each of the countries in the world. http://www.who.int/en/

World Association of Cultural Psychiatry: free newsletter and journal on cross cultural psychiatry issues at: http://www.waculturalpsychiatry.org/

National Institute for Clinical Excellence in the UK has developed a pathways approach to improving mental health services for people with schizophrenia backed by algorithms and research evidence: http://www.schizophreniaguidelines.co.uk/nice_implementation/pathways_to_care.php.

References

1. Rogler, L.H. and Cortes, D.E. (1993). Help-seeking pathways: a unifying concept in mental health care. *American Journal of Psychiatry*, **150**, 554–61.
2. Gater, R., de Almeida e Sousa, B., Barrientos, G., *et al.* (1991). The pathways to psychiatric care: a cross-cultural study. *Psychological Medicine*, **21**, 761–74.
3. Üstün, T.B. and Sartorius, N. (eds.) (1995). *Mental illness in genera health care*. Wiley, New York.
4. Gater, R., Jordanova, V., Maric, N., *et al.* (2003). Pathways to psychiatric care in eastern Europe. *British Journal of Psychiatry*, **186**, 529–35.
5. Bhui, K., Stansfeld, S., Hull, S., *et al.*(2003). Ethnic variations in pathways to and use of specialist mental health services in the UK. Systematic review. *British Journal of Psychiatry*, **182**, 105–16.
6. Wheeler, A., Robinson, E., Robinson, G., *et al.* (2005). Admissions to acute psychiatric inpatient services in Auckland, New Zealand: a demographic and diagnostic review. *NZ med J*, **118** (1226): u1752.
7. Malla, A. and Norman, R.M.G. (1988). Involuntary admissions in a Canadian province: the influence of geographic and population factors. *Social Psychiatry and Psychiatric Epidemiology*, **23**, 247–51.
8. Engberg, M. (1991). Involuntary commitment in Greenland, the Faroe Islands and Denmark. *Acta Psychiatrica Scandinavica*, **84**, 353–6.

9. Lin, T.Y., Tardiff, K., Donetz, G., *et al.* (1978). Ethnicity and patterns of help-seeking. *Culture in Medicine and Psychiatry*, **2**, 3–13.
10. Lin, K.M., Inui, T.S., Kleinman, A.M., *et al.* (1982). Sociocultural determinants of the help-seeking behavior of patients with mental illness. *Journal of Nervous and Mental Disorders*, **170**, 78–85.
11. King, G. (1996). Institutional racism and the medical health complex: a conceptual analysis. *Ethnicity and Disease*, **6**, 30–46.
12. Kleinman, A. (1991). *Rethinking psychiatry*. Free Press, New York.
13. Weiss, M.G., Raguram, R., and Channabasavanna, S.M., *et al.* (1995). Cultural dimensions of psychiatric diagnosis. A comparison of DSM-IIIR and illness explanatory models in south India. *British Journal of Psychiatry*, **166**, 353–9. Erratum: *British Journal of Psychiatry*, **167**, 119 (1995).
14. Jacob, K.S., Bhugra, D., Lloyd, K.R., *et al.* (1998). Common mental disorders, explanatory models and consultation behaviour among Indian women living in the UK. *Journal of the Royal Society of Medicine*, **91**, 66–71.
15. Cornwell, J. (1998). Do general practitioners prescribe antidepressants differently for South Asian patients? *Family Practice*, **15**, S16–18.
16. Ying, Y.-W. and Hu, L.-T. (1994). Public outpatient mental health services: use and outcome among Asian Americans. *American Journal of Orthopsychiatry*, **64**, 448–55.
17. Kilic, C., Rezaki, M., Üstün, T.B., *et al.* (1994). Pathways to psychiatric care in Ankara. *Social Psychiatry and Psychiatric Epidemiology*, **29**, 131–6.
18. Makanjuola, J.D. and Olaifa, E.A. (1987). Masked depression in Nigerians treated at the Neuro–Psychiatric Hospital Aro, Abeokuta. *Acta Psychiatrica Scandinavica*, **76**, 480–5.
19. Handelman, L. and Yeo, G. (1996). Using explanatory models to understand chronic symptoms of Cambodian refugees. *Family Medicine*, **28**, 271–6.
20. Banerjee, G. and Roy, S. (1998). Determinants of help-seeking behaviour of families of schizophrenic patients attending a teaching hospital in India: an indigenous explanatory model. *International Journal of Social Psychiatry*, **44**, 199–214.
21. Fosu, G.B. (1995). Women's orientation toward help-seeking for mental disorders. *Social Science and Medicine*, **40**, 1029–40.
22. Wells, K.B., Hough, R.L., Golding, J.M., *et al.* (1987). Which Mexican-Americans underutilize health services? *American Journal of Psychiatry*, **144**, 918–22.
23. Ying, Y.-W. and Miller, L.S. (1992). Help-seeking behavior and attitude of Chinese Americans regarding psychological problems. *American Journal of Community Psychology*, **20**, 549–56.
24. Callan, A. and Littlewood, R. (1998). Patient satisfaction: ethnic origin or explanatory model? *International Journal of Social Psychiatry*, **44**, 1–11.
25. Wahass, S. and Kent, G. (1997). A cross-cultural study of the attitudes of mental health professionals towards auditory hallucinations. *International Journal of Social Psychiatry*, **43**, 184–92.
26. Angermeyer, M.C. and Matschinger, H. (1996). Public attitude towards psychiatric treatment. *Acta Psychiatrica Scandinavica*, **94**, 326–36.
27. van Os, J., Galdos, P., Lewis, G., *et al.* (1993). Schizophrenia sans frontieres: concepts of schizophrenia among French and British psychiatrists. *British Medical Journal*, **307**, 489–92.
28. van Os, J. and Neeleman, J. (1994). Caring for mentally ill people. *British Medical Journal*, **309**, 1218–21.
29. Department of Health (2007) regulatory impact assessment of mental health bill available at dh.gov.uk/prod_consum_dh
30. Dunn, J. and Fahy, T.A. (1990). Police admissions to a psychiatric hospital. Demographic and clinical differences between ethnic groups. *British Journal of Psychiatry*, **156**, 373–8.
31. Rawaf, S. and Bahl, V. (1998). *Assessing health needs of people from minority ethnic groups*. Royal College of Physicians of London.
32. Lloyd, K.R., Jacob, K.S., Patel, V., *et al.* (1998). The development of the Short Explanatory Model Interview (SEMI) and its use among primary-care attenders with common mental disorders. *Psychological Medicine*, **28**, 1231–7.
33. Gallo, J.J., Marino, S., Ford, D., *et al.* (1995). Filters on the pathway to mental health care. II. Sociodemographic factors. *Psychological Medicine*, **25**, 1149–60.
34. Cuffe, S.P., Waller, J.L., Cuccaro, M.L., *et al.* (1995). Race and gender differences in the treatment of psychiatric disorders in young adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, **34**, 1536–43.
35. Scheffler, R.M. and Miller, A.B. (1989). Demand analysis of mental health service use among ethnic subpopulations. *Inquiry*, **26**, 202–15.
36. Brown, D.R., Ahmed, F., Gary, L.E., *et al.* (1995). Major depression in a community sample of African Americans. *American Journal of Psychiatry*, **152**, 373–8.
37. Sussman, L.K., Robins, L.N., Earls, F., *et al.* (1987). Treatment-seeking for depression by black and white Americans. *Social Science and Medicine*, **24**, 187–96.
38. Neighbors, H.W. (1986). Ambulatory medical care among adult black Americans: the hospital emergency room. *Journal of the National Medical Association*, **78**, 275–82.
39. Bhui, K., Brown, P., Hardie, T., *et al.* (1998). African-Caribbean men remanded to Brixton Prison—psychiatric and forensic characteristics and outcome of final court appearance. *British Journal of Psychiatry*, **172**, 337–44.
40. Morgan, C., Mallett, R., Hutchinson, G., *et al.* (2005). On behalf of the ÆSOP Study Group Pathways to care and ethnicity II; Source of referral and help seeking. A Report From the ÆSOP (Aetiology and Ethnicity in Schizophrenia and Other Psychoses) Study. *British Journal of Psychiatry* ,**186**, 290–6.
41. Cole, E., Leavey, G., King, M., *et al.* (1995). Pathways to care for patients with a fi rst episode of psychosis. A comparison of ethnic groups. *British Journal of Psychiatry*, **167**, 770–6.
42. Littlewood, R. and Lipsedge, M. (1978). Migration, ethnicity and diagnosis. *Psychiatric Clinics of Basel*, **11**, 15–22.
43. Amaddeo, F., Gater, R., Goldberg, D., *et al.* (1995). Affective and neurotic disorders in community-based services: a comparative study in south Verona and south Manchester. *Acta Psychiatrica Scandinavica*, **91**, 386–95.
44. Sytema, S., Micciolo, R., Tansella, M., *et al.* (1997). Continuity of care for patients with schizophrenia and related disorders: a comparative south Verona and Groningen case-register study. *Psychological Medicine*, **27**, 1355–62.
45. Kessler, R.C., Frank, R.G., Edlund, M., *et al.* (1997). Differences in the use of psychiatric outpatient services between the United States and Ontario. *New England Journal of Medicine*, **336**, 551–7.
46. Katz, S.J., Kessler, R.C., Frank, R.G., *et al.* (1997). The use of outpatient mental health services in the United States and Ontario: the impact of mental morbidity and perceived need for care. *American Journal of Public Health*, **87**, 1136–43.
47. Bijl, R.V., Ravelli, A., Van Zessen, G., *et al.* (1998). Prevalence of psychiatric disorder in the general population: results from The Netherlands mental health survey and incidence study. *Social Psychiatry and Psychiatric Epidemiology*, **33**, 587–95.
48. Katz, S.J., Kessler, R.C., Lin, E., *et al.* (1998). Medication management of depression in the United States and Ontario. *Journal of General Internal Medicine*, **13**, 77–85.
49. Meltzer, H., Gill, B., Petticrew, M., *et al.* (1995). *OPCS surveys of psychiatric morbidity. Report 1: the prevalence of psychiatric morbidity among adults aged 16–64 living in private households in Great Britain*. HMSO, London.
50. Katz, S.J., Kessler, R.C., Frank, R.G., *et al.* (1997). The use of outpatient mental health services in the United States and Ontario: the impact of mental morbidity and perceived need for care. *American Journal of Public Health*, **87**, 1136–43.

51. Frank, R.G. and McGuire, T.G. (1986). A review of studies of the impact of insurance on the demand and utilization of specialty mental health services. *Health Services Research*, **21**, 241–65.
52. Citrome, L., Levine, J., Allingham, B., *et al.* (1996). Utilization of depot neuroleptic medication in psychiatric inpatients. *Psychopharmacology Bulletin*, **32**, 321–6.
53. Segal, S.P., Bola, J.R., Watson, M.A., *et al.* (1996). Race, quality of care, and antipsychotic prescribing practices in psychiatric emergency services. *Psychiatric Services*, **47**, 282–6.
54. Lawson, W.B. (1996). Clinical issues in the pharmacotherapy of African-Americans. *Psychopharmacology Bulletin*, **32**, 275–81.
55. Price, N., Glazer, W., Morgenstern, H., *et al.* (1985). Demographic predictors of the use of injectable versus oral antipsychotic medications in outpatients. *American Journal of Psychiatry*, **142**, 1491–2.
56. Shubsachs, A.P., Huws, R.W., Close, A.A., *et al.* (1995). Male Afro-Caribbean patients admitted to Rampton Hospital between 1977 and 1986—a control study. *Medicine, Science and Law*, **35**, 336–46.
57. Strakowski, S.M., Shelton, R.C., Kolbrener, M.L., *et al.* (1993). The effects of race and comorbidity on clinical diagnosis in patients with psychosis. *Journal of Clinical Psychiatry*, **54**, 96–102.
58. Rosenblat, R. and Tang, S.W. (1987). Do oriental psychiatric patients receive different dosages of psychotropic medication when compared with occidentals. *Canadian Journal of Psychiatry*, **32**, 270–4.
59. Bond, W.S. (1991). Ethnicity and psychotropic drugs. *Clinical Pharmacology*, **10**, 467–70.
60. Benson, P.R. (1984). Drug information disclosed to patients prescribed antipsychotic medication. *Journal of Nervous and Mental Disease*, **172**, 642–53.
61. Olfson, M. and Pincus, H.A. (1994). Outpatient psychotherapy in the United States. II. Patterns of utilization. *American Journal of Psychiatry*, **151**, 1289–94.
62. Littlewood, R. and Lipsedge, M. (1997). *Aliens and alienists*. Routledge, New York.
63. Zito, J. M., Safer, D. J., dos Reis, S., *et al.* (1997). Methylphenidate patterns among Medicaid youths. *Psychopharmacology Bulletin*, **33**, 143–7.
64. Zito, J. M., Safer, D. J., dos Reis, S., *et al.* (1998). Racial disparity in psychotropic medications prescribed for youths with Medicaid insurance in Maryland. *Journal of the American Academy of Child and Adolescent Psychiatry*, **37**, 179–84.
65. Bussing, R., Schoenberg, N.E., Perwien, A.R., *et al.* (1998). Knowledge and information about ADHD: evidence of cultural differences among African-American and white parents. *Social Science and Medicine*, **46**, 919–28.
66. McKenzie, K. and Murray, R.M. (1998). *Risk factors for mental illness in African-Caribbeans.* Department of Health Conference on Culture, Ethnicity and Mental Health. Department of Health, London.
67. Ying, Y.W. and Hu, L.T. (1994). Public outpatient mental health services, use and outcome among Asian Americans. *American Journal of Orthopsychiatry*, **64**, 448–55.
68. Bebbington, P.E. (1995). The content and context of compliance. *International Clinics in Psychopharmacology*, **5**, 41–50.
69. Balabil, S. and Dolan, B. (1992). A cross-cultural evaluation of expectations about psychological counselling. *British Journal of Medical Psychology*, **65**, 305–8.
70. Yamamoto, J., Acosta, F. X., Evans, L.A., *et al.* (1984). Orienting therapists about patients' needs to increase patient satisfaction. *American Journal of Psychiatry*, **141**, 274–7.
71. McKenzie, K., van Os, J., Fahy, T., *et al.* (1995). Psychosis with good prognosis in Afro-Caribbean people now living in the United Kingdom. *British Medical Journal*, **311**, 1325–8.
72. Birchwood, M., Cochrane, R., Macmillan, F., *et al.* (1992). The influence of ethnicity and family structure on relapse in first-episode schizophrenia. A comparison of Asian, Afro-Caribbean, and white patients. *British Journal of Psychiatry*, **161**, 783–90.
73. Parkman, S., Davies, S., Leese, M., *et al.* (1997). Ethnic differences in satisfaction with mental health services among representative people with psychosis in south London: PRiSM Study 4. *British Journal of Psychiatry*, **171**, 260–4.
74. Becker, T., Kilian, R. (2006). Psychiatric services for people with severe mental illness across western Europe: what can be generalized from current knowledge about differences in provision, cost and outcomes of mental health care *Acta Psychiatr Scan,* Suppl 429, 9–16
75. Morgan, C., Fearon, P., Hutchinson, G., AESOP Study Group., *et al.* (2006) Duration of untreated psychosis and ethnicity in the AESOP fi rst-onset psychosis study. *Psychological Medicine,* **36**(2), 239–47
76. EPIC group (2007). ww.wolfson.qmul.ac.uk/psychiatry/epic/docs/SRper cent20Keyper cent20Summary.pdf accessed
77. Lin, K.M., Anderson, D., Poland, R.E., *et al.* (1995). Ethnicity and psychopharmacology. Bridging the gap. *Psychiatr Clin North Am*. **18**(3), 635–47.
78. Sue, S., Fujino, D. C., Hu, L. T., *et al.* (1991). Community mental health services for ethnic minority groups: a test of the cultural responsiveness hypothesis. *J Consult Clin Psychol*, **59**(4), 533–40.
79. Rosemarie McCabe, PhD and Stefan Priebe, MD. (2004). Explanatory models of illness in schizophrenia: comparison of four ethnic groups *The British Journal of Psychiatry,* **185**, 25–30
80. Lelliott, P. and Audini, B. (2003). Trends in the use of Part II of the Mental Health Act 1983 in seven English local authority areas. Br J Psychiatry, **182**, 68–70.
81. Gibbs, A., Dawson, J., Forsyth, H., (2004). Maori experience of community treatment orders in Otago, New Zealand. *Aust N Z J Psychiatry*, **38**(10), 830–5s
82. McKenzie, K. (2000). Neighborhood, safety and mental health outcomes. Postingnumber 28, 2000. Social Capital Lets Talk. Socialcapital@tome. worldbank.org

7.4

Primary prevention of mental disorders

J. M. Bertolote

Despite the demonstration of the possibility of preventing some forms of mental disorders, many mental health professionals continue to underestimate the possibilities of primary prevention in their field. This is due to:

1 a lack of clear concepts when referring to this issue;

2 the fact that the effective prevention of mental and neurological disorders often falls outside the usual remit of mental health professionals (in many cases it falls outside the health sector altogether).

These two factors are discussed below, in addition to an indication of actions which effectively prevent some forms of mental disorders.

Prevention

In the late 1950s, Leavell and Clark[1] proposed a three-level concept of prevention (primary, secondary, and tertiary), covering almost all medical actions. Their innovative approach must be understood in relation to what they also called the horizon of the natural history of the disease process: under natural circumstances, a disease will proceed from a prepathological period through its early stages, evolving either to partial or full recovery (or cure), or to death; in the case of partial recovery, there may be chronification or sequelae. Prevention, in this sense, refers not only to the appearance of the disease but also to any further worsening or complication of it once it has appeared.

Primary prevention

The primary prevention level covers what is otherwise referred to as both health promotion and specific protection, and is best exemplified by, for example, adequate nutrition and immunization against specific diseases by vaccines. Whereas, in this example, adequate nutrition is totally non-specific (it contributes to enhancing the overall resistance to several diseases without conferring any specific protection against any), vaccination is highly specific in relation to a single condition. Rational-specific protection is fully dependent on a reasonable knowledge of the aetiology of the disease (or, at least, its mode of transmission) in order to be effective.

Secondary prevention

The secondary prevention level refers to early detection and treatment of diseases. Usually the bulk of the medical activity, its main preventive goal is to avoid chronicity and the establishment of irreversible sequelae. It is dealt with more specifically in Part 6 of this book.

Tertiary prevention

The tertiary prevention level largely corresponds to rehabilitation. It enters into operation once the disease process has been established and aims at reducing as much as possible damages caused by the disease process, preserving intact functions, and restoring and/or compensating impaired functions, disabilities, and handicaps.

On one hand, this conceptual model was highly instrumental in providing an impetus towards preventive activities in the medical field as a whole but, on the other hand, it so popularized the term prevention that it almost lost its powerful message. Therefore, it is important to retain the idea of primary prevention as a synonym of specific protection, referring to methods designed to avoid the occurrence of a specific disorder or groups of disorders. It comprises those measures applicable to a particular disease or group of diseases in order to intercept their causes before they affect people, and should be differentiated not only from treatment and rehabilitation, but also from mental health promotion.

The main obstacle for the prevention of many mental disorders is the limited knowledge about their aetiology. Admittedly, there are very promising and exciting hypotheses concerning the causes of three mental disorders which represent the greatest burden, namely, depression, schizophrenia, and dementia. They are, nevertheless, nothing more than hypotheses. The most successful examples of prevention of diseases refer to those whose aetiology (cause and/or mode of transmission) is relatively well-known. There are historical examples of the prevention of some conditions based on false assumptions about or without a good knowledge of their aetiology (for example, the eradication of malaria in ancient Rome, and the control of the London cholera epidemics by John Snow in the nineteenth century). However, it does not seem appropriate for health professionals and scientists to base their actions on chance or false assumptions, even though the result might be opportune to the population.

As implied by the need to intercept causes of a particular disease or groups of diseases (with a common cause), the concept of prevention calls for a high degree of specificity concerning the target condition or conditions. In the medical field, it led to successful programmes for the prevention of, for example, diarrhoeal diseases (such as typhoid), hypertension, coronary heart disease, breast cancer, and unwanted pregnancies, rather than of infectious diseases, cardiovascular diseases, cancer, or obstetrical problems. Unfortunately, in the mental health field there has not been a great concern with the specification of the target condition, and the prevention of 'mental disorders' (as a whole) became a label soon associated with failure and disinterest.

Mental disorders

What is understood as 'mental disorders' comprises a variety of quite diverse clinical conditions in terms of aetiology, symptomatology, clinical course, prognosis, and response to treatment. Therefore, whenever the prevention of mental disorders is referred to, an effort must be made to obtain some precision.

From a nosological point of view, most of the mental disorders are conceptually at a syndromal level; depression, schizophrenia, and dementia are appropriate examples. In this respect, dementia is one step ahead of the other two, in so far as vascular dementia is now clearly differentiated from Alzheimer's disease, with important implications for prevention.

Therefore a strategic shift is necessary in order to obtain greater efficiency in the successful prevention of some mental disorders. The first step is for an effort to be as specific as possible in relation to the target condition: for instance Down syndrome or phenylketonuria instead of intellectual disorder, foetal alcohol syndrome, delirium tremens instead of alcoholism, and vascular dementia and dementia following brain injury instead of dementia in general.

The second step applies to those conditions which cannot be meaningfully broken down into more specific conditions, such as schizophrenia or depression. In these cases, the target is displaced from the appearance of the conditions towards future relapses, once a first episode has occurred; this conveniently applies to schizophrenia, depression, and dependence on alcohol and other drugs.

Finally, there are some violent behaviours, such as suicide, parasuicide, and violence against others, the control (and prevention) of which are largely expected by society to come from the field of mental health. They do not characterize a mental disorder in particular, but are frequently associated with one or more of them. Their prevention, therefore, requires specifically dedicated interventions.

With this wide range of issues considered as mental disorders, it becomes clear that the coverage of their prevention goes well beyond the limits of this chapter. A detailed conceptual approach to the prevention of mental and psychosocial disorders can be found in a recent publication of the World Health Organization (**WHO**).[2]

Prevention of mental disorders

From a practical point of view there are three groups of conditions for which efficient preventive action has been documented.

1 Mental disorders with known aetiology: this mostly includes those disorders demonstrated to have an organic basis, ranging from the 'historical' general paresis and dementing disorders (e.g. vascular dementia, pellagra, and dementias associated with infectious and parasitic diseases such as malaria and HIV infection) to several forms of intellectual disorder (Down syndrome, foetal alcohol syndrome, phenylketonuria, and intellectual disorder due to iodine deficiency).

2 Mental disorders without a well-established aetiology but with a relatively predictable course: these are chronic disorders with a recurrent relapsing fluctuating pattern, such as schizophrenia, mood disorders (unipolar and bipolar), and alcohol dependence syndrome.

3 Psychosocial problems strongly associated with mental disorders: these range from violence (domestic and other) to suicide and staff burnout.

Mental disorders with known aetiology

(a) Infectious diseases

Prevention of this group of disorders has by far yielded the greatest success. The demonstration in 1911 by Noguchi and Moore of the brain infection by *Treponema pallidum* as the cause of general paresis[3] opened the way in 1917 to its treatment by malaria therapy, and later to its prevention with penicillin; this is now a landmark in the history of medicine. The discovery of the aetiology of pellagra also led to its prevention and control, leading to the prevention of one type of dementia associated with alcoholism and avitaminosis.

These two once very frequent diseases have almost completely disappeared and there are many experienced psychiatrists who never come across a single case of either; with them also disappeared the history of their successful control. Although the same success has not yet been achieved in relation to vascular dementia, the control of hypertension and atherosclerosis (e.g. through the reduction of salt and fat intake) can significantly reduce brain damage and ensuing dementia (vascular or multi-infarct dementia).

In some developing countries, meningitis and malaria (and, to a lesser extent, inadequately treated epilepsy) are important causes of permanent brain damage which can also lead to dementing disorders. The environmental control of malaria and other brain infections, of which bacterial meningitis is the most important, and their early and prompt treatment can reduce the impact of the infection on the brain and prevent these forms of dementia (or intellectual disorder, depending on the age of onset).

More recently, it has been demonstrated that in some people infected with HIV, the initial manifestations of AIDS are accompanied by some forms of mental disorder, such as mood disorders or dementia.[4] The prevention of these forms of mental disorders follow the same measures as for the prevention of AIDS in general. However, it is not yet certain if the newer combined treatments (bi- and tritherapy) can alter the course of AIDS when brain damage due to HIV has been confirmed.

(b) Intellectual disorder (mental retardation)

Up to 15 per cent of cases of intellectual disorder could be prevented by dealing with the causes that lead to it. A recent WHO publication[2] has set detailed guidelines for the prevention of some forms of this condition, namely, Down syndrome, foetal alcohol syndrome, phenylketonuria, and iodine deficiency syndrome.

These preventive actions are both efficient and affordable even in very poor regions of the world.

- *Down syndrome*—The primary prevention of Down syndrome can be successfully achieved through the control of the age at which women become pregnant: ideally, the age range during which the risk is minimal is between 16 and 35 years, after which the risk increases almost exponentially, as shown in Fig. 7.4.1. Amniocentesis is a procedure that can be very useful for the *in utero* diagnosis of Down syndrome (as well as of other problems and malformations). Where it is culturally and morally acceptable, and legally permitted, a therapeutic abortion is viewed by some as another primary prevention measure.
- *Iodine deficiency*—The world population at risk of intellectual disability due to iodine deficiency is approximately 1 billion and it still occurs in large numbers in some regions of the globe.[6] However, it can be very efficiently and cheaply prevented through the addition of iodine to salt, milk, flour, or water, or, in special situations, through injections of an oily solution containing iodine.[7]
- *Phenylketonuria*—Intellectual disorder due to phenylketonuria can also be successfully prevented through the early identification of children at risk who then receive a phenylalanine-free diet throughout their lives.[8]
- *Foetal alcohol syndrome* (FAS)—Intellectual disorder and malformations seen in FAS syndrome can be prevented if women stay away from alcohol during pregnancy, more particularly during the first trimester, or at least keep their alcohol intake below the dangerous limit of 15 g of ethanol per day.[9]

Table 7.4.1 summarizes actions which can effectively prevent some forms of intellectual disorder. Prevention of intellectual disorder is discussed further in Chapter 10.3.

Table 7.4.1 Action to prevent mental retardation

Condition	Preventive action
Iodine deficiency disorders	Iodize salt or water supply Treat individuals at risk with iodized oil or Lugol's solution
Down syndrome	Discourage pregnancies in women over the age of 35 years If appropriate, provide amniocentesis to women over the age of 35 years
Foetal alcohol syndrome	Use simple screening test to identify women at risk Discourage women from drinking alcohol during pregnancy Alert women that drinking around the time of conception increases the risk to the child
Phenylketonuria	Screen all newborn babies for phenylketonuria Treat with a special low-phenylalanine diet Discourage pregnancies in women with phenylketonuria

Mental disorders without a well-established aetiology but with a relatively predictable course

In this group of disorders, the target for prevention is not the disorder itself, whose aetiology is not clearly established, but the occurrence of further episodes of the disorder, since these are chronic recurrent disorders. Three good examples of this are mood disorders, schizophrenia, and alcohol dependence syndrome.

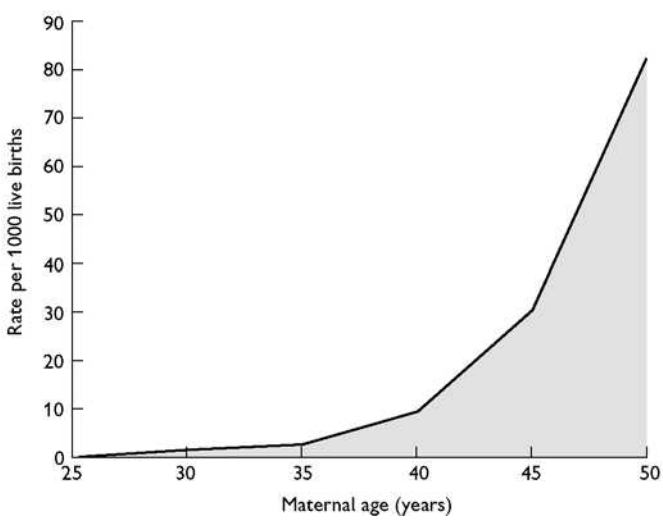


Fig. 7.4.1 Estimated risk of Down syndrome by related age. (Data from Gottesman.[5] Taken from the contribution of genetic factors to the common psychopathologies © World Health Organization, www.who.int)

(a) Mood disorders

In addition to the use of antidepressants and neuroleptics to treat episodes of depression and mania, respectively, it has been demonstrated that the appropriate use of lithium salts can prevent the reappearance of new episodes of disease, or at least increase the disease-free periods and reduce both their duration and severity. The use of lithium salts, also called psychoprophylactics, is now a standard procedure for the treatment and management of mood disorders and this use can be considered as a form of prevention of mood disorders (see Chapter 4.5.8).

(b) Schizophrenia

Although no evidence of a definite cause of schizophrenia is yet available, it can be reasonably controlled through the use of psychopharmacotherapy and psychosocial interventions. Relapse rates (measured by number of hospital readmissions or days in hospital) decrease significantly when persons with schizophrenia adhere to some specific pharmacological regimens and are exposed, together with their relatives, to psychoeducational programmes.[10] These usually include some social skills training for the patient and information about the disease and the management of expressed emotions for the relatives. Unfortunately, schizophrenia is a very long-lasting condition, even lifelong, and most studies on the combination of those two approaches have not yet gone beyond a follow-up of 48 months, thus limiting the full appreciation of its long-term efficiency.

(c) Alcoholism

This term refers to both the alcohol dependence syndrome and other forms of problematic use of alcohol. For people in either category, whenever total abstinence is an unachievable or undesirable goal, the intake reduction, particularly when in connection with risk or dangerous situations, may become the target for prevention. There are several therapeutic techniques usually revolving around brief interventions that are useful to help people significantly to reduce their alcohol and their exposure to problems

associated with it (harm reduction). One should neither forget nor minimize the positive impact of self-help groups (such as Alcoholics Anonymous) in helping some people to achieve sobriety.

Psychosocial problems strongly associated with mental disorders

(a) Violent behaviour

In this category, the immediate target condition is not strictly a mental disorder. However, its close connections with some forms of mental disorder or symptoms make their prevention of immediate concern in the mental health field. Domestic violence (spouse and child beating) is strongly associated with substance abuse;[11] some types of mental disorders are potent risk factors for suicide (see below), and stress and anxiety are at the roots of the burnout syndrome seen in health workers (see below).

Extremely violent behaviour (and crime) is not associated with severe mental disorders (such as schizophrenia) as usually portrayed by the media.[12,13] On the contrary, it is more frequently associated with some types of personality disorders (e.g. antisocial, poor impulse control, aggressive) and with substance use, of which the most prevalent is alcohol. Hence, the control of substance use disorders (in itself a case of secondary prevention) can be seen as an example of primary prevention of domestic violence. The focus of prevention should be broadened from the individual perspective (the substance user) to the social group (the family) which is not exposed to the direct organic effects of alcohol, but nonetheless suffers from its indirect effects.

(b) Suicide

Despite the long-standing sociological tradition of considering suicide as a phenomenon associated with the social condition known as anomia,[14] the medical profession tends to see it primarily as a medical problem. Probably both are right. Among the demonstrated risk factors for suicide there are both social factors (anomy, old age, masculine gender, social isolation) and medical problems (chronic, painful, and incurable diseases and, most of all, psychiatric disorders and psychological problems). Several European studies have indicated that approximately 80 per cent of all cases of suicide are associated with alcohol use and depression combined.[2] This indicates the appropriateness of targeting the treatment of these two conditions (another secondary prevention intervention) regarding the prevention of suicide. However, given the broad social implications of suicide, treatment of mental disorders alone has not yet produced a significant reduction of suicide rates.

Hence the importance of a paradigmatic change which views suicide on a ecological perspective that considers the suicidal act as the immediate target for prevention and not just suicidal intention or ideation as subsumed by the traditional medical (and psychiatric) model.[2] It is now firmly established that the control of the availability of means to commit suicide can greatly contribute to the reduction of mortality rates from suicide. According to WHO,[2] steps to prevent suicide can be taken in the following areas:

- psychiatric treatment
- gun control
- gas detoxification
- control of toxic substances
- responsible media reporting.

Corresponding actions to prevent suicide include:

- identification and treatment of people suffering from depression
- restriction of access to guns
- detoxification of domestic gas and car emissions
- controlling the availability of toxic substances and medicines
- downtoning reports of suicide in the media
- erection of physical barriers to deter jumping from high places.

A more detailed discussion about the prevention of suicide can be found in Chapter 4.15.4.

(c) Staff burnout

Freudenberger[15] first used the term 'staff burnout' to describe a 'syndrome of exhaustion, disillusionment and withdrawal in voluntary mental health workers'. The concept has aroused considerable interest in the caring professions, and the publication of a large number of articles and books on the subject suggests that burnout is a major problem in health services.[2]

Indeed, with the general trend towards community-based care in many parts of the world, burnout is now a problem faced by all caregivers, including the relatives of people suffering from chronic disorders.

There is no single accepted definition of burnout; however, there is general agreement that the syndrome has three major characteristics, which are observed in various caregivers, particularly health workers and family members:

- emotional exhaustion
- depersonalization
- a reduced feeling of personal accomplishment.

The two approaches most frequently employed to prevent staff burnout are:

- Stress management (at the individual level);[16] and
- Supervisor training (at the organizational level).[17]

Table 7.4.2 summarizes actions which contribute to the prevention of burnout among health care staff, at different levels of intervention.

Who is responsible for primary prevention?

As indicated in the introduction of this section, one of the obstacles to a greater involvement of mental health professionals in preventive action stems from the fact that the effective actions for the

Table 7.4.2 Action to prevent staff burnout

Avoid making unrealistically high demands of caregivers
Ensure that all workers have some rewarding tasks
Train caregivers in time-management and relaxation techniques
Modify jobs that are proving too stressful
Encourage the formation of support groups
Consider the possibility of part-time employment
Encourage workers to participate in decisions which affect them

prevention of several mental disorders often falls outside their usual remit and in many cases falls outside the health sector altogether, as indicated in Table 7.4.3. Although this is true, it does not mean that mental health professionals do not have a major role to play in three areas: advocacy, information generation, and supervision.

Perhaps mental health professionals need to reconsider their potential role in primary prevention; for instance, they could develop their potential to act as advocates and advisers to professionals in other sectors. As Eisenberg[18] has argued:

> what matters is not the mode of action of the agent, the venue in which it is applied, or the academic discipline of the practitioner, but the effectiveness of the measure in preventing diseases manifested by disturbances in mental function.

At any rate, prevention is unquestionably a public health priority. Accordingly, WHO has recently published a book on this aspect of prevention, whose reading is strongly recommended.

Table 7.4.3 Agents for prevention

Condition	Effective agents for prevention
Mental retardation	
Down syndrome	Family planning professionals Obstetricians/midwives Women's associations
Foetal alcohol syndrome	General health-care personnel Substance abuse professionals Family planning professionals Obstetrictans/midwives . Educators Women's associations
Phenylketonuria	Obstetricians/midwives Paediatricians Nutritionists
Iodine deficiency	General health-care personnel Nutritionists Groups involved in salt production/trade Water supply personnel Educators
Depression	Mental health workers Rehabilitation officers
Schizophrenia	Mental health workers Rehabilitation officers
Alcohol dependence syndrome	General health workers Mental health workers Self-help groups
Violence	General health workers Mental health workers Paediatricians Justice officers Police officers
Suicide	Mental health workers General health workers Agricultural and environment authorities Journalists Pharmaceutical industry Car industry Traffic authorities Authorities in charge of provision of domestic gas Gun control authorities (including legislators)
Burnout	Occupational health workers General health workers Staff counsellors Trade unions Staff associations Personnel officers Job supervisors Self-help groups

Further information

Bertolote, J.M. (ed.) (1965). *Primary prevention of mental, neurological and psychosocial disorders*. World Health Organization, Geneva.

Leavell, H.R. and Clark, E.G. (1965). *Preventive medicine for the doctor in his community: an epidemiological approach* (3rd edn). McGraw-Hill, New York.

World Health Organization. (2004). *Prevention of mental disorders: effective interventions and policy options*. World Health Organization, Geneva.

References

1. Leavell, H.R. and Clark, E.G. (1965). *Preventive medicine for the doctor in his community: an epidemiological approach* (3rd edn). McGraw Hill, New York.
2. Bertolote, J.M. (ed.) (1998). *Primary prevention of mental, neurological and psychosocial disorders*. World Health Organization, Geneva.
3. Schneck, J.M. (1960). *A history of psychiatry*. C.C. Thomas, Springfield, IL.
4. Maj, M. (1990). Psychiatric aspects of HIV-1 infection and AIDS. *Psychological Medicine*, **20**, 547–63.
5. Gottesman, I. (1982). *The contribution of genetic factors to the common psychopathologies*. World Health Organization, Geneva.
6. Hetzel, B.S. (1986). Mental defect due to iodine deficiency: a major international public health problem that can be eradicated. In *Science and service in mental retardation* (ed. J.M. Berg), pp. 297–306. Methuen, London.
7. Dunn, J.T. and van der Haar, F. (1990). *A practical guide to the correction of iodine deficiency*. International Council for Control of Iodine Deficiency Disorders, Adelaide.
8. World Health Organization. (1985). *Community approaches to the control of hereditary diseases: report of a WHO advisory group*. World Health Organization, Geneva.
9. Smith, I.E., Lancaster, J.S., Moss-Wells, S., *et al.* (1987). Identifying high-risk pregnant drinkers: biological and behavioural correlates of continuous heavy drinking during pregnancy. *Journal of Studies on Alcohol*, **48**, 304–9.
10. Leff, J. and Waughn, R. (1981). The role of maintenance therapy and relative expressed emotion in relapse of schizophrenia: a 2-year follow up. *The British Journal of Psychiatry*, **139**, 102–4.
11. Regier, D.A., Farmer, M.E., Rae, D.S., *et al.* (1990). Comorbidity of mental disorders with alcohol and other drug abuse: results from the Epidemiological Catchment Area (ECA) study. *The Journal of the American Medical Association*, **264**, 2511–18.
12. Barbato, A. (1998). Psychiatry in transition: outcomes of mental health policy shift in Italy. *The Australian and New Zealand Journal of Psychiatry*, **32**, 673–9.

13. Aldige, H.V. (1992). Civil commitment and arrests: an investigation of the criminalization thesis. *The Journal of Nervous and Mental Disease*, **180**, 184–91.
14. Durkheim, E. (1990). *Le suicide*. Presse Universitaire de France, Paris.
15. Freudenberger, H.J. (1974). Staff burn out. *Journal of Social Sciences*, **30**, 159–65.
16. Meichenbaum, D. and Jaremko, M.E. (1983). *Stress reduction and prevention*. Plenum Press, New York.
17. Kilburg, R.C., Nathan, P.E., and Thoreson, R.W. (eds.) (1986). *Professionals in distress: issues, syndrome and solutions in psychology*. American Psychological Association, Washington, DC.

7.5

Planning and providing mental health services for a community

Tom Burns

Introduction

The aim of this chapter is to assist clinicians and managers review and plan services effectively for their local population. Severe psychiatric disorders manifest themselves in social relations and often disrupt social structures; they have wide-ranging consequences and services need to be comprehensive. Health and social care have been intertwined in psychiatry from its origins—it is neither feasible nor sensible to ignore the wider context of their management.

Mental health services research

The last 30 years have seen an explosion of Mental Health Services Research alongside the shrinking and closure of mental hospitals (see Chapter 7.6). Policy considerations, particularly cost containment and public safety, have influenced the research agenda which is disproportionately Anglophone (from the United States, United Kingdom, and Australasia) and focused on new services developed as alternatives to institutional care with staffing and motivation that are not easily generalizable. More routine practices, crucial for safe and effective care, have been relatively neglected by researchers.

Scope of chapter

This chapter is mainly devoted to describing the essential components of a mental health service—its 'building blocks'. It will then consider how they relate to one another, how they can be prioritized, and how integrated into an effective local service linking into other essential services. Lastly it will stress how their inevitable evolution should be monitored.

Services for adults (increasingly referred to as 'adults of working age' indicating 18–65 years) will be used as the template. In many settings these may be the only services, stretching to accommodate all comers. In better resourced health care systems a range of specialized services have evolved from this basic model and are described elsewhere in this section (refugees 7.10.1, homeless 7.10.2, and ethnic minorities 7.10.3).

Building blocks of mental health services: care and treatment

Most mental health treatments (whether psychological, pharmacological, or social) are based on face-to-face interviews and do not require sophisticated equipment or buildings. Institutions (the asylums) evolved for social care of disabled individuals, to protect them while they recovered and, sometimes, to protect society from them. Patients needing long-term institutional care are now relatively few but psychiatry is judged on how they are managed and service planners must pay them due attention.

Inpatient beds

No comprehensive service can survive without access to 24 h nursing supervision for acute episodes of severe illness. These include patients at risk from neglect or suicide or those lacking insight. Wards usually accommodate 10–20 patients. It is rarely possible to effectively staff and run stand-alone units of less than 3–4 such wards (30–60 beds). Ward size is a trade-off between privacy and domesticity against effective supervision. Single rooms are preferable, affording maximal privacy and, while initially expensive, improve flexibility and reduce conflict.

Smaller, more flexible, units such as 'crisis houses' offering 24 h care are a useful complement to inpatient wards, but not a replacement. Ward design and management are increasingly crucial as improved community care concentrates involuntary and disturbed inpatients in them.

How many acute beds?

'How many beds do we need for our local population?' is often the first question asked by planners or managers. Unfortunately there is no reliable or precise answer to this. We know that supply will drive use (perceived as need)—beds are rarely left unfilled despite enormous variation in their availability. It is also surprisingly difficult to collect useable figures on bed usage nationally or internationally because of differences in methods of reporting and also the profusion of overlapping and rarely defined local terms

(e.g. night hospitals, crisis homes, step-down wards). The levels of external accommodation provision (e.g. hostels, day care) clearly also impact the need for acute beds. Similarly need for beds will reduce as community services become more comprehensive and robust.

European provision of general acute beds in 2000 in the public sector ranged from 128 per 100 000 in the Netherlands to 6 per 100 000 in Northern Italy. However, unless we know the pattern of care (in particular the level of private and social services care) these figures tell us relatively little. The United Kingdom has little parallel private care and here acute beds needed for a population of 250 000 have been estimated to range from 50 to 150 plus 5 to 20 secure or intensive care beds[1] dependent on morbidity (generally much higher in large urban settings). London figures for the mid-1990s were very close to this range, averaging 73 for outer and 110 for inner London, but with increased secure provision, particularly in the deprived inner city. The authors predicted a similar range of 24 h supervised hostel need (40–150 per 250 000 population) and London use was somewhat higher (99 and 162 per 250 000, respectively) but with a markedly wider range.

Current bed usage in the United Kingdom is closer to the 50 per 250 000 and well below this in stable communities. This reflects both the establishment of specialized home-treatment and assertive outreach teams and the expansion of forensic care but also a shift in expectations and practice. The average duration for admissions has been steadily reducing over the last three decades. Figures can be misleading as they are heavily skewed by short (1–2 day) admissions but the current admission for an uncomplicated psychotic relapse is likely to be between 3 and 6 weeks.

Longer inpatient care

Acute inpatient wards admit patients for weeks or a couple of months. Rapid discharge is anticipated and regimes emphasize openness and independence. Even within a local service some patients will require longer or more secure care because of illness severity or for legal reasons. Modern rehabilitation practice restricts long-stay wards to patients whose behaviour is persistently unacceptable to local communities. Forensic and secure services are usually a regional or national rather than local responsibility.

Diagnosis-specific wards

Alcohol and substance abuse wards have been long established (especially in Scandinavia and Central Europe) and diagnosis- or disorder-specific wards are increasingly common. Wards for specialized patient groups such as anorexia nervosa or resistant schizophrenia provide highly specific regimes. These are generally an adjunct to acute admission wards rather than an alternative. Some services are organized in disorder-specific wards (e.g. a psychosis unit, a psychosomatic ward) *instead* of general wards. Such specialization is not possible in comprehensive services for populations of less than about 1 000 000. For smaller populations this increased specialization must be balanced against reduced flexibility and energy wasted in 'boundary disputes'.

Day care

Day care is provided either in day-hospitals or day-centres, with little consistency in the terms or practices. Patients attend usually from 1 to 5 days a week for a half or whole day before returning to their homes in the evening. It is particularly valuable when families are out at work but can offer support at evenings and weekends or for very isolated patients.

Generally day-hospitals are provided by health services, include medical and nursing staff and can offer treatments (e.g. the prescription and monitoring of medication, psychotherapies). Day-hospitals were a significant feature in the move of mental health services from mental hospitals to District General Hospital sites. However their role has been more uncertain since community teams have expanded and taken on much of their therapeutic role. Many services have scaled down or even closed their day-hospitals relying more on social services for day care. Day-hospitals have had a problem of isolating themselves from service needs, locked in time with a static patient group. Comprehensive services can, undoubtedly, survive well without them, so if they are to be established it is essential that there are strong links into local teams who can exercise some control over their clientele and their activity.

Day-centres, provided by social care organizations, can rarely provide treatments or employ clinical staff. However overlap is wide with services highly specific to local context (e.g. a drop-in day-centre may be the main provider of psychiatric assessment and treatment in areas of high social mobility and homelessness). Generally day-centres provide long-term social support and day-hospitals focused interventions and treatments.[2] The 'Club House' is a specialized rehabilitation day centre, popularized in the United States, which emphasizes useful normal work and where members take responsibility for running the centre with minimal supervision. Many day units now function in the evening and at weekends.

Acute day-hospitals in Europe and partial hospitalization in the United States have been energetically proposed as alternatives to inpatient care[3] but have had little impact. While day-hospitals never achieved their anticipated prominence they serve specific groups well (e.g. mothers with small children or protracted treatment of eating disorders or personality problems). Day care is problematic in rural settings but adaptations such as travelling day-centres (i.e. a team that moves from setting to setting on specific days) or a weekly open day run by the community team are worth considering.

Supported accommodation and residential care

Many patients remain well outside hospital only with adequate support. At its most basic this implies stable, affordable accommodation. For many, however, supervision is needed to ensure self care, continued medication, and to anticipate and defuse crises. This can be provided by voluntary agencies, social services, or health services. Voluntary agencies tend to be more efficient at providing long-term residential care[4] but they may be reluctant to accept risky patients (e.g. with a history of violence or substance abuse). A mixed economy works best and the need for health services supported accommodation depends on the vigour of local voluntary and social services. While some purpose built units exist, the accommodation is usually shared adapted houses to promote integration and reduce stigma.

Supported or sheltered accommodation is subject to a bewildering terminology but can be considered at four basic levels of increasing need:

1 ***Group homes***. These have no regular staff and are reserved for relatively independent patients visited by staff from their own community teams.

2 ***Day-staffed hostels***. One or two staff are present each day to support and monitor patients (encouraging cooking and cleaning, etc). They would usually not provide specific treatment but liaise with the community team about it.

3 ***Night-staffed hostels***. Non-clinical staff sleep over in the hostel to provide greater safety and availability.

4 ***24 h staffed/nursed hostels***. On-site clinical staff are available overnight either sleeping in or, sometimes, awake. These are expensive hostels and generally restricted to patients with long-term severe illnesses (including sometimes those compulsorily detained). Night-staffed hostels tend to be larger usually with 10–20 residents as opposed to 4–8 in day staffed ones.

Most comprehensive local services provide levels 1 and 2 and most social services undertake to provide level 3. Level 4 is relatively rare and would usually serve a population of 500 000–1 000 000.

Office-based care and outpatient clinics

In insurance-based systems many psychiatrists run individual office practices and manage patients on their own. In state-funded systems this is rare; most work in outpatient clinics or mental health teams. Both approaches should be considered when planning and providing public mental health services, paying particular attention to financial regulations that can inhibit integration and development (comprehensive planning may pose a significant threat to their livelihood and be resisted). Office-based practice remains widespread but neglected in academic and policy publications. It tends to be narrow in remit (usually either psychotherapy or pharmacotherapy) and is poorly equipped for managing severe disorders.

Outpatient clinics ('polyclinics' or 'dispensaries') are an essential part of modern services increasingly replacing office practice. Psychiatrists and psychologists may still operate independently within them but with access to enhanced resources and second opinions. In the public sector outpatient clinics may operate either alongside community mental health teams (CMHTs) or as part of them (which works better for severe illness).[5] They provide an efficient, predicable format for assessments, treatment, and monitoring.

Community mental health centres (CMHCs)

Mental hospitals, for all their faults, had no problems coordinating care; what little was available was all in the same place. Outpatient clinics expanded to Community Mental Health Centres (CMHCs) providing a wide range of services located in shared buildings (e.g. depot clinics, a day-hospital, psychotherapy services). The failure of the early US CMHCs demonstrated that relying entirely on patients to attend fails to engage the more ill and also that down-playing the 'medical model' made it impossible to recruit psychiatrists, further distancing practice from the severely ill.

Most CMHTs are based in CMHCs sharing accommodation with other CMHTs and services (e.g. day care). They provide an important safeguard in sustaining clinical standards and reducing the professional isolation in dispersed community services. This is a particularly important safeguard for community teams which can otherwise easily become idiosyncratic and rigid in their practice if not forced into regular contact with others.

Multidisciplinary Community mental health teams (CMHTs)

Most community mental health services consist of varied forms of multidisciplinary CMHT consisting of psychiatrists, nurses, social workers and often psychologists, and occupational therapists. The staffing of these teams will vary but their strength is that regular meetings to assess and review the management of patients incorporates their varied professional perspectives and allocates tasks based on skills and needs. Developed initially in France and the United Kingdom and championed latterly in Italy they have seen further specialization from North America and Australia.

The generic sector CMHT ('The CMHT')

Who it is for

The CMHT is *the* fundamental building block of modern community mental health services. It originated as mental hospital catchment areas (often covering a whole city or county) were divided into sectors of 50–100 000 inhabitants to permit ongoing care. The aim was that it should be possible for most of the team to have some familiarity with most of its complex and long-term patients and to have some *personal* knowledge of its referrers and community resources. Current sector size in Western Europe ranges from 20–50 000 population, determined both by resources (shrinking as investment increases) and by the local configuration. As more specialized teams are established the CMHTs remit may be reduced and sector size consequently increased keeping its caseload fairly constant. 200–250 is considered the maximum for most teams to exploit multidisciplinary working. The number is less in services for highly complex and difficult patients.

CMHTs offer assessment and care for patients discharged from psychiatric units and those who cannot be adequately treated in primary care or in the private sector. They should prioritize severe mental illnesses (SMI—e.g. psychoses and severe affective disorders). However diagnosis is not all—complications from social adversity, personality difficulties, or substance abuse can make secondary mental health care necessary even for apparently 'minor' disorders. Tools to clarify this threshold[6] have been of limited use and most teams rely on clinical assessments. In countries with limited private care CMHTs also treat mild and transient disorders. CMHTs can be remarkably inefficient if little thought is given to their structure and thresholds. To work well, there needs to be agreement on their purpose, clientele and systems of management and they have often suffered from lack of clarity and leadership.

Staffing and management

CMHT staffing varies enormously and there is no uniform model. Teams of less than 6 can rarely provide comprehensive care or cross-cover while teams of more than about 12–15 start to become

unwieldy, overwhelmed with management and information transfer. CMHTs emphasize skill-sharing and a degree of generic working and have evolved an informal, democratic style[7] which often means confusion over clinical leadership (originally provided informally by senior medical staff). With increased staff numbers and treatment complexity 'team managers' now coordinate workload with a role which varies from the purely administrative to setting clinical priorities and supervising staff. Establishing a clear understanding of clinical leadership in CMHTs (without inhibiting initiative and creativity) is essential for effective functioning. If leadership and management are separated (common with a strong medical presence) the roles need to be well defined and relationships good.

Assessments

The key to good care is accurate assessment (see Chapter 1.8.1). Most commonly psychiatrists conduct initial assessments (usually in an outpatient clinic) and involve the team members in treatment. Increasingly other team members have taken a role in assessments, either individually or jointly with the psychiatrist. Although this issue generates strong feelings there is surprisingly little research into it. With highly developed primary care non-medical assessments may be effective but otherwise medical time should prioritize assessments. With severely ill patients home-based assessments pay considerable dividends.[8,9]

Case management

Most CMHT staff act as clinical case managers[10,11] with responsibility for coordination, delivery, and review of care for their patients. The caseloads of staff members should be explicitly limited (usually 15 to 30) and reviews recorded and systematic. In the United Kingdom this has been formalized as the Care Programme Approach.[12] Fig. 7.5.1 shows a care plan indicating a patient's needs or problems, the interventions proposed to meet them, who is responsible and who is informed, plus an agreed date for review. Such concise structured paperwork (as with the risk assessment and contingency plan (Fig. 7.5.2)) can be adapted to any service, coordinates complex care and serves as a natural focus for clinical reviews. The level of detail needs to be clinically (not managerially) determined.

Team meetings

CMHTs need 1–2 regular meetings (each usually 1.5–2 h) per week for both clinical and administrative business. The degree of structure depends on team style and remit.

(a) Allocation of referrals

Referrals can be allocated by who is first available or by matching the clinical problem against available skill and training. Time discussing allocations before assessment is generally unprofitable and most well-established teams delegate the task to the manager or a senior clinician.

(b) Patient reviews

*R*eviews should be held for (i) *new patients*, (ii) *routine monitoring*, and (iii) *discharge*. Reviews can range from simply reporting the problem and proposed treatment in uncomplicated cases through to detailed, structured, multidisciplinary case-conferences including other services (e.g. GP, housing, child protection). *New patient* reviews are an excellent opportunity for providing a broad, experienced overview, and ensuring rational and fair allocation to caseloads. *Routine monitoring* is often overlooked yet probably the most important for team efficiency. It should be systematic and not only responsive to crises and problems. It shapes and redirects treatment and identifies patients ready for discharge. The burden on individual staff members is regularly monitored. Routine monitoring is a legal requirement of the Care Programme Approach and good practice in all case management. *Discharge reviews* are an excellent opportunity for audit and learning within the team.

(c) Managing waiting lists and caseloads

Effective CMHTs need to guarantee prompt access. *Routine assessments* should be within 2–4 weeks. Sooner is rarely productive and delays above 3 weeks result in a rapidly rising rate of failed appointments.[13] *Urgent assessments* (most psychotic episodes) need to be seen within a week, usually within a couple of days. *Emergency assessments* are for those associated with immediate risk (e.g. hostile behaviour or suicidal intent) and need to be seen the same day.

A practical approach to waiting lists is to count the assessments in the preceding year and allocate routine appointments for 20 per cent more. Thus a team with 400 assessments the preceding year allocating nine slots a week will have one available weekly for emergencies. Rapid routine assessment reduces pressure for urgent and emergency referrals more efficiently than emergency rotas.

Communication and liaison

Team meetings ensure internal communication but CMHTs need good links with a wide network of professional colleagues. Structured liaison is advisable with primary care and general hospitals in addition to routine letters. Hospital links may be between specific CMHTs and wards or CMHTs may provide input to patients from their sectors in the absence of dedicated liaison psychiatry services.

(a) General practice liaison

Much of mental health care is delivered in primary care (see Chapter 7.8) and effective coordination is essential. GP liaison systems range from informal contact through to shared care and co-location of CMHTs in GP Health Centres.[14] An effective system comprises regular, timetabled meetings between the two teams or a 'link' CMHT member attending the GP health centre. Monthly meetings where shared and complex patients are discussed are highly time-efficient because of prompt problem solving and crisis anticipation. However it is important to be clear about responsibilities, fudging boundaries is risky.

(b) Liaison with other agencies

The same principles apply to liaison with other agencies (social services, housing, charitable, and voluntary sector providers). Whether regular meetings are cost-effective will depend on the volume of shared work but showing up and meeting people (even just once) pays enormous dividends in improved relationships and understanding. Professional confidentiality and information sharing is more sensitive.

Assertive Outreach (AO) Teams

The most replicated and researched specialist CMHT is the AO Team. The original US model[9] improved clinical and social outcomes with substantially reduced hospitalization at slightly lower cost.

CPA REVIEW

Patient's name: Jenny T	**CMHT: West Central**
Address: 56 Acacia Avenue	**Phone:**
Phone:	**New patient: ~~YES~~/NO**
Date of birth: 09.06.61	**If NO, date of review: 20.10.07**
GP: Dr Findlay	**Diagnosis:**
Phone:	**1…Major depressive disorder..... F 32 .0**
	2…… F __ __.__

You must consider the following: 1) Mental health, including indicators of relapse; 2) Physical health; 3) Medication; 4) Daytime activity; 5) Personal care / living skills; 6) Carers, family, children and social network; 7) Forensic history; 8) Alcohol or substance misuse 9) Cultural factors; 10) Housing/finances/legal issues.

Complete a **risk assessment** and include: **i) a crisis plan; ii) a contingency plan**

Assessed needs or problem	Intervention	Resp.of
1. Depressed mood, apathetic and self critical	• Regular home visits, assess mental state • Encourage compliance with antidepressants • Encourage activity – take to shops etc	BJ
2. Suicidal thoughts	• Explore severity (+/- plans) at each visit • Support mother and husband who are scared of suicidal thoughts	BJ/ Cons
3. Daughter's school problems	• Maintain links with class teacher • Keep family informed of her progress	BJ
4. Plan for recovery	• Link with support group when mood lightens • Help reapply for part-time cleaning job	BJ

Professionals involved in care:	✓ Dr	Psychologist	✓ CPN	OT	✓ SW	Ward Nurse	ACT	Other
Present at planning meeting:	✓ Dr	✓ Psychologist	✓ CPN	OT	✓ SW	Ward Nurse	ACT	Other

Copy given to patient? **YES/~~NO~~** Copy sent to GP? **YES/~~NO~~**

Care co-ordinator(print): **Billie Jarvis (BJ)** **Phone**

Care co-ordinator (signature): ..... **Date of next review: 20.04.08.**

Job title: **CPN** **Patient's signature:……**

On Supervision Register? **~~YES~~/NO** Care management? **YES/~~NO~~** Risk history completed? **YES/~~NO~~**

On Supervised Discharge? **~~YES~~/NO** Relapse + risk plan required? **YES/~~NO~~**

Fig. 7.5.1 Care programme review document.

CONFIDENTIAL: RELAPSE AND RISK MANAGEMENT PLAN			
Name: Alastair W			
Categories of Risk Identified:			
Aggression and violence	~~YES~~/NO	Severe self-neglect	YES/~~NO~~
Exploitation (self or others)	YES/~~NO~~	Risk to children & young adults	YES/~~NO~~
Suicide and self-harm	YES/~~NO~~		
Other (please specify} ……			
Current factors which suggest there is significant apparent risk: **(For example: alcohol or substance misuse; specific threats; suicidal ideation; violent fantasies; anger; suspiciousness; persecutory beliefs; paranoid feelings or ideas about particular people)** Continued excessive drinking—especially when depressed. Makes him more suspicious and hostile.			
Clear statement of anticipated risk(s): **(Who is at risk; how immediate is that risk; how severe; how ongoing)** Clear risk to strangers (not family or staff), usually in bars. Often when poor medication compliance.			
Action Plan: **(Including names of people responsible for each action and steps to be taken if plan breaks down)** Relapse plan discussed and agreed—to increase antipsychotics and contact when concerned with people plotting ('to help you cope with them'). If he feels seriously threatened to seek admission through the emergency room			
Date Completed:	**xx/xx/xx**	**Review date:**	**xx/xx/xx**

Fig. 7.5.2 Risk assessment and contingency plan.

AO teams (Box 7.5.1) are costly and consequently reserved for the most difficult ('hard to engage' or 'revolving-door') psychotic patients with frequent, often dangerous, relapses and poor medication compliance plus alcohol or drug abuse, significant personality difficulties, and offending behaviour.

AO emphasizes proactive outreach—visiting patients at home even when they are reluctant. It exploits enhanced team working with daily meetings and several members actively involved with most patients both for safety considerations and also reflecting patients' extensive needs. The culture is of very practical working (taking patients shopping, sorting out accommodation, delivering medicines daily if need be) well beyond traditional professional boundaries.

Despite strongly expressed convictions there is little evidence that AO teams need to slavishly follow the original model[15,16] and local clinical adjustments are both sensible and justified. If embedded in a comprehensive system there is little need for a 24 h service, most staff establish strong individual relationships with patients and caseloads are usually more than the recommended 1:10. Where CMHTs function well AO teams take only patients who cannot be stabilized despite their support.

If CMHTs provide outreach and a comprehensive treatment then the extra AO resources may add little. Improved outcomes follow only from additional effective treatments (e.g. daily Clozapine visits in resistant schizophrenia); it is not outreach itself that is therapeutic. Whether AO will improve care (and, if so, how many teams are needed for how many patients) will depend on current services.

Box 7.5.1 ACT core components

- Assertive follow-up.
- Small caseloads (1:10–1:15).
- Regular (daily) team meetings.
- Frequent contact (weekly to daily).
- *In vivo practice* (treatment in home and neighbourhood).
- Emphasis on engagement and medication.
- Support for family and carers.
- Provision of services using all team members.
- Crisis stabilization 24 h a day, 7 days a week.

Ethics in community mental health care

Balancing patients' welfare with their autonomy and their rights with those of their families and the wider community are sharply revealed in AO teams. These teams regularly visit patients who vigorously and clearly reject them. When does intensive support become intrusion? When does professional persistence tip over into coercion or disrespect?

Compulsion was traditionally identified with the buildings of the old asylums or left to the family (as it still is in many parts of the world). With expanded community care, compulsion, and coercion (either explicitly in the form of legal requirements or informally through professional or social pressure[17]) are now a pervasive feature of practice. Improved legal and professional scrutiny makes compulsory treatment possible in the community. Most developed countries have enacted forms of community treatment order ('mandated community treatment', 'outpatient committal') mainly for the care of young psychotic individuals

without insight into their need for ongoing treatment. The introduction of these provisions has generally been controversial but their operation not so.

Community treatment orders have the advantage of legal scrutiny unlike most of the ethical dilemmas facing CMHT and AO staff in their day-to-day work. These require discussion case-by-case. How proper is it to inform neighbours if a patient may pose a risk to them but will not give consent? Is it right for a patient, heavily dependent on his parents, to deny them information on his treatment? What does a case manager do when they know their client is doing something illegal? Most professionals share the goal of maximizing their patient's autonomy while minimizing significant risks. Guidelines exist only for extreme circumstances. Teams should be encouraged in regular discussion of these issues.

Crisis teams

Crisis teams play a crucial role where local services are poorly developed (they may be the *only* community services) or in city centres with many transient and homeless patients. They must prioritize rapid response and accessibility. Most teams will see patients immediately, certainly the same day. Their clinical aims and staffing are essentially similar to the acute functioning in CMHTs. They are best located alongside CMHTs or in the emergency rooms of general hospitals with 24 h availability. Liaison services (see Chapter 5.7) can often incorporate and manage hospital-based crisis teams.

Crisis Resolution/Home Treatment (CR/HT) Teams

The CR/HT team model, developed in Australia[18] and currently implemented in the United Kingdom and Europe, reflects increased consumer demand for access in crises and a desire to reduce inpatient care costs. It draws heavily on AO practice with limited, shared caseloads, flexible working, extended access, and an emphasis on outreach. Reduction in hospitalization offsets much of their cost[19] but this needs to be considerable as with two daily shifts and on call overnight needs a staff of about 15 for a caseload of 30 proposed for a population of 150 000.[20] They target patients who would otherwise be in hospital and focus on the severely mentally ill with intensive visiting (usually daily for a limited period) and considerable practical support and work with patients' social networks; most aim for a maximum of 6 weeks involvement. Such intensive team working requires highly effective communication and the teams meet daily (often twice at shift handovers). Information transfer is burdensome and liaison with CMHTs complex requiring absolute clarity on local arrangements for clinical responsibility.

Variations in practice and sustainability

The CR/HT teams may reduce the need for hospital care[18,21] but how much will vary. The UK model is precisely specified (including who it should and should not care for, Box 7.5.2) but practice varies considerably. A full 24 h service is rarely needed, an on-call facility to the emergency room and police station at night usually suffices. Contact frequencies are generally lower, patients stay with the service longer than anticipated and they are inevitably referred individuals recurrently in crisis (often with alcohol and relationship problems) who cannot easily be refused care but would not be 'otherwise in hospital'. Good medical staffing is needed and CMHT responsibilities need to be carefully negotiated, mutually agreed and crystal clear if to be avoided. These realities need to be carefully considered before deciding to establish such teams.

Box 7.5.2 Remit of UK Crisis Resolution/Home Treatment Teams[20]

'Commonly adults (16 to 65 years old) with severe mental illness (schizophrenia, manic depressive disorders, severe depressive disorder) with an acute psychiatric crisis of such severity that, without the involvement of the CR/HT team hospitalisation would be necessary'.

'The service is not usually appropriate for individuals with:

- *Mild anxiety disorders.*
- *Primary diagnosis of alcohol or other substance abuse.*
- *Brain damage or other organic disorders including dementia.*
- *Learning disabilities.*
- *Exclusive diagnosis of personality disorder.*
- *Recent history of self harm but not suffering from a psychotic or serious depressive illness.*
- *Crisis related solely to relationship issues'.*

Crisis services may have a relatively limited lifespan[22] but can be a very successful way to improve local access, gain familiarity with at-risk populations, and then consolidate as a more comprehensive service. Sometimes, however, they become overwhelmed with inappropriate referrals or patients who cannot be referred on and close.

Crisis houses and respite care

Crisis houses allow admission with a minimum of formality and often with reduced supervision compared to hospitals. They are usually small (4–8 beds) in a domestic setting and take people for days, occasionally a week or two. They are favoured for vulnerable women and early intervention services. Most have one staff member sleeping in overnight and a couple on during the day with support from patients' case managers. They are very welcome for a minority of patients but do not replace inpatient care and need careful supervision to avoid becoming chaotic or blocked.

Adjunct or replacement for CMHTs?

The three teams outlined above comprise the fundamental building blocks of most community services. AO and CR/HT teams have been proposed as substitutes for CMHTs, particularly when there are problems with local CMHTs. However both experience and research evidence[23] suggests that they are rarely durable without effective CMHTs to relate to. They should be considered to improve the quality of care in otherwise well-functioning services rather than cost-saving shortcuts.

Highly specialized and diagnosis-specific teams

There are various specialized teams, generally organized on a regional level. These are not essential local services but impact on

them—both in terms of removing some of the clinical obligations and the need to ensure clear and negotiable thresholds.

Early Intervention Teams (EIS)

Concern that a long duration of untreated psychosis (DUP) confers poorer prognosis[24] has led to the development of EIS teams which many would now argue should be standard provision. Developed mainly from Australian and UK models[25,26] they vary remarkably even despite a detailed prescription for UK teams. Some down-play diagnosis in favour of easy access, others restrict to schizophrenia, some emphasize a 'youth service' while others take all first episode patients irrespective of age.[7] Even more confusing there are three quite different activities which may, or may not, be part of the service (Box 7.5.3).

The core of EIS is a specialized CMHT which case-manages first episode psychosis patients protecting social networks and functioning (keeping patients at college or work, an emphasis on family interventions, etc.) assuming a return to premorbid functioning. Crisis and respite houses are preferred to hospital. Some EIS teams conduct public awareness campaigns, lecturing in schools and colleges.[27] A minority of research teams attempt to identify and treat 'ultra-high risk' patients to prevent progression to psychosis.[28]

Forensic and rehabilitation teams

Community-focused services face particular difficulties in treatment-resistant patients, particularly those with socially unacceptable or offending behaviour. Such patients fit poorly into open wards and specialized forensic teams provide care where offending behaviour and danger to others predominates. Some provide community services (intensive case management of dangerous patients) with an emphasis on risk assessment and management. Integrating them with general services can be problematical.

Rehabilitation

A significant number of patients remain disabled despite best treatments and require long-term management of disability rather than episode-based care. Rehabilitation teams generally serve patients who cannot survive without supervised accommodation even when at their best. They include the diminishing cohort of old long-stay patients and increasingly a very disturbed 'new long-stay' population with comorbid substance abuse and behavioural disturbances.

Diagnostic-specific teams

Highly specialized teams for individual disorders (e.g. eating disorders, personality disorders, bipolar patients) concentrate specific skills and provide specialized treatments and are usually provided at regional level. They usually have stronger advocates than CMHTs, both from professionals and families of sufferers, and the opportunity costs (see below) of establishing them need careful thought.

Box 7.5.3 Components of early intervention teams

- **Case management**—ongoing care of identified patients
- **Early Identification**—awareness-raising campaigns for psychoses
- **High risk and prodromal** patient identification and treatment

Planning services

Step-wise planning and adaptation

Planning mental health services for a given community rarely starts with a clean sheet. In such circumstances there are excellent texts, both general and specific to mental health. Tansella and Thornicroft's 'matrix' model[29] is particularly thorough and structured (see Chapter 7.2). It covers the process from establishing service principles and needs assessment (at national, regional, and local levels) through to monitoring and reviewing the cycle of planning and provision. They propose a hierarchical approach depending on the level of mental health spend.[30] Case identification and outpatient treatments in primary care are the priority for low income countries and only with increased resources the establishment of a secondary care mental health service (usually a form of generic CMHT). Not until these are well-established are specialist and inpatient services indicated. This process must, however, take account of what is already in place.

Local population needs assessment

Psychiatric morbidity varies considerably with social deprivation and is much higher in cities than in stable rural or suburban settings (see Chapter 2.7). At the regional and national level comparative need can be predicted fairly well from established indexes incorporating levels of migrants, overcrowding, poverty, etc. Catchment areas should broadly reflect these differences. At the more local level these figures are of limited value. How does one factor in travelling time or known differences in the quality of primary care? A process of negotiation is best to agree local allocation following these general guidelines. However a concentration of hostels for the mentally ill or homeless or the presence of a railway station or international airport may swamp these differences. These should be provisionally estimated in planning but then regularly monitored and reviewed.

Opportunity costs and unintended consequences

Planning mental health services has become based on international evidence, often including cost-effectiveness analyses, i.e. is there more overall patient health gain for the same input (see Chapter 7.7)? However these rarely address the opportunity costs across a whole system. For example one form of day hospital may be more cost-effective for its patients than another day hospital, but is diverting nurses from an inpatient unit to staff that day hospital a net gain? Rigorous intervention studies of large systems are formidably difficult to conduct and even more so to interpret.

The impact of enthusiasm and the migration of the best staff to such research services can be especially misleading.[31] Successful new services are always reported but there is less consistency in reporting when a service may have lost its efficacy or was abandoned.[5,22] It is best to visit examples of services that are proposed and not to always assume that what works for them will necessarily work for you.[15]

Manpower is often as significant a resource limitation as funding, hence the need for a system-wide appraisal. Also, though expressed as costs, these decisions include wider judgements of values and expectations rather than simply outcomes.[30] Local and national objectives also matter, not just clinical ones (e.g. public safety now dominates much mental health planning).

Cultures and funding

Health care cultures, their structures and their funding systems vary enormously. Services must be congruent with them otherwise they will not survive. Occasionally service planners can influence the system but more often have to adapt to it. Obtaining relatively small changes in funding arrangements (or external governance) can deliver quite major improvements. However caution is needed as unintended consequences and perverse incentives may arise. Enthusiastic clinicians are often blind to the risks of over-prescription which can lock in outmoded practices (e.g. a highly specific form of day hospital or crisis facility) that after opening is found not to have the level of need predicted but may be so rigidly prescribed that it cannot be adapted.

How 'integrated' mental health care should be is a highly local decision. Well-established systems often strive for integration with general medicine or, increasingly, with social services to reduce both discrimination and administrative barriers to integrated care. For less confident services the value of a distinct, separate, identity can be considerable—not least the ability to protect its resources. The 19th century British mental health reformer Lord Shaftsbury wrestled with the same dilemma.

The task of integration is easily underestimated, particularly the energies required to accommodate contrasting health and social care cultures. Social care, for example, tends to rely on very detailed paperwork and a highly structured system of intensive supervision (reflecting political accountability and previously only minimally trained workforce) as opposed to the high levels of professional autonomy in health care. Agreed compromises need to be reached before integration and the negotiations can be exhausting, but preferable to misunderstandings. The costs and benefits will depend very much on the local situation, relationships and history and should be very carefully weighed up.

Relationships with the voluntary sector and patients movement

Relations with local non-health statutory services (housing, education, police) and with the voluntary sector will determine much of the success of MH services. Voluntary and private sectors may fill specialist niches left by a monolithic public health care system (as with the NHS in the United Kingdom) or conversely the public system may act as a safety net, for charitable and private provision. Service planners need to exploit the strengths of local providers. Non-statutory services are often more efficient but less comprehensive and may also be less reliable over the long-term. Both may be equally reliant on public funding, differing only in contracts and degrees of independence. Patient and carer advocacy and support organizations are now a major force; effective working with them significantly will enhance both the design and delivery of services (see Chapter 7.9).

Monitoring and review

Careful monitoring and review are as important as careful planning for several reasons. The long-term, fluctuating nature of disorders, the subjective nature of diagnoses (increasingly self-ascribed) complicate outcome measures and the targeting of services; services can easily drift to those who demand them from those who most need them; treatments (particularly psychological and psychotherapeutic treatments) may evolve over time losing their effective characteristics; needy patients are rarely demanding or well-informed; engaging and motivating them to collaborate in long-term, treatments is difficult.

A consistent effort to deliver even the most basic, proven interventions will make a substantial difference to patient welfare. The PORT study in the United States demonstrated how schizophrenia care was strikingly inconsistent and fell below accepted essentials.[32] Regular audit ensures that services remain targeted on those for whom they were developed and that their application and quality remains good. Monitoring can vary in sophistication—from a simple head count of who is getting what through to careful evaluation of care pathways. The audit review process feeds back into the development process, adjusting and refining it. Even in the most hard-pressed services audit more than rewards its investment.

Routine outcome measures (ROM)

Audit brings rigour and reflection (often lost in the immediacy of the therapeutic relationship) into the care process. Measurement also serves a training purpose by benchmarking interdisciplinary understandings of symptoms and outcomes. Systematic, periodic recording of patients' clinical or social status is increasingly used in both planning and research. Structured outcomes can be generic (e.g. HoNOS[33]) or for specific disorders (e.g. the Brief Psychiatric Rating Scale[34]) or even locally developed. Their value lies in their consistency of use.

Conclusions

Planning and providing mental health services requires flexibility and compromise. Epidemiological and service statistics translate poorly to local planning; this remains primarily a practical and political, rather than academic, activity. The scientific evidence is dominated by Anglophone alternatives to hospital care studies but local history, culture, mental health law, and political imperatives cannot be ignored. Mediterranean societies, for example, with strong family supports and fewer isolated psychotic individuals have less interest in AO teams. A high-profile patient homicide, or a strong public endorsement of services by politician or celebrity, can derail years of careful planning.

This chapter has attempted to draw out some principles (Box 7.5.4) for the process but these can only be guidelines. Inevitably the decision will be based on what is *possible* locally. Despite this most solutions draw on a limited number of tried and tested structures described here in some detail. Their balance and configuration depend on what is available for them and around them. Above all the mentally ill and their families deserve reliable and predictable services. Not all change is innovation and research findings should be judged in terms of their sustainability and demonstrated translation from research efficacy to clinical effectiveness.

Box 7.5.4 Developing local community mental health services

- Make a careful inventory of what services exist, and any special local needs.
- Consult locally and invest heavily in building coalitions with policy makers, statutory services, and voluntary groups.
- Test research evidence for durability and relevance. If possible visit established services and ask 'around the service'.
- Monitor and review regularly. Improved consistency of current practice often delivers more than introducing new treatments.
- Consider carefully opportunity costs. Include both the impact of a specific improvement across the whole service and the costs of system change itself.
- Avoid excessive reorganization—not all change is innovation.

Further information

Burns, T. (2004). *Community mental health teams*. Oxford University Press, Oxford.

The Mental Health Policy Implementation Guide. (2001). Department of Health, London.

Thornicroft, G. and Tansella, M. (2004). Components of a modern mental health service: a pragmatic balance of community and hospital care: overview of systematic evidence. *The British Journal of Psychiatry*, **185**, 283–90.

Thornicroft, G. and Tansella, M. (1999). *The mental health matrix: a manual to improve services*. Cambridge University Press, Cambridge.

References

1. Strathdee, G. and Thornicroft, G. (1992). Community sectors of need-lead mental health services. In *Measuring mental health needs* (eds. G. Thornicroft, C.R. Brewin, and J. Wing). Gaskell, London.
2. Catty, J., Goddard, K., and Burns, T. (2005). Social services day care and health services day care in mental health: do they differ? *International Journal of Psychoanalysis*, **51**(2), 151–61.
3. Marshall, M. (2003). Acute psychiatric day hospitals. *British Medical Journal*, **327**, 116–7.
4. Knapp, M., Hallam, A., Beecham, J., *et al.* (1999). Private, voluntary or public? Comparative cost-effectiveness in community mental health care. *Policy and Politics*, **27**(1), 25–41.
5. Wright, C., Catty, J., Watt, H., *et al.* (2004). A systematic review of home treatment services. Classification and sustainability. *Social Psychiatry and Psychiatric Epidemiology*, **39**, 789–96.
6. Slade, M., Powell, R., Rosen, A., *et al.* (2000). Threshold assessment grid (TAG): the development of a valid and brief scale to assess the severity of mental illness. *Social Psychiatry and Psychiatric Epidemiology*, **35**(2), 78–85.
7. Burns, T. (2004). *Community mental health teams*. Oxford University Press, Oxford.
8. Burns, T., Beadsmoore, A., Bhat, A.V., *et al.* (1993). A controlled trial of home-based acute psychiatric services. I: clinical and social outcome. *The British Journal of Psychiatry*, **163**, 49–54.
9. Stein, L.I. and Test, M.A. (1980). Alternative to mental hospital treatment. I: conceptual model, treatment program, and clinical evaluation. *Archives of General Psychiatry*, **37**(4), 392–7.
10. Intagliata, J. (1982). Improving the quality of community care for the chronically mentally disabled: the role of case management. *Schizophrenia Bulletin*, **8**(4), 655–74.
11. Holloway, F., Oliver, N., Collins, E., *et al.*(1995). Case management: a critical review of the outcome literature. *European Psychiatry*, **10**, 113–28.
12. Department of Health. (1990). The care programme approach for people with a mental illness referred to the special psychiatric services. Department of Health, London. Report No.: Joint Health/Social Services Circular HC (90) 23/LASS (90) 11.
13. Burns, T., Raftery, J., Beadsmoore, A., *et al.* (1993). A controlled trial of home-based acute psychiatric services. II: treatment patterns and costs. *The British Journal of Psychiatry*, **163**, 55–61.
14. Burns, T. and Bale, R. (1997). Establishing a mental health liaison attachment with primary care. *Advances in Psychiatric Treatment*, **3**, 219–24.
15. Fiander, M., Burns, T., McHugo, G.J., *et al.* (2003). Assertive community treatment across the Atlantic: comparison of model fidelity in the UK and USA. *The British Journal of Psychiatry*, **182**, 248–54.
16. Burns, T., Marshall, M., Catty, J., *et al.* (2005). *Variable outcomes in case management trials—an exploration of current theories using meta-regression and meta-analysis*: Final Report. Department of Health, London.
17. Monahan, J., Redlich, A.D., Swanson, J., *et al.* (2005). Use of leverage to improve adherence to psychiatric treatment in the community. *Psychiatric Services*, **56**(1), 37–44.
18. Hoult, J. (1986). Community care of the acutely mentally ill. *The British Journal of Psychiatry*, **149**, 137–44.
19. Smyth, M.G. and Hoult, J. (2000). The home treatment enigma. *British Medical Journal*, **320**(7230), 305–9.
20. Department of Health. (2001). The mental health policy implementation guide. Department of Health, London.
21. Johnson, S., Nolan, F., Pilling, S., *et al.* (2005). Randomised controlled trial of acute mental health care by a crisis resolution team: the north Islington crisis study. *British Medical Journal*, **331**(7517), 599.
22. Cooper, J.E. (1979). Crisis admission units and emergency psychiatric services. Public Health in Europe, No. 2. Copenhagen: World Health Organisation.
23. Burns, T., Catty, J., Watt, H., *et al.* (2002). International differences in home treatment for mental health problems. Results of a systematic review. *The British Journal of Psychiatry*, **181**, 375–82.
24. Marshall, M., Lewis, S., Lockwood, A., *et al.* (2005). Association between duration of untreated psychosis and outcome in cohorts of first-episode patients: a systematic review. *Archives of General Psychiatry*, **62**(9), 975–83.
25. Edwards, J., McGorry, P.D., and Pennell, K. (2000). Models of early intervention in psychosis: an analysis of service approaches. In *Early intervention in psychosis: a guide to concepts, evidence and interventions* (eds. M. Birchwood, D. Fowler, and C. Jackson). John Wiley & Sons, New York.
26. Birchwood, M., Todd, P., and Jackson, C. (1998). Early intervention in psychosis. The critical period hypothesis. *The British Journal of Psychiatry—Supplement*, **172**(33), 53–9.
27. McGorry, P. and Jackson, H. (1999). Recognition and management of early psychosis. A preventative approach. Cambridge University Press, Cambridge.
28. McGorry, P.D., Yung, A.R., Phillips, L.J., *et al.* (2002). Randomized controlled trial of interventions designed to reduce the risk of progression to first-episode psychosis in a clinical sample with subthreshold symptoms. *Archives of General Psychiatry*, **59**(10), 921–8.
29. Tansella, M. and Thornicroft, G. (1998). A conceptual framework for mental health services: the matrix model. *Psychological Medicine*, **28**(3), 503–8.

30. Thornicroft, G. and Tansella, M. (2004). Components of a modern mental health service: a pragmatic balance of community and hospital care: overview of systematic evidence. *The British Journal of Psychiatry*, **185**, 283–90.
31. Coid, J. (1994). Failure in community care: psychiatry's dilemma. *British Medical Journal*, **308**(6932), 805–6.
32. Lehman, A.F. and Steinwachs, D.M. (1998). Translating research into practice: the Schizophrenia Patient Outcomes Research Team (PORT) treatment recommendations. *Schizophrenia Bulletin*, **24**(1), 1–10.
33. Orrell, M., Yard, P., Handysides, J., *et al.* (1999). Validity and reliability of the health of the nation outcome scales in psychiatric patients in the community. *The British Journal of Psychiatry*, **174**, 409–12.
34. Overall, J.E. and Gorham, D.L. (1962). The brief psychiatric rating scale. *Psychological Reports*, **10**, 799–812.

7.6

Evaluation of mental health services

Michele Tansella and Graham Thornicroft

Introduction

Evaluation is the basis for improving care to people with mental illness. It is vital to know whether interventions are beneficial or harmful, and whether they offer value for money. Mental health interventions need to be understood both in terms of their active ingredients and how they fit within their context.[1] Such combined interventions, often including pharmacological, psychological, and social elements, are the epitome of 'complex interventions'[2] and their evaluation poses considerable challenges. In this chapter we shall discuss *definitions* of evaluation, and go on to discuss *why* evaluate, *what* to evaluate, and *how* to evaluate mental health services. In our conclusion we shall offer an indication of the most important trends in this field in the coming years. The overall approach that we take is centred upon the idea that ongoing evaluative research is of fundamental importance in discovering which interventions are effective, neutral, or harmful, and that such information is essential to deliver better mental health care.

Evaluation: definitions and conceptual framework

The *Concise Oxford English Dictionary*[3] gives the following definitions of 'evaluation':

evaluate (*verb transitive*) **1**. assess, appraise; **2a**. find or state the number or amount of; **2b**. find a numerical expression for.

evaluation (*noun*) **1**. appraisal, valuation, assessment; **2**. estimate, estimation, approximation, rating, opinion, ranking, judgement, reckoning, figuring, calculation, computation, determination.

The etymological root of the word therefore refers directly to 'value', although in common usage 'evaluation' now has a more technical connotation. In our view evaluation necessarily requires both the precise measurement of the effects of treatments or services, alongside a contextual understand of the meaning, and value of such results.

A conceptual model that can be used to clarify key issues related to the evaluation of mental health services is the Matrix Model.[4,5] The two *dimensions* of this model are place and time (see Table 7.6.1). Place refers to three geographical levels: (1) country/regional, (2) local, and (3) individual. Time refers to three phases: (A) inputs, (B) processes, and (C) outcomes. In this framework *inputs* relates to all those resources which are necessary before health care can take place (such as financial and human resources, policies, and treatment guidelines), *processes* refers to all those activities which constitute the delivery of health care (such as outpatient consultations, or hospital admissions), while *outcomes* refers to the consequences of health care (such as changes in symptoms, disability, and quality of life). In relation to the evaluation of mental health services, we shall illustrate in this chapter how inputs and processes need to be measured and understood in their contribution to the outcomes of care.

Historically, the first attempts to evaluate psychiatric practice originated in the mid-nineteenth century as the tabulation of admissions, discharges, and deaths in mental hospitals, simply describing the inputs and processes of care. In recent decades, as more sophisticated research methodologies and more valid and reliable research measures have been developed, so the evaluation of mental health services has increasingly focussed upon the analysis of the outcomes of care. As Sartorius has put it, 'In its most classical form, evaluation denotes a comparison between results and goals of activity',[6] indicating that evaluation has now become a purposeful exercise in which measurements are used as tools to answer specific questions, usually defined a priori at the beginning of a scientific study.

Why evaluate mental health services?

In our view, the main purposes of mental health service evaluation are to assess the effectiveness and cost-effectiveness of care, either at the organizational (local) or at the patient (individual) level. In the long-term such evidence can be used to provide better services for people with mental illness. For example, evaluation can be applied to comparing differing models of care, such as studies in England showing that home-treatment teams can provide a realistic alternative to emergency hospital admission.[7–9] Evaluation therefore measures the impact of care (outcomes) and also aims to increase understanding of the active ingredients (inputs and processes) which contribute to better outcomes.[1] In fact, a wider range of purposes can be served by the evaluation of mental health services, as shown in Table 7.6.2.

What to evaluate in mental health services?

In our view the most important focus of evaluation is upon the *outcomes* of care.[10,11] The outcome chosen for any particular

Table 7.6.1 Overview of the Matrix Model, with examples of inputs, processes, and outcomes

Place dimension	Time dimension		
	(A) Input phase	(B) Process phase	(C) Outcome phase
(1) Country/regional level	1A Mental health budget allocation Mental health laws Government directives and policies Training plans for mental health staff Treatment protocols and guidelines	1B Performance/activity indicators (e.g. admission rates, compulsory treatment rates)	1C Overall suicide rates Homelessness rates Imprisonment rates Years lived with disability
(2) Local level	2A Local service budgets and balance for hospital and community services Local population needs assessment Staff numbers and mix Clinical and non-clinical services Working relationships between teams	2B Service contacts and patterns of service use Pathways to care and continuity Targeting of services to special groups	2C Suicide rates among people with mental illness Employment rates Physical morbidity rates
(3) Individual level	3A Assessments of individual needs made by staff, service users, and by families Therapeutic expertise of staff Information for service users Information for family members	3B Content of therapeutic interventions (both psychological, social, and pharmacological) Continuity of clinical staff Frequency of appointments	3C Symptom severity Impact on caregivers Satisfaction with services Quality of life Disability Met and unmet needs

evaluation will depend upon the central question addressed and the level at which outcomes are assessed, as shown in Table 7.6.3.

Directly in relation to the population level, a frequently used outcome measure is suicide rate (see cell 1C in Table 7.6.1). Rates of homelessness among mentally ill people (or rates of mental illness among the homeless) can also be used as an outcome indicator of the effectiveness of mental illness policies at the national (or regional) level.

At the local level, outcome indicators useful for evaluation can be made in three ways: (i) by interpolating from regional/national data; (ii) by measuring directly at the local level; and (iii) by aggregating individual-level information up to the local level. For example, rates of suicide and unemployment can be estimated using the first method, or directly measured using the second approach if the appropriate data and resources exist, which will provide more accurate and up-to-date information. The third approach is to aggregate up to the local level information gathered from individual patients, if institutions providing care to those local patients are willing to cooperate in integrating their datasets.

At the individual level mental health service evaluation increasingly acknowledges the importance of outcomes other than symptom severity.[10,11] Traditionally, **symptom severity measures** have been used most often to assess the effectiveness of the early, mental health treatments. Psychiatrists and psychologists have contributed to the early development of such assessment scales to allow this

Table 7.6.2 Main purposes of mental health service evaluation

- To assess the outcomes of services in experimental conditions (*efficacy*)
- To investigate whether interventions which have demonstrated *efficacy* under experimental conditions are also *effective* in ordinary, routine clinical conditions
- To understand the mechanism of action (i.e. active ingredients) of interventions
- To inform mental health service investment decisions, for example using health economic data on cost-effectiveness
- To raise awareness among planners, policy makers, and politicians of service gaps
- To test a priori or to check *post hoc* the value of planning decisions (for example, the closure of mental hospitals)

Table 7.6.3 Outcome measures suitable for use in routine clinical practice

Outcome measure	Place dimension		
	Country level	Local level	Individual level
Employment status	√	√	√√
Physical morbidity	√	√	√
Suicide and self-harm	√√	√	√√
Homelessness		√	√√
Standardized mortality ratios	√	√	
Symptom severity		√	√√
Impact on caregivers		√	√
Satisfaction with services		√	√√
Quality of life		√	√
Disability		√	√√
Met and unmet needs for care		√	√

Key: √ = suitable for use as an outcome, √√ = commonly used as an outcome.

research to take place.[10,11] While the primary symptoms are clearly important, for most of the more severe mental disorders there is symptom persistence, and, at present, it is unrealistic to see symptom eradication as the sole aim of treatment. Therefore, very often, after the point of maximum symptom relief, when the extent of the ongoing impairments is clear, then the clinical task becomes one of attempting to minimize the consequent disability and handicap.

The importance of the **impact of caring** for people with mental illnesses upon family members and others who provide informal care has long been recognized, but has only been subjected to concerted research relatively recently.[11–13] Such research has shown that it is common for carers themselves to suffer from mental illnesses, most commonly depression and anxiety, and to worry about the future when they may no longer be able to cope. Moreover, many family members are most distressed by the patient's underactivity, and are often poorly informed about the clinical condition, its treatment, and the likely prognosis, as well as being inadequately provided with a practical action plan of what to do in the future should a crisis occur. Indeed, some services continue to convey to families the outmoded idea that carers, especially parents, are in some way to blame for the disorder or for relapses of the condition. The regular provision of information sessions for family members is now a hallmark of a good practice.[14,15]

Patients' **satisfaction with services** is a further domain that has recently become established as a legitimate, important, and feasible area of outcome assessment.[16] This is a recognition of the contribution that service users and their carers can make to outcome assessment. Psychometrically adequate scales in this field are those that adopt a multidimensional approach, assess the full range of service characteristics, are independently administered (so that patient ratings have no consequences upon their future clinical care), and have established validity and reliability.[17]

Quality of life ratings have also become prominent during the last decade, and several scales have been constructed that reflect different basic approaches to the topic.[18] The first distinction is between scales that address subjective well-being, compared with those that also measure objective elements of quality of life. The second main point of differentiation is between scales constructed for the general population and those designed for patients suffering from specific disorders, including the more severe mental illnesses.[19] One advantage of quality of life data is that they tend to be popular with politicians, for whom the concept often has powerful face validity.

Among people with longer-term or more complex mental illnesses, the measurement of **disability** is often an important consideration.[20] Increasing importance is also being attached to the **needs** of people with mental illness, where met needs are difficulties faced by people with mental illness in the presence of appropriate interventions.[21] Needs (both met and unmet) may be defined by professionals/experts, or by service users, and in fact there is emerging evidence that service user ratings may be more informative, for example in predicting quality of life.[22–24]

Psychometric properties of outcome measures

Establishing the psychometric qualities of scales used for service evaluation is a central issue.[4] Among the most important characteristics of outcome scales are validity and reliability. Validity refers to whether a scale actually measures what it is intended to measure. It is conventionally assessed in terms of face validity, content validity, consensual validity, criterion-related validity, and construct validity.

In addition, a rating scale must give repeatable results for the same subject when used under different conditions, i.e. it must be reliable. There are four widely used methods to gauge reliability: inter-rater reliability, test–retest reliability, parallel-form reliability, and split-half reliability. The main issue for the evaluation of mental health services is to use wherever possible scales with known and adequate psychometric properties.

How to evaluate mental health services

In this section we consider research designs that may be applicable to the range of contexts used in mental health service evaluation.[1] Different types of evidence produced using these designs cannot be considered as equivalent. A hierarchical order has been proposed by Geddes and Harrison[25] as shown in Table 7.6.4.

In terms of research methods or designs which can be used to produce such evidence, they can be considered as: (i) randomized controlled trial (RCT), (ii) quasi-experimental studies, (iii) case-control studies, (iv) cohort studies (prospective or retrospective), (v) cross sectional studies, and (vi) case series and single case studies. Since evaluations of mental health services are usually concerned with complex interventions, it is helpful to have an overall scheme linking different stages of research to test treatment interventions. The Medical Research Council (MRC) framework for the evaluation of complex interventions sets out one such sequence, as shown in relation to anti-stigma interventions in Table 7.6.5. The elements in this scheme can be considered as sequential, or stages 0, 1, and 2 can be seen as one larger iterative activity.[1] Nevertheless, although this gives salience to randomized controlled trial designs, it is important to appreciate that research study designs need to be matched to the purpose of each type of evaluation, as shown in Table 7.6.6.

Evidence from a meta-analysis of randomized controlled trials

Meta-analysis can be defined as 'the quantitative synthesis of the results of systematic overviews of previous studies', while systematic overviews, in turn, are methods of collating and synthesizing all the available evidence on a particular scientific question.[26] Since randomized controlled trials are often considered to produce the most sophisticated evidence on the efficacy of medical treatments,

Table 7.6.4 Hierarchy of evidence

1a	Evidence from a meta-analysis of RCTs
1b	Evidence from at least one RCT
2a	Evidence from at least one controlled study without randomization
2b	Evidence from at least one other type of quasi-experimental study
3	Evidence from non-experimental descriptive studies, such as comparative studies, correlation studies, and case-control studies
4	Evidence from expert committee reports or opinions and/or clinical experience of respected authorities

(Reproduced from J.R. Geddes, and P.J. Harrison, Closing the gap between research and practice, *The British Journal of Psychiatry*, **171**, 220–5, copyright 1997, The Royal College of Psychiatrists.)

Table 7.6.5 Phases of the Medical Research Council framework for the evaluation of complex interventions[1,2]

0 Preclinical	1 Modelling/manualization	2 Exploratory	3 Definitive trial	4 Long-term implementation
Explore relevant theory to ensure best choice of intervention and hypothesis and to predict major confounders and strategic design issues	Identify the components of the intervention and the underlying mechanisms by which they will influence outcomes to provide evidence that you can predict how they relate to and interact with each other	Describe the constant and variable components of a replicable intervention and a feasible protocol for comparing the intervention with an appropriate alternative	Compare a fully defined intervention with an appropriate alternative using a protocol that is theoretically defensible, reproducible, and adequately controlled in a study with appropriate statistical power	Determine whether others can reliably replicate your intervention and the results in uncontrolled settings over the long-term
Example: anti-stigma intervention in schools study				
Social contact theory[69]	Yes	Completed[70]	Planned	Potential if preceding phases successful

a meta-analysis conducted on well selected and relevant randomized controlled trials can be seen as the highest order of knowledge. It follows that the quality of systematic overviews is limited by the quality and quantity of the contributory trials (see Table 7.6.7).[27]

Cochrane was the first to emphasize the need to bring together, within specific categories, the results of randomized controlled trials.[28] This approach is now central to evidence-based medicine. Within psychiatric evaluation the first meta-analyses were conducted in the late 1970s, and more information is given on systematic reviews in Chapter 1.10 and 6.1.

An illustration of such an exercise is the systematic overview and meta-analysis is that which reviewed RCTs comparing the outcomes of community mental health teams with those of standard care for patients with severe mental illness and disordered personalities.[29,30] They found 1200 citations using the search strategy: 70 appeared relevant to the review, but only four studies satisfied the inclusion criteria. The main results of this systematic review are that community mental health team management is associated with fewer deaths by suicide, with fewer people being dissatisfied with services or leaving the studies early. No clear difference was found in admission rates, overall clinical outcomes, or in the duration of inpatient hospital treatment. The authors concluded that community mental health team management is not inferior to non-team standard care in any important respects, and is superior in promoting greater acceptance of treatment. It may also be superior in reducing hospital admissions and avoiding deaths by suicide.

The randomized controlled trials

The importance of randomized controlled trials within medical research has been expressed by Korn and Baumrind:[31] 'Randomized clinical trials are the sine qua non for evaluating treatment in man'. According to Barker and Rose: 'the essence of the randomized controlled trial is that the outcome of the treatment given to one group of patients is compared with one or more other groups who are given different treatments or none at all. Allocation of individuals to the treatment and comparison groups is by random selection'.[32]

Table 7.6.6 Research aims and appropriate study designs

Research aim	Appropriate study designs
Service description	Cross sectional survey
Assess intervention	Quasi-experimental study (e.g. controlled before–after comparison) Randomized controlled trial
Identify prognosis for a condition	Cohort study
Establish aetiology of a condition	Cohort study Case-control study

The advantages of the research design of these studies have been extensively described[33] and are shown in Table 7.6.8 (see also Chapter 1.10 in this book). However, the design limitations of such trials also need to be appreciated, particularly in relation to health service research, as shown in Table 7.6.9.[34]

In addition to the technical limitations of the trial design, there are also situations where randomized controlled trials are not applicable to specific research questions. These can be summarized as conditions in which randomized controlled trials are inadequate, impossible, inappropriate, or unnecessary, as shown in Table 7.6.10. Nevertheless, where they are appropriate, it will be necessary to use explicit criteria to assess the quality of such trials, such as those shown in Table 7.6.11.[35,36]

It is now common to distinguish between *efficacy trials* (which tend to be explanatory) and *effectiveness trials* (sometimes otherwise called large simple, pragmatic, practical, or management trials).[33,37–39] This categorical distinction has its uses, although

Table 7.6.7 Characteristics of systematic overviews

Questions to ask about papers for potential inclusion in a systematic overview

- Were the questions and methods clearly stated?
- Were comprehensive search methods used to locate the relevant articles?
- Were explicit methods used to determine which articles were included in the review?
- Was the methodological quality of the primary studies assessed?
- Were the selection and assessment of the primary studies reproducible and free from bias?
- Were the differences in individual study results adequately explained?
- Were the results of the primary studies combined appropriately?
- Were the reviewer's conclusions supported by the data cited?

(Reproduced from S.I. Sackett and J.E. Wennberg, Choosing the best research design for each question, *British Medical Journal*, **315** (7123), 1636, copyright 1997, BMJ Publishing Group Ltd.)

Table 7.6.8 Advantages of RCTs

- Controls for many confounding variables which may exist
- Eliminates the effects of spontaneous remission
- Eliminates regression to mean
- Eliminates placebo effect
- Independent of rater bias if blindness maintained
- Basis for systematic reviews

for some purposes we may rather see efficacy and effectiveness trials as falling along a continuum. Efficacy trials, which usually precede effectiveness studies, refer to those conducted under more ideal, experimental conditions, while effectiveness trials are RCTs carried out in more routine clinical conditions.[28,40–42] Nevertheless, some important questions, for example the impact of clinical guidelines, may only be researchable in real world settings, and will therefore bypass the efficacy study stage.[43]

Cochrane has defined effectiveness, at the patient level, as assessing whether an intervention does more good than harm when provided under usual circumstances of health care practice.[28] At the level of service provision, Wells has defined effectiveness trials as those which 'duplicate as closely as possible the conditions in the target practice venues to which study results will be applied'.[44] The key differences between efficacy and effectiveness trials are shown in Table 7.6.12, although in practice the differences between these types of trial may not necessarily be as great as the differences between pharmacological, psychological, and service interventions

Table 7.6.9 Limitations and disadvantages of RCTs designs

1. Difficulties in choosing the unit or level of random allocation
   - Should allocation be made at the patient level, the clinician level, the clinical team/practice level, or the locality level?
2. Difficulties in achieving random allocation
   - Randomization not possible
   - Particular patient groups excluded
   - Self-exclusion because of non-consent
3. Difficulties in obtaining consent and in maintaining motivation
   - Consent may be inversely proportional to severity of condition
   - Consent may be refused because of patient treatment preferences
   - Retention with the trial may be affected by patient motivation
4. Difficulties in establishing and maintaining blindness
   - Degree of blinding of subjects
   - Degree of blinding of staff
   - Degree of blinding of raters
   - Deactive Hawthorne effect (the effect of being studied upon those being studied)
5. Difficulties related to the experimental conditions
   - Concurrent multiple interventions in health service research trials (without a single potentially active ingredient)
   - Interactions between treatment components
   - Consistency of control ('usual treatment') conditions
   - High attrition rates or loss to follow-up

   Large differences between conditions in which trials can take place and those of routine practice

Table 7.6.10 Situations when RCT designs are not applicable

1. Situations in which experimentation is inadequate
   - Poor generalizability—low external validity
   - Unrepresentative staff included
   - Atypical patients included
   - Treatments not standardized
2. Situations in which experimentation is impossible
   - Refusal of clinicians to take part
   - Ethical objections to the study
   - Political barriers
   - Legal objections
   - Contamination between experimental and control conditions
   - Scale of task—trials are required for too many treatments
3. Situations in which experimentation is inappropriate
   - Studies conducted to reduce the occurrence of events of very low frequency
   - Studies to prevent unwanted outcomes in the distant future
4. Experimentation unnecessary
   - When benefit/risk ratio is dramatic
   - When there is a small likelihood of confounders

(Reproduced from N. Black (1996), Why we need observational studies to evaluate the effectiveness of health care, *British Medical Journal*, **312** (7040), 1215–18, copyright 1997, BMJ Publishing Group Ltd.)

(such as the dissemination, and related barriers, of proven interventions).[45–47]

In planning effectiveness RCTs, seven sets of issues need to be carefully considered: (i) study question (e.g. is the study question expressed in an answerable way?), (ii) reference population (e.g. what is the reference group or subgroup to which the trial results should be generalized?), (iii) patient sample (e.g. how far does the sample reflect the target population?), (iv) study settings (e.g. how representative are the study settings of routine clinical sites?), (v) study interventions (e.g. is the study intervention manualized, acceptable to patients, and suitable for widespread use?), (vi) control condition (e.g. are the key characteristics of the control condition well described, and do they vary within and between sites?), and (vii) bias (e.g. attrition, blinding, concealment, consent, and

Table 7.6.11 Criteria to evaluate the quality of an RCT

Criteria
1. Is the hypothesis clearly defined?
2. Is the study population representative?
3. Was patient assignment randomized?
4. Were patients, practitioners, and assessors blind to the experimental intervention?
5. Were the groups similar at the start of the trial?
6. Were the groups treated equally apart from the experimental intervention?
7. Were all those who entered the trial accounted for at its conclusion?
8. Was this in the groups to which they were originally allocated?
9. Are all clinically important outcomes considered?
10. Whose perspective do they reflect?
11. Is the data analysis appropriate?
12. What is the size and precision of the treatment effect?
13. Do the likely benefits outweigh the harms and risks?
14. Is the conclusion supported by the results?

Table 7.6.12 Key differences between efficacy and effectiveness trials

	Efficacy trials	Effectiveness trials
Goal	To estimate efficacy and safety (if relevant) usually of a specific clinical intervention	To estimate relative benefits and risks of approved treatments, clinical interventions, programmes, or policies
When	Usually before an intervention is introduced	Post-implementation
Diagnosis	Diagnosis by structured interview	Clinical diagnosis or structured interview
Inclusion and exclusion criteria	Strict and multiple inclusion and exclusion criteria, typically excluding patients with comorbid physical and psychiatric disorders	Relatively few inclusion and exclusion criteria to optimize external validity of sample
Patient sample	Typically enrol highly motivated patients	Attempt to include more representative patients, including those who are ambivalent and who may not adhere to the allocated treatment regime
Sample size	At most a few hundred, more often less than 100	Often larger to enable smaller effect sizes to be identified in heterogeneous populations (e.g. in large simple trials with dichotomous outcomes)
Comparator	Placebo and/or single active comparator (for drug trials) Treatment as usual or active control (for psycho-social interventions)	One or more active comparators (for drug trials) Treatment as usual or active control (for psycho-social interventions)
Dosing	Fixed or flexible	Flexible dosing in clinically used range
Blinding	Triple-blind (i.e. patients, staff, and researchers blind), or double-blind	Double-blind (where possible), or single-blind
Duration	1–4 months	6 months or more
Research sites	Small number of experienced research sites	Dozens of routine treatment sites
Delivery of intervention	According to manual or protocol, not a focus of research	Fidelity to manual a key variable, and study may consider barriers to delivery of intervention in routine practice[47]
Research protocol	Strictly defined	Deliberately similar to usual practice
Adjunctive treatments	Not allowed or strictly limited	Allowed as in usual practice
Outcomes	Symptom rating scales and other clinical parameters	A single well-defined, clinically important outcome (for large simple trials) and multiple secondary outcomes, including safety and costs (for practical trials)

(Reproduced from S. Stroup, Practical clinical trials for schizophrenia, *Epidemiologia e Psichiatria Sociale*, **14**, 132–6, copyright 2005 Il Pensiero Scientifico Editore.)

contamination). These issues are shown in more detail in Tables 7.6.13 and 7.6.14.[48]

Quasi-experimental studies

The term 'quasi-experiment' was first used to refer to a situation in which the decision about whether an individual does or does not receive the intervention to be evaluated is not under the investigator's control.[49] Random allocation of patients is therefore not made, so selection bias may occur. In other respects, a quasi-experiment aims to apply the logic of randomized controlled trials to the study design, and the researcher tries to reduce this bias by making the study units, in the groups being compared, as alike as possible in terms of the most important characteristics. This approach is known as ***matching***. Characteristics chosen for matching are those expected to influence the outcome (i.e. confounding factor). Therefore, the overriding aim of matching is to reduce the contribution made by the matched variables to the selection bias, although this method is inferior to randomization in that it cannot reduce the selection bias from all other variables.

There are two main approaches to matching. ***Paired matching*** consists of selecting individuals for the comparison group (or groups) who have closely similar characteristics to those included in the experimental group, for example in terms of age, gender, and occupation. This form of prestratification will need to be taken into account at the data analysis stage. A less rigorous variant is *group matching*, which only ensures that there are similar overall proportions of people, for both the experimental and comparison groups, in the various age bands, occupational groups, or other predefined strata used for the variables chosen for matching.

Non-experimental descriptive studies

The next type of research design included in the hierarchy of evidence is non-experimental descriptive studies. For the sake of clarity and brevity we shall distinguish two types of descriptive study: structured clinical practice and everyday unstructured clinical practice. An example of a descriptive evaluation design is the South Verona Outcome Study.[50,51] This is a prospective study, which aims to

Table 7.6.13 Criteria to plan effectiveness randomized controlled trials[48]

1. Study question
   - Who defines the aim of the study?
   - What process is used to identify the question addressed?
   - Is the study question expressed in an answerable way? (as a clear hypothesis)
   - Prior evidence of intervention effect size
   - Is the answer to this question really unknown?
   - Why is this question important now?
   - Is there initial evidence from efficacy trials or effectiveness studies? (observational or trials)
   - What is the public health importance of the policy or practice question addressed?
   - What is the clinical necessity of the question?
   - Sample size and statistical power for primary/secondary aims and related hypotheses
2. Reference population
   - What is the reference group (or subgroup) to which the trial results should be generalized?
   - What are their socio-demographic and clinical characteristics?
   - What are the ethnic and cultural characteristics of the target group?
   - What is resource level in this population?
   - What is the nature and standard and coverage of health and social care?
   - At what time point is population identified?
3. Patient sample
   - What are their socio-demographic, and clinical characteristics?
   - What are the inclusion criteria?
   - Not invited to participate rate
   - Non-participation rate
   - Patient preferences
   - What are the exclusion criteria?
   - How far does the sample reflect the target population?
   - What level of heterogeneity is there?
   - Selection of incident or prevalent cases (true incidence/prevalence or treated incidence/prevalence)
   - What are the rates of adherence and non-adherence to treatment as recommended?
4. Study settings
   - Characteristics and representativeness of professional staff
   - Levels of resources available
   - Research oriented culture
   - Staff morale and sustainability of intervention
   - Incentives for research collaboration
   - Opportunities for data linkage
   - Centre/professional non-participation
5. Study intervention
   - Is intervention acceptable?
   - Total time needed to deliver intervention
   - Frequency of interventions
   - Simplicity/complexity of the intervention
   - Single/multi-component intervention
   - Is intervention manualized?
   - Do usual professional staff deliver the intervention during the study?
   - Can treatment process be measured? (fidelity)
   - Degree of fit/feasibility for current practice
   - Exit strategy, who pays after the end of study
6. Control condition
   - Treatment as usual or specific control
   - Acceptability to patients of control condition
   - Cost and feasibility of control condition
   - Variation between control condition within and between sites (fidelity)
   - Are the key characteristics of the control condition well described?
7. Bias
   - Does contamination take place?
   - Degree of blinding
   - Choice of primary and secondary outcomes
   - Perspectives prioritized in outcome choice
   - Time(s) at which outcomes measured
   - Total length of follow-up and late effects
   - Sources of outcome data
   - Respondent burden
   - Consent rate
   - Recruitment rate
   - Attrition/drop-out and follow-up rates

(Reproduced from Tansella, M. Thornicroft, G., Barbui, C. *et al.* Seven criteria for improving effectiveness trials in psychiatry. *Psychological Medicine*, **36**(5) 711–200, copyright 2006, Cambridge University Press.)

assess the outcome of mental health care. Data from this study have been analysed using a multidimensional perspective. Among 354 patients followed up after 6 years of treatment in routine clinical settings the study revealed a complex pattern of emerging and disappearing clinical and patterns of exacerbation and remissions, with both changing frequently over time, but changes in both clinicial and social domains were not associated with diagnosis.[52]

Conclusions

Over the course of the next decade we expect that the following key trends will be of paramount importance. In relation to the focus of research, in many countries a degree of contestability may well develop, in which those who have traditionally identified questions to be addressed by research (investigators) will be challenged by research funders (such as governments and charities) and by the intended beneficiaries of the research (people with mental illness and their family members) to set the research agenda.[53] We can expect governments to direct their research investment towards policy challenges, such as barriers to the implementation of evidence-based practice.[54–56] This may well include the commissioning of research not just on new treatments and services, but also to evaluate already or even long-established models of care. For example, there is relatively little mental health service evaluation about: outpatient services (clinics), inpatient services, or forensic service provision.[57,58] In future, it will be important to evaluate *post hoc* current but unproven service configurations, particularly those that are widespread and expensive, as well as innovative interventions. This will necessitate providing sufficient long-term funding for health service research.

In terms of study design, we anticipate that there will be a relative growth of effectiveness studies, especially RCTs, which attempt to balance internal and external validity.[48] These will more often than in the past specify the precise nature of the control condition, use representative patient samples, and standardized outcomes measures. Less common study designs, such as cluster and preference RCTs will be necessary to tackle complex interventions.[59,60] The recent trend to more often include qualitative assessments

within RCTs we expect to accelerate, for example to identity the acceptability of interventions to patients, and to identify the active ingredients (and barriers) to treatment effectiveness.[61–65]

How will the conduct of evaluations of mental health services evolve in the coming years? We can identify a trend for study interventions to be increasingly often manualized. At the same time the patient populations treated will be more often similar to those treated in routine clinical practice. As a consequence, more attention will need to be paid to ways to incentivize clinical staff to participate in research.

Although traditionally research scientists have seen the dissemination of research findings largely in terms of publications in scientific journals,[66] it is likely that research funders will increasingly encourage or even insist upon using effective and target-specific communication methods to reach key audiences with the results of research and their implications, including non-traditional methods such as social marketing.[67] This will be intended to alter the behaviour of service planners, commissioners, practitioners, and service users, so that the results of research do influence the behaviour of these key groups in terms of service decisions, so that they increasingly reflect the evidence both of effective interventions and where the evidence shows lack of effect then to stop ineffective practices and services. Such behaviour change (and an established evidence base for this) may include methods as social marketing, as well as carefully combined interventions such as those used in case management in the treatment of depression.[68]

Table 7.6.14 Key challenges to the evaluation of mental health services

Focus of research
◆ Clarifying who is defining research questions
◆ Including service users and family members in setting research questions
◆ Asking clear research questions to answer important clinical challenges
◆ Evaluating already established as well new services
◆ Providing sufficient funding for long-term health service research
Study design
◆ Balancing internal and external validity of mental health service evaluation
◆ Moving from efficacy to effectiveness trials
◆ Specifying the precise nature of the control condition
◆ Using representative patient samples
◆ Using standardized outcome measures
◆ Combining qualitative and quantitative information
Conduct of research studies
◆ Manualizing the interventions to be evaluated
◆ Specifying the key characteristics of the patient groups to be treated
◆ Incentivizing clinical staff to participate in research
Data analysis and interpretation of findings
◆ Identifying the active ingredients of effective interventions
Dissemination of research findings
◆ Using effective and target-specific communication methods to reach key audiences with the results of research and their implications
Implementation of effective interventions
◆ Implementing the results of evaluation when the evidence is strong enough and decommissioning ineffective practice

Further information

Thornicroft, G. and Tansella, M. (1999). *The mental health matrix: a manual to improve services*. Cambridge University Press, Cambridge.

Thornicroft, G. and Szmukler, G. (2001). *Textbook of community psychiatry*. Oxford University Press, Oxford.

Tansella, M. and Thornicroft, G. (2001). *Mental health outcome measures* (2nd edn). Gaskell, Royal College of Psychiatrists, London.

Thornicroft, G. (2001). *Measuring mental health needs* (2nd edn). Gaskell, Royal College of Psychiatrist, London.

Fulop, N. and Allen, P. (2002). *Studying the organization and the delivery of the health services: research methods*. Routledge, London.

Thornicroft, G. and Tansella, G. (2004). The components of a modern mental health service: a pragmatic balance of community and hospital care. *The British Journal of Psychiatry*, **185**, 283–90.

Thornicroft, G., Becker, T., Knapp, M., *et al.* (2006). International outcome measures in mental health. Quality of life, needs, service satisfaction, costs and impact on carers. Gaskell, Royal College of Psychiatrists, London.

Slade, M. and Priebe, S. (2006). *Choosing methods in mental health research*. Routledge, London.

Knapp, M.J., McDaid, D., Mossialos, E., *et al.* (2007). *Mental health policy and practice across Europe*. Open University Press, Buckingham.

References

1. Campbell, N.C., Murray, E., Darbyshire, J., *et al.* (2007). Designing and evaluating complex interventions to improve health care. *British Medical Journal*, **334**(7591), 455–9.
2. Campbell, M., Fitzpatrick, R., Haines, A., *et al.* (2000). Framework for design and evaluation of complex interventions to improve health. *British Medical Journal*, **321**, 694–6.
3. Soanes, C. and Stevenson, A. (2003). *Concise Oxford English dictionary* (11th edn). Oxford University Press, Oxford.
4. Thornicroft, G. and Tansella, M. (1999). *The mental health matrix: a manual to improve services*. Cambridge University Press, Cambridge.
5. Thornicroft, G. and Tansella, M. (2007). *Better mental health care*. Cambridge University Press, Cambridge.
6. Sartorius, N. (1997). Evaluating mental health services. A world perspective. *Epidemiologia e Psichiatria Sociale*, **6**(Suppl. 1), 239–45.
7. Johnson, S., Nolan, F., Hoult, J., *et al.* (2005). Outcomes of crises before and after introduction of a crisis resolution team. *The British Journal of Psychiatry*, **187**, 68–75.
8. Killaspy, H., Bebbington, P., Blizard, R., *et al.* (2006). The REACT study: randomised evaluation of assertive community treatment in north London. *British Medical Journal*, **332**(7545), 815–20.
9. Glover, G., Arts, G., and Babu, K.S. (2006). Crisis resolution/home treatment teams and psychiatric admission rates in England. *The British Journal of Psychiatry*, **189**(5), 441–5.
10. Tansella, M. and Thornicroft, G. (eds.) (2001). *Mental health outcome measures*. Royal College of Psychiatrists, Gaskell, London.
11. Thornicroft, G., Becker, T., Knapp, M., *et al.* (2006). International outcome measures in mental health. Quality of life, needs, service satisfaction, costs and impact on carers. Gaskell, Royal College of Psychiatrists, London.
12. Schene, A., Tessler, R.C., Gamache, G.M., *et al.* (2001). Measuring family or care giver burden in severe mental illness: the instruments. In *Mental health outcome measures* (2nd edn) (eds. M. Tansella and G. Thornicroft), pp. 48–71. Royal College of Psychiatrists, Gaskell, London.
13. Joyce, J., Leese, M., and Szmukler, G. (2000). The experience of caregiving inventory: further evidence. *Social Psychiatry and Psychiatric Epidemiology*, **35**(4), 185–9.
14. Thornicroft, G. and Tansella, M. (2004). The components of a modern mental health service: a pragmatic balance of community and hospital care. *The British Journal of Psychiatry*, **185**, 283–90.

15. Szmukler, G., Kuipers, E., Joyce, J., *et al.* (2003). An exploratory randomised controlled trial of a support programme for carers of patients with a psychosis. *Social Psychiatry and Psychiatric Epidemiology*, **38**(8), 411–8.
16. Ruggeri, M. (2001). Measuring satisfaction with psychiatric services: towards a multi-dimensional, multi-axial assessment of outcome. In *Mental health outcome measures* (2nd edn) (eds. M. Tansella and G. Thornicroft), pp. 34–47. Royal College of Psychiatrists, Gaskell, London.
17. Ruggeri, M., Dall'Agnola, R., Agostini, C., *et al.* (1994). Acceptability, sensitivity and content validity of the VECS and VSSS in measuring expectations and satisfaction in psychiatric patients and their relatives. *Social Psychiatry and Psychiatric Epidemiology*, **29**(6), 265–76.
18. Lehman, A. (2001). Measures of quality of life for people with severe mental disorders. In *Mental health outcome measures* (2nd edn) (eds. M. Tansella and G. Thornicroft), pp. 72–92. Royal College of Psychiatrists, Gaskell, London.
19. Ware, J. and Sherbourn, C. (1992). The MOS, 36 item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, **30**, 473–83.
20. Wiersma, D. (2001). Measuring social disabilities in mental health. In *Mental health outcomes measures* (2nd edn) (eds. M. Tansella and G. Thornicroft), pp. 118–32. Royal College of Psychiatrists, Gaskell, London.
21. Thornicroft, G. (2001). *Measuring mental health needs* (2nd edn). Royal College of Psychiatrists, Gaskell, London.
22. Slade, M., Thornicroft, G., Loftus, L., *et al.* (1999). *CAN: the Camberwell Assessment of Need*. Gaskell, Royal College of Psychiatrists, London.
23. Lasalvia, A., Bonetto, C., Malchiodi, F., *et al.* (2005). Listening to patients' needs to improve their subjective quality of life. *Psychological Medicine*, **35**(11), 1655–65.
24. Slade, M., Leese, M., Ruggeri, M., *et al.* (2004). Does meeting needs improve quality of life? *Psychotherapy and Psychosomatics*, **73**(3), 183–9.
25. Geddes, J.R. and Harrison, P.J. (1997). Closing the gap between research and practice. *The British Journal of Psychiatry*, **171**, 220–5.
26. L'Abbe, K.A., Detsky, A.S., and O'Rourke, K. (1987). Meta-analysis in clinical research. *Annals of Internal Medicine*, **107**(2), 224–33.
27. Sackett, D.L., Rosenberg, W.M., Gray, J.A., *et al.* (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal*, **312**(7023), 71–2.
28. Cochrane, A. (1972). *Effectiveness and efficiency: random reflections on health services*. Nuffield Provincial Hospitals Trust.
29. Tyrer, P., Coid, J., Simmonds, S., *et al.* (2000). Community mental health teams (CMHTs) for people with severe mental illnesses and disordered personality. *Cochrane Database of Systematic Reviews*, (2), CD000270.
30. Simmonds, S., Coid, J., Joseph, P., *et al.* (2001). Community mental health team management in severe mental illness: a systematic review. *The British Journal of Psychiatry*, **178**, 497–502.
31. Korn, E.L. and Baumrind, S. (1991). Randomised clinical trials with clinician-preferred treatment. *Lancet*, **337**(8734), 149–52.
32. Barker, D. and Rose, G. (1979). *Epidemiology in medical practice* (2nd edn). Churchill Livingstone, London.
33. Everitt, B. and Wessely, S. (2004). *Clinical trials in psychiatry*. Oxford University Press, Oxford.
34. Black, N. (1996). Why we need observational studies to evaluate the effectiveness of health care. *British Medical Journal*, **312**(7040), 1215–18.
35. Marriott, S. and Palmer, C. (1996). Clinical practice guidelines: on what evidence is our clinical practice based? *Psychiatric Bulletin*, **20**, 363–6.
36. Macfarlane, W., Dushay, R., Stastny, P., *et al.* (1996). Comparison of two levels of family-aided assertive community treatment. *Psychiatric Services*, **47**, 744–50.
37. Schwartz, D. and Lellouch, J. (1967). Explanatory and pragmatic attitudes in therapeutic trials. *Journal of Chronic Diseases*, **20**(20), 637–48.
38. Peto, R., Collins, R., and Gray, R. (1993). Large scale randomised evidence. *Annals of the New York Academy of Science*, **703**, 314–40.
39. Oliver, S. (1997). Exploring lay perspectives on questions of effectiveness. In *Non random reflections on health services research* (eds. A. Maynard and I. Chalmers), pp. 272–91. British Medical Journal Publications, London.
40. Haynes, B. (1999). Can it work? Does it work? Is it worth it? The testing of healthcare interventions is evolving. *British Medical Journal*, **319**(7211), 652–3.
41. Lilienfeld, A. (1982). The Fielding H. Garrison lecture: ceteris paribus: the evolution of the clinical trial. *Bulletin of the History of Medicine*, **56**, 1–18.
42. Pocock, S. (1983). *Clinical trials: a practical approach*. Wiley, London.
43. Andrews, G. (1999). Randomised controlled trials in psychiatry: important but poorly accepted. *British Medical Journal*, **319**(7209), 562–4.
44. Wells, K.B. (1999). Treatment research at the crossroads: the scientific interface of clinical trials and effectiveness research. *The American Journal of Psychiatry*, **156**(1), 5–10.
45. Proudfoot, J., Goldberg, D., Mann, A., *et al.* (2003). Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. *Psychological Medicine*, **33**(2), 217–27.
46. Wells, K.B., Sherbourne, C., Schoenbaum, M., *et al.* (2000). Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. *The Journal of the American Medical Association*, **283**(2), 212–20.
47. Campbell, M., Fitzpatrick, R., Haines, A., *et al.* (2000). Framework for design and evaluation of complex interventions to improve health. *British Medical Journal*, **321**(7262), 694–6.
48. Tansella, M., Thornicroft, G., Barbui, C., *et al.* (2006). Seven criteria for improving effectiveness trials in psychiatry. *Psychological Medicine*, **36**(5), 711–20.
49. Campbell, D.T. and Stanley, J.C. (1996). *Experimental and quasi-experimental designs for research*. Rand-McNally, Chicago, IL.
50. Ruggeri, M., Biggeri, A., Rucci, P., *et al.* (1998). Multivariate analysis of outcome of mental health care using graphical chain models. The south-Verona outcome project 1. *Psychological Medicine*, **28**(6), 1421–31.
51. Lasalvia, A. and Ruggeri, M. (2007). Multidimensional outcomes in 'real world' mental health services: follow-up findings from the south Verona project. *Acta Psychiatrica Scandinavica*, (Suppl.).
52. Lasalvia, A., Bonetto, C., Cristofalo, D., *et al.* (2007). Predicting clinical and social outcome of patients attending 'real world' mental health services: a 6 year multi-wave follow-up. *Acta Psychiatrica Scandinavica*, **116**, (s437), 16–30.
53. Chamberlin, J. (2005). User/consumer involvement in mental health service delivery. *Epidemiologia e Psichiatria Sociale*, **14**(1), 10–4.
54. Magnabosco, J.L. (2006). Innovations in mental health services implementation: a report on state-level data from the U.S. Evidence-based practices project. *Implemention Science*, **1**, 13.
55. Drake, R.E., Becker, D.R., Goldman, H.H., *et al.* (2006). Best practices: the Johnson & Johnson—Dartmouth community mental health program: disseminating evidence-based practice. *Psychiatric Services*, **57**(3), 302–4.
56. Aarons, G.A. and Sawitzky, A.C. (2006). Organizational culture and climate and mental health provider attitudes toward evidence-based practice. *Psychological Services*, **3**(1), 61–72.
57. Szmukler, G. and Holloway, F. (2001). In-patient treatment. In *Textbook of community psychiatry* (eds. G. Thornicroft and G. Szmukler), pp. 321–37. Oxford University Press, Oxford.
58. Becker, T. (2001). Out-patient psychiatric services. In *Textbook of community psychiatry* (eds. G. Thornicroft and G. Szmukler), pp. 277–82. Oxford University Press, Oxford.

59. Medical Research Council. (2002). *Cluster randomised trials: methodological and ethical considerations*. Medical Research Council, London.
60. Howard, L. and Thornicroft, G. (2006). Patient preference randomised controlled trials in mental health research. *The British Journal of Psychiatry*, **188**, 303–4.
61. Lester, H., Tritter, J.Q., and Sorohan, H. (2005). Patients' and health professionals' views on primary care for people with serious mental illness: focus group study. *British Medical Journal*, **330**(7500), 1122.
62. Pope, C., Mays, N., and Popay, J. (2006). How can we synthesize qualitative and quantitative evidence for healthcare policy-makers and managers? *Healthcare Management Forum*, **19**(1), 27–31.
63. Mays, N., Pope, C., and Popay, J. (2005). Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. *Journal of Health Services Research & Policy*, **10**(Suppl. 1), 6–20.
64. Pope, C., Ziebland, S., and Mays, N. (2000). Qualitative research in health care. Analysing qualitative data. *British Medical Journal*, **320**(7227), 114–16.
65. Mays, N. and Pope, C. (2000). Qualitative research in health care. Assessing quality in qualitative research. *British Medical Journal*, **320**(7226), 50–2.
66. Lewison, G., Thornicroft, G., Szmukler, G., *et al*. (2007). The fair assessment of the merits of psychiatric research. *The British Journal of Psychiatry*, **190**, 314–18.
67. Kotler, P., Roberto, E.L., and Lee, N. (2002). *Social marketing: improving the quality of life*. Sage, New York.
68. Wells, K., Miranda, J., Bruce, M.L., *et al*. (2004). Bridging community intervention and mental health services research. *The American Journal of Psychiatry*, **161**(6), 955–63.
69. Thornicroft, G. (2006). *Shunned: discrimination against people with mental illness*. Oxford University Press, Oxford.
70. Pinfold, V., Toulmin, H., Thornicroft, G., *et al*. (2003). Reducing psychiatric stigma and discrimination: evaluation of educational interventions in UK secondary schools. *The British Journal of Psychiatry*, **182**, 342–6.
71. Sackett, D.L. and Wennberg, J.E. (1997). Choosing the best research design for each question. *British Medical Journal*, **315**(7123), 1636.
72. Stroup, S. (2005). Practical clinical trials for schizophrenia. *Epidemiologia e Psichiatria Sociale*, **14**, 132–6.

7.7

Economic analysis of mental health services

Martin Knapp and Dan Chisholm

Introduction

Economics is concerned with the use and distribution of resources within a society, and how different ways of allocating resources impact on the well-being of individuals. Economics enters the health sphere because resources available to meet societal needs or demands are finite, meaning that choices have to be made regarding how best to allocate them (typically to generate the greatest possible level of population health). Economics provides an explicit framework for thinking through ways of allocating resources.

Resource allocation decisions in mental health are complicated by the fact that disorders are common, debilitating, and often long-lasting. Epidemiological research has demonstrated the considerable burden that mental disorders impose because of their prevalence, chronicity, and severity: globally, more than 10 per cent of lost years of healthy life and over 30 per cent of all years lived with disability are attributable to mental disorders.[1] Low rates of recognition and effective treatment compound the problem, particularly in poor countries.

However, disease burden is not in itself sufficient as a justification or mechanism for resource allocation or priority-setting. A disorder can place considerable burden on a population but if appropriate strategies to reduce this burden are absent or extremely expensive in relation to the health gains achieved, large-scale investment would be considered misplaced. The reason is that scarce resources could be more efficiently channelled to other burdensome conditions for which cost-effective responses *were* available. For priority-setting and resource allocation, it is necessary to ask what amount of burden from a disorder can be avoided by using evidence-based interventions, and at what relative cost of implementation in the target population.

Cost and cost-effectiveness considerations enter into health care reform processes, priority-setting exercises within and across health programmes, and regulatory decisions concerning drug approval or pricing. Two broad levels of economic analysis can be distinguished: macro and micro.

Economic analyses at macro level: the mental health system

Macro-level economic analyses are concerned with how health systems function and what they achieve. What, for example, are the motivations and behaviour of key 'stakeholders', with what implications for access, quality, and costs? What roles do economic forces play and can they be shaped to improve health outcomes and cost-effectiveness? Do different organizational or financial arrangements produce different resource configurations? For example, do markets achieve fairer or more efficient allocations than state bureaucracies? Through their macro analyses, economists can contribute to a better understanding of how health systems can improve utilization of resources (for example, see Box 7.7.1).

While improved psychological well-being in the population is likely to represent the primary goal of the mental health system, there are other (social) goals that could also feature prominently, including quality improvements in service provision, and financial (as well as human rights) protection for people with mental health problems.[2] Meeting these goals is achieved via a number of key health system functions, including resource generation, their allocation via appropriate modes of financing, actual provision of services, and overall stewardship and evaluation of these various functions.[3] Economic analysis contributes to policy formation relating to each of these functions.[4–6]

There are barriers to implementation of evidence-based mental health care in even generously resourced health systems. An insidious barrier is resource insufficiency: mental health services are often grossly under-funded in comparison to both needs and the

Box 7.7.1 The relevance of a health systems perspective

The need for a systems approach to mental health policy and planning is made apparent from a simple illustration: cheap, effective drugs exist for key neuropsychiatric disorders, including tricyclic antidepressants, conventional neuroleptics, and antiepileptic drugs, which are affordable even to resource-poor countries. The availability and prescription of these drugs to those in need, however, are determined by the extent to which such drugs have been distributed and by the ability of health care providers to detect and appropriately treat the underlying condition. Access to and use of such medications may further be hampered by the private cost of seeking and receiving health care, particularly if it is out-of pocket. User fees, provider incentives, and clinical practice are in turn influenced by the availability of national legislation, regulation, and treatment guidelines.

cost-effectiveness of interventions to meet them. This is a major issue for countries where the proportion of national income devoted to health care is low, or where the proportion of the health budget allocated to mental health is minimal. With limited funds it is difficult to build any kind of service system, because it is difficult to recruit, train, and retain skilled staff.

Even when resources are committed, available services might be poorly distributed, available at the wrong place or time relative to the distribution of needs. They may only be delivered by specialist clinics or concentrated in big cities, or affordable only by wealthier individuals. Improvements to practice take time to work through to improved health outcomes, cost-effectiveness gains or fairer access, even when suitable professionals can be recruited or new facilities opened. Decision makers must think long-term, for the immediate consequences of many interventions could be modest but longer-term benefits immense.[7]

A more general difficulty is that available services do not match what is needed or preferred. Indeed, there may be scant information on population or individual needs, and patients may have few opportunities to participate in treatment decisions. Another problem could be poor coordination of services because of professional rivalry, stultifying bureaucracy or 'silo budgeting' (resources held in one agency's 'silo' cannot be allocated to other uses).

Economic analyses at micro level: cost-effectiveness

The most frequently posed micro questions relate to the cost-effectiveness of interventions, such as an emphasis on community-based care, the use of new drugs, or the development of secure accommodation. For example, consider what happens following development of a new treatment (say a new medication for schizophrenia). Decision makers want answers to two questions when considering whether to use or recommend this drug. The first is the clinical question: is it effective in alleviating psychotic symptoms and generally improving health-related quality of life? If the answer to the clinical question is 'yes', then there is a second question: is it cost-effective? That is, does the drug achieve the improved outcomes at a cost that is worth paying? The meaning of 'worth' is far from straightforward to establish and laden with controversy.

These two questions sit at the heart of economic evaluation: outcomes must be assessed (and compared between different treatments) and the (relative) costs of achieving them must be examined. Looking only at costs is *not* an economic evaluation.

There are different variants of economic evaluation. They share a common approach to the conceptualization, definition, and measurement of costs, but adopt different approaches when considering and measuring outcomes, primarily because they seek to answer slightly different questions. We set out these differences by discussing the questions a study might address, measuring costs and outcomes, making trade-offs between them, and utility and benefit measurement.

Question and perspective

The choice of evaluative approach depends on the question to be addressed. If the question is essentially clinical—what is the most appropriate treatment for someone with particular needs in particular circumstances—information is needed on the comparative costs of alternative treatments and comparative outcomes measured in terms of symptom alleviation, improved functioning, and so on. A cost-effectiveness analysis would be appropriate (see 'Effectiveness measurement' below).

If, to take a broader stance, the question is whether to treat depression rather than spending the funds elsewhere in the health system, then decision makers need to know the costs, but now need an outcome measure that uses a common metric across different health domains. The most common such metric is 'utility' and a cost-utility analysis would be undertaken (see 'Utility measurement' below).

To widen the perspective further, if the question is whether to increase expenditure in the health system or in (say) improving transport or launching a new environmental policy, then an evaluation needs to ask about the comparative costs and impacts of the different options, where 'impact' will need to be measured in a common unit across all public policy areas. The usual choice of broad measure is monetary, leading to cost–benefit analysis (see 'Benefit measurement' below).

The question to be addressed thus influences the type of evaluation needed, but the choices are not mutually exclusive: a single study can support more than one approach if the right measures are used. The broader the question, the lower the likelihood that the outcome measure will be sensitive to the particular circumstances of a specific disorder such as depression, but the greater the usefulness in terms of resource allocation decisions.

Linked to specification of the question to be addressed is the *perspective* of a study. Is the evaluation needed to help resource allocation within a particular agency (such as primary care clinic), or a particular system (such as the health care system), or the whole society? The perspective will determine the breadth of both cost and outcome measurement.

Cost measurement

Some costs are directly associated with a disorder or its treatment, such as the money spent on medications and services used by patients, and some are more indirect, measuring lost productivity because ill-health can disrupt someone's employment pattern or the social cost of unpaid care provided by families. How broadly the costs are measured will depend upon the purpose of the study.

In carrying out evaluations in practice, economists need data on service use patterns by patients. This information might come from organizational 'billing' systems (recording amounts transferred between purchasers and providers for services used), or from routine information systems that record service contacts, or from research instruments that specifically collect data on service use patterns through interviews with patients, caregivers, or service professionals. One widely used instrument is the Client Service Receipt Inventory.[8]

The next task is to attach unit cost estimates to these service use data. In England, there is an excellent annual compendium of health and social care unit costs, which provides just such figures.[9] In other countries, it might be necessary to estimate unit costs anew. A range of data sources could be used, including government statistics, health system expenditure figures, and specific facility or organization accounts. The main cost categories to be quantified would be:

- salaries of staff employed in patient treatment and care
- facility operating costs (e.g. cleaning, catering)

- overhead costs (e.g. personnel, finance)
- capital costs for buildings and durable equipment

Effectiveness measurement

The most intuitive mode of economic evaluation is cost-effectiveness analysis (CEA): it measures costs as set out above, and outcomes along the dimensions that would be recognized by clinicians and used in clinical studies (changes in symptoms, behaviour, functioning, and so on). A CEA can help decision makers choose between interventions aimed at specific health needs. A cost-effectiveness analysis looks at a single outcome dimension—such as change in symptoms—and computes and compares the difference in costs between two treatments and the difference in this (primary) outcome. If one treatment is both more effective and less costly than another, then it would clearly be the more cost-effective of the two. But if it is more effective and *more* costly then a trade-off is needed (see below).

Often the economist will compute cost differences and a range of effectiveness differences (one for each outcome dimension)—an approach sometimes called cost-consequences analysis—which has the advantage of breadth but poses a challenge if one outcome is better and another worse for a particular treatment. It is then not always obvious which treatment is to be preferred, and the decision maker must weigh up the strength of evidence.

Making trade-offs

If an evaluation finds a new intervention to be both more effective but simultaneously more expensive than an older intervention, which is the more cost-effective of the two? A trade-off must be made between the better outcomes and the higher costs necessary to achieve them.

The classical way of determining this trade-off has been via the derivation of a incremental cost-effectiveness ratio (CER), which divides the extra cost associated with a new intervention by its additional effect (see Box 7.7.2).

More recently, health economists have developed the 'net benefit approach' to explicate the nature of the trade-off (see Box 7.7.3 for an example). It is commonly seen today in the construction of *cost-effectiveness acceptability curves* (CEACs). These curves show the probability that an intervention will be cost-effective for each of a number of pre-specified or implicit valuations of an outcome improvement by the decision maker.

Utility measurement

One way to overcome the potential problem of different outcome dimensions pointing in different directions is to employ a single, over-arching measure. A preference-weighted, health-related quality of life measure could be used. The value of the quality of life improvement is gauged in units of 'utility', usually expressed by a combined index of the mortality and quality of life effects of an intervention. The best known such index is the Quality Adjusted Life Year (QALY).

A cost-utility analysis (CUA) measures the outcome difference between two interventions in terms of QALY gain, and compares this with the difference in costs. CUAs have a number of attractions, including a unidimensional, generic outcome measure that allows comparisons across diagnostic groups, based on an explicit methodology for weighting preferences and valuing health states. But the utility measure may be too reductionist and insufficiently sensitive to changes expected in a particular clinical area such as depression treatment.[11] Nevertheless, cost-utility analyses produce estimates of cost-per-QALY gain from one therapy over another, which can then inform health care resource allocation decisions, such as by the National Institute for Health and Clinical Excellence (NICE) in England and Wales.

Box 7.7.2 Calculation of the incremental cost-effectiveness ratio

For example, evaluation of a new antidepressant may show that the cost per treated case of major depression is an extra £500 per year but also results in a lower average symptom score over this period (such as a ten point reduction on the Beck Depression Inventory); the resulting incremental CER would therefore be £50 [£500/10], meaning that each additional unit improvement on the BDI cost £50.

Box 7.7.3 Example of the net benefit approach in mental health services research

An example of the use of cost-effectiveness acceptability curves comes from a study of computer-delivered cognitive behavioural therapy (CCBT) for anxiety and depression.[10] CCBT was more expensive in health service terms than standard primary care services, but more effective in reducing symptoms. The fitted CEACs showed that, even if the value placed by society on a unit reduction in the Beck Depression Inventory (the primary clinical measure used in the trial) was as little as £40, there was an 81 per cent probability that CCBT would be viewed as cost-effective. Similarly, assigning a societal value of just £5 to each additional depression-free day would result in an 80 per cent probability that CCBT would be cost-effective. The CEAC makes transparent the trade-offs faced by decision makers.

Benefit measurement

Cost–benefit analysis asks whether the benefits of a treatment or policy exceed the costs, helping decision makers to allocate resources across a wide area, for example comparing health care with housing, education, or defence. All costs and outcomes (benefits) are valued in the same (monetary) units. If benefits exceed costs, the evaluation would provide support for the intervention or programme, and vice versa. With two or more alternatives, the intervention with the greatest net benefit would be deemed the most efficient. Cost–benefit analyses are thus intrinsically attractive, but conducting them is especially problematic because of the difficulties associated with attaching monetary values to health outcomes, and especially mental health outcomes. Methodological advances in health economics offer ways to obtain direct valuations of health outcomes by patients, families, or others,[12] but they will not be easy to apply in mental health contexts.

Design issues

As in clinical evaluation, an important consideration for the review, assessment, and interpretation of economic evidence is research

design. Generally speaking, the ideal type of study upon which to base decisions on cost-effectiveness and resource allocation is the one conducted prospectively with two (or more) appropriately sized randomly allocated groups of patients, for whom all conceivable costs and outcomes are measured appropriately, including a comparable measure of outcome (monetized benefits or, more realistically, a utility metric).

Looking across the mental health field, the accumulation of new cost-effectiveness evidence has been uneven, tending to be greater in diagnostic areas where new classes of medication have been launched: the pharmaceutical industry looks to economic evidence to support its marketing. At the same time, health care funding and delivery bodies also want their own independent evidence on new therapies. Consequently, a lot of economic studies of depression followed the licensing of the early selective serotonin-reuptake inhibitors (SSRIs) and later antidepressants with other mechanisms of action. Similarly, the arrival of the atypical antipsychotics and the cholinesterase inhibitors for Alzheimer's disease stimulated a lot of economics research.

We cannot cover all mental health areas here. Instead, in the next three sections, we look at areas where there has been some interesting activity:

- cost-utility analysis of depression treatment in primary care
- cost-effectiveness of interventions for child and adolescent mental health problems
- sectoral cost-effectiveness analysis of mental health interventions in developing countries.

Cost-utility analysis of depression treatment in primary care

Ten years ago, examples of the application of the cost-utility approach to mental health were hard to find.[11] Since then, there has been an increasing use of the so-called cost-per-QALY approach in mental health evaluations, following recommendations for such analyses by regulatory bodies in Australia, Canada, the United Kingdom, and the United States. In the field of depression, for example, cost-utility analyses have now been carried out for: screening in primary care[13] (annual and periodic screening cost more than $50 000 per QALY, one-off screening below this threshold); newer versus older antidepressant drugs[14]; maintenance treatment for recurrent depression[15,16]; guideline-concordant primary care treatment for women[17]; primary care practice-initiated quality improvement programmes[18]; computerized cognitive behavioural therapy[10]; and ECT versus transcranial magnetic stimulation for 'treatment-resistent' depression.[19] Many of the cost-utility analyses carried out to date employ secondary data and modelling techniques to estimate costs and effects, others have constructed cost-per-QALY estimates alongside clinical trials.

An example of a cost-utility analysis using modelling is that by Revicki *et al.*[14] who compared treatment for major depression with (a) newer antidepressants (nefazodone and fluoxetine) (b) tricyclics (imipramine) and, for treatment failures, (c) a step approach involving initial treatment with imipramine followed by nefazodone. A decision analysis model was developed to simulate the clinical management pathways and pattern of recurrences of major depression for these alternative treatment strategies to estimate lifetime medical costs and health outcomes (expressed as QALYs). There were only minor differences in costs and QALYs between nefazodone and fluoxetine, and both these newer antidepressants were estimated to be cost-effective compared to imipramine treatment and the imipramine step approach. The ratios of cost to QALYs gained for these newer antidepressants were deemed to be sufficiently low (below $20 000 per QALY gained) to merit adoption of these treatments in the health system. For example, the extra lifetime cost of nefazodone over imipramine ($1321) resulted in 0.32 added QALYs, giving a ratio of $4065 per QALY gained. Since decision models and their findings are only as good as their underlying assumptions and the quality of the data used to estimate key model parameters, extensive sensitivity analyses were conducted, but these did not alter the conclusions. However, the results did not include indirect costs such as changes in work productivity—important for a societal perspective—and are not readily generalizable to groups other than the targeted population (in this study, 30-year old women with one previous depressive episode).

An example of the empirically based generation of cost-per-QALY information is the randomized controlled trial of practice-initiated quality improvement (QI) for depression,[18] which involved group-level randomization of 46 clinics in six community-based US managed care organizations, either to medication or psychotherapy quality improvement programme (in addition to training and enhanced educational resources). Two QALY measures were derived, one from the Short-From, 12-Item Health Survey (SF-12) plus a standard gamble utility weighting exercise among a local convenience sample, the other with reference to estimated time spent depressed plus values from the literature for lost utility due to depression. Relative to usual care, average health care costs increased by $400–500 per treated patient, while QALY gains were less than 0.025, resulting in an estimated cost per QALY of $15 000–36 000 (QI-medication) and $9500–21 500 (QI-therapy). In addition to these health gains, patients exposed to the quality improvement programmes were employed more days than those receiving usual care.

The envisaged benefit of expressing the results of economic evaluation in these terms lies in the ability to line-up cost-per-QALY estimates for a range of different interventions and disorders, with a view to determining acceptable efficiency against a pre-defined threshold (of, say, $50 000 in the US context), or even constructing 'league tables' summarizing best and worst buys in the health sector. In practice, there remain significant problems in relying on league tables for allocating resources (due to the heterogeneous and context-specific nature of cost-utility studies), while there may be criteria unrelated to efficiency that determine whether a particular intervention is deemed acceptable for reimbursement or inclusion in a defined package of basic health care.

Cost-effectiveness of interventions for child and adolescent mental health problems

There are hundreds of completed economic evaluations in the depression field, almost all confined to adults of 'working age'. But there is surprisingly little economic evidence on child and adolescent mental health interventions. A systematic review a few years ago found only 14 published economic evaluations, some of rather poor quality.[20] Common problems included small sample sizes, narrow cost measures, short follow-ups, and limited outcome measures. (Guidelines and quality checklists are available for health economist researchers and readers of their outputs.)[21]

Another drawback is that most of the completed economic studies have been undertaken in North America, the United Kingdom, and Australia. But the results of economic evaluations generally do not transfer easily from one health system to another because of differences in system structure and financing, leading to differences in relative costs. It is infeasible and certainly unnecessary to carry-out an evaluation every time a policy decision needs to be taken, but it is also difficult to assess the relevance of economic evidence from another country, especially if its mental health system is markedly different.

An example of a well-conducted cost-effectiveness analysis is the evaluation of a home-based social work intervention for children and adolescents who have deliberately poisoned themselves.[22] The researchers measured suicidal ideation, hopelessness, and family functioning as the main outcomes, and costs were based on patterns of utilization of health, education, social care, and voluntary sector services. Within a randomized controlled trial, involving 162 children aged 16 years or under, they found no significant difference in the main outcomes or costs, although parental satisfaction with treatment was significantly greater in the group that received a new social work intervention compared to those who received routine care.

In another pragmatic randomized trial, a parenting intervention for parents of children at risk of developing conduct disorder (the Incredible Years programme) was compared to wait-list controls. The perspective for cost measurement was the public sector (health, social care, special education); effectiveness was measured by reductions in intensity of behaviour problems.[23] The Incredible Years programme was more effective but also more costly. The researchers found that it would cost £1344 to bring the average child in the intervention group (in terms of behaviour intensity score) to below the clinical cut-off point. A cost-effectiveness acceptability curve was plotted to show the trade-offs between cost and effectiveness.

Sectoral cost-effectiveness analysis of mental health interventions in developing countries

Cost-effectiveness analysis can also be used to evaluate mental health programmes for whole populations (countries or even world regions). While the burden of neuropsychiatric disease is very high, the resources available to address that burden are extremely low. Given the consequent tension between the need for and the availability of mental health care, plus the fact that effective interventions do exist, the job of cost-effectiveness analysis is to show how much of the burden can be reduced or averted, by doing what, and at what cost.

Through its CHOICE project (choosing interventions that are cost-effective), WHO embarked on an initiative to assemble databases on cost-effectiveness of key health interventions in 14 epidemiological sub-regions of the world.[24] A comparative cost-effectiveness analysis of interventions for reducing the burden of major neuropsychiatric disorders formed part of this programme.[6,25,26] WHO-CHOICE advocates a 'generalized' form of cost-effectiveness analysis, in which costs and effects of current and new interventions are compared to the starting point of 'doing nothing'. Accordingly, the costs and effectiveness of pharmacological and psychosocial interventions in primary care or outpatient settings for psychiatric disorders were compared in a population model to an epidemiological situation representing the untreated natural history of these disorders. Effects are measured as disability adjusted life years (DALYs) averted (i.e. reduced burden), and costs in international dollars (I$; one international dollar should buy the same quantity of health care resources in China as in the United States).

Compared to no treatment (natural history), the most cost-effective strategy for averting the burden of psychosis and severe affective disorders in developing regions of the world is a combined intervention of first-generation antipsychotic or mood-stabilizing drugs with adjuvant psychosocial treatment delivered by community-based outpatient services, with cost-effectiveness ratio of I$4200–5500 in Sub-Saharan Africa and South Asia, rising to more than I$10 000 in middle-income regions[25] (see Fig. 7.7.1). Currently, the high acquisition price of second-generation antipsychotic drugs makes their use in developing regions questionable on efficiency grounds alone, although this situation stands to change as these drugs come off patent. By contrast, evidence indicates that the relatively modest additional cost of adjuvant psychosocial treatment reaps significant health gains, thereby making such a combined strategy for schizophrenia and bipolar disorder treatment more cost-effective than pharmacotherapy alone.

For more common mental disorders treated in primary care settings (depressive and anxiety disorders), the single most cost-effective strategy is the scaled-up use of older antidepressants (due to their lower cost but broadly similar efficacy to newer antidepressants). However, as the price margin between older and generic newer antidepressants continues to narrow, generic SSRIs should be at least as cost-effective and may therefore represent the treatment of choice in the future. Since depression is commonly recurring, there are also grounds for thinking that proactive care management, including long-term maintenance treatment with antidepressant drugs, represents a cost-effective (if more resource-intensive) way of significantly reducing the enormous burden of depression in developing regions.[27]

The purpose of such an exercise is to locate the relative position of effective and applicable interventions within a wider cost-effectiveness and priority-setting framework. Using the affordability criteria of the WHO Commission for Macroeconomics and Health,[28] this analysis indicates that (a) the most efficient interventions for common mental disorders can be considered very cost-effective (each DALY averted costs less than 1 year of average per capita income), and (b) community-based interventions for severe mental disorders using older antipsychotic and mood-stabilizer drugs meet the criterion for being cost-effective (each DALY averted costs less than three times GDP per capita). These findings therefore provide relevant information regarding the relative value of investing in neuropsychiatric treatment and prevention, and so may help to remove one of many remaining barriers to a more appropriate public health response to mental health needs.

Conclusion

Economic evaluation provides a means of comparing the costs and outcomes of mental health interventions or programmes, enabling decision makers to assess whether they offer good use of (scarce) resources. An analysis of costs alone, or indeed of outcomes alone,

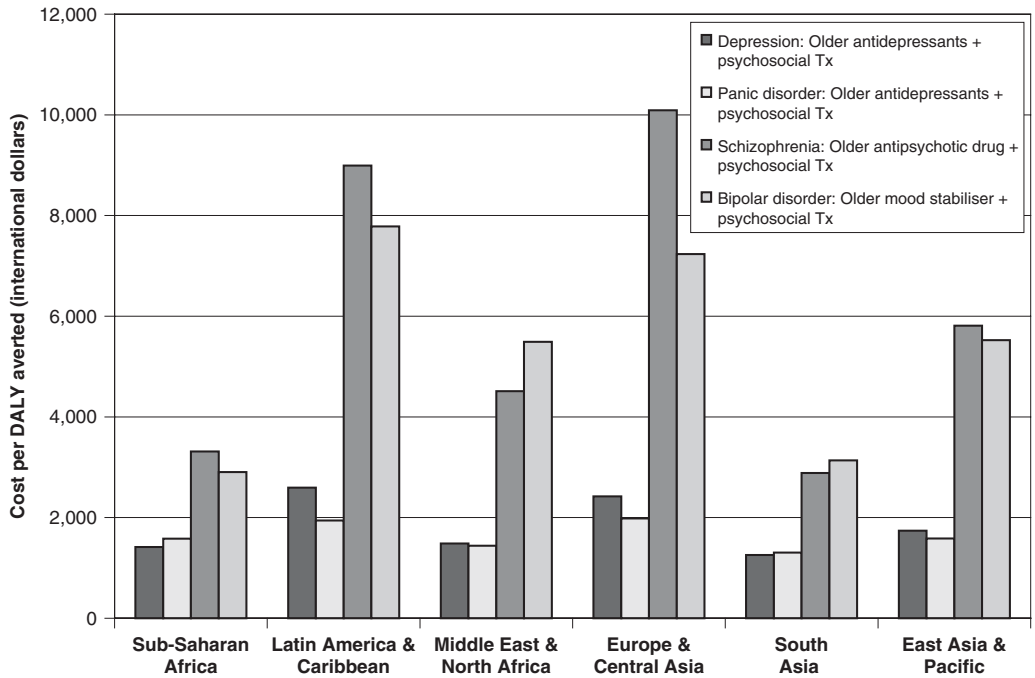


Fig. 7.7.1 Cost-effectiveness ratios for a basic mental health package in low and middle-income regions of the world.

does not provide such information. The results of well-conducted economic evaluations can be channelled into decision-making processes at a succession of levels.

Patients, users, and caregivers

Economic evidence can complement clinical decision-making at the patient, user, and caregiver level by comparing costs and consequences of particular treatments. One very pertinent question about any treatment is whether the additional acquisition costs associated with (say) newer antidepressants or second-generation antipsychotics are compensated by better symptomatic response or fewer side effects. Our earlier examples of economic evaluations of treatments for depression and for child and adolescent mental health problems provide this kind of evidence.

Purchasers and providers

At another decision-making level, those who commission or purchase mental health services need economic data. A core element of local needs assessment and strategic service development by (say) a state health care system or a health maintenance organization concerns the resource implications of changes to, for instance, the hospital/community balance or investment in a new clinic or training of therapists.

Government and society

Economic evaluations should influence national-level policy and resource allocation decisions. Such evaluations have influenced policy with respect to the substitution of community-based for long-term hospital care, the development of 'assertive outreach' models, the expansion of early intervention initiatives for psychosis, and the overall level of funding. The CHOICE programme aims to provide this kind of evidence.

While adding economic analysis to mental health evaluations introduces an extra dimension that offers a wider assessment of the implications of new or existing courses of action, there can also be limitations. Many economic evaluations fall short of the ideal, whether in terms of sample size, comprehensiveness of cost measurement, outcome assessment, or evidence interpretation. Conclusions based on small-sample randomized trials can often only be tentative, while failure to measure the wider (non-health and non-service) costs associated with two or more treatments may produce misleading and partial results.

Even when it overcomes these limitations, an economic evaluation can never resolve difficult allocative and policy issues; rather, it is one additional tool that, together with evidence on the clinical and social dimensions, can facilitate explicit evidence-based decision-making.

Further information

Knapp, M.R.J., Funk, M., Curran, C., *et al.* (2006). Economic barriers to better mental health practice and policy. *Health Policy and Planning*, **21**, 157–70. http://heapol.oxfordjournals.org/cgi/content/full/21/3/157

WHO. (2006). *Dollars, DALYs and decisions: economic aspects of the mental health system*. WHO, Geneva, Switzerland. http://www.who.int/mental_health/evidence/dollars_dalys_and_decisions.pdf

Website of WHO's cost-effectiveness work programme (CHOICE): http://www.who.int/choice

References

1. WHO. (2001). *The World health report 2001; mental health: new understanding, new hope*. WHO, Geneva.
2. WHO. (2004). *Mental health policy, plans and programmes. Mental health policy and service guidance package*. WHO, Geneva.
3. WHO. (2000). *The world health report 2000; health systems; improving performance*. World Health Organization, Geneva.
4. Dixon, A., McDaid, D., Knapp, M., *et al.* (2006). Financing mental health services in low- and middle-income countries. *Health Policy and Planning*, **21,** 171–82.

5. Knapp, M.R.J., Funk, M., Curran, C., *et al.* (2006). Economic barriers to better mental health practice and policy. *Health Policy and Planning*, **21**, 157–70.
6. WHO. (2006). *Dollars, DALYs and decisions: economic aspects of the mental health system*. WHO, Geneva.
7. Scott, S., Knapp, M., Henderson, J., *et al.* (2001). Financial cost of social exclusion: follow-up study of antisocial children into adulthood. *British Medical Journal*, **323**, 191–4.
8. Beecham, J.K.J. and Knapp, M.R.J. (2000). Costing psychiatric interventions. In *Measuring mental health needs* (2nd edn) (eds. G. Thornicroft, C. Brewin, and J.K. Wing), pp. 200–24. Gaskell, London.
9. Curtis, J. and Netten, A. (2005). *Unit costs of health and social care*. Personal Social Services Research Unit, University of Kent, Canterbury.
10. McCrone, P., Knapp, M., Proudfoot, J., *et al.* (2004). Cost-effectiveness of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. *The British Journal of Psychiatry*, **185**, 55–62.
11. Chisholm, D., Healey, A., and Knapp, M.R.J. (1997). QALYs and mental health care. *Social Psychiatry and Psychiatric Epidemiology*, **32**, 68–75.
12. Olsen, J.A. and Smith, R.D. (2001). Theory versus practice: a review of 'willingness-to-pay in health and health care. *Health Economics*, **10**, 39–52.
13. Valenstein, M., Vijan, S., Zeber, J.E., *et al.* (2001). The cost-utility of screening for depression in primary care. *Annals of Internal Medicine*, **134**, 345–60.
14. Revicki, D., Brown, R., Keller, M., *et al.* (1997). Cost-effectiveness of newer antidepressants compared with tricyclic antidepressants in managed care settings. *The Journal of Clinical Psychiatry*, **58**, 47–58.
15. Kamlet, M.S., Wade, M., Kupfer, D.J., *et al.* (1992). Cost-utility analysis of maintenance treatment for recurrent depression: a theoretical framework and numerical illustration. In *Economics and mental health* (eds. R.G. Frank and W.G. Manning), pp. 267–91. Johns Hopkins University Press, Baltimore, MD.
16. Hatziandreu, E.J., Brown, R.E., Revicki, D.A., *et al.* (1994). Cost-utility of maintenance treatment of recurrent depression with sertraline versus episodic treatment with dothiepin. *Pharmacoeconomics*, **5**, 246–64.
17. Pyne, J.M., Smith, J., Fortney, J., *et al.* (2003). Cost-effectiveness of a primary care intervention for depressed females. *Journal of Affective Disorders*, **74**, 23–32.
18. Schoenbaum, M., Unutzer, J., Sherbourne, C., *et al.* (2001). Cost-effectiveness of practice-initiated quality improvement for depression; results of a randomized clinical trial. *The Journal of the American Medical Association*, **286**, 1325–30.
19. Knapp, M., Romeo, R., Mogg, A., *et al.* (in press). Cost-effectiveness of transcranial magnetic stimulation vs. electroconvulsive therapy for severe depression: a multi-centre randomised controlled trial. *Journal of Affective Disorders*.
20. Romeo, R., Byford, S., and Knapp, M. (2005). Annotation: economic evaluations of child and adolescent mental health interventions: a systematic review. *Journal of Child Psychology and Psychiatry*, **46**, 919–30.
21. Drummond, M.F. and Jefferson, T.O. (1996). Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ economic evaluation working party. *British Medical Journal*, **313**, 275–83.
22. Byford, S., Harrington, R., Torgerson, D., *et al.* (1999). Cost-effectiveness analysis of a home-based social work intervention for children and adolescents who have deliberately poisoned themselves. Results of a randomised controlled trial. *The British Journal of Psychiatry*, **174**, 56–62.
23. Edwards, R., Ceilleachair, A., Bywater, T., *et al.* (2007). Parenting programme for parents of children at risk of developing conduct disorder: cost effectiveness analysis. *British Medical Journal*, **334**, 682–7.
24. Tan Torres, T., Baltussen, R.M., Adam, T., *et al.* (2003). *Making choices in health: WHO guide to cost-effectiveness analysis*. WHO, Geneva.
25. Chisholm, D. (2005). Choosing cost-effective interventions in psychiatry. *World Psychiatry*, **4**, 37–44.
26. Hyman, S., Chisholm, D., Kessler, R., *et al.* (2006). Mental disorders. In *Disease control priorities in developing countries* (2nd edn) (eds. D. Jamison, J. Breman, A. Measham, *et al.*). Oxford University Press, New York.
27. Chisholm, D., Sanderson, K., Ayuso-Mateos, J.L., *et al.* (2004). Reducing the global burden of depression: a population-level analysis of intervention cost-effectiveness in 14 epidemiologically-defined sub-regions (WHO-CHOICE). *The British Journal of Psychiatry*, **184**, 393–403.
28. Commission on Macroeconomics and Health. (2001). *Macroeconomics and health: investing in health for economic development*. WHO, Geneva.

7.8

Psychiatry in primary care

David Goldberg, André Tylee, and Paul Walters

Epidemiology

In recent years, major epidemiological surveys have been carried out in the community in many different countries, in the United Kingdom most recently by the Office of National Statistics (to find this and other surveys go to http://www.statistics.gov.uk/STATBASE/Product.asp?vlnk=8258). The findings in such community surveys can be compared with findings in primary care surveys, when it will be found that the list of common mental disorders is not quite the same, although conditions characterized by symptoms of depression and anxiety are the most common. Rather than considering the detailed diagnoses, it can be helpful to distinguish between 'internalizing disorders', which besides anxiety states and depression, also include the fear disorders like phobias and panic disorder, obsessive-compulsive disorder and many cases of somatization disorder; and 'externalizing disorders' consisting of conduct disorder in childhood, and antisocial behaviour, as well as drug and alcohol disorders in adult life. The former group of disorders are characterized by subjective distress, and typically high levels of anxious and depressive symptoms; while in the latter group abnormalities are in externally observed behaviour.[1]

In community surveys it can be seen that rates of internalizing disorders rise sharply after puberty, are highest between the ages of 35 and 55 and fall thereafter, and that females rates are higher than males at all ages., while rates of externalizing disorders reach their maximum between the ages of 15 and 34, and fall sharply after that, with males rates much higher than female rates at all ages. This is shown in Table 7.8.1 (which does not include antisocial behaviour as reliable data are not available in the community).

The Goldberg-Huxley Model[2]

This was devised as a framework for comparing the characteristics of patients seen in the community with those in other medical settings, and describing the pathway which people usually follow to mental health care in places where GPs act as 'gatekeepers'. It consists of five levels, separated by four filters. The figures for psychiatric morbidity over 1 year necessitate using estimates of incidence rates, and are therefore much higher than the point prevalence rates reported in community surveys. The essence of the model is the demonstration that most distressed patients will see a doctor over the course of 1 year (filter 1), but only about half of them will have their distress detected (filter 2). Most common mental disorders are treated in primary care, so filter 3 is relatively impermeable, only allowing one in five to pass. Psychiatrists only have any part in the process with the fourth filter, which also holds back most patients. Psychiatrists therefore form their ideas about mental disorders from a highly skewed section of all those with disorders.

Prevalence of psychiatric disorder in primary care

In the United Kingdom, about 80 per cent of the population consult their doctor in the course of a year, and prevalences among attenders are higher than among the general population.[2] In contrast, specialist mental health services see between 1 and 2 per cent of the population in the course of a year, and admit only about 0.5 per cent to inpatient care, so that primary care deals with the major part of the burden of common mental disorders.

The World Health Organization (WHO) carried out the largest primary care survey in 14 countries[3] but for purposes of comparison only the UK data will be shown here. Table 7.8.2 compares the frequencies and types of mental disorders seen in the community, in primary care, and in psychiatric practice. Mental disorders

Table 7.8.1 Annual prevalence of mental disorders in the community by type and age, rates per 100 at risk

Disorder	Gender	5 to 16	to 34	to 54	to 74	All (16–75)
Internalizing	Male	3.1	11.70	16.75	9.87	13.5
	Female	4.3	20.55	21.35	14.80	19.4
Externalizing	Male	10.05	11.6; 18.9	2.25; 10.4	0.4; 3.8	6.0; 11.9
	Female	4.35	5.3; 5.7	0.75; 2.1	0.4; 0.5	2.3; 2.9
Other	Male	1.9	0.33	0.78	0.23	0.5
	Female	0.75	0.42	0.73	0.52	0.6

Internalizing = any neurotic disorder. *Externalizing* = conduct disorder for age 5–16; for the remaining age groups the rate for drug dependence is shown first, followed, after the semicolon, by the rate for alcohol dependence. *Other* = psychotic disorders in adults. Source: National Statistics website: www.statistics.gov.uk Crown copyright material is reproduced with the permission of the Controller Office of Public Sector Information (OPSI).

Table 7.8.2 Prevalence of mental disorder by gender for the community, for primary care attenders, and for admissions to psychiatric beds

	The community annual prevalence (%)		Primary care cases consecutive attenders (%)		Mental hospital inpatients (%)	
	Males	Females	Males	Females	Males	Females
Mixed anxiety depression	6.8	10.8	2.1	4.5	9.8	17.6
GAD	4.3	4.6	4.9	14.9		
Panic	0.7	0.7	3.4	3.6		
Phobias	1.3	2.2	2.1	4.6		
Neurasthenia	–	–	6.1	21.7		
Somatoform disorder	–	–	–	0.5		
OCD	0.9	1.3	–	–		
Depression	2.3	2.3	13.9	18.3	17.9	27.3
Alcohol dependence	11.9	2.9	5.3	0.8		
Drugs dependence	5.4	2.1	–	–	30.1	14.3
Schizophrenia	0.6	0.5	–	–	20.4	13.7
Organic, dementia	–	–	–	–	10.3	15.9
Subnormality	–	–	–	–	7.1	5.4
Developmental disorders	–	–	–	–	5.1	5.4
Any Dx	14.1%	19.9%	23.5%	27.5%	100%	100%

Sources: National Statistics website: www.statistics.gov.uk Crown copyright material is reproduced with the permission of the Controller Office of Public Sector Information (OPSI).

seen in primary care settings are more severe on average than those seen in community surveys, and different disorders predominate. The figures shown are for practices in Manchester with a fairly high prevalence of mental disorders, but the spread of diagnoses is fairly similar in other countries. The ICD-10 criteria only counted somatoform disorders if they were severe and long-standing, and do not count the many patients presenting with unexplained somatic symptoms, which are often accompanied by symptoms typical of anxiety or depression. A more recent study from Denmark[4] has estimated that almost a half of their patients were diagnosed cases of mental disorders, with somatoform disorders being found in about one-third. Patients with established physical illnesses are also at greater risk of mental disorders, and this is especially so if they are disabled by their illness. It can be seen from Table 7.8.3 that disorders admitted to psychiatric hospitals in the United Kingdom are different again from those typically seen in primary care, with organic states, drug and alcohol dependence, schizophrenia and severe depressive states accounting for the majority of cases (Source: http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=202).

A study in 10 European countries shows 28 per cent of consecutive attenders in the United Kingdom to be distressed on a screening interview, but only 6 per cent presented psychological symptoms to their GP. Most of these (5.5 per cent) received a psychiatric diagnosis, but the GPs also diagnosed others as 'psychiatric'—so that their total rate was 15 per cent.[5] These figures are fairly similar to those in Switzerland and the Netherlands, but in stark contrast to those in Eastern Europe. In the Russian Federation, for example, 27 per cent were distressed, but none reported psychological distress to their GPs, and none were diagnosed: however, the GPs identified 3 per cent of their patients as 'psychiatric'. Fairly similar figures were reported in Estonia, Poland, and Belgium; while figures in Germany, Spain, and Sweden are intermediate (*ibid* 2007; see Fig. 7.8.1). This study also showed that GPs who discuss psychosocial matters with their patients, and look at them are better at diagnosing them—a finding that echoes previous research in the United Kingdom.[2]

It can be seen from Fig. 7.8.1 that most distressed patients—who may well be found to have a mental disorder if interviewed with a research interview—do not mention their distress to their doctor, and that this accounts for failure to diagnose the disorder. Many patients who have not endorsed feelings of distress are nonetheless assessed as mental unwell, either because they are presenting with

Table 7.8.3 The Goldberg-Huxley Model: Data for Manchester, UK

Level 1: Community samples *First Filter: The decision to consult*	**250–315/1000/year**
Level 2: All those seeing GPs found to have a mental disorder *Second filter: GPs ability to detect*	**210–230/1000/year**
Level 3: Cases recognized by the GP As 'mental disorders' *Third filter: GPs decision to refer*	**101/1000/year**
Level 4: All those seeing mental health professionals *Fourth filter: Psychiatrist's decision to admit to hospital*	**20.6/1000/year**
Level 5: Those admitted as inpatients	**3.4/1000/year**

(Note that these estimates are annual period prevalences, and depend on estimates of annual incidence rates in addition to point prevalence rates. Source: Reproduced from D.P. Goldberg and P.J. Huxley, *Common mental disorders—a bio-social model*, copyright 1992, The Tavistock Institute, London.)

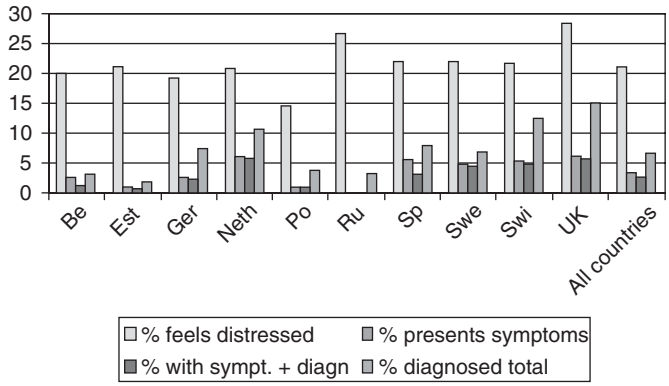


Fig. 7.8.1 Proportions of population that feel distressed, present symptoms of distress, are diagnosed if they do so, and total number of patients diagnosed, in 10 European countries (Reproduced from P. Verhaak, J. Bensing, and A. Brink-Nuinen, Primary mental health care in 10 European countries: patients' demands and GPs' responses. *European Journal of Psychiatry*, **21**, (1), Zaragoza Jan–Mar 2007, copyright 2007, INO REPRODUCCIONES, S.A.).

unexplained somatic symptoms or because they are under treatment for a mental disorder that has responded to treatment.

Clinical presentations

Somatization

Somatization is broadly defined as the expression of psychological distress through physical symptoms. In primary care, most patients present physical symptoms at the onset of an episode of anxiety and depression. This is actually the usual way that new episodes of common mental disorders present in general medical settings, as only about 15 per cent of new episodes present purely in psychological terms.[2]

Primary care somatizers can be subdivided into 'facultative somatizers', who admit to their psychological symptoms and accept a mental disorder diagnosis if appropriately interviewed, and 'pure somatizers' who, despite such an enquiry, still deny the presence of psychiatric symptoms. People have many reasons for preferring to present somatic symptoms to their doctors—they are understandably worried that they may have a new physical disease, they give priority to pains over other symptoms because it is pain that hurts, and they often wish to avoid the stigma of being thought psychologically distressed.

Somatizers may be compared with 'psychologisers', who directly present their psychological problems to their doctors. The former are more likely to report adverse childhood events and periods of childhood illnesses, while the latter have more abnormal attachment behaviours.

Hidden versus conspicuous morbidity

The ability of GPs to detect psychological disorders among their patients forms the second filter that patients must pass through in order to receive a diagnostic label. This ability varies a good deal between places as well as between disorders. In the WHO study[3] the overall average detection rate was 48.9 per cent, but rates varied from 75 per cent for Verona to 15.9 per cent for Shanghai (62.9 per cent for Manchester). Recognition rates for individual diagnoses followed these overall rates, with somatization disorder being best recognized, followed by depression. However, these detection rates do not reveal whether the GP is identifying the same patients as the research assessment. In fact, the exact agreement between the two (measured by κ) is rather poor, at only +0.18 for all centres (+0.38 for Manchester).

The second filter is passed when the GP recognizes a mental health problem in the patient, although this will often be without a precise ICD-10 diagnosis. Those recognized by the GP make up the '**conspicuous' morbidity**—in fact, just under half of that estimated to be present in the waiting room population by a two-stage case-finding procedure: so that the patients who are not identified can be thought of as the '**hidden morbidity'**. These undetected patients continue to consult, but an outsider's inspection of notes and prescriptions, or even discussion with the relevant doctor, will not identify them as patients with psychiatric morbidity. In practice, the 'conspicuous' morbidity may be greater than 50 per cent, as the longitudinal nature of primary care means that patients may be diagnosed in subsequent visits, and this is missed in cross-sectional waiting room studies.[6,7] Kessler *et al.* followed up a cohort of primary care patients over 3 years and found only 14 per cent of patients with depression remained unrecognized at the end of this period.[8] Rost *et al.* followed up 98 depressed patients who had made at least one visit to their GP and found 32 per cent were undetected at 1 year.[9] Despite this, GPs are good at recognizing severe depression, and unrecognized depression tends to be mild.[10–12] The severity of depression in primary care, rather than being defined categorically, may therefore be better conceptualized as running along a continuum from mild to severe. Using a dimensional approach Thompson *et al.* calculated GPs only miss one 'probable' case of depression every 29 consultations.[13]

Doctors better able to detect disorder have the following characteristics:[2]

- Make eye contact with the patient
- Make empathic comments
- Pick up verbal cues
- Pick up non-verbal cues
- Ask directive questions, with a psychological content
- Do not read notes, or look at their computer, while the patient is speaking
- Deal with over-talkativeness
- Deal with today's problem

Data from the WHO study indicate that these 'undetected illnesses' are on an average less severe than those detected by GPs and have a somewhat better outlook. However, the data does not support the view that failure to detect these less severe disorders has serious long-term consequences for the patient[14] although this does not mean that there are not individuals would be better served if their distress was acknowledged.

The elderly

Mental illness in the elderly is common in primary care. Between 5 and 10 per cent of older adults attending primary care will suffer from depression, though this may be higher in areas of socio-economic

deprivation.[15] Older people are less likely to admit to psychiatric problems and more likely to emphasize somatic concerns and present behavioural changes. One study found that only 38 per cent of those identified as depressed through screening in the community had discussed feelings of depression with their GP.[16] This may lead to under-recognition in primary care. Crawford *et al*. found that only 52 per cent of 62 patients identified with clinical depression from a community survey were correctly diagnosed by their GP.[17] The elderly may attribute their symptoms to 'normal ageing', grief, or physical illness, or may fear sigmatization more than younger patients making recognition more difficult. Elderly depressed men may be particularly likely to go unrecognized.[17]

A systematic approach using a collaborative care model may improve depression management for the elderly in primary care. The Improving Mood-Promoting Access to Collaborative Treatment (IMPACT) trial in the United States demonstrated that a primary care collaborative model for late-life depression was more effective than usual care in improving depressive symptoms. It also decreased pain due to osteoarthritis, increased functional abilities, and improved quality of life.[18] The collaborative care model was highly cost-effective[19] and continued to show benefits over a 2-year follow-up period.[20]

The other major condition with which the GP will be involved is dementia. In the United Kingdom, the prevalence rate is approximately 5 per cent for all those over 65, but there is an age-related rise within this band to 25 per cent for those aged over 85. Consultations for organic psychoses reflect this: 370 consultations per 10 000 years at risk for those aged over 75, and 888 per 10 000 years at risk for those aged over 85. However, the management of people with dementia in primary care has been criticized.[21] Up to 75 per cent of patients with moderate to severe dementia and up to 97 per cent of patients with mild cognitive impairment go unrecognized by their GP.[22] Again there may be a number of reasons for this. Dementias have an insidious onset, and sometimes the doctor's familiarity with the patient can militate against spotting change. If a relative also accepts that the changes associated with dementia reflect normal ageing, the diagnosis may be delayed or never made. Only 40 per cent of GPs in the United Kingdom use a specific test to detect dementia, and in a survey of 8051 GPs in England 40 per cent thought an early diagnosis of dementia was not important.[23] Turner *et al*. surveyed GPs knowledge, confidence, and attitudes about dementia and found that despite GPs overall knowledge about diagnosis and management being good, a third lacked confidence in their diagnostic skills and two-thirds lacked confidence in their management of behavioural problems and other associated problems in dementia.[24]

Downs and others conducted a trial of an educational package to improve detection and management of dementia in primary care.[25] In the United States, a trial of collaborative care versus care as usual for patients with dementia in primary care has produced encouraging results.[26] Compared with care as usual, collaborative care (consisting of case management though a senior practice nurse working with the patient's family and integrated in the primary care team, and the use of standard protocols to guide and monitor treatment) resulted in significant improvements in the quality of care and in the symptoms of dementia, without an increase in psychotropic medication use.

Classification of mental disorders in primary care

Difficulties with conventional psychiatric taxonomies

The main problem with the International Classification of Diseases, 10th Edition (ICD-10)[27] or Diagnostic and Statistical Manual, 4th Edition (DSM-IV)[28] classifications used by psychiatrists is that they were devised to describe a very different consulting population, they are needlessly complicated, and they do not lead directly to management. Patients usually present a mixture of physical, psychological, and social symptoms expressed in any order, although somatic symptoms are usually first. Some symptoms are repeatedly mentioned and some are mentioned only in passing. Symptoms left to the end may be the most important of all. Symptoms may not fit a psychiatrist's taxonomy. In primary care, patients often have several concurrent problems of a medical, psychological, and social nature.

Most psychologically distressed patients show symptoms of both anxiety and depression. The ICD-10 classification has a mild disorder called 'mixed anxiety depression', since some patients have symptoms of each which together seem sufficient for a diagnosis, although not satisfying the criterion for either disorder on its own. However, this does not solve the problem of the many patients who are above the threshold for both disorders, who are declared 'co-morbid' for two different disorders by conventional psychiatric taxonomy.

An alternative view points out that the two groups of symptoms are strongly correlated with one another (about +0.7) in the consulting population,[2] and thus views them as two related dimensions of symptomatology which tend to co-vary over time. GPs themselves rarely emphasize the distinction between the two groups of symptoms, and there appears to be no evidence that any adverse consequences follow this neglect. For the GP, the diagnostic task can be one of separating the symptoms of depression and anxiety from those of an accompanying physical illness, or of probing for psychiatric morbidity in patients where apparent physical symptoms do not have an organic cause. In primary care settings, 'co-morbidity' refers to patients who have both physical disorders and mental disorders.

Solutions to the classification problem

Both the major psychiatric classifications—ICD-10 and DSM-IV—offer special versions produced in collaboration with primary care physicians which are deemed suitable for this setting, and roughly correspond to the parent classification. The WHO offers '**ICD10-PHC**'[29] which consists of 26 common conditions, with advice on how they present, the diagnostic features, the differential diagnosis of similar symptoms, essential information for the patient and the carer, advice and support for the patient and the carer, the role of medication, and indications for referral to the mental health services. Information is given about national organizations, and self-help materials, to assist people with particular diagnoses, and advice is given to GPs in particular areas about customizing the system by including information about self-help and support groups for local people.

The '**DSM-IVPC**'[30] on the other hand is organized by symptoms that branch out into diagnostic algorithms. The GP assesses the

patient's symptoms and, in workbook fashion, determines the relevant psychiatric diagnoses. The manual is formatted to be concise and practical, with limited use of psychiatric jargon. As a key feature, the chapter devoted to 'Algorithms for Common Primary Care Presentations' presents nine algorithms, headed by the presenting symptoms, for the most common psychiatric concerns encountered in primary care. No advice is given on management or on information to be given to carers.

Given the quantity of printed material that floods into GPs offices, it is doubtful whether many GPs keep such systems on their desks—although they may be incorporated in their computer programs. It is more realistic to use such systems in training new doctors, so that ways of approaching patients and their carers become incorporated in their usual routines, with the system only consulted when an unusual problem presents itself. It can be seen that whereas the DSM-IV system is aimed at formal diagnosis, the ICD10-PHC system is aimed at management once an assessment has been completed.

The International Classification for Primary Care, (ICPC-2-R)[31]

The classification most widely used by GPs is of course their own, devised under the auspices of the World Organization of National Colleges & Academies ('WONCA'), called the International Classification for Primary Care, '**ICPC-2-R**'. This is a system which classifies all patient data and clinical activity in primary care, taking into account the frequency distribution of problems commonly encountered.

It allows classification of

- the patient's reason for encounter (RFE),
- the problems/diagnosis managed,
- interventions,
- test results, and the
- ordering of these data in an episode of care structure

It has a biaxial structure and consists of 17 chapters, each divided into seven components dealing with

- symptoms and complaints
- diagnostic, screening, and preventive procedures
- medication, treatment, and procedures
- test results
- administrative
- referrals and other reasons for encounter, and
- diseases

It is not clear to what extent all GPs work their way through this complex system, but should they reach the seventh component, there is a rough correspondence with the ICD. Note that multiple disorders can be coded in the same episode, but the extent to which this is done is not clear. Nor does the system provide advice on the management of the various conditions—it is assumed that the clinician knows how to do this.

The Read codes[32]

The most widely used system in the United Kingdom since the advent of computerization, are the Read codes. This meets the needs of the generalist by including diagnoses, symptoms, and problems. Some very broad reasons for consultation—such as 'anxiousness', 'depressed', and 'headache' are provided. A letter of the alphabet is followed by up to four numerical codes, and there are also about 50 codes for diagnoses such as E204. 'Depression' or E2003 'anxiety with depression', which correspond very roughly to ICD diagnoses. However, no criteria are given, nor any advice on management for these various diagnoses.

None of the above four systems take into account functional impairment and disability, yet clinicians need to consider this in conjunction with the set of symptoms presented by an individual patient. Both diagnosis and current impairment are essential, and may help to explain why up to a quarter of patients with schizophrenia are managed solely in primary care settings[33] whilst some patients with adjustment disorder need referral to the community mental health team.

Improving the identification of some common disorders

(a) Aids to accurate detection of depression

Rather than using routine screening questionnaires to all patients, it is more practical to ask the following questions in six groups of patients:

- all those who look or sound depressed, or mention depressive symptoms
- all those with a past history of depression
- all those with significant physical illness causing disability
- all those with diabetes and coronary heart disease, where the risk is higher
- all those with other mental health problems, such as dementia or heavy drinking
- mothers of infants who are either single or are unsupported by a partner or family

Both these questions should be asked:

- During the past month, have you been feeling down, depressed or hopeless?
- During the last month, have you often been bothered by having little interest or pleasure in doing things?

(b) Aids to accurate assessment of severity of depression

If positive replies are obtained to either of these, assess severity of using the Patient Health Questionnaire-9 (PHQ-9) a questionnaire with nine questions validated for use in primary care,[34] which may assist in ensuring that antidepressant medication is better targeted on those with moderate and severe degrees of depression.

(c) Aids to the accurate detection of alcohol problems

Other, more recent tools, such as the Drug Abuse Problem Assessment for Primary Care (DAPA-PC),[35] were designed specifically to be administered via computer. While currently popular in the areas of depression and substance abuse, the audience for assessment tools is expanding for a number of reasons, including: convenience, privacy, high-patient satisfaction,[36] decreased provider time, improved validity, and reliability,[37] and decreased expense.

The management of mental disorders within primary care

(a) Stepped care

With the introduction of mental health guidelines there has been a shift towards stepped care for mental health problems.[38] Stepped care 'provides a framework in which to organize the provision of services supporting both patients and carers, and healthcare professionals in identifying and accessing the most effective interventions' (see Fig. 7.8.2). Stepped care allows treatment to be provided in steps according to the severity of problems and/or response to treatment, aiming to provide the greatest benefit to most of the people from the resources available. Patients who fail to respond are given the next step on the treatment plan. Stepped care can be delivered by starting at the least invasive step and gradually 'climbing' the steps according to response, thus targeting more intensive treatments to those that that need them, or by 'stratifying' care so that the first treatment step is determined by severity of disorder and thereafter by response.[39] Stepped care was developed in the United States, initially for the management of depression.[40] In the United Kingdom, this model of care has now been recommended for depression,[38] anxiety disorders,[41] obsessive-compulsive disorder,[42] and self-harm.[43]

(b) Mental health workers based in primary care

Several new possibilities have emerged recently with the advent of the new *graduate mental health workers* in primary care. The NHS plan (DH 2000) was to have 1000 such workers in England and indications are that there are 600–700 in post. Mostly, the posts are similar to assistant psychologists and they generally are trained on a one-day release scheme at local university level in primary care mental health. Many are conducting initial assessment, providing brief psychological interventions such as brief cognitive behaviour therapy (CBT) or problem-solving for common mental health problems and medication management. Many are overseeing the use of self-help written materials and computerized help such as 'Beating the Blues', a computerized CBT programme for depression, which has been shown to be cost-effective.[44] Graduate primary care mental health workers (PCMHWs) also act as advisors on the range of other local services (for example, local library bibliotherapy schemes, and support groups). They have been proved to be effective at increasing patient satisfaction with episodes of care but were found not to improve mental health symptoms or to use the voluntary sector more than usual care.[45] Many PCMHWs are psychology graduates, and they tend to move on after a year or two into much wanted clinical psychology posts, so corporate memory can be diminished. Many other interventions described in level two of the National Institute for Clinical Excellence (NICE) Depression guideline[38] are increasingly applied in primary care—such as exercise schemes, befriending schemes, brief problem-solving, brief CBT, self-help materials, and sleep restoration.

Practice primary care counsellors may in reality be psychologists, psychotherapists, or counsellors depending on their training, expertise, and accreditation. They mostly see adults with common mental health disorders which may include adjustment disorders and losses. They often integrate different models of brief psychotherapeutic treatment depending on their training and the particular patient, and they receive regular supervision. These services may be provided in-house or in a neighbouring practice depending on

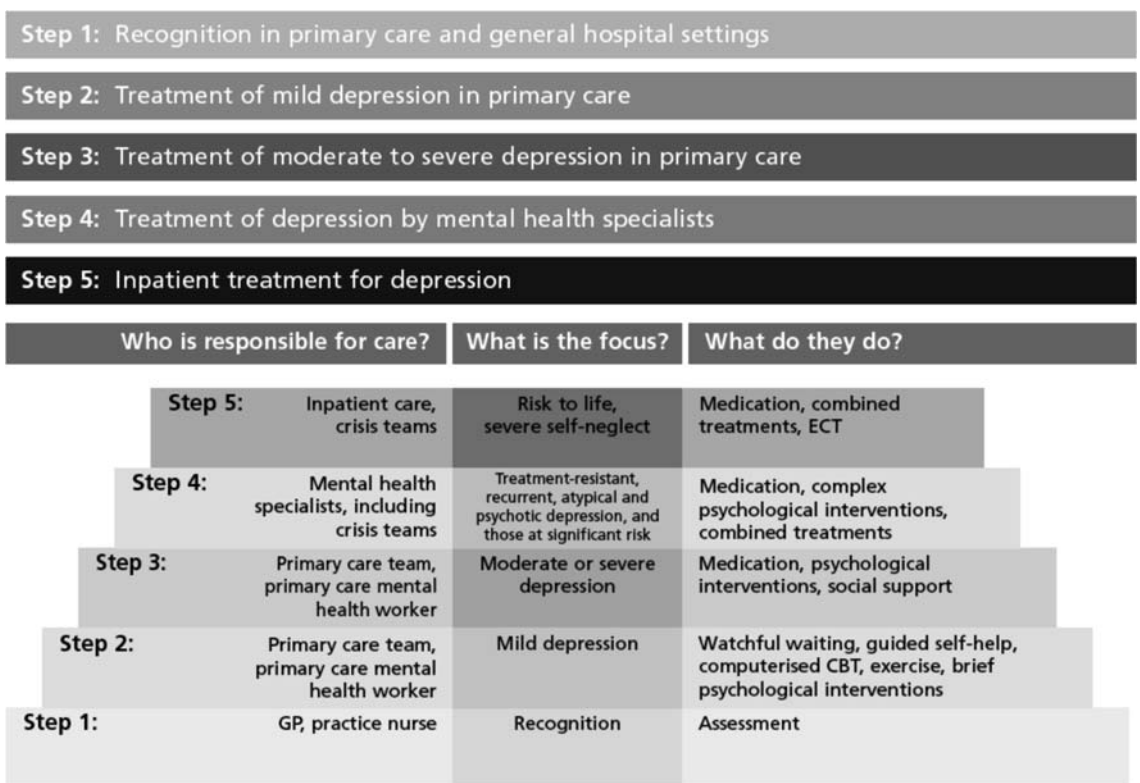


Fig. 7.8.2 The stepped care model applied to depression. Reproduced with permission.[38]

available space. Some counsellors specialize in providing services for non-English speaking groups and certain patient subgroups such as eating disorders, domestic violence, etc.). Whilst resources are usually limited on occasions they may see patients for longer courses of treatment.

(c) The role of the primary care nurse

Practice nurses are increasingly involved in chronic disease management within practices and this includes managing mental health conditions whether severe or common. From April 2006, practices have been rewarded in England under the Quality and Outcomes Framework of the General Services Medical Contract[46] for screening their patients with coronary heart disease (CHD) and diabetes for depression and with any newly depressed patients for using a recommended patient self-report instrument to assess baseline severity of depression. The recommended instruments are the Patient Health Questionnaire-9 item (PHQ9),[34] the Hospital Anxiety and Depression Scale (HADS),[47] or the Beck Depression Inventory (BDI).[48] Practice nurses are expected to undertake much of this work as they see the patients with CHD or diabetes for review and therefore they need proper training and supervision. Also rewarded in the GP contract is an annual physical review of all patients on the Severe and Enduring Mental Illness registers (SEMI or SMI)—again usually conducted by practice nurses. Such nurses have traditionally administered depot medication for such patients without much training on the procedure or the accompanying health checks. This training has been improving lately. Expecting an untrained practice nurse to administer depot without support and training is unacceptable, and training is needed in how to give the injection, how to review the suitability of this form of medication, how to monitor for side-effects, how to assess mental state for signs of deterioration, when to involve colleagues or refer to the community mental health team, and crisis management.

Practice nurses have been successfully trained in the assessment and management of depression[49,50] and the use of problem-solving in major depression.[51] Practice nurses have been successfully trained in chronic disease management of the severely mentally ill,[51] whereby they constructed a disease register, a care plan, and 3-monthly structured reviews encompassing psychological, physical, and social well-being. Whilst the practice nurses were diligent and detected problems, there was little evidence that they communicated these to the GPs. There was also a failure to increase health promotion in this group, despite their recognized increased standardized mortality.

Health visitors trained in counselling have been shown to benefit women with postnatal depression,[52] and Appleby *et al.*[53] have targeted the same group of patients with a nurse-led intervention.

Primary care mental health services

The primary care team provides the majority of care to people suffering from mental health problems. Though many services remain GP-led, many are becoming increasingly sophisticated with a growing number of professionals involved in the care of people with mental health problems. In England, the Department of Health has produced a guide for improving primary care mental health services.[39]

A number of models for primary care mental health services are described reflecting the need for services to provide care to populations with different needs, resource availability, and relationships with secondary care. Examples include the 'opt-in' or self-referral model, proactive models such as a collaborative care model, 'one-stop shop' models, and focused service models (i.e. services are developed for a specific remit such as common mental disorder or psychosis). The 'opt-in' model is one in which the patient initiates care and follow-up and is the traditional primary care model. In a collaborative care model there is active follow-up and collaboration between patient, primary care health professionals, and when required secondary care professional with the health professional coordinating care. A 'one-stop shop' model provides a single point of access to all mental health services with a triage system determining resource allocation. There is no perfect model, and all have advantages and disadvantages depending on the population needs and availability of resources.

As services develop so the range of primary care mental health professionals is expanding. As well as the traditional primary care professionals—the GP and practice nurse—many practices now have others, including primary care graduate mental health workers and health visitors. Their roles include making assessments, facilitating referrals and movement through the health care system, and delivering treatments. In the United Kingdom, the availability of counselling services is now commonplace in primary care and up to 50 per cent of general practices have on-site counselors.[54] They are referred people with a wide range of common mental disorders and other psychosocial problems such as adjustment disorders. Bower and Roland reviewed the evidence for the effectiveness of counselling in primary care and found it was associated with a modest improvement in short-term outcomes compared to usual care, but was no better than usual care over the longer-term. It did not appear to be any more cost-effective than usual care but patients appeared satisfied with it.[55]

The primary–secondary care interface

What GPs expect of psychiatrists

GPs expect psychiatrists to possess and exhibit specialized skills of assessment and management not possessed within the primary health care team. The primary care team may include primary care mental health workers ranging from graduate workers to counsellors, psychotherapists, or psychologists, although if present these are usually extremely thinly spread (i.e. 1–2 sessions per practice per week). GPs do expect their psychiatrists or a named person from the mental health team to be available when they are needed, and they should also make themselves available for incoming telephone calls from the key contact. Because of sheer numbers (in England and Wales there are 12 times as many GPs as psychiatrists) the GP must protect this valuable resource by not overloading it with inappropriate referrals and by obtaining and maintaining certain assessment and management skills that can be used in primary care as well as sharing the care of certain patients under the leadership of secondary care. Referrals will vary widely from practices depending on their own 'in-house' expertise and the presence or absence of any primary care mental health workers or local counselling services. Larger practices often have more in-house expertise and are more likely to have at least one of their GPs who possess more specialized mental health skills. Small practices with one or two partners are less likely to have such skills and may need

more support. GPs expect the psychiatrist to provide inpatient care when needed (e.g. serious self-neglect, suicide intent, etc.) and day-patient facilities to provide a place of care, respite, and safety. They can also expect the psychiatrist to use diagnostic facilities and investigations (e.g. scans) as necessary and to provide highly specialized treatments when indicated (e.g. electroconvulsive therapy). Less frequently, the GP may need respite from particular doctor–patient relationships for the longer-term good of both. Referral is also sometimes a result of pressure by the relatives or patient. Where good communication exists between primary and secondary care, these often 'covert' reasons for referral can be openly discussed.

Access to specialist assessment when appropriate is paramount for a primary care service. GPs are not usually trained in specialist assessment and therefore, to match need to services, a psychiatrist, community psychiatric nurse, or psychologist from the community mental health team can perform this function, often in a primary care setting or the patient's home. Other community mental health teams operate an outpatient clinic (which may be moved into the surgery). Often 'true consultancy' is being sought by the GP, whereby he or she may receive advice only. Other practices operate a joint consultation system whereby the specialist and generalist see the patient together and formulate a plan. With recent changes in the configuration of mental health teams in the United Kingdom to provide acute care, early intervention and assertive outreach, in some places the more traditional community mental health teams are being reduced which is often confusing and can be unsettling for practices which have developed relationships with individual psychiatrists and colleagues. Also, if crisis teams work 9 to 5 p.m., yet the GP surgeries are open until 6 or 7 p.m., this can create difficulties for making urgent referrals.

Ways of organizing the interface

The interface between primary and secondary care is of key importance in the delivery of mental health care. Bower and Gilbody[56] have described a continuum of specialist involvement in primary care mental health with least involvement of secondary care professionals in the education/training model, and increasing involvement through the consultation-liaison model, the collaborative care and the replacement/referral model. In practice, these models are not mutually exclusive. They can complement one another and be adapted to take into account local workforce issues and staff availability.

There are at least four models of working across the interface between primary and secondary care:

1 A **replacement/referral model** is the traditional way in which primary care interfaces with secondary care. In this model, the patient's care is handed over to specialist services by way of a referral, the specialist service only relinquishing care when the patient had been treated. However, over the last 25 years other models have developed.[57] These include:

2 The **consultation-liaison model** allows secondary care professionals to develop ongoing relationships with primary care professionals, not only providing expert advice but also actively liaising with the primary care team, and often attending team meetings or seeing patients jointly with the GP.

3 The **collaborative care model**, as described above has primary care mental health professionals working between primary and secondary care to improve the overall care of patients. These link-workers can provide active follow-up and access to specialist advice and care as needed.[59]

4 In the **training/education model**, secondary care provides education and training to the primary care team which otherwise functions autonomously utilizing secondary care services when needed via a referral system.

How health services negotiate the interface between primary and secondary mental health services is likely to get increasingly complex as secondary care moves away from centralized care through the community mental health teams, to specialist teams such as assessment and brief treatment teams, early intervention teams, and continuing care teams. It is likely that each of these teams will develop its own model of interfacing with primary care. This may allow for a more fluid interface and closer working relationships between primary care and secondary care teams.[59] However, it needs to be managed appropriately or could lead to confusion of roles and responsibilities.

Shared care registers and shared care plans

A **shared care register** is usually a computerized record of all patients jointly cared for by the two services. It might consist of all those who have been discharged from hospital in the past 2 years, all those who have been on a psychotropic drugs for longer than a year, and all psychotic patients known to the GP who have not had an admission to the hospital. The record gives information about the key worker, outpatient clinics are held in the surgeries, and 'good practice protocols' can be developed, so that the case register can be audited against what other teams agree is good clinical practice.

Shared care plans follow on from this development. Such a plan gives the primary care staff information about symptoms which they may expect while the patient is well, likely symptoms in relapse, the name of the key worker, and full details of whom to contact in an emergency both during the day and at night. The plan makes clear who is responsible for medication, and gives an acceptable alternative should the GP find it necessary to vary the medication. It is essential that these plans are mutually agreed between the two teams, rather than being imposed by one team on the other. GPs in England are now being remunerated for keeping registers of patients with psychotic illnesses, dementia, and learning disabilities.

Improving the mental health skills of GPs

About half of GP trainees have a 6-month psychiatry hospital attachment, many of which are considered to be unhelpful for a future generalist career.[59] Many GPs have had no higher professional training in mental health and are not required to do so. There are several GPs who may have previously trained in psychiatry or psychotherapy (e.g. cognitive analytic psychotherapy, CBT, family therapy, etc.) before entering general practice and there is a growing national network of GPs with a special interest in psychiatry (GPsis) with the development of a national course for GPsis to Diploma/Masters level organized by PRIMHE, the National Primary Care Mental Health charity for professionals (www.primhe.org). The GPs network covers most regions and greatly overlaps with the National Trailblazer network (www.iop.kcl.ac.uk). Trailblazers was developed 10 years ago by one of the authors (AT) to bring together professionals from primary and secondary care to work together in pairs to build bridges and

enhance local services by working on a local service development project together with supervision. Trailblazer courses bring pairs together for tutor and peer supervision of projects for up to a year. To date nearly 1000 GPs, psychiatrists, CPNs, PNs, etc., have participated in this way and trailblazer training centres now run in every region in England, supported by the Regional Care Services Improvement Partnership Development Centres (www.csip.org.uk/regions) of the UK Department of Health. International trailblazers is now part of the International Initiative for Mental Health Leaders (www.IIMHL.org) and runs in New Zealand, United States, and England where there is the added interest of comparing service systems by participants who have exchange visits as modules rotate in the three countries. Trailblazers has been positively evaluated for the adult-centred learning approach.[60] Another well-recognized Quality Improvement Programme with a long history of working in primary care has recently focused its methods on common mental health disorders. The Improvement Foundation (www.improvement foundation.org.uk) formerly known as the National Primary Care Development Team (www.npdt.org.uk) have been working with 20 PCTs in England using a Plan Do Study Act cycle (PDSA) to help practices improve their depression care.

As there are few opportunities for primary care workers to obtain mental health skills training, one successful distance learning method involves the use of training DVDs. 'Micro-skills' of assessment or treatment of mental health disorders can be demonstrated by real-life general practitioners with actor–patients in 10 min consultations. The learner is then encouraged to practice these skills using role-plays supplied with the DVDs. A series of existing training materials have been put together for the World Psychiatric Association (WPA) which involves two of the authors (DG and AT) and colleagues from the Institute of Psychiatry, and Prof. Linda Gask and her colleagues at Manchester University. The materials in this WPA package cover depression, somatization, chronic fatigue, schizophrenia, anxiety, and dementia (see www.iop.kcl.ac.uk or www.man.ac.uk for further details).

Summary

At one time, it was asserted that the 'worried well' were treated in primary care, while true mental illnesses were seen by the mental illness services. This was not true when it was asserted, and is even less true now. The great majority of patients with common mental disorders are cared for within primary care, and many of those with severe mental illnesses are only seen in primary care. 'Stepped care' is a model for distributing clinical problems between the services, and 'shared care' refers to the care of patients seen by both primary care and specialist mental health services. Many other workers in primary care now assist GPs with the treatment of mental disorders, and special administrative arrangements within primary care are necessary to ensure that clinical services are available to those with special needs.

In summary, mental disorders in primary care:

- Are an important public health problem
- Frequently present with somatic symptoms
- Are more likely to be detected if the doctor has better communication skills
- Those with disabling physical illnesses are also at greater risk
- Are on average less severe than those seen in specialist care

Further information

Care Services Improvement Partnership (CSIP). Improving Primary care Mental Health Services. National Institute for Mental Health England, Department of Health. 2006. (www.csip.org.uk/resources/publications/primary-care.html) Accessed 22/3/07. A practical guide to developing and improving mental health services in primary care.

http://guidance.nice.org.uk/topic/behavioural. National Institute for Health and Clinical Excellence's website for guidance on mental and behavioural disorders.

Jenkins, R. (ed.) (2004). *WHO guide to mental and neurological health in primary care. A practical guide to the assessment and treatment of mental disorders in primary care* (2nd edn). Royal Society of Medicine Press, London.

References

1. Krueger, R.F. (1999). The structure of common mental disorders. *Archives of General Psychiatry*, **56**, 921–6.
2. Goldberg, D.P. and Huxley, P.J. (1992). *Common mental disorders—a bio-social model*. Tavistock, London.
3. Ustun, B. and Sartorius, N. (1995). *Mental disorders in general health care: an international study*. John Wiley, Chichester.
4. Toft, T., Fink, P., Oernboek, E., *et al.* (2005). Mental disorder in primary care: prevalence and co-morbidity among disorders. *Psychological Medicine*, **35**, 1173–84.
5. Verhaak, P., Bensing, J., and Brink-Muinen, A. (in press). Primary mental health care in 10 European countries: patients' demands and GPs' responses. *European Journal of Psychiatry*.
6. Ormel, J. and Tiemens, B. (1995). Recognition and treatment of mental illness in primary care. Towards a better understanding of a multifaceted problem. *General Hospital Psychiatry*, **17**, 160–4.
7. Freeling, P. and Tylee, A. (1992). Depression in general practice. In *Handbook of affective disorders* (ed. E.S. Paykel), pp. 651–6. Churchill Livingstone, Edinburgh.
8. Kessler, D., Bennewith, O., Lewis, G., *et al.* (2002). Detection of depression and anxiety in primary care: follow up study. *British Medical Journal*, **325**, 1016–7.
9. Rost, K., Zhang, M., Fortney, J., *et al.* (1998). Persistently poor outcomes of undetected major depression in primary care. *General Hospital Psychiatry*, **20**, 12–20.
10. Simon, G.E., Goldberg, D., Tiemens, B.G., *et al.* (1999). Outcomes of recognized and unrecognized depression in an international primary care study. *General Hospital Psychiatry*, **21**, 97–105.
11. Dowrick, C.F. (1995). Case or continuum? Analysing GPs ability to detect depression in primary care. *Primary Care Psychiatry*, **1**, 255–7.
12. Wittchen, H.U., Hofler, M., and Meister, W. (2001). Prevalence and recognition of depressive syndromes in German primary care settings: poorly recognized and treated? *International Clinical Psychopharmacology*, **16**, 121–35.
13. Thompson, C., Ostler, K., Peveler, R.C., *et al.* (2001). Dimensional perspective on the recognition of depressive symptoms in primary care: the Hampshire Depression Project 3. *The British Journal of Psychiatry*, **179**, 317–23.
14. Goldberg, D.P., Privett, M., Üstün, B., *et al.* (1998). The effects of detection and treatment on the outcome of major depression in primary care: a naturalistic study in 15 cities. *The British Journal of General Practice*, **48**, 1840–4.
15. Unutzer, J. (2002). Diagnosis and treatment of older adults with depression in primary care. *Biological Psychiatry*, **52**, 285–92.
16. Blanchard, M., Waterreus, A., and Mann, A. (1994). The nature of depression among older people in inner London and the contact with primary care. *The British Journal of Psychiatry*, **164**, 396–402.
17. Crawford, M.J., Prince, M., Menezes, P., *et al.* (1998). The recognition and treatment of depression in older people in primary care. *International Journal of Geriatric Psychiatry*, **13**, 172–6.

18. Unutzer, J., Katon, W., Callahan, C.M., *et al.* (2002). Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *The Journal of the American Medical Association*, **288**, 2836–45.
19. Katon, W.J., Schoenbaum, M., Fan, M.Y., *et al.* (2005). Cost-effectiveness of improving primary care treatment of late-life depression. *Archives of General Psychiatry*, **62**, 1313–20.
20. Hunkeler, E.M., Katon, W., Tang, L., *et al.* (2006). Long term outcomes from the IMPACT randomised trial for depressed elderly patients in primary care. *British Medical Journal*, **332**, 259–63.
21. Audit Commission. (2002). *Forget me not: developing mental health services for older people in England*. Audit Commission, London.
22. Gifford, D.R. and Cummings, J.L. (1999). Evaluating dementia screening tests: methodologic standards to rate their performance. *Neurology*, **52**, 224–7.
23. Woods, R.T., Moniz-Cook, E., Iliffe, S., *et al.* (2003). Dementia: issues in early recognition and intervention in primary care. *Journal of the Royal Society of Medicine*, **96**, 320–4.
24. Turner, S., Iliffe, S., Downs, M., *et al.* (2004). General practitioners' knowledge, confidence and attitudes in the diagnosis and management of dementia. *Age and Ageing*, **33**, 461–7.
25. Downs, M., Turner, S., Bryans, M., *et al.* (2006). Effectiveness of educational interventions in improving detection and management of dementia in primary care. *British Medical Journal*, **332**, 692.
26. Callahan, C.M., Boustani, M.A., Unverzagt, F.W., *et al.* (2006). Effectiveness of collaborative care for older adults with Alzheimer disease in primary care: a randomized controlled trial. *The Journal of the American Medical Association*, **295**, 2148–57.
27. ICD-10. (1993). *The ICD-10 classification of mental & behavioural disorders*. WHO, Geneva.
28. American Psychiatric Association. (1994). *Diagnostic & statistical manual of mental disorders* (4th edn). American Psychiatric Association, Washington.
29. WHO. (2001). *Guide to mental health in primary care*. Royal Society of Medicine Press, London.
30. Pingitore, D. and Sansone, R.A. (1998). *Using DSM-IV primary care version: a guide to psychiatric diagnosis in primary care*. *American Family Physician* **58**, 1347–52. Online at: http://www.aafp.org/afp/981015ap/pingitor.html.
31. WONCA. (2005). *International classification committee ICPC--R: International Classification of Primary Care*. OUP, Oxford.
32. The Clinical Terms Version 3 (The Read Codes) NHS Information authority. (2000). Available at http:\\www.coding.nhsia.nhs.uk.
33. Melzer, D., Hale, A., Malik, S., *et al.* (1991). Community care for patients with schizophrenia one year after hospital discharge. *British Medical Journal*, **303**, 1023–6.
34. Kroenke, K., Spitzer, R.L., and Williams, J.B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, **16**, 606–13.
35. Holtz, K., Landis, R.D., Nemes, S., *et al.* (2001). DAPA-PC: development of a computerized screening system to identify substance abuse in primary care. *Journal of Health Quality*, **23**, 34–7.
36. Kobak, K.A., Taylor, L.H., Dottl, S.L., *et al.* (1997). A computer-administered telephone interview to identify mental disorders. *The Journal of the American Medical Association*, **278**, 905–10.
37. Greist, J.H. (1998). Clinical computing: the computer as clinician assistant: assessment made simple. *Psychiatric Services*, **49**, 467–72.
38. National Institute for Health and Clinical Excellence (NICE) (2007) CG23 Depression (amended). *Management of depression in primary and secondary care*. NICE, London. Available from www.nice.org.uk/CG23.
39. Raistrick, H. and Richard, D. (2006). *Designing primary care mental health services: guidebook*. Care Services Improvement Partnership, Department of Health, Leeds.
40. Katon, W., Von Korff, M., Lin, E., *et al.* (1999). Stepped collaborative care for primary care patients with persistent symptoms of depression: a randomized trial. *Archives of General Psychiatry*, **56**, 1109–15.
41. National Institute for Clinical Excellence. (2004). *Anxiety: management of anxiety (panic disorder, with or without agoraphobia, and generalised anxiety disorder) in adults in primary, secondary and community care*. Clinical Guideline 22. National Institute for Clinical Excellence, London.
42. National Institute for Clinical Excellence. (2006). *Obsessive-compulsive disorder: core interventions in the treatment of obsessive-compulsive disorder and body dysmorphic disorder*. National Clinical Practice Guideline Number 31. National Institute for Clinical Excellence, London.
43. National Institute for Clinical Excellence. (2004). *Self harm: short-term treatment and management*. National Clinical Practice Guideline Number 16. National Institute for Clinical Excellence, London.
44. McCrone, P., Knapp, M., Proudfoot, J., *et al.* (2004). Cost-effectiveness of computerized cognitive behaviour therapy for anxiety and depression in primary care randomised controlled trial. *The British Journal of Psychiatry*, **185**, 55–62.
45. Lester, H., Freemantle, N., Wilson, S., *et al.* (2007). Cluster randomised controlled trial of the effectiveness of primary care mental health Workers. *The British Journal of General Practice*, **57**, 196–203.
46. Quality and Outcomes Framework Information available at http://www.ic.nhs.uk/services/qof.
47. Zigmond, A.S. and Snaith, R.P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, **67**, 361–70.
48. Beck, A.T., Mendelson, M., and Mock, J. (1961). Inventory for measuring depression. *Archives of General Psychiatry*, **4**, 561–71.
49. Wilkinson, G. (1992). The role of the practice nurse in the management of depression. *International Review of Psychiatry*, **4**, 311–6.
50. Mann, A.H., Blizard, R., Murray, J., *et al.* (1998). An evaluation of practice nurses working with general practitioners to treat people with depression. *The British Journal of General Practice*, **48**, 875–9.
51. Mynors-Wallis, L.M., Gath, D.H., Lloyd-Thomas, A.R., *et al.* (1995). Randomised controlled trial comparing problem-solving treatment with amitriptyline and placebo for major depression in primary care. *British Medical Journal*, **310**, 441–5.
52. Holden, J.M., Sagovsky, R., and Cox, J.L. (1989). Counselling in a general practice setting: controlled study of health visitor intervention in treatment of postnatal depression. *British Medical Journal*, **298**, 223–6.
53. Appleby, L., Warner, R., Whitton, A., *et al.* (1997). A controlled study of fluoxetine and cognitive-behavioural counselling in the treatment of postnatal depression. *British Medical Journal*, **134**, 932–6.
54. Mellor-Clark, J., Simms-Ellis, R., and Burton, M. (2001). *National survey of counsellors in primary care: evidence for growing professionalisation*. Royal College of General Practitioners, London.
55. Bower, P. and Rowland, N. (2006). Effectiveness and cost effectiveness of counselling in primary care.[update of Cochrane Database of Systematic Reviews, 2002; (1): CD001025; PMID: 11869583]. *Cochrane Database of Systematic Reviews*, **3**, CD001025.
56. Bower, P. and Gilbody, S. (2005). Managing common mental health disorders in primary care: conceptual models and evidence base. *British Medical Journal*, **330**, 839–42.
57. Gask, L., Sibbald, B., and Creed, F. (1997). Evaluating models of working at the interface between mental health services and primary care. *The British Journal of Psychiatry*, **170**, 6–11.
58. Von Korff, M. and Goldberg D. (2001). Improving outcomes in depression: the whole process of care needs to be enhanced. *British Medical Journal*, **323**, 948–9.
59. Turton, P., Tylee, A., and Kerry, S. (1995). Mental health training needs in general practice. *Primary Care Psychiatry*, **1**, 197–9.
60. Brown, C., Wakefield, S., Bullock, A., *et al.* (2003). A qualitative evaluation of the Trailblazers teaching the teachers programme in mental health. *Learning in Health and Social care*, **2**, 74–82.

7.9

The role of the voluntary sector

Vanessa Pinfold and Mary Teasdale

What is the mental health voluntary sector?

The voluntary sector plays an important role in the mental health field across the world. Originally set up and run by volunteers, the prime motivation and purpose for this sector is improving the lives of people affected by mental health problems by doing things differently, tirelessly pushing for change, never giving up hope and working alongside service users (also known as patients and consumers) and their families every step of the way. In some countries it has been labelled the 'third sector' to distinguish it from other organizational sectors namely industry (private sector) and government (public or statutory sector).

The voluntary sector is not, however, a cohesive group of organizations and across the mental health community each one operates with its own specific remit. Some of these organizations have become large businesses providing a wide range of services under contract with statutory agencies. Others choose to avoid employing staff and are still run entirely by committed volunteers. Many rely on voluntary donations in order to remain fiercely independent of government. Each has its own aims and mission, core stakeholder group, trustee and membership structures, management systems, governance procedures, and a unique portfolio of activities.

Some mental health organizations focus activities on mental health or emotional well-being specifically (e.g. Finnish Association for Mental Health). Others target a social problem and support all those affected such as charities working with the homeless, refugees, victims of domestic violence, or young offenders, including people with mental health problems. There are organizations that primarily campaign, educate, advocate, lobby, and promote self-help resources such as EUFAMI—an association for families across Europe and SANE Australia. In parts of the world, including Eastern Europe, there are particular challenges in mental health resulting from poverty, dislocation of the population, and insufficient resources for health. Some states, like Armenia, have no mental health services. Here, the Catholic Agency for Overseas Development (CAFOD) is working with Armenia's Association of Child Psychiatrists and Psychologists to provide therapy and mental health care and to increase awareness and understanding in order to overcome prejudice.

Wherever they work, key characteristics of voluntary sector organizations include an independent position, a strong values base, empowerment principles, non-profit distributing of resources, passionate commitment to the work focus, rooted to service user, and carer experiences and they are always striving for changes to provide people with mental health problems better provision and opportunities. Most started as local support groups but some have grown into large national organizations with considerable political leverage. The National Schizophrenia Fellowship (known today as Rethink) was founded in 1972 by a group of families concerned that relatives of people with schizophrenia had no support for themselves. In this chapter we draw on the Rethink experience to illustrate how the voluntary sector contributes to, and shapes, modern psychiatry. Although the English experience does not directly map onto those in other countries across the world, there are similarities and we seek to highlight these through the use of international examples where possible.

Rethink severe mental illness

Rethink is a membership charity with 7500 members (service users, carers, mental health professionals, the general public) whose mission is 'to support everyone affected by severe mental illness recover a better quality of life'. It adopts a recovery-orientated approach to supporting the individual and their family through periods of ill health and their journey of recovery. This perspective is significantly different from that of clinicians and statutory providers, who have not been through the experiences common to people who suddenly have to cope with severe mental illness. Rethink staff descriptions of their role include:

> *Bridging: linking a person with non-judgemental delivery of services connected with service user and carer experience*
>
> *Ensuring service users are heard and needs met more holistically*

Initially providing mutual support, the organization later offered information resources which address the problems commonly encountered by service users and carers, like difficulty in gaining access to services or funding for appropriate care. The emphasis is on finding successful strategies which achieve solutions. Advocacy is provided for individuals and families whose needs are not being met. The experiences of service users struggling to cope provide detailed evidence which is used to develop Rethink's policy on the mental health issues which reach the political arena and also as the basis of campaigns on stigma and discrimination. Research and

surveys of service user and carer views form the basis of reports on vital issues, like how information can be provided to carers with due respect for the service user's privacy and autonomy. Guidance on good practice may be developed and sometimes training for professionals. Media activity has publicized Rethink's campaigns and using the internet has made dissemination of information cheaper and easier than it used to be.

Rethink's activities also include the provision of front-line services in partnership with the National Health Service and Social Services (statutory sector) for example supported housing schemes, advocacy projects, community resource centres, carer support services, employment and training programmes, school education projects, and mentoring programmes for young people. In 2007, Rethink ran 350 front-line mental health services and employed approximately 1300 staff.

Does the voluntary sector make a difference?

The voluntary sector makes a substantial contribution to both the image of psychiatry and its practices. For example in New Zealand, the Mental Health Foundation has worked in partnership with the Ministry of Health and other agencies to run a successful anti-discrimination campaign—like minds, like mine (whakaitia te whakawhiu i te tangata) for the past 10 years. This is an internally renowned mental health awareness programme that is transforming how the New Zealand public engages with mental health issues. Non-Governmental Organizations (NGOs) can also bring influence to bear at a national level and bring people together to plan new service models. The World Fellowship for Schizophrenia and Applied Disorders (WFSAD) supported a workshop in East Africa in 2003 where service users and families could meet with government ministers and medical professionals to discuss plans for achieving effective health care delivery. In England, the voluntary sector has collectively ensured that the published clinical guidelines describing best practice for the treatment of schizophrenia in 2002 took note of service user and carer treatment preferences. In India, Action for Mental Illness (ACMI), an advocacy initiative has achieved tax concessions for those with mental illness and their carers and also maintenance allowances equivalent to those provided to people with physical disabilities.

Standards of mental health care delivery vary dramatically—with cases of human rights abuses in psychiatry being documented in some countries and innovative services emerging in others in response to local demands. The voluntary sector can, and does, bring the spotlight on both ends of the service delivery spectrum (good and bad) and demand better for everyone. It also leads the way by developing innovative solutions both in terms of service delivery, public education, and self-management techniques. In the United States, NAMIs Peer-to-Peer Education Course is a 9-week experiential education course on the topic of recovery for anyone with serious mental illness who is interested in establishing and maintaining wellness. In Canada, a family to family network was established for first episode psychosis families by the Canadian Mental Health Association. The recovery model is being pioneered and embraced by the voluntary sector across the world. However, this relatively new approach requires a change of attitude by both service users and professionals as shown by the Scottish recovery network programme.

Common themes tend to emerge through the campaigns of voluntary organizations across the world, in spite of differences in wealth and varying stages of development. In most places the demands are for earlier intervention, better crisis response, more support for families, and less use of physical restraint on the ward when coping with challenging behaviour. The transition to care in the community presented a new set of problems, not least the need for different professional skills and adequate resources. And in the later part of the last century, the controversy over community treatment orders was raging in many countries. Changes in government policy and concerns over access to newer treatments resulted in new alliances in many countries, and the voluntary sector, clinicians and other professionals often learned to work together, recognizing how in alliance they could lobby more effectively for better law and more resources.

An excellent example of successful partnership working in England and Wales has been the formation of a coalition known as the Mental Health Alliance involving 80 organizations. Professional bodies have joined, including the Royal College of Psychiatrists as prominent and active members. The Mental Health Alliance has opposed the government's proposals for reforming the Mental Health Act 1983 for the past 8 years and has managed to achieve some change in the content of the legislation as well as delaying the whole process.

The role of critical friends

The independence of organizations is an essential characteristic of the voluntary sector. These bodies do occupy the territory of 'critical friends' to both the statutory and private sectors of the mental health community, monitoring activity and speaking out in praise of positive developments but also highlighting when things are wrong.

For example at Rethink experience shows that misdiagnosis can result in inappropriate care and treatment with tragic consequences like imprisonment, suicide, or even homicide. Therefore Rethink supports service users in obtaining expert second opinions. Providing families or individual service users with accurate but understandable information enables them to challenge the opinion of a psychiatrist or medical team if this seems appropriate.

Rethink also advocates for families by using the complaints procedures or by providing legal representation at inquest hearings in order to draw attention to deficiencies in support, care, and treatment. They focus efforts on cases where systemic problems played a part, like a poor approach to risk assessment or refusal to accept information from families. The aim is to persuade the Ombudsmen or coroners to recommend improvements in the local policies and procedures in order to improve the quality of services.

In recent times there has been a spotlight on mental health services' engagement with people from Black and Minority Ethnic groups (BME) in England. The BME voluntary mental health organizations have formed a network which aims to reduce inequalities and promote good practice in mental health for racialized groups. The Network has been very critical of the Government's proposals to amend mental health legislation and has criticized the statutory health sector for failing to meet legal requirements on race equality. Similar issues arise in the United States where the National Council of La Raza (NCLR)—the largest national Hispanic civil rights and advocacy organization in the United States—works to improve

opportunities for Hispanic Americans. They report that Latinos are at a disproportionately high risk for depression and other conditions associated with mental illness, and are also much less likely to seek treatment or receive quality culturally and linguistically competent care.

Conclusion

The voluntary sector is a dynamic and vital part of any mental health system. Rooted in the experiences of mental health service users and carers, voluntary sector organizations across the world ensure that the voices of 'experts by experience' directly influence campaigns, policy debates, service redesign, and project planning and treatment guidelines. The sector is, however, fragile and in some countries organizations are increasingly dependent on state funding which could undermine their autonomy and independence. Psychiatrists can support their local voluntary organizations by joining them—as members, as campaigners, and as educators. The sector can also support psychiatrists, helping to transform the public image of psychiatry and encouraging young people to take an interest in mental health as a career option. The alliances forged with psychiatrists and their representative bodies are crucial for improving the quality of mental health services and to effectively tackle stigma and discrimination. We do need each other in order to deliver better outcomes for mental health service users and their families.

Further information

The websites for some of the voluntary organizations referenced in the chapter are:

Rethink: www.rethink.org

Canadian Association for Mental Health: www.camh.ca

EUFAMI—European Federation of Associations of Families of People with Mental Illness: www.eufami.org

Mental Health Foundation in New Zealand: www.mentalhealth.org.nz

National Alliance on Mental Illness: www.nami.org

SANE Australia: www.sane.org

Scottish recovery network: www.scottishrecovery.net

The World Fellowship for Schizophrenia and Allied Disorders (WFSAD): www.world-schizophrenia.org

7.10

Special problems

Contents

7.10.1 The special psychiatric problems of refugees

Richard F. Mollica, Melissa A. Culhane, and Daniel H. Hovelson

While the forced displacement of people from their homes has been described since ancient times, the past half-century has witnessed an expansion in the size of refugee populations of extraordinary numbers.[1,2] In 1970, for example, there were only 2.5 million refugees receiving international protection, primarily through the United Nations High Commission for Refugees (**UNHCR**). By 2006, UNHCR was legally responsible for 8.4 million refugees. In addition, it is conservatively estimated that an additional 23.7 million people are displaced within the borders of their own countries. Although similar in characteristics to refugees who have crossed international borders, internally displaced persons do not receive the same protection of international law. Adding all refugee-type persons together, the world is forced to acknowledge the reality that over the past decade more than 10 000 people per day became refugees or internally displaced persons.

The sheer magnitude of the global refugee crisis, the resettlement of large numbers of refugees in modern industrial nations such as Canada, the United States, Europe, and Australia, and the increased media attention to civil and ethnic conflict throughout the world has contributed to the medical and mental health issues of refugees becoming an issue of global concern. This chapter will focus on a comprehensive overview of the psychiatric evaluation and treatment of refugees and refugee communities. Although this mental health specialty is in its infancy, many scientific advances have been made that can facilitate the successful psychiatric care of refugee patients.

Definition

The definition of a refugee as outlined in the 1951 Convention and 1967 Protocol relating to the Status of Refugees is presented in Box 7.10.1.1.[3] A person or persons who has passed over from one country into another seeking protection from violence and who cannot return to his country of origin because of fear of persecution or injury is considered a refugee according to international law.

The 1951 Convention Relating to the Status of Refugees was drawn up by the United Nations parallel to the creation of UNHCR. This Convention and the subsequent 1967 Protocol establishes

Box 7.10.1.1 Definations according to the 1951 convention and 1967 protocol relating to the status of refugees

Article 1—Definition of the term 'refugee' A(2) [Any person who] . . . owing to well-founded fear of being persecuted for reasons of race, religion, nationality, membership of particular social group or political opinion, is outside the country of his nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country; or who, not having nationality and being outside the country of his former habitual residence . . . , is unable or, owing to such fear, is unwilling to return to it. (As amended by Article 1(2) of the 1967 Protocol.)

Article 33—Prohibition of expulsion or return (refoulement)
(1) No contracting state shall expel or return (refouler) a refugee in any manner whatsoever to the frontiers of territories where his life or freedom would be threatened on account of his race, religion, nationality, membership of a particular social group or political opinion.

international law for the definition of refugees as well as the protection accorded to them. It also articulates the important principle of *non-refoulement* (Box 7.10.1.1), which states that no refugee can be returned to his or her country of origin or any other location where there is any probability that he or she will be harmed. These legal definitions indicate that a refugee is not an economic migrant or a traditional immigrant. Sadruddin Aga Khan, in a seminal report, was one of the first High Commissioners to acknowledge the human rights violations that are primarily responsible for the generation of refugee populations.[4] Corresponding to these international covenants, the international community has focused on the protection of refugees. There are two components to protection that are classically viewed by UNHCR as an essential aspect of its mandate. These two elements include protection against:

1 ongoing violence and potential injury to the refugee including being denied proper asylum and involuntary repatriation;

2 lack of adequate food, water, clothing, and other forms of material assistance.

The UN Declaration of Human Rights adopted by the United Nations in December 1948 and the United Nations Convention against Torture and Other Cruel Inhuman or Degrading Treatment or Punishment adopted in December 1984 extends the basic principles of refugee protection and asylum, and reaffirms the principle of *non-refoulement*. In most refugee crises, not withstanding the political and military barriers to protection, UNHCR and the international community strive to offer refugees a safe asylum and basic humanitarian aid.

Trauma and torture

By definition, most refugees have experienced traumatic life events of extraordinary brutality. Since the Second World War, empirical studies have investigated the relationship between mass violence, the refugee experience, and psychiatric morbidity. The earliest research focused on survivors of the Nazi concentration camps.[5–8] Shortly after the Second World War, Eitinger and his colleagues gave a detailed account of their medical and psychiatric examinations of concentration camp survivors. They postulated that the traumatizing process had a dual nature. They described the somatic traumas of captivity, such as head injury, hunger, and infections, as leading to a 'psycho-organic syndrome', and the predominately psychological traumas as leading to other psychiatric disorders such as depression. Thygesan's studies of concentration camp survivors in Denmark revealed similar results.[9,10] These early pioneering investigations of the psychosocial sequelae of the Nazi concentration camps established a preliminary baseline of traumatic outcomes for future generations of refugees, many of whom had experienced the trauma of similar experiences in Cambodia, Bosnia-Herzegovina, and elsewhere.

Increasingly, civilian populations carry the burden of ethnic conflict and mass violence. It is now estimated that more than 80 per cent of casualties caused by the recent violence in Africa, Asia, and Europe have primarily affected non-combatants.[11] Extensive research has revealed the major trauma events experienced by refugee populations fall into the eight groups below:

1 material deprivation

2 war-like conditions

3 bodily injury

4 forced confinement and coercion

5 forced to harm others

6 disappearance, death, or injury of loved ones

7 witnessing violence to others

8 brain injury

Every refugee situation will have a range of traumatic events that will fall into each of these categories that are unique or characteristic of a specific conflict. It is essential that the specific types of violence experienced by a given refugee population are well known to the psychiatric clinician who can use this knowledge to assess potential traumatic outcomes.[12] In addition to many unique forms of violence occurring in different refugee settings, the meaning of violent events also differs across cultures. Anecdotal, clinical, and epidemiological evidence suggests that certain categories of refugee trauma are more potent than others in producing psychiatric morbidity and other traumatic outcomes. Brain injury, sexual violence, torture and other forms of bodily injury, coercion, and forced confinement have great potential of causing psychiatric harm in refugees exposed to these events. Consistent with indicators of the 'potency' of specific trauma events, there evidence of a dose–effect relationship between cumulative trauma and psychiatric symptoms.[13] The personal aspects of human suffering associated with specific types of trauma, such as the murder of a child or the disappearance of a family member are still relatively undefined but obviously very difficult.

Many refugees have actually experienced torture. Recent research states that the most significant finding in the last 7 years may be that either torture has become more prevalent worldwide or the total number of events reported has increased, likely as a result of advocacy and elevated media attention.[14] After 25 years of research on treatment work with torture survivors, still no consensus exists for effective interventions within the field.

Conceptual model of traumatic outcomes

Research on refugees has revealed the persistence of negative health and social outcomes decades after their initial experience of violence and dislocation. Emergence of standardized criteria for psychiatric diagnoses and disability and the demonstrated ability to elicit trauma events through simple screening instruments in culturally diverse populations have allowed evidence to accumulate suggesting a model of traumatic outcomes associated with the refugee experience. This model is primarily based upon the classic epidemiological triad which describes the interaction between host (i.e. the refugee), agent (i.e. traumatic life experiences), and environment (e.g. refugee camp) in the pathogenesis of psychiatric disorders.[15,16] This model, illustrated in Fig. 7.10.1.1, allows equal attention to be given to all aspects of the refugee experience.

The model in Fig. 7.10.1.1 has three major elements. First, it suggests that the major medical outcomes associated with the refugee experience are medical illness, psychiatric disorders, and disability. Second, trauma and the personal and environmental characteristics of the refugee describe the major risk factors associated with violent outcomes. Third, the direction of the causal arrows in the model do not imply a lack of reciprocal relationships where none is indicated; instead they indicate what most investigations consider

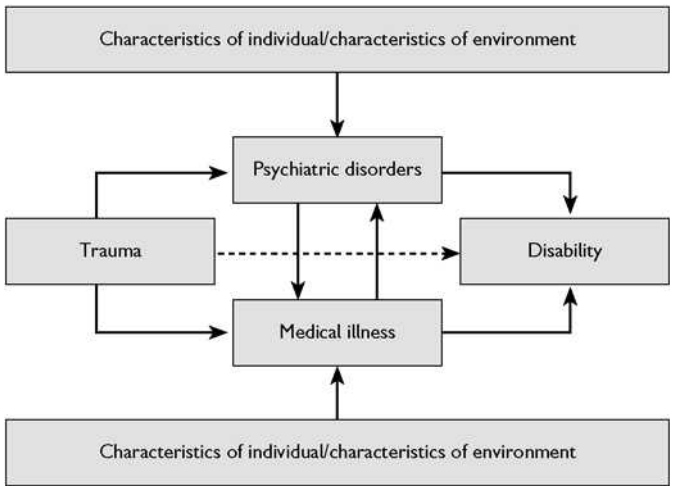


Fig. 7.10.1.1 Conceptual model of refugee risk factors and traumatic outcomes.

to be the most dominant causal relationship. Despite its limitations, this simple conceptual model can provide the psychiatric professional with a scheme for approaching the refugee patient from either a clinical or public health perspective. The importance of the socio-cultural and political context unique to each refugee situation and its impact on each of the model's pathways cannot be overstated, since refugees come from diverse cultural groups and political experiences.

Health status and medical illness

Refugees experience many diseases and chronic debilitating conditions, such as starvation and landmine injuries that have both immediate and long-term effects on their physical health. Every refugee situation involves many unique but common acts of violence leading to major medical sequelae. For example, gender-based violence and rape, which were instruments of ethnic cleansing in the Balkans, resulted in pregnancy, medically complicated self-administered abortions, and sexually transmitted diseases.[17] Extensive documentation of refugee survivors of torture describes the medical sequelae of torture and the causal link between refugee trauma and medical mortality and morbidity over time.

Throughout the process of migration, refugees are frequently exposed to infectious diseases and physical or psychological trauma that can have a profound impact on their overall health and well-being. Refugees are at increased risk for diseases such as obesity, diabetes, and heart disease, and often encounter significant barriers to employment and health care when attempting resettle. For these reasons, programmes implemented in the refugees' country of resettlement designed to improve health-seeking behaviours and overall self-care can have a positive, lasting impact on refugees' physical and mental health.[18]

Psychiatric symptoms and illness

Observations since Kinzie *et al.*[19] and Mollica *et al.*[20] first diagnosed PTSD in Cambodian refugees, have made the cultural validity of PTSD seem almost certain. However, this reality does not negate the importance of culture-specific symptoms related to trauma that are independent of PTSD criteria. Recent large-scale epidemiological studies of refugee populations have confirmed the high prevalence of major depression and PTSD in Western (e.g. Bosnian[21]) and non-Western (e.g. Cambodian[22] and Bhutanese[23]) refugee communities. The mental health impact of major depression, which presents both as a comorbid disorder with PTSD and alone, is chronic, severely disabling and demands the attention of the clinician working with refugees. Longitudinal data indicates that 45 per cent of Bosnian refugees who met criteria for PTSD, depression or both continued to meet criteria for these disorders 3 years later.[24] Similar results were found in a longitudinal study of Cambodian refugees 20 years after resettling in the United States.[25] As with Western populations, depression in refugees tends to be under-diagnosed and can be expressed as somatic complaints. A study of Vietnamese refugees showed high prevalence based on self-report, but high rate of physician under-diagnosis. Most patients with depression (95 per cent) presented with physical complaints.[26] These finding underscore the importance of depression screening especially in the primary care setting.

New research indicates that there may be memory problems in refugees with PTSD. When asked to recall traumatic events refugees with PTSD reported an increased number of traumatic or torture events over their baseline report, as compared to those with other psychiatric disorders who showed no change or decreases in number of events reported.[27] Substance use disorders are often overlooked in refugee populations but are often comorbid with PTSD.[28] Early in the 1980s reports of Hmong refugees using opium emerged.[29] Substance use disorders have been shown to have a delayed presentation of 5 to 10 years after the initial settlement of the refugee.[30] However, substance use disorders may vary by population. One recent study of Cambodian refugees in the United States reports low rates of alcohol use in the past 30 days.[31] Screening for substance use disorders especially in primary care is important. Complex grief reaction and chronic insomnia are also prevalent in this population.

(a) Head injury

Clinical evidence is also emerging identifying head injury as a cause of significant psychopathology in refugee survivors. Head trauma is one of the most common forms of torture, so much so that reports of torture almost always imply that some kind of head trauma occurred. Many head injury survivors experience seizures and headaches, as well as behavioural disturbances such as aggressiveness, irritability, and sleep disturbances. Recent research on Vietnamese ex-political detainees who experienced head trauma while in captivity indicates that the number of head injuries is related to a decrease in executive functioning, those with head injury have increased risk of developing PTSD, and had decreased cortical thickness in several brain regions (Mollica *et al.* 2007, unpublished). The long-lasting effects of head trauma are serious and pervasive, affecting both the injured individual and their family. The presence of head trauma in torture victims has been overlooked in past research, but awareness is growing.

(b) Functional status and disability

Although the functional status of refugees at the emergency and long-term ends of the continuum of the refugee experience has received little attention, the significance of this traumatic outcome is beginning to emerge. Until recently, the standard operating model of refugee protection has not been asked to determine the long-term socio-economic damage caused by the refugee experience. The answer to this question is extremely important to the

recovery of societies in which the majority of the population has been displaced. In many societies, the refugee experience is the majority experience, which has strong implications for future socio-economic development. While the prevalence of functional impairment and disability is unknown in refugee populations, a recent epidemiological study of Bosnian refugees in Croatia reveals that functional disability may, in fact, be extremely high, especially in elderly refugees.[21] Furthermore, disability may be exacerbated in refugee survivors who have both chronic medical and psychiatric disorders.

Psychiatric assessment

This section reviews key factors unique to the psychiatric evaluation and diagnosis of the refugee patient.

Primary care: the proper setting for a refugee clinic

The psychiatric literature has generally stressed the importance of evaluating and treating refugee patients in a primary health care setting, whether in a refugee camp or in a country of resettlement. Four factors seem to support this viewpoint:

1 refugee patients seldom self-refer to psychiatry;

2 in many societies considerable stigma is associated with psychiatry but not with primary care medicine;

3 the majority of refugees seek out the care of their local medical doctors and indigenous traditional healers for the relief of their emotional suffering;

4 most refugees have associated medical and psychiatric disorders.

Considerable field experience has shown that establishing a mental health programme within a health facility where refugees already seek medical care can result in the highly successful utilization of psychiatric professionals and treatment.

Cross-cultural psychiatric assessment and diagnosis

Early research describing the psychiatric status of refugee survivors, especially those who had been tortured, refrained from the use of psychiatric diagnoses because of a prevailing perception that the observed symptoms were a normal response to horrific life experiences.[12] Similarly, many medical anthropologists believed that Western psychiatric diagnostic classifications were not relevant to the assessment of suffering in non-Western populations.[32] Despite these reservations, the emergence of standardized diagnostic criteria for major depression and post-traumatic stress disorder (PTSD) have allowed for the cultural validity of these diagnoses to be tested in a number of refugee settings. Cross-cultural research suggests that assessments of psychiatric illness should begin with phenomenological descriptions of folk diagnoses or culture-specific syndromes.[33] Important methods of exploring the validity of DSM-IV diagnoses in cross-cultural settings have included using culturally valid definitions of functioning and mental health problems based on local views of maladaptive thoughts and behaviour in response to distress, not preconceived Western diagnostic categories[34]; however, to date, not a single culture-specific illness associated with the mass violence and torture experienced by refugees has been defined.[35] On the contrary, the criteria for the two major diagnoses associated with violence in Western society, i.e. major depression and PTSD, have been successfully applied to refugees from many parts of the world. While high rates of PTSD can be measured in refugee patients and traumatized civilians, it is not known if other culture-specific symptoms not part of the DSM-IV criteria that may have greater clinical relevance and meaning to a specific refugee group. A general principle demonstrated by the World Health Organization cross-cultural study of depression,[36] i.e. that while some depressive symptoms may be present across cultures, they may not be the symptoms most strongly endorsed by the patient; this principle may also apply to PTSD. Figure 7.10.1.2 provides an illustration, which can help to address the problem of psychiatric diagnoses in refugee patients. This figure suggests that, until further research is forthcoming, the psychiatric provider needs to determine clinically whether the refugee patient is presenting with scenario A, B, or C.

The high prevalence of psychiatric symptoms associated with trauma in refugee populations neither affirms nor negates the 'normalization' of these symptoms. A narrow medical viewpoint could create a psychiatric redefinition of refugee mental health problems that would place the majority of refugees in a 'mentally ill' box without any access to individual psychiatric care; on the other hand, the hostility of many humanitarian aid workers toward psychiatry has denied the seriously mentally ill refugee legitimate access to psychiatric treatment. There will probably be a compromise at the intersection of public health objectives and the protection goals of humanitarian aid workers. In future, the presence of chronic and severe disability in refugee survivors will be the gold standard which drives the psychiatric and humanitarian rehabilitation of refugee survivors.[37,38]

Psychiatric screening

If mental health practitioners are to treat refugee patients, they must be able to assess the refugee's major risk factors and traumatic outcomes. Simple screening instruments culturally adapted to the language, trauma, and symptoms of refugee patients have been found to be extremely effective as well as being well received by refugees themselves.[39–41] For example, a simple well-known

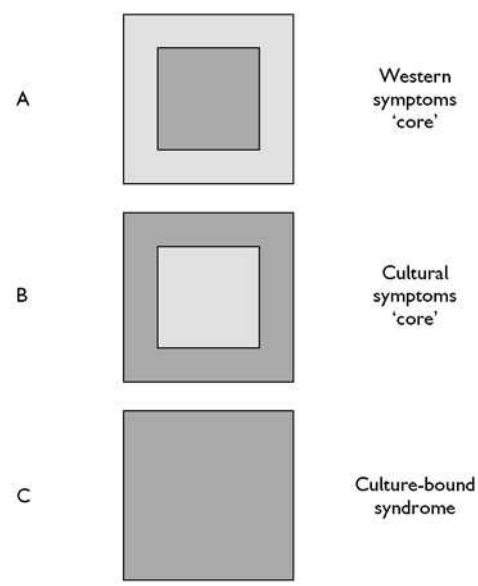


Fig. 7.10.1.2 Comparison of diagnostic classifications.

screening instrument adapted from the Hopkins Symptom Checklist allowed Indochinese patients for the first time to provide their symptoms with little distress. The development of the Harvard Trauma Questionnaire further revealed that these same patients could provide answers to lists of possible trauma events and symptoms without becoming retraumatized. Overall, two major lessons have been learned from the use of simple checklists with refugee survivors.

1 The checklist acknowledges the traumatic life experiences of the refugee survivors and *de facto* gives them permission to elaborate on the details of their trauma.

2 The checklist, as a simple medical test, helps refugees to 'put words around' events and symptoms that would be too emotionally overwhelming for them in an open-ended interview.

Extensive research regarding appropriate methods of development, translation, and validation of the clinical screening instruments such as the Hopkins Symptom Checklist and Harvard Trauma Questionnaire has been widely disseminated.[29,42]

Understanding somatic complaints

For years it was considered standard psychiatric wisdom that emotional distress, especially in non-Western patients, was primarily somatic in character; and that, because of their dominant expression of suffering through somatization, these patients were not psychologically minded, making them incapable of participating in psychotherapy or counselling. Extensive clinical experience and scientific studies with refugee patients throughout the world have led to a revision of these previously dominant professional attitudes. While it may be valid that some refugees primarily present their emotional suffering through somatic complaints, it has also been found that it is fairly easy for the medical practitioner to obtain from refugee patients deeper insights into their feelings and their beliefs as to the causes of their emotional distress. Often because of severe victimization, refugees as a group will not readily share their experiences of trauma and related medical and psychiatric injuries unless they are in a highly confidential environment where it is clear that the medical team can be trusted. In fact, the ability of refugee patients to extend themselves beyond their initial somatic complaints, as well as participating in clinical dialogue with the psychiatric practitioner as to the nature of their mental distress, is more the rule than the exception and is a major goal of assessment.[43]

Psychiatric treatment

The trauma story

The trauma story emerges as the centrepiece of any treatment approach. Every refugee patient has at least one traumatic experience that figures prominently in his or her life history. The trauma story is a living reality and is present for every patient. The trauma story is also present for the clinician. Yet, the story can be elusive and difficult for both patient and doctor to share. Often times when the patient is ready to tell the trauma story, the clinician is not ready to hear it. More frequently, when the clinician is ready for the patient to tell his or her story, the patient is unwilling. Therefore, one of the major goals of psychotherapy with refugees is to allow the trauma story to emerge gently and become a familiar and acceptable theme, which the refugee is not shamefully hiding from his or her therapist and ultimately his or her family and community.[44]

(a) Debriefing

Although commonly conducted within populations who have experienced trauma, typically by relief workers or crisis counsellors, debriefing has been shown to have negative effects on psychiatric outcomes and should be avoided at all costs.[45] As described above, the trauma story should be allowed to emerge naturally and at the patient's own pace.

(b) Treatment

Psychiatric treatment of refugee patients should primarily be based upon standard psychiatric practices for mentally ill patients, including the appropriate use of psychotropic medicines. However, a number of novel therapeutic approaches have been used with refugees. Recent research in refugee populations has indicated that cognitive behavioural therapy (CBT) is efficacious for treatment-resistant PTSD and panic attacks[46,47] and in refugees who have experienced torture.[48] Interpersonal psychotherapy for depression has been found to be useful in rural Uganda.[49] This approach may also be effective with resettled refugees in Western countries. In female veterans of war, prolonged exposure therapy was more effective over time than other interventions for reducing symptoms of PTSD in female veterans of war. It may be possible and beneficial to use prolonged exposure therapy for others with PTSD.[50] Similarly, although no research has specifically examined the effects of therapies such as eye movement desensitization and reprocessing (EMDR) for refugees, positive results have been seen in a variety of different studies done with traumatized populations, including individuals with PTSD. Traditional medicine, which includes diverse health practices, knowledge, and beliefs is widely accepted and practiced worldwide. Traditional medicine typically relies upon local classifications of emotional distress and its treatment can incorporate plant-, animal- or mineral-based medicines, as well as spiritual therapies or massage.[45] Traditional healing approaches were widely used for the Cambodian refugee crisis of the 1990s.[51–53] Although widely used, no conclusive research has yet been done using complementary treatments such as massage, acupuncture or herbal medicine in the treatment of torture survivors.[14] Recent research with Cambodian refugees it was shown that interventions supporting work, altruism, and spirituality for refugees might serve as protective factor for the onset of psychiatric disability.[54]

Special considerations

A number of special considerations have emerged from the many therapeutic approaches that have been tried with refugees.

Family focus

Treatment should be directed at the entire family of the identified refugee patient. The refugee crisis tends to initiate a process in which surviving family members become each other's major social support system. The disruption and disintegration of stable communities and traditional social supports place the entire burden of survival upon family members. Sometimes in extreme situations that affect the family unit such as when a refugee is experiencing symptoms of depression, psychosis, or social withdrawal, the psychiatrist must reach out to the patient and establish a one-to-one therapeutic relationship until family members can be of assistance.

Cultural sensitivity

The cultural sensitivity of refugee mental health practitioners is essential to the proper therapeutic relationship to the refugee patient. This means that the refugee mental health practitioner must be informed as to the type of trauma experienced by the refugee and its socio-cultural meaning, the cultural idioms by which human suffering is expressed in a given community, and the social stigma associated with mental illness. Despite the refugees' own adversity, long-standing social prejudices against mental illness will persist throughout the refugee crisis.

Little cross-cultural literature exists on the relationship between refugee patient, Western professional, and bilingual interpreter. The use of refugees as interpreters can be problematic. Ideally, health professionals from the refugee communities should be recruited to participate in psychiatric intervention. These individuals have had professional medical training as to the importance of confidentiality and can also provide insight into the cultural nuances of the doctor–patient relationship. The use of untrained interpreters from the refugee community should be avoided if at all possible. This is especially true of family members or members of a community or government institution that has previously threatened the security of the refugees. In addition, medical practitioners must respect patient dignity by not allowing young people to interview community elders or males from the community to ask refugee women explicit sexual questions and/or witness or be exposed to an undressed refugee woman during a medical examination. The more the bicultural interpreter can function in the role of a trained mental health paraprofessional, the more successful will be the therapeutic experience of all involved.

Caring for the seriously mentally ill

Despite numerous biases against psychiatric intervention in refugee populations, psychiatrists are well positioned to treat the seriously mentally ill refugee. In most refugee camps, those in immediate danger are the psychotic and depressed refugees with suicidal ideation. These individuals have difficulty coping with their ongoing crisis and have high mortality rates. Once resettled, these refugees with serious mental health issues need appropriate medical and mental health services.[55]

Special treatment needs of gender-based violence

Many colleagues in the field of refugee mental health have sought to break the 'conspiracy of silence related to sexual violence'.[32] While this issue has been well documented in refugee communities, it was not until the Bosnian conflict that rape was finally accepted by the international community not as a criminal act but as a crime against humanity. This recognition of gender-based violence as the most common type of torture of women, as well as a major terrorist instrument of war, has contributed to the protection and psychiatric care of sexually abused refugee survivors. While the political will now exists to condemn rape, and to protect refugees from it, the cultural stigma and corresponding social punishments of women considered 'tainted' by sexual violence, sometimes including their murder ('honour killing') by relatives, continues to make the psychiatric care of these survivors extremely difficult. In most cultures, rape remains a secret issue caused by the refugee's extreme resistance to reveal any details to the physician. Psychiatric practitioners must approach this issue with extreme caution in a strictly confidential manner. They must also be aware of the severe consequences to the patient if the rape experience becomes public knowledge, even to the patient's family members.

Risk and resiliency factors

The pre- and post-conflict personality characteristics of refugees that increase resiliency and reduce psychiatric distress and disability are not known. While certain demographic characteristics have been associated with negative traumatic outcomes, these characteristics may be confounded by other risk factors. Women, especially widows, seem especially vulnerable to negative refugee effects. UNICEF has extensively reviewed those risk factors associated with the vulnerability of refugee children and adolescents.[56] Data from Bosnian refugees correspondingly reveal the high rates of disability associated with trauma and psychiatric comorbidity in the elderly.

Lessons learned from studies of political prisoners in Turkey reveal the importance of a well-established political world view as a major protection against the long-term human suffering associated with torture.[57] Studies of Bhutanese refugees in Nepal confirm the possible protective function of Buddhism in devout refugee practitioners of this religion.[23] Anecdotal reports by refugees themselves consistently confirm the emotional safety that they have found in their spiritual and religious beliefs and practices. Finally, recent research findings concur with the earlier research in concentration camp survivors and prisoners of war that prior psychiatric history and premorbid personality factors may have little effect on the psychiatric sequelae of traumatic refugee experiences. In resettlement countries, opportunities related to family unification, learning to speak the new country's language, and employment have clearly been shown to be associated with decreased psychiatric morbidity over time. Those who provide psychiatric care for refugees must keep in mind the environmental opportunities that can increase their overall resiliency.

Reducing risk and maximizing resiliency

Figure 7.10.1.1 provides a readily accessible model for psychiatric practitioners to determine how they can reduce the risk factors associated with psychiatric disorders as well as promote the resiliency of the refugee patient. The section on risk factors provides many useful insights into the importance of enhancing the refugee's active role in work, spiritual participation, involvement in altruistic behaviour, and the many other personality and environmental factors that can directly reduce depression and other forms of psychiatric distress.

Further information

www.hprt-cambridge.org <http://www.hprt-cambridge.org> (Harvard Programme in Refugee Trauma)

http://mentalhealth.samhsa.gov/cmhs/SpecialPopulations/refugmhnew.asp (U.S. Substance Abuse and Mental Health Services Administration, Refugee Mental Health Division)

www.unhcr.org (United Nations High Commission on Refugees)

References

1. Office of the United Nations High Commissioner for Refugees. (1993). *The state of the world's refugees, 1993: the challenge of protection*, Vol. 9, p. 191. Penguin Books, New York.
2. Office of the United Nations High Commissioner for Refugees. (1997). *The state of the world's refugees, 1997–1998: a humanitarian agenda*, Vol. 12, p. 298. Oxford, England; Oxford University Press, New York.

3. Office of the United Nations High Commissioner for Refugees. (1992). *Handbook on procedures and criteria for determining refugee status: under the 1951 convention and the 1967 protocol relating to the status of refugees*, p. 93. Office of the United Nations High Commissioner for Refugees, Geneva.
4. Independent Commission on International Humanitarian Issues. (1986). *Refugees: the dynamics of displacement: a report for the independent commission on international humanitarian issues*, Vol. 18, p. 152. Zed Books, Atlantic Highlands, NJ, London.
5. Eitinger, L. (1961). Pathology of the concentration camp syndrome. Preliminary report. *Archives of General Psychiatry*, **5**, 371–9.
6. Eitinger, L. (1972). *Concentration camp survivors in Norway and Israel*, p. 199. Martinus Nijhoff, The Hague.
7. Eitinger, L. and Strom, A. (1973). *Mortality and morbidity after excessive stress*. Oslo University Press, Oslo, Norway.
8. Strøm, A.C.S. (1968). *Norwegian concentration camp survivors*. Universitetsforlaget; Humanities Press, Oslo, New York.
9. Thygesen, P. (1980). The concentration camp syndrome. *Danish Medical Bulletin*, **27**(5), 224–8.
10. Thygesen, P., Hermann, K., and Willanger, R. (1970). Concentration camp survivors in Denmark: persecution, disease, disability, compensation. A 23-year follow-up. A survey of the long-term effects of severe environmental stress. *Danish Medical Bulletin*, **17**(3), 65–108.
11. Levy, B.S. and Sidel, V.W. (1997). *War and public health*, Vol. 19, p. 412. Oxford University Press, New York.
12. Mollica, R.F. and Caspi-Yavin, Y. (1992). Overview: the assessment and diagnosis of torture events and symptoms. In *Torture and its consequences* (ed. M. Basoglu), pp. 253–74. Cambridge University Press, Cambridge.
13. Mollica, R.F., *et al.* (1998). The dose-effect relationships between torture and psychiatric symptoms in Vietnamese ex-political detainees and a comparison group. *The Journal of Nervous and Mental Disease*, **186**(9), 543–53.
14. Quiroga, J. and Jaranson, J. (2006). Politically-motivated torture and its survivors: a desk study review of the literature. *Torture*, **15**(2–3), 1–111.
15. Dohrenwend, B.P. (1998). *Adversity, stress, and psychopathology*, Vol. 15, p. 567. Oxford University Press, New York.
16. Susser, M. (1981). The epidemiology of life stress. *Psychological Medicine*, **11**(1), 1–8.
17. Swiss, S. and Giller, J.E. (1993). Rape as a crime of war. A medical perspective. *The Journal of the American Medical Association*, **270**(5), 612–15.
18. Research Triangle Institute. (2005). *Refugee health promotion and disease prevention toolkit*. SAMHSA Center for Mental Health Services, Editor. Rockville, MD.
19. Kinzie, J.D., *et al.* (1990). The prevalence of posttraumatic stress disorder and its clinical significance among Southeast Asian refugees. *The American Journal of Psychiatry*, **147**(7), 913–17.
20. Mollica, R.F., Wyshak, G., and Lavelle, J. (1987). The psychosocial impact of war trauma and torture on Southeast Asian refugees. *The American Journal of Psychiatry*, **144**(12), 1567–72.
21. HPRT. (1998). *Trauma and disability: long-term recovery of Bosnian refugees*. Harvard Program in Refugee Trauma, Cambridge, MA.
22. Mollica, R.F., *et al.* (1993). The effect of trauma and confinement on functional health and mental health status of Cambodians living in Thailand-Cambodia border camps. *The Journal of the American Medical Association*, **270**(5), 581–6.
23. Shrestha, N.M., *et al.* (1998). Impact of torture on refugees displaced within the developing world: symptomatology among Bhutanese refugees in Nepal. *The Journal of the American Medical Association*, **280**(5), 443–8.
24. Mollica, R., Sarajlic, N., Chernoff, M., *et al.* (2001). Longitudinal study of psychiatric symptoms, disability, mortality, and emigration among Bosnian refugees. *The Journal of the American Medical Association*, **286**(5), 546–4.
25. Marshall, G., Schell, T.L., Elliott, M.N., *et al.* (2005). Mental health of Cambodian refugees 2 decades after resettlement in the United States. *The Journal of the American Medical Association*, **294**(5), 571–9.
26. Lin, E., Ihle, L.J., and Tazuma, L. (1985). Depression among Vietnamese refugees in a primary care clinic. *The American Journal of Medicine*, **78**(1), 41–4.
27. Mollica, R., Caridad, K., and Massagli, M. (in press). Longitudinal study of posttraumatic stress disorder, depression, and changes in traumatic memories over time in Bosnian refugees. *The Journal of Nervous and Mental Disease*.
28. Brune, M., *et al.* (2003). Treatment of drug addiction in traumatised refugees. A case report. *European Addiction Research*, **9**(3), 144–6.
29. Westermeyer, J., Lyfoung, T., and Neider, J. (1989). An epidemic of opium dependence among Asian refugees in Minnesota: characteristics and causes. *British Journal of Addiction*, **84**(7), 785–9.
30. Westermeyer, J. (1995). Cultural aspects of substance abuse and alcoholism. Assessment and management. *The Psychiatric Clinics North America*, **18**(3), 589–605.
31. D'Amico, E., Schnell, T.L., Marshall, G.N., *et al.* (2007). Problem drinking among Cambodian refugees in the United States: how big of a problem is it? *Journal of Studies on Alcohol*, **68**(1), 11–17.
32. Goldfeld, A.E., *et al.* (1988). The physical and psychological sequelae of torture. Symptomatology and diagnosis. *The Journal of the American Medical Association*, **259**(18), 2725–9.
33. Westermeyer, J. (1981). Lao folk diagnosis for mental disorders: comparison with psychiatric diagnosis and assessment with psychiatric rating scales. *Medical Anthropology*, **5**, 425–43.
34. Bolton, P. (2001). Local perceptions of the mental health effects of the Rwandan genocide. *The Journal of Nervous and Mental Disease*, **189**(4), 243–8.
35. Simons, R.C. and Hughes, C.C. (1985). *The Culture-bound syndromes: folk illnesses of psychiatric and anthropological interest*. Culture, illness, and healing, Vol. 15, p. 516. D. Reidel; Sold and distributed in the U.S.A. and Canada by Kluwer Academic Publishers, Boston, Hingham, MA.
36. Jablensky, A., *et al.* (1981). Characteristics of depressive patients contacting psychiatric services in four cultures. A report from the WHO collaborative study on the assessment of depressive disorders. *Acta Psychiatrica Scandinavica*, **63**(4), 367–83.
37. Ingstad, B. and Whyte, S.R. (1995). *Disability and culture*, Vol. 10, p. 307. University of California Press, Berkeley.
38. Ormel, J., *et al.* (1994). Common mental disorders and disability across cultures. Results from the WHO collaborative study on psychological problems in general health care. *The Journal of the American Medical Association*, **272**(22), 1741–8.
39. Mollica, R., Caspi-Yavin, Y., and Lavelle, J. (1996). The Harvard Trauma Questionnaire (HTQ) manual: Cambodian, Laotian, and Vietnamese versions. *Torture—Quarterly Journal on Rehabilitation of Torture Victims and Prevention of Torture*, (Suppl. 1), 19–42.
40. Mollica, R.F., *et al.* (1992). The Harvard Trauma Questionnaire. Validating a cross-cultural instrument for measuring torture, trauma, and posttraumatic stress disorder in Indochinese refugees. *The Journal of Nervous and Mental Disease*, **180**(2), 111–6.
41. Willis, G.B. and Gonzalez, A. (1998). Methodological issues in the use of survey questionnaires to assess the health effects of torture. *The Journal of Nervous and Mental Disease*, **186**(5), 283–9.
42. Mollica, R.F., McDonald, L.S., Massagli, M.P., and Silove, D.M. (eds.) (2004). *Measuring trauma, measuring torture*. Harvard Program in Refugee Trauma, Cambridge, MA.
43. Mollica, R.F. and Lavelle, J. (1988). Southeast Asian refugees. In *Clinical guidelines in cross-cultural mental health* (eds. L. Comas-Diaz and E. Griffith), pp. 262–304. Wiley, New York.
44. Mollica, R.F. (2006). *Healing invisible wounds: paths to hope and recovery in a violent world*. Harcourt Press, New York.

45. Mollica, R.F., *et al.* (2004). Mental health in complex emergencies. *Lancet*, **364**(9450), 2058–67.
46. Hinton, D.E., *et al.* (2005). A randomized controlled trial of cognitive-behavior therapy for Cambodian refugees with treatment-resistant PTSD and panic attacks: a cross-over design. *Journal of Traumatic Stress*, **18**(6), 617–29.
47. Hinton, D.E., *et al.* (2004). CBT for Vietnamese refugees with treatment-resistant PTSD and panic attacks: a pilot study. *Journal of Traumatic Stress*, **17**(5), 429–33.
48. Basoglu, M., *et al.* (2004). Cognitive-behavioral treatment of tortured asylum seekers: a case study. *Journal of Anxiety Disorders*, **18**(3), 357–69.
49. Bolton, P., *et al.* (2003). Group interpersonal psychotherapy for depression in rural Uganda: a randomized controlled trial. *The Journal of the American Medical Association*, **289**(23), 3117–24.
50. Schnurr, P.P., *et al.* (2007). Cognitive behavioral therapy for posttraumatic stress disorder in women: a randomized controlled trial. *The Journal of the American Medical Association*, **297**(8), 820–30.
51. Lavelle, J., Tor, S., Mollica, R.F., *et al.* (eds.) (1996). *Harvard guide to Khmer mental health*. Harvard Program in Refugee Trauma, Cambridge.
52. Mollica, R.F., Tor, S. and Lavelle, J. (1998). *Pathway to healing*. Harvard Program in Refugee Trauma, Cambridge.
53. Heigel, J. (1994). Use of indigenous concepts and healers in the care of refugees: some experiences from the Thai border camps. In *Amidst peril and pain: the mental health and well-being of the world's refugees* (ed. A. Marcella). American Psychological Association, Washington, DC.
54. Mollica, R.F., *et al.* (2002). Science-based policy for psychosocial interventions in refugee camps: a Cambodian example. *The Journal of Nervous and Mental Disease*, **190**(3), 158–66.
55. Silove, D., Ekblad, S., and Mollica, R. (2000). The rights of the severely mentally ill in post-conflict societies. *Lancet*, **355**(9214), 1548–9.
56. UNICEF. (1990). *Children and development in the 1990s: a UNICEF source book on the occasion of the World Summit for Children*. UNICEF, New York.
57. Basoglu, M. (1992). *Torture and its consequences: current treatment approaches*, Vol. 13, p. 527. Cambridge University Press, Cambridge, New York.

7.10.2 Mental health services for homeless mentally ill people

Tom K. J. Craig

Definition and demography of homelessness and its link to mental illness

The term 'homeless' has been used to describe populations as diverse as those sleeping in the shelter of a cardboard box, to those sleeping on a friend's floor. Given such wide definition, it is not surprising that estimates of the numbers involved vary greatly from survey to survey and from one country to another. But regardless of the definition, there is consensus that the numbers of homeless people in most Western urban areas increased during the past two decades, reflecting a scarcity of low-cost housing, the erosion of traditional family networks, and downsizing in the organization and delivery of supportive services. Of all these factors, the shortage of affordable accommodation is the most important. For example, in England there has been a 40 per cent increase since 2002 in the number of households on waiting lists for social housing with estimates that a minimum of 20 000 housing units above current government targets are required to simply meet newly arising urgent need.[1]

Compared with a domiciled population, homeless people are less likely to have completed basic education, less likely to have ever held employment, and more likely to have experienced parental neglect and abuse in their childhood.[2]

Given the evidence linking homelessness to poverty and social disadvantage, it is hardly surprising that homeless people report higher rates of psychiatric disorder relative to the general population. While rates vary depending on the particular measure of mental illness adopted by each study and by the homeless population being investigated, most report major psychiatric disorder in 30 to 60 per cent of those using emergency shelters and sleeping rough. The prevalence of schizophrenia and other psychoses is particularly high amongst the middle-aged residents of long-stay hostels, while depression, generalized anxiety, and impulsive self-harm are more typically encountered in younger runaways and adolescent populations. Alcoholism and drug dependency are present in as many as two-thirds of men and a third of homeless women. Co-morbidity of mental illness and substance use disorder is the rule rather than the exception as are the co-occurrence of respiratory disease, infections, trauma, and the physical consequences of poor diet, poor hygiene, and the complications of substance abuse. Of growing concern is the accumulation of older multiply disabled populations in some North American cities.[3]

The typical pattern of service utilization of the severely mentally ill among the homeless population is one of extremes—bursts of involuntary hospital admissions and compulsory treatment interspersed with long periods of neglect and isolation. Many of those who are found sleeping rough or resident in temporary shelters have found their way to these locations as a conscious effort to avoid contact with health and social care professionals and remain unwilling to be part of any structured rehabilitation programme.

Barriers to care

Poverty and isolation

Very few homeless mentally ill people have satisfactory links to family or other supportive social groups. Unemployment is the norm and many have histories of contact with the criminal justice system. The lack of supportive kinship networks mean that there is seldom anyone who has an interest in their welfare and no one on whom services can rely for informal care giving. Affordable housing is likely to be of poor quality and unsupervised. Landlords are reluctant to rent property to someone with a history of destructive behaviour, a criminal record, or manifest mental illness.

Barriers arising from the illness

Severe mental illness contributes to incompetence in many aspects of daily life, with impaired social function and problems initiating and executing daily living tasks that require a degree of forward planning. Co-morbid cognitive impairment or substance abuse compound these problems. Many homeless patients will have lost their accommodation as a direct result of their illness, being evicted

for failing to keep up with rent payments, neglecting or damaging the property, or following complaints from neighbours.

Barriers put up by services

The lack of common purpose and co-ordination of social welfare, health, and criminal justice agencies lies at the heart of many difficulties faced by homeless mentally ill people. A young homeless person, for example, may be too chaotic to undertake the retraining programme that is his only route to welfare support, may be unable to register with a local family health centre because of his lack of a permanent address, and may be summarily rehoused without reference to health services involved in his care. Finally, the prejudices of professionals can make services unacceptable and the emphasis on treatment is seldom attractive to a homeless person whose immediate needs are for food, shelter, and security.

Principles of service organization and delivery

To state the obvious, the solution to problems of homelessness lies in the provision of suitable accommodation, targeted efforts to re-house the most vulnerable, and sufficient longer-term tenancy support to prevent a return to the streets. While many countries have social welfare legislation to assist homeless people, only a minority provide a legally enforceable right to suitable accommodation for vulnerable populations. The Rough Sleepers Initiative in England, provides emergency accommodation for the roofless population and follows this emergency re-housing with a Tenancy Sustainment Programme of flexible practical and emotional support to prevent future accommodation breakdown.[4] This has been very successful in reducing the numbers of rough sleepers though gaps remain, particularly for people suffering from severe mental illness and substance dependency. For these populations, a further tier of service involving specialized mental health provision is needed either as part of an intensive initial stabilization[5] or in the longer-term. While the detail of specialized services varies according to local circumstances, they are all based on a small number of ideological and organizational principles (Table 7.10.2.1).

Improved inter-agency co-operation

As a first step, most involve a steering group comprising senior representatives of the key stakeholders in health, housing, social services, police, and voluntary sectors. These groups oversee the development of services across a wide geographical area—a large sector of a city or state. The members carry sufficient political and managerial authority to be effective in dealing with bureaucratic obstacles that are bound to arise from time to time.

Table 7.10.2.1 Services for homeless mentally ill people

Essential components for rehousing
Availability of temporary accommodation with a pathway to permanent housing
Capacity to deliver basic needs (shelter, food, income support)
Ongoing practical support to maintain tenancy
Specialized mental health service
Steered by a partnership of key stakeholders (housing, health, welfare, etc.)
Multidisciplinary front-line team
Assertive outreach model
Capable of managing mental illness and substance-use disorder
Wider context
Community-orientated mainstream psychiatric services
Influence of central and local government policies on health, welfare, and criminal justice

Providing local multidisciplinary specialist teams

This co-ordination is replicated at the local level through multidisciplinary clinical teams, joint working, and case management. Such partnership ventures have been established in several cities in North America, Europe, and Australia using a variety of organizational approaches ranging from a single multidisciplinary team through 'one-stop shops' where professionals from a variety of backgrounds come together to provide services at a common location.

Essential components of the specialist service

The management of a homeless mentally ill person involves stages of engagement, stabilization, resettlement, and the eventual transfer of care to mainstream providers. Engagement can take a long time, staff must be prepared to leave the clinic and go to where homeless people congregate. Help with welfare and practical problems may be all that can be done at first but the duty to maintain a therapeutic focus must always be maintained. Stabilization requires the specialist assessment and treatments provided to any mentally ill person, including hospital admission if necessary. The task of resettlement typically involves a compromise between personal preferences, available resources, and the level of support needed to promote rehabilitation. For example, independent accommodation may be a person's first choice but may only be a viable prospect if it can be backed up by weekly visits from the mental health team. Core and cluster arrangements, in which residents have their own flat but receive supervision from an on-site warden within the complex is a particularly effective model for those who have failed in independent accommodation but who reject shared facilities.

The eventual transfer of care to mainstream services can be quite difficult to manage and most follow-up studies suggest that fewer than half of those transferred remain in treatment.

Conclusion

Specialist multidisciplinary teams for homeless mentally ill people provide an essential safety net for those who have fallen out of the wider mental health care system. They offer distinct advantages in terms of their capacity to work across traditional geographical and bureaucratic barriers, to take the longer-term view of the task of engagement, and to bring together the multiple strands of care across different provider agencies. Introduced as a temporary measure over a decade ago, they are still with us and likely to remain a permanent fixture of urban mental health care.

Further information

Access to mental health services for people who are homeless or living in temporary or insecure accommodation. A good practice guide: http://www.communities.gov.uk/index.asp?id=1162512

Essential statistics on homelessness in Britain: http:// www.homeless.org.uk/policyandinfo/facts/statistics

References

1. National Housing Federation. (2006). *England's housing time-bomb: affordability and supply 2006–11*. National Housing Federation, London. Website: http://www.housing.org.uk/uploads/file/campaigns/tb_england.pdf
2. Bhugra, D. (ed.) (1996). *Homelessness and mental health*. Cambridge University Press, Cambridge.
3. Hahn, J.A., Kushel, M.B., Bangsberg, D.R., *et al.* (2006). The aging of the homeless population: fourteen-year trends in San Francisco. *Journal of General Internal Medicine*, **21**, 775–8.
4. Lomax, D. and Netto, G. (2007). Evaluation of tenancy sustainment teams. Department of communities and local government. http://www.odpm.gov.uk/index.asp?d=1505917
5. Susser, E., Valencia, E., Conover, S., *et al.* (1997). Preventing recurrent homelessness among mentally ill men: a 'critical time' intervention after discharge from a shelter. *American Journal of Public Health*, **87**, 256–62.

7.10.3 Mental health services for ethnic minorities

Tom K. J. Craig and Dinesh Bhugra

Ethnicity, culture, and health care need

Services aimed at minority ethnic populations are all too often developed on the basis of conspicuous morbidity than on any real understanding of the diversity of ethnic minority communities and their wider health needs. For example, in England, while much has been written about ethnicity and psychiatric morbidity, the literature remains largely focused on African-Caribbeans and Asians, while the needs of the Irish, who comprise the largest ethnic community by migration in many parts of the UK are seldom explicitly addressed despite evidence of high rates of suicide and unexplained death many times in excess of the indigenous population.[1] In addition, the large numbers of asylum seekers and refugees who move around the world, brings an increased need for culturally sensitive services. But very few models exist for developing these. The principles of good practice indicate that the start has to be a clear knowledge of the population that will be accessing services and an appreciation of the complicating factors of social disadvantage, material deprivation, and poverty.

There is no doubt that social disadvantage and racial prejudice whether real or perceived are pivotal in determining not only the mental health of minority populations but also the pathways individuals and their families use in seeking help for ill health. Delays in help-seeking can also be due to the stigma of mental illness and to sufferers' fears that they will be misunderstood and mistreated because of differences in culture, language, and racist attitudes within the services. These factors may be more apparent in older individuals and those who were born outside the country who may not be aware of various options available to them. Studies over the past 30 years or more in Britain, the Netherlands, the United States, Canada, and Australia have shown that minority groups have lower access to mental health services, are less likely to receive care, and when they do this is more likely to be of a lower quality. Black people in the UK and the United States are more likely than white people to be compulsorily detained in hospital, to be screened for drug abuse, to receive higher doses of medication and physical rather than psychological therapies. They are over-represented compared with their numbers in the general population, whether in general wards, locked wards, secure units, court diversion schemes, special hospitals, or prisons.[2]

Some of these problems can be attributed to a lack of understanding on the part of mental health practitioners of the cultural beliefs, values, and practices of minority groups with consequent shortcomings in assessment, diagnosis, and the provision of care. While language can be a major obstacle, for many people from minority ethnic groups who speak English the problem is of communication rather than language. The power dynamic that is always present in any clinical consultation is magnified and both patient and doctor will have predetermined expectations of how their interaction will turn out depending upon their experience of previous consultations. Problems in the interaction are likely to be interpreted as arrogance and racism on the one hand and indifference, wariness, or docility on the other. Thus, both missed diagnoses and misdiagnosis may result. A lack of recognition of the personal, social, and cultural problems which influence the presenting patterns of symptoms in different ethnic groups can contribute to the tendency of clinicians to make assumptions and listen out for stereotypical triggers which then prompt a particular therapeutic response. Such triggers include religious euphoria, use of Cannabis in African-Caribbeans, and the 'fatalistic attitude' attributed to Asian patients.

Culturally competent services

The past decade has seen an emerging consensus that the way forward lies in the development of fully integrated multicultural services with good working links with the local minority community rather than separate services. Such services would provide staff who can understand their client's cultural background and the ways in which this influences the presentation of distress and disorder. There would be closer working links with religious leaders and healers of local ethnic minority communities, female-only areas on wards, and a greater involvement and support of the family in understanding the problems and developing solutions.

In North America,[3,4] Britain,[5] and Australia[6] there has been a significant 'top–down' pressure to shift health care organizations in this direction and several large-scale programmes such as the European 9-country 'Migrant-Friendly Hospitals' initiative[7] have been reported. Although differing in detail, all these programmes share common elements. These principles are outlined in Table 7.10.3.1 setting out the main conclusions in a 'bottom–up' approach in order to emphasize the importance of changes in the attitudes, knowledge, and skills of front-line managerial and clinical staff.

Commonly referred to as 'cultural competence' these attributes include attention to obvious language differences but go further to include history, traditions, beliefs, and values even if the latter differ from those held by the professional. A culturally competent clinician is sensitive to a patient's cultural influences, expressions of distress and help-seeking, and is also aware of their own attitudes and prejudices and how these are in turn shaped by their own cultural background. This objective is achieved through

Table 7.10.3.1 Organizational steps towards a culturally competent mental health service

1 **Workforce level**
Training in 'cultural competence' should be mandatory for all mental health professionals
(a) Undergraduate programmes
(b) Post-graduate as continuing professional development
2 **Health care provider level**
Provide accommodation, washing, and living space facilities that take into account different cultural and gender definitions of ordinary social behaviour, dignity, and respect.
Senior management responsibility and accountability for:
(a) Active race equality policies
(b) Recruitment policy—to increase presence of minority staff and provision of training and support as needed
(c) Ensure staff have received relevant cultural competency training
Ensure adequate data collection includes a robust estimate of the numbers of ethnic minorities using services with a focus on key areas such as disparities by ethnicity in the use of coercive treatments, dropout from follow-up; differences in the uptake of psychological and pharmacological treatments.
Specialized outreach services targeting mentally ill people in the criminal justice system, homeless, and refugee populations.
Partnership arrangements with NGOs including the provision of volunteer advisors and of translation services that do not rely on relatives or other informal carers.
3 **Wider health service level**
At the appropriate Regional or National HMO or Statutory Organization:
(a) Policy and practice commitment to removing barriers to access, e.g. extend health insurance to the uninsured and closer integration of primary and secondary mental health care
(b) Collection of good quality demographic information and ethnic monitoring for planning and overseeing services
(c) Minority representation at all levels of health service planning and delivery
Continue to expand the science base to determine what works best for whom

training and a plethora of different cultural competency/diversity courses have sprung up in recent years. A search of the Internet identifies courses in undergraduate nursing, medical, and pharmacy programmes, in post-graduate continuing professional development and as part of wider organizational change and development. While the moral argument for improved cultural competence is hardly contestable, whether or not these training courses are sufficient to effect lasting change in behaviour and the delivery of health care is less certain. The very limited empirical evidence base is predominantly from the United States and shows efficacy in terms of short-term changes in attitudes and knowledge. Evidence is still lacking on which elements of this training are essential and on the downstream effects on service quality that is the real target.

Table 7.10.3.1 also shows the other key steps that have been taken towards developing more culturally sensitive services. Most health service providers have gone some way to providing female-only inpatient wards, more community-orientated service settings and addressing the need for expanding the numbers of people from minority backgrounds in their workforce though it is still all too easy to think that problems with cultural sensitivity can be solved with this alone. Simply hiring people on the basis of nationality, ethnicity, or skin colour will not necessarily ameliorate problems in the service. Typically these staff are the lowest in the hierarchy of power and have the least capacity for influencing either the care of the patient or wider attitudes within the institution. Even when employed in a position of power there is a danger of their appointment being seen as tokenistic.

Another important step involves the elaboration of performance indicators to assess the impact of training and service changes and ethnic monitoring with a focus on key areas such as the use of coercive treatments, treatment discontinuation, and ethnic differences in the uptake of psychological and pharmacological treatments. In the most sophisticated systems this includes both quantitative epidemiological and economic data as well as qualitative inputs from consultation with users, carers, and the general public including community and religious leaders.[8]

Finally, at the wider community level, NGOs have often become the champion of good practice, seen by their users as an antidote to inadequate mainstream care. These small organizations are generally based on consultation with users, carers, and local mental health professionals and are more in tune with the expressed needs of the community. The best have good working relations with mainstream services, and are generally seen as making an important contribution to wider community care. They provide a range of supportive services and are ideal vehicles for health promotion and dissemination of health-related information. Given the need to develop ways of working which promote inclusion of patients, their families, and the community in general, it is impossible to overstate the importance of effective liaison between the voluntary sector and mainstream psychiatric services. People often come into contact with the emergency services because there is a lack of knowledge of the availability of community alternatives and of where to go when distressed. Working with voluntary services will not only contribute to their longevity but also ensure that the complementary treatment modalities they offer are part of the service provided.

Even where these system-wide initiatives are absent or not yet fully implemented, individual clinicians can do a lot to improve the care of minorities within their own service. Finding out the scale of the problem is a good place to start. Are population estimates available from a recent census, is information available from a recent census or from local government sources? More importantly, are there known problems of access for these populations in general or particularly within local mental health services? Are significant numbers not treated because they are held in 'inappropriate' settings in the criminal justice system or because they are mobile populations? In terms of delivering the service, outreach is likely to be the key and this will almost always involve partnership arrangements, developing understanding of key cultural influences including the importance of alternative health perspectives, spirituality and traditional healing. Several innovative services have been developed alongside organizations that 'hold' significant numbers of the minority population such as the Church or the local Mosque or NGOs dealing with specific communities, which can provide a wider social or housing service to minority populations. It is the near universal experience of these services that initial progress is slow with many barriers of mutual misunderstanding and suspicion to overcome. Developing a relationship or even employing someone from the minority culture as a go-between, advisor, or 'cultural consultant' can be helpful though it needs careful preparation to avoid selecting a consultant from the wrong tribal

background or at the wrong level of seniority, gender, or language group. Culture broker models have been used in the United States, whereas cultural consultation services happen in parts of the UK and cultural liaison officers are used in parts of Australia. Whatever the choice, it is likely to take time and patience to develop a high level of 'visibility' in the target community as well as a sound understanding of its culture, taboos, and historical context. Where available, cultural supervision from a more experienced practitioner is also helpful. The underlying principle has to be a two-way process, which deals with information from the patient's community and the service provider to ensure that communications are clear.

A few further points can usefully be taken into account in planning the service. First an understanding of local models of illness that determine when and how and of whom help will be sought, second use of local epidemiological data to assess the impact of age and gender on service demand and finally the involvement of the local community to promote a sense of ownership and involvement in the delivery of services.

In conclusion, the diagnosis and treatment of mental ill health among multiethnic populations is probably one of the most complex and contentious challenges in psychiatric service provision. At the heart of this complexity are problems of ignorance, attitude, and failures of communication on all sides. It is also potentially one of the most rewarding endeavours if got right. No one service model is likely to apply to every community, even if people belong to the same ethnic group. The development of specialist psychiatric services may not always be possible or even essential. Instead, the requirements are for approaches that are flexible, sensitive, accessible, and accountable to the people they serve.

Further information

American Psychological Association guidelines for providers of services to ethnic linguistic and culturally diverse populations: http://www.apa.org/pi/guide.html

Sainsbury Centre for Mental Health. (2002). Breaking the circles of fear http://www.scmh.org.uk

References

1. Bhugra, D. (2004). *Culture and self-harm*. Psychology Press, Hove.
2. Lipsedge, M. (1993). Mental health: access to care for black and ethnic minority people. In *Access to health care for people from black and ethnic minorities* (eds. A. Hopkins and V. Bahl), pp. 169–83. Royal College of Physicians, London.
3. Office of Minority Health (USA). Assuring cultural competence in health care: recommendations for national standards and an outcome-focused research agenda. www.omhrc.gov/Assets/pdf/checked/Assuring_Cultural_Competence_in_Health_Care-1999.pdf
4. Mental Health: Culture Race and Ethnicity. A supplement to Mental Health: A Report of the Surgeon General. www.surgeongeneral.gov/library/mentalhealth/cre/
5. Department of Health. (2005). *Delivering race equality in mental health care: an action plan for reform inside and outside services and the government's response to the independent inquiry into the death of David Bennett*. HMSO, London. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4100773
6. Royal Australasian College of Physicians. (2005). Policy statement: aboriginal and Torres Strait Islander Health. Available at: http://www.racp.edu.au/index.cfm?objectId=49F4E2A9-2A57-5487-D0597D1ED8218B61
7. Migrant-Friendly Hospitals Initiative. (2004). Available at: http://www.mfh-eu.net/public/home.htm
8. Jordan, J., Dowswell, T., Harrison, S., *et al*. (1998). Health needs assessment. Whose priorities? Listening to users and the public. *British Medical Journal*, **316**, 1668–70.

SECTION 8

The Psychiatry of Old Age

This page intentionally left blank

8.1

The biology of ageing

Alan H. Bittles

Introduction

Although old age is readily recognizable, methods to define and measure the underlying biological processes are much less amenable to study. For this reason, **life expectancy** has been widely used as a surrogate measure of ageing, as well as to monitor economic progress at national and regional levels. It is generally acknowledged that lifespan is a constitutional feature of the human phenotype, and twin studies have indicated that 25–33 per cent of the variance in human **longevity** is genetic in origin.[1,2] External factors including lifestyle can also exert a major influence, as illustrated by the current mean life expectancies of 79 and 86 years for males and females in Japan, whereas the comparable figures for Botswana are 35 and 33 years, respectively.

The importance of genetic inheritance as a determinant of extended survival has been illustrated by population level studies in Okinawa, an island prefecture of southern Japan with a very high prevalence of long-lived individuals. On the island, the mortality rates of the male and female siblings of centenarians were approximately half those of birth cohort-matched, non-centenarian siblings.[3] These findings parallel an earlier study of the family of Jeanne Calment, who died in France in 1997 aged 122 years. Of her 55 relatives, 24 per cent had lived to >80 years compared to just 2 per cent of a matched control group.[4] However, it remains unclear whether the enhanced lifespan of individuals who exhibit above average longevity is due to a slowing of the overall ageing process or is primarily associated with resistance to major life-threatening pathologies.

The concept of an '**allostatic load**', potentially involving the neuroendocrine, sympathetic nervous, immune and cardiovascular systems, and metabolic pathways, has been advanced to describe the lifetime costs of adapting to physical and psychological stresses. According to this hypothesis, while the actions of biological mediators of **stress** can be initially beneficial to health, chronic stimulation results in regulatory imbalance and subsequent pathophysiological changes.[5] Empirical studies have indicated increased physiological dysregulation and functional decline at >70 years of age, which would imply that predicted global increases in the numbers of older persons will be accompanied by disproportionately larger groups of individuals with major age-related pathologies.

Theories of ageing

While initially popular, it became apparent that single, 'magic bullet' causes of ageing were inappropriate to complex biological species, and with this recognition earlier organ- and system-based theories have gradually been discounted. Conversely, the observation that ageing appears to be initiated at different ages and can proceed at different rates in individual members of a species provides presumptive evidence for the interaction of multiple genetic and non-genetic influences. Two main groups of theories have been formulated, genomic and stochastic, each subdivided into a number of discrete topic headings.

Genomic theories of ageing

Genomic theories premise that ageing is primarily associated with changes in the genetic constitution of the organism. Support for genomic theories stems largely from the characteristic life expectancies of mammalian and non-mammalian species, and theories proposing a primarily genetic basis for ageing were greatly strengthened by the demonstration that human diploid cells exhibited a highly reproducible lifespan when cultured in the laboratory. Although strong evolutionary advantages can be envisaged for genetic control of developmental changes up to and including reproductive adulthood, the existence of genes uniquely encoding ageing seems improbable since few free-living animals or humans have ever succeeded in attaining the maximum lifespan of their species.

(a) Information transfer

The ability to synthesize functional proteins is dependent on the fidelity of genetic information encoded in the DNA, its unimpaired transcription from DNA to RNA, and translation into peptides and proteins. As each of these processes is subject to inaccuracy, and during the life course of an organism the sequence of information transfer steps is continuously operational, the error potential is large. With increasing chronological age the probability of errors increases, resulting in the accumulation of deleterious mutations late in life.[6] Since a number of the proteins synthesized may be involved as surveillance enzymes to maintain the accuracy of the entire system, feedback mechanisms could lead to its collapse, resulting in a phenomenon initially termed **error catastrophe**.

(b) Somatic mutation

With the demonstration of an inverse correlation between the lifespan of mammalian species and the incidence of chromosome abnormalities, age-related physiological changes were originally ascribed to accumulated mutations in the nuclear DNA (**nDNA**) of somatic cells. Findings of this nature could, however, be explicable in terms of the ability of an organism to tolerate DNA damage via the repair of damaged molecules, with more than 130 human **DNA repair genes** identified.[7] The capacity of nDNA to resist attack by endogenous reactive species and environmental agents is therefore considerable, which casts doubt on the general applicability of the theory.

(c) Epigenetic mechanisms

Epigenetic errors, i.e. errors in the control of gene expression rather than mutations in DNA or protein, have been proposed as major primary causal factors in senescence.[8] In promoter regions of genes, hypermethylation silences a gene whereas the hypomethylation of previously methylated sequences permits their expression. The pattern of **DNA methylation** is established during development and is cell type-specific, and changes in methylation can occur both during ageing and in cancer cells. The advantage of epigenetic models of ageing is their lack of requirement for the evolutionary preservation of genes encoding ageing, which in former generations would seldom have been expressed.

(d) Mitochondrial decline

Mitochondria are subcellular organelles responsible for aerobic energy production in humans and many other species. The mitochondrial genome of ~16.5 kb is characterized by its extremely compact organization, with no protective histones, a lack of excision or recombinational repair mechanisms, and a virtual absence of introns, all of which make it highly susceptible to mutation. Mitochondrial DNA (**mtDNA**) plays a central role in mitochondrial propagation and the maintenance of cellular respiration, but a majority of proteins involved in the regulation of mtDNA transcription, translation and replication, and the mitochondrial respiratory chain, are encoded in the nuclear genome. This design requires the operation of a highly coordinated mechanism for the expression of the nuclear and mitochondrial genomes. The central role of mitochondria in energy production means that defects may be of major metabolic significance, and the demonstration of increased levels of mtDNA deletions and base-substitutions in aged human neurones, heart, and skeletal muscle suggest a causative role for mtDNA mutations in ageing.[9]

(e) Telomere loss

Telomeres are specialized structures located at the terminus of the DNA helix and critical to the maintenance of DNA stability and replication. The enzyme telomerase which is responsible for telomere synthesis is active during early embryonic and foetal development but its activity is down-regulated in all human somatic cells before birth. As human diploid fibroblasts in culture were shown to progressively lose telomeres, it was hypothesized that telomere length could act as a predictor of the potential *in vitro* lifespan achievable by a cell strain.[10] Humans have a common telomere profile found on lymphocytes, amniocytes, and fibroblasts which appears to be preserved throughout life. However, the rate of telomere loss with ageing varies between chromosomes and there is evidence that, in addition to the common human telomere profile, each person exhibits an individual profile.[11]

Besides ageing, telomere loss has been implicated in a wide range of disease states, including heart disease, stroke, infection, long-term chronic stress, and obesity. Given the apparent relationship between telomere loss and both ageing and age-related pathologies, pharmacological activation of telomerase has been proposed as a potential treatment for chronic or degenerative diseases.[12] As tumour tissue and transformed cells constitutively produce telomerase, any therapeutic intervention of this nature would require careful monitoring.

Stochastic theories of ageing

Stochastic theories of ageing propose that cumulative adverse random changes at the cellular level ultimately overwhelm the capacity of an organism to survive, with ageing representing the preceding period of functional decline.

(a) Rate of living

An optimum lifespan was achieved by a variety of non-mammalian species when the organisms were maintained at suboptimal temperatures. The further demonstration of an inverse relationship between basal metabolic rate and longevity in mammals was interpreted as evidence that a species lifespan was governed by its **rate of living**, which in turn was correlated with its level of energy expenditure. Theories of this type tend to be imprecise in defining the nature of the factor(s) controlling ageing and lifespan, although it was subsequently proposed that the rate of living theory could be reformulated as a **stress** theory of ageing, with stress resistance and longevity positively correlated.

(b) Waste product accumulation

Ageing has been ascribed to interference by accumulated waste products in normal cellular metabolism and function, ultimately resulting in dysfunction and death at cellular and organ levels. Lipofuscin, a highly insoluble, pigmented compound derived by auto-oxidation from incompletely degraded cellular materials and detected with advancing age in neurones, cardiac muscle fibres, and the adrenal cortex, has been particularly implicated. Alternatively, the build-up of lipofuscin in older organisms may be secondary to an age-related decline in the function of cellular catabolic processes.

(c) Macromolecule cross-linkage

Many macromolecules of biological importance develop cross-links with increasing chronological age. The establishment of cross-linkage, whether covalent in nature or due to hydrogen bonding, alters the chemical and physical properties of molecules. Thus cross-linkage of the extracellular protein collagen is believed to be responsible for the loss of elasticity in mammalian blood vessels and skin with advancing age, even though collagen is subject to turnover throughout the lifespan. DNA and RNA also are believed to be potential intracellular targets for cross-linking agents, and changes in their structure could have serious functional implications for cellular information flow.

(d) Post-synthetic modification

In addition to cross-linkage, molecular aggregation and immobilization that compromises cellular metabolism and function could be caused by post-synthetic modification of proteins, with non-enzymic glycosylation (glycation) particularly associated with ageing. Glycation is initiated by the reaction of glucose with the amino group of lysine residues, which then proceed to form a Schiff

base, and progressively more complex compounds collectively termed advanced glycosylation end (AGE) products. As little variation was found in the glycation levels of lens crystallin proteins in subjects aged between 10 and 80 years, post-synthetic mechanisms may be as much an effect as a cause of ageing.

(e) Free radical damage

The role of **free radicals** in ageing was first proposed over 50 years ago.[13] A wide range of highly reactive free radicals are derived from molecular oxygen, including the superoxide and hydroperoxyl radicals, hydrogen peroxide, hydroxyl radical, and singlet oxygen. The polyunsaturated fatty acid side chains of cell and organelle membranes form highly susceptible targets for the action of **reactive oxygen species (ROS)**, and the resulting lipid peroxidation can result in severe membrane damage and eventual death of the cell. DNA may also be a critical target molecule for free radical damage, with mtDNA especially susceptible because of its proximity to the site of free radical production in the inner mitochondrial membrane.[9] Although a wide variety of antioxidants have been identified in humans, including ascorbate, α-tocopherol, β-carotene, glutathione, and the enzymes superoxide dismutase, peroxidase, and catalase, there has been little experimental evidence that these antioxidants can produce a significant extension in maximum lifespan. However, transgenic mice expressing the free radical scavenger enzyme catalase targeted to mitochondria showed an approximately 20 per cent extension in their mean and maximum lifespans, and concomitant delays in cardiac pathology and cataract development.[14]

Ageing as an energy crisis

From an evolutionary perspective, it was suggested that senescence was the end-result of an energy conservation strategy operating in somatic cells. During the course of a lifespan, total available energy has to be allocated to a variety of functions, including macromolecular synthesis and degradation, cell and organ maintenance, and reproduction of the species. Since the energy supply is finite, and to ensure propagation of the species by the successful transmission of genes to future generations, a compromise has to be reached between the energy made available for each of these functions. According to the **disposable soma theory**, this accommodation in energy saving is achieved by maintenance of absolute or near absolute accuracy in germ cell replication but less rigorous error correction in somatic cells.[15]

As an organism ages, the demands placed on the free energy pool alter and increase from a primarily anabolic role to meeting the requirements of ever-increasing repair and catabolic functions, including those imposed by specific disease-related insults. If the mitochondrial inner membrane and/or mtDNA is damaged, an organism must increasingly rely on alternative, less efficient pathways for its energy needs, ultimately resulting in a critical shortfall in the energy supply needed to sustain life. In such an **energy crisis**, the somatic cells primarily affected would be post-mitotic cells with high energetic demands, typified by the heart, skeletal muscle, and the brain.

Dietary modification of ageing

Inherited factors clearly play a major role in ageing, and in determining the human lifespan. But if ageing also is stochastic in nature then it should be possible to modify development of the ageing phenotype by altering the relative influence of contributory environmental variables, including diet.

Dietary (or calorie) restriction, based on a diet reduced in total amount but otherwise nutritionally adequate, is the only method so far proven to increase maximum lifespan in mammals. The original dietary restriction experiments conducted in the 1930s resulted in animals that remained prepubertal as a result of their retarded growth and development.[16,17] In more recent food restriction experiments, rodents have typically been fed a diet corresponding to approximately 60 per cent of the food ingested by *ad libitum* fed controls, commencing either soon after weaning or in young adulthood. Under these circumstances, besides a increase in maximum lifespan the development of tumours and other chronic diseases of late adulthood was slowed.

DNA microarray studies into the effects of **calorie restriction (CR)** in mice have indicated a shift in transcriptional patterns towards increased protein turnover and decreased macromolecular damage. Rats maintained on a restricted diet (representing 60 per cent of the control diet) for 36 weeks showed increased transcription of muscle genes involved in ROS scavenging, tissue development, and energy metabolism, with decreased expression of genes involved in signal transduction, stress response, and structural and contractile proteins.[18] CR also was able to maintain ATP production but at the same time reduce age-dependent endogenous oxidative damage.[19]

Preliminary studies conducted on rhesus macaque monkeys aged approximately 20 years and maintained on a reduced caloric intake for 9–10 years have suggested that long-term CR produced beneficial alterations in glycogen metabolism and mitigated the development of insulin resistance in older animals. Although restricted numbers of primates have been studied and few have attained extreme old ages, a wide range of potentially beneficial outcomes have been reported, in particular an improvement in glucose tolerance, a lower core body temperature, an attenuated decline in dehydroepiandrosterone (DHEA) sulphate levels, decreased triglycerides and increased HDL2b, in combination with lower weight, lean body mass and fat, and lower energy expenditure (reviewed in Bittles 2008).

A number of small-scale human studies have been reported, including a 6-day investigation on eight adults and eight pubertal children involving a 50 per cent caloric reduction that resulted in a significant reduction in the nitrogen balance of both adults and children and a decrease in their insulin-like growth factor-1 (IGF-1) levels. Individuals who had voluntarily adopted a restricted food intake for 6 years displayed a wide variety of physiological, metabolic, and biochemical changes, all of which would be consistent with protection against atherosclerosis,[20] and CR alone or with accompanying exercise regimes significantly increased the numbers of mitochondria in skeletal muscle cells while decreasing both energy expenditure and the frequency of mtDNA damage in overweight healthy adults.[21]

Ageing and the concept of healthy life expectancy

Active Life Expectancy (ALE) is defined as the period of life free of disabilities which interfere with basic Activities of Daily Living (ADL), e.g. eating, getting in and out of bed, bathing and toiletry needs, dressing, and indoor mobility. The concept of healthy life expectancy has been extended by weighting specific physical and

cognitive dysfunctions to measure Disability-Adjusted Life Years (DALY) and Quality-Adjusted Life Years (QALY). **Disability Adjusted Life Expectancy (DALE)** is now widely used in epidemiological studies to estimate the number of years that might be expected to be spent in 'full health'. A common finding in developed countries was that although females enjoyed higher DALE scores they also could expect more years of disability at advanced ages. Although measures such as DALE, DALY, and QALY have been criticized on methodological grounds, given global increases in the numbers of elderly individuals, the concept of healthy life expectancy may be increasingly useful in identifying the health and support needs of the aged.

Discussion

As in other areas of medical science, the Human Genome Project has impacted strongly on research into biological aspects of ageing, and DNA analysis now offers major insights into the development of the ageing phenotype. During the last decade, DNA microarray studies have been adopted to investigate changes in gene expression that accompany ageing.

Initial studies on rodent tissue showed differential gene expression patterns with advancing age, indicative of a marked stress response and lower expression of metabolic and biosynthetic genes. An 'ageing transcriptome' conserved across mammalian species has been identified, comprising deregulation of mitosis, cell adhesion, transport, signal transduction, mitochondrial function, and inflammatory response, and accompanied by a reduction in processes dependent on energy metabolism and mitochondrial function.[22] Subsequent analysis of human tissue based on large-scale DNA microarrays has revealed diverse patterns of both increased and decreased gene expression. However, a study of ~32 000 muscle tissue genes obtained from volunteers aged 16 to 89 years confirmed the existence of a common ageing signature, with altered levels of expression in 250 age-regulated genes and three genetic pathways that correlated both with chronological and physiological age.[23]

Research on adult stem cells may provide key future insights into ageing. Adult stem cells mainly undergo chronological ageing, as in skeletal muscle, or exhibit a combination of chronological and replicative ageing, typified by haematopoietic stem cells.[24] What remains to be determined is whether the overall decline in tissue regenerative capacity with advancing age is caused by intrinsic ageing of stem cells, or is due to increasing impairment of stem cell function in an aged tissue environment. Until this basic question is resolved, the prospect of stem cell therapy as a potential 'treatment' to correct the functional declines and degenerative diseases typical of human ageing will remain a theoretical possibility.

Further information

Agren, G. and Berennson, K. (eds.) (2007). *Healthy ageing—a challenge for Europe*. Swedish National Institute of Public Health, Stockholm; ISBN 91 7257 481 X.

http://www.healthyageing.nu.

Bittles, A.H. (2008). The biology of human ageing. In *Psychiatry in the elderly* (4th edn) (eds. R. Jacoby, C. Oppenheimer, T. Dening, and A. Thomas). Oxford University Press, Oxford, in press.

Johnson, M., Coleman, P.G., and Bengtson, V.L. (eds.) (2007). *Cambridge handbook of age and ageing*. Cambridge University Press, Cambridge; ISBN 052 182 6322.

National Institute of Aging (NIA); http://www.grc.nia.nih.gov/

Baltimore Longitudinal Study of Aging.

NIA Intramural Research Program 2006 Factbook.

References

1. Herskind, A.M., McGue, M., Holm, N.V., *et al.* (1996). The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870–1900. *Human Genetics*, **97**, 319–23.
2. Ljungquist, B., Berg, S., Lanke, J., *et al.* (1998). The effect of genetic factors for longevity: a comparison of identical and fraternal twins in the Swedish twin registry. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, **53**, M441–6.
3. Willcox, B.J., Willcox, D.C., He, Q., *et al.* (2006). Siblings of Okinawan centenarians share lifelong mortality advantages. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, **61**, 345–54.
4. Robine, J.M. and Allard, M. (1998). The oldest old. *Science*, **279**, 1834–5.
5. Seeman, T.E., McEwen, B.S., Rowe, J.W., *et al.* (2001). Allostatic load as marker of cumulative biological risk: MacArthur studies of successful aging. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4770–5.
6. Hughes, K.A., Alipaz, J.A., Drnevich, J.M., *et al.* (2002). A test of evolutionary theories of aging. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 14286–91.
7. Wood, R.D., Mitchell, M., Sgouros, J., *et al.* (2001). Human DNA repair genes. *Science*, **291**, 1284–9.
8. Holliday, R. (1987). The inheritance of epigenetic defects. *Science*, **238**, 163–70.
9. Kujoth, G.C., Bradshaw, P.C., Haroon, S., *et al.* (2007). The role of mitochondrial DNA mutations in mammalian aging. *PLoS Genetics*, **3**, 0161–73.
10. Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature* (*London*), **345**, 458–60.
11. Graakjaer, J., Londono-Vallejo, J.A., Christensen, K., *et al.* (2006). The pattern of chromosome-specific variations in telomere length in humans shows signs of heritability and is maintained through life. *Annals of the New York Academy of Sciences*, **1067**, 311–16.
12. Harley, C.B. (2005). Telomerase therapeutics for degenerative diseases. *Current Molecular Medicine*, **5**, 205–11.
13. Harman, D. (1992). Free radical theory of aging. *Mutation Research*, **275**, 257–66.
14. Schriner, S.E., Linford, N.J., Martin, G.M., *et al.* (2005). Extension of murine life span by overexpression of catalyse targeted to mitochondria. *Science*, **308**, 1909–11.
15. Kirkwood, T.B.L. (1977). Evolution of ageing. *Nature*, **270**, 301–4.
16. McCay, C.M., Crowell, M.F., and Maynard, L.A. (1935). The effect of retarded growth upon the length of lifespan and upon the ultimate body size. *The Journal of Nutrition*, **10**, 63–79.
17. McCay, C.M., Ellis, G.H., Barnes, L.L., *et al.* (1939). Clinical and pathological changes in aging and after retarded growth. *The Journal of Nutrition*, **18**, 15–25.
18. Sreekumar, R., Unnikrishnan, J., Fu, A., *et al.* (2002). Effects of caloric restriction on mitochondrial function and gene transcripts in rat muscle. *American Journal of Physiology. Endocrinology and Metabolism*, **283**, E38–43.
19. Lopez-Lluch, G., Hunt, N., Jones, B., *et al.* (2006). Calorie restriction induces mitochondrial biogenesis and bioenergetic efficiency. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 1768–73.
20. Fontana, L., Meyer, T.E., Klein, S., *et al.* (2004). Long-term calorie restriction is highly effective in reducing the risk for atherosclerosis in

humans. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 6659–63.
21. Civitarese, A.E., Carling, S., Heilbronn, L.K., *et al.* (2007). Calorie restriction increases muscle mitochondrial biogenesis in healthy humans. *PLoS Medicine*, **4**, 0485–94.
22. Wennmalm, K., Wahlestedt, C., and Larsson, O. (2005). The expression signature of *in vitro* senescence resembles mouse but not human aging. *Genome Biology*, **6**, R109.
23. Zahn, J.M., Sonu, R., Vogel, H., *et al.* (2006). Transcriptional profiling of aging in human muscle reveals a common aging signature. *PLoS Genetics*, **2**, e115.
24. Rando, T.A. (2006). Stem cells, ageing and the quest for immortality. *Nature Genetics*, **441**, 1080–6.

8.2

Sociology of normal ageing

Sarah Harper

Introduction

Research on the sociology of normal ageing has focused on understanding the paradigms of '*successful ageing*'. In an apparent reaction to '*disengagement theory*'[1] which proposed that to withdraw from roles and relationships in old age was normal, a new conceptual framework was developed in the late 1960s and 1970s which attempted to explain how individuals adapted to the constraints of ageing and old age. This has been variously measured in terms of good health, high levels of physical and mental functioning, and active engagement with one's social and physical environment. While post-modernism and critical gerontology have attempted to refocus the debate, the emphasis of most research and writing has remained within the framework of understanding, explaining, and even facilitating, 'success' in old age.

There is also a body of research which recognizes the importance of the *life course perspective*, and that throughout an individual's life, he or she is faced with *continuities* and *discontinuities* which have to be negotiated and resolved. Old age is but part of this lifelong process. Changes which occur in later life, such as retirement and widowhood, will lead to discontinuities in roles and relationships, other aspects of our lives will undergo little change allowing continuity. Alongside this, perspectives from anthropology, history and the social constructionist school of thought have also been recently influential.

This chapter will discuss concepts of age, generation, and cohort. It will consider the contribution of the life course approach to understanding ageing, and the manner in which other perspectives, such as social constructionism, narrative psychology and anthropology, have contributed to the sociology of normal ageing.

Structuring the life course through age

According to Hazelrigg,[2] the concept of age introduces signposts which link memory and anticipation, an iteratively remembered past and an iteratively expected future. Age classification is thus integral to normal organization of consciousness. As Mead's extensive work on life history, reminiscence and autobiography informs us,

> *one interacts retrospectively with one's younger selves, recalling earlier states of selfhood in the productive functioning of memory, and interacts prospectively with ones' older selves, anticipating conditions, actions, goal realizations and the like of late states of selfhood.*[1]

For both the individual and society, age conveniently dissects the life course into more manageable components. As a capitalist, industrial system emerged, and individuals moved from domestic units to bureaucratically organized corporations, so age was used to define adulthood and thus labour force participation. Age became the basis for regulating a large population. It defined the responsibilities of citizenship, and for each age related transition there is a stage of preparation, a stage of participation, and a stage of retirement.

Various anthropological studies[2] have highlighted alternative ways in which the life course might be structured. One of the most influential anthropological studies on the sociology of ageing was Cowgill and Holmes[3] work on ageing and modernization, which argued that the marginalization of older people was directly linked to modernization. While extensively debated ever since, this work highlighted the importance and complexity of cultural diversity. The burgeoning of anthropological studies around the concept of age and ageing since the Cwgill and Holmes study have contributed significantly to our understanding of this diversity.

Neither the !Kung nor Herero, hunter-gather and Bantu pastoralist peoples respectively of Botswana, have a concept of chronological age, marking age by physical transitions. Alternatively, the Tuareg, a semi-nomadic peoples in northern Niger, noted age by social transitions—courtship, marriage, childbirth, and grandchildren. Here, life transitions defining the ageing process are predominantly social rather than biological. A girl becomes a women not at menstruation, but at marriage; a women becomes an older women not at menopause but on having a child marry. For the Sukama of north-west Tanzania, ageing is defined through life course events. This emphasizes the social status of elderhood, measured by the wealth of alliances, offspring and livestock, which could not be diminished through ill health or loss of mental capacity. The Gussui of south-western Kenya have a similar notion of elderhood. However, they have adapted this traditional seniority gradation based on networks and affiliations to modern demands,

---

1 G.H. Mead, quoted in Hazelrigg p. 105[2]

2 See Harper for a full reference list to these studies[5]

incorporating such aspects as the role of entrepreneur to the criteria for achieving successful seniority status.

Modern Japanese society still applies a wide variety of terms to different points of the life course indicating complex relationships between chronological age and life transitions and physical appearance. For example, mid-life men and women with children whether or not they are married, will commonly be referred to as uncle and aunt, (*oji-san* and *oba-san*). Similarly, old men and women are frequently given the name of grandfather or grandmother, (*ojii-san* and *obaa san*) regardless of the presence of grandchildren, a characteristic also found in some European countries such as Greece. It is therefore clear that the domination of chronological age, has less salience in some other cultures.

Generation and cohort

Two further important concepts are generation and cohort. Individuals born within the same time period may be perceived as having a shared history and a common biography. The concept of *generation* is thus the link between an individual life course and the social changes that occur during the historical time of that life course. A generation may thus be thought of as *embodied history*.

Many of these draw on ideas from Mannheim who explored the creation of society through the continuous emergence of new age groups or generations. He argued that if social processes were always carried on and developed by the same individuals then once established, any fundamental social pattern, attitude or intellectual trend would probably be perpetuated. Culture was thus developed by individuals who come into contact anew with the accumulated heritage, that is the role of generations and while the continuous emergence of new individuals results in some loss of accumulated possessions it facilitates re-evaluation of our inventory and teaches us both to forget that which is no longer useful and to covet that which has yet to be won.

The problem for quantitative social scientists is how to disentangle those factors pertaining to the individual life course from those emerging from the historical context. It is here that the concept of cohorts, and cohort analysis has been refined by some to form a more analytical tool in the understanding of age and generational change.

A cohort begins with a particular demography at birth, that is its sex, race and economic composition. Differential mortality may lead to a higher proportion of some sub-groups surviving to old age; social mobility may lead to changes in cohort social status composition; and different historical periods will allow or enhance differential migration in and out of specific cohorts. A more sophisticated analysis places cohorts within specific historical contexts.[4] The *life-stage principle* suggests that disruptive social changes have enduring consequences on the subsequent lives, a particularly marked effect on those vulnerable at the time of occurrence.

Life course perspective

The life course perspective views old age as part of a life-long process of continuity and change. These can be addressed within four main frameworks: *context, transitions, roles*, and *relationships*.

Context

A starting point for life course analysis is the acknowledgement of the historical *context* within which different cohorts experience different aspects of the life course to life course perspective. As Harper[5] explores, while, most older men experienced a long period of economic activity followed by abrupt retirement, many older Western women experienced their younger lives within a framework of primary domestic duties, supplemented by intermittent economic activity. As a result, most older women replaced low earning capacity or economic dependence in younger life, with low incomes in old age. Cohorts in mid life, however, have had very different social and economic frameworks within which to live out their lives. Half the labour force in many countries is now female and full-time economic employment, with or without domestic, in particular childcare responsibilities, is becoming a widespread experience for many women. Despite this, there are still considerable income disparities in earning capacity of mid-life men and women. However, it is likely that future cohorts of older women will have higher incomes relative to older cohorts, and a lower gender income disparity.

Transitions

The processes which occur within these contexts can be understood as a series of life transitions.[6] Key transitions associated with later life are the end of active parenting, grandparenthood, widowhood and retirement. Each of these phases of life which may overlap, may be understood in relation to prior phases, and are mediated by other variables such as gender, class, and race. The transition to grandparenthood, for example, is experienced very differently by men and women, while the end of active parenting and transition to parent of a non-dependent child, the so-called *empty nest syndrome*, is mediated both by gender and by the experience of active parenting itself.

The transition to *widowhood* is one of the most stressful events of later life, with a high prevalence of depression both immediately before (presumably due to anticipation of the event and/or associated care giving) and in the first year following bereavement. Widowhood is likely to lead to lower income but higher social contacts for women, while men maintain their income, but are more likely to lose social contacts, unless they remarry. Over half women over 65 are widowed, rising to four-fifths at 85. Only 17 per cent of men are widowed over 65, rising to 43 per cent by their late 80s. Nearly three-quarters of older men in the UK are married, compared with less than a half of older women. This is explained both by differential life expectancy and the tendency of men to remarry following divorce or widowhood.[7]

The transition to *Grandparenthood* is the current normal experience of old age. US data suggests that more than half the population aged over 55 are in four-generation families and three-quarters of this population are or can expect to be as grandparents, with a prediction that one-third of current grandparents will live to be great grandparents, and one-fifth of all women who reach 80 will spend some time in a five-generation family as great-great-grandmothers. A similar picture may be found in the UK with estimates that three-quarters of adults over 66 years of age are grandparents. The transition from parenthood to grandparenthood, and even great-grandparenthood, determines both an individual's self-identity and subsequent roles and functions as

grandparenthood. In addition, the experience of the relationship that the grandchild has with his or her grandparent earlier will partially determine the way they take on the role and relate to their own grandchildren later on in life. Other sociological theories have been applied to the study of grandparenthood. *Role theory* suggests that a successful transition to grandparenthood requires both some socialization to the role, and appropriate life course timing. *Social stress theory* is used to argue that stress associated with transition to grandparenthood is related to the number, type, and context of the transitions and moderated by gender, education, income, and race.[9]

For many individuals, especially men, the transition to *retirement* is abrupt. Although early retirement has increased in Europe over the past twenty five years, most men still retire from full time work in their early-to mid-60s. A successful transition to retirement requires securing both financial security and personal adjustment. Atchley[10] identifies several phases of retirement. Pre-retirement, which may include a combination of both negative and positive feelings towards the impending event; a honeymoon period immediately following the event, which may extend for several years depending on the adjustment and resources, social, financial, and personal, available to the individual; disenchantment and reorientation; and eventually (if successful) stability. This latter stage occurs with the development of a well established set of criteria for making choices and dealing with the challenges and opportunities of this new life phase.

Roles and relationships

The above transitions—retirement, widowhood, grandparenthood - are also, of course, phases of life with specific roles and relationships. Thus, the transition of retirement also has an associated phase of being retired; that of grandparenthood of being a grandparent; that of widowhood of being a widow or widower.

We can examine two of these—late-life parenting and grandparenting—in the context of negotiating transitions and continuities in family roles and relationships as individuals age.[11] Intergenerational solidarity—shared values, normative obligations and enduring ties—and intergenerational conflict—whereby issues are resolved and relationships move on have been long seen as important components of this. More recently the concept of intergenerational ambivalence has been introduced. This, it is argued, reflects the contradictions which occur with ageing within family relationships These arise both through the desire of parents and children for both help and freedom, and conflicting norms regarding family relationships especially around the issue of care giving.

(a) Late-life parenting

Increasing longevity also means that most parent-child relationships will be lived out as predominantly non-dependent adult dyads, this is despite the delaying of child birth. The common experience for many parents and children is around 60 years of joint life, of which under one-third is spent in the traditional parent/dependent-child relationship. Around one-quarter of UK women and nearly 40 per cent of US women aged 55–63 still have a surviving parent. These women have thus spent around 60 years a child, some 40 of them in an adult relationship with a living parent. This relies on *re-bonding* in adulthood sometimes also referred to as 'reverse bonding'.[12] Under such experiences we see a loosening of the association between marital and parental roles. As the common experience of parenthood moves to more than 50 years of shared life, parents and children are adjusting to spending most of their relationship as independent adults. Similarly, husbands and wives are spending fewer of their joint lives as parents of young children. Relationships which have been historically based on a hierarchy which existed in part to support successful reproduction must move to greater equality, both child-parent, and husband-wife, as traditional roles based on parenthood give way to companionate relationships.[12]

(b) Grandparenting roles[3]

Currently, women can expect to become grandmothers in their 50s and 60s due to the early first age of births in the 1960s and 1970s. In addition, grandparental roles are lasting far longer due to increased longevity, the grandparent is thus more likely to be able to build a relationship with their grandchild into their adulthood. As a result many grandmothers, in particular, now face simultaneous demands as children of frail and dependent parents, mothers and grandmothers, as well as still being in full or part time economic employment.[13]

Grandmothers, in particular maternal grandmothers, are repeatedly attributed with having more influence in almost every value domain over their grandchildren than grandfathers. Research into the role of grandfathers has been limited.[14] However, it has been proposed that men become more nurturing as they get older and it could be hypothesized that these qualities might be expressed in relationships with their grandchildren. Similarly, the need to consider grandfathers as important resources for teenage mothers who are rearing their children, has been stressed. Harper also found that grandfathers could act as replacement partners and replacement fathers in female single parent households.[13]

Various roles of grandparenthood have been identified.[15] Bengtson (1985) for example, identifies five separate symbolic functions of grandparents: being there; grandparents as national guard; family watchdog; arbiters who perform negotiations between (family) members; and participants in the social construction of family history. Harper's study of grandmothers identifies grandmother as carer, replacement partner (confidante, guide and facilitator), replacement parent (listener, teacher and disciplinarian), and as family anchor (transferring values, attitudes and history).[13]

Complementary perspectives

Our understanding of ageing from a life course perspective has drawn on valuable insights from narrative gerontology and the social constructionist perspective.

Narrative gerontology

Narrative gerontology's main contribution to the field of ageing has been to the role of the life story in the development of theoretical and empirical approaches to ageing. It presumes that individuals think and act on the basis of stories, which have an external structure and an internal reality. Individuals retell their stories as they progress through their lives to make sense of their lives, and from a sociological perspective this retelling when carried out publicly provides considerable insights into various aspects of an individual's experience of ageing across the life course. There are

[3] A full reference list on grandparenthood may be found in Harper.[5]

four dimensions to this: the structural story, which reflects the wider societal context inhabited across the life course; the sociocultural story, which reflects other identities such as ethnicity and gender; the situational story which reflects an individual's roles and relationships; and the interpersonal story, the meanings which the individuals place themselves on their life story.

Writers in this genre of theory have further suggested that individuals experience two types of time. Achenbaum[16] describes this as a *physical outer time* and *a psychological inner time*, or as Kenyon and Randall[17] state *clock time* and *story time*. There may be a tension between the two types of time described as *on time* and *off time*.[18] Hazelrigg[19] suggests that the tension between the two times has been become more extreme within modern society, so dominated by rigid timekeeping. He argues that modern life is lived in two separate registers. On the one hand, most of a life experience is formed directly and indirectly in a highly standardized sequence of institutionalized events—schooling, work, parenting, retirement. These events are regulated by procedural rules and recognized routines, with predictable durations and regulated transitions between events. On the other hand, those aspects of life experiences that are not institutionalized and structurally stabilized in recognized life course sequences tend to have little or no connection to status dimensions or specific locations in the life course. These would include self-image, personal satisfaction, existential aesthetics etc.

Tensions arise when the two registers fail to coincide—*off time*.[18] Examples include middle-aged couples falling in love and publicly exhibiting displays of physical affection and romance, or older people adopting student style lives. Off time may also include the experience of being externally forced, through illness for example, to fall outside the normal behaviour range as defined for one's age. This also includes examples where society takes an individual and places them within a situation which is unusual for their chronological age–for example, a very young person being rapidly promoted within an institution which is very highly age regulated, such as a university, and taking on a professorial mantle.

Social constructionist approaches

The sociology of ageing has also recently enveloped social constructionist approaches, including phenomenology, symbolic interactionism, and ethnomethodology. These approaches share a subjective orientation to social reality, focusing on describing how individuals negotiate their worlds, rather than trying to explain why. A key area of interest here is the management of identity across the life course and in particular as we age. Examples include, Matthew's[20] seminal work using symbolic interactionist approaches to explore how old women negotiate their own identities when they are continually deluged with negative public stereotypes of infirmity and worthlessness. This is a theme taken up later by Featherstone and Hepworth[21] in their work on the Masks of Ageing. Similarly, Karp[22] uses symbolic interactionist approaches to explore the impact of social messages on the emerging consciousness of men and women in their 50s of their own ageing.

Conclusion

The sociology of, so-called, normal ageing thus combines perspectives from diverse traditions of thought. This actual measurement of ageing, in terms of health, physical and mental functioning, and active engagement with one's social and physical environment, has combined quantitative social science, epidemiology and political economy. The understanding of these interactions, draws on perspectives from sociology, psychology, and anthropology. In both cases, however, there is now recognition of the importance of context and process, and that, in reality, there is perhaps no single construct of 'normal' ageing. Given this perspective, it is essential that care for older adults is person-centred rather than imposed by professionals.

Further information

Johnson, M., Bengtson, V., Coleman, P. *et al.* (eds) (2005) *The Cambridge Handbook of Age and Ageing*. Cambridge University Press, New York.

Ageing Horizons: http://www.ageing.ox.ac.uk/ageinghorizons/index.htm

Binstock, R. and George, L. (2005). *Handbook of Ageing and the Social Sciences*. Elsevier, London.

References

1. Cumming, C., and Henry, W. (1961). *Growing Old*, New York Basic Books.
2. Hazelrigg, L. (1997). On the importance of Age. In *Studying aging and social change: conceptual and methodological issues* (ed. M. A. Hardy). pp. 93–128. Sage Publications, Thousand Oaks.
3. Cowgill, D. and Holmes, L. (eds.) (1972). *Ageing and Modernization*, New York, Apple Century Crofts.
4. Elder, G. H., and O'rand, A. M. (1995). Adult lives in a changing society. In *Sociological perspectives on social psychology* (eds. K. S.Cook, G. A. Fine, & J. S.House), pp. 452–475. Allyn and Bacon, Boston.
5. Harper, S. (2006). *Ageing Societies*. Hodder Arnold, London.
6. Elder, G. H. (1985). Perspectives on the life course. In *Life course dynamics: trajectories and transitions* (ed. G. H. Elder) *1968–1980*. pp.23–49. Cornell University Press, Ithaca.
7. Arber, S. Davidson, K. and Ginn, J. (eds.) (2003). *Gender and Ageing: changing Roles and Relationships*, Maidenhead, Open University Press.
8. Kornhaber, A. (1996). *Contemporary Grandparenting*, New York, Sage.
9. Szinovacz, M. (1997). Grandparents today: a demographic profile. *The Gerontologist,* **38**, 37–52.
10. Atchley, R. (1989). A continuity theory of normal aging, *The Gerontologist,* **29**, 183–90.
11. Giarrusso, R., Silverstein, M., Gans, D., *et al.* (2005). Ageing parents and adult children: new perspectives on intergenerational relationships. In *The Cambridge handbook of age and ageing* (eds. M.L. Johnson, and V. L. Bengston). Cambridge University Press, New York.
12. Harper, S. (2006). *Ageing Societies*. Hodder Arnold, London.
13. Harper, S. (2005). Understanding Grandparenthood. In *The Cambridge handbook of age and ageing* (eds. M.L. Johnson, and V. L. Bengston), Cambridge University Press, New York.
14. Mann, R. (2007). 'Out of the shadows?: Grandfatherhood, age and masculinities', *Journal of Aging Studies*, **21**(4), 271–81.
15. Bengtson, V. (1985) Diversity and symbolism in grandparental roles. In *Grandparenthood* (eds. V. Bengtson, and J. Robertson) Beverly Hills, Sage.
16. Achenbaum, W. A. (1991) Time is the messenger of the Gods: a gerontological metaphor. In *Metaphors of Aging in Science and Humanities* (eds. *G.* Kenyon, J. Birren, and J. Schroots), pp. 83–101. New York, Springer.

17. Keynon, G., and Randall, W., (1997). *Restorying our Lives: personal growth through autobiographical reflection*. Westport, CT, Praeger.
18. Hagestad, G. O. (1986). Dimension of time and the family. *American Behavioral Scientist*, **29**, 679–94.
19. Hazelrigg, L. (1997). On the importance of Age. In *Studying aging and social change: conceptual and methodological issues* (ed. M. A. Hardy), pp. 93–128. Sage Publications, Thousand Oaks.
20. Matthews, S. (1979). *The Social World of Older Women*. Newbury park, Sage.
21. Featherstone, M., and Hepworth, M., (eds.) (1995). *Images of Ageing*. Sage, London.
22. Karp, D. (1988). A decade of reminders: changing age consciousness between fifty and sixty years old, *The Gerontologist*, 727–38.

8.3

The ageing population and the epidemiology of mental disorders among the elderly

Scott Henderson and Laura Fratiglioni

In the last decades the ageing of the populations has become a worldwide phenomenon.[1] In 1990, 26 nations had more than 2 million elderly citizens aged 65 years and older, and the projections indicate that an additional 34 countries will join the list by 2030. In 2000, the number of old persons (65+ years) in the world was estimated to be 420 million and it was projected to be nearly 1 billion by 2030, with the proportion of old persons increasing from 7 to 12 per cent.[2] The largest increase in absolute numbers of old persons will occur in developing countries; it almost triples from 249 million in 2000 to an estimated 690 million in 2030. The developing regions' share of the worldwide ageing population will increase from 59 to 71 per cent. Developed countries, which have already seen a dramatic increase in people over 65 years of age, will experience a progressive ageing of the elderly population itself (see Fig. 8.3.1). The global trend in the phenomenon of population ageing has dramatic consequences for public health, health care financing, and delivery systems in the whole world. The absolute number of chronic diseases as well as psychiatric disorders is expected to increase. In this chapter, the epidemiological aspects of the most common psychiatric disorders of the elderly are summarized and discussed.

Depressive disorders

The epidemiology of depression in the elderly can be approached at three levels: its occurrence in the elderly living in the community, in those reaching primary care, and in the residents of hostels and nursing homes.

The community

It might be expected that, overall, the prevalence of depressive symptoms and disorders might increase in old age due to the loss of partners, friends, social status, retirement, income, and, above all, declining health. It is surprising, therefore, that surveys of the elderly in the general population have recurrently found rates that are significantly lower than in younger adults. Many of the large national surveys have not included persons aged over 65 years, but two exceptions are Australia, which found a 12-month prevalence of 1.7 per cent for the 65 years and over group compared with 5.8 per cent for all adults; and the New Zealand survey with rates of 2.0 and 8.0 per cent, respectively. It must be emphasized that these data refer to depressive symptoms in the elderly living in the community.

What is so far unproven is that such findings are indeed valid, and if they are, what might explain them.[3] They could be due to sample bias, in which elderly respondents with depressive symptoms may be more likely to decline to be interviewed than younger depressed people. Selective mortality has also been proposed, but cannot account for the size of the difference. It could be due to an error in case ascertainment, by which the interview instrument is not equally valid across age groups. For example, questions about depressive symptoms may be responded to differently by persons aged 20 and 80 years. Another possibility is a cohort effect in much of the Western world, where people born in the second half of the twentieth century have higher rates for depression.[4,5] This seems increasingly likely and may be due to a combination of social and environmental factors.

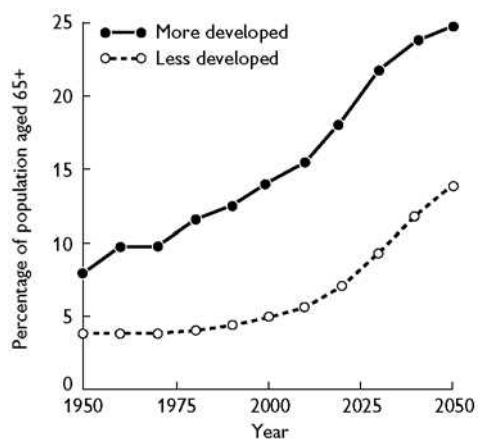


Fig. 8.3.1 Percentage of the population aged 65+ for more developed and less developed countries. (Reproduced from Kinsella, K. and Velkroff, V.A. The demographics of aging. *Aging Clinical and Experimental Research*, **4**, 59–69, copyright 2002 with permission from Editrice Kurtis S.r.l.)

Primary care

Unsurprisingly, the prevalence of depressive symptoms is considerably higher in elderly persons consulting their doctor than in the general community. One study in London found a point prevalence of about 30 per cent. Where it has been possible to compare the rates for those cases recognized by their doctor with cases independently ascertained by a research measure, such as a screening instrument or standardized interview, a typical finding is that the general practitioner recognizes about two thirds of the mild cases, rising to some 90 per cent of the moderate to severe ones. Some cases are considered to be depressed when they are not. Another finding is that elderly persons with depressive symptoms may not mention them to their doctor, attributing them to their age and circumstances. These findings have led to programmes offering additional training for GPs and to recommending the use of brief screening tests in primary care. Because diagnosis would lead to appropriate treatment being given earlier in an episode of depressive disorder, its duration would be shortened. Since prevalence is the product of incidence and duration, the prevalence of depression would therefore be expected to fall. This is an example of the application of epidemiology to prevention, which is its ultimate service.

Depression in hostels and nursing homes

Prevalence rates are also higher in hostels and nursing homes. In the United States, levels as high as 30 to 50 per cent have been reported. It might be thought that the context of living in a nursing home would account for having depressive symptoms. But one study found that the excess over the general population rates was largely accounted for by medical disorders, environmental factors contributing little to the variance. This needs to be studied further because it seems counterintuitive that the social and physical environment, both of which can be modified, could be of little relevance.

What is important is that only about one quarter of cases are recognized. To compound the situation, those cases that are recognized tend to be treated with too low doses of antidepressants. Depressive symptoms are well known to occur comorbidly with cognitive decline and the dementias. These findings from clinical epidemiology have pointed to the need for better case recognition through education of medical and nursing staff, and to the use of routine screening of residents in such settings.

Suicide

For many decades across the world, the traditional pattern has been for the highest rates of suicide to be in elderly men. This has now changed. In over a third of countries, both developed and less developed, it is younger people who have come to carry the highest rates. The World Health Organization provides a valuable resource for such data, showing rates by age and gender for nearly all countries from 1950 onwards.[6] The pattern varies considerably between countries. For example, in the United Kingdom in 2002, men aged 75 and over had a rate of 10 per 100 000, whereas the highest rate was 18 per 100 000, in men aged 35–44 years. In the United States and in the Russian Federation, the rates for men aged 75 and over were 41 and 89 per 100 000, respectively.

The main risk factors for suicide in the elderly are a past history of an attempt, depressive disorder, physical illness or disability, chronic pain, recent losses, social isolation, and access to lethal means. While universal interventions are more powerful than selected factors in prevention,[7] these attributes can be used in selective intervention to identify groups at increased risk. Furthermore, being multiplicative, these markers are of great value in individual cases by alerting the clinician to a person needing particular attention. Here is another example of the use of epidemiology for prevention. A systematic review of suicide prevention strategies for all age groups concluded that two interventions did reduce rates: physician education in recognizing and treating depression; and restricting access to lethal means.[8] Both of these interventions have close relevance to the elderly.

Personality disorders

The subject matter here refers to older people who have enduring attitudes and behaviour that bring difficulties for themselves or for others.[9] There is only sparse information on the prevalence of personality disorders in the general population, let alone specifically in the elderly. One exception, based on a national survey of mental health, found a lifetime prevalence of 6.5 per cent across all age groups with a trend towards lower rates with increasing age.

In clinical practice, it has long been suggested that traits such as impulsivity and externalizing behaviours tend to become less frequent in later life, whereas anxiety-prone, dependent, schizoid, paranoid, or obsessional persons are likely to change little as they age, or to become more so. Bergmann's pioneering enquiries among the elderly of Newcastle upon Tyne found that it was the anxiety-prone and insecure types that had late-onset neurotic disorders. A more recent study of late-life depression found an overall prevalence of comorbid personality disorder of 10–30 per cent. The group formerly known as neurotic and more recently as Cluster C in the DSM classification, had the higher prevalence. The Cluster B group, those with borderline, narcissistic, histrionic, and antisocial traits, were rare. What is not yet established, however, is if this lower prevalence also exists in the general population of the elderly, not just among cases with depressive disorder who have reached treatment in specialist services.

The epidemiology of personality disorders in later life is therefore significant for two reasons. First, some types are associated with increased risk of anxiety, depression, or paranoid states (*vide infra*). Second, there remains much yet to understand about the natural history of the personality disorders across the lifespan.

Psychosis of late onset

For the functional psychoses of late life, epidemiological information comes from two sources: studies of persons who have reached psychiatric services; and surveys of elderly persons living in the general community.[10] Psychotic symptoms probably exist as a continuum of severity, with only the more developed cases meeting diagnostic criteria. These often, but not always, reach psychiatric services, not uncommonly through being brought to the attention of the police. States phenomenologically similar to those found in clinics do occur in the community in non-trivial numbers. For cases that reach the threshold for a diagnosis by virtue of the range and severity of symptoms and behaviour, it has been proposed that cases with onset after the age of 60 years be called 'very-late-onset schizophrenia-like psychosis'. The syndrome has a 1-year prevalence

of 0.1 to 0.5 per cent. For advancing knowledge about the aetiology of schizophrenia, any information on it might be useful in explaining why people with this syndrome have reached the seventh decade or later in life without becoming psychotic, and only then develop it. It is more common in women. This is unlikely to be due to different social visibility or access to services. It is associated with a better premorbid level of social and occupational functioning. Premorbid paranoid or schizoid traits have been implicated and both clinical and community-based studies have found an association with sensory impairment such as deafness or poor eyesight. Personal and environmental factors associated with ageing have been considered, such as physical ill health, bereavement, loss of friends, and loss of income, but these have not been shown to contribute significantly. Genetic factors appear to be less important than in earlier onset schizophrenia.

Alcohol and drug dependence

It is generally believed that the prevalence of alcohol abuse and dependence declines during adult life and that the elderly have low rates in most communities. This may well be the case, but some other factors have to be considered. Whatever the prevalence, the absolute numbers will rise in the future because of the unprecedented growth in the elderly population. Next, the assumption may be false. In community surveys, errors in the ascertainment of alcohol abuse may lead to an underestimate for older persons. Most screening instruments were developed for use on younger adults, so their validity in the elderly is largely undetermined. Measures of the quantity drunk may mislead because smaller amounts may have an intoxicating effect in persons whose body fat, lean tissue, cerebral reserve, and metabolic function have declined. So the usual cut-off for problem drinking may be set too high for the elderly. One review of screening instruments concluded that the CAGE and MAST-G scales were appropriate, whereas other widely used instruments were not.[11] Next, all the studies have been cross-sectional. The elderly may have lower rates because of a cohort effect, whereby people born in the first half of the twentieth century may have been more moderate drinkers for all their life, compared to the high levels of consumption that are now found in the young of both sexes.

The actual values for prevalence are dependent on the instrument used and the definition used to define problem drinking, alcohol abuse, or dependence.[12] One review of community studies gives a figure of 5.1 per cent using various definitions. Invariably, men have higher rates than women. There is also considerable variation between countries and across different cultures. In identifying cases, a distinction of clinical significance needs to be made between late-onset and long-standing alcohol abuse. In primary care, accident and emergency departments, hospital in-patients, and nursing homes, the prevalence is much higher, yet cases are consistently under-recognized. The use of screening instruments in all of these settings has been advocated to improve this.

Alcohol abuse carries important comorbidity. In addition to all the established medical complications, it is associated with falls, subclinical delirium, cognitive decline, and depression. One study demonstrated a five-fold increase in the risk of developing a psychiatric disorder, especially depression and dementia. Simultaneous use of benzodiazepines, itself common in older persons, is clearly an additional and important factor. Against all this, it should be recalled that moderate alcohol use has been found in population studies to be associated with better mental and cardiovascular health, as well as being subjectively enjoyable.

Alzheimer's disease and other dementias

In the last two decades the dementia field has registered a tremendous scientific progression in many research areas including aetiology, pathogenesis, clinical aspects, treatment, and prevention. These advances have opened new perspectives, especially concerning definitions and diagnostic criteria, which have a relevant impact on epidemiological research.

Dementia is still defined as a syndrome which includes memory deficits and disturbance of other higher cortical functions; these major symptoms are commonly accompanied, and occasionally preceded, by deterioration in emotional control, social behaviour or motivation. However, it has become apparent that memory impairment may not necessarily be the major or first symptom for dementia subtypes such as frontotemporal dementia (FTD) and vascular dementia (VaD). Furthermore, as the current definition requires impairment severe enough to interfere with daily functioning, in several cases a delay of the diagnosis occurs. For that reason, a new research line has emerged with the aim to detect early Alzheimer's disease (AD) and other dementias, and the terms mild cognitive impairment (MCI) and cognitive impairment no-dementia (CIND) have been proposed to identify those subjects that show a clear cognitive deficit but do not fulfil diagnostic criteria for dementia. Finally, it is well known that dementia syndrome can be induced by many different underlying diseases, and that a differential diagnosis may be difficult for several reasons. AD as well as other dementia subtypes shows heterogeneity with distinct clinical and pathological characteristics; many different dementing disorders overlap in clinical and pathologic features; and different dementing disorders may make a common contribution or interact in causing dementia symptoms. Thus, rather than viewing, for example, AD and VaD as dichotomous entities, it may be more relevant to consider the role of their additive or synergistic interactions in producing a dementia syndrome.[13,14]

Following these new perspectives, in this chapter we will summarize the major findings from the most recent epidemiological research according to three major topics: early detection of AD and other dementias, incidence and risk factors for AD and dementia, and prevalence and impact of the dementing disorders at the individual and societal levels.

Early detection

As diagnostic criteria for AD require gradual onset of cognitive deficits, it is expected that cognitive disturbances are present already before the diagnosis can be rendered. Cognitive deficits are observable up to 10 years before dementia diagnosis with a sharp decline more evident in the final 3 years,[15] and occurring in episodic memory as well as in other cognitive domains such as executive functioning, verbal ability, visuospatial skills, attention, and perceptual speed.[16] However, our capability to use such early disturbances as a predictive tool of incipient dementia is strongly limited by several concomitant facts: (1) cognitive decline is also present as a function of the normal ageing process; (2) several conditions other than AD may lead to cognitive disturbances in the elderly; and (3) dementia-free patients with cognitive impairment observed in specialized clinical settings are different from cognitively impaired persons detected in the general population.[17]

To overcome these difficulties, different definitions have been proposed, with MCI and CIND being the most commonly used. MCI definition was originally derived in a clinical setting to identify subjects with isolated memory loss (now referred to as the 'amnestic' type) who may be in a preclinical phase of AD. Since then, the view has widened to cover a broader range of cognitive disturbances, and other MCI subtypes have been proposed.[18,19] CIND is derived essentially from population-based studies, and operationalized in slightly different items. Unfortunately, not one of the proposed definitions has shown a sufficiently good predictivity at the community level. Even a highly selected algorithm including subjective memory complaints, and global, and specific (memory/language) cognitive deficits could identify only 18 per cent of the incipient AD cases.[20,21] Although elderly persons with cognitive impairment have a high risk of developing dementia with a rate of about 11 to 50 per cent over 1 to 5 years, not all persons with CIND or MCI develop dementia. A substantial proportion of these persons (24–42 per cent) even improve in their cognitive performances over time.[22] This diverse prognosis supports the notion that AD is not the only causal mechanism underlying cognitive impairment in the non-demented elderly population. Cognitive psychologists have detected lower cognitive performances in elderly persons with deficiency in vitamin B_{12} and folate, elevated homocysteine, thyroid stimulating hormone deficiency, and cardiovascular disease. Physicians found an association between CIND and a number of factors including frailty-related factors such as history of hip fracture and high consumption of multiple drugs, history of psychoses, and depressed mood occurring 3 years before CIND development.[23] Other studies have identified older age, low education, depression, APOE ε4 allele, medicated hypertension, mid-life elevated serum cholesterol, and high diastolic blood pressure, as well as diabetes and anticholinergic medication use as risk factors for MCI.[17]

The prevalence of cognitive impairment—no dementia varies depending on diagnostic criteria with a maximum of 30 per cent.[24] In the younger elderly (e.g. 65–75 years old) cognitive impairment is actually more frequent than dementia disorders.[25] Annual incidence rates vary from 15 per 1000 among persons aged 75–79 years to 98 per 1000 among nonagenarians, when the estimates are corrected for dropouts due to death (Fig. 8.3.2; Ref.[26]). Regardless of the aetiology underlying the cognitive deficits, the high prevalence and incidence highlight the importance of this syndrome in the ageing population. Given that the criteria for cognitive impairment in non-demented persons are still under construction, and that there is no efficacious treatment to stop the possible progression to AD, at the moment we must be cautious with diagnosing MCI, which may cause an unnecessary burden on patients and relatives due to the unclear prognosis. However, it is clinically relevant to identify all persons with cognitive deficits due to treatable conditions such as depression, low-level vitamin B_{12}, and use of drugs.

Incidence and risk factors

Dementia incidence is similar in all continents and from different regions of the world (Table 8.3.1). Slightly lower rates detected in US in comparison with those in Europe and Asia are likely to be due to differences in study designs and case ascertainment.[27] AD accounts for 60–70 per cent and VaD accounts for 15–20 per cent of all dementia cases. The incidence of both AD and dementia increases almost exponentially with age. However, there are inconsistent findings regarding whether the rates continue to increase even in more advanced ages. The apparent decline found in some studies may be an artifact of the poor response rates, survival effects, and nature of populations previously sampled in these very old age groups.[28]

Age is the strongest risk factor for dementia and AD, suggesting that ageing-related biological processes could be implicated in the etiopathogenesis of AD. Further, the strong association with increasing age can be, at least partially, explained by a lifetime cumulative risk to different risk factors. Using this approach, the risk of dementia in late life is considered as a result of complex

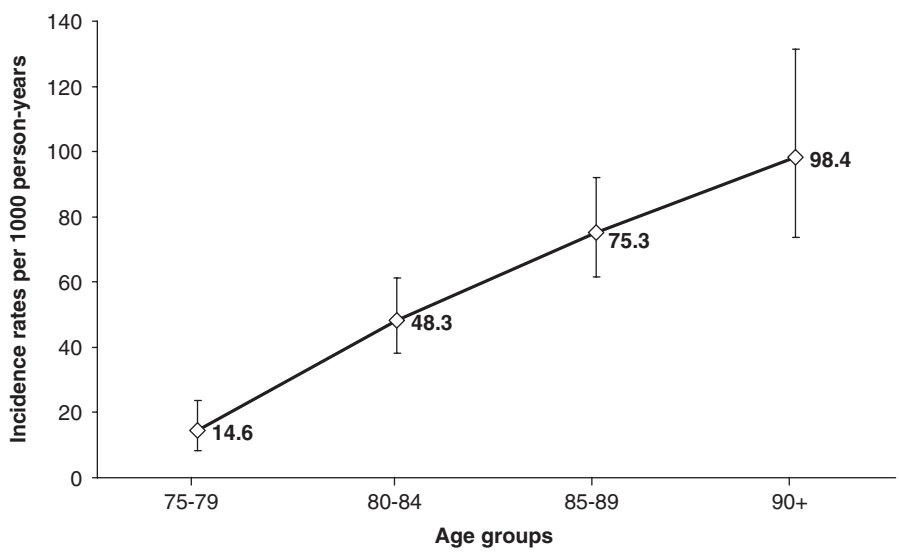


Fig. 8.3.2 Corrected age-specific incidence rates with 95 per cent confidence intervals of non-dementia cognitive impairment, including the two mutually exclusive definitions of amnestic MCI and CIND. (Reprinted from Alzheimer's and Dementia, 2, L., Fratiglioni, C. Qui, and K. Palmer, vascular cognitive impairment: time for prevention?, 202–4, copyright 2006, with permission from Elsevier.)

Table 8.3.1 Age-specific prevalence and incidence rates for dementia in the world and different regions: estimated from meta-analyses

Age groups	Incidence rate (per 1000 person-years)				Prevalence (per 100 population)			
	Worldwide (Gao *et al.* 1998)	Europe (Fratiglioni *et al.* 2000)	USA (Jorm and Jolley, 1998)	East Asia (Jorm and Jolley, 1998)	Worldwide (Jorm *et al.* 1987)	Worldwide (Fratiglioni *et al.* 1999)	Europe (Lobo *et al.* 2000)	China (Liu *et al.* 2003)*
60–64	1.1	—	—	—	0.7	0.9	—	0.3
65–69	3.3	2.4	2.4	3.5	1.4	1.6	0.8	0.7
70–74	8.4	5.5	5.0	7.1	2.8	3.5	3.0	1.3
75–79	18.2	16.0	10.5	14.7	5.6	6.9	5.8	2.8
80–84	33.6	30.5	17.7	32.6	11.1	13.0	12.0	5.6
85–89	53.3	48.6	27.5	72.1	23.6	25.2	17.4	11.8
90–95	72.9	70.2	—	—	—	35.8	28.5	23.7
95+	86.8	—	—	—	—	48.1	—	—

* The dementia cases in this meta-analysis include only Alzheimer's disease and vascular dementia.

(Reproduced from Backman, L., Small, B.J. and Fratiglioni, L. Cognitive deficits in preclinical Alzheimer's disease: current knowledge and future directions. In *New frontiers in cognitive aging* (eds. R.D. Dixon, L. Backman and L.G. Nilsson), pp. 161–77, copyright 2004, with permission from Oxford University Press.)

interactions of genetic susceptibility, biological factors, and environmental exposures experienced over the lifespan. A summary of all these factors is reported in Table 8.3.2, according to different biological mechanisms and grade of scientific evidence. As a part of an initiative of the Swedish Council on Technology Assessment in Health Care, specific criteria to summarize the scientific evidence concerning risk and protective factors for dementia have been proposed. Similar to criteria adopted for other diseases, these criteria first integrate the internal validity with basic causal criteria to weight the study quality, and then they take into account also number and proportion of the included studies reporting a specific association.[29] Moderate or strong evidence supports several genetic, vascular, and psychosocial factors as significantly related to both AD and dementia risk. Whereas implementing preventive strategies targeting the genetic susceptibility is limited, the other two hypotheses can easily lead to prevention programmes.

(a) Genetic hypothesis

First-degree relatives of AD patients have a higher lifetime risk of developing AD than the general population or relatives of non-demented subjects. Both genetic and environmental factors contribute to the phenomenon of familial aggregation. Twin studies have shown that heritability of AD is about 58 per cent, whereas other variance may be attributable to non-genetic factors.[30] The APOE ε4 allele is the only established susceptibility gene factor for both early- and late-onset AD. The risk effect of APOE ε4 allele decreases with increasing age, and after age 75 years, 15–20 per cent of AD cases are attributable to APOE genotype.[31] Familial aggregation of dementia and AD can be only partially explained by APOE polymorphism, implying that other genetic factors may be active and need to be detected.[32] Several other genes have been examined as possible candidates, but the reports are sporadic or the results are inconsistent.[33]

(b) Vascular hypothesis

Whereas the reported association between vascular risk factors and dementia risk is expected, due to vascular dementia, several explanations have been proposed for the association between vascular risk factors and Alzheimer-type dementia: (1) coexistence of vascular factors and AD pathology in the elderly; (2) precipitating effect of cerebrovascular disease or interactive effect between Alzheimer-type and vascular lesions in the brain; and (3) misclassification of mixed dementia as AD. Even if the mechanisms are still not fully understood, prevention may be possible as most vascular risk factors and diseases are modifiable or amenable to prevention and treatment. Controlling high blood pressure in middle age, avoiding mid-life obesity, and appropriately treating diabetes are the major intervention actions. Some studies also show that people who maintain tight control over their blood glucose levels tend to score better on tests of cognitive function than those with poorly controlled diabetes. Indeed, borderline diabetes or impaired glucose tolerance is also linked to an increased risk of dementia and AD in very old people.[34] Finally, to postpone clinical expression of the dementia syndrome in old people, preventing recurrent cerebrovascular disease as well as maintaining sufficient cerebral perfusion by adequately managing heart failure and avoiding very low blood pressure seems to be critical.

(c) Psychosocial hypothesis

Evidence from both epidemiological and biological studies indicates that factors acting at different periods across life course and having an intellectually stimulating nature may contribute in increasing the neural reserve and therefore promote functionally more efficient cognitive networks to cope with brain pathology and delay the onset of clinical manifestations of dementia. These factors include education, adult-life occupational work complexity as well as late-life social network and intellectually stimulating activities.[29,35] Although physical exercise may reduce the risk of brain damage due to atherosclerosis, the relevance of physical activity itself remains in debate, as most physical activities include also social and mental components. Complex leisure activities with physical, mental, and social components seem to have the most beneficial effect.[36] In addition to the reserve hypothesis, other mechanisms such as premorbid cognitive ability, vascular damage,

Table 8.3.2 Scientific evidence supporting risk and protective (in italic) factors of dementia and AD by different aetiological hypotheses.

	Risk and *protective* factors			
	Vascular	**Psychosocial**	**Genetic**	**Others**
Insufficient or limited scientific evidence	High cholesterol Cigarette smoking Obesity Late life high BP Late life low BP Heart failure Silent stroke *Moderate alcohol intake* *Dietary factors (e.g. fish and vegetables)*	Depression Low SES *Midlife physical activity* *Late life social network*	Several susceptibility genes	Head trauma Inflammatory markers *NSAIDs* *HRT* Folate and vitamins B12, A, E, and C deficiency Occupational exposure to toxics *Antioxidants*
Moderate or strong scientific evidence	Midlife high BP Diabetes mellitus Clinical stroke Atherosclerosis *Antihypertensive drugs*	Low education *Late life mentally stimulating activities* *Physical activity*	APOE ε4 allele* Familial aggregation	

Abbreviations: AD = Alzheimer's disease; APOE = apolipoprotein E gene; BP = blood pressure; HRT = hormone replacement therapy; NSAIDs = non-steroidal anti-inflammatory drugs; SES = socioeconomic status.

(Reproduced from Backman, L., Small, B.J. and Fratiglioni, L. Cognitive deficits in preclinical Alzheimer's disease: current knowledge and future directions. In *New frontiers in cognitive aging* (eds. R.D. Dixion, L. Backman and L.G. Nilsson), pp. 161–77, copyright 2004, with permission from Oxford University Press, New York)

neuroprotection, or detection bias may be possible explanations. The most likely effect of a mentally, physically, and socially active life is to postpone the onset of clinical dementia: even delaying dementia onset by 5 years would halve dementia prevalence and substantially decrease the number of dementia cases in the community.

Prevalence and impact

Despite different inclusion criteria, several meta-analyses of prevalence studies have resulted in strikingly similar results (Ref.[27]; Table 8.3.1). Currently, more than 24 million people in the world have dementia and this number will double in 20 years.[37]

The prognosis of dementia is dramatic. In 3 years, more than 50 per cent of the dementia cases progress to the severe stage. In the Kungsholmen Project, the proportion of severe dementia among prevalent cases increased from 19 per cent at baseline to 48 per cent after 3 years, and to 78 per cent after 7 years. This progression is due to both cognitive and functional decline. The mean annual rate of cognitive decline as measured with the MMSE varies from –4.0 to –2.0 points. Many predictors of a more rapid cognitive decline have been reported such as initial higher cognitive function, functional disability, and brain lesions. The APOE ε4 allele seems not to act as relevant prognostic factor.[38]

Dementia is strongly associated with disability being the major determinant of developing dependence and functional decline over 3 years. Approximately half of the persons who developed functional dependence in a 3-year period can be attributable to dementia.[39] In industrialized countries, mental disease and cognitive impairment are the most prevalent disorders among older adults living in nursing homes or other institutions. However, institutionalization of dementia patients varies depending on age structure, urban or rural residence, and other cultural aspects. In the 75+-year-old population, 70 per cent of incident dementia cases die during the 5 years following the diagnosis, accounting for a mortality rate specific for dementia of 2.4 per 100 person-years. Dementia triplicates the risk of death.[40]

Conclusion

Mental disorders are common chronic conditions among the elderly people, and the absolute number of subjects with psychiatric disorders will increase dramatically worldwide in the near future due to the ageing of the populations. In addition, the mental disorders have a high impact both at the individual and societal level. Prevention may represent one answer to these challenging conditions. The scientific advances of the last few years have provided sufficiently strong evidence supporting two possible preventative strategies: an active and stimulating lifestyle in late life as well as optimal control of other chronic disease both at middle and late age may decrease the risk of relevant psychiatric disorders such as AD and other dementias.

Further information

Hybels, C.F. and Blazer, D.G. (2003). Epidemiology of late-life mental disorders. *Clinics in Geriatric Medicine*, **19**, 663–96.

Blazer, D.G. and Hybels, C.F. (2005). Origins of depression in later life. *Psychological Medicine*, **35**, 1241–52.

Jorm, A.F. (2001). History of depression as a risk factor for dementia: an updated review. *The Australian and New Zealand Journal of Psychiatry*, **35**, 776–81.

Ferri, C.P., Prince, M., Brayne, C., *et al*. (2005). Global prevalence of dementia: a Delphi consensus study. *Lancet*, **366**, 2112–17.

Qiu, C., De Ronchi, D., and Fratiglioni, L. (2007). The epidemiology of the dementias: an update. *Current Opinion in Psychiatry*, **20**, 380–5.

Fratiglioni, L. and Wang, H.X. (2007). Brain reserve hypothesis in dementia. *Journal of Alzheimer's Disease*, **12**(1): 11–22.

Fratiglioni, L., von Strauss, E. and Qiu, C., (2007). Epidemiology of the dementias of old age. In *Oxford textbook of old age psychiatry* (eds. R. Jacoby, T. Dening, A. Thomas, and C. Oppenheimer), pp. 391–406. Oxford University Press, Oxford.

References

1. Kinsella, K. and Velkoff, V.A. (2002). The demographics of aging. *Aging Clinical and Experimental Research*, **4**, 59–69.

2. The US Centers for Disease Control and Prevention. (2003). Public health and aging: trends in aging—United States and worldwide. *The Journal of the American Medical Association*, **289**, 1371–3.
3. Henderson, A.S. (1994). Does ageing protect against depression? *Social Psychiatry and Psychiatric Epidemiology*, **29**, 107–9.
4. Klerman, G.L. and Weissman, M.M. (1989). Increasing rates of depression. *The Journal of the American Medical Association*, **261**, 2229–35.
5. Frombone, E. (1995). Depressive disorders: time trends and possible explanatory mechanisms. In *Psychosocial disorders in young people* (eds. M. Rutter and D.J. Smith). John Wiley & Sons, Chichester.
6. World Health Organization. (2007). Suicide prevention and special programmes. http://www.who.int/mental_health/prevention/suicide/country_reports/en/index.html.
7. Rose, G. (1993). Mental disorder and the strategies of prevention. *Psychological Medicine*, **23**, 553–5.
8. Mann, J.J., Apter, A., Bertolote, J., *et al.* (2004). Suicide prevention strategies: a systematic review. *The Journal of the American Medical Association*, **294**, 2064–74.
9. Burns, A., Bergmann, K., and Lindesay, J. (1998). Key papers in geriatric psychiatry. *International Journal of Geriatric Psychiatry*, **13**, 199–202.
10. Howard, R., Rabins, P.V., Seeman, M.V., *et al.* (2000). Late-onset schizophrenia and very-late-onset schizophrenia-like psychosis: an international consensus. The International Late-Onset Schizophrenia Group. *The American Journal of Psychiatry*, **157**, 172–8.
11. Beullens, J. and Aertgeerts, B. (2004). Screening for alcohol abuse and dependence in older people using DSM criteria: a review. *Aging and Mental Health*, **8**, 76–82.
12. Johnson, I. (2000). Alcohol problems in old age: a review of recent epidemiological research. *International Journal of Geriatric Psychiatry*, **15**, 575–81.
13. Agüero-Torres, H., Kivipelto, M., and von Strauss, E. (2006). Rethinking the dementia diagnoses in a population-based study: what is Alzheimer's disease and what is vascular dementia? A study from the Kungsholmen project. *Dementia and Geriatric Cognitive Disorders*, **22**, 244–9.
14. Fratiglioni, L., Qiu, C., and Palmer, K. (2006). Vascular cognitive impairment: time for prevention? *Alzheimer's & Dementia*, **2**, 202–4.
15. Bäckman, L., Small, B.J., and Fratiglioni, L. (2004). Cognitive deficits in preclinical Alzheimer's disease: current knowledge and future directions. In *New frontiers in cognitive aging* (eds. R.D. Dixon, L. Bäckman, and L.-G. Nilsson), pp. 161–77. Oxford University Press, Oxford.
16. Bäckman, L., Jones, S., Berger, A.K., *et al.* (2005). Cognitive impairment in preclinical Alzheimer's disease: a meta-analysis. *Neuropsychology*, **19**, 520–31.
17. Palmer, K. and Fratiglioni, L. (2006). Is mild cognitive impairment a distinct clinical entity? *Aging Health*, **2**, 763–9.
18. Gauthier, S., Reisberg, B., Zaudig, M., *et al.* (2006). Mild cognitive impairment. *Lancet*, **367**, 1262–70.
19. Portet, F., Ousset, P.J., Visser, P.J., *et al.* (2006). For the MCI working group of the European Consortium on Alzheimer's Disease (EADC). Mild cognitive impairment (MCI) in medical practice: a critical review of the concept and new diagnostic procedure. *Journal of Neurology, Neurosurgery, and Psychiatry*, **77**, 714–18.
20. Palmer, K., Bäckman, L., Winblad, B., *et al.* (2003). Detection of Alzheimer's disease and dementia in the preclinical phase: population-based cohort study. *British Medical Journal*, **326**, 245.
21. Sacuiu, S., Sjogren, M., Johansson, B., *et al.* (2005). Prodromal cognitive signs of dementia in 85-year-olds using four sources of information. *Neurology*, **65**, 1894–900.
22. Palmer, K., Wang, H.X., Bäckman, L., *et al.* (2002). Differential evolution of cognitive impairment in nondemented older persons: results from the Kungsholmen project. *The American Journal of Psychiatry*, **159**, 436–42.
23. Monastero, R., Palmer, K., Qiu, C., *et al.* (2007). Heterogeneity in risk factors for CIND. A population-based longitudinal study from the Kungsholmen project. *The American Journal of Geriatric Psychiatry*, **15**, 60–9.
24. Panza, F., D'Introno, A., Colacicco, A.M., *et al.* (2005). Current epidemiology of mild cognitive impairment and other predementia syndromes. *The American Journal of Geriatric Psychiatry*, **13**, 633–44.
25. De Ronchi, D., Berardi, D., Menchetti, M., *et al.* (2005). Occurrence of cognitive impairment and dementia after the age of 60: a population-based study from Northern Italy. *Dementia and Geriatric Cognitive Disorders*, **19**, 97–105.
26. Caracciolo, B., Palmer, K., Monastero, R., *et al.* (2008). Occurrence of cognitive impairment and dementia in the community: a 9-year long prospective study. *Neurology*, **70**, 1778–85.
27. Fratiglioni, L., von Strauss, E., and Qiu, C. (2007). Epidemiology of the dementias of old age. In *The Oxford textbook of old age psychiatry* (eds. R. Jacoby, T. Dening, A. Thomas, and C. Oppenheimer), pp. 391–406. Oxford University Press, Oxford.
28. Matthews, F. and Brayne, C. (2005). Medical Research Council Cognitive Function and Ageing Study Investigators. The incidence of dementia in England and Wales: findings from the five identical sites of the MRC CFA study. *PLoS Medicine*, **2**, e193.
29. Fratiglioni, L. and Wang, H.-X. (2007). Brain reserve hypothesis in dementia. *Journal of Alzheimer's Disease*, **12**(1), 11–22.
30. Gatz, M., Reynolds, C.A., Fratiglioni, L., *et al.* (2006). Role of genes and environments for explaining Alzheimer disease. *Archives of General Psychiatry*, **63**, 168–74.
31. Qiu, C.X., Kivipelto, M., Agüero-Torres, H., *et al.* (2004). Risk and protective effects of APOE gene towards Alzheimer's disease in the Kungsholmen project: variation by age and sex. *Journal of Neurology, Neurosurgery, and Psychiatry*, **75**, 828–33.
32. Huang, W.Y., Qiu, C.X., von Strauss, E., *et al.* (2004). APOE genotype, family history of dementia, and Alzheimer disease risk: a 6-year follow-up study. *Archives of Neurology*, **61**, 1930–4.
33. D'Introno, A., Solfrizzi, V., Colacicco, A.M., *et al.* (2006). Current knowledge of chromosome 12 susceptibility genes for late-onset Alzheimer's disease. *Neurobiology of Aging*, **27**, 1537–53.
34. Xu, W., Qiu, C., Winblad, B., *et al.* (2007). The effect of borderline diabetes on the risk of dementia and Alzheimer's disease. *Diabetes*, **56**, 211–16.
35. Fratiglioni, L., Paillard-Borg, S., *et al.* (2004). An active and socially integrated life in late life might protect against dementia. *Lancet Neurology*, **4**, 3343–53.
36. Karp, A., Paillard-Borg, S., Wang, H.X., *et al.* (2006). Mental, physical and social components in leisure activities equally contribute to decrease dementia risk. *Dementia and Geriatric Cognitive Disorders*, **21**, 65–73.
37. Ferri, C.P., Prince, M., Brayne, C., *et al.* (2005). Global prevalence of dementia: a Delphi consensus study. *Lancet*, **366**, 2112–17.
38. Agüero-Torres, H., Qiu, C., Winblad, B., *et al.* (2002). Dementing disorders in the elderly: evolution of disease severity over 7 years. *Alzheimer Disease and Associated Disorders*, **16**, 221–7.
39. Agüero-Torres, H., Fratiglioni, L., Guo, Z., *et al.* (1998). Dementia is the major cause of functional dependence in the elderly: 3-year follow-up data from a population-based study. *American Journal of Public Health*, **88**, 1452–6.
40. Agüero-Torres, H., Fratiglioni, L., Guo, Z., *et al.* (1999). Mortality from dementia in advanced age: a 5-year follow-up study of incident dementia cases. *Journal of Clinical Epidemiology*, **52**, 737–43.

8.4

Assessment of mental disorder in older patients

Robin Jacoby

The assessment of older people is not fundamentally different from that of younger patients. The principles of taking history and mental-state examination are the same at any age. But if the goals are common, the routes taken to reach them are not necessarily so. For example, an assessment adequate enough to begin treatment of a 30-year-old woman presenting to an outpatient clinic with a depressive illness might take about an hour and involve speaking only to the patient and perhaps briefly to her partner, whereas the equivalent assessment of an 81-year-old woman in whom uncertainty exists as to whether the diagnosis is that of a depressive or a dementing illness may require more than one interview and necessitate enquiry from several informants. This section will not repeat what can be found in Chapter 1.8.1, but cover only those points which are specific to or need to be emphasized for older patients.

The referral process

Who refers?

Whilst the referral process might be the same as for younger patients, it is more often different. In many cases the patient has no idea why, or indeed does not even know or has forgotten that she has been referred. (The feminine gender is used in this chapter because older women are more likely to develop a mental illness and to survive longer than men. However, what is written applies also to men.)

The process has most often been initiated by family members who might not have discussed it with the patient. Many old people live alone with no relatives nearby or even in the same country or state, so that referrals are frequently initiated by friends, neighbours, or other acquaintances, such as local shopkeepers, social services care workers, and people who run luncheon clubs.

Reasons for referral

In the case of a woman of 30 with a depressive illness, she is referred to a psychiatrist for treatment to effect a remission. However, an older woman of 80 with a similar condition may be referred for a variety of reasons including the following: the primary care doctor might be uncertain of the diagnosis, that is whether it could be dementia; the grown-up children might have removal from home to residential or nursing care as the first item on their agenda; the patient's condition may not be the primary issue—there may be greater concern for her husband who is failing to cope, perhaps to the extent of physically abusing her.

The informants

A large number of older people seen by psychiatrists are unable to give complete or reliable information about themselves. Frequently, but not invariably, there is a spouse or adult offspring living with the patient. In other cases, however, it is necessary to track down someone less obvious. Neighbours are often helpful at relating recent history, but may know little of past personal or family history. Effort spent in telephoning relatives, even those on the other side of the world, can be invaluable in giving an account of such items as family history or premorbid personality. If an informant is not readily available, for example, because it is night-time in Australia, the psychiatrist should not shelve the task of phoning, but only defer it to the next available opportunity.

Where conflicting information is given by a variety of informants it might be necessary to weigh up the particular 'hidden agenda' of each one. For example, the husband of a demented woman may minimize his wife's behaviour disturbance for fear that she would be 'put away'; whereas the daughter may overstate it in order to support a case for her mother's transfer to a nursing home because her father repeatedly phones her for assistance at all hours of the day and night. Each one of the two informants has cogent reasons for weighting the information, but the psychiatrist and his or her team cannot help to resolve the situation until they understand those reasons.

Professional informants

Psychogeriatrics is as dependent on multi-disciplinary working as any other branch of psychiatry. Many patients seen for the first time will already be well known to their primary care doctor who will be able to provide invaluable information. The same frequently applies to community psychiatric nurses who now take referrals directly from general practitioners and may themselves be making referrals to the old-age psychiatry service. The psychogeriatrician can save a great deal of time and effort by consulting community psychiatric nurses and general practitioners before seeing the patient or relatives.

Where to assess the patient

The patient needs to be placed at her maximum advantage to provide clinical information in whatever setting the assessment takes place. This has to be stated explicitly because the doctor is often required to take active steps to ensure it. Account has to be taken of special sensory impairment. Poor vision may need lights to be switched on so that the patient can see who is asking her questions. Distracting noises will make it even more difficult for someone with hearing impairment to grasp what is said. Surprisingly often, this may require a request that the television be switched off. Most importantly, examiners need to sit facing the patient with the lips visible, to speak slowly, and to enunciate words carefully. The patient should then be asked if she can hear properly. Simply shouting at her is not a substitute for these simple steps.

Social customs vary within and between societies. For instance, in the United Kingdom and the United States the use of first names is much more acceptable with younger adults than it was 40 years ago. With the current generation of older patients it is not. For them to be called by their first names unbidden is disrespectful and infantilizing. Even if nurses and other non-medical staff do so, doctors should not use first names, unless specifically invited. Instead, the surname plus appropriate title (Mr, Mrs, etc.) is correct.

At home

The preferred place to assess older patients is in their own homes, although circumstances sometimes dictate that it will be elsewhere. At home patients feel less intimidated and can be seen within an environment which tells the psychiatrist a great deal that he cannot know in the clinic. If a house is filthy and cold and the patient in a similar state, and if there is reliable information that this is only a recent phenomenon, then it is a powerful descriptor of the patient's inability to cope. However, the converse is not always true; a clean and tidy home may only reflect someone else's willingness to support and care for the patient who could not otherwise do it herself (e.g. a daughter or neighbour). Another advantage of a home assessment is that cognitive disabilities, such as dyspraxia and agnosia, can be tested in an ecologically valid way (making tea, recognition of family members from photographs) that is more acceptable to a patient than being formally tested with the Mini-Mental State Examination.[1]

Assessments at home require more preparation for the doctor than is necessary at outpatient clinics where equipment for physical examination and blood tests are available, for example. It is an obvious courtesy to the patient to let her know of the visit beforehand, but it is also wise to arrange for a suitable informant to be present. Furthermore, some older patients are incapable of letting visitors into their houses and the informant might well first have to facilitate the doctor's access. Elderly patients are much more likely to be suffering from comorbid physical illness which may be the fundamental cause of the mental disorder, for example, pneumonia or a urinary tract infection manifesting delirium. The old-age psychiatrist does not therefore need to adhere rigidly to lines of specialty demarcation but rather be aware of the possibility of and prepared to search for physical illness. The basic equipment for a medical examination, such as a stethoscope, sphygmomanometer, and patellar hammer are items to be taken on home visits. Urine testing strips and a thermometer, especially a low-reading thermometer, are also sometimes useful.

In a psychiatric hospital

Patients who are assessed after admission to psychiatric beds lack the advantages of being in their own environment, although the opportunity for physical examination is much easier. Another advantage for hospital inpatients is that the assessment can be carried out over a longer period of time, since older people tire more easily and cooperation varies from day-to-day. For example, some demented patients will object to undergoing full cognitive assessment in one go, especially because they are often aware that they are failing. If a few questions are asked in the course of several short sessions, a more accurate and complete picture of the patient's abilities eventually emerges. If the Mini-Mental State Examination is administered in this way, a higher total can be achieved than if an attempt to administer it all at once meets with sullen refusal after the first few questions, with all subsequent ones having to be scored zero.

Information from other informants is as crucial for hospital inpatients as it is for those seen at home. It is usually the responsibility of the house officer or resident to collect the history, and they may be required to telephone several informants in distant and local parts to obtain a full picture which the patient is incapable of providing.

Liaison visits in general hospitals

Liaison visits to patients in general hospitals make up a considerable part of the old-age psychiatrist's work because comorbid mental and physical illnesses are very common. In spite of the fact that the host nurse's instinct is to lead the visiting psychiatrist straightaway to the patient's bedside, the latter should insist on first reading the case notes (charts) and speaking to the nursing staff who know her best. From the case notes and the prescription cards (medication orders) invaluable information on current and past drug therapy as well as details of the patient's medical history are obtained. Clues as to the patient's mental state are often best gleaned from the records written by the nurses. Nevertheless, non-psychiatrist doctors, surgeons, and nurses are not accustomed to assessment of the mental state and statements such as 'confused' should not be taken at face value, since they stand for anything from slight difficulty in answering complex questions due to anxiety at being in a strange environment to major mental disturbance. As in most other settings, time spent telephoning informants from the general hospital ward is well invested and may permit the visiting old-age psychiatrist to express an opinion on the patient's condition more firmly than would otherwise have been possible.

A useful final step before going to talk to the patient is, if possible, to observe her from a suitable distance. In this way signs of delirium, disruptive behaviour, social interaction, and other phenomena such as dyskinesias may be seen.

When seeing the patient herself, wherever possible she should be taken to a separate room and not examined in an open ward where there are other patients. If it is impossible for the patient to leave her bed, then it is usually feasible to move the bed to a more private place.

Nursing and residential homes

Much of that which is required for liaison visits to general hospitals applies to assessment in residential or nursing homes, most notably trying to see the patient in a private room away from other residents. Since abuse of elderly people is sometimes an issue in these

settings, it is preferable to have at least some time completely alone with the patient first, and if indicated, to check for bruises or other injuries, and secondly to allow the patient to tell the doctor things which she might be frightened to do in front of the staff of the home. Another problem in some nursing and residential homes is that the psychiatrist finds that an untrained or unqualified member of staff accompanies him, the quality of whose information may not be at the level of trained nurses. Careful questioning of several members of staff, attention to written records, and telephone calls to appropriate informants should all improve the quality of the assessment.

The history

Family and personal history

As has already been made clear, for many older patients a complete history may have to be obtained from a variety of informants. With the patients themselves a more flexible approach than is taken with younger ones is often needed. Whether intellectual failure is obvious and global or there is only relatively mild cognitive impairment (MCI), for some to give a history that is fully chronologically correct can be too great an effort. The examiner must accept these limitations and try to keep the atmosphere as relaxed as possible. Much more than the young, elderly psychiatric patients perceive the psychiatric interview as an ordeal or a form of trial in which it is easy for them to acquire a sense of failure. This in turn induces anxiety and a vicious spiral of ever worsening performance. One way in which the patient can be put at ease is to reassure her that you will come to her main problem in due course but that it would be good to hear something of her background first. For most older patients the family and personal histories are easier to recall than the confusing events which have led up to the referral. This is not simply good for the patient but for the examiner as well. Amongst the most profitable of pleasures in old-age psychiatry are the life stories of people who have lived during some momentous periods of world history. Furthermore, these stories put patients into a context which makes it much easier to understand why and how they have reacted to the mental illness with which they have presented.

As regards the family history specifically, the examiner needs to be alert to mistakes which could indicate cognitive impairment. A patient may confuse family relationships or misidentify family members quite early in the course of a dementing illness. In other words, inaccurate information from the patient can be as clinically informative as that which is correct. Older patients are not necessarily as sophisticated in medical vocabulary as their children and grandchildren. Therefore, to obtain facts about a possible family history of dementia (an important issue), it may be necessary to ask if any blood relative had 'memory problems' in late life or 'had to go into a home'.

In eliciting the personal history the examiner might need to be aware of the historical context at the time in question. Some older patients, however affluent they may be now, grew up in poverty or other adverse circumstances (e.g. a parent died of tuberculosis during their childhood) which could still be affecting their psychological lives. Similarly, education may have been disrupted in a way that is more unusual nowadays. Some patients, notably women, relate how they missed education because they had to look after their younger siblings after mother died or father was killed, because the remaining parent had to go out to work for the family to survive. A precise enquiry should be made as to educational attainment, and especially the level of literacy and numeracy which may be the only 'baseline' appraisal obtainable for a patient with current dyslexia or dyscalculia.

Whilst it is often very obvious to the examining doctor that a patient has cognitive impairment because of her errors and inconsistencies in giving her personal history, it is very unwise to foreclose on a diagnosis at this stage. Patients with severe depressive illnesses can be even more hesitant than those with, for example, Alzheimer's disease, or show such retardation and lack of concentration on what is being asked that they can portray a clinical picture that is not easily distinguishable from a state of advanced dementia.

Medical and psychiatric history

The nature of the information required is no different from that in younger patients. However, of particular importance in the older population is past and present medication. Drugs taken at the prescribed dose, at a wrong dose (due to dementia), or drugs no longer intended to be taken and prescribed sometimes quite a long time ago but of which a residual supply remains, are all potent causes of confusion and even frank delirium in old people. It is therefore good practice to ask to see where all medication is kept and to examine each pack or bottle to check that the amount left is approximately proportionate to that which one might expect given the date the drugs were dispensed. Since elderly people are frequently the victims of clinically injudicious polypharmacy, it is common to find a large quantity of current and prescription-expired drugs which the patient is taking on a random basis. To counter this problem proprietary boxes which dispense drugs in daily amounts, such as the Dosette or Nomad systems, are used and can more easily be checked for compliance or overdosing.

Premorbid personality

Personality is one of the prime determinants of outcome in mental illness at any age, but where older people are concerned too little effort is made, partly for lack of reliable informants, to give a valid assessment of personality. False assumptions are made that someone has always been awkward or cantankerous, whereas it is shown later that they have become so because of frontal-lobe impairment. Similarly, it is sometimes assumed that a woman has always exhibited attention-seeking or manipulative behaviour, when to the surprise of doctors and nurses alike such behaviour disappears following effective treatment for a depressive illness. Every effort should be made to find a reliable informant before reaching such conclusions.

The mental state examination

Appearance, behaviour, and the environment

A great deal can be learnt from the appearance of the patient and her home environment. Signs of neglect in both are commonly found. A brief tour of the home may reveal rotten food, little or no food, empty bottles of liquor, evidence of poor hygiene or incontinence, and inadequate heating. The patient may be dirty and unkempt, and clothes may have been put on in the wrong order (dressing dyspraxia). Particular attention should be paid to relatively mild impairment of attention and concentration, since it

might betray a delirium which can be treated to achieve a remarkable improvement in the patient's condition. Agitation is another sign to be carefully sought. Sometimes agitation is obvious with behaviour such as pacing and sighing which make it almost impossible to communicate with the patient, but at other times it can be allied to psychomotor retardation and perceptible only in tireless movements of the fingers.

Talk

Dysphasia in any of its guises is a frequent manifestation of dementia. Sometimes it is obvious, but not always. For example, it may be difficult at first to differentiate between an expressive dysphasia (Broca's dysphasia) and retardation. It is relatively mild or moderate receptive dysphasia (Wernicke's dysphasia), however, which traps the unwary clinician. In such cases the patient may appear to be obtuse, unintelligent, or hard of hearing until it is appreciated that she simply does not understand a considerable proportion of what is said to her.

Even if the patient is not dysphasic, she may be evasive and given to circumlocution. Again, this should not be taken automatically as evidence of a premorbid lack of intelligence or pompousness, since it is frequently used as a camouflage for cognitive impairment. A patient with dementia might say 'Oh of course I know that' or 'I never paid any attention to that sort of thing' when asked to give an item of current affairs.

In manic or hypomanic illness in old-age slow flight of ideas is sometimes missed by inexperienced clinicians or mistaken for evidence of cognitive impairment. Here the normal coherence of thought is disrupted because the patient is distracted from one idea to another, just as in characteristic flight of ideas, but they are delivered at a normal or even slower pace. The latter can occur if a mixed affective state is present.

Thought content

Much of that which applies to younger patients applies also to older ones and does not need to be mentioned here. However, subtle changes in thought content amounting to a restriction in breadth and a repetitiveness of themes may be noticed. Formal thought disorder which is found in young patients with schizophrenia is extremely rare in the old.

Mood

Depressive illness is commonly missed in older patients. This is partly because the clinical picture can mimic a dementing illness, so that a history from a reliable informant, as has already been stressed, is mandatory. Another reason is so-called *masked depression* in which the patient denies depressed mood but presents with other symptoms, such as those of apparent physical illness. In this sort of case a reliable account of sleep and appetite disturbance, weight loss, and anhedonia give clues as to the presence of an affective disorder. It should also be remembered that dementia and depressive illness are common disorders and not infrequently occur together, so that the diagnosis of one does not rule out the other.

Older men are still the group most at risk of suicide in most countries of the world where statistics are recorded (see Chapter 4.15.1). The psychiatrist does not shrink from specific enquiry about suicidal thinking in patients of any age, but it can sometimes be difficult to differentiate between a rational desire to die when the time comes and active suicidal ideation. In-depth probing is therefore mandatory and the examiner should not be put off by the patient's attempts to leave the topic, if he or she thinks that there is likely to be risk of self-harm.

Cognitive examination

In assessing elderly patients more emphasis is usually placed on the cognitive examination than in younger patients. In theory it can be as exhaustive and thorough as assessment by a neuropsychologist, but in the routine practice of old-age psychiatry it usually has to be feasible within the constraints of a consultation which lasts about an hour.

It is in this part of the assessment that the examiner is most likely to lose the patient's cooperation, principally because of the humiliation experienced by some at their own failures. Some patients become angry, indignant, or defensive. Others become anxious and their performance deteriorates. One way to pre-empt this is to preface testing by stressing that this is not a competitive examination and that most people have difficulty answering some of the questions. Correct answers are praised without excessive emphasis and incorrect ones are either treated in a neutral way or given a positive spin by saying, for example, 'Well it's…, but you weren't far off'.

Many old-age psychiatrists and other members of their multidisciplinary team prefer to use standardized questionnaires, such as the Mini-Mental State Examination,[1] the questionnaire in widest use. It has the advantage that results between and within patients can be compared and progress can be monitored. However, no off-the-shelf test is exhaustive and none produces an adequate cognitive assessment by itself. The clinician therefore needs to have some sort of schema for covering the main areas of cognitive function which would include: memory (in its various aspects) and general information, and naming; the understanding and production of language; praxis (ideomotor and constructional); sensory recognition (gnosis); abstract reasoning; verbal fluency; calculation; left/right orientation; and executive function (the ability to integrate mental processes for goal-directed activity). The list is not exhaustive and some areas may need to be covered in greater detail as the clinical situation demands. Table 8.4.1 gives a guide to cognitive examination based on an extended Mini-Mental State Examination. Clinicians vary in the order and way in which they test individual cognitive functions, and the list in Table 8.4.1 is not intended to be prescriptive.

Other aspects of the mental state examination

These do not differ in essence from that in younger adults and are not covered in this section.

Physical assessment

Since physical comorbidity is extremely common, the old-age psychiatrist needs to be able and willing to conduct a basic physical examination. In the patient's home this may not always be easy. For instance, it may not be kind to ask a frail person who takes an hour or more to dress and come downstairs in the morning to return to her bed and undress, but it is usually possible to make a reasonable examination of the arterial pulse, blood pressure, and jugular venous pulse, and to auscultate the heart and lungs when

Table 8.4.1 Schema for testing cognitive functions based on an extended mini-mental state examination

Function	Subfunction	Examples
Orientation	Time	Year, month, date, day, season
	Place	Own address or that of hospital, city, county/state, country
	Person	Own name (married women sometimes cannot give married name); recognize others by name or function (e.g. you are a doctor)
Memory	Immediate recall	Immediate repetition of three objects or a name and (local) address
	Delayed recall	Repetition as above but after a distractor task
	Long-term recall	Give historical or personal events (that can be verified)
	General information	Names of politicians or other VIPs
Concentration		Months of the year in reverse order, counting from 20 back to 1; spelling WORLD forwards then backwards
Praxis	Construction	Copy diagram of interlocking pentagons
	Ideomotor	Draw a clock and set the hands at a specified time (also a test of executive function)
	Dressing	Put on a jacket; undo, and refasten buttons
Sensory recognition (gnosis)	Visual including prosopagnosia	Recognize photographs taken from unusual angles and of familiar faces
	Auditory	Recognize the doorbell
	Tactile	Recognize objects placed in the palm, e.g. coins
	Reading	Any sample *but* use large print, e.g. newspaper headlines
	Olfactory	Recognize something from the kitchen, e.g. coffee
Language	Expressive	Repeat 'no ifs ands or buts'
	Understanding	Carry out a three-stage command
	Naming	Naming objects of increasing complexity
Verbal fluency		List as many items from a category as possible in 1 min, such as boys' and girls' forenames, or as many words beginning with a specified letter
Writing		Write an ordinary English sentence
Calculation		Not too complex—subtraction of serial sevens from 100 is too difficult for many. A simple sum involving money is better
Left–right orientation		Face–hand test (e.g. left hand to left ear, right hand to left ear); finger recognition on own and examiner's hand
Abstract reasoning		'In what way are an apple and a banana alike?'
		'In what way are a boat and a car similar?'
		Interpretation of simple proverbs

Testing one function usually depends on one or more others; for example, most tests depend on understanding of language. Examples are neither prescriptive nor exhaustive.

the patient is seated. Similarly, a partial neurological examination for signs of focal deficits is also possible. However, if something alerts the doctor to the need to examine an undressed and supine patient, the duty should not be shirked, lest a hitherto unsuspected abdominal mass, or a strangulated hernia are missed.

Patients admitted to psychiatric hospital or nursing home beds should all undergo a physical examination. Focal neurological signs may indicate the cause of dementia. Carcinoma of the breast which either dementia or fear has prevented the patient from declaring may be much more treatable than she has believed. All the physical disorders which may be revealed are too numerous to mention, but their detection and treatment nearly always contribute to an improvement in mental function.

Laboratory investigations

Owing to tight budgetary constraints it is sometimes argued that routine laboratory investigations, such as full blood count and chest radiography, are unnecessary for younger adults. Whether or not this is true, with older patients such tests are strongly advised because the treatment of comorbid physical illness improves mental disorder. Furthermore, a treatable or arrestable cause of dementia may be found. For patients with dementia the following are recommended: full blood count; serum electrolytes, and creatinine; liver and thyroid function tests; syphilis serology; vitamin B_{12} and red cell folate; chest radiography. Medical and nursing staff should also have a low threshold for sending urine for microbiological examination (see superimposed delirium below). The vexed question of neuroimaging, an expensive procedure, is much discussed and the debate is not easily summarized or resolved. Most space-occupying lesions can be detected, as can many vascular changes. However, there is nothing pathognomonic for Alzheimer's disease on CT, magnetic resonance imaging, or single-photon emission tomography. In Alzheimer's disease, scans may support the diagnosis but not establish it.

General considerations in the assessment of older psychiatric patients

Falls

The causes of falls in older people are many and the reader is referred to textbooks of geriatric medicine for a full discussion. However, the old-age psychiatrist needs to be aware that many

psychotropic drugs precipitate falls through postural hypotension, with tricyclic antidepressants and neuroleptics being particular offenders. Neuroleptics induce parkinsonism, putting the patient at risk of tripping against rugs or items of furniture. The clinical implications of falls, especially in older women, are serious because patients are at risk of a fractured neck of femur or (less commonly) a subdural haematoma, both of which carry a high mortality.

Nutrition

Poor nutrition is commonly found in patients presenting to old-age psychiatry services. Even in affluent societies many older people are amongst the most impoverished or feel that they cannot afford good food. Some are too frail to get out to the shops. Others lack motivation to shop and eat because of depressive illness. Patients suffering from dementia may be incapable of shopping and preparing food. Widowers might never have learned to cook. A vitamin B_{12} or folate level at the lower end of the reference range, or even below it, is as often a reflection of poor nutrition as an indication of pernicious anaemia.

Superimposed delirium

Here it is mentioned only that subacute delirium superimposed on another condition is at risk of not being recognized because it is taken to be a manifestation of the underlying illness, usually dementia. This is particularly the case when a subclinical urinary tract infection occurs in a demented patient. It is of clinical relevance because treatment of the urinary tract infection results in a great improvement in the patient's mental state. On routine assessment particular attention should therefore be paid in the history to evidence of sudden worsening of a stable or only slowly deteriorating condition, and to nocturnal disturbance especially with (usually visual) hallucinations. On examination of the patient herself the level of consciousness, awareness of the environment, attention, and concentration should be noted.

Delirium is described in Chapter 4.1.1, and special features in older people are considered in Chapter 8.5.1.

Reassessment after treatment

Because so much of psychiatry is practised in the community, it is impossible for all patients to be reassessed by a psychiatrist after treatment. In hospital or outpatient clinics it is feasible, but not for the majority of older patients who live at home. Furthermore, in areas where the population is geographically widespread, it is very difficult if not impossible for many older people, who may be frail and infirm, to travel long distances to attend clinics. Much of the follow-up assessment in old-age psychiatric services is therefore carried out by other members of the multi-disciplinary team, such as psychologists or occupational therapists, but mostly by community psychiatric nurses. It is essential for the psychiatrist to meet regularly with members of the team seeing patients in the community to discuss the progress of individual patients following treatment.

Further information

Goldberg, D. and Murray, R. (eds.) (2006). *The Maudsley handbook of practical psychiatry* (5th edn). Oxford University Press, Oxford.

Hodges, J.R. (2007). *Cognitive assessment for clinicians* (2nd edn).Oxford University Press, Oxford.

Jacoby, R., Oppenheimer, C., Dening, T. *et al*. (eds.) (2008). *The Oxford textbook of old age psychiatry*. Chap. 10–12. Oxford University Press, Oxford.

References

1. Folstein, M.F., Folstein, S.E., and McHugh, P.R. (1975). 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **12**, 189–98.

8.5

Special features of clinical syndromes in the elderly

Contents

8.5.1 Delirium in the elderly

James Lindesay

Note Dementia in people of all ages is considered in Part 4, Section 4.1, where the following topics are considered: Chapter 4.1.2 Dementia: Alzheimer's disease; Chapter 4.1.3 Frontotemporal dementias; Chapter 4.1.4 Prion disease; Chapter 4.1.5 Dementia with Lewy bodies; Chapter 4.1.6 Dementia in Parkinson's disease; Chapter 4.1.7 Dementia due to Huntington's disease; Chapter 4.1.8 Vascular dementia; Chapter 4.1.9 Dementia due to HIV disease; Chapter 4.1.10 The neuropsychiatry of head injury; Chapter 4.1.11 Alcohol-related dementia (alcohol-induced dementia; alcohol-related brain damage); Chapter 4.1.13 The management of dementia.

Introduction

Although delirium occurs at all ages, it is most frequently encountered in late life. This is because delirium is the result of an interaction between individual vulnerability factors (e.g. brain disease, sensory impairment) and external insults (e.g. physical illness, medication), the rates of which both increase with age. Our current concept of delirium derives principally from the florid clinical stereotype that has evolved from centuries of clinical observations on younger patients, and it may not be applicable to our historically unique ageing population. In younger adults, a major physical insult is usually necessary to precipitate delirium, which is often a dramatic disturbance. This is not the case in vulnerable elderly patients when relatively mild physical, psychological, or environmental upsets may be sufficient to bring about acute disturbances of mental functioning. These disturbances may be less obvious than in younger patients, particularly if they occur in the context of pre-existing cognitive impairment. Consequently, despite being common and problematic, delirium in elderly patients is frequently missed or misdiagnosed as dementia or depression by medical and nursing staff.[1] This is unfortunate, because delirium is an important non-specific sign of physical illness or intoxication, and if left untreated there may be costly consequences, both for the patient and for health services.

Clinical features

The clinical features of delirium are described in Chapter 4.1.1. Most delirium in elderly patients is of the quiet hypoactive variety, lacking the more florid disturbances in mood, perception, and behaviour that bring the disorder to clinical notice. Reversible cognitive impairment in elderly patients is associated with reduced conscious level, poor attention, poor contact with the patient, incoherent speech, reduced psychomotor activity, lack of awareness of surroundings,

poor orientation, and poor memory.[2] Hyperactive delirium does occur in elderly patients, but it is less pronounced, with the overactivity usually confined to purposeless behaviour such as pulling at the bedclothes. Violent behaviour is uncommon; elderly patients are more likely to injure themselves than others.

Classification

The ICD-10 and DSM-IV diagnostic criteria for delirium are described in Chapter 4.1.1. They are not entirely concordant; ICD-10 is more restrictive, resulting in the diagnosis of fewer cases.[3] However, the two systems agree on four essential features: disturbance of consciousness, disturbance of cognition, rapid onset/fluctuating course, and evidence of an external cause. Unfortunately, none of these features is specific for delirium as opposed to dementia, and the current diagnostic criteria are poor predictors of outcome, defined in terms of improvement in cognitive function. Reversibility of cognitive impairment may be the most discriminating feature of delirium,[2] but is problematic as a diagnostic criterion since outcome is unknown at the outset.

Another shortcoming of the current classifications of delirium is that they do not recognize the partial and transitory disturbances that are commonly observed in elderly patients. Subsyndromal delirium is common, and is part of a continuum between normality and the full syndrome. Subsyndromal cases are clinically significant, since they have the same risk factors and the same increased mortality as syndromal cases.[4]

Diagnosis and differential diagnosis

The diagnosis of delirium is a two-stage process: first, diagnose the delirium, and second, identify the underlying cause or causes. The diagnosis of delirium in elderly patients can be problematic, given the predominantly hypoactive clinical picture and the unreliability of 'positive' symptoms. However, it is important to consider the possibility if cognitive decline is rapid, and if any of the recognized signs and symptoms are present. A good informant history from relatives or ward staff is essential to establish the onset and course of the disorder. Routine screening procedures may be useful in identifying patients who develop delirium while in hospital. Brief instruments such as the Mini-Mental State Examination[5] are not diagnostic, but will alert the clinician to any sudden decline in cognitive function. More extended diagnostic instruments are also available, such as the Delirium Rating Scale,[6] the Confusion Assessment Method,[7] and the Delirium Symptom Interview.[8] Another approach to screening for delirium is to identify those at particular risk of developing the disorder. Predictive factors related to the patient include: visual impairment, severity of illness, cognitive impairment, and a blood urea nitrogen/creatinine ratio of 18 or more.[9] Hospital- and treatment-related factors include: use of restraints, malnutrition, use of more than three medications, bladder catheterization, and the number of iatrogenic events.[10] These factors are multiplicative in their effect.

The differential diagnosis of delirium includes most other psychiatric disorders in this age group. These disorders are themselves risk factors for delirium, so the possibility of co-morbidity must always be considered. When in doubt, investigate and manage as delirium until the situation is clear.

Dementia

Dementia is a major risk factor for delirium, and in practice co-morbidity commonly occurs. However, differential diagnosis is important, as episodes of delirium need to be identified in order for them to be managed effectively. Recent onset and rapid decline of cognitive functioning, from whatever baseline, indicate an episode of delirium until proved otherwise. Delirium in elderly patients can be prolonged, and failure to recover quickly following treatment of the cause does not necessarily indicate an underlying dementia. It is important to have a good history of pre-morbid functioning.

Depression

Delirium can be difficult to distinguish from severe depression in elderly patients, cognitive impairment associated with severe depression is usually relatively mild in comparison with the affective disturbance, whereas the reverse is true of delirium. The pattern of diurnal variation also varies in the two disorders, with depressed patients tending to be worse in the mornings, and delirious patients in the evenings. Elderly depressed patients are at increased risk of delirium, either through self-neglect or because of the antidepressant treatment they are receiving. Anticholinergic tricyclic drugs are particularly troublesome in this respect. Adverse life events, such as bereavement, may precipitate both depression and delirium in vulnerable individuals.

Mania

Mania is much less common than delirium in old age, and is often mistaken for it. There may be a previous history of manic-depressive illness, but a proportion of cases of mania in late life are first presentations, usually in association with underlying organic brain disease. Elderly manic patients are often exhausted and dehydrated, and so 'manic delirium' is a common presentation.

Other disorders

Anxiety states in elderly patients are unlikely to be mistaken for delirium, unless they are particularly severe. Similarly, paranoid states and schizophrenia rarely lead to diagnostic difficulty, although it should be noted that patients with these disorders are at an increased risk of developing delirium, either through self-neglect, or the effects of neuroleptic and anticholinergic medications. A number of other rare conditions in which cognitive, perceptual, affective, and behavioural disturbances occur, such as amnesic syndromes, epilepsia partialis continua, twilight states, the Charles Bonnet syndrome, neuroleptic malignant syndrome, and catatonia, may also resemble delirium. If the history and clinical examination are inconclusive, EEG may be helpful in making the diagnosis.

Epidemiology

The community prevalence of delirium increases with age, rising to 14 per cent in those aged 85 years and older. In medical and surgical inpatients, the rates of delirium vary considerably (prevalence, 10–30 per cent; incidence, 4–53 per cent), because of methodological and population differences. Similar rates are also found in studies of acute psychogeriatric admissions.[11] Some patient groups, such as those with hip fractures, have consistently higher rates. Other at-risk populations, such as nursing home residents, have received less systematic investigation, but the available evidence

suggests that they also have rates of delirium comparable to those found in elderly inpatients.

Aetiology

Almost any physical illness can give rise to delirium in elderly patients. The most common physical causes are listed in Table 8.5.1.1. In many cases the underlying cause is not obvious, and the delirium may be the most prominent presenting feature. The aetiology is commonly multi-factorial, and all contributory factors need to be identified and treated. As a rule, hyperactive delirium is more commonly due to infection and toxic/withdrawal states, whereas hypoactive delirium is more commonly due to metabolic abnormalities.

Table 8.5.1.1 Common causes of delirium in elderly patients

Drugs
Psychotropics
Hypnotics
Anticonvulsants
Anticholinergic drugs
Dopamine agonists
Analgesics
Anaesthetics
Alcohol withdrawal
Infection
Urinary tract infections
Pneumonia
Septicaemia
Ulcers, pressure sores, gangrene
Endocarditis
Postsurgical wound infection
Metabolic and endocrine
Electrolyte abnormalities
Uncontrolled diabetes
Hyper/hypothyroidism
Renal failure
Hepatic failure
Hypothermia
Malnutrition
Cardiovascular
Cardiac failure
Myocardial infarction
Vascular disease
Anaemia/polycythaemia
Respiratory
Pulmonary embolism
Pneumothorax
Pleural effusion
Intracranial
Trauma
Subdural haematoma
Stroke
Tumour
Epilepsy
Gastrointestinal
Perforation
Pancreatitis
Cholecystitis/cholangitis
Haemorrhage
Constipation

Drugs are an important cause of delirium in elderly patients, due to age-associated changes in their distribution, metabolism, and excretion. These pharmacokinetic changes are very variable, with the result that toxicity at apparently therapeutic doses is unpredictable. Certain drugs are particularly prone to cause delirium in elderly patients, for example those with anticholinergic activity. Tricyclic antidepressants, thioridazine, and benzhexol are particularly toxic in this respect, but many of the drugs commonly prescribed to elderly patients have some degree of anticholinergic activity, for example, digoxin, prednisolone, cimetidine, ampicillin, and warfarin. Individually, this activity may be small, but the cumulative effect can be significant if patients are on multiple medications.[12] Patients with Alzheimer's disease are particularly prone to develop delirium when given anticholinergic drugs, perhaps because their central cholinergic function is already impaired. In a minority of particularly vulnerable elderly patients, purely environmental and psychological insults are sufficient to cause delirium. The mechanisms of action in these cases are not known, but may involve factors such as sensory deprivation and stress responses via the hypothalamic-pituitary-adrenal axis.

Course and prognosis

Traditionally, delirium has been regarded as a transient condition that proceeds to either recovery or death. In the majority of cases, the delirium is brief, but about one-third of patients have prolonged or recurrent episodes.[13] Delirium is associated with increased short-term mortality in elderly patients, mainly because of the severity of the underlying illness. Delirium interferes with the processes of diagnosis, treatment, and rehabilitation, and as a result patients have longer hospital stays and higher rates of functional decline and discharge to nursing homes.[14] Increased length of stay and mortality are particularly associated with hypoactive delirium. In general, patients with hyperactive delirium appear to be less severely ill than those with hypoactive delirium; this may be due to differences in the cause of the delirium, or to the fact that hyperactive delirium is more likely to be identified and the causes treated.

Prospective studies have shown that the prognosis, in terms of persistent or recurrent symptoms, is relatively poor in elderly patients.[15] This is probably because those who experience delirium are a vulnerable group who are likely to develop the condition provided there is sufficient external insult. A proportion will also be suffering from a form of dementia, which will increase the vulnerability to delirium as it progresses. There is evidence that delirium is followed by persistent cognitive decline,[16] which raises the possibility that it (or the underlying cause) is a risk factor for the development or exacerbation of dementia.

Evaluation of treatment

Evidence regarding the efficacy of treatments for delirium is sparse (Chapter 4.1.1). The cholinergic hypothesis of delirium raises the possibility that cholinergic agonists, such as the cholinesterase inhibitors licensed for the treatment of Alzheimer's disease, may be of value in the prevention and treatment of delirium.

Management

There are four important steps in the management of delirium[17]:

Address the underlying causes (see above)

Behavioural control

This aspect of delirium management can be divided into pharmacological and non-pharmacological strategies. Non-pharmacological interventions in delirium are aimed at reducing the confusing, frightening, and disorienting aspects of the hospital environment in which most patients find themselves. They have received little formal evaluation, but features such as good lighting, low noise levels, a visible clock, and the reassuring presence of personal possessions and familiar individuals, such as relatives, are thought to be helpful. Any invasive intervention, including personal care tasks, should be introduced and explained simply, slowly, clearly, and repeatedly before it is carried out. Holding the patient's hand while talking helps to focus attention and provides reassurance.

The drug treatment of the symptoms and behaviours of delirium in the elderly is similar to that of younger patients, although it is necessary to start with lower doses, such as haloperidol 0.5 to 2 mg orally, or intramuscularly if necessary, repeated until the disturbance is controlled. Prescriptions should be for short periods only (up to 24 h) to encourage review of the effects and the necessary dosage. Once the delirium has resolved, the medication should be reduced/discontinued over a period of 3 to 5 days. If the patient cannot tolerate typical or atypical neuroleptic drugs, then a benzodiazepine (for instance, diazepam, lorazepam, or alprazolam) should be used instead.

Prevent/treat complications

The complications that befall patients with delirium probably contribute to the adverse outcomes associated with this condition. For example, hyperactive delirium is associated with falls during the hospital admission, whereas hypoactive delirium is associated with the development of pressure sores. Other complications of delirium include urinary incontinence, sleep disturbance, malnutrition, and immobilization; all of these problems should be anticipated and prevented where possible.

Rehabilitation and family support

Given the risk of functional decline following delirium, every effort should be made to return the patient to their pre-morbid level of functioning. ADL capacity should be assessed regularly, and independence encouraged where possible. The patient's family need to be involved in the rehabilitation process, as they will be largely responsible for aftercare following discharge. They should know that delirium is often recurrent, and be advised about the early signs of this. Indeed, delirium is a useful marker of vulnerability, and of the need for more intensive community aftercare.

Prevention

The modern hospital environment contributes significantly to the development of delirium in elderly patients, and multi-component interventions to improve poor clinical practice have been shown to reduce cost-effectively the incidence of delirium in elderly inpatients.[18] The following areas are important:

- *Prescribing*

Avoid where possible any drugs with known deliriogenic potential, particularly in at-risk individuals such as those with Alzheimer's disease. There should be regular review of the drug chart, with the aim of keeping the number of drugs to the minimum necessary. Non-pharmacological sleep-promotion strategies should be used in preference to hypnotic drugs.

- *Ward environment and routines*

These should aim to minimize disorientation, sensory impairment, and sleep deprivation. Patient mobility should be encouraged, as should adequate food and fluid intake. Medical and nursing staff should be trained to recognize and manage delirium.

- *Surgical routines*

Good preoperative, perioperative, and postoperative care (especially with regard to infection control, blood pressure, and oxygenation) will reduce the risk of postoperative delirium.

Further information

American Psychiatric Association. (1999). *Practice guideline for the treatment of patients with delirium*. American Psychiatric Association, Washington, DC.

British Geriatrics Society. (2006). *Guidelines for the prevention, diagnosis and management of delirium in older people in hospital*. British Geriatrics Society, London.

Byrne, E.J. (1994). *Confusional states in older people*. Edward Arnold, London.

Lindesay, J., Rockwood, K., and Macdonald, A. (2002). *Delirium in old age*. Oxford University Press, Oxford.

Lipowski, Z.J. (1990). *Delirium: acute confusional states*. Oxford University Press, New York.

References

1. Bowler, C., Boyle, A., Branford, M., *et al.* (1994). Detection of psychiatric disorders in elderly medical in-patients. *Age and Ageing*, **23**, 307–11.
2. Treloar, A.J. and Macdonald, A.J.D. (1997). Outcome of delirium: Parts 1 and 2. *International Journal of Geriatric Psychiatry*, **12**, 609–18.
3. Liptzin, B., Levkoff, S.E., Cleary, P.D., *et al.* (1991). An empirical study of diagnostic criteria for delirium. *The American Journal of Psychiatry*, **148**, 451–7.
4. Levkoff, S.E., Liptzin, B., Cleary, P.D., *et al.* (1996). Subsyndromal delirium. *The American Journal of Geriatric Psychiatry*, **4**, 320–9.
5. Folstein, M.F., Folstein, S.E., and McHugh, P.R. (1975). Mini-mental state-a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **2**, 1891–8.
6. Trzepacz, P.T., Baker, R.W., and Greenhouse, J. (1988). A symptom rating scale for delirium. *Psychiatry Research*, **23**, 89–97.
7. Inouye, S.K., van Dycke, C.H., Alessi, C.A., *et al.* (1990). Clarifying confusion: the confusion assessment method. *Annals of Internal Medicine*, **113**, 941–8.
8. Albert, M.S., Levkoff, S.E., Reilly, C., *et al.* (1992). The delirium symptom interview: an interview for the detection of delirium symptoms in hospitalized patients. *Journal of Geriatric Psychiatry and Neurology*, **5**, 14–21.
9. Inouye, S.K., Viscoli, C.M., Horowitz, R.I., *et al.* (1993). A predictive model for delirium in hospitalised elderly medical patients based on admission characteristics. *Annals of Internal Medicine*, **119**, 474–81.

10. Inouye, S.K. and Charpentier, P.A. (1996). Precipitating factors for delirium in hospitalized elderly persons. Predictive model and interrelationship with baseline vulnerability. *The Journal of the American Medical Association*, **275**, 852–7.
11. Lindesay, J., Rockwood, K., and Rolfson, D. (2002). The epidemiology of delirium. In *Delirium in old age* (eds. J. Lindesay, K. Rockwood, and A. Macdonald), pp. 27–50. Oxford University Press, Oxford.
12. Tune, L., Carr, S., Hoag, E., *et al.* (1992). Anticholinergic effects of drugs commonly prescribed for the elderly: potential means of assessing risk of delirium. *The American Journal of Psychiatry*, **149**, 1393–4.
13. Rudberg, M.A., Pompei, P., Foreman, M.D., *et al.*(1997). The natural history of delirium in older hospitalized patients: a syndrome of heterogeneity. *Age and Ageing*, **26**, 169–74.
14. Inouye, S.K., Rushing, J.T., Foreman, M.D., *et al.* (1998). Does delirium contribute to poor hospital outcome? A three-site epidemiologic study. *Journal of General Internal Medicine*, **13**, 234–42.
15. Levkoff, S., Evans, D., Liptzin, B., *et al.* (1992). Delirium, the occurrence and persistence of symptoms among elderly hospitalised patients. *Archives of Internal Medicine*, **152**, 334–40.
16. Jackson, J.C., Gordon, S.M., Hart, R.P., *et al.* (2004). The association between delirium and cognitive decline: a review of the empirical literature. *Neuropsychological Review*, **14**, 87–98.
17. Marcantonio, E. (2002). The management of delirium. In *Delirium in old age* (eds. J. Lindesay, K. Rockwood, and A. Macdonald), pp. 123–51. Oxford University Press, Oxford.
18. Inouye, S.K., Bogardus, S.T., Charpentier, P.A., *et al.* (1999). A multicomponent intervention to prevent delirium in hospitalized older patients. *The New England Journal of Medicine*, **340**, 669–76.

8.5.1.1 Mild cognitive impairment

Claudia Jacova and Howard H. Feldman

Introduction

Within the cognitive functioning continuum from normal ageing to dementia three broad states can be distinguished: normal functioning for age, clear-cut impairment meeting diagnostic criteria for dementia, and mild cognitive impairment (MCI), which falls below normal but short of dementia in severity (Fig. 8.5.1.1.1). There is active debate over what MCI is, how to define and classify this state, and where to set its borders on the described continuum.[1] Some definitions depict MCI as the tail-end of normal cognitive ageing whereas in other definitions MCI embodies the early clinical manifestation of Alzheimer Disease (AD) and other dementias. In 2003, the key elements of different MCI definitions were integrated into a consensus diagnostic and classification framework,[2] thus establishing some common ground in a field that is still evolving. MCI has also been positioned as a potentially important target for early treatment interventions to delay progression to dementia.

Nosologically, MCI is not currently included as a diagnostic entity in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR)[3] and the *International Classification of Diseases*, 10th revision.[4] The diagnostic categories of *Mild Neurocognitive Disorder* (DSM-IV-TR) and *Mild Cognitive Disorder* (ICD-10) are similar to MCI because they require the presence of cognitive impairment but these categories can only be assigned if a specific neurological or general medical condition can be identified to account for the cognitive symptoms. Much of the current condition

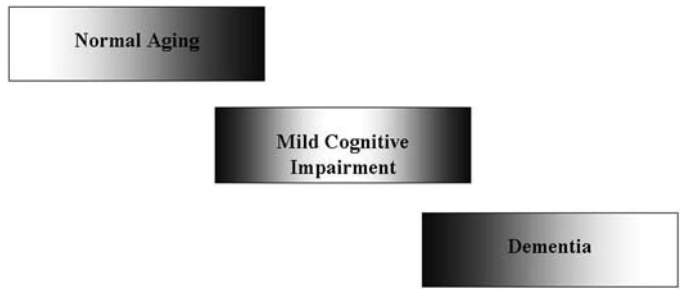


Fig. 8.5.1.1.1 Theoretical continuum from normal ageing to dementia: darker shading indicates areas of overlap between adjacent states and increased diagnostic challenge. (Reproduced from R. Petersen *et al.* Apolipoprotein E status as a predictor of the development of Alzheimer's disease in memory-impaired individuals, *The Journal of the American Medical Association*, **273**, 1274–78, copyright 1995, The American Medical Association.)

of MCI does not fit as it has no aetiologic specification. Nevertheless, MCI is increasingly a presenting condition in primary and specialized settings of care. Medical practice guidelines have recognized MCI as a risk state for dementia and recommend careful clinical evaluation and monitoring of individuals with this diagnosis.[5,6]

Nosology

The current nosological entities within the general MCI framework include a variety of definitions and capture overlapping but not identical conditions in the ageing population (Fig. 8.5.1.1.2).

Age-associated memory impairment (AAMI)

AAMI describes healthy individuals over the age of 50 that experience memory decline. Formal diagnostic criteria require complaints of memory loss, performance on objective memory tests falling at least 1 SD below norms for young individuals, and intellectual functioning normal. AAMI cannot be diagnosed if there is a

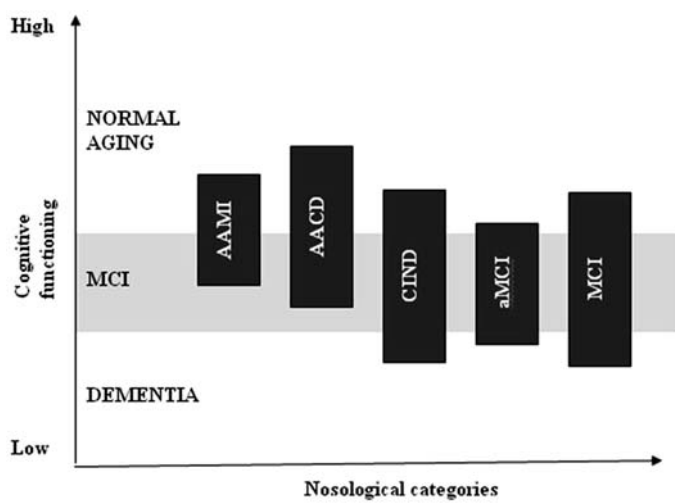


Fig. 8.5.1.1.2 MCI nosological entities on the continuum from normal ageing to dementia. (Reproduced from H.H. Feldman and C. Jacova, Mild cognitive impairment, *The American Journal of Geriatric Psychiatry*, **13**(8), 645–55, copyright 2005, American Association of Geriatric Psychiatry, Lippincott Williams & Wilkins.)

neurological, psychiatric, or medical condition that can account for the impairment.[7]

Age-associated cognitive decline (AACD)

The AACD category covers impairments in any domain affected by ageing, including learning, memory, attention, thinking, language, and visuospatial function. There must be self- or informant-reported cognitive decline over at least 6 months and performance on objective cognitive tests at least 1 SD below age- and education-appropriate norms. The cognitive impairment cannot fulfil dementia criteria and is not accounted for by systemic, neurological, or psychiatric disorders.

Cognitive impairment not dementia (CIND)

CIND includes all individuals that cannot be classified as cognitively normal or as demented. CIND has been applied both in population- and clinic-based studies.[8,9] This diagnostic label is assigned by clinical judgement when there is memory and/or cognitive impairment insufficient to meet DSM criteria for dementia, without exclusions related to underlying aetiologies. There are to date no operational criteria for this category. Because of its inclusiveness CIND encompasses a range of aetiologies that must be disentangled to be clinically meaningful.[8,9]

Amnestic mild cognitive impairment (aMCI)

This amnestic condition is defined as a clinical disorder that describes a transitional state between normal ageing and AD. It is characterized by memory impairment in the context of otherwise preserved abilities. The diagnostic criteria for aMCI require memory complaint preferably corroborated by an informant, objective memory impairment for age and education, largely normal general cognition, essentially intact activities of daily living (ADLs), and the absence of dementia. Objective memory impairment, though not anchored to a specific cut point, is generally ≥1.5 SD below appropriate norms.[10]

Mild cognitive impairment (MCI): international working group criteria

The MCI concept has recently been broadened to encompass multiple patterns of cognitive impairment including amnestic, non-amnestic, single- or multiple-domain deficits.[2] In this framework the classification of MCI requires multiple steps. First, individuals should be judged as neither normal nor demented. Second, there should be evidence of cognitive decline, supported by self and/or informant reports, impairment on objective cognitive tests, or evidence of decline over time on these tests. Third, activities of daily living should be mainly preserved, with the provision that complex ADLs can be minimally impaired.[2] Like CIND, this MCI category recognizes multiple aetiologies underlying impairment, and requires their identification.

Clinical staging scales

Studies of MCI frequently utilize clinical staging scales both to define the inclusion criteria as well as to track outcomes. The Clinical Dementia Rating (CDR) scale[11] distinguishes five stages of dementia severity, with a stage of questionable dementia (CDR 0.5), between the stages of healthy (CDR 0) and mild dementia (CDR 1). CDR 0.5 is most often applied to MCI; however, this stage also can include those with functional impairment who meet dementia criteria. Similarly, the Global Deterioration Scale (GDS),[12] which distinguishes seven stages of impairment, overlaps with MCI at stage 2 (normal with a subjective complaint) or stage 3 (subtle deficits in cognition and occupational/social activities), whereas individuals with mild dementia may receive a GDS stage 3 or 4. The mapping of MCI onto these staging scales has not yet been fully reconciled.

Epidemiology

Prevalence

Prevalence estimates for MCI will naturally vary according to the definition, to age and to the setting. In population-based studies, AAMI has been estimated to affect up to 38.4 per cent, and AACD between 21 and 35.2 per cent, of individuals aged 60 or older.[13] CIND has been reported to affect between 16.8 and 23.4 per cent of individuals aged 65 or older. The prevalence of aMCI has been much lower, at 3 to 6 per cent in similar age groups.[13] Within the broad MCI classification, the multiple domains and single non-memory domain subtypes have been described as roughly twice as frequent as the amnestic subtype, with multiple domain impairment estimated to affect 16 per cent of individuals.[14,15] The prevalence of AACD, aMCI, MCI, and CIND varies with age, with a two- to threefold increase from age 65–74 to >85 in CIND (Fig. 8.5.1.1.3).[8,16] The prevalence of MCI and related conditions within the referral clinic setting is much higher than population estimates.[17]

Natural history

(a) Progression rates to dementia

While the rate of progression to dementia is 1 to 2 per cent per year for cognitively normal individuals aged 65 or older, the rates for all MCI entities are systematically higher and quite variable (Fig. 8.5.1.1.4). Whereas AAMI has low progression rates (1 to 3 per cent per year) and is closest to a normal population, for all other categories most studies report rates between 10 and 15 per cent per

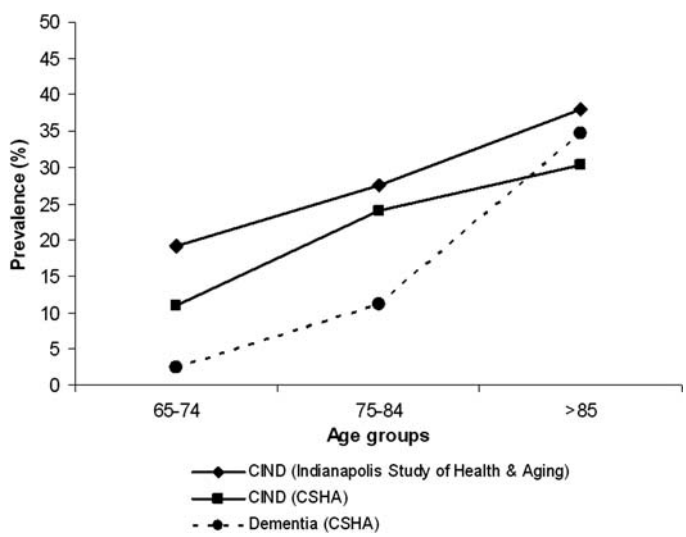


Fig. 8.5.1.1.3 The prevalence of CIND with increasing age reported in the Canadian study of health and ageing (CSHA)[8] and the Indianapolis study of health and ageing.[16]

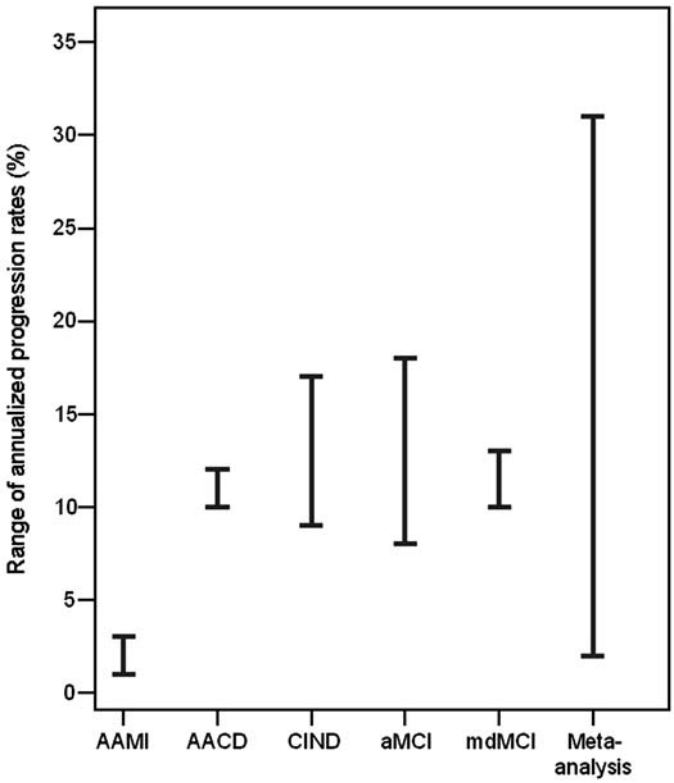


Fig. 8.5.1.1.4 Range of annualized estimates of progression to dementia in various MCI groups, reported in Refs.[13,14,18]

annum, generally over 5-year periods. A meta-analytic study indicated a mean annual conversion rate of 10.24 per cent, with a range of 2 to 31 per cent. The single most important factor accounting for this heterogeneity was the source of the participants, with subjects referred to specialist services, either geriatric or memory/dementia clinics, progressing to dementia at roughly twice the rate found for community volunteers.[18] In studies of CIND annual rates have been 17 per cent for a dementia clinic-referred cohort, 9 per cent for a population-based cohort.[1,9]

An unresolved question is whether all subjects with MCI eventually develop dementia. In a highly selected aMCI sample with CDR 0.5 followed over 10 years, all subjects eventually progressed to meet dementia criteria.[19] This evidence has been proposed to support the belief that aMCI represents early-stage AD.[19] Other studies with similar length of follow-up have not found as high rates of progression and neuropathologic evidence also suggests more heterogeneity, with up to 50 per cent of subjects with aMCI meeting criteria for AD, and the remainder having varied non-AD abnormalities including vascular lesions, argyrophilic grain disease, or other conditions.[20]

(b) Reversion rates to normal

There are consistently proportions of individuals with MCI that will revert back to normal cognitive functioning during follow-up. In population-based studies, this backcrossing has sometimes been particularly high (range 4 to >40 per cent after 1.5 to 5 years).[1] In clinic-referred samples, reversion rates for CIND appear to be lower (14 per cent after 2 years).[9] Reversion may in part be a reflection of an inherent instability in the MCI condition, particularly when it is defined exclusively by psychometric cut points.[1] Reversion appears to be more common when there is no clearly identified aetiology for MCI.[9]

(c) Mortality risk

The mortality risk of MCI is near twice the risk of cognitively normal individuals. The estimated relative risk has been 1.5 for AACD, 1.5–1.9 for CIND, and 1.3–1.7 for aMCI while there has been no increased risk reported for AAMI. There is no clear explanation for this increased mortality risk. It is independent of health conditions such as cardiac disease, cerebrovascular disease, diabetes, and malignancies, and it may be related to incipient and eventually full-blown dementia.[21]

Diagnosis: clinical approach

Overview

Figure 8.5.1.1.5 depicts a flow chart for the clinical diagnosis of MCI.[2] The diagnostic process begins with an expressed concern about cognitive functioning from the patient and/or informant. The assessment then requires a careful history from the patient and informant, as well as the mandatory administration of objective cognitive testing. An evaluation of social function and ADLs, both instrumental and basic, is performed to determine whether there is impairment sufficient for a dementia diagnosis. A clinical judgement must be made of whether the impairment falls outside of the normal range for age and whether it falls within MCI or dementia. The cognitive profile can further be classified into single amnestic, multiple domains, or single non-memory domain. The final step in the diagnostic process is the determination of the aetiology of MCI.

Cognitive assessment

The identification of MCI requires evidence of impairment on objective cognitive testing. Generally, a cognitive screening test is needed as a first step. It should be recognized that the most widely used Mini Mental State Examination (MMSE) lacks sensitivity for MCI diagnosis. In turn, two novel instruments (the Montreal Cognitive Assessment, MoCA, and the DemTect) have been shown to reliably discriminate MCI from normal ageing.[22] The effects of age and educational achievement on test performance and the consequent risk of misclassifications must be kept in mind in the clinical assessment. Irrespective of the instrument that is chosen, there should be coverage of episodic memory, executive functioning and language, which are the most frequent presenting problems in MCI.

Neuropsychological testing (NPT) can be helpful where a clear-cut determination of MCI is difficult following the initial assessment. NPT provides information on the pattern of impairment, with identification of the domains affected and reference to standardized scores. Serial NPT may be particularly useful in defining the progression from MCI to mild dementia. The rate of decline on tests of memory, executive functioning and language accelerates during the 3- to 5-year time window before diagnosis of full-blown dementia.[22] NPT also has some predictive utility in addressing the risk of progression from MCI to dementia. Deficits on tests of episodic memory and executive functioning have consistently been found to characterize those with MCI that will develop dementia.[22]

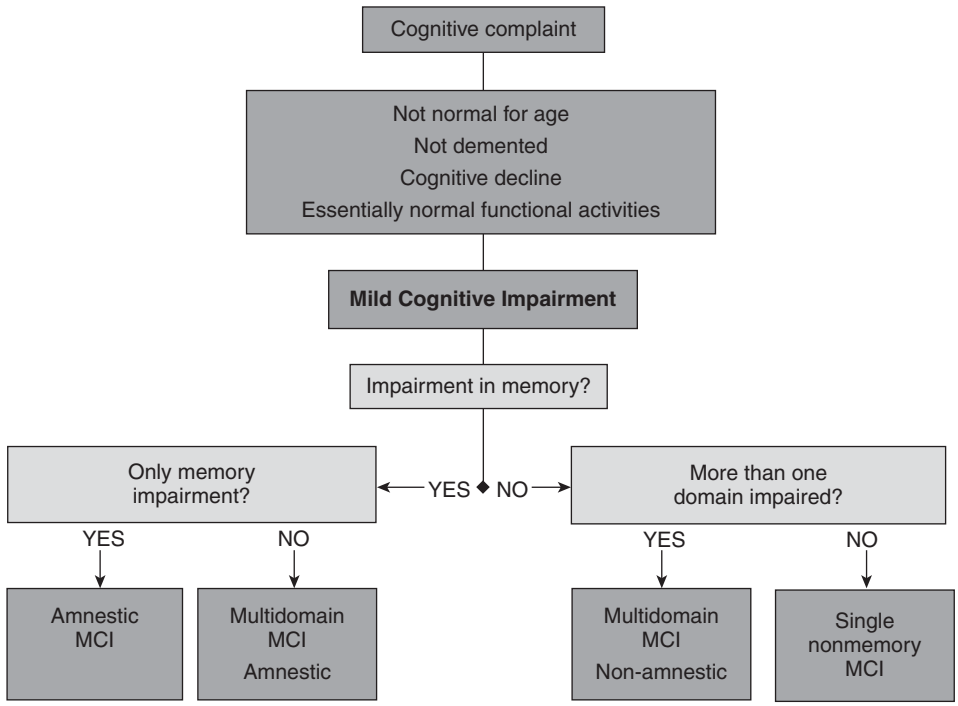


Fig. 8.5.1.1.5 Diagnostic flow chart proposed by the International Working Group on MCI. (Reproduced from B. Winblad *et al.* (2004), Mild cognitive impairment—beyond controversies, towards a consensus: report of the International Working Group on mild cognitive impairment, *Journal of Internal Medicine*, **256**, 240–6, copyright 2004, John Wiley & Sons, Inc.)

Neuropsychiatric symptoms (NPS)

NPS may provide a very useful additional domain for evaluation. An estimated 50 to 70 per cent of subjects with MCI have informant-reported NPS. Symptoms of depression, agitation/aggression, anxiety, apathy, and irritability, while at a low level of intensity, nevertheless each have a frequency of >30 per cent. NPS are associated with greater MCI severity and may be predictive of progression to dementia.[23] The evaluation of NPS is recommended within an MCI assessment.

Aetiology

The determination of aetiology is a final key step in the assessment of MCI. It requires additional laboratory studies as well as consideration of neuroimaging results. Within a cohort of individuals with CIND, the most prevalent aetiologic subtypes were pre-AD, vascular cognitive impairment, cognitive impairment with psychiatric illness, and not otherwise specified (NOS) (Table 8.5.1.1.1). These CIND aetiologic subtypes differed in their functional and psychobehavioural profiles, and in their 2-year prognosis. Pre-AD and vascular CIND had the highest rates of progression to dementia (~40 per cent), with pre-AD subjects developing exclusively probable AD. Psychiatric CIND and CIND NOS had the highest rates of reversion to normal (20 to 30 per cent). Progression to dementia occurred in all aetiologic subtypes (Table 8.5.1.1.1).[9]

Biomarkers

There are no widely accepted MCI biomarkers at the present time. Research has focused on neuroimaging with MRI and FDG-PET as well as on tau and β-amyloid (Aβ) protein levels in cerebrospinal fluid (CSF).

Neuroimaging

On structural MRI, individuals with MCI may show volume loss in the medial temporal lobe (MTL) that may presage AD. The key structures of the MTL include the hippocampus and the entorhinal cortex. MTL structures can be rated on visual scales that may

Table 8.5.1.1.1 Aetiologic subtypes of CIND and their 2-year progression to AD/dementia

Subtype	Prevalence (% within cohort)	Dementia at 2-years (% within subtype)	AD at 2-years (% within dementia)
Pre-AD	24.6	40.9	100
Vascular	18.1	40.0	33.3
Psychiatric	17.3	25.0	75
Non-AD degenerative	2.3	33.3	0
Neurological	7.3	25.0	25
Medical	3.5	60.0	100
Mixed	7.6	33.3	100
Not otherwise specified	19.3	24.2	62.5

(Reproduced from G.Y. Hsiung et al. (2006), Outcomes of cognitively impaired not demented at 2 years in the Canadian cohort study of cognitive impairment and related dementias. *Dementia and Geriatric Cognitive Disorders*, **22**(5–6), 413–20, with permission from S. Karger A.G, Basel.)

reasonably predict AD.[24] MRI-based predictive algorithms may be better when MTL measures are combined with measures of lateral temporal lobe or anterior cingulate structures, or with performance scores on episodic memory.[24] Both whole brain and MTL atrophy rates are greater in MCI subjects that progress to AD than in those who do not.[25]

The characteristic pattern of AD is to have metabolic reductions in temporoparietal and posterior cingulate regions on [^{18}F-]-fluorodeoxyglucose (FDG) PET. This pattern may be seen in individuals with MCI ahead of full-blown disease. Serial PET may show further metabolic deterioration in these areas as well as abnormalities in the ventrolateral prefrontal cortex in subjects who progress to AD.[1] Recently, PET radioligands that bind to cerebral amyloid and potentially to tau proteins have been developed. These include Pittsburgh compound B (PiB) (*N*-methyl-[^{11}C]2-(4′-methylaminophenyl)-6-hydroxybenzothiazole) for amyloid and FDDNP (2-(1-[6-[(2-[^{18}F]fluoroethyl)(methyl)amino]-2-naphthyl]ethylidene)malononitrile) for amyloid and tau. Studies with these ligands have revealed higher than normal retention in subjects with MCI, with a pattern that follows the anatomical distribution of AD pathology.[26,27] If longitudinal follow-up confirms that these MCI cases develop AD, PET with PiB or FDDNP could become a diagnostic test for the early identification of AD.[24]

Cerebrospinal fluid markers

CSF markers can reflect the AD pathogenic process including a reduction in Aβ42 as it aggregates into senile neuritic plaques, as well as an increase in total tau (t-tau) and phospho-tau (p-tau), which signals the hyperphosphorylated state of tau. In MCI, the combination of abnormal Aβ42, t-tau, and p-tau 181 is associated with a 17–20 increased risk of developing AD over 4–6 years. Currently, the utility of these markers is limited by the lack of a standardized assay and the variability in measurements obtained at different laboratories.[28]

Management and treatment

There are no standard therapies for MCI. A clinical management plan is formulated on an individual basis in consideration of the cognitive pattern of MCI and its aetiology. There are two goals for treatment: first, to alleviate the cognitive symptoms of MCI, and second, to attempt to delay the onset of dementia in those at risk.

Symptomatic treatment

(a) Pharmacotherapy

A 24-week trial of the cholinesterase inhibitor (ChEI) donepezil did not benefit subjects with aMCI on its primary endpoints of delayed recall and global impression of change. There were benefits on secondary measures including the ADAS-cog, neuropsychological tests of attention, and patient-rated global function (Patient Global Assessment, PGA).[29]A longer term trial of donepezil with subjects with aMCI (the Alzheimer's Disease Cooperative Study-Memory Impairment Study, ADCS-MIS) demonstrated benefits on the ADAS-cog, memory and language scores, and global measures including the CDR sum of boxes but these were confined to the first 18 months of treatment.[30]

(b) Non-pharmacologic therapies

Cognitive and lifestyle interventions may help the cognitive and behavioural difficulties in MCI. An 8-week cognitive intervention programme to improve memory strategies produced benefits on tests of delayed recall and face-name association, and on self-assessed everyday memory function in subjects with aMCI.[31] A 14-day healthy lifestyle programme with memory training, physical conditioning, relaxation techniques, and a diet plan, showed benefits in subjects with mild self-reported memory complaints. These benefits included improved word fluency and metabolic changes on FDG-PET in dorsolateral prefrontal cortex.[32] Therapeutic approaches of this type require considerable resources, and confirmation in randomized controlled trials (RCTs) is needed before large-scale implementation in MCI can be recommended. Nevertheless these approaches hold some promise.

Delaying the onset of AD

(a) Pharmacotherapy

A considerable number of long and large trials of ChEIs and non-steroidal anti-inflammatory drugs (NSAIDs) have been directed at delaying the time to diagnosis of AD in those with aMCI. In the 3-year ADCS-MIS trial, treatment with donepezil did not have an effect on the primary outcome measure of progression to AD after 36 months. An interesting observation in the study was that carriers of the apolipoprotein E ε4 allele treated with donepezil had a reduced risk of progression to AD at all time points.[30] Similarly, rivastigmine was unsuccessful in delaying the time to diagnosis of AD.[33] A 4-year trial of the COX-2 inhibitor rofecoxib did not show treatment benefits on the primary endpoint of percentage of subjects that were diagnosed with AD nor on secondary measures of cognition and global function.[1] Based on the available data, the long-term use of ChEIs or NSAIDs in MCI, with the goal of delaying AD onset, cannot be recommended.

(b) Non-pharmacologic therapies

An expanding literature supports the hypothesis that a cognitively, socially, and physically active lifestyle in late life may reduce the risk for AD.[34] Evidence from RCTs to support this hypothesis is not currently available and it is premature to recommend organized interventions. There would be little perceived harm in promoting an engaged lifestyle as part of the management of MCI.

The future of MCI

MCI has been a useful construct to focus attention on the cognitive impairment that is going to increase exponentially within our greying societies. It is recognized that MCI is a risk state for developing AD and other dementias. A natural next step will be to develop criteria to diagnose AD earlier. A proposal of research criteria for the early diagnosis of AD has recently been published.[24] The diagnosis builds on a clinical core of early and significant episodic memory impairment and requires in addition the presence of at least one biological footprint of the disease: medial temporal lobe atrophy on structural MRI, abnormal CSF, or reduced temporoparietal glucose metabolism on FDG-PET.[24] Similar frameworks are already in place or will be developed for non-AD dementias. As these frameworks will advance, the concept of MCI will likely be significantly refined and could look quite different by the next edition of this textbook.

Further information

Petersen, R. (ed.) (2003). *Mild cognitive impairment*. Oxford University Press, New York.

Chertkow, H., Nasreddine, Z., Joanette, Y., *et al.* (2007). Mild cognitive impairment and cognitive impairment, no dementia: part A, concept and diagnosis. *Alzheimer's & Dementia*, **3**, 266–82.

Massoud, F., Belleville, S., Bergman, H., *et al.* (2007). Mild cognitive impairment and cognitive impairment, no dementia: part B, therapy. *Alzheimer's & Dementia*, **3**, 283–91.

Dubois, B., Feldman, H.H., Jacova, C., *et al.* (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurology*, **6**(8), 734–46.

References

1. Feldman, H.H. and Jacova, C. (2005). Mild cognitive impairment. *The American Journal of Geriatric Psychiatry*, **13**(8), 645–55.
2. Winblad, B., Palmer, K., Kivipelto, M., *et al.* (2004). Mild cognitive impairment—beyond controversies, towards a consensus: report of the International Working Group on mild cognitive impairment. *Journal of Internal Medicine*, **256**, 240–6.
3. American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (DSM-IV-TR) (4th text revised edn). APA, Washington, DC.
4. World Health Organization. (1992). ICD-10: international statistical classification of diseases and related health problems (10th revision edn): based on recommendations of the Tenth Revision Conference, 1989 and adopted by the Forty-third World Health Assembly.
5. Petersen, R.C., Stevens, J.C., Ganguli, M., *et al.* (2001). Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review). Report of the quality standards subcommittee of the American academy of neurology. *Neurology*, **56**, 1133–42.
6. Chertkow, H., Nasreddine, Z., Joanette, Y., *et al.* (2007). Mild cognitive impairment and cognitive impairment, no dementia: part A, concept and diagnosis. *Alzheimer's & Dementia*, **3**, 266–82.
7. Crook, T., Bartus, R., Ferris, S.H., *et al.* (1986). Age-associated memory impairment: proposed diagnostic criteria and measures of clinical change—report of a National Institute of Mental Health Work Group. *Developmental Neuropsychology*, **2**, 261–76.
8. Graham, J.E., Rockwood, K., Beattie, B.L., *et al.* (1997). Prevalence and severity of cognitive impairment with and without dementia in an elderly population. *Lancet*, **349**, 1793–6.
9. Hsiung, G.Y., Donald, A., Grand, J., *et al.* (2006). Outcomes of cognitively impaired not demented at 2 years in the Canadian cohort study of cognitive impairment and related dementias. *Dementia and Geriatric Cognitive Disorders*, **22**(5–6), 413–20.
10. Petersen, R.C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, **256**(3), 183–94.
11. Morris, J.C. (1997). Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *International Psychogeriatrics*, **9**(Suppl. 1), 173–6.
12. Reisberg, B., Sclan, S.G., Franssen, E. *et al.* (1997). Clinical stages of normal aging and alzheimer's disease: the GDS staging system. *Neuroscience Research Communications*, **13**(Suppl. 1), 51–4.
13. Panza, F., D'Introno, A., Colacicco, A.M., *et al.* (2005). Current epidemiology of mild cognitive impairment and other predementia syndromes. *The American Journal of Geriatric Psychiatry*, **13**(8), 633–44.
14. Busse, A., Bischkopf, J., Riedel-Heller, S.G., *et al.* (2003). Subclassifications for mild cognitive impairment: prevalence and predictive validity. *Psychological Medicine*, **33**(6), 1029–38.
15. Lopez, O.L., Jagust, W.J., DeKosky, S.T., *et al.* (2003). Prevalence and classification of mild cognitive impairment in the cardiovascular health study cognition study: part 1. *Archives of Neurology*, **60**(10), 1385–9.
16. Unverzagt, F., Gao, S., Baiuewu, O., *et al.* (2001). Prevalence of cognitive impairment: data from the Indianapolis study of health and aging. *Neurology*, **57**(9), 1655–62.
17. Feldman, H., Levy, A.R., Hsiung, G.Y., *et al.* (2003). A Canadian cohort study of cognitive impairment and related dementias (ACCORD): study methods and baseline results. *Neuroepidemiology*, **22**, 265–74.
18. Bruscoli, M. and Lovestone, S. (2004). Is MCI really just early dementia? A systematic review of conversion studies. *International Psychogeriatrics*, **16**, 129–40.
19. Morris, J.C., Storandt, M., Miller, J.P., *et al.* (2001). Mild cognitive impairment represents early-stage Alzheimer disease. *Archives of Neurology*, **58**(3), 397–405.
20. Petersen, R.C., Parisi, J.E., Dickson, D.W., *et al.* (2006). Neuropathologic features of amnestic mild cognitive impairment. *Archives of Neurology*, **63**(5), 665–72.
21. Guehne, U., Angermeyer, M.C., and Riedel-Heller, S. (2006). Is mortality increased in mildly cognitively impaired individuals? A systematic literature review. *Dementia and Geriatric Cognitive Disorders*, **21**(5–6), 403–10.
22. Jacova, C., Kertesz, A., Blair, M., *et al.* (2007). Neuropsychological testing and assessment for dementia. *Alzheimer's & Dementia*, **3**, 299–317.
23. Feldman, H., Scheltens, P., Scarpini, E., *et al.* (2004). Behavioral symptoms in mild cognitive impairment. *Neurology*, **62**, 1199–201.
24. Dubois, B., Feldman, H.H., Jacova, C., *et al.* (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurology*, **6**(8), 734–46.
25. Jack, C.R. Jr, Shiung, M.M., Gunter, J.L., *et al.* (2004). Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology*, **62**(4), 591–600.
26. Klunk, W.E., Engler, H., Nordberg, A., *et al.* (2004). Imaging brain amyloid in Alzheimer's disease with Pittsburgh compound-B. *Annals of Neurology*, **55**, 306–19.
27. Small, G.W., Kepe, V., Ercoli, L.M., *et al.* (2006). PET of brain amyloid and tau in mild cognitive impairment. *The New England Journal of Medicine*, **355**(25), 2652–63.
28. Hansson, O., Zetterberg, H., Buchhave, P., *et al.* (2006). Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study. *Lancet Neurology*, **5**(3), 228–34.
29. Salloway, S., Ferris, S., Kluger, A., *et al.* (2004). Efficacy of donepezil in mild cognitive impairment: a randomized placebo-controlled trial. *Neurology*, **63**, 651–7.
30. Petersen, R.C., Thomas, R.G., Grundman, M., *et al.* (2005). Vitamin E and donepezil for the treatment of mild cognitive impairment. *The New England Journal of Medicine*, **352**(23), 2379–88.
31. Belleville, S., Gilbert, B., Fontaine, F., *et al.* (2006). Improvement of episodic memory in persons with mild cognitive impairment and healthy older adults: evidence from a cognitive intervention program. *Dementia and Geriatric Cognitive Disorders*, **22**(5–6), 486–99.
32. Small, G.W., Silverman, D.H., Siddarth, P., *et al.* (2006). Effects of a 14-day healthy longevity lifestyle program on cognition and brain function. *The American Journal of Geriatric Psychiatry*, **14**(6), 538–45.
33. Feldman, H.H., Ferris, S., Winblad, B., *et al.* (2007). Effect of rivastigmine on delay to diagnosis of Alzheimer's disease from mild cognitive impairment: the InDDEx study. *Lancet Neurology*, **6**(6), 501–12.
34. Fratiglioni, L., Paillard-Borg, S., and Winblad, B. (2004). An active and socially integrated lifestyle in late life might protect against dementia. *Lancet Neurology*, **3**, 343–53.

8.5.2 Substance use disorders in older people

Henry O'Connell and Brian Lawlor

Introduction

This chapter is divided into three main sections, focussing respectively on alcohol use disorders (AUDs), medication use disorders (MUDs), and use of illegal substances and nicotine, in older people. In each section we focus in detail on definitions and diagnosis, epidemiology, aetiology, clinical features, investigations, screening, management, and prognosis. More is known about AUDs in older people, hence this section is the longest, but MUDs in older people is also a significant problem and abuse of illegal drugs may become increasingly important in future years.

Alcohol use disorders (AUDs) in older people

Introduction

The ageing of populations worldwide means that the already significant problem of alcohol use disorders (AUDs) in older people is likely to become even more important in future years. However, AUDs in older people are neglected and underdiagnosed, for the reasons outlined in Table 8.5.2.1, and unless these factors are tackled proactively there exists a real danger of AUDs in older people becoming a silent epidemic, with negative impacts on all aspects of health and well-being[1] (see Table 8.5.2.2).

Definitions and diagnosis

AUD is a general and broad term, used to include a wide range of alcohol-related problems, as outlined in Table 8.5.2.3. Alcohol status may also change throughout life, with one-third of older people with AUDs developing such problems for the first time in later life (late-onset AUDs). A more severe course of AUD, higher levels of antisocial personality and stronger family histories of AUDs are seen in those with early-onset AUDs.

Because of the effects of physical and cognitive ageing, pharmacokinetic changes, the increased prevalence of comorbid illness, and interactions with prescribed medication, older people are likely to encounter AUDs at levels of intake lower than the general population. Therefore, the recommended levels of intake for the general population (i.e. up to 21 and 14 units per week for men and women, respectively[2]) may be inappropriately high for older people. However, apart from the NIAAA recommendations of no more than one drink per day for older people,[3] there is a lack of guidance on safe levels of alcohol intake, and the pursuit of obvious and 'down and out' drinkers may lead to a significant amount of more subtle and clinically 'silent' AUDs being missed.

Furthermore, the diagnostic criteria used by ICD-10 and DSM-IV used to describe harmful use and alcohol dependence syndrome may not be applicable to older people, as evidence of diagnostic criteria such as craving, compulsion, tolerance, and withdrawal features may be less clear-cut and masked by other medical conditions, and older people may be less likely to encounter the financial, occupational, family and legal consequences of AUDs (see also Table 8.5.2.1 and section on screening).

Table 8.5.2.1 Reasons for the neglect and underdiagnosis of AUDs in older people

Patient factors
Older people may be less likely to volunteer information on alcohol intake/AUDs
Recall of alcohol intake may be inaccurate due to cognitive impairment
Features of AUDs may be atypical or masked (e.g. presenting as falls, confusion)
Pharmacokinetic changes, comorbid illness, and drug-interactions mean alcohol-related problems may arise even at relatively low levels of intake
Health service factors
Health care professionals less likely to ask older people about alcohol intake and AUDs
Health care professionals less likely to refer older people for treatment, even when AUD detected
Inappropriate screening and diagnostic tools used
Therapeutic pessimism in treating older people
Inappropriately high levels for 'recommended' or 'healthy' levels of intake
Family and societal factors
Family members may be less likely to perceive AUD as a problem in older relatives
Ageist attitudes lead to risk of AUDs in older people being perceived as 'understandable'
AUDs in older people less 'noisy', with less impact on absenteeism, antisocial behaviour, crime

Epidemiology

The prevalence of AUDs in older people varies depending on the screening and diagnostic criteria used, clinical and socio-demographic characteristics (men having levels 4–6 times higher than women) and the level of severity of AUD being defined. In community-based studies, for example, 2–4 per cent of older people have been estimated to have alcohol misuse or dependence,[4] with higher rates of 16 per cent (men) and 2 per cent (women) when looser criteria such as excessive alcohol consumption are used.[5] Clinical populations of older people have higher levels of AUDs, with emergency department, nursing home and psychiatric inpatients being described as having levels of 14, 18, and 23 per cent, respectively.[6–8]

The true prevalence of AUDs in older people is often underestimated, for the reasons outlined in Table 8.5.2.1. It is likely, however, that the actual levels of alcohol consumption and AUDs do decline with age.[9,10] This decline may be due to factors such as premature death of those with AUDs, reduced physiological reserve and comorbid medical illness leading to reduced alcohol intake, age-cohort effects and age-related changes in social networks, occupational, and financial status.

Aetiology, risk factors, and associations

These factors can be broadly described as being biological/medical, social, and psychological in nature. Genetic factors are likely to be important in relation to both early-onset[11] and late-onset AUDs.[12] The genetic risk for AUDs may also overlap with risk for other mental disorders such as antisocial personality disorder, other drug use problems, anxiety disorders, and mood disorders.[13] AUDs may have a cause and effect relationship with medical illness.

Important social aetiological factors are likely to include male gender, bereavement, age-cohort effects, culture and ethnicity, religion, and marital status (higher levels of AUDs in divorced and single). Some social factors, such as marital problems, may have a two-way relationship with AUDs.

Table 8.5.2.2 Physical, neuropsychiatric, and socio-demographic aspects of AUDs in older people

1. Physical factors
Gastrointestinal
Hepatic problems: elevated liver enzymes; fatty liver; alcoholic hepatitis; cirrhosis; malignancy
Gastritis, peptic ulcer disease, and bleeding
Oesophageal varices
Acute and chronic pancreatitis
Malignancies
Mouth, pharynx, larynx, oesophagus, hepatic, colorectal, pancreatic
Cardiovascular
Ischaemic heart disease
Hypertension
Alcohol-induced arrhythmias
Congestive heart failure
Alcoholic cardiomyopathy
Haematological
Macrocytosis (acute effect of alcohol intake and due to vitamin B12 and folate deficiency in chronic AUD)
Anaemia (due to gastrointestinal problems)
Musculoskeletal
Falls and fractures
Reduced bone density
Myopathy
Metabolic
Hypoglycaemia
Hyperuricaemia
Elevated lipids
Diabetes more difficult to control
2. Neuropsychiatric factors
Cognitive impairment and dementia
Frontal lobe impairment
Wernicke–Korsakoff syndrome
Cerebellar cortical degeneration
Central pontine myelinosis
Marchiafava–Bignami disease
Depression
Psychosis
Intoxication
Withdrawal syndrome (may be more difficult to treat in older people)
Suicide
3. Socio-demographic
Male gender
Divorced, widowed, and single status
Social isolation
Upper and lower ends of socio-economic spectrum
4. Other
Alcohol–drug interactions
Aspiration pneumonia
Road traffic and other accidents

Relevant personality factors include the stronger association between antisocial personality, hyperactivity, and impulsivity in 'early-onset' compared to late-onset AUDs, who may have higher levels of 'neuroticism' and depression.[14]

AUDs in older people, as in all populations, may also have a two-way relationship with psychiatric disorders such as depression and anxiety disorders. For example, an older person may begin drinking in an effort to self-medicate depressive symptoms, or they may become depressed because of their drinking.[15]

Table 8.5.2.3 Types/levels of severity of AUDs

- Excessive alcohol consumption (i.e. drinking above recommended levels)
- Binge drinking (i.e. episodic bouts of excessive alcohol intake)
- Problem drinkers/harmful use/abuse
- Alcohol dependence syndrome
- 'Early-onset' versus 'late-onset'

Clinical features and comorbidity

AUDs in older people are linked to significant morbidity and mortality, affecting practically all aspects of physical, neuropsychiatric and social health and well-being,[1,16] as summarized in Table 8.5.2.2.

Pharmacokinetic changes (reduced physiological reserve, reduced metabolic efficiency, and increased volume of distribution due to a higher fat to lean muscle ratio, leading in turn to relatively higher blood alcohol concentrations in older people) along with the general effects of physical and cognitive ageing, increasing frailty, reduced functional ability, and higher levels of concomitant prescription drug use means that alcohol is relatively more toxic to older people than younger people. Furthermore, as outlined earlier, such toxic effects may be subtle and may be missed or mistaken for other conditions.

AUDs in older people are associated with a wide range of mental disorders, such as depression, psychosis, withdrawal syndromes, cognitive impairment, and dementia[17] (see Table 8.5.2.2) and are also associated with an increased risk of suicide.[18] The relationship between alcohol use and brain damage and dementia is complex,[19] in that AUDs may increase the risk for different types of dementia[20] and there also exist diagnostic entities known as 'amnesic syndrome associated with alcohol use' (ICD-10) or 'alcohol-induced persisting dementia' (DSM-IV). In contrast, light to moderate alcohol use may protect against dementia.[21]

(a) Clinical assessment

The assessment of AUDs in older people begins with a thorough clinical interview and history of alcohol use (quantity and frequency of drinking, beverage type, drinking context), mental state examination, physical examination and collateral history if available, and with the patient's consent. If indicated by the initial history, additional questions should be asked about features of alcohol dependence syndrome, as they relate to the patient's physical and psychosocial health. Questions should be framed in a sensitive and non-judgemental way, as patients may disengage and be lost to treatment and follow-up if they feel threatened by the assessment procedure.

(b) Further investigations in AUDs

Following a detailed history and examination, other investigations may be indicated and should be directed by the patient's clinical status. These may include: blood tests to check the following: urea and electrolytes; full blood count; liver function tests; vitamin B12 and folate levels; neuroimaging (CT or MRI brain); gastrointestinal investigations such as ultrasound, CT, or MRI examinations of the abdomen, upper gastrointestinal endoscopy, and liver biopsy; basic cardiovascular investigations such as electrocardiogram and other

more detailed investigations if indicated, e.g. echocardiogram and 24 h blood pressure monitoring.

(c) Screening for AUDs in older people

Screening programmes should aim to detect both clear-cut and subtle cases of AUDs in older people. It must be remembered that screening tests are not diagnostic in themselves, but positive results should lead on to further investigations. Screening methods may be based on self-report alcohol screening instruments such as the CAGE,[22] the AUDIT,[23] and biophysical measures such as blood tests checking mean corpuscular volume and liver function tests.

A systematic review of self-report alcohol screening instruments in older people[24] revealed that the CAGE was the most widely studied, but that sensitivity and specificity varied depending on the clinical characteristics of the population in question. However, the CAGE is the most well recognized alcohol screening instrument and is quickly and easily administered, so the authors would recommend use of at least this instrument, along with further investigations and assessment scales if problems are detected.

The utility of biophysical screening measures such as carbohydrate-deficient transferrin, liver function tests or the mean corpuscular volume may be less reliable in older people,[25] because of higher levels of comorbid physical illnesses leading in themselves to false-positive and abnormal results. However, they may prove useful when combined with other clinical information, both in the detection of AUDs and monitoring of progress through treatment.

Management and prevention

Primary prevention strategies should focus on individual older people (especially considering that one-third of older people have late-onset AUDs) and can also be directed at the entire population, targeting factors such as ease of access to alcohol, restrictions on alcohol advertising, and education about the adverse effects of drinking. Such primary prevention and public health initiatives tend to be directed towards younger individuals, but they should also take into account the more clinically 'silent' AUDs that may develop in older people.[26]

Secondary prevention strategies should focus on older people who already have 'at-risk' drinking, either currently or in the past, and who are at risk of developing worsening problems in the context of diverse factors such as bereavement, social isolation, adjustment to retirement, and physical or psychiatric health problems.

Tertiary prevention involves treatment of existing AUDs. Treatment modalities can be divided into biological/medical, social, and psychological. Biological/medical treatments are most important in the acute setting, where detoxification may be required.

Care should be taken with benzodiazepine-assisted withdrawal in older people, in view of the elevated risk of oversedation, confusion, and falls. There are no elderly-specific guidelines on benzodiazepine-assisted alcohol withdrawal in older people. Lorazepam has been identified in one review[27] as the safest choice of benzodiazepine for treatment of alcohol withdrawal in older people, in view of the fact that advancing age and liver disease have little impact on its metabolism, and absorption by the intramuscular route is predictable. In practice, however, it is likely that there is more clinical experience with use of long-acting benzodiazepines such as chlordiazepoxide. The choice of benzodiazepine used should be based on individual patient characteristics such as previous treatments, current medical status (e.g. degree of hepatic impairment) and an objective measure of withdrawal may help guide the dosing regimen (see Table 8.5.2.4).

Parenteral or oral thiamine should be given to prevent development of the Wernicke–Korsakoff syndrome. A recent review has concluded that, in the emergency department setting, oral thiamine administration is as effective as parenteral administration.[28] Again, however, there are no elderly-specific guidelines, and individual patient characteristics must be taken into account, such as general health, ability to take oral medication, and compliance.

The three medications that are approved by the US Food and Drug Administration to promote abstinence and reduce relapse are Disulfiram, Acamprosate, and Naltrexone.[29] However, the limited efficacy of Disulfiram, combined with the potential for a more severe side-effect profile means it is best avoided in this age group. In contrast, Naltrexone and Acamprosate have been suggested as suitable agents for use in older people.[30]

Psychosocial aspects of treatment should also be explored. This may include addressing social circumstances that may be contributing to the AUD (e.g. personal finances and housing). There is a dearth of evidence on psychotherapeutic approaches to AUDs in older people, but there is some evidence that older people may respond better to psychotherapy in same-age settings,[31,32] and consideration should also be given to support groups such as Alcoholics Anonymous.

Prognosis

The available literature on the topic suggests that older people are at least as likely, if not more likely, to benefit from treatment of AUDs as younger people.[33,34] However, prognosis in older people is likely to vary widely depending on a number of factors relating to the individual themselves and the nature of their AUD, the presence of family and other support systems and the availability of treatment services, particularly services that are tailored to older people.

Table 8.5.2.4 The Clinical Institute Withdrawal Assessment for Alcohol-Revised Version (CIWA-Ar) (Reproduced from The South London and Maudsley NHS Trust Prescribing Guidelines, 2005–2006, copyright South London and Mandsley NHS Foundation Trust)

1. Nausea and vomiting
2. Tremor
3. Paroxysmal sweats
4. Anxiety
5. Agitation
6. Tactile disturbances
7. Auditory disturbances
8. Visual disturbances
9. Headaches and fullness in head
10. Orientation and clouding of sensorium

Severity of alcohol withdrawal	
Mild:	<10
Moderate:	10–20
Severe:	20+

(Items 1–9 are scored from 0–7 and item 10 from 0–4. Maximum possible score is 67)

Medication use disorders (MUDs) in older people

Introduction

High levels of prescribing of all types of medications for older people, which may at times be inappropriate, along with factors such as variable compliance, altered pharmacokinetics, reduced functional ability, and increased levels of physical, psychiatric, and cognitive morbidity mean that older people are at higher risk of developing MUDs than any other age group.[35] As with AUDs, clinical features of MUDs may be atypical and masked by other conditions and thus go undetected and untreated.[36]

Definitions

As with AUDs, older people are affected by a wide range of types and severity of MUD.

The ICD-10 uses the same general principles of intoxication, harmful use, dependence, and withdrawal state that apply to alcohol for use of sedative and hypnotic medications. As with AUDs, elderly-specific criteria are not cited, but the same general principles apply: older people are likely to experience harm at lower levels of use and clinical features guiding diagnosis are more likely to be atypical and masked by other health problems.

Iatrogenic factors are also important, as drugs may be inadequately or underused for treating or preventing conditions, or drugs may be overused, leading to unnecessary exposure of the older individual to adverse effects.[36]

Epidemiology

Older people comprise 13 per cent of the US population, but they have been estimated to use more than 30 per cent of prescription[37,38] and 35 per cent of OTC drugs:[37] it has been estimated that older people use prescription and OTC medications approximately three times as much as the general population. Furthermore, we know that the risk of MUDs increases with polypharmacy, which is common in older people.[39,40]

Benzodiazepines are the most commonly prescribed psychotropic drugs in older people, with one study of community-dwelling older people in Ireland demonstrating that 17 per cent of participants were prescribed benzodiazepines, with use in females being twice that in males, and 18 per cent of benzodiazepine users taking at least one other psychotropic drug. Furthermore, 52 per cent of benzodiazepine users were prescribed a long-acting benzodiazepine.[41] It has also been reported that depression in older community-dwelling people is more likely to be detected if accompanied by anxiety symptoms, and such individuals are at risk of inappropriate treatment with benzodiazepines.[42]

Use of opiate analgesia is common in older people and is liable to give rise to MUDs. Therefore, use of these medications should be carefully monitored, with due consideration of dose and careful tapering.[43]

Aetiology, risk factors, and associations

The general principles for aetiology, risk factors, and associations for substance misuse outlined above (see Table 8.5.2.5) also apply to MUDs. Further MUD-specific factors are outlined below in Table 8.5.2.6.

Table 8.5.2.5 Risk factors for substance abuse in the elderly (Reproduced from R.M. Atkinson (2002), Substance abuse in the elderly, In *Psychiatry in the elderly* (3rd edn.) (eds. R. Jacoby and C. Oppenheimer), copyright 2002, with permission from Oxford University Press).

Predisposing factors
Family history (alcohol)
Previous substance abuse
Previous pattern of substance consumption (individual and cohort effects)
Personality traits (sedative–hypnotics, anxiolytics)
Factors that may increase substance exposure and consumption level
Gender (men-alcohol, illicit drugs; women-sedative–hypnotics, anxiolytics)
Chronic illness associated with pain (opioid analgesics), insomnia (hypnotic drugs), or anxiety (anxiolytic)
Long-term prescribing (sedative–hypnotics, anxiolytics)
Caregiver overuse of 'as needed', medication (institutionalized elderly)
Life stress, loss, social isolation
Negative affects (depression, grief, demoralization, anger) (alcohol)
Family collusion and drinking partners (alcohol)
Discretionary time, money (alcohol)
Factors that may increase the effects and abuse potential of substances
Age-associated drug sensitivity (pharmacokinetic, pharmacodynamic factors)
Chronic medical illnesses
Other medications (alcohol–drug, drug–drug interactions)

Clinical features and comorbidity

Clinical features and comorbidities associated with MUDs in older people will vary widely depending on the drug being used and patient characteristics such as age, gender, and presence of other physical and neuropsychiatric problems. An outline of clinical features and comorbidities are listed in Table 8.5.2.7.

(a) Clinical assessment

As with AUDs, a standard clinical assessment involving a history, mental state, and physical examinations and collateral history will form the basis of an MUD assessment. A list of all prescribed and over the counter medications being used, along with their indications for use, should be recorded. Ideally, the patient should be asked to bring with them all medications in their containers, as this will also give an indication as to levels of adherence or compliance. Any reported adverse effects should be recorded, along with

Table 8.5.2.6 Aetiology, risk factors, and associations of MUDs in older people

Biological/medical factors
Genetic predisposition
Chronic medical conditions (e.g. pain)
Age-related pharmacokinetic changes
Interactions: other medications and alcohol
Type of medication (e.g. benodiazepines, analgesics)
Psychosocial factors
Depression
Anxiety disorders
Personality disorder
Older age
Female gender
Lower educational level
Separated or divorced status

Table 8.5.2.7 Clinical features and comorbidity associated with MUDs in older people

Neuropsychiatric (all psychotropic drugs; benzodiazepines may be particularly problematic)
Delirium
Daytime drowsiness
Sleep disturbance
Depression
Anxiety
Physical
Falls
Fractures
Drug–drug and drug–alcohol interactions
Problems related to drug metabolism (e.g. renal and hepatic impairment)

symptoms and signs indicating underuse, overuse, or intermittent use of medication.

(b) Investigations in MUDs

Blood levels of the patient on some prescribed medications may be checked in order to assess levels of compliance and to establish if the blood level is within the therapeutic window for the drug in question (e.g. lithium, carbamazepine). Other biophysical measures may also be indicated that provide proxy measures of medication compliance, such as random or fasting glucose levels and levels of glycosylated haemoglobin, to assess for level of diabetes control and compliance with hypoglycaemic agents or insulin.

(c) Screening

There are no routinely used screening measures for MUDs in older people. However, use of a measure such as Beers' criteria[44,45] may be a useful addition to the overall assessment of an older person if an MUD is suspected.

Management and prevention

(a) Primary and secondary prevention

Along with patients themselves, health care workers, family members, and carers all have important roles in the primary and secondary prevention of MUDs in older people. Prescriptions should be reviewed regularly with a view to simplification and rationalization if possible, and the practice of giving 'repeat prescriptions' without clinical assessment should be discouraged.

Community pharmacists have an important role in providing education and advice on the use of both prescription and over the counter medications.

As older people may have physical and cognitive disabilities that interfere with appropriate use of medication,[46] devices such as dosette boxes and combination packs may be helpful.[47,48] Secondary prevention of MUDs in older people should focus on those with a past history of MUD.

(b) Tertiary prevention

Tertiary prevention of MUDs in older people will depend on the medication in question and the clinical and socio-demographic profile of the patient. Admission to a medical or psychiatric ward may be required to facilitate reduction or stopping of certain medications, e.g. benzodiazepine detoxification, as outpatient detoxification in older people may be hazardous.

Prognosis in MUDs

Similar prognostic indicators that apply to AUDs are likely to be relevant to MUDs, and centre on the individual's clinical and socio-demographic characteristics, levels of support, and available services. The duration of abuse and the medication or medications being abused is also of relevance.

Illicit drug use and nicotine use in older people

Illicit drug use in older people is far less of a problem in comparison to AUDs and MUDs. Lifetime prevalence rates for illicit drug dependence have been estimated as 17 per cent for 18–29 year olds, 4 per cent for 30–59 year olds, and less than 1 per cent for those over the age of 60.[49] Epidemiological Catchment Area data suggest a lifetime prevalence rate for illegal drug use of only 1.6 per cent for older people.[50]

Several other sources of data suggest similarly low rates of illegal drug use among older people. However, the ageing of the 'Baby-boomer' generation is likely to result in a cohort of older people who are healthier and have higher life expectancies than previous generations of older people, but who also carry with them higher rates of illegal drug use.[51]

Principles similar to those seen with AUDs and MUDs apply, in that lower levels of drug intake are required to cause harm and presentation may be atypical and thus go undetected. There is a dearth of evidence in the literature on detoxification and opiate replacement therapies in older populations.

Nicotine use (primarily through cigarette smoking) in older people arguably causes more significant morbidity and mortality than AUDs and MUDs, but the problem tends not to be addressed by psychiatrists, because of a lack of significant neuropsychiatric effects of nicotine use, and a more obvious impact on many aspects of physical health. As with AUDs and MUDs, smoking in older people is treatable, and any comprehensive approach to improving the health of older people, at the individual clinical level or public health level, should involve education about the adverse effects of smoking and efforts at active treatment through the use of nicotine replacement therapies[52] or antidepressants such as nortriptyline or buproprion.[53] However, an important and circular relationship has been described between depression, smoking, and medical illness that complicates smoking cessation in those who have a history of depression.[54]

Conclusions

In this chapter we have highlighted the importance of AUDs and MUDs in older people, in terms of their prevalence and their important but often underrecognized contribution to morbidity and mortality. AUDs are underdetected, misdiagnosed, and often completely missed in older populations. However, despite ageist and therapeutically pessimistic assumptions, AUDs in older people are as amenable to treatment as in younger people, and treating an AUD in an individual of any age can lead to significant benefits in their quality of life.

Likewise, the wide variety of MUDs in older people may be associated with addiction to medication and the undertreatment and inappropriate treatment of medical and psychiatric conditions. Considering that older people are the highest consumers of

prescription medications, screening and treatment programmes for MUDs should also lead to considerable improvements in quality of life, along with financial and other savings.

Misuse of illicit drugs by older people is not generally a major problem at present, but it is virtually certain that consumption of illegal substances by people over 65 will increase in the future.

Greater awareness amongst physicians and other health care providers of the possibility of AUDs and MUDs in their older patients should lead to the development of more comprehensive and age-appropriate prevention and treatment strategies. At the levels of everyday clinical practice and public health policy, greater emphasis should be placed on AUDs and MUDs in older people and further evaluation of dedicated 'same-age' treatment services and settings should be performed.

Further information

Website of National Institute on Alcohol Abuse and Alcoholism (NIAAA): www.niaaa.nih.gov

Website of Royal College of Psychiatrists: www.rcpsych.ac.uk

O'Connell, H., Chin, A.V., Cunningham, C., *et al.* (2003a) Alcohol use disorders in elderly people—redefining an age old problem in old age. *British Medical Journal*, **327**, 664–7; www.bmj.com.

Gurnack, A.M., Roland, R., Atkinson, M.D., *et al.* (eds.) (2001). *Treating alcohol and drug abuse in the elderly*. Springer, New York.

Barry, K.L., Oslin, D.W., and Blow, F.C. (eds.) (2001). *Alcohol problems in older adults: prevention and management*. Springer, New York.

References

1. O'Connell, H., Chin, A.V., Cunningham, C., *et al.* (2003a). Alcohol use disorders in elderly people—redefining an age old problem in old age. *British Medical Journal*, **327**, 664–7.
2. The Royal Colleges Report. (1995). Alcohol and the heart in perspective: sensible limits reaffirmed. A working group of the Royal Colleges of Physicians, Psychiatrists and General Practitioners. *Journal of the Royal College of Physicians of London*, **29**, 266–71.
3. National Institute on Alcohol Abuse and Alcoholism. (1998). *Alcohol alert No. 40*. NIAAA, Bethesda, MD. www.niaaa.nih.gov/publications/aa40.htm
4. Adams, W.L. and Cox, N.S. (1995). Epidemiology of problem drinking among elderly people. *International Journal of the Addictions*, **30**, 1693–716.
5. Greene, E., Bruce, I., Cunningham, C., *et al.* (2003). Self-reported alcohol consumption in the Irish community dwelling elderly. *Irish Journal of Psychological Medicine*, **20**, 77–9.
6. Adams, W.L., Magruder-Habib, K., Trued, S., *et al.* (1992). Alcohol abuse in elderly emergency department patients. *Journal of the American Geriatrics Society*, **40**, 1236–40.
7. Joseph, C.L., Ganzini, L., and Atkinson, R.M. (1995). Screening for alcohol use disorders in the nursing home. *Journal of the American Geriatrics Society*, **43**, 368–73.
8. Speer, D.C. and Bates, K. (1992). Comorbid mental and substance disorders among older psychiatric patients. *Journal of the American Geriatrics Society*, **40**, 886–90.
9. Adams, W.L., Garry, P.J., Rhyne, R., *et al.* (1990). Alcohol intake in the healthy elderly. *Journal of the American Geriatrics Society*, **38**, 211–6.
10. Temple, M.T. and Leino, E.V. (1989). Long-term outcomes of drinking: a twenty year longitudinal study of man. *British Journal of Addiction*, **84**, 889–93.
11. Dahmen, N., Volp, M., Singer, P., *et al.* (2005). Tyrosine hydroxylase Val-81-Met polymorphism associated with early-onset alcoholism. *Psychiatric Genetics*, **15**, 13–6.
12. Tiihonen, J., Hammikainen, T., Lachman, H., *et al.* (1999). Association between the functional variant of the catechol-*O*-methyltransferase (COMT) gene and type 1 alcoholism. *Molecular Psychiatry*, **4**, 286–9.
13. Nurnberger, J.I., Wiegand, R., Bucholz, K., *et al.* (2004). A family study of alcohol dependence: coaggregation of multiple disorders in relatives of alcohol-dependent probands. *Archives of General Psychiatry*, **61**, 1246–56.
14. Mulder, R.T. (2002). Alcoholism and personality. *The Australian and New Zealand Journal of Psychiatry*, **36**, 44–52.
15. Davidson, K.M. and Ritson, E.B. (1993). The relationship between alcohol dependence and depression. *Alcohol and Alcoholism*, **28**, 147–55.
16. Reid, M.C. and Anderson, P.A. (1997). Geriatric substance use disorders. *The Medical Clinics of North America*, **81**, 999–1016.
17. Saunders, P.A., Copeland, J.R., Dewey, M.E., *et al.* (1991). Heavy drinking as a risk factor for depression and dementia in elderly men. Findings from the Liverpool longitudinal community study. *The British Journal of Psychiatry*, **159**, 213–6.
18. Waern, M. (2003). Alcohol dependence and misuse in elderly suicides. *Alcohol and Alcoholism*, **38**, 249–54.
19. Neiman, J. (1998). Alcohol as a risk factor for brain damage: neurologic aspects. *Alcoholism, Clinical and Experimental Research*, **22**(Suppl. 7), 346S–51S.
20. Letenneur, L. (2004). Risk of dementia and alcohol and wine consumption: a review of recent results. *Biological Research*, **37**, 189–93.
21. Ruitenberg, A., van Sweiten, J.C., Witteman, J.C.M., *et al.* (2002). Alcohol consumption and risk of dementia: the Rotterdam study. *Lancet*, **359**, 281–6.
22. Ewing, J.A. (1984). Detecting alcoholism: the CAGE questionnaire. *The Journal of the American Medical Association*, **252**, 1905–7.
23. Saunders, J.B. (1993). Development of the alcohol use disorders identification test (AUDIT). *Addiction*, **88**, 791–804.
24. O'Connell, H., Chin, A.V., Hamilton, F., *et al.* (2004). A systematic review of the utility of self-report alcohol screening instruments in the elderly. *International Journal of Geriatric Psychiatry*, **19**, 1074–86.
25. Luttrell, S., Watkin, V., Livingston, G., *et al.* (1997). Screening for alcohol misuse in older people. *International Journal of Geriatric Psychiatry*, **12**, 1151–4.
26. O'Connell, H., Chin, A.V., and Lawlor, B.A. (2003b). Alcohol use in Ireland—can we hold our drink?' *Irish Journal of Psychological Medicine*, **20**, 109–10.
27. Peppers, M.P. (1996). Benzodiazepines for alcohol withdrawal in the elderly and in patients with liver disease. *Pharmacotherapy*, **16**, 49–57.
28. Jackson, R. and Teece, S. (2004). Oral or intravenous thiamine in the emergency department. *Emergency Medicine Journal*, **21**, 501–2.
29. Williams, S.H. (2005). Medications for treating alcohol dependence. *American Family Physician*, **72**, 1775–80.
30. Barrick, C. and Connors, G.J. (2002). Relapse prevention and maintaining abstinence in older adults with alcohol use disorders. *Drugs & Aging*, **19**, 583–94.
31. Schonfeld, L. and Dupree, L.W. (1995). Treatment approaches for older problem drinkers. *International Journal of the Addictions*, **30**, 1819–42.
32. Kofoed, L.L., Tolson, R.L., Atkinson, R.M., *et al.* (1987). Treatment compliance of older alcoholics: an elder-specific approach is superior to "mainstreaming". *Journal of Studies on Alcohol*, **48**, 47.
33. Curtis, J.R., Geller, G., Stokes, E.J., *et al.* (1989). Characteristics, diagnosis and treatment of alcoholism in elderly patients. *Journal of the American Geriatrics Society*, **37**, 310–6.
34. Oslin, D.W., Pettinati, H., and Volpicelli, J.R. (2002). Alcoholism treatment adherence: older age predicts better adherence and drinking outcomes. *The American Journal of Geriatric Psychiatry*, **10**, 740–7.

35. Chutka, D.S., Takahashi, P.Y., and Hoel, R.W. (2004). Inappropriate medications for elderly patients. *Mayo Clinic Proceedings. Mayo Clinic*, **79**, 122–39.
36. Beers, M.H., Baran, R.W., and Frenia, K. (2000). Drugs and the elderly. Part 1: the problems facing managed care. *The American Journal of Managed Care*, **6**, 1313–20.
37. Williams, L. and Lowenthal, D.T. (1992). Drug therapy in the elderly. *The Southern Medical Journal*, **85**, 127–31.
38. Avorn, J. (1995). Medication use and the elderly: current status and opportunities. *Health Affairs*, **14**, 276–86.
39. Chrischilles, E.A., Segar, E.T., and Wallace, R.B. (1992). Self-reported adverse drug reactions and related resource use. A study of community-dwelling persons 65 years of age and older. *Annals of Internal Medicine*, **117**, 634–40.
40. Cadieux, R.J. (1989). Drug interactions in the elderly: how multiple drug use increases risk exponentially. *Postgraduate Medicine*, **86**, 179–86.
41. Kirby, M., Denihan, A., Bruce, I., *et al.* (1999a). Benzodiazepine use among the elderly in the community. *International Journal of Geriatric Psychiatry*, **14**, 280–4.
42. Kirby, M., Denihan, A., Bruce, I., *et al.* (1999b). Influence of anxiety on treatment of depression in later life in primary care: questionnaire survey. *British Medical Journal*, **318**, 579–80.
43. Schneider, J.P. (2005). Chronic pain management in older adults: with coxibs under fire, what now? *Geriatrics*, **60**, 26–8, 30–1.
44. Beers, M.H. (1997). Explicit criteria for determining potentially inappropriate medication use by the elderly. An update. *Archives of Internal Medicine*, **157**, 1531–6.
45. Fick, D.M., Cooper, J.W., Wade, W.E., *et al.* (2003). Updating the Beers criteria for potentially inappropriate medication sue in older adults: results of a US consensus panel of experts. *Archives of Internal Medicine*, **163**, 2716–24.
46. Beckman, A., Bernsten, C., Parker, M.G., *et al.* (2005). The difficulty of opening medicine containers in old age: a population-based study. *Pharmacy World & Science*, **27**, 393–8.
47. Levings, B., Szep, S., and Helps, S.C. (1999). Towards the safer use of dosettes. *Journal of Quality in Clinical Practice*, **19**, 69–72.
48. Ringe, J.D., van der Geest, S.A., and Moller, G. (2006). Importance of calcium co-medication in bisphosphonate therapy of osteoporosis: an approach to improving correct intake and drug adherence. *Drugs & Aging*, **23**, 569–78.
49. Hinkin, C.H., Castellon, S.A., Dickson-Fuhrman, E., *et al.* (2002). Screening for drug and alcohol abuse among older adults using modified version of CAGE. *The American Journal on Addictions*, **10**, 319–26.
50. Anthony, J.C. and Helzer, J.E. (1991). Syndromes of drug abuse and dependence, in psychiatric disorders in America: In *The epidemiologic catchment area study* (eds. L.N. Robins and D.A. Regier). Free Press, New York.
51. Patterson, T.L. and Jeste, D.V. (1999). The potential impact of the baby-boom generation on substance use among elderly persons. *Psychiatric Services*, **50**, 1184–8.
52. Silagy, C., Lancaster, T., Stead, L., *et al.* (2002). Nicotine replacement therapy for smoking cessation. *Cochrane Database of Systematic Reviews*, (4), CD000146.
53. Hughes, J., Stead, L., and Lancaster, T. (2004). Antidepressants for smoking cessation. *Cochrane Database of Systematic Reviews*, (4), CD000031.
54. Wilhelm, K., Arnold, K., Niven, H., *et al.* (2004). Grey lungs and blue moods: smoking cessation in the context of lifetime depression history. *The Australian and New Zealand Journal of Psychiatry*, **38**, 896–905.

8.5.3 Schizophrenia and paranoid disorders in late life

Barton W. Palmer, Gauri N. Savla, and Thomas W. Meeks

Introduction

Estimates of the point-prevalence of paranoia and other psychotic symptoms among persons age ≥ 65 years have ranged from approximately 4 per cent to 6 per cent,[1–3] and may be as high as 10 per cent among those age ≥ 85 years.[4] Although the majority of these symptoms occur as secondary psychoses in the context of Alzheimer's disease or related dementias,[5] the population of people with schizophrenia is ageing along with the general 'greying' of the industrialized world, and mental health care for older adults with schizophrenia is expected to be an increasingly important public health concern.[6]

Clinical features

Schizophrenia is typified by the presence of two or more of the following categories of core symptoms: delusions, hallucinations, disorganized or catatonic behaviour, disorganized speech (or formal thought disorder), and negative symptoms (such as affective flattening, avolition, or social withdrawal).[7] Older patients tend to have less severe positive symptoms (hallucinations, delusions, disorganized behaviour) than their younger counterparts, but there are few age-related differences in presence or severity of negative symptoms.[8,9]

Most patients with schizophrenia and related primary psychotic disorders also have mild to moderate neurocognitive deficits.[10] There is considerable interpatient heterogeneity in terms of the severity of neuropsychological deficits, but the level of these deficits is a consistent and strong determinant of impairments in everyday functioning[11] and competence or decisional capacity.[12]

In terms of late-life schizophrenia, one common division is between those with earlier onset in adolescence or early adulthood (prior to age 40 or 45 years) versus later-onset (onset ≥ age 40 or 45 years). The latter group may comprise as many as 24 per cent of people with late-life schizophrenia.[13] Relative to similarly aged patients who had earlier onset, those with later-onset schizophrenia tend to have a higher prevalence of paranoid subtype and persecutory delusions, but better premorbid social-occupational functioning, fewer current disorganized symptoms, less severe (although not an absence of) negative symptoms, and less severe neuropsychological impairment. They are also more likely to be women, and tend to respond to lower doses of antipsychotic medication.[3,14,15] The two groups are similar in terms of severity of thought disorder,[16] although patients with very late onset schizophrenia-like psychosis (age of onset ≥ 60 years) tend to have less severe formal thought disorder.[17]

Classification systems

The term 'schizophrenia' was coined by Eugen Bleuler in the early 20th century, but he wrote of 'the schizophrenias' (plural)[18] as an

explicit acknowledgement of the substantial heterogeneity that characterizes this condition. Efforts to group 'the schizophrenias' into meaningful subtypes have been a key part of efforts to define the syndrome itself.[19] Most of the terms describing different subtypes of schizophrenia in the current *Diagnostic and Statistical Manual* (DSM-IV-TR)[7] and in the International Classification of Diseases (ICD-10)[20] [such as paranoid, catatonic, hebephrenic (disorganized), and undifferentiated (simple) subtypes] overlap with the subtypes identified by Kraepelin and or E. Bleuler a century ago. Other subtyping efforts have focused on a variety of dimensions such as positive and negative symptoms, cognitive functioning and/or course, but as true of the clinical subtypes in the DSM-IV-TR and ICD-10, there is invariably substantial intrasubtype heterogeneity.[21]

In regard to late-life schizophrenia, one of the key nosological controversies over the past century has been whether or not the late onset form is actually schizophrenia. Kraepelin's conception of *dementia praecox* in 1896 was that the disorder was defined by onset in adolescence or early adulthood. By 1913, Kraepelin came to acknowledge that early onset was not a universal feature, but the emphasis on early onset remained a potent belief in the field throughout most of the 20th century.[22] On the other hand, interest in late-onset schizophrenia has a long history, including seminal work by Manfred Bleuler, begun in the early 1940s with patients whose symptoms emerged at or after age 40 years.[23]

The term 'late-onset schizophrenia' has occasionally been used interchangeably with the term *late paraphrenia*, although the latter was originally conceptualized as a more circumscribed psychosis with onset at age 60 or 65.[24,25] Unfortunately, the terms 'late-onset schizophrenia,' and *paraphrenia* (with or without the epithet 'late'), and a variety of age cut-offs have been used interchangeably and inconsistently over the years, resulting in considerable confusion in the literature.[22,25] In a 1998 international consensus meeting on this topic, the group consensus suggestion was that the term 'late onset schizophrenia' be reserved for those with onset between ages 40 and 59 years, whereas the term 'very late onset schizophrenia-like psychosis' be used with those whose symptoms first manifest at age 60 or later.[3]

None of the above schizophrenia onset-related categories is represented in the contemporary formal diagnostic systems. The 1980 version of the American Psychiatric Association's *Diagnostic and Statistical Manual* (DSM-III)[26] arbitrarily excluded the diagnosis of schizophrenia if symptoms did not emerge prior to age 45. This exclusion was dropped from the subsequent revision (DSM-III-R),[27] although the DSM-III-R required the specification of 'late onset' if the prodromal phase of illness developed after 45. The latter is the only instance of 'late-onset schizophrenia' appearing as a named condition in one of the major nosological systems. Based on mounting empirical evidence that 'real' schizophrenia could manifest after age 45,[13] the age of onset restrictions as well as the 'late onset' specifier were dropped in the DSM-IV [28] and DSM-IV-TR.[7] Similarly, there is no age-of-onset related restriction or specification under the ICD-10.[20]

Diagnosis and differential diagnosis

The diagnostic criteria for schizophrenia in the DSM-IV-TR and ICD-10 mention neither current age nor age of onset.[7,20] A key differential diagnosis with older adults is to rule out presence of a secondary psychosis.[29] For instance, among elderly patients, psychotic symptoms most commonly present in the context of dementia, such as Alzheimer's disease, Parkinson's disease, or dementia with Lewy Bodies.[5] The pattern in any one patient may of course vary from normative trends, but in general among those with dementia-related psychotic symptoms, there is a greater propensity for visual over auditory hallucinations, and bizarre content is less common in the delusions than in those of patients with primary psychotic disorders such as schizophrenia.[30]

Delirium may also present as acute psychosis;[31] as with dementia, visual hallucinations and delusions tend to be more common than auditory hallucinations, but the psychotic symptoms associated with delirium can be of any form.[32] Given the high rates of polypharmacy among the elderly as well as age-related changes in pharmacokinetics, it is also important to consider potential acute mental effects of the medications in isolation and in combination.[33] Other differential diagnoses to consider among elderly patients are non-psychotic hallucinations related to bereavement or sensory deprivation.[34,35]

Among the primary psychotic conditions, the standard differential diagnoses and considerations apply in terms of differentiating among schizophrenia, schizoaffective disorder, delusional disorder, brief psychotic disorder, substance-induced psychotic disorder, bipolar disorder with psychotic features, and major depressive disorder with psychotic features.[7]

Epidemiology

As was noted above, prevalence estimates of paranoia and other psychotic symptoms among persons age ≥ 65 years have ranged from approximately 4 per cent to 6 per cent,[1–3] but these symptoms are most commonly in the context of a dementia or other medical condition. Estimating the lifetime prevalence of schizophrenia is a methodologically complex endeavour needing additional research attention; recent estimates have ranged from approximately, 0.4 per cent to 1.0 per cent although estimates as high as 1.6 per cent have also been reported.[36,37] The lifetime prevalence of schizophrenia is similar among men and women, and the majority of patients of either gender experience onset in adolescence or early adulthood.[37] However, a consistent finding noted a century ago by E. Bleuler,[18] is that women tend to show later onset than men.[38]

Estimates of the lifetime prevalence of schizophrenia for persons over age 65 have also varied, although the 95 per cent confidence interval estimate from one recent comprehensive study was 0.58 to 1.45 per cent.[37] There have been some epidemiological studies suggesting that the prevalence (current and lifetime) of schizophrenia among elderly persons is lower than that for the younger population. People with schizophrenia have higher mortality due to suicide and physical disorders,[39,40] so there are probably proportionally fewer people with schizophrenia who survive to older age. However, the prevalence of schizophrenia in elderly patients may also have been underestimated in some of the earlier major epidemiological studies.[41] For instance, the Epidemiologic Catchment Area study used the DSM-III criteria, but as was noted above, the DSM-III criteria for schizophrenia arbitrarily required onset of prodromal symptoms prior to age 45, so any cases of later-onset schizophrenia would have been excluded.[41]

Aetiology

The cause(s) of schizophrenia remains unknown. Both Kraepelin and E. Bleuler correctly suspected that there is a heritable vulnerability to schizophrenia, confirmed by the substantially higher concordance rates among monozygotic twins (estimated at 40 to 50 per cent) relative to dizygotic twins (estimated at 5 per cent to 25 per cent).[42,43] The elevated risk of schizophrenia among first-degree relatives is present among those with schizophrenia onset in middle-age as well as those wither earlier onset relatives.[3] Although some candidate genes have been identified,[44] these efforts remain in an early stage of development. Also, given that even the monozygotic twin concordance rate is substantially below 100 per cent, non-genetic factors clearly have a role in the ultimate expression of the schizophrenia phenotype.

At present, the prevailing model of schizophrenia is that of neurodevelopmentally based aberrations in connectivity of key brain regions and systems.[45,46] Evidence for the neurodevelopment component includes an elevated risk of schizophrenia associated with certain pre- or peri-natal insults or stresses, an increased prevalence of minor facial anomalies among patients with schizophrenia, and an increased prevalence of subtle childhood abnormalities in motor, cognitive, and/or psychosocial development among those who later develop schizophrenia.[45] At the level of neuropathology, Kraepelin expressed some suspicion of involvement of the prefrontal and temporal lobes;[47] these remain areas of interest in schizophrenia research although the current focus is on functional (rather than gross structural) impairments related to the connections among such brain regions or systems.[46]

Course and prognosis

Schizophrenia is generally a chronic condition, but not necessarily a constantly sustained one in that many patients experience one or more periods of several years of sustained recovery over their lifespan, but periods of relapse are also common.[48] As with other dimensions of this disorder, the long-term course of schizophrenia is also characterized by heterogeneity among patients, as well as methodological challenges in interpreting varied findings in the empirical literature.[49] Some of the factors that have been cited as associated with worse prognosis include poor premorbid functioning, very early and gradual/insidious onset of symptoms, male gender, and a relative prominence of negative symptoms.[49,50] In a recent review of ten long-term longitudinal outcome studies, Jobe and Harrow[48] found that the estimates of 'good outcome' ranged from 21 to 57 per cent. Since most of the long-term research on the course of schizophrenia has been of younger patients as they age, there remains a clear need for longitudinal research to document changes among patients as they age from their 60s, 70s, and 80s. Nonetheless, empirical data do not seem to support Kraepelin's initial suggestion that *dementia praecox* is characterized by a course of progressive decline.[49] In fact, there may be some modest age-related improvements in positive symptoms and perhaps other aspects of psychopathology.[8] Although there is a small subset of 'poor outcome'/ chronically institutionalized patients who seem to be at added risk for cognitive and functional decline in older age,[51] for most patients there is generally no increased (beyond age normal) decline in cognitive functioning among those with early or middle-age onset.[52,53]

Treatment

Contemporary treatment guidelines for schizophrenia in older adults parallel those for younger adults in that a combination of pharmacological and psychosocial interventions is recommended.[54] Although treatment with antipsychotic medications is a mainstay of effective treatment and management of late-life schizophrenia, selection of the appropriate type and dose of antipsychotic medication in the elderly may be complicated by age-related factors. One of the primary concerns with conventional neuroleptic medications is that they can cause iatrogenic motor abnormalities; older adults are at even greater risk than younger patients to develop tardive dyskinesia and extrapyramidal symptoms, or EPS, (especially parkinsonism) from conventional neuroleptics.[55] The newer ('atypical' or 'second generation') antipsychotic medications have lower (although not absent) associated risk of tardive dyskinesia and EPS, though risks for these motor side effects also vary from one atypical antipsychotic to another.

The current APA treatment guidelines for schizophrenia indicate '*Second-generation antipsychotics are generally recommended over first-generation agents because of their significantly lower risk of inducing extrapyramidal symptoms and tardive dyskinesia in older persons . . . However, the second-generation agents have other clinically significant and common side effects* (pp. 33–34).' In addition to concerns about potential sedation, orthostatic hypotension, and other potential physical or medical side-effects from the newer medications, there has been recent concern about the potentially serious metabolic and cardiovascular side effects.[56] Overall, atypical agents (with the exception of clozapine) have not shown superior efficacy for schizophrenia in direct comparisons with typical antipsychotics.[57] This fact, combined with the lower cost of typical agents, has prompted some experts to question the use of atypical antipsychotics as first-line pharmacotherapy for schizophrenia. In the absence of more definitive evidence, clinicians should discuss these various advantages and disadvantages of specific drugs with patients when choosing drug therapy.

For most antipsychotics (typical or atypical), older adults with schizophrenia generally respond to 50–75 per cent of the doses needed in younger patients and often cannot tolerate the full young adult dose.[58] Side effects that are especially problematic in older adults include anticholinergic effects (e.g. constipation, confusion, urinary retention), orthostatic blood pressure changes (which may lead to falls), and reported increases in stroke and death among persons with dementia receiving antipsychotics (as dementia, of course, becomes more prevalent with increasing age).

Antipsychotic medications are helpful in managing the psychopathologic symptoms of schizophrenia, especially the so-called 'positive symptoms' such as delusions and hallucinations, but they are not a cure. Medications also tend to have little benefit for the 'negative' (e.g. apathy, anhedonia) and cognitive symptoms of the illness. Furthermore, many patients continue to have residual functional disability despite resolution of positive symptoms of psychosis. Thus, the importance of adjunctive treatment with evidenced-based psychosocial interventions in schizophrenia is being increasingly recognized.[59] For instance, investigators at our Research Center have developed or adapted, and validated a number of effective adjunctive psychosocial interventions for older patients with schizophrenia or related psychoses; these include Cognitive Behavioral Social Skills Training,[60]

Functional Adaptation Skills Training,[61] diabetes management/lifestyle modification,[62] and vocational rehabilitation/supported employment.[63] These efforts address a number of dimensions of schizophrenia that are unaffected by pharmacologic treatment alone.

Management

As noted above, schizophrenia is a chronic condition. In addition, to specific therapies previously described, managing this complex and devastating illness requires careful attention to clinician-patient rapport. This is especially true in persons with severe paranoia and in those who lack insight into their illness. Providing care for someone who does not trust anyone or who sees no reason for treatment can be challenging, but these obstacles can often be overcome by involving family, listening empathically, avoiding premature confrontation about delusions, respecting the increased interpersonal distance many patients require, and identifying the patient's goals and priorities. Management of schizophrenia may often be optimized by multi- and interdisciplinary care to address the many ways in which the illness affects patients' lives. Because there tends to be a high degree of medical comorbidity in persons with schizophrenia (especially older adults), integration of primary care and psychiatric care is often needed. Due to normal age-related physical changes, polypharmacy, and the higher risk of medication side-effects, the long-term management of schizophrenia in older adults demands frequent monitoring of symptoms, overall health, and side-effects. Cross-discipline collaborative care and continuity with the same clinicians can help ensure older persons with schizophrenia achieve the best possible outcomes.

Further information

Cohen, C. I. (ed.). (2003). *Schizophrenia into later life: Treatment, research, and policy*. American Psychiatric Publishing, Washington, DC.

Lehman, A.F., Lieberman, J.A., Dixon, L.B., *et al.* (2004). Practice guidelines for the treatment of patients with schizophrenia, second edition. *American Journal of Psychiatry*, **161** (2 Suppl), 1–56.

Sharma, T, and Harvey, P. (eds.). (2006). *The early course of schizophrenia*. New York: Oxford University Press.

References

1. Henderson, A.S., Korten, A.E., Levings, C., *et al.* (1998). Psychotic symptoms in the elderly: a prospective study in a population sample. *International Journal of Geriatric Psychiatry* , **13**(7), 484–92.
2. Forsell, Y. and Henderson, A.S. (1998). Epidemiology of paranoid symptoms in an elderly population. *British Journal of Psychiatry*, **172**, 429–32.
3. Howard, R., Rabins, P.V., Seeman, M.V., *et al.*(2000). Late-onset schizophrenia and very-late-onset schizophrenia-like psychosis: an international consensus. The International Late-Onset Schizophrenia Group. *American Journal of Psychiatry*, **157**(2), 172–8.
4. Ostling, S. and Skoog, I. (2002). Psychotic symptoms and paranoid ideation in a nondemented population-based sample of the very old. *Archives of General Psychiatry*, **59**(1), 53–9.
5. Mintzer, J. and Targum, S.D. (2003). Psychosis in elderly patients: classification and pharmacotherapy. *Journal of geriatric psychiatry and neurology*, **16**(4), 199–206.
6. Palmer, B.W., Heaton, S.C., Jeste, D.V., *et al.* (1999). Older patients with schizophrenia: challenges in the coming decades. *Psychiatric services*, **50**(9), 1178–83.
7. American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders - Text Revision.* (4th edn). American Psychiatric Association, Washington, DC.
8. Jeste, D.V., Twamley, E.W., Eyler Zorrilla, L.T., *et al.* (2003). Aging and outcome in schizophrenia. *Acta Psychiatrica Scandinavica*, **107**(5), 336–43.
9. Ciompi, L. (1985). Aging and schizophrenic psychosis. *Acta psychiatrica Scandinavica. Supplementum*, **319**, 93–105.
10. Heinrichs, R.W. and Zakzanis, K.K. (1998). Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology*, **12**(3), 426–45.
11. Green, M.F., Kern, R.S., Braff, D.L., *et al.* (2000). Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the 'right stuff '? *Schizophrenia bulletin*, **26**(1), 119–36.
12. Palmer, B.W., Savla, G.N. (in press). The association of specific neuropsychological deficits on capacity to consent to research or treatment. *Journal of the International Neuropsychological Society*, in press.
13. Harris, M.J., Jeste, D.V. (1998). Late-onset schizophrenia: an overview. *Schizophrenia bulletin*, **14**(1), 39–55.
14. Jeste, D.V., Symonds, L.L., Harris, M.J., *et al.* (1997 Fall). Nondementia nonpraecox dementia praecox? Late-onset schizophrenia. *The American journal of geriatric psychiatry* , **5**(4), 302–17.
15. Sato, T., Bottlender, R., Schroter, A., *et al.* (2004). Psychopathology of early-onset versus late-onset schizophrenia revisited: an observation of 473 neuroleptic-naive patients before and after first-admission treatments. *Schizophrenia research*, **67**(2-3), 175–83.
16. Palmer, B.W., McClure, F.S., Jeste, D.V., *et al.* (2001). Schizophrenia in late life: findings challenge traditional concepts. *Harvard Review of Psychiatry* , **9**(2),51–8.
17. Castle, D.J., Wessely, S., Howard, R., et al. (1997). Schizophrenia with onset at the extremes of adult life. *International Journal of Geriatric Psychiatry*, 12(7), 712–7.
18. Bleuler, E. (1950/1911). *Dementia praecox; or, The group of schizophrenias*. International Universities Press, New York.
19. Berrios, G. E., Luque, R., & Villagrán, J. M. (2003). Schizophrenia: A conceptual history. *International Journal of Psychology & Psychological Therapy*, **3**(2), 111–40.
20. World Health Organization. International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Version for 2007. Available via internet at: http://www.who.int/classifications/icd/en/[Accessed June 15, 2007].
21. Palmer, B.W., Nayak, G.V., Jeste, D.V., *et al.* (2003). A comparison of early- and late-onset schizophrenia. In *Schizophrenia into later life* (ed. C., Cohen), pp. 43–75. American Psychiatric Publishing, Washington, DC.
22. Adityanjee, Aderibigbe, Y.A., Theodoridis, D., *et al.* (1999). Dementia praecox to schizophrenia: the first 100 years. *Psychiatry and Clinical Neurosciences*, **53**(4),437–48.
23. Bleuler, M. (1978). *The schizophrenic disorders: Long-term patient and family studies*. Yale University Press, New Haven, CT.
24. Roth, M. (1995). The natural history of mental disorder in old age. *The Journal of Mental Science* , **101**(423), 281–301.
25. Berrios, G.E. (2003). The insanities of the third age: a conceptual history of paraphrenia. *The Journal of Nutrition, Health & Aging*, **7**(6), 394–9.
26. American Psychiatric Association. (1980). *Diagnostic and Statistical Manual* (3rd edn.). Author, Washington, DC.
27. American Psychiatric Association. (1987). *Diagnostic and Statistical Manual* (3rd edn.). Author, Washington, DC.
28. American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th edn.). American Psychiatric Association, Washington, DC.
29. Jeste, D.V. and Palmer, B.W. (1998) Secondary Psychoses: An overview. *Seminars in Clinical Neuropsychiatry*, **3**, 2–3.

30. Jeste, D.V., Finkel, SI. (2000 Winter) Psychosis of Alzheimer's disease and related dementias. Diagnostic criteria for a distinct syndrome. The American journal of geriatric psychiatry , **8**(1),29–34.
31. Cole, M.G. Delirium in elderly patients. *The American journal of geriatric psychiatry*.
32. Webster, R. and Holroyd, S. (2000). Prevalence of psychotic symptoms in delirium. *Psychosomatics*,**41**(6),519–22.
33. Marsh, C.M. (1997). Psychiatric presentations of medical illness. *The Psychiatric Clinics of North America*, **20**(1), 181–204.
34. Grimby, A. (1993). Bereavement among elderly people: grief reactions, post-bereavement hallucinations and quality of life. *Acta psychiatrica Scandinavica*, **87**(1), 72–80.
35. Menon, G.J., Rahman, I., Menon, S.J., *et al.* (2003). Complex visual hallucinations in the visually impaired: the Charles Bonnet Syndrome. *Survey of Ophthalmology*, **48**(1), 58–72.
36. Saha, S., Chant, D., Welham, J., *et al.* (2005).A systematic review of the prevalence of schizophrenia. *PLoS Medicine*, **2**(5), e141.
37. Perala, J., Suvisaari, J., Saarni, S.I., *et al.* (2007). Lifetime prevalence of psychotic and bipolar I disorders in a general population. *Archives of General Psychiatry*, **64**(1), 19–28.
38. Lindamer, L.A., Lohr, J.B., Harris, M.J., *et al.* (1999). Gender-related clinical differences in older patients with schizophrenia. *The Journal of Clinical Psychiatry*, **60**(1), 61–7; quiz 8–9.
39. Black, D.W. and Fisher, R. (1992). Mortality in DSM-IIIR schizophrenia. *Schizophrenia research* , **7**(2), 109–16.
40. Joukamaa, M., Heliovaara, M., Knekt, P., *et al.* (2006). Schizophrenia, neuroleptic medication and mortality. *British Journal of Psychiatry*, **188**, 122–7.
41. Hybels, C.F. and Blazer, D.G. (2003). Epidemiology of late-life mental disorders. *Clinics in Geriatric Medicine*, **19**(4), 663–96, v.
42. Cardno, A.G. and Gottesman, II. (2000 Spring) Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *American Journal of Medical Genetics*, **97**(1), 12–17.
43. Walker, E., Kestler, L., Bollini, A., *et al.* (2004).Schizophrenia: etiology and course. *Annual review of psychology*, **55**, 401–30.
44. Ross, C.A., Margolis, R.L., Reading, S.A., *et al.* (2006) Neurobiology of schizophrenia. *Neuron*, **52**(1), 139–53.
45. Rapoport, J.L., Addington, A.M., Frangou, S., *et al.* (2005). The neurodevelopmental model of schizophrenia: update 2005. *Molecular Psychiatry*, **10**(5), 434–49.
46. Friston, K.J. and Frith, C.D. (1995). Schizophrenia: a disconnection syndrome? *Clinical neuroscience* , **3**(2), 89–97.
47. Kraepelin E. (1971/1919). *Dementia praecox and paraphrenia*. New York: Krieger.
48. Jobe, T.H. and Harrow, M. (2005). Long-term outcome of patients with schizophrenia: a review. *Canadian Journal of Psychiatry*, **50**(14), 892–900.
49. Riecher-Rossler, A. and Rossler, W. (1998). The course of schizophrenic psychoses: what do we really know? A selective review from an epidemiological perspective. *European Archives of Psychiatry and Clinical Neuroscience* , **248**(4), 189–202.
50. Ram, R., Bromet, E.J., Eaton, W.W., *et al.* (1992)The natural course of schizophrenia: a review of first-admission studies. *Schizophrenia bulletin* , **18**(2), 185–207.
51. White, L., Friedman, J.I., Bowie, C.R., *et al.* (2006). Long-term outcomes in chronically hospitalized geriatric patients with schizophrenia: retrospective comparison of first generation and second generation antipsychotics. *Schizophr Res*, **88**(1–3), 127-34.
52. Heaton, R.K., Gladsjo, J.A., Palmer, B.W., *et al.* (2001). The stability and course of neuropsychological deficits in schizophrenia. *Archives of General Psychiatry* , **58**(1), 24–32.
53. Palmer, B.W., Bondi, M.W., Twamley, E.W., *et al.* (2003 Winter) Are late-onset schizophrenia spectrum disorders neurodegenerative conditions? Annual rates of change on two dementia measures. *The Journal of Neuropsychiatry and Clinical Neurosciences*; **15**(1), 45–52.
54. Lehman, A.F., Lieberman, J.A., Dixon, L.B., *et al.* (2004)Practice guideline for the treatment of patients with schizophrenia, second edition. *American Journal of Medical Genetics* , **161**(2. Suppl.), 1–56.
55. Jeste, D.V. (2000). Tardive dyskinesia in older patients. *The Journal of clinical psychiatry*. **61** (Suppl 4), 27–32.
56. Newcomer, J.W. and Lieberman, J.A.(2007). Comparing safety and tolerability of antipsychotic treatment. *The Journal of clinical psychiatry* , **68**(3), e07.
57. Lieberman, J.A., Stroup, T.S., McEvoy, J.P., *et al.* (2005). Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *The New England Journal of Medicine* , **22**;353(12), 1209–23.
58. Jeste, D.V., Dolder, C.R., Nayak, G.V., *et al.* (2005). Atypical antipsychotics in elderly patients with dementia or schizophrenia: review of recent literature. *Harvard Review of Psychiatry* , **13**(6), 340–51.
59. Dickerson, F.B., Lehman, A.F. (2006). Evidence-based psychotherapy for schizophrenia. *The Journal of Nervous and Mental Disease* , **194**(1), 3–9.
60. Granholm, E., McQuaid, J.R., McClure, F.S., *et al.* (2005). A randomized, controlled trial of cognitive behavioral social skills training for middle-aged and older outpatients with chronic schizophrenia. *American Journal of Psychiatry* , **162**(3), 520–29.
61. Patterson, T.L., Mausbach, B.T., McKibbin, C., *et al.* (2006). Functional adaptation skills training (FAST): a randomized trial of a psychosocial intervention for middle-aged and older patients with chronic psychotic disorders. *Schizophrenia research* , **86**(1–3), 291–9.
62. McKibbin, C.L., Patterson, T.L., Norman, G., *et al.* (2006).A lifestyle intervention for older schizophrenia patients with diabetes mellitus: a randomized controlled trial. *Schizophrenia research* , **86**(1–3), 36–44.
63. Twamley, E.W., Padin, D.S., Bayne, K.S., *et al.* (2005).Work rehabilitation for middle-aged and older people with schizophrenia: a comparison of three approaches. *The Journal of Nervous and Mental Disease*, **193**(9), 596–601.

8.5.4 Mood disorders in the elderly

Robert Baldwin

Introduction

This chapter considers some of the commonly asked questions about mood disorders in later life. Is depression in later life a distinct clinical syndrome? How common is it? Is there an organic link, for example to cerebral changes, and if so, is there an increased risk of later dementia? Is it more difficult to diagnose and treat late-life depression, and once treated, is the outcome good, bad, or indifferent? The emphasis will be on depression but bipolar disorder and mania will also be considered.

Classification

The main mood disorders which older people suffer are classified as: depressive episode, dysthymia, bipolar disorder, and organic mood disorder. Depressive illness and major depression are terms often used synonymously with depressive episode. Current classificatory systems, notably The World Health Organization ICD10 and the DSM Version IV of the American Psychiatric Association, are described in Chapter 4.5.3.

Depressive episode

Clinical features

Other than more frequent somatic and hypochondriacal complaints, patients with depression in later life are little different, symptomatically, to younger adults.[1] An exception may be the recently described 'vascular depression' (depression linked to small vessel disease of the brain), in which depressive ideation is less but cognitive impairment and apathy greater. In most cases then the pathoplastic effects of ageing and ill-health are what mainly influence the presentation of depression in later life (Table 8.5.4.1).

An overlap of symptoms due to *associated physical ill-health* may lead to diagnostic difficulty, and determining whether a symptom has arisen predominantly because of affective disorder or a medical condition can be difficult for those without the necessary experience.

Older depressed patients may *minimize feelings of sadness* and instead become *hypochondriacal* (morbidly preoccupied with a fear of illness).[1] Late-onset *neurotic symptoms* (*severe anxiety*, *phobias*, *obsessional compulsive phenomena or hysteria*) are usually secondary to depressive illness. Any act of *deliberate self-harm* suggests depression, as elderly people rarely take 'manipulative' overdoses. An overdose in an older person should never be dismissed because its effects, in purely medical terms, were trivial. *All* require psychiatric assessment. Severe depression may mimic dementia. Table 8.5.4.2 highlights the main differences between progressive dementia and the pseudodementia of depression. Pseudodementia is a term which is perhaps waning in use as it has become clear that depressive disorder is commonly associated with cognitive impairment, which may not be reversible, even with adequate treatment of depression. Pseudodementia is often applied to an older depressed patient who, on presentation, appears very confused, with frequent 'don't know' responses. However, the onset of confusion is acute, easily dated, the patients convey their despair non-verbally, and, unlike the person with degenerative dementia, complain vociferously about their memory. Cortical signs (aphasia, apraxia, etc.) suggest a primary dementia with a super-added depression rather than depressive pseudodementia. Wandering off and getting lost suggests dementia but occasional cases are seen of fugue states caused by severe depression mimicking disorganized behaviour in dementia. The key is a good history.

An unusual *behavioural disturbance* may occasionally be a leading symptom of depression. Examples include the onset of incontinence in an older person who feels trapped in a situation of resented dependency in a residential or nursing home, late-onset alcohol abuse or, rarely, shoplifting.

Table 8.5.4.1 Factors influencing the presentation of depression in older people

Overlap of symptoms of physical disorder with those of the somatic symptoms of depression
Tendency of older people to minimize a complaint of sadness and instead become hypochondriacal
Late-onset neurotic symptoms (severe anxiety, obsessional compulsive symptoms, hysteria) which mask depression
Deliberate self-harm which seems medically trivial
Pseudodementia
Behavioural disturbance such as alcohol abuse or shoplifting

Table 8.5.4.2 Characteristics distinguishing depression ('pseudodementia') from dementia

Dementia	Depression
Insidious	Rapid onset
Symptoms usually of long duration	Symptoms usually of short duration
Mood and behaviour fluctuate	Mood is consistently depressed
'Near miss' answers typical	'Don't know' answers typical
Patient conceals forgetfulness	Patient highlights forgetfulness
Cognitive impairment relatively stable	Cognitive impairment fluctuates greatly
Higher cortical dysfunction evident	Higher cortical dysfunction absent

(a) Vascular depression

In vascular depression, vascular disease is judged to predispose, precipitate, or perpetuate depressive symptoms. Evidence (summarized by Baldwin[2]) includes the following. There is a high rate of structural brain abnormalities in both white matter and basal ganglia grey matter on imaging and on post-mortem examination of older patients with depressive disorder, notably with a late age of onset. Psychomotor change, apathy, and executive dysfunction (leading to slowed responses, failure of initiation, impersistence in tasks, and inefficient memory) occur characteristically in such patients. Strategic lesion location, sufficient to disrupt subcortical-frontal circuitry, is associated with poorer depression outcomes, and progression of such lesions is associated with later incident cases of depression in those not already depressed. The concept of vascular depression is discussed critically under aetiololgy.

Diagnosis and differential diagnosis

(a) Assessment

The psychiatric history should include a collateral history as well as drug evaluation (prescribed, 'borrowed', and over-the-counter) and alcohol intake. A cognitive screening test should always be undertaken. A physical evaluation should focus on possible disorders causing an organic mood disorder (Table 8.5.4.3), including medication. Non-selective β-blockers, calcium antagonists, benzodiazepines, and systemic corticosteroids were the main culprits in one study.[3]

Screening questionnaires can be used to help diagnose depression, especially in settings such as medical wards where the prevalence is high, but their results must be informed by clinical judgement. The Geriatric Depression Scale (GDS) (Geriatric Depression Scale website http://stanford.edu/~yesavage/GDS.html) is widely used. It focuses on the cognitive aspects of depressive illness rather than physical depressive symptomatology, and has a simple 'yes/no' format (Table 8.5.4.4). It loses specificity in severe dementia but performs reasonably well in mild to moderate dementia. For rapid screening four questions (1, 3, 8 and 9) can be used.

(b) Investigations

Table 8.5.4.5 summarizes investigations appropriate for a first episode of depression and a recurrence. A guiding principle is that elderly people are in a more precarious state of homeostasis with their environment because they have less physiological reserve.

Table 8.5.4.3 Common medical illnesses and drugs that may cause organic mood syndromes

Medical conditions	Central-acting drugs
Endocrine/metabolic	*Anti-hypertensive drugs*
Hypo/hyperthyroidism	β-blockers (especially non-selective)
Cushing's disease	Methyldopa
Hypercalcaemia	Reserpine
Sub-nutrition	Clonidine
Pernicious anaemia	Nifedipine, calcium channel agents
Organic brain disease	Digoxin
Cerebrovascular disease/stroke	*Steroids*
CNS tumours	*Analgesic drugs*
Parkinson's disease	Opioids
Alzheimer's disease and vascular dementia	Indomethacin
Multiple sclerosis	*Anti-parkinson*
Systemic lupus erythematosus	L-Dopa
Occult carcinoma	Amantadine
Pancreas	Tetrabenazine
Lung	*Psychiatric drugs*
Chronic infections	Neuroleptics
Neurosyphilis	Benzodiazepines
Brucellosis	*Miscellaneous*
Neurocysticercosis	Sulphonamides
Myalgic encephalomyelitis	Alcohol
AIDS	Interferon

Severe depression in a 75-year-old may lead to quite serious metabolic derangement which would be unlikely in a fit 35-year-old.

An electroencephalogram (EEG) can help in differentiating depression from an organic brain syndrome such as delirium or an early dementia. A brain scan is only performed if clinically indicated, for example a rapid-onset depression with neurological symptoms or signs. The Dexamethasone Suppression Test (DST) is less specific for depressive illness than was first thought. It cannot reliably differentiate dementia from depression.

(c) Differential diagnosis

Organic mood disorder is diagnosed when a direct aetiological link can be established between the onset of the mood disorder and an underlying systemic or cerebral disorder (including dementia), or an ingested substance such as medication or alcohol.

Bipolar disorder is covered later. **Psychotic illness** (schizophrenia or delusional disorder) may present with marked depressed affect but other symptoms are present. A common depressive delusion in old age is hypochondriasis and sometimes it is difficult to decide whether the patient has a psychotic or an affective disorder. Interpretation depends on which symptoms predominate; if they occur together, it may be appropriate to use the term schizoaffective disorder.

Table 8.5.4.4 Geriatric Depression Scale

Instructions: Choose the best answer for how you have felt over the past *week*.

1. **Are you basically satisfied with your life?** No
2. **Have you dropped many of your activities and interests?** Yes
3. **Do you feel your life is empty?** Yes
4. **Do you often get bored?** Yes
5. Are you hopeful about the future? No
6. Are you bothered by thoughts you can't get out of your head? Yes
7. **Are you in good spirits most of the time?** No
8. **Are you afraid something bad is going to happen to you?** Yes
9. **Do you feel happy most of the time?** No
10. **Do you often feel helpless?** Yes
11. Do you often get restless and fidgety? Yes
12. **Do you prefer to stay at home, rather than going out and doing new things? Yes**
13. Do you frequently worry about the future? Yes
14. **Do you feel you have more problems with your memory than most?** Yes
15. **Do you think it is wonderful to be alive now?** No
16. Do you often feel downhearted and blue (sad)? Yes
17. **Do you feel pretty worthless the way you are?** Yes
18. Do you worry a lot about the past? Yes
19. Do you find life very exciting? No
20. Is it hard for you to start on new projects (plans)? Yes
21. **Do you feel full of energy?** No
22. **Do you feel that your situation is hopeless?** Yes
23. **Do you think most people are better off (in their lives) than you are?** Yes
24. Do you frequently get upset over little things? Yes
25. Do you frequently feel like crying? Yes
26. Do you have trouble concentrating? Yes
27. Do you enjoy getting up in the morning? No
28. Do you prefer to avoid social gatherings (get-togethers)? Yes
29. Is it easy for you to make decisions? No
30. Is your mind as clear as it used to be? No

Notes: (1) Answers refer to responses which score '1'; (2) bracketed phrases refer to alternative ways of expressing the questions; (3) questions in bold are for the 15-item version. Threshold for possible depression: >/=11 (GDS30); >/=5 (GDS15); >=2 (GDS4).

Dysthymia chiefly occurs in younger adults but may occur in later life in association with chronic ill-health. Where there is a clear onset of depressive symptoms within 1 month of a stressful life event without the criteria for a depressive episode being met, then an *adjustment disorder* may be diagnosed.

Epidemiology

In the United Kingdom, pervasive depression (a term denoting a depressive syndrome that a psychiatrist would regard as warranting intervention) is found between 8.6 and 14.1 per cent of elderly people living at home. The prevalence of a depressive episode is between 1 and 4 per cent of elderly people living at home.[4] The finding of a high rate of depressive symptoms but a much lower rate of depressive episodes is an epidemiological dilemma which is discussed in Chapter 4.5.4. It is likely that current classification

Table 8.5.4.5 Investigations for depression in later life

Investigation	First episode	Recurrence
Full blood count	Yes	Yes
Urea and electrolytes	Yes	Yes
Calcium	Yes	Yes
Thyroid function	Yes	If clinically indicated, or more than 12 months elapsed
B_{12}	Yes	If clinically indicated, or more than 12 months elapsed
Folate	Yes	If clinically indicated (for example recent poor diet)
Liver function	Yes	If indicated (for example suspected or known alcohol misuse)
Syphilitic serology	If clinically indicated (for example relevant neurological symptoms)	Only if clinically indicated
CT (brain)	If clinically indicated	If clinically indicated
EEG	If clinically indicated	If clinically indicated

systems overlook many of the late-life depressions found in community studies.

Depression in later life, whether major or 'minor', is associated with worsened medical morbidity, disability, and increased health utilization.[1] Co-morbidity from physical disorder or cognitive impairment is the main determinant of prevalence. Handicap, the disadvantage imposed by a physical impairment and attendant disability, is a further strong predictor of depression.[5] This matters because handicap is amenable to social intervention.

In residential care, nursing homes and medical wards the rates are between 20 and 40 per cent.

Aetiology

The risk factors for depression in later life are discussed in Chapter 4.5.5. A depressive episode usually arises from a combination of vulnerability factors along with a triggering (precipitating) adverse life event. Avoidant and dependent personality types are associated with late-life depression, and a lifelong lack of a capacity for intimacy is another risk factor.[6] Precipitating life events occur at a similar frequency to other age groups, although health-related events are more common among older people.[6]

The concept of vascular depression suggests new aetiological insights.[2] However, criticisms against vascular depression as a distinct subtype include the difficulty in establishing a temporal link between depression onset and vascular disease and that, if causal, vascular disease is likely to have been present well before old age. Furthermore, in studies of vascular depression the direction of causality is unclear since patients with vascular disease have a high rate of depression, and depression appears to worsen vascular disease, perhaps by direct effects on blood vessel endothelial function and indirectly through poor self-monitoring of health and poor adherence to medical drugs.

If confirmed, the vascular depression hypothesis could lead to antidepressant strategies aimed at improving the underlying vascular impairment as well as mood. In the meantime, a patient presenting late-onset depression should be thoroughly investigated for vascular disease as a potential aetiological contributory factor which should be optimally treated along with the depression.

Course and prognosis

Across all age groups depressive disorder is prone to persistence. Beekman *et al.*[7] followed 277 community-dwelling subjects, aged over 55 (most over 75), from a Dutch epidemiological survey. Using multiple assessments of mood over a 6-year period, they found almost half the sample was depressed for more than 60 per cent of the time. Twenty three per cent had true remissions, 12 per cent remissions with recurrence, a third a chronic-intermittent course; another third had chronic depression. Outcomes reported from the community and medical ward patients are worse than those of depressed patients under psychiatric care,[8] possibly linked to undertreatment in the non-specialist settings.

(a) Comparative outcome

Mitchell and Subramaniam[9] reviewed the literature between 1966 and July 2004, finding 24 publications which could be used to assess outcomes between different age groups. Overall the authors concluded that episodes of depression remitted in later life as well as in other times but with a greater risk of relapse. Two factors seemed to explain this: age of onset (recurrent depression from earlier life conferring a poorer prognosis) and medical comorbidity (with a worse prognosis linked to a later onset). Although these mechanisms differ, the high risk of relapse in late-life depression highlights the need for effective continuation and maintenance phases of treatment, regardless of age of onset.

Mortality

A number of studies show that depression is an independent risk factor for increased mortality,[1] not accounted for by suicide. Following a cohort of 652 depressed and non-depressed subjects over 3.5 years, Geerlings[10] found that duration, chronicity, and increasing symptoms from baseline were all linked to a higher risk of death, leading them to suggest that adequate treatment may reduce this effect. Physical illness, occult disease (e.g. a carcinoma), poor self-monitoring of health, inactivity, poor adherence to treatments, and effects on the hypothalamic-pituitary-adrenal axis or other endocrine systems are possible factors.

Factors predictive of outcome

Clinical features that have been shown to be associated with a poorer outcome, include a slower initial recovery, more severe initial depression, duration of illness for more than 2 years, three or more previous episodes, a previous history of dysthymia, psychotic symptoms, and cerebrovascular disease (including vascular depression). Other factors that may affect outcome adversely are chronic stress associated with a poor environment, crime, and poverty as well as a new physical illness, becoming a victim of crime, and poor perceived social support.

The practical message is that to improve the prognosis of depression one must treat episodes early and vigorously and attend to the patient's social supports, and milieu.

(a) Does depressive disorder predispose to later dementia?

There is growing evidence from epidemiological studies that depression is a risk factor for later cognitive impairment or dementia.[11] Why this might be so is not clear but it is known that chronic depression is associated with hippocampal atrophy in older patients,[12] and that depressed patients may adopt unhealthy lifestyles, which can aggravate the risk factors for vascular disease and later dementia.

Treatment

Multi-modal management (pharmacological, psychological, and social) within a multidisciplinary framework is as important in late-life as it is at other times of life. Attending to an elderly depressed person's physical health needs, physical environment, and social needs is essential. The goal is remission of all symptoms and not merely improvement, as residual symptoms increase the chance of chronicity. Ageing and frailty result in increased dependency and less ability to adapt in a flexible way to the kinds of adversity that maintain depression. To give a simple example, good chiropody aimed at optimizing mobility can have a major positive impact alongside medical intervention, but no one aspect of treatment should be prioritized over another. Another important principle is patient-centredness. This should include giving the patients as much choice as possible regarding their treatment. Many older depressed patients, if asked, would prefer a psychological approach to medical management and there is good evidence that psychological treatments work well in older adults.[13] Resource limitations are understood, but age alone should not determine the likelihood of receiving a psychological intervention, if preferred.

Collaborative care, whereby, a depression care manager (usually a nurse, psychologist, or social worker) coordinates the care and works closely with both primary care physician and psychiatrist, is an important model for improving outcomes. In the largest study to date, 'IMPACT' from the United States, involving 1801 depressed primary care patients (major depression, 17 per cent; dysthymia, 30 per cent; or both, 53 per cent), at 12 months the Number-Needed-to-Treat (NNT) was highly significant at four in those receiving the intervention.[14]

General principles of pharmacological management include: building a therapeutic partnership with the patient; explanation of how and when antidepressants work; addressing adherence through building therapeutic concordance; tailoring antidepressants drug to the patient; arranging appropriate follow-up.

Evidence of efficacy

(a) Antidepressants

Altered pharmacokinetics, different pharmacodynamics, a greater chance of polypharmacy and hence drug interactions and reduced compensatory mechanisms are all important factors which bear upon treatment response in late-life depression.[15] An important practical consequence is to be mindful of greater inter-individual variation in drug-handling in older patients. Patients worry about antidepressants being addictive, which they are not, and that depression means 'senility', which it does not. Psychoeducation is important and can improve adherence to medication recommendations. Recommended starting and therapeutic dosages are listed in Table 8.5.4.6. These are average doses. For some drugs, notably the tricyclics, there is a very wide therapeutic range and higher levels may be required, provided they are tolerated.

A Cochrane systematic review of Randomized Controlled Trials (RCT)[16] found efficacy for antidepressants over placebo in late-life major depression. However, the analysis could only find 17 suitable studies (two of SSRIs). The NNT averaged four, but was higher for SSRIs. In a further Cochrane systematic review of studies including patients aged over 55 (29 trials), there was no difference in efficacy between tricyclics and SSRIs, but tricyclics were associated with higher withdrawal rates due to side-effects. Patients receiving tricyclic-related antidepressants (Mianserin or Trazodone) had a similar withdrawal rate to SSRIs, leading the authors to conclude that tricyclic-related antidepressants may offer a viable alternative to SSRIs in older depressed patients. The analysis also investigated 'atypical' antidepressants but it was not possible to make recommendations because of low statistical power. Atypicals included the important antidepressants reboxetine, venlafaxine, and mirtazapine.[17]

More recent RCTs have suggested less positive results.[18] Trials involving older adults and with venlafaxine, fluoxetine, citalopram, and escitalopram showed these drugs to be no more effective than placebo. Sertraline and duloxetine were superior to placebo but remission rates, as opposed to response rates, were relatively low in all these studies. Trials which are of insufficient duration and the use of antidepressants for milder depressions, for which they are ineffective, are possible factors.

(b) Depression in special patient groups

A Cochrane systematic review of antidepressants for depression in dementia found 'weak' evidence for their effectiveness,[19] but there were few admissible trials. This does not mean

Table 8.5.4.6 Suggested starting and therapeutic doses for antidepressants in the United Kingdom

Drug	Average therapeutic doses[a]	Average starting doses
Tricyclics		
Amitriptyline	75–100	25
Clomipramine	75–100	10
Dosulepin (dothiepin)	75–150	25–50
Imipramine	75–100	10–25
Nortriptyline	75–100	10–30
Lofepramine	140–210	70–140
SNRIs		
Venlafaxine	150	37.5 bd
Duloxetine	60	30–60
SSRIs		
Fluoxetine	20	20
Fluvoxamine	100–200	50–100
Paroxetine	20	20
Sertraline	50–150	50
Citalopram	20–40	20
Escitalopram	10	5
Reversible monoamine oxidase inhibitor (RIMA)		
Moclobemide	150–600	150 bd
5-HT2 receptor blocker		
Trazodone	100–300	100
NASSA		
Mirtazapine	15–30	30
Others		
Mianserin	30–90	30

[a] See text.

antidepressants are ineffective in depression-related dementia but there is a high rate of spontaneous recovery of depression complicating dementia so that 'watchful waiting' is reasonable for mild-to-moderate cases.

There is little trial data for individual antidepressants in patients with common medical disorders, but in the previously mentioned IMPACT research, both physical function and pain were improved in participants who received active case management compared to usual care.[20,21]

Depression post-stroke is common. A high rate of spontaneous recovery occurs, especially in the first 6 weeks. Studies are underway to assess whether repetitive Transcranial Magnetic Stimulation (rTMS) may be helpful in post-stroke depression. TCAs and SSRIs in standard dosages are effective in post-stroke emotionalism.

(c) Choice of antidepressant

Tricyclics often cause postural hypotension, which may lead to unpleasant dizziness or dangerous falls. Secondary amine tricyclics are generally safer in this respect than tertiary drugs. Poor left ventricular function is a risk, and so are diuretics or antihypertensive medication. Delirium is more likely in medically ill patients.

Behavioural toxicity (affecting vigilance, reaction times, etc.) has been largely ignored in the elderly. Now that so many older people drive and pursue other activities demanding high levels of vigilance, this must be addressed. The *SSRIs* are safer in this respect than tricyclics, except for lofepramine which causes less impairment than the older tricyclics. *Mirtazepine* enhances noradrenergic and serotonergic function via antagonism at the pre-synaptic α_2 receptor. Differences in pharmacodynamics and pharmacokinetics are minimal with age. The side effect profile is similar to tricyclics; weight gain and sedation can be troublesome. *Duloxetine*, like venlafaxine is a dual-acting drug. The latter has been subject to cautions, regarding heart disease, at present in the United Kingdom, although recent restrictions have been lifted. Duloxetine has some RCT evidence in older depressed adults.[22] Although a special diet with *moclobemide* is not required, patients should be aware of drug interactions with painkillers and other antidepressants. Co-prescriptions of tricyclic and SSRIs should be avoided. A wash-out period of around 4–5 half-lives of the drug and any active metabolite is advised when transferring from a tricyclic or SSRI to moclobemide (but not from moclobemide to a tricyclic or SSRI).

(d) Failure to respond to initial treatment

If a patient has made minimal or no recovery after 4 weeks of treatment at optimal dose, then the chances of recovery are slim.[23] The antidepressant may be changed to one of another class but if the patient shows at least 25 per cent improvement and is on an improving trajectory, then augmentation with lithium or a psychological intervention should be considered (see below). Electroconvulsive Treatment (ECT) is safe and effective in older patients. It is recommended when drug treatment has failed, when the patient is in danger of inanition or is acutely suicidal and is probably the treatment of choice for psychotic depression.

(e) Continuation treatment

Most relapses occur in the first 12 months,[24] so that this is a reasonable time for continuation therapy. Patients must be educated about why they should continue to take medication even when feeling better. In psychotic depression anti-psychotic medication should be continued for 6 months and gradually withdrawn if the patient is well. Following ECT, medication should be continued to avoid relapse. Limited evidence suggests either continuing antidepressants at the acute treatment dose or using lithium.

Older people have been shown to benefit substantially from maintenance therapy,[25] even after a first episode.[24] Maintenance treatment is considered later.

(f) Resistant depression

Data in the elderly are sparse but the most important consideration is a rational stepped care approach. Before moving up the steps of treatment, the following should be addressed: is the diagnosis correct (for example has a psychotic depression been overlooked)? Is poor tolerance a reason for non-recovery? Does the patient take the tablets? Have psychosocial reinforcing factors (for example, family conflict) been addressed? The steps themselves include optimizing the dose of antidepressant (relevant mainly for older tricyclic antidepressants), changing from one class of antidepressant to another, augmentation with lithium or a psychological intervention and combining antidepressants (for example, a SSRI plus mirtazepine). Finally ECT should be considered as it remains the most effective antidepressant treatment. Using such an approach 80 per cent of patients in one study responded.[26]

(g) Psychological therapies

Cognitive behavioural therapy (CBT), Interpersonal Psychotherapy (IPT), and psychodynamic psychotherapy have been shown to work in older people,[13] including in group format. CBT and IPT along with family interventions are discussed elsewhere. Problem-solving Treatment (PST) addresses the here and now, focusing on current difficulties and setting future goals.

Psychological interventions may also be important in relapse prevention. Reynolds *et al.*[27] showed in a study of older patients that monthly IPT given in the continuation phase of treatment was more effective than routine care, with combined IPT and antidepressant therapy being the most effective strategy. However, the same group were unable to replicate this in a later study with a group of patients who were somewhat older.[25]

Anxiety management can be an effective adjunctive treatment for depressed patients, especially those recovering from depression but left with residual anxiety, low confidence, or phobic avoidance, any of which can undermine functional improvement. Techniques include progressive relaxation, either alone with a commercial tape, or in groups. Exercise and activity are important both to avoid depression and counter it. Behavioural activation is a technique which can overcome the withdrawal and apathy that so often exists in late-life depression. It works by helping the patient develop a schedule of activities, agreed with the patient, with or without a written diary to support implementation.

Work to support the family and main caregivers is also important.

Prevention

(a) Primary prevention

Many prevalent diseases of later life are associated both with depression as well as lifestyle factors: diet, exercise, and obesity. Cole and Dendukuri[28] carried out a meta-analysis of risk factors for late-life depression, finding five which were robustly linked to it. These were bereavement, sleep problems, disability, prior depression, and female gender. Some of these are amenable to a public health preventative approach.

Those most vulnerable to depression will often be in touch with a home care services. This is one area where education about detection could be usefully targeted. Another example is the staff of nursing homes where depression is highly prevalent. Postgraduate training of general practitioners is another way of improving detection via the 'filter' of primary care.

(b) Secondary prevention

Maintenance treatment with a tricyclic,[24] the SSRIs citalopram[29] and paroxetine[25] or a combining medication with a psychological treatment[27] are effective prevention strategies.

Expert guidelines[30] recommend a minimum of 12 months continuation treatment for a first episode, 24 months for a second, and at least 3 years for three or more episodes. Some clinicians recommend lifelong treatment with antidepressants following even a single episode of major depression on the grounds that a substantial later period morbidity might be reduced. This must be balanced against an increased risk of side effects as patients age.

(c) Tertiary prevention

Often the emphasis is on basic explanations and on simple instructions about how to manage problems such as frequent hypochondriacal complaints or apathy. Although respite care is usually associated with dementia, there is occasionally a case for it in those with chronic treatment-resistant depression, in order to allow the relative(s) a break.

Bipolar disorder

Practice guidelines for bipolar disorder are available from the internet. Two are: *British Association of Psychopharmacology* (2003)[31]—http://www.bap.org.uk/consensus/bipolar_disorder.html and the *National Institute for Clinical Excellence (NICE) (2006)*.[32] http://www.nice.org.uk/page.aspx?o=CG38.

Mania

Clinical features

Clinical descriptions of mania in late life often portray it as atypical. However, Broadhead and Jacoby[33] found few clinical differences between 35 manic patients over the age of 60 compared to 35 younger manic patients, aged below 40. The younger manic patients were more severely ill but there was no support for the often-held view that there is a greater depressive admixture in older patients.

Diagnosis and differential diagnosis

The main differential diagnosis of mania in late life lies between a late-onset manic episode and bipolar disorder. In later presentations of bipolar disorder the time between depression and first manic episode can be us much as 15 to 40 years or more. Given this long latency it is easy to overlook bipolar disorder unless a thorough history is taken.

Epidemiology

Although only about 10 per cent of new onset cases of bipolar disorder occurs after the age of 50 they account for proportionally greater morbidity. Changing demography makes it likely that there will be more cases of bipolar disorder in later life. Episodes can be misdiagnosed; for example a depressive mood swing presenting with withdrawal or a manic one with irritability.

Aetiology

The phenomenon of conversion to bipolarity after many years of unipolar depression has led to speculation that cerebral organic factors may play a part in the aetiology of late-onset mania. In support of this, cognitive function is significantly impaired in between a fifth and a third of elderly manics.[33,34] Furthermore studies[35] have shown a high rate of neurological disturbance, cerebral deep white matter lesions, and reduced heritability in late-life mania.

The term 'secondary mania' denotes manic illness which starts without a prior history of affective disorder in close temporal relationship to a physical illness or drug treatment and often in the absence of a family history of affective illness. A large number of conditions have been associated with secondary mania, including stroke, head injury, tumours, and non-specific lesions to the right side of the brain.

Course and prognosis

The prognosis for mania is similar to that for late-life depression; that is, there is an 80 to 90 per cent recovery in the acute phase but relapses and/or recurrences occur over time in about 50 per cent of cases.[35]

Treatment

As with younger patients, the mainstays of acute treatment are neuroleptics and mood stabilizers which include lithium and anticonvulsants, with ECT reserved for refractory cases. Increasingly atypical antipsychotics are used. These include olanzapine, risperidone, and quetiapine with aripiprazole as a possible new contender. Valproate preparations are the most widely used anticonvulsant in bipolar disorder.

Some general points to consider are: (1) a greater inter-individual variability in drug metabolism, which makes predicting the therapeutic dose difficult. Emergency rapid tranquillization with haloperidol 5 to 10 mg (often with 1 to 2 mg of lorazepam) can be used, but haloperidol has a long half-life and may lead to sudden immobility after a few days; (2) balancing risks caused by overactivity and exhaustion against an increased risk of falls in the elderly when using sedative tranquillizers; (3) an increased risk of confusion and delirium if anticholinergic drugs are given to counteract side effects; and (4) the higher risk of side effects and toxicity from lithium in older patients (including a risk even at what are considered therapeutic doses in younger patient).[35] The optimal treatment dose of lithium is not known. NICE recommends levels of 0.6 to 0.8 mmol/L for adults requiring maintenance lithium. Some old age psychiatrists use lower dosages but the evidence for low dose treatment in older patients is mixed.

The NICE Bipolar Guidance also covers the treatment of bipolar depression. The main message is that antidepressants, if used, should be combined with a mood stabilizer or an antipsychotic because of the risk of a manic switch.

Prevention

Given the high rate of relapse or recurrence, prophylaxis should be considered in all patients. There has been a steady shift away from lithium to valproate over recent years, although some argue this has occurred ahead of evidence.[36] Also, there is some evidence that lithium may reduce the risk of suicide in bipolar disorder as

well as being neuroprotective, possibly reducing the risk of dementia which some population studies have been shown to be raised in affective disorder.[31]

Further information

CRUSE Bereavement Centre (helpline@crusebereavementcare.org.uk, 0870 167 1677).

Bipolar Organisation (formerly the Manic Depressive Fellowship, http://www.mdf.org.uk/; telephone 08456 340 540 [UK Only]; 0044 207 793 2600 [Rest of world]).

MIND National Association for Mental Health, http://www.mind.org.uk.

Depression Alliance (http://www.depressionalliance.org).

References

1. Blazer, D.G. (2003). Depression in late life: review and commentary. *Journal of Gerontology: Medical Sciences*, **56A**, 249–65.
2. Baldwin, R.C. (2005). Is vascular depression a distinct sub-type of depressive disorder? A review of causal evidence. *International Journal of Geriatric Psychiatry*, **20**, 1–11.
3. Dhondt, T.D.F., Beekman, A.T.F., Deeg, D.J.H., *et al.* (2002). Iatrogenic depression in the elderly results from a community-based study in the Netherlands. *Social Psychiatry and Psychiatric Epidemiology,* **37**, 393–8.
4. Copeland, J.R.M., Beekman, A.T.F., Dewey, M.E., *et al.* (1999). Depression in Europe: geographical distribution among older people. *The British Journal of Psychiatry*, **174**, 312–21.
5. Prince, M.J., Harwood, R.H., Thomas, A., *et al.* (1998). A prospective population-based cohort study of the effects of disablement and social milieu on the onset and maintenance of late-life depression. The Gospel Oak Project VII. *Psychological Medicine*, **28**, 337–50.
6. Murphy, E. (1982). Social origins of depression in old age. *The British Journal of Psychiatry*, **141**, 135–42.
7. Beekman, A.T., Geerlings, S.W., Deeg, D.J., *et al.* (2002). The natural history of late-life depression. A 6-year prospective study in the community. *Archives of General Psychiatry*, **59**, 605–11.
8. Cole, M.G. and Bellavance, F. (1997). The prognosis of depression in old age. *The American Journal of Geriatric Psychiatry*, **5**, 4–14.
9. Mitchell, A.J. and Subramaniam, H. (2005). Prognosis of depression in old age compared to middle age: a systematic review of comparative studies. *The American Journal of Psychiatry*, **162**, 1588–601.
10. Geerlings, S.W., Beekman, A.T.F., Deeg, D.J.H., *et al.* (2002). Duration and severity of depression predict mortality in older adults in the community. *Psychological Medicine*, **32**, 609–18.
11. Green, R.C., Cupples, L.A., Kurz, A., *et al.* (2003). Depression as a risk factor for Alzheimer disease: the MIRAGE study. *Archives of Neurology*, **60**, 753–9.
12. Bell-McGinty, S., Butters, M.A., Meltzer, C.C., *et al.* (2002). Brain morphometric abnormalities in geriatric depression: long term neurobiological effects of illness duration. *The American Journal of Psychiatry*, **159**, 1424–7.
13. Gatz, M., Fiske, A., Fox, L.S., *et al.* (1998). Empirically validated psychological treatments for older adults. *Journal of Mental Health and Aging*, **4**, 9–46.
14. Unűtzer, J., Katon, W., Callahan, C., *et al.* (2002). Collaborative care management of late-life depression in the primary care setting. *The Journal of the American Medical Association*, **288**, 2836–45.
15. Lotrich, F.E. and Pollock, B.G. (2005). Aging and clinical pharmacology: implications for antidepressants. *Journal of Clinical Pharmacology*, **45**, 1106–22.
16. Wilson, K., Mottram, P., Sivanranthan, A., *et al.* (2001). Antidepressant versus placebo for depressed elderly (Cochrane Review). In *The cochrane library*. Update Software, Oxford.
17. Mottram, P., Wilson, K., and Strobl, J. (2006). Antidepressants for depressed elderly. *Cochrane Database of Systematic Reviews*, (1): CD003491.
18. Roose, S.P. and Schatzberg, A.F. (2005). The efficacy of antidepressants in the treatment of late-life depression. *Journal of Clinical Psychopharmacology*, **25**, S1–7.
19. Bains, J., Birks, J.S., and Dening, T.D. (2002). Antidepressants for treating depression in dementia. *Cochrane Database of Systematic Reviews*, (4): CD003944.
20. Callahan, C.M., Kroenke, K., Counsell, S.R., *et al.* For the IMPACT investigators. (2005). Treatment of depression improves physical functioning in older adults. *Journal of the American Geriatrics Society*, **53**, 367–73.
21. Lin, E.H., Katon, W., Von Korff, M., *et al.* (2003). IMPACT investigators. Effect of improving depression care on pain and functional outcomes among older adults with arthritis: a randomized controlled trial. *The Journal of the American Medical Association*, **290**, 2428–9.
22. Nelson, J.C., Wohlreich, M.M., and Mallinckrodt, C.H. (2005). Duloxetine for the treatment of major depressive disorder in older patients. *The American Journal of Geriatric Psychiatry*, **13**, 227–35.
23. Sackeim, H.A., Roose, S.P., and Burt, T. (2005). Optimal length of antidepressant trials in late-life depression. *Journal of Clinical Psychopharmacology*, **25**, S34–7.
24. Old Age Depression Interest Group. (1993). How long should the elderly take antidepressants? A double blind placebo-controlled study of continuation/prophylaxis therapy with dothiepin. *The British Journal of Psychiatry*, **162**, 175–82.
25. Reynolds, C.F. III, Dew, M.A., Pollock, B.G., *et al.* (2006). Maintenance treatment of major depression in old age. *The New England Journal of Medicine*, **354**, 1130–8.
26. Flint, A.J. and Rifat, S.L. (1996). The effect of sequential antidepressant treatment on geriatric depression. *Journal of Affective Disorders*, **36**, 95–105.
27. Reynolds, C.F. III, Frank, E., Perel, J.M., *et al.* (1999). Nortriptyline and interpersonal psychotherapy as maintenance therapies for recurrent major depression: a randomized controlled trial in patients older than 59 year. *The Journal of the American Medical Association*, **281**, 39–45.
28. Cole, M.G. and Dendukuri, N. (2003). Risk factors for elderly community subjects: a systematic review and meta-analysis. *The American Journal of Psychiatry*, **160**, 1147–56.
29. Klysner, R., Bent-Hansen, J., Hansen, H.L., *et al.* (2002). Efficacy of citalopram in the prevention of recurrent depression in elderly patients: placebo-controlled study of maintenance therapy. *The British Journal of Psychiatry*, **181**, 29–35.
30. Alexopoulos, G.S., Katz, I.R., Reynolds, C.F., *et al.* (2001). The expert consensus guideline series: pharmacotherapy of depressive disorders in older patients *Postgrad Med Special Report*; (October):1–86, Expert Knowledge Systems, L.L.C, McGraw-Hill Healthcare Information Programs, Minneapolis, US.
31. Goodwin, G.M. (2003). For the consensus group of the British association for psychopharmacology. Evidence-based guidelines for treating bipolar disorder: recommendations from the British association for psychopharmacology. *Journal of Psychopharmacology*, **17**, 149–73.
32. National Institute for Health and Clinical Excellence. (2006). *NICE clinical guideline 38 bipolar disorder: the management of bipolar disorder in adults, children and adolescents, in primary and secondary care*. National Institute for Health and Clinical Excellence, London, UK.
33. Broadhead, J. and Jacoby, R. (1990). Mania in old age: a first prospective study. *International Journal of Geriatric Psychiatry*, **5**, 215–22.
34. Stone, K. (1989). Mania in the elderly. *The British Journal of Psychiatry*, **155**, 220–4.

35. Shulman, K. (1996). Recent developments in the epidemiology, co-morbidity and outcome of mania old age. *Reviews in Clinical Gerontology*, **6**, 249–54.
36. Shulman, K.L., Rochon, P., and Sykora, K. (2003) Changing prescription patterns for lithium and valproic acid in old age: shifting practice without evidence. s, **326**, 960–1.

8.5.5 Stress-related, anxiety, and obsessional disorders in elderly people

James Lindesay

Stress-related, anxiety, and obsessional disorders in elderly people are common, distressing, costly to services, and potentially treatable. However, despite their clinical importance, many patients still go untreated, or are treated inappropriately. The specific conditions covered here are described in detail elsewhere; this chapter focuses on the differences and difficulties that are encountered when they occur in old age.

Classification

The ICD-10 and DSM-IV diagnostic classifications are described in Chapter 1.11. Although the term 'neurotic disorder' is not used in DSM-IV, it is retained in ICD-10 as a collective term for the disorders considered in this chapter. The extensive comorbidity between these conditions and their diagnostic instability over time are also apparent in elderly people, which supports the idea that they are better considered as aspects of a general neurotic syndrome than as discrete diagnostic categories.[1] This model is particularly applicable to elderly patients, whose illnesses are often the result of a long interaction between individual vulnerability, circumstances, and maladaptive responses to distress. Obsessive–compulsive disorder (OCD) is probably not part of a general neurotic syndrome. Although classified with the anxiety disorders in ICD-10 and DSM-IV, it has a number of features that suggest it is a distinct and stable condition with a different aetiology (see Chapter 4.8).

Clinical features

These disorders have psychological, somatic, and behavioural features. In elderly people, these symptoms and behaviours are similar to those seen in younger patients, but there are some important differences in how they manifest themselves or are perceived by others. Although most neurotic disorders in elderly patients are long standing, an important minority of cases have their onset in old age, and it is these that usually cause the greatest diagnostic difficulties.

Psychological symptoms

Symptoms of anxiety and depression occur to some extent in all of these disorders in late life. Depressive symptomatology in old age is described elsewhere (see Chapter 8.5.4). Regarding anxiety, the focus of the worries and fears of elderly people is on those issues that are of general concern in this age group (health, finances, crime). The phobias described by elderly people are similar to those seen in younger adults,[2] although some, such as the fear of falling, are more commonly seen in old age. Clinically significant anxieties and fears in elderly people are often dismissed as reasonable purely on grounds of age. In fact, it is physical frailty and the availability of social support that determine elderly people's perceptions of vulnerability and risk, and these rather than age should be considered when deciding whether or not concerns are reasonable.

The clinical features of OCD in old age are similar to those seen in younger patients. Obsessional symptoms rarely appear for the first time after the age of 50 years, and in such cases the possibility of an organic cause such as dementia or a space-occupying lesion should be investigated. They may also form part of a primary affective disorder.

Somatic symptoms

The somatic symptoms of anxiety are similar at all ages, but in elderly patients there is a greater likelihood of misdiagnosis and inappropriate investigation and treatment. This is particularly true of elderly patients experiencing panic attacks, who tend to be misdirected to cardiologists, neurologists, and gastroenterologists.

Behavioural disturbance

The psychological and somatic symptoms of anxiety have several adverse behavioural consequences, for example, phobic avoidance, the abuse of sedative drugs and alcohol, and the development of troublesome abnormal illness behaviours such as somatization and hypochondriasis. In elderly patients these behaviours are usually of long standing, but they can develop following the onset of anxiety or depression in old age. In cognitively impaired patients, disturbed behaviour may be the main presenting feature.

Diagnosis and differential diagnosis

In old age, these disorders usually present in primary care and the general hospital, and clinicians working in these settings need to be able to identify them, and to distinguish them from the other mental and physical disorders that they may accompany or mimic.

Depression

There is extensive comorbidity between neurotic disorders and depression, and depressive symptoms are an integral component of many neurotic disorders, particularly in old age. It is therefore important to assess to what extent depression forms part of the clinical picture, as this may require treatment in its own right. Depressive disorder that is comorbid with anxiety responds less well to antidepressant treatment, and there is a greater likelihood of relapse and recurrence.

Dementia

In the early stages, dementia may present with symptoms such as anxiety, and obsessionality. More commonly, anxiety and depression cause subjective cognitive impairment, which may be the presenting symptom. Dementia is associated with higher rates of anxiety, unrelated to severity of cognitive impairment. Patients with vascular dementia may be more vulnerable in this respect.

This anxiety may be associated with the implications of the diagnosis in those patients who retain insight, or a response to psychotic symptoms or misinterpretations of the external environment in those who are more severely affected. The caregivers of people with dementia are also vulnerable to developing depressive and anxiety disorders, particularly if they have a previous psychiatric history.

Delirium

Although delirium is a relatively quiet disorder in elderly patients (see Chapter 8.5.1), it may be associated with significant affective disturbances, often in response to frightening visual hallucinations and imagined assaults. Conversely, in vulnerable individuals, severe anxiety may be sufficient to precipitate delirium.

Paranoid states and schizophrenia

Patients suffering from these disorders may experience significant fear and anxiety in response to their psychotic experiences, but this rarely causes diagnostic difficulty. Unusual hypochondriacal ideas may sometimes be difficult to distinguish from monosymptomatic delusional disorders.

Physical illness

There is an important association between physical illness and neurotic disorders in old age. As a life event, an episode of physical illness may be the cause of neurotic disorder, particularly if it is severe or has sinister implications. For example, mild anxiety symptoms are common following myocardial infarction in old age, and vulnerable individuals may develop a disabling 'cardiac neurosis' focused on their somatic anxiety symptoms. Most cases of agoraphobia that develop after the age of 65 years are not induced by panic but arise following an alarming experience of physical ill health.[1] Follow-up studies of stroke survivors show that conditions such as agoraphobia and generalized anxiety are common, tend to become chronic in a significant proportion of cases, and are associated with poor functional recovery.[3] Chronic disabilities that limit mobility and independence, such as arthritis, balance disorders, and sensory impairments, increase the patient's sense of personal vulnerability and are also associated with elevated rates of anxiety and secondary avoidance.

Neurotic disorders can also cause physical illness by direct or indirect effects on the body. In elderly people, this may come about as the result of many years of harmful anxiety-driven behaviours such as smoking and alcohol abuse.

In terms of differential diagnosis, there is also the problem that a wide range of physical disorders may present with neurotic symptoms, and vice versa. In particular, a number of important cardiovascular, respiratory, and endocrine disorders may present with anxiety or depression and little else in old age.[4] Anxiety symptoms may also be caused by prescribed drugs such as oral hypoglycaemics and corticosteroids, or by excessive intake of caffeine and preparations containing sympathomimetics. In view of this, the clinical assessment should always include a drug history and a physical examination. A physical cause for neurotic symptoms should be considered if there is no past psychiatric history and no life event or other circumstances to account for their onset.

Epidemiology

While neurotic disorders are relatively uncommon in clinical populations, there are significant prevalence rates in community samples, indicating that they do not pass easily through the filters on the pathway to care. Surveys using different diagnostic criteria produce different rates of disorder, which makes comparisons difficult. However, some general findings include a female preponderance for most disorders, and a fall in prevalence and incidence rates with age. Most elderly people with neurotic disorders developed them before their fifties, but elderly cases of phobic disorder, panic, and OCD tend to be of later onset.[5,6]

Aetiology

The acquisition and subsequent loss or elaboration of the symptoms of anxiety and depression are determined by the patient's premorbid vulnerability, the particular factors that precipitate the episode of illness (destabilization), and the measures taken by the patient or the doctor to control it (restitution).[7]

Biological factors

Most of the evidence that biological factors play a role in the development of neurotic disorders derives from studies in younger subjects (see Chapters 4.7.1–4.7.3). These studies indicate that genetic factors contribute significantly to premorbid vulnerability, but the role that they play in old age is not known. Neuroimaging studies of elderly depressed patients provide only limited information about neurotic disorders, suggesting that patients with milder forms of depression and higher anxiety scores are more likely than severely depressed patients to have normal CT scans. It has been suggested that some anxiety disorders following stroke may be related to lesion location.[8]

Psychosocial factors

(a) Social adversity

This may have its effect through higher rates of physical illness and exposure to adverse life events, or by inculcating a sense of poor self-esteem. However, the impact of adversity on self-esteem in old age is not clear, since a hard life may in fact equip one to cope better with the difficulties of old age.

(b) Life events

Adverse life events have an important role in determining the onset of depressive and anxiety disorders (see Chapters 4.5.5.1 and 4.7.1–4.7.3). It is the meaning of the event to the individual that is important; loss events lead to depression and threatening events to anxiety. Some types of life events (physical illness, bereavement, retirement, institutionalization) are more common in old age, and are associated with psychiatric morbidity in vulnerable individuals.

Extreme trauma and catastrophe are well known to have adverse psychological consequences, and post-traumatic stress disorder (PTSD) is a recognized diagnosis in ICD-10 and DSM-IV (see Chapter 4.6.2). It is clear from studies of elderly survivors of traumatic experiences such as war and the holocaust that PTSD is often a persistent disorder, and that its onset or recurrence may be precipitated by events many years after the original traumatic

experience. PTSD may also develop following trauma in late life, and as in younger subjects it tends to persist.

(c) Early experience

Early parental loss and childhood physical and sexual abuse are associated with the development of mental disorder in adult life, an effect that persists into old age.[1]

(d) Relationships

There is evidence that both the quantity and quality of social relationships are important determinants of psychological well-being in old age, and that factors such as smaller social networks are associated with anxiety disorders.[9]

Course and prognosis

The limited evidence suggests that neurotic disorders in elderly people tend to become chronic and that older age of onset is a predictor of poor outcome, particularly in men.[10] The pattern of symptoms may change over time.[11] Elderly patients with anxiety disorders have an excess mortality.

Treatment

Evidence

There is little high-quality evidence available about the effects of treatments for these disorders in elderly patients. Cognitive behaviour therapy (CBT) is of proven benefit in younger adults (see Chapter 6.3.2.1), and there is some evidence that it is also effective in later life.[12] There have been surprisingly few trials of anxiolytic drugs in elderly patients, but there is evidence that antidepressant drugs such as SSRIs and venlafaxine are effective in the treatment of generalized anxiety. In the absence of good evidence for a particular treatment, patient preference, and choice are important considerations.

Management

Since most patients are seen in primary care and general medical settings, this is where the focus of management should be. The role of specialist old age psychiatry services should be to provide any advice and support that is necessary, which may include assuming responsibility for the most complex cases. Wherever the patient is seen and treated, the following should need consideration from the outset (see also Chapter 8.6):

- a thorough assessment and accurate diagnosis as the basis for the management plan
- the full range of physical and psychological treatment options, including patient education, lifestyle advice, bibliotherapy, and supportive counselling
- clear goals for the treatment plan, agreed if possible with the patient
- an adequate trial of treatment
- the likely duration of treatment
- frequency of review
- any possible adverse consequences, such as dependency, adverse side effects, risk of self-harm.

(a) Psychological treatments

For the most part, the goals and techniques of CBT are the same for elderly patients as they are for younger adults (see Chapter 6.3.2.1). However, these may require some adaptation to accommodate sensory impairments, physical illness and disability, and cognitive dysfunction.[13] The need to tailor treatment to the individual may limit the value of CBT in a group setting, although this has to be set against the benefits of shared experience and peer support. Group treatment is probably more straightforward with task-centred activities such as anxiety management.

The use of formal psychodynamic approaches to management is currently limited by economic constraints, and the lack of evidence regarding their effectiveness. However, health professionals should have some knowledge of the psychodynamics that underlie the concerns of elderly patients, and the mental defences that they use. They also need to be aware of their own preconceptions and cognitive distortions regarding the experience of old age, the psychological sophistication of elderly people, and their capacity for growth and change.

(b) Physical treatments

Chapter 8.6 describes the general principles of drug treatment in old age. None of the drugs used to treat anxiety in elderly people is entirely without problems, so they should be prescribed with care.[14] Benzodiazepines are the most commonly used drugs, but they are often prescribed inappropriately. Elderly patients are particularly sensitive to their adverse effects, and drug accumulation may lead to delirium, incontinence, and falls. Compounds with short half-lives and no active metabolites, such as oxazepam, are least problematic, although patients may develop withdrawal symptoms if they are discontinued, or taken erratically. Long-term benzodiazepine use should be avoided where possible, although it may be necessary in a few patients unresponsive to other forms of treatment.

Antidepressant drugs are now the first choice in generalized anxiety and panic, particularly if depressive symptoms are prominent. The use of antidepressants in elderly patients is discussed in Chapter 8.5.4. SSRIs also have a specific effect in OCD. Neuroleptics have only a limited role in the management of anxiety, given their potentially disabling extrapyramidal side effects. However, a short course of low-dose treatment with a drug such as haloperidol or zuclopenthixol may be considered in those unable to tolerate benzodiazepines. Alternatively, sedative antihistamine drugs such as hydroxyzine may be useful. β-blockers are used in younger adults to control the sympathetic somatic anxiety symptoms, but contraindications such as chronic obstructive airways disease, sinus bradycardia, and heart failure limit their use in elderly patients. Buspirone is an azapirone anxiolytic that is well tolerated by elderly patients, but it takes about 2 weeks to become effective, so is not useful for the management of acute episodes. It is indicated for severe chronic generalized anxiety and in patients where there is risk of dependence or abuse.

Prevention

The possibilities for primary prevention are limited at present. However, in view of the association with physical illness, there may be an opportunity to intervene and prevent the development of chronic neurotic disability following strokes, heart attacks, and

falls. It remains to be seen if the improved management of these disorders earlier in life will result in lower rates of chronicity and recurrence as cohorts age.

Further information

Lindesay, J. (ed.) (1995). *Neurotic disorders in the elderly*. Oxford University Press, Oxford.

Lindesay, J. (2008). Neurotic disorders. In *The Oxford textbook of old age psychiatry* (eds. R. Jacoby, T. Dening, A. Thomas, and C. Oppenheimer). Oxford University Press, Oxford.

National Institute for Clinical Excellence. (2004). *Anxiety: management of anxiety (panic disorder with or without agoraphobia and generalised anxiety disorder) in adults in primary, secondary and community care*. Clinical Guideline 22. NICE, London (www.nice.org.uk/CG022NICEguideline).

Ruskin, P.E. and Talbot, J.A. (eds.) (1996). *Aging and posttraumatic stress disorder*. American Psychiatric Press, Washington, DC.

Salzman, C. and Lebowitz, B. (eds.) (1991). *Anxiety in the elderly*. Springer, New York.

References

1. Tyrer, P. (1989). *Classification of neurosis*. Wiley, Chichester.
2. Lindesay, J. (1991). Phobic disorders in the elderly. *The British Journal of Psychiatry*, **159**, 531–41.
3. Astrom, M. (1996). Generalized anxiety disorder in stroke patients. A 3-year longitudinal study. *Stroke*, **27**, 270–5.
4. Pitt, B. (1995). Neurotic disorders and physical illness. In *Neurotic disorders in the elderly* (ed. J. Lindesay), pp. 46–55. Oxford University Press, Oxford.
5. Robins, L. and Regier, D. (eds.) (1991). *Psychiatric disorders in America*. Free Press, New York.
6. Krasucki, C., Howard, R., and Mann, A. (1998). The relationship between anxiety disorders and age. *International Journal of Geriatric Psychiatry*, **13**, 79–99.
7. Goldberg, D. and Huxley, P. (1992). *Common mental disorders: a bio-social model*. Tavistock/Routledge, London.
8. Robinson, R.G. (1998). *The clinical neuropsychiatry of stroke*. Cambridge University Press, Cambridge.
9. Beekman, A.T.F., Bremmer, M.A., Deeg, D.J.H., *et al.* (1998). Anxiety disorders in later life: a report from the longitudinal aging study Amsterdam. *International Journal of Geriatric Psychiatry*, **13**, 717–26.
10. Noyes, R. and Clancy, J. (1976). Anxiety neurosis: a 5-year follow-up. *The Journal of Nervous and Mental Disease*, **162**, 200–5.
11. Larkin, A.B., Copeland, J.R.M., Dewey, M.E., *et al.* (1992). The natural history of neurotic disorder in an elderly urban population. Findings from the Liverpool longitudinal study of continuing health in the community. *The British Journal of Psychiatry*, **160**, 681–6.
12. Mohlman, J. (2004). Psychosocial treatment of late-life generalized anxiety disorder: current status and future directions. *Clinical Psychology Review*, **24**, 149–69.
13. Koder, D.A. (1998). Treatment of anxiety in the cognitively impaired elderly: can cognitive-behavior therapy help? *International Psychogeriatrics*, **10**, 173–82.
14. Suribhatla, S. and Lindesay, J. (2005). Treatment of anxiety disorders. In *Practical old age psychopharmacology*, Chap. 10 (eds. S. Curran and R. Bullock). Radcliffe, Abingdon.

8.5.6 Personality disorders in the elderly

Suzanne Holroyd

Introduction

The study of personality disorder (PD) in late life presents conceptual, diagnostic, and methodological difficulties. By definition, PD is considered a group of personality traits that relatively persistent through adulthood. However, the concept of PD persisting throughout the lifespan contradicts widespread clinical belief that they become less severe with ageing. For example, DSM-IV[1] notes that 'some types of personality disorders . . . tend to become less evident or remit with age'.

There are difficulties in studying PD in the elderly. One is the instability of the definition of PD over time, making it difficult to relate earlier studies to those using current definitions of PD. In addition, diagnostic criteria are subject to criticism when applied to the elderly, in that they may be 'age-biased'.[2,3] Finally, the methodology used to diagnose PD has been highly variable and difficult to interpret between studies.

A major issue is whether personality is fully developed by early adulthood and then remains unchanged, or whether personality continues to develop and change throughout life. The work of McCrae and Costa[4] demonstrated that personality characteristics are relatively stable within individuals over a 30-year period with correlations ranging from 0.7 to 0.8. However, this also demonstrates that complete stability is not there and suggests certain aspects of the personality may still develop and change with ageing. This suggests that PD may also change over the lifespan.

Another issue is whether underlying traits that persist throughout the lifespan can rise to the level of a PD depending on the environment. For example, traits that may be personality disordered in young adult life, such as extreme dependency, may be an appropriate and adaptive trait for an older individual with multiple physical disabilities.[5] Conversely, an individual may have a trait of extreme independence that may be adaptive in earlier life but which leads to distress and maladaptive functioning in a setting requiring dependence, such as a nursing home. Thus it is possible to have a PD diagnosed for the first time in late life, which goes against the very definition of lifelong PD.

Clinical features

Clinical features of PD and details of their classification are reviewed in Chapters 4.12.2 and 4.12.3. However, there is difficulty in simply relating these criteria, which were developed for younger individuals, to the elderly. Typical diagnostic behaviours that clinicians associate with PD in younger adults may present as different behaviours in the elderly. This may lead clinicians to overlook personality traits and disorders in older individuals. For example, a criterion of antisocial personality disorder is the repeated failure to sustain consistent work, behaviour not applicable to the older retired individual. Yet it is possible that the personality trait of irresponsibility, which led to the loss of jobs earlier in life continues, now appearing as a behaviour such as medicine non-compliance. Some authors have thus argued that the clinical

features for some PD are age-biased since certain behaviours are less likely to occur in elderly persons despite the persistence of personality traits.[2,3]

Diagnosis of personality disorders

Diagnostic criteria for PD are discussed in Chapter 4.12.3. Because features of PD may change with ageing, diagnosis can be difficult. Overlap with Axis I diagnoses such as depression or dementia make the diagnosis even more challenging. For example, depressed elderly people have symptoms normally associated with PD as they may be more dependent, avoidant, resistant, negative, and somatic.[6] In addition, depressed elderly people may view their lives negatively and overestimate personality psychopathology.[7,8]

Clinicians may be reticent to give a personality diagnosis to an individual with multiple medical problems to which maladaptive behaviours may be attributed even if a lifelong history of personality pathology is established.[8] They may also be concerned about the validity of historical information needed to make a PD diagnosis. Therefore, in making a diagnosis of PD, a clinician should take a thorough history from as many reliable outside informants as possible. If the patient is in a state of acute distress with a current Axis I diagnosis such as depression, it is best to defer diagnosis of the PD until the illness is in remission. Otherwise, it is especially important to ask outside informants to think back to when the individual was a younger person, as current symptoms can colour the perception of lifelong personality traits. Asking for specific examples of history such as, details of relationships and job history, legal history, and the like, will be more helpful than just general descriptions of personality.

If behavioural difficulties and personality problems are found to be recent, the clinician needs to search carefully for a superimposed medical condition, or a psychiatric condition such as depression. Clinicians should carefully screen for illnesses such as dementia, stroke, or other neurological disease, or a systemic medical illness. Those with frontal lobe dementia, Alzheimer's disease, or vascular dementia may have personality changes early in their life course.[9–11]

Epidemiology and aetiology

The prevalence of PD in the elderly varies as to the methodology used and the population studied. It should be noted that no assessment instrument for PD in the elderly has been validated.

Community studies

Community studies have been the most useful to date. A community study,[12] using the Epidemiologic Catchment Area (**ECA**) data, had 841 subjects examined by psychiatrists using the semi-structured Standardized Psychiatric Examination with DSM-III criteria. Comparing those over the age of 55 with those under 55, older individuals were found significantly less likely to have a PD (6.6 to 10.5 per cent) as compared with younger individuals. This finding was almost entirely due to a three-fold higher prevalence of cluster B PD in those under the age of 55, especially antisocial and histrionic PD. Interestingly, in this study none of the older individuals were found to have cluster A PD. Table 8.5.6.1 summarizes the findings of this large community study. The strengths of this study were that it was a community rather than a clinical sample, and subjects were evaluated by psychiatrists using a structured questionnaire. Limitations of this study were those inherent to the study of PD in late life, in that older subjects may have been inaccurate in recalling maladaptive behaviours, outside informants were not used, and lack of non-validated instruments for diagnosing PD in the elderly.

Table 8.5.6.1 Weighted prevalence (%) of DSM-III personality disorders in a large community study

	Age < 55 years	Age > 55 years
Cluster A	0.1	0.0
Paranoid	0.0	0.0
Schizoid	0.1	0.0
Schizotypal	0.1	0.0
Cluster B	6.8	2.2*
Antisocial	2.7	0.1*
Borderline	0.8	0.0
Histrionic	4.3	2.2*
Narcissistic	0.0	0.0
Cluster C	3.8	4.3
Avoidant	0.0	0.0
Dependent	0.2	0.1
Obsessive–compulsive	3.6	3.3
Passive-aggressive	0.0	1.0
Any personality disorder	10.5	6.6*

* $p<0.05$

Reproduced from B.J. Cohen *et al.* (1994). Personality disorders in later life. A community study. *British Journal of Psychiatry*, **165**, 493–9, copyright 1994, The Royal College of Psychiatrists.

A community study of 43 093 persons, examining alcohol and related conditions across the life span, confirmed that those over 65 years had significantly lower rates of all studied PD including avoidant, obsessive–compulsive, paranoid, schizoid, histrionic, and antisocial, using DSM-IV criteria.[13]

A community survey study of DSM-III PD traits using the Personality Diagnostic Questionnaire revealed that 'dramatic' and 'anxious' personality traits declined up to 60 years of age with a slight increase thereafter, but that 'odd' or 'eccentric' traits showed no change with age.[14]

Psychiatric populations

In addition to community samples, specific clinical samples have been examined. Limitations of these studies are the possibility of over diagnosis of PD due to symptoms of Axis I diagnoses.

A retrospective study of 2322 psychiatric inpatients with major depression found the prevalence of PD to be 11.2 per cent in those over the age of 65 as compared with 17.2 per cent for those under 65.[15]

Psychiatric inpatient studies suffer from the limitation of diagnosing PD in the face of an acute psychiatric illness requiring hospitalization, making them likely to overdiagnose. Such studies are of very limited value.[16,17]

Unfortunately, outpatient studies have similar limitations when examining those with concurrent Axis I diagnoses. A study of 36 psychiatric outpatients, including those with bipolar disorder, delusional disorder, and schizophrenia, revealed that 58 per cent had a diagnosis of personality disorder.[18] Arguably, diagnosing

personality disorder in the face of these disorders is likely to be difficult and result in an overestimation.

Prevalence summary

A meta-analysis of 11 articles published from 1980 to 1994 of personality disorders based on DSM-III or DSM-IIIR criteria revealed an approximately 10 per cent prevalence of personality disorders in those aged 50 and over.[19] In comparing these studies it was noted that the method of diagnosis affected the prevalence of personality disorder. In conclusion, the authors felt that there was a definite need for well-designed studies using statistically robust samples to assess the true prevalence of personality disorders in late life.

Taking the best studies together—the ECA community study and the meta-analysis—the prevalence of personality disorder in the elderly, as currently defined, ranges from 7 to 10 per cent, with a decrease in prevalence of cluster B diagnoses.

Course and prognosis

Longitudinal community studies of PD have not been performed. With longitudinal data lacking, only cross-sectional studies are available. However, cross-sectional studies have a variety of limitations, including the possibility of a cohort effect explaining changes in prevalence in late life.

Antisocial personality disorder has been the best studied. The ECA study revealed that antisocial personality disorder declined from a 1-month prevalence of 0.9 per cent for individuals between 25 and 44 years of age to 0 per cent for those over the age of 65. When considering men only, the rate fell from 1.5 per cent in those aged 22 to 44 to 0.1 per cent in those over 65.[20] Supporting this is a study revealing the decline in lifetime prevalence in antisocial personality disorder from between 2.1 and 3.3 per cent to between 0.2 and 0.8 per cent in those aged 65 and older.[21] In addition, antisocial traits, as measured by the Minnesota Multiphasic Personality Index, reveal a decline with ageing.[22] However, a forensic centre study revealed while antisocial PD declined after the age of 27, one-third remained criminally active throughout their lives.[23]

Several hypotheses exist to explain this apparent decline in antisocial personality disorder with ageing. Personality may continue to mature and develop. Early death due to high-risk behaviour or a change in antisocial behaviours to other symptoms including hyperchondriasis, depression, or alcoholism may occur.[23] Also, behaviours such as criminality may decrease in older individuals, but antisocial personality traits remain and are simply more difficult to measure using current diagnostic criteria. Also, decrease in impulsive and aggressive behaviours may correlate with full myelination of frontal, temporal, and parietal cortices that does not occur until 30 or 40 years of age.[24] Changes in brain neurochemistry with ageing, including serotonin and dopamine, may also result in decreased impulsiveness or aggressiveness.[12] Decreased testosterone levels in men with ageing may contribute to a decline in these traits.[25]

Other Cluster B disorders may decline with ageing. The ECA study previously reviewed found a decline in antisocial and histrionic disorder.[23] Another community study supported a decline in histrionic PD with ageing.[26] Interestingly, the pattern of decline varied with gender, with rates remaining constant in women but declining in men. Similarly, a diagnosis of borderline PD is rare in elderly individuals, with only two case reports in the literature.[5] There is conflicting data regarding a decline in cluster A or C diagnoses. A large community study revealed lower rates of both Cluster A and C diagnoses in the elderly.[13] However, a study of schizotypal PD revealed all cases began before 40 years of age and continued lifelong.[27]

Some work has been done on the interaction of PD with Axis I diagnoses. Studies of depressed elderly patients suggest PD is associated with earlier age of depression onset, chronicity, and severity of dysthymia[28,29] however depression may exacerbate or conceal personality traits, thus making firm conclusions of PD in such individuals difficult. In the ECA study, certain Axis I diagnoses were found to be more common with a PD diagnosis. For example, all cases of obsessive–compulsive disorder in older individuals occurred concurrent with a PD.[12] Both generalized anxiety disorder and substance use disorders were more common in the presence of a PD. There were no differences in the prevalence of schizophrenia and major depression in those with or without a PD. The findings need to be confirmed since the group of elderly with PD was small.

A recent interesting study has revealed any PD (DSM-IV criteria) is associated with increased risk of stroke and ischemic heart disease, adding to the possible morbidity of these disorders.[30] However, such results should be viewed with caution as no screening was done to rule out depression or other associated factors that are prevalent in this population and may have led to over diagnosis of PD.

Treatment and management

Given the relative lack of data regarding PD in the elderly, it is not surprising there is a corresponding lack of information regarding treatment and management. In general, clinicians should have a low threshold for suspecting concurrent psychiatric diagnosis, as major depression, anxiety, substance use disorders or dementia may mimic or exacerbate a personality disorder.[5] Physical and medical problems should be thoroughly evaluated and treated to minimize any associated complaints.

Social and family supports should be explored and maximized. Firm and consistent limits must be set by the clinician for both patients and their families in regard to inappropriate behaviour.[5] Clinicians should also try to determine why the disordered behaviour is occurring at a particular time. For example, placement in a nursing home may be stressful to an individual who has had difficulty forming relationships and is now dependent on a group of caregivers. Psychotherapy with the goal of focusing on current life stresses, the individual's vulnerabilities, and adaptive strategies can help the patient adjust to the current circumstance.[5]

Psychotherapeutic treatments used for personality disorders in younger individuals may be tried although little data exists on their effectiveness for elderly. A study of Dialectical Behavior Therapy (DBT) used in combination with medication to treat elderly depressives with personality disorders revealed better results than just medication. The study is limited in that such results are common in other studies using any psychotherapy with medication versus medication alone and does not support a specific usefulness of DBT in elderly personality disorder. However, the study is the first to use DBT in an older population and shows it is tolerable in this age group.

If possible, psychiatric medication should be avoided unless there is a specific diagnosed condition. This will minimize the possibility

of side effects in elderly individuals and avoid dependency and control issues.[5] This is important as elderly individuals with abnormal personality traits have been found to be at higher risk of receiving psychotropic medication.[31]

Possibilities for prevention

There are no data available for preventing the development of a PD in late life. Clearly, more information is needed on the longitudinal course of diagnosed PD in late life so that information regarding treatment and prevention may be realized.

Further information

At the time of writing, there are no books or reports that give an overview of personality disorder in the elderly. Further information about specific aspects of the subject can be obtained from the relevant references in the text.

References

1. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorder* (4th edn). American Psychiatric Association, Washington, DC.
2. Agronin, M.E. and Maletta, G. (2000). Personality disorders in late life: understanding and overcoming the gap in research. *The American Journal of Geriatric Psychiatry*, **8**, 4–18.
3. Kroessler, D. (1990). Personality disorder in the elderly. *Hospital & Community Psychiatry*, **41**, 1325–9.
4. McCrae, R.R. and Costa, P.T. (1984). *Emerging lives, enduring dispositions*. Little Brown, Boston, MA.
5. Holroyd, S. and Rabins, P.V. (1994). Personality disorders. In *Principles of geriatric medicine and gerontology* (3rd edn) (eds. W.R. Hazzard, E.L. Bierman, J.P. Blass, W.H. Ettinger, and J.B. Hatter), pp. 1131–6. McGraw–Hill, NY.
6. Thompson, L.W., Gallagher, D., and Czirr, R. (1988). Personality disorder and outcome in the treatment of late-life depression. *Journal of Geriatric Psychiatry*, **21**, 133–46.
7. Hirschfeld, R.M.A., Klerman, G.L., and Clayton, P.J. (1983). Assessing personality: effects of the depressive state on trait measurement. *The American Journal of Psychiatry*, **140**, 695–9.
8. Reich, J., Noyes, R., Hirschfeld, R.M.A., *et al.* (1987). State and personality in depressed and manic patients. *The American Journal of Psychiatry*, **144**, 181–7.
9. Petry, S., Cummings, J.L., and Hill, M.A. (1990). Personality alterations in dementia of the Alzheimer type. *Archives of Neurology*, **45**, 1187–90.
10. Neary, D., Snowden, J., and Mann, D. (2005). Frontotemporal dementia. *Lancet Neurology*, **4**, 771–80.
11. Dian, L., Cummings, J.L., Petry, S., *et al.* (1990). Personality alterations in multi–infarct dementia. *Psychosomatics*, **31**, 415–19.
12. Cohen, B.J., Nestadt, G., Samuels, J.F., *et al.* (1994). Personality disorders in later life. A community study. *The British Journal of Psychiatry*, **165**, 493–9.
13. Grant, B.F., Hasin, D.S., Stinson, F.S., *et al.* (2004). Prevalence, correlates and disability of personality disorders in the United States: results from the national epidemiologic survey on alcohol and related conditions. *The Journal of Clinical Psychiatry*, **65**, 948–58.
14. Reich, J., Nduaguba, M., and Yates, W. (1988). Age and sex distribution of DSM–III personality cluster traits in a community population. *Comprehensive Psychiatry*, **29**, 298–303.
15. Fogel, B.S. and Westlake, R. (1990). Personality disorder diagnoses and age in inpatients with major depression. *The Journal of Clinical Psychiatry*, **51**, 232–5.
16. Molinari, V. (1993). Personality disorders and relapse rates among geropsychiatric inpatients. *Clinical Gerontologist*, **14**, 49–52.
17. Molinari, V., Ames, A., and Essa, M. (1994). Prevalence of personality disorders in two geropsychiatric inpatient units. *Journal of Geriatric Psychiatry and Neurology*, **7**, 209–15.
18. Molinari, V. and Marmion, J. (1993). Personality disorders in geropsychiatric outpatients. *Psychological Reports*, **73**, 256–8.
19. Abrams, R.C. and Horowitz, S.V. (1996). Personality disorders after age 50: a meta–analysis. *Journal of Personality Disorders*, **10**, 271–81.
20. Regier, D.A., Farmer, M.E., Rae, D.S., *et al.* (1988). One–month prevalence of mental disorders in the United States. *Archives of General Psychiatry*, **45**, 977–86.
21. Robins, L.N. (1984). Lifetime prevalence of specific psychiatric disorders in three sites. *Archives of General Psychiatry*, **41**, 949–58.
22. Weiss, J.M. (1973). The natural history of antisocial attitudes–what happens to psychopaths? *Journal of Geriatric Psychiatry*, **6**, 236–42.
23. Arboleda–Florez, J. and Holley, H.L. (1991). Antisocial burnout: an exploratory study. *Bulletin of the American Academy of Psychiatry and the Law*, **19**, 173–83.
24. Elliott, F.A. (1992). Violence, the neurologic contribution: an overview. *Archives of Neurology*, **49**, 595–603.
25. Gray, A., Jackson, A.N., and McKinlay, J.B. (1991). The relation between dominance anger and hormones in normally aging men: results from the Massachusetts Male Aging Study. *Psychosomatic Medicine*, **53**, 375–85.
26. Nestadt, G., Romanoski, A.J., Chahal, R., *et al.* (1990). An epidemiologic study of histrionic personality disorder. *Psychological Medicine*, **20**, 413–22.
27. Baron, M., Given, R., Asnis, L., *et al.* (1983). Age of onset in schizophrenia and schizotypal disorder. *Neuropsychobiology*, **10**, 199–204.
28. Abrams, R.C., Rosendahl, E., and Card, C. (1994). Personality disorder correlates of late and early onset depression. *Journal of the American Geriatric Society*, **42**, 727–31.
29. Devanand, D.P., Turret, N., and Moody, B.J. (2000). Personality disorder in elderly patients with dysthymic disorder. *The American Journal of Geriatric Psychiatry*, **8**, 188–95.
30. Moran, P., Stewart, R., Brugha, T., *et al.* (2007). Personality disorder and cardiovascular disease: results from a national household survey. *The Journal of Clinical Psychiatry*, **68**, 69–74.
31. Mann, A.H. (1981). The twelve–month outcome of patients with neurotic illness in general practice. *Psychological Medicine*, **11**, 535–50.

8.5.7 Suicide and deliberate self-harm in elderly people

Robin Jacoby

Introduction

Although in some countries suicide rates in young males have risen dramatically in the last decade or so, suicide in old age is important because rates in older people, especially those over 74, are still proportionately higher in most countries of the world where reasonably reliable statistics can be obtained.[1] For example, in 2004 in Lithuania where suicide incidence is currently the highest, the

overall rate in males per 100 000 total population was 70.1, but in men over 74 the rate was 80.2. In the United States, where suicide is neither especially common nor rare, in 2002 the overall rate for males per 100 000 total population was 17.9, but 40.7 in men over 74. Rates for older women are nearly always much lower than for their male counterparts.

A second reason for the importance of suicide in old age is that the proportion of older people in the population is rising worldwide. Indeed, the increase in developing countries is likely to be even greater than in developed countries. Although rates vary from year to year and birth cohort to cohort, it is highly likely that unless suicide prevention becomes a great deal more effective than at present, more and more older people will kill themselves in the coming years.

As with younger people, completed suicide in old age may be seen as part of a continuum from suicidal thinking through deliberate self-harm (which does not lead to death), to completed suicide. An added component within this continuum for older people is that of 'indirect self-destructive behaviour', such as refusal to eat and drink or 'turning one's face to the wall' which is clearly intended to hasten death. Finally, although this section does not deal with euthanasia and related issues, assisted suicide in people with terminal illness such Alzheimer's disease and cancer may also be seen as part of the suicide continuum.

Suicidal thinking in community-dwelling elderly people

A number of studies have explored this issue. Fleeting thoughts of suicide or the idea that life is not worth living occur in up to about 15 per cent of community-dwelling older people,[2] but serious consideration of suicide is very much less.[3] It is those older people with mental disorders, mostly depressive, who show a higher frequency of thoughts that life is not worth living and harbour ideas of committing suicide. It seems logical to suppose, therefore, that depressed elders should be the target of suicide prevention strategies.

Indirect self-destructive behaviour

Unlike the young, some elderly people have the possibility open to them of behaving passively in such a way as to hasten death. This may happen either by refusing medical treatment essential to maintain life, or simply by declining to eat and drink—'turning one's face to the wall'. As regards the latter, many people, especially non-medical, believe that this is reasonable behaviour akin to so-called 'rational suicide', and court rulings have sanctioned it. There is no doubt that there are several cases in which a person's right to refuse treatment or nutrition, for example during the terminal phase of cancer, should and would be respected. However, it has been argued that many of such cases suffer from undiagnosed but treatable depressive illnesses. Some support for this point of view was provided by a questionnaire study of more than 1000 residential and nursing home administrators in the United States.[4] Cognitive impairment, loss events, refusing medication, food, and drink, loneliness, feeling rejected by families are all risk factors for indirect suicidal behaviour in residential homes.[5] It is wise, therefore, that no one should be permitted to turn his or her face to the wall before assessment for the presence of a treatable depressive disorder.

Deliberate self-harm

Incidence

It is less possible to make a clear distinction between deliberate self-harm (DSH) and completed suicide in older than younger people. DSH at all ages has been quite extensively studied, but for obvious reasons mainly in hospital samples, and it is possible that several cases are undetected in the community. Broadly speaking the incidence curve for DSH is highest for the young and declines with age, whereas that for completed suicide rises with age. By the same token suicidal intent behind acts of DSH in older people is significantly greater than in younger adults.[6] In clinical practice it is therefore wise to consider deliberate self-harm in those over 75 as failed suicide.

Sex

As with completed suicide, rates for DSH differ quite widely from country to country. As with younger attempters, females outnumber males at a raw number ratio of approximately 3:2, but the *proportionate* gender ratio is approximately unity because fewer males survive into old age. Contrast this with completed suicide where men clearly outnumber women.

Methods

Deliberate drug overdose is the favoured method for DSH at all ages in Western countries; in some others, corrosive poisons or detergents are used. The most common types of drug for overdose are benzodiazepines, analgesics, and antidepressants. After drugs, self-cutting is the next most frequent method.

Psychiatric diagnosis

Older people are more likely to be assigned a psychiatric diagnosis after DSH, about half suffering from major depressive disorder, up to about a third from alcohol abuse, and under 10 per cent from other disorders.[7] Only about 10 per cent have no psychiatric diagnosis at all. Alcohol abuse together with depressive disorder augments the risk of DSH in older people. The status of cerebral organic disorder is uncertain because selection bias in reported case series reduces comparability. However, mild cognitive impairment and a co-morbid depressive disorder have been considered risk factors, and should be borne in mind by the clinician, if only on common-sense grounds. Personality factors have been implicated in DSH in older people, but research data are too poor and too few to make reliable statements on the subject.

Risk factors

Risk factors for deliberate self-harm in elderly people include: physical illness; widowhood and divorce or separation from a cohabitee; social isolation and loneliness (not the same thing); or simply living alone.[6,7] Unresolved grief, usually after death of a spouse, is a commonly found risk factor. The threat of transfer to a nursing home is, unsurprisingly, a precipitant of deliberate self-harm, although once an elderly patient is transferred to institutional care the risk of an overdose or some other attempt at suicide is reduced, probably because of lower access to the means and higher supervision. Surprisingly perhaps, terminal illness is not commonly found in older patients who attempt suicide but

fail, although hitherto undiagnosed but treatable physical disorders are sometimes revealed.

In keeping with the fact that more older suicide attempters are assigned a psychiatric diagnosis than younger ones is the fact that about 50 to 90 per cent, depending on the case series, undergo some form of psychiatric treatment as a result of the act of deliberate self-harm. Although fewer older people commit DSH than younger ones (about 5 per cent compared with 12 per cent) the risk of subsequent completed suicide is higher, compared with people of all ages (about 7 per cent compared with 3 per cent). Individual risk factors for later successful suicide include being male, having a prior psychiatric history, divorce, and current treatment for a persistent depressive illness.[6,8]

Completed suicide

Rates

The point has already been made in the opening paragraph of this chapter that suicide rates are still highest in the oldest old in most countries. Men outnumber women by about three or four to one in most countries; the exception being rural China.[1] However, suicide rates in the old have in fact been declining in many industrialized countries over the past 25 years, whilst those in young males have been rising—a reminder of the maxim that rates in all groups can and do vary over time and between countries, so that general conclusions about suicide should always take context into account. The reasons for incidence variations are discussed elsewhere, but socio-economic conditions and access to means play their part with the old as well as younger suicide victims.

Suicide at all ages is associated with divorce, widowhood, and single marital status. Widowers are more likely to kill themselves than widows, which has relevance for old age psychiatry, since in the overall population there are more old widowers than young ones.

Methods

Methods of suicide chosen by older people depend to a great extent on availability. In the United States, firearms are used by the majority of older men who kill themselves.[9] Shooting is also commonly chosen in Australia and Finland. In the United Kingdom, which has more stringent firearms control, drug overdose, especially in women and frequently with combination analgesics, hanging (especially in men), suffocation, or jumping from tall structures are preferred methods.[10] In Japan hanging, in Hong Kong jumping from one of the many very high buildings, and in Sri Lanka organophosphate poisoning are the commonest means in use.

Planning

Suicide in older persons is marked by careful planning and about half of the victims leave a note to indicate why or to confirm that they have killed themselves.[10] Suicide pacts are generally rare, but half of those that do occur involve people over 65. A previous history of a suicide attempt (DSH) is found in about a third of those older people who kill themselves.

Psychiatric diagnosis

Studies, including case-control, in the United States, Scandinavia, and the United Kingdom have found that 70 per cent of older suicide victims suffer from a mental illness, most commonly a major depressive disorder at the time they die.[11–14] Chronic symptoms of depression and a first depressive illness in later life are associated with a greater risk of suicide. Untreated or inadequately treated depressive illness is also found more commonly in elderly suicides. By contrast with younger suicide victims, alcohol and drug abuse rates are lower in elderly people, although co-morbid depression and alcohol abuse do occur more frequently than by chance. Schizophrenia or schizophrenia-like disorders are found less commonly in older than younger suicides. Similarly, cerebral organic impairment or dementia are infrequent and even absent from some series of cases. There has been more recent interest in the role of personality in suicide in older people. In various studies obsessional or anankastic traits which researchers have called 'low openness to experience' have been shown to predispose to suicide.[14,15]

Co-morbid physical illness

Co-morbid physical illness is, on common-sense grounds alone, likely to be a risk factor for suicide in older people and this has been confirmed in a number of studies.[16–18] Also, older people are much more likely to have visited their primary care doctor in the month prior to killing themselves than are younger suicide victims, and furthermore more likely to complain of physical than mental symptoms. Nevertheless, suicide to bring about the premature ending of a terminal illness or the avoidance of pain, although found to be a factor in studies of a variety of specific diseases, is not as common as one might imagine.

Social risk factors

Social risk factors for suicide in old age have been found to include: isolation and poor social integration; lack of a person to confide in; and concerns over dependence or a move from home to residential care.[13,18–20] Bereavement by itself is no more of a risk factor in older than the younger suicides, but a grief reaction prolonged for more than a year has been found to increase the risk.

Risk assessment

The study of suicide at any age is primarily for the purpose of prevention. In older people this means that, episodes of deliberate self-harm need to be considered as serious, even and perhaps especially when they do not appear to be so. A quantitatively small overdose of a relatively less lethal drug is frequently no indication of the seriousness of suicidal intent. Primary care doctors should be aware of the suicide risk in those attending with physical disorders, especially where the patient's complaints seem to be out of proportion to the actual evidence of disease. Nor is the identification of a physical illness a reason to relax vigilance over suicide risk. Dismissal of an older person's wish to die as 'rational' is probably wrong in the great majority of cases, but in any case should never be done before a thorough assessment of the mental state concentrating in particular on depressive disorder. Anxiety is frequently so prominently a presenting symptom of depressive disorder in elderly people, that other manifestations, such as suicidal thinking, may be overlooked. Whilst elderly people respond well to antidepressant medication, many live alone. Thus, a prescription for perhaps a month of treatment might be an enhancement of suicide risk. It is therefore prudent either to arrange close supervision or administration of medication by a carer rather than the patient

themselves, or for no more than a week's supply to be dispensed at a time. Pharmacists may be willing to assist in this by providing proprietary boxes with compartments for each dose.

Further information

Harwood, D. (2008). Suicide in older persons. In *The Oxford textbook of old age psychiatry* (eds. R. Jacoby, C. Oppenheimer, T. Dening, and A. Thomas) Chapter 29 iii. Oxford University Press, Oxford.

Hawton, K. and van Heeringen, K. (eds.) (2000). *The international handbook of suicide and attempted suicide*. Wiley and Sons, Chichester.

Pirkis, J. and Burgess, B. (1998). Suicide and recency of health care contacts: a systematic review. *The British Journal of Psychiatry*, **173**, 462–74.

References

1. http://www.who.int/mental_health/prevention/suicide/country_reports/en/index.html
2. Kirby, M., Bruce, I., Radic, A., *et al.* (1997). Hopelessness and suicidal feelings among the community dwelling elderly in Dublin. *Irish Journal of Psychological Medicine*, **14**, 124–7.
3. De Leo, D., Cerin, E., Spathonis, K., *et al.* (2005). Lifetime risk of suicide ideation and attempts in an Australian community: prevalence, suicidal process, and help-seeking behaviour. *Journal of Affective Disorders*, **86**, 215–24.
4. Osgood, N.J. and Brant, B.A. (1990). Suicidal behavior in long-term care facilities. *Suicide & Life-threatening Behavior*, **20**, 113–22.
5. Draper, B., Brodaty, H., Low, L.F., *et al.* (2002). Self-destructive behaviors in nursing home residents. *Journal of the American Geriatrics Society*, **50**, 354–8.
6. Hawton, K. and Harriss, L. (2006). Deliberate self-harm in people aged 60 years and over: characteristics and outcome of a 20-year cohort. *International Journal of Geriatric Psychiatry*, **21**, 572–81.
7. Draper, B. (1996). Attempted suicide in old age. *International Journal of Geriatric Psychiatry*, **11**, 577–87.
8. Hepple, J. and Quinton, C. (1997). One hundred cases of attempted suicide in the elderly. *The British Journal of Psychiatry*, **171**, 42–6.
9. Kaplan, M.S., Adamek, M.E., Geling, O., *et al.* (1997). Firearm suicide among older women in the US. *Social Science & Medicine*, **44**, 1427–30.
10. Harwood, D.M.J., Hawton, K., Hope, T., *et al.* (2000). Suicide in older people: mode of death, demographic factors, and medical contact before death. *International Journal of Geriatric Psychiatry*, **15**, 746–3.
11. Conwell, Y., Duberstein, P.R., Cox, C., *et al.* (1996). Relationships of age and axis I diagnoses in victims of completed suicide: a psychological autopsy study. *The American Journal of Psychiatry*, **153**, 1001–8.
12. Waern, M., Runeson, B.S., Allebeck, P., *et al.* (2002). Mental disorder in elderly suicides: a case-control study. *The American Journal of Psychiatry*, **159**, 450–5.
13. Beautrais, A.L. (2002). A case-control study of suicide and attempted suicide in older people. *Suicide & Life-threatening Behavior*, **32**, 1–9.
14. Harwood, D.M.J., Hawton, K., Hope, T., *et al.* (2001). Psychiatric disorder and personality factors associated with suicide in older people: a descriptive and case-control study. *International Journal of Geriatric Psychiatry*, **16**, 155–65.
15. Duberstein, P.R., Conwell, Y., and Caine, E.D. (1994). Age differences in the personality characteristics of suicide completers: preliminary findings from a psychological autopsy study. *Psychiatry*, **57**, 213–24.
16. Conwell, Y., Lyness, J.M., Duberstein, P., *et al.* (2000). Completed suicide among older patients in primary care practices: a controlled study. *Journal of the American Geriatrics Society*, **48**, 23–9.
17. Waern, M., Rubenowitz, E., Runeson, B., *et al.* (2002). Burden of illness and suicide in elderly people: a case-control study. *British Medical Journal*, **324**, 1355–7.
18. Harwood, D.M.J., Hawton, K., Hope, T., *et al.* (2006). Life problems and physical illness as risk factors for suicide in older people: a descriptive and case-control study. *Psychological Medicine*, **36**, 1265–74.
19. Turvey, C.L., Conwell, Y., Jones, M.P., *et al.* (2002). Risk factors for late-life suicide: a prospective, community-based study. *The American Journal of Geriatric Psychiatry*, **10**, 398–406.
20. Duberstein, P.R., Conwell, Y., Conner, K.R., *et al.* (2004). Poor social integration and suicide: fact or artefact? A case-control study. *Psychological Medicine*, **34**, 1331–7.

8.5.8 Sex in old age

John Kellett and Catherine Oppenheimer

Introduction

A Darwinian sees man as devoted to reproducing himself. The decline of fertility with age makes one question the biological purpose of sexuality in the senium. Is it simply the remains of a once useful behaviour, a vestigial characteristic? The fact remains that sexual interest and sexual activity, both as sources of enjoyment and as important components of pair bonding, continue among men and women even into extreme old age. However, as Alex Comfort memorably remarked, 'old people give up sex for the same reasons that they give up cycling—general infirmity, fear of looking ridiculous, no bicycle'.[1] Or, more soberly: the common obstacles to the continued enjoyment of sex in old age are illness, attitudes, and demography.

Surveys of sexuality in old age

A comprehensive discussion of surveys in this field can be found in Bouman.[2] These vary widely in their focus, setting, methods, and target age groups. The details are of great interest but only the main themes and a few illustrative examples can be described here.

Methodology

As with all surveys, one has to consider what factors influenced the selection (and self-selection) of the responders. In general, participants in these studies tend to be better educated and more liberal in their attitudes than their contemporaries; and important groups, such as people with chronic illness, may be under (or over) represented, depending on the setting of the survey.

Cross-sectional and longitudinal studies

Attempts to assess the effects of ageing on sexuality by surveying different age-groups at a single point in time are vulnerable to *cohort effects*—the sexual experience of those brought up before the Second World War is different from that of people whose childhood was in the 1950s, and need have nothing to do with age. Longitudinal studies are demanding and costly, but they reveal the effects of ageing more clearly, and they also allow individual patterns, stable over time, to be identified. For example, the series

of studies conducted at Duke University showed that although the *prevalence* of sexual interest and activity in both men and women decreases with age, *individual* patterns of sexuality tend to be stable until some event (such as illness or loss of a partner) disrupts the pattern.[3]

Sexual interest and sexual pleasure

There is more to sex than actual intercourse, especially in old age. A consistent finding is that in both sexes interest in sex is more resistant to ageing than is sexual activity, even when non-coital activity is included. For example, in a cross-sectional study of volunteer respondents aged 80 or over, living in a residential home in the United States,[4] 88 per cent of the men and 71 per cent of the women enjoyed daydreams and fantasies about sex, 82 per cent of the men and 61 per cent of the women engaged in touching and caressing, while 63 per cent of the men and 30 per cent of the women had sexual intercourse 'at least sometimes'. Janus and Janus[5] surveyed 2765 subjects in the United States by questionnaire, supplemented by 125 interviews. They found less reduction of activity in the older groups than had been found in earlier surveys, probably because they did not confine the question to coitus. In every age group men thought that they were more active than 3 years previously, but women, particularly those over 50, noted a decline. Unlike earlier studies foreplay was discussed, and the authors concluded that older men gain greater pleasure and experience more intimacy and warmth after coitus than younger men.

Obstacles to continued sexual activity

The commonest reasons given by people for a decline in their accustomed level of sexual activity were illness, and the loss of their partner.[3] Interestingly, women (but not men) also gave illness of their partner as a reason. This may reflect the traditional male role in initiating sexual activity, and also perhaps the fact that husbands tend to be older (therefore more at risk of illness) than their wives. For example, in a Swedish community study of 85-year olds,[6] participation in sexual intercourse was reported by 10 per cent of married women but only 1 per cent of the unmarried women, and by 22 per cent of married men compared to 13 per cent of the unmarried men. The rates for sexual interest (as opposed to activity) were higher: 46 and 37 per cent for married and unmarried men respectively; 24 and 15 per cent for married and unmarried women.

Attitudes to sexuality

Attitudes towards sexuality in old age—as revealed in responses to systematically varied vignettes—have become generally more positive over the last half-century, among both younger and older people.[2,7] Probably this is true also of professional attitudes, and where this matters most is among the staff caring for older people in institutional settings. The evidence suggests that care staff who are older, better educated and have had vocational training, and who have more experience of caring, are likely to be more open to the sexual needs of their residents. However as Bouman[2] points out, the same cannot be said of health policies and strategy. Government guidance on the care of older people has ignored their sexuality, and in most official surveys and policies on sexual issues, people aged 60 or more are excluded from consideration.

Demography

With increasing life-expectancy many marriages continue well into the 9th decade of one or both partners. Divorce has now overtaken death as the main cause for the ending of a marriage, though this does not necessarily mean increasing numbers of people left single in old age. Many older people experience second or third marriages (or cohabitations), and the age gap between the partners in these new relationships tends to be larger than that between partners still in their first marriage. The complex effects on family structure of these different social trends are analysed by Harper.[8] She shows that married older people have higher levels of health, social participation, and life satisfaction than those not married, and they live longer; while divorced men (compared to women, and to widowers) are the most disadvantaged in those respects. To this we can add (based on the survey data mentioned above) that in old age married people also enjoy greater opportunities than the unmarried for sexual expression.

Sexual orientation

Much less is known and written about the sexual lives of older people who are not heterosexual in their orientation. Despite the prevailing trend in Western societies towards valuing diversity, most older homosexual people in their earlier lives will have feared—or faced—stigma, discrimination, even the threat of criminal procedure, and may still face such discrimination. Further discussion of this important group of people can be found in Bouman.[2]

Sexuality and dementia

The effect of dementia on sexual interest and activity is unpredictable. Most often it is associated with a decline in interest, but sometimes (probably in less than 10 per cent of cases) a person with dementia may become more sexually demanding, or may lose the ability to judge when the expression of sexual interest is unwanted or out of place.[2,9] The effect of the patient's dementia on the spouse is also difficult to predict. In some couples the physical relationship continues as an important expression of their affection, support, and concern for each other. More commonly, sexual activity declines. For example, Wright[10] followed a group of couples in which one partner had dementia, alongside a control group of couples without dementia. Only 27 per cent of the afflicted couples continued sexual contact over the 5 years after diagnosis, compared to 82 per cent of the control couples.

Patients with dementia who have altered sexual behaviour, or who can no longer make reliable judgments about potentially sexual social situations, become vulnerable to misunderstanding, exploitation, or censure. In marriages the impact of these changes is often borne alone by the spouse, who also keeps them hidden from others. But people with dementia who live alone or who go into institutional care (even if only temporarily, into hospital perhaps) are not so protected, and clinical staff may be asked to intervene—to control behaviour that has offended or is harmful to others, or to reduce risk to a vulnerable patient. It may be crucial to establish whether such a patient is making an autonomous choice (for example, to engage in sexual contact with a fellow resident) or whether their failure of understanding is being exploited; and Lichtenberg and Strzepek[11] describe a helpfully structured approach to this question.

Normal sexual function in old age

The normal human sexual response cycle and the physiology of sexual intercourse are described in Chapter 4.11.1 in this textbook. These descriptions hold true for sexually active older people, although age does bring some relatively minor physical changes (see Table 8.5.8.1, based on data from Masters and Johnson[12])—changes experienced by some but not necessarily by all. The most noticeable of these are probably the much longer refractory period in older men (the interval which follows ejaculation before renewed erection is possible), and the oestrogen-dependent changes in vaginal tissue and lubrication in older women. However, according to the self-reports of older sexually active people, the capacity for sexual pleasure and the quality of orgasms are not at all affected by age.

Sexual dysfunction in old age

The range of sexual dysfunctions is discussed in detail in Chapter 4.11.2 in this textbook. In old age, as mentioned earlier, physical illnesses and their treatments assume increasing importance in curtailing the normal sexual activity and interest of an established couple. Likewise, with ageing, the balance between psychological and physical factors in the causation of a sexual problem, tips towards the physical. An example of this is erectile dysfunction, where a notable increase in research into the physiology of penile erection has led to a number of effective physical treatments, especially relevant to erectile problems associated with illnesses common in old age (such as diabetes and vascular disorders).[13,14] However, a clearer understanding of the physical components of a problem does not diminish the importance of psychological factors. Myocardial infarction, probably accompanied by hypertension and atherosclerosis, provides a good example. After the infarction the couple may misperceive the breathlessness of orgasm as cardiac distress, and they may need encouragement to resume sexual relations which may reduce the risk of further infarction.[15] Even a hospital admission interrupts the couple's sexual routine which may then be difficult to resume.

Some of the most common physical causes of sexual dysfunction are listed in Tables 8.5.8.2, 8.5.8.3, and 8.5.8.4.

Table 8.5.8.1 Physical changes of ageing

	Male	Female
Retained	Nil	Nipple erection Clitoral tumescence and retraction
Reduced	Penile and nipple erection Testicle elevation Power of ejaculation Rectal contractions	Vaginal lubrication and expansion Uterine elevation Bartholin gland secretion Orgasmic contractions
Lost	Flush Scrotal swelling Re-erection Ejaculatory inevitability Prostate contractions	Breast engorgement Flush Swelling of labia majora
Other	Refractory period >24 h	

Table 8.5.8.2 Medical factors affecting sexual function

Drugs reducing sexual drive	Drugs reducing testosterone	Drugs blocking physical arousal
Dopamine antagonists	Digoxin	Thiazide diuretics
Major tranquillizers	Cimetrdine	Some β-blockers
Metaclopamide	Cyproterone	
5-HT_2 agonists–SSRIs except nefazodone, trazodone, fluvoxamine	Finasteride	
Benzodiazepines	Oestrogens Progesterone	

Treatment of sexual problems in old age

The treatment of sexual dysfunction is founded on a comprehensive understanding of the problem. It begins with listening. Perhaps this is a statement of the obvious—except that the emotional power of sexuality makes it hard for patients and their partners to speak about sex, and for others to listen properly. Sexual histories are rarely taken as a routine part of the assessment of older people,[2] and even when patients disclose a sexual problem, their doctor may shrink from embarking on an exploration of the difficulty. Yet there is good evidence that opening up communication about a sexual problem—not only between clinician and patient but also (helped by the clinician) between sexual partners—forms a large part of successful treatment, and sometimes may be all that is needed. Even patients attending a specialized clinic may be satisfied by receiving assessment and information, without necessarily wanting active treatment.[13]

Sexual problems come to the notice of many different medical and surgical specialties (gynaecology, urology, genitourinary medicine, diabetology, endocrinology) who have expertise in the physical treatments now available, and sometimes also in the psychological and relationship components of the dysfunctions that they treat. In other cases the psychiatrist may be helping the couple to work with a physical treatment that one of them is receiving, and to make the most of the improvements in sexual function that it offers.

Behavioural treatment for sexual dysfunction, first described by Hunter,[16] was developed by Masters and Johnson.[17] Their *Sensate focus exercise* (see Chapter 4.11.2) is a simple behavioural technique that can be very effective in helping couples in whom sexual intercourse has become (for whatever reason) difficult, painful, or disappointing, and who have then retreated from all pleasurable physical contact with each other. In this technique attention is removed from intercourse (in fact intercourse is forbidden), and instead the focus is put on renewing the partners' pleasure in mutual touch and caressing for its own sake.

Table 8.5.8.3 Surgical procedures affecting sexual function

Transurethral prostatectomy leads to retrograde ejaculation
Pararectal surgery damages nervi erigentes
Indwelling catheters and pessaries
Mutilation affecting body image
Surgery to genitalia

Table 8.5.8.4 Diseases affecting sexual function

Diabetes
Myxoedema, pituitary tumours
Neurofibromatosis, paralysis, myotonia dystrophica, autonomic neuropathy
Peyronie's disease
Malignancy and infections of genitalia and prostate, vaginal fistulas
Liver failure leading to higher oestrogens
Arthritis
Hypertension, vascular diseases, myocardial infarction
Respiratory distress
Depression, schizophrenia, cortical dementias

There is little scope for pharmacological treatment of most sexual difficulties arising in old age, other than erectile dysfunction.[13,14] Hormonal treatments are no longer thought to be helpful except where deficiencies have been clearly demonstrated. The commonest hormonal treatment in old age is probably the use of oestrogen (topically) in dyspareunia. Treatment of hypersexuality and sexual aggression in the context of dementia is difficult: pharmacological methods have been reviewed by Series and Degano.[18]

Conclusion

Ageing brings increasing diversity. This certainly applies to sexual behaviour. Those who work in the field of old age psychiatry can help their patients by understanding this diversity, making it safe and acceptable for patients to talk about whatever sexual concerns they have, and helping them in the acquisition of all the information they need. Sex in old age is not the frightening imperative of the teenager, but it can still contribute greatly to the quality of life.

Further information

Bouman, W.P. (2008). Sexuality in later life. In *The Oxford textbook of old age psychiatry* (eds. R. Jacoby, C. Oppenheimer, T. Dening, and A. Thomas) (chapter 37). Oxford University Press, Oxford.

Gott, M. (2005). *Sexuality, sexual health and ageing*. Open University Press, Maidenhead, Berkshire.

Greengross, W. and Greengross, S. (1989). *Living, loving and ageing. Sexual and personal relationships in later life*. Age Concern England, Mitcham, Surrey.

Leiblum, S.R. and Rosen, R.C. (eds.) (2000). *Principles and practice of sex therapy*. The Guildford Press, New York.

Warner, J. (2000). Sexuality in dementia. In *Dementia* (2nd edn) (eds. J. O'Brien, D. Ames, and A. Burns), pp. 267–71. Arnold, London.

References

1. Comfort, A. (1980). *Practice of geriatric psychiatry*. Elsevier, New York.
2. Bouman, W.P. (2008). Sexuality in later life. In *The Oxford textbook of old age psychiatry* (eds. R. Jacoby, C. Oppenheimer, T. Dening, and A. Thomas) (chapter 37). Oxford University Press, Oxford.
3. George, L.K. and Weiler, S.J. (1981). Sexuality in middle and later life; the effects of age, cohort and gender. *Archives of General Psychiatry*, **38**, 919–23.
4. Bretschneider, J.G. and McCoy, N.L. (1988). Sexual interest and behavior in healthy 80- to 102-year olds. *Archives of Sexual Behavior*, **17**, 109–29.
5. Janus, S.S. and Janus, C.L. (1993). *The Janus report on sexual behavior*. Wiley, New York.
6. Skoog, J. (1996). Sex and Swedish 85-year olds. *The New England Journal of Medicine*, **334**, 1140–1.
7. Gott, M. and Hinchliff, S. (2003). How important is sex in later life? The views of older people. *Social Science & Medicine*, **56**, 1617–28.
8. Harper, S. (2008). Sociological approaches to age and ageing. In *The Oxford textbook of old age psychiatry* (eds. R. Jacoby, C. Oppenheimer, T. Dening, and A. Thomas) (chapter 2). Oxford University Press, Oxford.
9. Haddad, P.M. and Benbow, S.M. (1993). Sexual problems associated with dementia: part 1. Problems and their consequences. *International Journal of Geriatric Psychiatry*, **8**, 547–51.
10. Wright, L.K. (1998). Affection and sexuality in the presence of Alzheimer's disease: a longitudinal study. *Sexuality and Disability*, **16**, 167–79.
11. Lichtenberg, P.A. and Strzepek, D.M. (1990). Assessment of institutionalized dementia patients' competencies to participate in intimate relationships. *The Gerontologist*, **30**, 117–20.
12. Masters, W. and Johnson, V. (1966). *Human sexual response*. Churchill, London.
13. Melman, A. and Christ, G.J. (2002). The hemodynamics of erection and the pharmacotherapies of erectile dysfunction. *Heart Disease*, **4**, 252–64.
14. Trost, L., Kendirci, M., and Hellstrom, W.J.G. (2006). Erectile dysfunction in the aging man: current options for treatment. *Aging Health*, **2**, 71–86.
15. Kellett, J. (1987). Treatment of sexual disorder: a prophylaxis for major pathology? *Journal of the Royal College of Physicians*, **21**, 58–60.
16. Hunter, J. (1782). *Venereal disease*. London.
17. Masters, W.H. and Johnson, V.E. (1970). *Human sexual inadequacy*. Little Brown, Boston.
18. Series, H. and Degano, P. (2005). Hypersexuality in dementia. *Advances in Psychiatric Treatment*, **11**, 424–31.

8.6

Special features of psychiatric treatment for the elderly

Catherine Oppenheimer

Introduction

Three themes underlie the topics in this chapter.

Old age is a time of multiple problems

Physical, psychological, and social problems often occur together, linked by chance or causality in the life of the old person. Very rarely can one problem be dealt with in isolation, and many different sources of expertise may be engaged with a single individual. Therefore good coordination between different agents is essential in old age psychiatry, both for the individual patient and in the overall planning of services.

Clear boundaries between 'normality' and 'disease' are rare in old age

Many of the pathologies characteristic of old age are gradual in onset and degenerative in nature, and more due to failures in processes of repair than to an 'external foe', so the distinction between disease and health is often quantitative rather than qualitative. 'Normality' becomes a social construct with fluid borderlines, containing the overlapping (but not identical) concepts of 'statistically common' and 'functionally intact'. Thus the popular perception of normal old age includes the 'statistically common' facts of dependence and failing function, whereas 'intactness' (excellent health and vigorous social participation) is seen as remarkable rather than the norm. But the boundaries of 'old age' are also socially constructed—in developed countries good health at the age of 65 would nowadays be regarded as a normal middle-aged experience, whereas superb health at 95 would still be something noteworthy.

Since some degree of physical dependence, forgetfulness, and vulnerability to social exclusion is expected in old age, meeting those needs is also regarded as a 'normal' demand on families and community agencies such as social services, rather than the responsibility of health care providers. As the severity of the needs increases, however, so also does the perceived role of health professionals, both as direct service providers and in support of other agencies.

Lack of competence is common in old age

Because of the high prevalence of cognitive impairment in old age (especially among the 'older old'), questions frequently arise as to the competence of patients to make decisions. Older people who cannot manage decisions alone may come to depend increasingly on others for help; or, resisting dependence, they become vulnerable through neglect of themselves or through the injudicious decisions they make. When an incompetent person is cared for by a spouse or family member, the danger of self-neglect or of ill-considered decisions is lessened, but instead, there are the risks of faulty decisions by the caregiver (whether through ignorance or malice), and also risks to the health of the caregiver from the burden of dependence by the incompetent person. Legal mechanisms, differing from one country to another, exist to safeguard the interests of incompetent people.

These three themes will be developed further, and with them the following special topics:

1 multiple problems: including sleep disorders in old age, medication in old age psychiatry, and psychological treatments in old age psychiatry;

2 blurred boundaries of normality: including the role of specialist services and support between agencies;

3 incapacity and dependence: including balancing the needs of patients and caregivers, abuse of older people, ethical issues, and medico-legal arrangements for safeguarding decisions.

Multiple problems

Sleep disorders in old age

Useful reviews of this topic may be found in Anconi-Israel and Ayalon,[1] Sivertsen and Nordhus,[2] and Mosimann and Boeve.[3] More detailed general discussion of sleep and its disorders will be found in Chapters 4.14.1–4.14.4 of this textbook.

(a) Normal changes with age

With age, the architecture of sleep changes—in fact, most of the change occurs before the age of 60. Sleep is divided into shorter periods interspersed with wakefulness or brief arousals, there is a decrease in total sleep time and in sleep efficiency (the ratio of time asleep to time in bed), and there is less stage 4 (deep) and more stage 1 and 2 (shallow) sleep, without an increase in the proportion of rapid-eye-movement (REM) sleep. This change in sleep architecture is conventionally associated with changes in circadian rhythms with age, such as decreased amplitude and phase length of these rhythms (but see Monk[4] for a critical review).

Many people adapt to these changes, but others find the altered pattern distressing. Thus the borderline between normal and problematic sleep is blurred, because subjective assessments of sleep quality are not necessarily matched by objective measures (such as polysomnography); consequently the definition of 'insomnia' hinges not only on features of night-time sleep, but on impaired functioning in the daytime.

(b) Comorbidity

The majority of healthy older people have no complaints about their sleep, but there is a strong association between poor sleep and other health problems, and sleep problems make a material contribution to the impaired quality of life suffered by people with comorbid illness. In a study of patients in primary care,[5] a positive answer to even one question about sleep ('do you feel excessively sleepy during the day?') predicted the quality of life related to physical or mental health problems. Attention to improving sleep in these patients can improve their well-being, but too often the sleep problem is missed in the general assessment of the patient. Impaired sleep can have serious consequences: it is associated with symptoms of anxiety and depression, an increased risk of falls, and diminished memory and cognitive functioning.[1]

(c) Causes of disordered sleep

These include the following:

1 **Environmental causes**: e.g. a strange bed, noise, cold or heat, or loss of a familiar bed companion (e.g. through bereavement).

2 **Physical causes**: sleep can be broken by pain, stiffness (e.g. Parkinson's disease or arthritis), limb movement (restless legs syndrome, periodic movements of sleep), breathlessness (cardiac failure or sleep apnoea), the need to urinate (prostatic disease or urinary tract infection), eating too close to bedtime, or dehydration (e.g. voluntary restriction of fluids to prevent nocturia).

3 **Medication**: alcohol, especially if taken to relieve anxiety or to assist sleep (since the rapid metabolism of alcohol leads to rebound anxiety and wakefulness), and antidepressants such as selective serotonin reuptake inhibitors (**SSRIs**) can cause wakefulness or nightmares. Information on the numerous other medications which may impair sleep can be found in Anconi-Israel and Ayalon.[1]

4 **Psychological causes**: for example, anxiety, depression, hypomania, and paranoid illness. Often a sleep problem is triggered initially by a physical cause, but is then maintained by the patient's anxiety about wakefulness.

5 **Sleep in dementia**: changes in sleep rhythm in dementia are similar to those of normal old age, but often more severe: daytime drowsiness or napping, difficulty in falling asleep at night, decreases in slow wave sleep and in rapid-eye-movement sleep. However, another common cause of sleep impairment is the use of benzodiazepines or major tranquillizers to treat behaviour disturbances: the patient may end up drugged in the day and wakeful at night. A partial remedy may be to create an overriding diurnal rhythm (e.g. by attendance at day care or a programme of physical activity in the day), together with the minimal use of medication at night. Patients in institutional care are particularly likely to be deprived of the normal cues for circadian rhythms: quiet and darkness at night, bright daylight in the morning, and physical activity in the day. On the other hand sleep time may be strikingly increased in dementia, especially in vascular dementia where it may be part of the apathy that is common in that disease.

(d) The parasomnias

The parasomnias that are common in old age are obstructive sleep apnoea (or sleep-disordered breathing); restless legs syndrome and periodic limb movements in sleep; and REM sleep behaviour disorder (RBD). They may range in severity from troublesome to severely disabling, and accurate recognition is important for all of them, because of the consequences both to the patients and to their bed partners if their diagnosis and treatment are missed. Further details can be found in the sources mentioned above,[1,3] but RBD warrants some further discussion here. This disorder is defined as an 'intermittent loss of the muscle atonia normally present during REM sleep, and episodes of elaborate motor activity associated with dream mentation'.[6] Typically, in the early hours the patient (usually male) shouts, thrashes around, and may attack his bed partner, without waking, and without any recollection of the episode when he does wake later. The importance of this condition lies in the fact that it is very distressing, and possibly dangerous, for the bed partner; it can often be treated effectively (with clonazepam); and it is strongly associated with the development (sometimes after a very long latent period, of a decade or more) of neurodegenerative disease—especially Lewy body dementia, Parkinson's disease, or multisystem atrophy—the alpha-synucleopathies.[6,7]

(e) Management of sleep disorders

(i) Psychological methods

These are now the methods of choice with insomnia in older adults.[2,8] Trials comparing psychological with pharmacological treatment, and with a combination of the two, show equivalent effects in the short-term, but a longer-term advantage to the psychological methods. More importantly, it has also been shown that in secondary insomnia ('insomnia occurring when a psychiatric condition, a medical condition, a non-insomnia sleep disorder, or a medication appears to precipitate and then appears to maintain insomnia') it is not necessary to wait until the primary condition has been resolved: the insomnia can usually be effectively treated in its own right, and even if it is not completely cured, valuable improvements can be achieved.[8] Likewise, psychological methods can be used to help in withdrawing medication in hypnotic dependent insomnia.[9]

The methods used have generally been 'multicomponent behavioural treatments'. The components include:

- Relaxation training
- Stimulus control (requiring the patient to leave the bedroom if they are not sleeping)
- Sleep restriction and sleep compression (using a fixed time for getting up, progressively shortening the time in bed until it matches the time asleep)
- Cognitive restructuring (modifying the pre-sleep thinking patterns, which in insomnia usually include negative thoughts about the effects of sleep loss, and the use of worry and self-blame as an attempted strategy for controlling thoughts[10])
- Sleep hygiene education (advice on the effects of tea, coffee, exercise, etc on sleep)

Such treatment packages need not be dependent on sources of specialist expertise: Sivertsen and Nordhus[2] discuss the feasibility of treating insomnia psychologically within primary care.

(ii) Treatment of sleep problems in dementia

When cognitive impairment makes psychological treatment difficult, pharmacological treatment may be necessary. Benzodiazepines (probably temazepam for preference) must be used very cautiously; they are dangerous in ambulant patients, though less so for a patient who is no longer mobile. Sedating antidepressants (e.g. trazodone) can be used instead. An atypical antipsychotic may be appropriate if there is severe anxiety and suspiciousness (sometimes of delusional intensity) of the carer at night.

The sleep problems of the carer of a patient with dementia also need to be taken very seriously. The carer may benefit from the psychological measures outlined above, and can institute some of the measures (such as sleep hygiene) on behalf of the patient. Insomnia in a caregiver, caused by the wakefulness of the person cared for, can lead to rapid breakdown of the support system. If the patient's sleep problem cannot be resolved then it is essential to give the carer the opportunity for uninterrupted sleep at times, through arranging residential respite care, a night-sitter, or some other form of relief.

The use of medication in old age psychiatry

Specific uses of medication for the various psychiatric disorders occurring in old age are dealt with in the relevant chapters. Here, three general principles will be discussed:

- medication as an experimental trial
- stopping medication
- compliance and concordance.

(a) Medication as an experimental trial

Starting medication for any condition ought to be treated as the test of a hypothesis. There should be a plan, shared with the patient, setting out the following:

1 how long the trial will last before a decision is made that the treatment is unsuccessful and should be ended;

2 which target symptoms will be monitored, and what records should be kept (by the patient or caregiver);

3 when progress will be reviewed;

4 what side-effects might be developing and which of these should alert the patient to stopping the drug and contacting the doctor;

5 what will follow if the trial succeeds;

6 what will be tried instead if the trial fails.

This watchful approach to medication is particularly important for people with dementia. Often the patient's confusion, lack of insight, and communication difficulties mean that a tentative diagnosis has to be based on scanty information. For example, disturbed behaviour in a patient in a nursing home may be due to depression, but the patient cannot describe the depressive symptoms. Clues to the diagnosis may come from the care staff (e.g. 'She never jokes with us now'), and an empirical trial of an antidepressant may result in a resolution of the symptoms. However, the trial must be set up carefully and medication not thoughtlessly continued for months without review, merely because no one has asked whether it is helpful or not.

(b) Stopping medication

The decision to withdraw medication can be as valuable and informative as the decision to start it, and documentation of the reasons for the decision is equally important (though often neglected). In delirium, medication may be contributing to the problem, particularly where a cocktail of medications has been built up over time—improvement after withdrawal of a drug gives valuable guidance for a future episode. Also, patients reaching the terminal stages of dementia should have their medication gradually decreased, to test whether it is still needed: drugs prescribed at an earlier stage to control behavioural syndromes are rarely, if ever, required for the entire course of the disease.

However, there are also reasons for being cautious about stopping medication. Some drugs have important withdrawal effects (e.g. paroxetine and benzodiazepines). Medication that has been used prophylactically may seem to be unnecessary while the patient is well, but when the drug is withdrawn the need for it is revealed. For example, a patient can remain well for years on an antidepressant following a severe depressive episode, or on a low dose of antipsychotic after a schizophrenic illness—until well-intentioned withdrawal of the drug, however carefully monitored, precipitates the return of the illness.

(c) Compliance and concordance

The word 'compliance' appears to imply that patients must be obedient in following instructions; but nowadays patients are seen more as partners in their own treatment, for which the word 'concordance' (when the partnership is successful) is more apt. The same principle of partnership in treatment decisions applies to older people, wherever possible. However, in addition to willingness to participate there needs to be the ability to do so, which in older people is often impaired by physical causes (such as arthritic hands which cannot open child-proof packaging, or poor vision which misreads instructions), and by psychological causes, especially memory loss and temporal disorientation. Therefore patients will often need help in maintaining their concordance with treatment, through suitable packaging and memory aids (such as calendar boxes), or through supervision or administration of the tablets by others. Families may take a long time to notice that their parent or grandparent, who has taken medication reliably for years, has started to miss or duplicate doses.

It is important to understand the feelings of a caregiver who takes responsibility for the medication prescribed for a patient with dementia, especially where it is being used to control behaviour. Anything prescribed on an 'as-needed' basis should be very carefully explained. The caregiver may be afraid of provoking the difficult behaviour by offering the medication, or conversely of overdosing her relative into stupor, or she may feel guilty about meeting her *own* needs by giving drugs to another. Such feelings make it hard for a caregiver to judge objectively when discretionary medication should be used.

Psychological treatments in old age

Many older people prefer a 'talking treatment' to medication, and there is increasing evidence of the success of psychological interventions with cognitively intact older people.[11] Brief focused interventions are particularly suitable.

However, even mild cognitive impairment, too slight to amount to dementia but enough to interfere with grasp and retention, may hamper psychological treatment. For example, such a patient being treated psychologically for anxiety may try hard to cooperate with her therapist, only to be made more anxious as she fails to remember the instructions, or to put them into practice, and so the sense of failure she feels elsewhere in her life is reinforced.

The principles of supportive psychotherapy and problem-solving are always relevant, both on their own or in conjunction with medication: providing 'unconditional positive regard' for the patient; accepting a degree of dependence by the patient, while limiting its consequences; setting appropriate expectations so that the patient is not unnecessarily exposed to a sense of failure; openly facing realities that cannot be altered (such as loss, disability, and death); helping to think through practical problems as they arise.

Family and systems therapy[12] has a particular role in old age psychiatry, because it explicitly recognizes the patient in his or her social context. Although family therapy requires special training, the principles of a family approach can be adopted by all those working with older people.

When patients are cognitively impaired, psychological approaches generally focus more on the interaction between the patient and the caregivers, who can be helped to understand the patient and the reasons for her behaviour, as well as their own actions and the reciprocal effects that these have on her. For example, reminiscence therapy helps professionals to respond to patients as 'whole people' with individual lives, as well as helping patients to reconnect with their former identities, and to recover some of the confidence they took for granted in the past.

Therapies based on music, dance, drama, art, and sensory stimulation are also important in old age.[13] They can help patients, individually or in groups, to express feelings and thoughts, bypassing the impairments of verbal skills which come with dementia.

The blurred boundaries of 'normality'

We consider two examples of the way in which 'normality' is not clearly defined in old age.

First, social expectations allow for increasing dependency in the old, and many people have personalities better suited for protective relationships than for solitude and independence. People who have struggled with anxiety and loneliness in middle age can experience the onset of physical dependence as a kind of relief: now they can allow themselves to be looked after. What was a dysfunctional personality structure in earlier life becomes adaptive in old age. In contrast, the person who has always jealously guarded his or her personal boundaries will find the adjustment to disability very hard. Determined refusal to accept necessary help converts a 'normal' preference for independence into a problem for others.

The second example concerns the idea of death. For a younger person to say that they were waiting for death would prompt a psychiatrist to look for evidence of a depressive illness. In old age this way of thinking may be found in someone with no disorder of mood, with the ability to enjoy life, and with a rational appreciation of life's inevitable limits. But even in old age, to see nothing valuable in one's future might be a sign of depression.

The role of specialist services and cooperation between agencies

This topic is dealt with more fully in the succeeding chapter (see Chapter 8.7).

The difficulty in separating the normal from the pathological, and in deciding when specialist intervention is required, is reflected at a structural level in shaping the responsibilities of the different bodies involved in the welfare of older people. Help required by older people can range from the simplest of neighbourly assistance to the full resources of specialist hospital care. As people live longer, despite some morbidity, the demand for care increases. The response to this demand is shaped by many different factors: market forces, public concern, private enterprise, voluntary or charitable groups which spring up to fill perceived gaps, and governments acting to remedy abuse or to limit demand on state-funded services. The divisions of responsibility are arbitrary, and people fall between services as often as within them. Systems change. For example, in the United Kingdom older people with permanent disability (especially dementia) used to be cared for in state-funded hospitals, but this care has now become the responsibility of independent-sector nursing homes, with means-tested financial support from the social security system.

The scope of specialist services for older people extends beyond the treatment of established illness. They have a role in the prevention and early detection of illness, in supporting and educating other service providers and in collaborating with other agencies in strategic planning. This span of interest can be represented as a pyramid (Fig. 8.6.1). The peak of the pyramid represents those requiring the scarcest and most intensive specialist services, and the base represents the population at large. In the middle are the people who need day care, support at home, social support, and help in primary care.

Between a population base of, say 20 000 older people and an inpatient provision of, say 10 beds lies the range of people to whom psychiatric services can offer significant help without assuming total responsibility for care. It is essential that the different agencies interact effectively and overlap their care, rather than leaving patients to negotiate their way from one island of provision to another.

Key components of such a 'mixed economy of care' should include:

1 A system for prompt and accurate assessment of people's needs which leads to offers of appropriate help, including reassessment. Assessment for specialist psychiatric care should be both 'generic' (using the skills that all disciplines share) and specific (using the particular skills of medicine, nursing, occupational therapy, etc.).

2 A clear understanding by staff working in the various agencies of what their tasks will be, and good training that will allow them to carry these out with confidence and satisfaction. When non-specialist care breaks down, it is usually because the carer do not expect to have to deal with the problem which faces them.

3 Systems ensuring that people are not obliged to make frequent and abrupt transfers between one setting and another as their disabilities increase. Assessment must therefore include prediction of future need; and each provider should offer a degree of

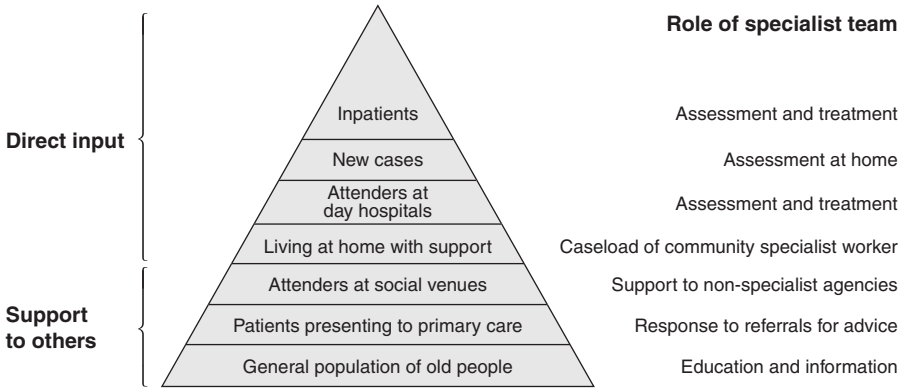


Fig. 8.6.1 The span of old age psychiatry services.

flexibility, to encompass people who will soon need what they provide, and others whose needs are greater than the norm. For some people, a change of setting as their condition progresses—for example, from a sheltered flat to care in a nursing home—will be inevitable, so care systems must have ways of making transitions as smooth as possible.

Loss of capacity and dependence on help

Mutual helpfulness is an accepted part of marital and family relationships. Family members often adapt unquestioningly to an older person's increasing dependence until the emotional pressures, or the adjustments they have to make in their day to day lives, become severe.

The 'needs of caregivers' are part of the currency of discussion amongst service providers. But people in everyday life do not necessarily perceive themselves that way: they think of themselves as simply participating in a normal aspect of family life. Younger people who have duties to their work, partner, and children—duties which conflict with the needs of their parents—may be readier to seek professional help than are the ageing spouses of a failing partner. Spouses often view caring as an intrinsic part of their lifelong relationship, and may resent offers from outside as an intrusion upon their privacy.

On the other hand, relationships can have malign as well as protective aspects, and sometimes the dependent partner in a caring relationship suffers more disadvantages than benefits. In extreme cases, abuse can occur.

Abuse of older people

This is much better understood now than a few decades ago, although systematic study is difficult because of the varieties of abuse or exploitation that arise, the ambivalent relationships that surround them, and the concealment often practised by abusers and victims alike.[14] Abuse can occur in any circumstances, any class, and any relationship, from blood ties and friendships to professional and commercial relations. Nevertheless pointers to risk have been identified (Table 8.6.1).

Typically, the victim is disabled, often but not always with a dementing illness, perhaps with impaired communication, and is unrewarding to look after. Typically, the abuser is also impaired in some way, isolated from support (often by refusal rather than the absence of offers of help), and may abuse alcohol. There may have been a long history of ambivalence or of mutual aggression between abuser and victim, or the victim's illness may have reversed a power relationship that previously operated (damagingly) in the opposite direction. The intention of the abuser is sometimes clearly to do harm. In other cases the abusive behaviour seems to be an impulsive response to emotional pressures which the abusers are poorly equipped to deal with because they do not understand them properly, do not know how to share the burden with others, or are overwhelmed by the patient's need for care. The abuse may range from a single assault by a caregiver, who is immediately horrified by what he or she has done, to systematic cold-blooded persecution. Abuse may be physical, emotional, psychological, sexual, or financial.[15]

Where harm to a vulnerable person is suspected, there is an obligation to report it to the appropriate authority—in the United Kingdom, the local social services. Sometimes they in turn will involve the police and a criminal prosecution may follow. But frequently the problem can be dealt with in other ways, such as arranging a different form of care for the vulnerable person. The emphasis will be on reducing and managing risk, and on enhancing

Table 8.6.1 Some risk factors for abuse in old age

Victim	Abuser	Relationship
Female	Family member in caregiver role	Previous relationship – not close, ambivalent, mutually abusive
Aged over 75	Psychiatric history	Role reversal
Physically dependent	History of abuse as a child	Power reversal
Cognitively impaired	Substance misuse	Lack of problem-solving skills in the relationship
Socially isolated (lacking external support)	Financial dependence on victim	Forced proximity
Sensory impairment	Dependence for housing on victim	Mutual dependence
Incontinence	Overburdened by caregiver role	
Abusive or unrewarding to caregiver	Unsupported, or rejecting support	
Ready to adopt sick role		

Reproduced from Fisk, J. Abuse of the elderly. In *Psychiatry in the elderly* (2nd edn.) (eds. R. Jacoby and C. Oppenheimer), copyright 1997, with permission from Oxford University Press.

the quality of life of the person concerned. Difficult ethical issues arise when an elderly couple both suffer from dementia and one of them (often the husband) insists on continuing to care alone for his wife, unaware of the loss of his competence to do so. For a time, professional staff will try to support both partners in their wish to remain together at home, and to mitigate the effects of the inadequate care by the husband. But at some point the professional duty may need to shift from supporting, to taking over legal responsibility for the neglected person—though the displacement of the caregiving spouse in this way is distressing for everyone concerned.

Ethical problems in old age psychiatry

The same classical principles of bioethics (beneficence, non-maleficence, autonomy, and confidentiality)[16] apply in old age psychiatry as in other age groups, but with differences in emphasis.[17] Since older people can rarely be considered in isolation, ethical principles have to be applied with the whole system in mind, and professionals often have ethical obligations to more than one person at a time.

(a) Values

In old age, as in youth, we should seek to produce benefit and not harm—but identifying which of these is which may be more difficult in old age. Death is usually thought of as harm, but sometimes may be regarded as a blessing. Therefore it is essential to understand what the patient, rather than the professional, sees as beneficent, and the patient's right to name the value that they set on something must be respected. Some older people would prefer their assets to go to their descendants rather than being spent on their own care at the end of their lives—but the state sets limits to such self-denial.

(b) Ambivalence

Impairments of thinking and communication can complicate ethically observant professional practice. People of all ages may have ambivalent feelings, or say things which seem belied by their actions, but this is particularly common among older people with cognitive impairment. Perhaps memory failure causes decisions to be swayed by feelings of the moment, so that decisions are not consistently maintained; and cognitive difficulty in marshalling complex information means that weighing alternative possible outcomes to a decision is much harder.

An illustration may make this clearer.

Case study: A woman with early cognitive impairment lives alone in the house where she brought up her children. She loves her house and garden, and tells her daughters that she never wants to live anywhere else, and she is convinced that she needs no help from them. At night, however, she becomes anxious and confused, and telephones her daughters, asking them why they 'haven't come home yet' and begging them not to abandon her.

Does the night or the day reveal her 'true wishes' better? Which should guide the decisions of her family, and of the professionals whose help they seek? Possibly, her considered daytime thoughts and her anxious actions at night are tapping different areas of her experience—the daytime communication reflects her aspirations and her lifetime self-perception, while her telephoning at night reflects the immediacy of her feelings and needs. Our duty is to give weight to both kinds of communication, helping the patient herself to understand what they mean, and to offer her the real-life opportunity (rather than theoretical discussion) of testing out the options she needs to consider.

(c) Giving information and safeguarding information

Ethical obligations here include truth telling, giving information to patients about themselves, and protecting patients' information from others.

'Loss of insight' is a feature of dementia, but good insight is also dependent on sound information. A person who is making significant errors, or beginning to fail in self-care, should not be left in ignorance of what is happening to them. In fact, public knowledge nowadays about Alzheimer's disease is so much greater that sufferers often recognize the early signs themselves. They are entitled to an open discussion with their doctors, in which full information is put before them. (Relatives often shy away from such open disclosure, although when asked in surveys what they would wish for themselves, they tend to say that they would rather be told).

When it comes to disclosing information about a patient to the people involved in his or her care, the arguments are different. It is generally accepted that better care must depend on the best information, and it is normal for information to be shared between members of a clinical team, where the members share also in the duty of confidentiality to their patient. But in old age it is harder to know where to set these boundaries—both as regards information that should be protected, and as regards recognition of who is a team member. For example, senior staff in some residential homes, strictly preserving confidentiality, may not share information about the residents with the untrained care staff. However, if they know nothing about a resident's former life, caregivers will tend to respond to her as 'a bundle of needs' rather than as a real person; and if she cannot tell her own story, others must do it for her. This is much better recognized now than it used to be, and homes may ask families to construct a diary or album of a resident's life, with photographs, mementos, and recollections by different relatives. The diary also acts as a memory prompt and trigger for enjoyable conversations between resident and caregiver, and it creates a domain of shared information about individuals within the institution.

(d) Autonomy

Healthy people strongly value the freedom to make their own decisions, to pursue their own aims, and to determine the course of their own lives. The onset of a physical illness may constrain this freedom, but people should still have as much influence as possible over decisions about their illness and its consequences.

Psychiatric disorder is different, because it may affect the powers by which that freedom is exercised, and may lead to decisions which would never have been made in health. From this comes the need, universally recognized although taking different legal forms in different countries, to set external controls over the decisions of people when they are mentally ill. Such legal controls are typically based on acute functional illness as the paradigm case. This gives them an 'all or nothing' character, envisaging hospital treatment of an illness capable of being relieved, so allowing patients to resume their autonomy when they recover their health.

This legal model is not really suited to cognitive impairment and its effect on decision-making in old age. The illness will not get better,

and patients are unlikely to give (as they might with an acute illness) later 'retrospective informed consent' to the treatment. The emphasis therefore has to be much more on **minimal necessary interference**—on setting up protective frameworks, in which as much autonomy as possible can be exercised; on supporting patients at home or in homely settings, rather than bringing them into a hospital where no effective treatment can be given; and on gathering the information (both from patients and from those who know them best) which will enable professional decisions to reflect the wishes that the patient would have expressed, had they been able to do so.

Looking after cognitively impaired patients requires us to try constantly to maximize the opportunity for autonomous decisions, while also being very clear when a patient lacks the capacity to engage with a more complex issue. At such times, the responsibility for the decision must be openly and seriously taken by others. Trying to circumvent the problem by concealment and persuasion is a greater affront to autonomy than is an honest explanation to the patient of the reasons why a decision has been taken out of his hands.

Medico-legal issues

'Doctors and lawyers have common responsibilities to ensure the protection of people who are incapable of deciding matters for themselves, and to promote the choice of those who can and should regulate their own lives. The careful assessment of whether individuals have or lack capacity is essential to protect their rights'.[18]

The legal framework in England and Wales for both financial and welfare decision-making where capacity is in doubt has been transformed by the Mental Capacity Act 2005, which came into force in April 2007. Some general principles of the Act are worth discussing here: more detail can be found in Lush.[19]

'Capacity' is a legal rather than a medical concept (see Chapter 11.1). Every adult is presumed to have full capacity, and a loss of capacity must be proved in relation to the particular decision being made: capacity is 'decision-specific'. For example, a person may *have* the capacity to choose another person to act for her in the management of her affairs, while *lacking* the capacity to manage those affairs herself. Although the final decision on capacity is made by the courts, they look to doctors to advise on whether a mental disorder has affected the individual's ability to make a particular decision or to carry out a specific task.

(a) Assessing capacity

The Act states that a person lacks capacity if ' . . . he is unable to make a decision for himself . . . because of an impairment of, or disturbance in, the functioning of the mind or brain'. Such a person lacks capacity in relation to a particular decision if he cannot:

- understand the information about the decision
- retain the information
- use or weigh the information in the process of making the decision
- communicate the decision

The information that he must understand should include the consequences of deciding one way or another, and of making no decision. The information must be retained only for long enough to allow a decision to be made, therefore memory loss does not automatically remove capacity. The focus of this test of capacity is on the *process* of decision-making: the Act explicitly states that 'a person is not to be treated as unable to make a decision merely because he makes an unwise decision'.

(b) Acting on behalf of a person who lacks capacity

Anyone acting on behalf of an incapacitated person must do so in the 'best interests' of that person. The action or decision in question should be delayed if there is a chance that the person may regain capacity; every effort should be made to enable them to participate in the decision-making process; the decision-maker should take into account the person's past and present wishes, beliefs, values, and feelings, especially any that had been written down when they had the capacity; the views of families, carers and anyone else interested in their welfare should be sought; and the least restrictive method for achieving the intended purpose should be chosen. The obligation to act in a person's best interests, and the protection from legal liability if they do so without negligence, extends to anyone carrying out acts of treatment and care, such as physical assistance, doing shopping, giving medical treatment, or nursing care.

(c) Lasting power of attorney

A person with capacity can choose to appoint someone to act on their behalf (their 'attorney'). The power to act continues to be valid even after the 'donor' loses capacity, provided that the document giving the power has been registered with the Public Guardian (a newly created statutory office, replacing the former Court of Protection). The donor can give authority to his attorney to take both financial decisions *and* decisions relating to his health and welfare, provided that he is shown to lack the capacity to make those decisions himself at the relevant time. The attorney has the duty to act in accordance with the principles of the Mental Capacity Act 2005 and the guidance of its Code of Practice, and must act in the best interests of the donor.

(d) Advance decisions to refuse treatment

The Mental Capacity Act 2005 also provides a framework and safeguards for a person who wishes to decide in advance what treatments should be withheld if they lose capacity in future. Advance decisions to refuse life-sustaining treatment have to be explicit, in writing, signed, and witnessed.

(e) Driving

A common problem concerns the ability of older patients with early cognitive impairment to drive safely.[20] Scores on simple cognitive tests (e.g. the Mini-Mental State Examination) are very poorly correlated with driving ability, except where there is severe and obvious impairment. No quick objective test of driving skills has yet been devised. A worried but competent patient can often be reassured by booking an hour with a driving instructor. Loss of the freedom to drive represents such a loss of independence and enjoyment, and such a blow to self-esteem, that advice to give up driving may be strenuously resisted. On the other hand, families and professionals are conscious of potential risks to the public even if the patient denies these risks. Legally, the position in England and Wales is straightforward: individuals have a duty to notify the Driving and Vehicle Licensing Authority (**DVLA**) if they have an illness which might impair their ability to drive. A doctor must tell the patient who has such an illness that they are under an

obligation to inform the DVLA; if it seems that this advice has been ignored, the doctor has an obligation to inform the DVLA himself—the duty of confidentiality is overruled by that requirement. Thereafter, the DVLA will arrange for the patient to receive an independent medical examination, and it is the DVLA which decides whether the driving licence should be withdrawn. If the patient fails to attend the medical examination (whether through forgetfulness or lack of insight), the licence is automatically withdrawn.

Further information

Baldwin, C. (2005). *Making difficult decisions. The experience of caring for someone with dementia.* Alzheimer's Society, Gordon House, 10 Greencoat Place, London SW1P 1PH. (ISBN 1 872874 94 0).

Alzheimer's Society website: www.alzheimers.org.uk.

Ballard, C.G., O'Brien, J., James, I., *et al.* (2001). *Dementia: management of behavioural and psychological symptoms.* Oxford University Press, Oxford.

Herr, J.J. and Weakland, J.H. (1979). *Counseling elders and their families—practical techniques for applied gerontology.* (Springer series on adulthood and aging, Vol. 2). Springer, New York.

Jacoby, J., Oppenheimer, C., Dening, T., *et al.* (eds.) (2008). *The Oxford textbook of old age psychiatry.* Oxford University Press, Oxford.

References

1. Anconi-Israel, S. and Ayalon, L. (2006). Diagnosis and treatment of sleep disorders in older adults. *The American Journal of Geriatric Psychiatry*, **14**, 95–103.
2. Sivertsen, B. and Nordhus, I.H. (2007). Management of insomnia in older adults. *The British Journal of Psychiatry*, **190**, 285–6.
3. Mosimann, U.P. and Boeve, B.F. (2008). Sleep disorders in older people. In *The Oxford textbook of old age psychiatry,* Chapter. 36. (eds. R. Jacoby, C. Oppenheimer, T. Dening, and A. Thomas). Oxford University Press, Oxford.
4. Monk, T.H. (2005). Aging human circadian rhythms: conventional wisdom may not always be right. *Journal of Biological Rhythms*, **20**, 366–74.
5. Reid, K.J., Martinovich, Z., Finkel, S., *et al.* (2006). Sleep: a marker of physical and mental health in the elderly. *The American Journal of Geriatric Psychiatry*, **14**, 860–6.
6. Uchiyama, M., Isse, K., Tanaka, N., *et al.* (1995). Incidental Lewy body disease in a patient with REM sleep behaviour disorder. *Neurology*, **45**, 709–12.
7. Ferini-Strambi, L., Fantini, M.L., Zucconi, M., *et al.* (2005). REM sleep behaviour disorder. *Neurological Sciences*, **26**, Part s3, s186–s92.
8. Nau, S.D., McCrae, C.S., Cook, K.G., *et al.* (2005). Treatment of insomnia in older adults. *Clinical Psychology Review*, **25**, 645–72.
9. Morin, C.M., Colecchi, C., Stone, J., *et al.* (1999). Behavioral and pharmacological therapies for late-life insomnia. *The Journal of the American Medical Association*, **281**, 991–9.
10. Zwi, R., Shawe-Taylor, M., and Murray, J. (2005). Cognitive processes in the maintenance of insomnia and co-morbid anxiety. *Behavioural and Cognitive Psychotherapy*, **33**, 333–42.
11. Wilkinson, P. (2008). Psychological treatments. In *The Oxford textbook of old age psychiatry*, Chapter 18i to iii (eds. R. Jacoby, C. Oppenheimer, T. Dening, and A. Thomas). Oxford University Press, Oxford.
12. Asen, E.K. (2008). Systemic interventions with older adults and their families. In *The Oxford textbook of old age psychiatry*, Chapter 18iv (eds. R. Jacoby, C. Oppenheimer, T. Dening, and A. Thomas). Oxford University Press, Oxford.
13. Evans, S. and Garner, J. (eds.) (2004). *Talking over the years. A handbook of dynamic psychotherapy with older adults.* Brunner-Routledge, Hove and New York.
14. Eastman, M. (1984). *Old age abuse.* Age Concern England, Mitcham.
15. Fisk, J. (1997). Abuse of the elderly. In *Psychiatry in the elderly* (2nd edn) (eds. R. Jacoby and C. Oppenheimer), pp. 736–48. Oxford University Press, Oxford.
16. Beauchamp, T.L. and Childress, J.F. (1989). *Principles of biomedical ethics* (3rd edn). Oxford University Press, New York.
17. Oppenheimer, C. (1999). Ethics in old age psychiatry. In *Psychiatric ethics* (3rd edn) (eds. S. Bloch, P. Chodoff, and S.A. Green). Oxford University Press, New York.
18. British Medical Association. (1995). *Assessment of mental capacity: guidance for doctors and lawyers.* British Medical Association, London.
19. Lush, D. (2008). The legal framework in the British Isles for making decisions on behalf of mentally incapacitated people. In *The Oxford textbook of old age psychiatry*, Chapter 43 (eds. R. Jacoby, C. Oppenheimer, T. Dening, and A. Thomas). Oxford University Press, Oxford.
20. O'Neill, D. (2008). Driving and psychiatric illness in late life. In *The Oxford textbook of old age psychiatry*, Chapter 43 (eds. R. Jacoby, C. Oppenheimer, T. Dening, and A. Thomas). Oxford University Press, Oxford.

8.7

The planning and organization of services for older adults

Pamela S. Melding

Introduction

When does an individual become an older adult? When they show signs of ageing? When they retire from work? When their health becomes frail? When they feel old? When society says they are old? Any of these indicators could define an 'older person' anywhere between 40 and 90 plus years! However, it was for statistical simplicity that many jurisdictions chose the chronological age of 65 to mark the change in status from mature adulthood to 'older person', mainly to establish an age for expected retirement and entitlement to certain benefits, including access to geriatric health services. When this arbitrary discriminator was instituted in the mid-twentieth century, 70 years was a good lifespan for most people. However, over the past 50 years, life expectancy steadily increased and is currently advancing at 6 weeks per annum, boosting the overall number of adults over 65 years and, particularly the over 80 years cohort.[1] Increasing life expectancy, due to improved health care and lowering birth rates, is causing worldwide 'population ageing'. This phenomenon will affect all health and mental health services in future years.

Already, health care resource and cost implications of population ageing for health services are enormous. Older adults occupy about two-thirds of general hospital medical, surgical, and orthopaedic beds; they are the greatest users of primary care and prescription medicines. Internationally, late-life illness takes up a considerable proportion of government or insurance funded health care budgets. As an example of the enormous costs involved, in 2003/2004, the United Kingdom's NHS spent around 43 per cent (£16.471 billion) of its hospital and community health services budget on people over the age of 65, and the cost of community and residential care for older people was 44 per cent (£7.38 billion) of all social welfare budgets.[2] These figures will rise dramatically over the next decades.

Mental Health Services have been slow to anticipate that population ageing will also increase the need for psychiatric services for older people. In many areas of the world, services are scarce, sporadic, or sub-standard. Even in developed nations, there is considerable variation in availability from one area to another. In the past decade, practically all OECD countries have promoted policies of de-institutionalization and community-based care for the elderly, in response to rising cost pressures associated with population ageing, plus a requirement to improve satisfaction for increasingly knowledgeable and assertive consumers, by providing better quality in all health services for older adults, including mental health.[3]

The need for services for older adults

Epidemiology

Whilst most people aspire to longevity, it can be a mixed blessing. Living longer increases risk of developing chronic degenerative diseases of body and brain, which can precipitate mental illness, possibly for the first time in life. Psychiatry services for older adults see a full range of new and chronic psychiatric disorders. However, the commonest new threats to mental health in late life are affective disorders and dementia.[4] Healthy, community-dwelling older adults are generally resilient, with a prevalence of major depressive illness of about 3 per cent, but studies of people in residential care find significant depressive symptoms in 14–42 per cent,[5] and in populations over 80 years, about 40 per cent.[6] Depression occurs in 15 to 40 per cent of medical inpatients[7] and in approximately 35.9 per cent of patients in geriatric rehabilitation units.[8] Among the many aetiological contributors to mental health problems in late life, physical illness and poor health are major risk factors, particularly for depression,[9] the risk increasing with disability.[10] Co-morbid physical disorders can mask depression or impede management with psychotropic medications. Although, pure anxiety disorders reduce in late life to about 1 per cent in community-dwelling older adults, anxiety co-morbid with other psychiatric disorders, particularly depression, is more common at approximately 4 per cent. Older people with anxiety disorders are high users of health care resources and may initially present with physical symptoms.[11] Affective disorders are often multi-dimensional, their treatment complicated, and frequently, they need joint management with geriatric colleagues.

Dementia affects 5–7 per cent of people at 65 and 20 per cent of those over age 80 years. By 2040, the number of people with the disorder will double in the developed countries of Europe and North America.[12] Older adults with dementia are an exemplar of consumers who require multi-disciplinary management. While geriatricians manage the majority of patients, those who exhibit behavioural and psychological symptoms of dementia (BPSD), approximately one-third, do best with additional specialist psychiatric

expertise and management.[13] People with dementia can live for many years as their disorder progresses, requiring increasing levels of support from family, social workers, nurses, residential care providers, and other community health care workers in addition to psychiatry services.

Why are specialist services required?

As indicated above, mentally or physically frail older adults have complex needs and frequently require a broad, multi-disciplinary approach in several domains, (a domain being a broad area of specialist services i.e. mental health, geriatrics, primary care, social services, etc.). This can be difficult to achieve if provision is by disjointed services. In many places in the world, lack of specialist services requires generic mental health services to treat older adults with dementia, depression, or other mental illnesses but this practice risks medical needs being unnoticed, or unappreciated. Frail older adults can find the experience of inpatient care in units with younger psychotic patients frightening and unacceptable. There are also different perspectives for working-age and geriatric psychiatry. Many working-age services promote a recovery model whereas care of older adults focuses on maintaining function, improving quality of life and paying more attention to the spiritual, environmental, and social influences on mental health. In addition, there is greater need to involve the social and family network of mentally frail elderly people than there is for working-age adults. For older adults, specialist services, with close collaboration with geriatricians and other geriatric providers, are preferred for optimal management.

Principles of good service delivery for older adults

Optimal service delivery starts by establishing the principles that services wish to adopt. These should govern the ethos of service delivery. For the World Health Organization (WHO), the mnemonic CARITAS (Latin for Compassion) summarizes a global consensus on specific values required for good service delivery for psychiatry of old age.[14] These principles, championed over many years by many international pioneers of mental health services for older adults, assert that optimal services are:

- **Comprehensive**

  They take all aspects of the patient's physical, psychological, and social needs and wishes into account i.e. are *patient-centred*.

- **Accessible**

  They minimize the geographical, cultural, financial, political, and linguistic obstacles to obtaining care.

- **Responsive**

  They act promptly and appropriately to a wide variety of patient needs.

- **Individualized**

  They focus on each person in her/his family and community context aiming, wherever possible, to maintain and support the person within her/his home environment.

- **Transdisciplinary**

  They optimize the contributions of people with a range of personal and professional skills and facilitate collaboration with voluntary and other agencies.

- **Accountable**

  They accept responsibility for assuring the quality of the service delivered, monitoring this in partnership with patients and their families. They are ethically and culturally sensitive.

- **Systemic**

  They work flexibly with all available services to ensure continuity of care.

Summarizing, good services provide *patient-centred* care with *easy access to a comprehensive range of services* delivered by *multi-disciplinary personnel* working in a *collaborative, responsive, respectful, and accountable* way.

Patient-centred care

The UK National Frameworks for Older People[15] and most OECD health administrations promote patient-centered care as a major means of improving quality of services and consumer satisfaction. Whilst most health professionals believe that they already practice patient-centred care, many patients would not agree. Predominantly, health systems for older people, particularly hospital-based, are far from patient-centered, being mostly organized around clinician or administrative requirements rather than patient needs. Patients encounter rigid appointment times, lack of evening or night services, inflexible boundaries between departments, inconvenient visiting hours, limited or expensive parking, and silo'd funding streams. Access or contact processes are obscure or difficult, information is inadequate, multiple assessments take place, and poor coordination between providers leads to treatment omissions or errors. Notably, older adults and their caregivers consider having multiple referrals to different specialists or providers and frequently repeating the same history to be a waste of time and resources.[16]

Consumer appraisals of their experience of services often result in common themes. Many experience poor communication—provider to patient, provider to caregiver, and provider to provider. Another common topic is lack of flexibility in developing management plans capable of involving several domains and dimensions of care (a dimension is a subset of a specialist domain, i.e. depression, dementia, continence, or mobility). As many elderly people have difficulties with mobility, or live far from services, transportation is another major issue.

In contrast, patient-centered services emphasize smoothing the progress of the patient 'journey' through the health system, eliminating duplication, matching care plans to patient needs, and generally making a demanding experience easier for the patient and their family. Unsurprisingly, the concept is appealing to patients and families. Increasingly, consumers, and their advocates, want to contribute to service planning, delivery, and evaluation.[17] So, what do older people want from their health care providers? Older people value their independence and being involved in the decision-making for their own care plans. Most want to remain in their own homes for as long as possible, but if that is detrimental for them, they want the right to relinquish decision-making, in various degrees, to other parties such as family members or their clinicians. They expect providers to treat them as an individual, to preserve their dignity, and to elicit and respect their preferences. Above all, they wish to be appropriately informed.[18] These desires are appropriate to all health care delivery for older adults, not just mental health. As most of us hope to grow old, we can empathize

with these wishes. Perhaps we need to remember that in planning and organizing services for older adults, we are potentially designing them for ourselves. The quality of care required is what we would be happy with, if we ever become clients.

Comprehensive and integrated services

Patient-centered care is more achievable with comprehensive or integrated services. The terms comprehensive and integrated are not interchangeable. A comprehensive service is one with a full range of inpatient and community services available within the same domain. An integrated service is one with a single point of entry capable of providing care plans that incorporate interventions and support in multiple domains and specific dimensions of care. Integrated services are characterized by a single point of entry, case management, geriatric, psychogeriatric and social assessments, and have multi-disciplinary teams.[3] They should have a seamless joining together of the various components of service, encompassing 'systemness' without diminishing component part identities.[19] Integration of different organizations is much easier if there are common administrative processes and financial systems. However, many mental health organizations have reporting and funding structures separate from other older adults' services (e.g. The Mental Health Trusts in the United Kingdom). Quasi-integration can be achieved by building effective *functional* links with a wide variety of health care professionals outside mental health e.g. primary care, geriatrics, acute medical and surgical care, social care, community health care, and non-clinical resources. For these collaborations to work, it is essential that bureaucratic processes enable easy transfer of funding and information across different entities and do not thwart clinicians' efforts to implement care plans for patients.

Integrated services are potentially more efficient as they should reduce duplication of assessments or investigations and service gaps. Currently, the majority of well-established services for older adults provide comprehensive rather than integrated services,[20] the latter being more ideology than practice, although this might be slowly changing. Research indicates that integrated care can delay institutionalization, reduce costs, and has benefits in consumer satisfaction[3] but is insufficient, as yet, to demonstrate that integrated services are more effective in achieving better health outcomes.

(a) The place of the common geriatric assessment (CGA)

An important tool for assisting integration is the common geriatric assessment (CGA). Different disciplines all have their own styles and foci for assessment but, despite individual differences, it is useful to have some common information for all teams and multidisciplinary groups, regardless of who takes the main responsibility for the patient.[21] Advocacy for the comprehensive geriatric assessment (CGA) covering all the main domains and dimensions of physical illness, mental health, disability, and social assessment, is increasing internationally, notwithstanding a lack of research on their effectiveness in improving health outcomes. They aim to save a patient from multiple repetitions of the same information.[16] To be useful, CGAs require personnel to work across professional and agency barriers, which can have benefits in creating relationships with allied colleagues, essential for developing integrated services. There needs to be agreement amongst the providers as to the applicability of the information required, agreed processes by which the CGA generates onward referral to the appropriate domains of care and procedures for updating and review. In some jurisdictions, (e.g. United Kingdom) CGAs or single assessment processes (SAPs) are mandatory for all older adults' services, in others, i.e. New Zealand and Australia, they are being trialed with a view to future obligatory use. SAPs and CGAs vary in their comprehensiveness and can aim at different levels, e.g. screening, proactive assessment, primary care, or secondary care services. They provide useful background information common to a range of providers but are not a substitute for specialized clinical assessment.

(b) The 'core business' of mental health services for older adults

Irrespective of whether a mental health service for older adults is part of a comprehensive or integrated system, their 'core *business*' is the:

- Diagnosis and management of new cases of mental illness arising in late life, often associated with the ageing process.
- Treatment of mental illness complicating physical illness and disability.
- Management of older adults with long-term mental illness complicated by ageing.
- Education and support for caregivers of older adults with mental illness.

Most psychiatry for older adults is about the management of chronic illness and care, rather than cure, is usually the main priority. An adaptation of the 5As model for patient-centred chronic illness management[22] is useful to describe the 'core *tasks*' of patient-centered psychiatry of old age. They are:

- Assessment of multiple care needs
- Advice on diagnosis and options for management
- Agreement with patient and caregiver on a care plan
- Assistance with implementation of care plan
- Assertive follow-up when needed

The 'core areas of *expertise*' for specialist services for older adults are the:

- Treatment of affective and psychotic disorders in late life
- Assessment of neurocognitive disorders and the management of the behavioural and psychological symptoms of dementia (BPSD)
- Rehabilitation of long-standing, chronic psychiatric disorders in patients whose disorder is complicated by physical illness or ageing
- Management of delirium in medically ill or complicating dementia
- Liaison with families, caregivers, and community providers

Core components of psychogeriatric service delivery

The evidence base for the 'core components'

Working with mentally ill older adults involves a variety of locations i.e. the patient's own residence, medical and surgical wards in a general hospital, psychiatry inpatient facilities, residential care facilities, outpatient clinics, outreach clinics, geriatric rehabilitation

units, or day hospitals. Several models of services for older adults have evolved over the past 30 or 40 years, shaping a degree of accord amongst clinicians on what are 'core components'. The evidence for these has been systematically reviewed by Draper and Low.[23] (See Table 8.7.1.)

Community old age psychiatry services

The lynchpin of geriatric psychiatry services is the community-based assessment and case management team. The community team model originated from the closure of the mental hospitals and the move into community-care in the 1980–90s. Their focus is domiciliary assessment and management. This offers the clinician opportunities to observe patients in their own environment, and promotes optimal cognitive functioning by decreasing stress for the individual. Home assessment avoids the sometimes-perceived stigma of attending a psychiatric clinic and eliminates transportation difficulties, as well as allowing ready access to family members or other caregivers. Treatment for most patients can take place at their residence (own home or nursing home), reducing reliance on inpatient or residential care. The model has proven to be efficient and highly acceptable to consumers and caregivers.

A typical multi-disciplinary team consists of at least one psychiatrist, psychiatric nurses, clinical psychologists, social workers, occupational, and other therapists plus support and administration staff. A psychiatrist traditionally leads a multi-disciplinary team but not necessarily so. As a whole, the team should be able to address the biopsychosocial, therapeutic, and psychoeducational requirements of a wide range of disorders and intervention settings. Team members case-manage depending upon their own special skills and expertise. Working collaboratively, with appropriate training, good supervision, and well-designed protocols and communication systems enhances the multi-disciplinary team. Some of the most effective teams are 'interdisciplinary' who develop flexible working patterns characterized by a non-hierarchical structure, and shared decision-making. They facilitate lateral communication between team members, and free exchange of ideas to develop optimal treatment and support management plans as a group. The evidence for community-delivered specialized multi- or interdisciplinary psychogeriatric assessment and management teams is strong and indicates consistently better outcomes than 'usual care' of primary care or generic mental health management.

Inpatient units

Inpatient units vary from the specialized older person's assessment, treatment, and rehabilitation (ATR) unit, similar to their geriatric counterparts, to dedicated beds in geriatric wards, or donated beds in working-age mental health units. Preferably, a specialized inpatient unit is purpose built or has the functionality to separate patients with functional and organic disorders, as each group has different clinical features, nursing needs, disabilities, and requirements for care plans. The evidence for specialized inpatient psychogeriatric units is positive but as there have been few random controlled trials (RCTs), it is less robust than the evidence for

Table 8.7.1 Level of evidence and study qualities for areas of service delivery

Area	No. of studies reviewed	No. of controlled trials	Quality range,* range (low) 0–1 (high)	Mean rating of quality	Level of evidence of effectiveness
Psychogeriatric day hospitals	10	0	0.43–0.82	0.57	Level IV (particularly depression)
Community old age psychiatry services	24	7	0.79–0.94	0.87	Level I for multi-disciplinary psychogeriatric teams, level IV for adult psychiatry teams
Integrated hospital and community-care	4	2	0.71–0.82	0.76	Level II for psychogeriatric services post-discharge care, no evidence for geriatric medical services (level I)
Primary care collaborations	3	2	0.89–0.94	0.92	Level II
Older people in general adult psychiatric wards	6	0	0.51–0.67	0.59	Level IV
Acute psychogeriatric wards	23	0	0.43–0.78	0.61	Level III-2
Hospital medical services	6	2	0.82–0.90	0.89	Level II for prevention of delirium without dementia, no evidence for other mental health outcomes (level 1)
Combined psychogeriatric and medical wards	3	0	0.52–0.53	0.52	Level IV
Hospital-based CL psychogeriatric service delivery	7	3	0.62–0.90	0.79	Level II effectiveness for reducing costs and length of stay
Long-term psychogeriatric care	11	0	0.58–0.71	0.66	Level III-2
Psychogeriatric outreach to long-term care	8	6	0.73–0.95	0.84	Level II for liaison style outreach services, Level III-2 for consultation style
Overall	108	25	0.62–0.95**	0.85**	

*If ≥2 controlled trials in service area, quality range reported for RCTs only, otherwise reported for all studies. **overall mean and range reported for controlled trials only.
(Reproduced from Draper, B. Melding, P. and Brodaty, H. (2005) Psychogeriatric Services Delivery: An international perspective, copyright 2005, with permission from Oxford University Press.)

community teams. Nevertheless, what evidence exists points to specialized psychogeriatric units having better outcomes for psychogeriatric patients than working-age mental health, or geriatric medical units.[23] Scientific evaluation of combined psychogeriatric and geriatric medical wards is inadequate but expert clinical opinion considers them useful in the management of co-morbid medical and psychiatric illness.

Consultation and liaison

Hospital-based consultation and liaison (CL) services are important components because of the high number of mental disorders in the physically ill, general hospital populations. However, while the evidence is relatively good for CL outcomes such as hospital stays and costs, it is only modestly positive for mental health outcomes.[24] Notwithstanding, in recent years, older adults' CL services have outreached from the general hospital to provider organizations in the community. The evidence for these liaison services, usually provided by community teams to psychogeriatric long-term care facilities and voluntary organizations, is relatively good.[23]

Day Hospitals

Day Hospitals, originally attached to the old psychiatric institutions, have mostly devolved into community facilities and many United Kingdom psychiatrists of old age consider them indispensable.[25] Specialist day hospitals provide care for people with moderate and severe needs, including people with functional mental illnesses such as depression, anxiety, and schizophrenia, who may need specific support with daily activities and people with moderate to severe dementia. The Day Hospital allows hospital level treatment while allowing patients to remain living in their own residence. The evidence for their effectiveness, while positive, is sparse. As they are less common, they are not as revered in other parts of the world as much as they are in the United Kingdom. Day Hospitals need to be differentiated from Day Centres, managed by the voluntary or welfare sectors, which usefully provide social activities to keep the patient involved with their community and much needed respite for caregivers.

Residential care

An elderly person unable to support themselves in their own home needing daytime or 24 h supervision requires a resident caregiver or residential care facility. A community may have a range of residential care facilities managed by a variety of agencies such as local authorities, voluntary organizations, for-profit or not-for-profit religious and welfare organizations in the independent sector. These non-government organizations (NGOs) provide care homes for older adults who need support because of physical or mental frailty but not to the extent of requiring hospital level care. Many of the residents have some cognitive impairment and, unsurprisingly, a relatively high number of patients have complicating depression and/or psychosis. Residential care homes vary markedly in their ability to support older adults with mental health problems. Evidence from The Netherlands[26] and Australia[27] indicates that mental health enhanced care in the form of regular psychogeriatric team liaison to these residential facilities has beneficial outcomes for patients for example less inpatient stays, less psychotropic medication, and improvement in depression and psychosis.

Long-term care

Progressive de-institutionalization since the 1980s saw the transfer of many long-stay, public sector beds, for patients with dementia or intractable mental illness, to community-based nursing homes or hospitals. These facilities were generally smaller, more home-like and provided additional activities for the residents than traditional long-term care did. The change required many old age psychiatry services to collaborate with independent sector stakeholders and evolve outreach and liaison models of care with the new long-term residential care facilities. This paradigm shift for service delivery was beneficial for patient outcomes.[23]

Disturbed behaviour resulting from dementia (BPSD) can be severe enough for patients to need Specialized Care Units (SCUs). Examples are in Italy, France (CANTOUs), Tasmania (ADARDS unit), and London (Domus units). Such units have a small number of beds, well-trained staff with high staff to patient ratios and access to ongoing specialized psychogeriatric team care. When compared with traditional psychogeriatric ward patients, unsurprisingly, those in a specialized care unit, show improvements in cognition, self-cares, activity participation, and behaviour. The key elements to success appear to be training staff to anticipate and recognize mental health problems and close liaison with specialist mental health teams.

Respite care

The community-care model relies on family, rather than professionals to attend to most of the needs of a mentally frail, elderly person. This is demanding, difficult work for caregivers who usually have other responsibilities of family, home, and jobs. The main focus of respite care is to give lay caregivers a break from caring every few weeks, so they may continue to provide care and thus delay the need for permanent nursing home care as long as possible. Respite can be in the patient's own home, by providing a 'live-in' professional caregiver, or in hospital or a residential care facility for a week or two, every couple of months, allowing the family caregiver to take a holiday. Whilst seeming an admirable concept, the effectiveness of respite care is doubtful and there is little evidence that it has a significant effect on caregivers' burden, psychiatric status or physical health, or on patients' cognition, function, physical health, or rate of institutionalization.[28] Respite care, even when available, may be poorly utilized. Caregiver barriers to using respite care include guilt, financial reasons, cultural attitudes, and fear of stigma. Access barriers include unavailability, lack of publicity about services, long waiting lists, and poor identification of at-risk caregivers. Patient barriers include severe problem behaviours, immobility, incontinence, wandering, and inability to communicate.[29]

Primary care collaborations

Despite older adults forming the majority of patients seen by general practitioners (GPs), they detect and treat less than half the number of older adults with mental disorders.[30] Short consultation times, with a concentration on physical symptoms, with few patients presenting explicitly with mental health complaints, plus a reluctance of older adults, especially men, to express psychological distress to their GP, leads to under-recognition of mental disorders in older adults. Depression may be erroneously attributed to loneliness or ageing and early dementia overlooked. Primary care practitioners also make fewer decisions to treat or refer patients

to specialist services and often preferred to monitor, or defer decisions.[31]

Mere cooperation between mental health services and primary care seems insufficient to improve matters. Collaborations such as mental health enhanced primary care looks more promising. Education of primary care nurses in recognition of mental disorders and use of screening instruments might be useful to improve identification of mental disorders.[31] Even better is the idea of 'embedding' nurses, who have the skills necessary to identify health problems and coordinate care for older adults, into primary care practices. One experimental scheme in New Zealand, 'The Coordinators of Services for the Elderly' (COSE project) works within a primary care small group of practices to coordinate care of the practices' elderly patients across health, mental health, community, and accident services. An RCT of the COSE project over usual care significantly demonstrated that patients in the COSE arm were less likely to be hospitalized, their residential care was delayed and morbidity reduced.[32]

The primary care physician is crucial to patient care from start to finish, not only in identifying people who are at risk but also for any post hospital discharge care as the majority of patients seen by specialist services eventually return to primary care for ongoing management, in conjunction with their family caregivers. Even if patients or their caregivers can self-refer to secondary services, it is important not to bypass the general practitioner, who has awareness of the patient's overall health care and context.

Special components of services

Memory clinics

The substantial growth in numbers of memory clinics, over the past 10–15 years, was stimulated by the licensing of cholinesterase inhibitor drugs for Alzheimer's disease.[33] Memory clinics offer a range of services from assessment of cognition to specialized treatment of memory problems. Dedicated memory services can improve diagnostic expertise and lessen stigma for the patient. Usefully, they often focus on education of patients and caregivers as well as monitoring medications. However, the concept of a 'memory clinic' flies in the face of the trend towards community-based services and integration with local services. Furthermore, the intervention base is often very narrow. Very few studies provide any evidence of increased mental health gain over other psychogeriatric services, but there is some evidence that the attention, communication, and counselling offered increases consumer satisfaction.[34]

Older adults with intellectual handicap

People with intellectual handicap are also living longer and they are at particular risk for developing dementia. Over 55 per cent of people with Down's syndrome between the ages of 60–70 have Alzheimer type dementia, which often begins to emerge in midlife rather than old age. Consequently, the intellectually handicapped older person has complex care needs that require dementia, mental health, and learning disability services to work together assiduously.[2]

Older prisoners

An often forgotten group is ageing prisoners, who are also increasing in number. Older prisoners have increased risk for depression and other mental health problems. Long-serving older prisoners may develop dementia and require the special challenge of care delivery whilst incarcerated. Specialist services for older prisoners are scarce and, if available, usually provided by visiting community assessment and treatment services in conjunction with local forensic psychiatry services.

Younger people with dementia

Early onset dementia is fortunately rare but when it occurs, it is devastating. The patients usually have family responsibilities with young children, jobs, and financial commitments. Early onset dementia usually has a more accelerated course and genetically related family members may be concerned for themselves or their offspring. Diagnosis is often delayed for younger people so considerable distress and problems have usually built up before they reach services. Their management needs may also be different as they are usually physically fitter and they may require more structured activities than their older counterparts do, but need similar levels of supervision. Appropriate services for younger people who develop dementia are often scarce and the patients may fall into an under-resourced gap between working-age and older adults' services.

Academic units

Although there has been a growth in the number of geriatric psychiatry academic units and positions worldwide, they are still under-represented in universities worldwide. They are important providers of under and postgraduate teaching and research.

Planning to commissioning

Commissioning is 'the process of specifying, securing, and monitoring services to meet the needs of a population at a strategic level'.[2] Proposed services require a 'business case' with the place of an intended service clearly demonstrated in the overall schemata of health provision for a population. Demonstration of the need for services and projections of likely demand is necessary and the chief scientific tools available to identify these are epidemiology, demography, and utilization studies.

Epidemiology predicts the likely problems in a population and demography the characteristics of the population that could increase risk. The older the population, the more likely it will have a high prevalence of dementia requiring services. An important demographic to consider when planning services is the socio-economic status of a population. Low socio-economic status increases the likelihood of poor physical health, poor functionality and deprivation causes stress, leading to poverty of control over one's life, low self-esteem, anxiety, insecurity, and depression.[35] Also important is the number of immigrant residents. Ethnic elders are more likely to have earlier social disadvantage compounding in later life into a multiple jeopardy of social disadvantage, poorer physical health, and mental health problems.[36]

Whilst epidemiological studies may predict potential need, and demographics highlight areas of possible risk, not all people with problems will demand services. Demand or utilization is considerably less than the potential need as predicted from the demographical and epidemiological data. International research consistently shows similar demand patterns. About half of identified patients obtain treatment from a health care provider. Only 10–16 per cent reaches specialist mental health services, and

primary care treats about 30–40 per cent. For older adults utilization is even less than for working-age adults.[37] Stigma can have a detrimental effect on willingness to access services[38] as may cultural barriers.[36] For individuals, utilization of mental health services is more likely if the disorder is severe, is adversely affecting the family or social network, or the patient is female.

Services are usually commissioned based on expected demand rather than predicted need. Demand often increases, outstripping supply once services are available and information about them permeates into the community. Resource review on a regular basis as demand rises is necessary.

Commissioning new or revised services involves stocktaking of available resources for older adults. Often these are inequitable, with urban areas enjoying a range of services that rural areas lack. Despite unique individualities of different countries of the world, some characteristics tend to be true of all rural communities. They tend to be older and poorer than urban populations with a higher percentage of females.[39] Stoical older adults living in rural areas have a high tolerance of distress and are often reluctant to seek help from mental health services.[40] Consequently, rural populations are even more vulnerable for late-life mental health disorders, yet outreach services are usually infrequent, and primary care limited. Some countries with very large rural areas, for example United States, Canada, Africa, Australia, and New Zealand commission novel methods of outreach, such as flying doctor services.

Ideally, service planning should be in conjunction with key stakeholders, that is consumers, other geriatric services and providers, including NGOs. For optimal patient-centred mental health care for older adults to be effective, individual specialist hospital services need to work with general practitioners and other community providers. Attention to administrative pathways and funding structures that promote collaboration between providers is important. Commissioning strategies that propose integration of existing services need to recognize that, whilst the end result *may* be beneficial to patients, the process *will* cause enormous upheaval and distress among the workforce involved, particularly if there is decommissioning or amalgamation of duplicate services.

Service development is contingent on there being a skilled, educated workforce. A development plan needs to consider the specialist disciplines required, based on expected case mix and work loads, staff availability, recruitment and retention, morale, job satisfaction, and very importantly, training issues. Clinical services have important roles in teaching and training ongoing professional development and in clinical evaluation, and systems research. These tasks can also be important means of cross-fertilization, dissemination of ideas between related services, teams, and disciplines. Involvement of different disciplines in undergraduate or postgraduate teaching programmes including psychiatry trainees can be useful for disseminating a broader perspective to prospective health professionals intending to work with older adults.

A vital aspect of commissioning is monitoring, that is, the methods of service review and evaluation. This is important for continuing quality initiatives and research but also to provide evidence for review and future commissioning of even better services.

Conclusion

Worldwide, population ageing is driving the development of mental health services for older adults. Finite resources, burgeoning costs, expanding therapeutic repertoires, and increasing consumer interest in involvement in health care is challenging health organizations to develop effective, efficient, and economic patient-centred services for older adults. Fundamentally, the quality of services is dependent on health personnel working with their patients and other providers, towards shared aims of improving health outcomes and quality of life. Clinicians' adaptability, flexibility, responsiveness, availability for patients, and willingness to collaborate are the keys to success in developing future services for older adults.

Further information

Draper, B., Melding, P., and Brodaty, H. (eds.) (2005). *Psychogeriatric service delivery: an international perspective*. Oxford University Press, Oxford, New York.

Department of Health and Care Services Improvement Partnership. (2005). *Everybody's business. Integrated mental health services for older adults: a service development guide*. www.everybody'sbusiness.org.uk

World Health Organization. (1997). *Organization of the care and the psychiatry of the elderly*. World Health Organization, Geneva.

Melding, P. and Draper, B. (eds.) (2001). *Consultation liaison geriatric psychiatry*. Oxford University Press, Oxford, New York.

References

1. Draper, B., Melding, P., and Brodaty, H. (2005). Preface. In *Psychogeriatric service delivery*, pp. 4–13. Oxford University Press, Oxford, New York.
2. Department of Health, Care Services Improvement Partnership. (2005). *Everybody's business. Integrated mental health services for older adults: a service development guide*. Older people and disability division, Directorate of Care Services, Department of Health, London.
3. Johri, M., Beland, F., and Bergman, H. (2003). International experiments in integrated care for the elderly: a synthesis of the evidence. *International Journal of Geriatric Psychiatry*, **18**(3), 222–35.
4. Riedel-Heller, S.G., Busse, A., and Angermeyer, M.C. (2006). The state of mental health in old-age across the 'old' European Union—a systematic review. *Acta Psychiatrica Scandinavica*, **113**(5), 388–401.
5. Djernes, J.K. (2006). Prevalence and predictors of depression in populations of elderly: a review. *Acta Psychiatrica Scandinavica*, **113**(5), 372–87.
6. Zarit, S., Femia, E., Gatz, M., *et al.* (1999). Prevalence, incidence and correlates of depression in the oldest old: the OCTO study. *Aging & Mental Health*, **3**(2), 119–28.
7. Gareri, P., Ruotolo, G., Curcio, M., *et al.* (2001). Prevalence of depression in medically hospitalized elderly patients. *Archives of Gerontology and Geriatrics*, **33**(Suppl.), 183–9.
8. Shah, D.C., Evans, M., and King, D. (2000). Prevalence of mental illness in a rehabilitation unit for older adults. *Postgraduate Medical Journal*, **76**(893), 153–6.
9. Cole, M.G. and Dendukuri, N. (2003). Risk factors for depression among elderly community subjects: a systematic review and meta-analysis. *The American Journal of Psychiatry*, **160**(6), 1147–56.
10. Braam, A.W., Prince, M.J., Beekman, A.T.F., *et al.* (2005). Physical health and depressive symptoms in older Europeans: results from EURODEP. *The British Journal of Psychiatry*, **187**(1), 35–42.
11. Flint, A.J. (2005). Generalised anxiety disorder in elderly patients: epidemiology, diagnosis and treatment options. *Drugs & Aging*, **22**(2), 101–14.
12. Ferri, C., Prince, M., Brayne, C., *et al.* (2005). Global prevalence of dementia: a Delphi consensus study. *Lancet*, **366**(9503), 2112–17.
13. Brodaty, H., Draper, B., and Low, L.F. (2003). Behavioural and psychological symptoms of dementia: a seven-tiered model of service delivery. *Medical Journal of Australia*, **178**(5), 231–4.

14. World Health Organization. (1997). *Organization of the care and the psychiatry of the elderly*. World Health Organization, Geneva.
15. Department of Health. (2001). *National service framework for older people*.
16. Stevenson, J. (1999). *Comprehensive assessment of older people. Kings fund rehabilitation programme. Developing rehabilitation opportunities for older people*. Kings Fund, London.
17. Dening, T. and Lawton, C. (1998). The role of carers in evaluating mental health services for older people. *International Journal of Geriatric Psychiatry*, **13**(12), 863–70.
18. Robb, G. and Seddon, M. (2006). Quality improvement in New Zealand healthcare. Part 6: keeping the patient front and centre to improve healthcare quality. *The New Zealand Medical Journal*, **119**(1242), U2174.
19. Melding, P. (2005). Integrating service delivery and quality of care. In *Psychogeriatric service delivery* (eds. B. Draper, P. Melding, and H. Brodaty), pp. 251–80. Oxford University Press, Oxford, New York.
20. Challis, D., Reilly, S., Hughes, J., *et al.* (2002). Policy, organisation and practice of specialist old age psychiatry in England. *International Journal of Geriatric Psychiatry*, **17**(11), 1018–26.
21. Stevenson, J. (1999). *Comprehensive assessment of older people. Kings fund rehabilitation programme. Developing rehabilitation opportunities for older people*. Kings Fund, London.
22. Sheridan, S.L., Harris, R.P., and Woolf, S.H.A. (2004). Shared decision making about screening and chemoprevention: a suggested approach from the U.S. preventive services task force. *American Journal of Preventive Medicine*, **26**(1), 56–66.
23. Draper, B., and Low, L.F. (2005). Evidenced-based psychogeriatric service delivery. In *Psychogeriatric service delivery: an international perspective* (eds. B. Draper, P. Melding, and H. Brodaty), pp. 75–122. Oxford University Press, Oxford, New York.
24. Draper, B. (2001). Consultation liaison geriatric psychiatry. In *Geriatric consultation liaison psychiatry* (eds. P. Melding and B. Draper), pp. 3–34. Oxford University Press, Oxford, New York.
25. Hoe, J., Ashaye, K., and Orrell, M. (2005). Don't seize the day hospital! Recent research on the effectiveness of day hospitals for older people with mental health problems. *International Journal of Geriatric Psychiatry*, **20**(7), 694–8.
26. Depla, M., Pols, J., De Lange, J., *et al.* (2003). Integrating mental health care into residential homes for the elderly: an analysis of six Dutch programs for older people with severe and persistent mental illness. *Journal of the American Geriatrics Society*, **51**(9), 1275–9.
27. Brodaty, H., Draper, B., Millar, J., *et al.* (2003). Randomized controlled trial of different models of care for nursing home residents with dementia complicated by depression or psychosis. *The Journal of Clinical Psychiatry*, **64**(1), 63–72.
28. Flint, A. (1995). Effects of respite care on patients with dementia and their caregivers. *International Psychogeriatrics*, **7**, 505–17.
29. Brodaty, H. and Gresham, M. (1992). Prescribing residential care for dementia- effects, side-effects, indications and dosage. *International Journal of Geriatric Psychiatry*, **7**, 357–62.
30. Speer, D. and Schneider, M. (2003). Mental health needs of older adults and primary care: opportunity for interdisciplinary geriatric team practice. *Clinical Psychology: Science and Practice*, **10**(1), 85–101.
31. Watts, S.C., Bhutani, G.E., Stout, I.H., *et al.* (2002). Mental health in older adult recipients of primary care services: is depression the key issue? Identification, treatment and the general practitioner. *International Journal of Geriatric Psychiatry*, **17**(5), 427–37.
32. Parsons, M., Anderson, C., Senior, H., *et al.* (2007). *ASPIRE: Assessment of Services Promoting Independence and Recovery in Elders*. University of Auckland, Auckland.
33. Lindesay, J., Marudkar, M., van Diepen, E., *et al.* (2002). The second Leicester survey of memory clinics in the British Isles. *International Journal of Geriatric Psychiatry*, **17**, 41–7.
34. Jolley, D., Benbow, S.M., and Grizzell, M. (2006). Memory clinics. *Postgraduate Medical Journal*, **82**(965), 199–206.
35. Marmot, M. (2005). Social determinants of health inequalities. *Lancet*, **365**(9464), 1099–104.
36. Silveira, E.R.T. and Ebrahim, S. (1998). Social determinants of psychiatric morbidity and well-being in immigrant elders and whites in east London. *International Journal of Geriatric Psychiatry*, **13**(11), 801–12.
37. Wang, P.S., Lane, M., Olfson, M., *et al.* (2005). Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **62**(6), 629–40.
38. Brodaty, H., Cathy, T., Claire, T., *et al.* (2005). Why caregivers of people with dementia and memory loss don't use services. *International Journal of Geriatric Psychiatry*, **20**(6), 537–46.
39. Allan, D. and Cloutier-Fisher, D. (2006). Health service utilization among older adults in British Columbia: making sense of geography. *Canadian Journal on Aging*, **25**(2), 219–32.
40. Judd, F., Jackson, H., Komiti, A., *et al.* (2006). Help-seeking by rural residents for mental health problems: the importance of agrarian values. *The Australian & New Zealand Journal of Psychiatry*, **40**(9), 769–76.

SECTION 9

Child and Adolescent Psychiatry

9.1

General issues

Contents

9.1.1 Developmental psychopathology and classification in childhood and adolescence

Stephen Scott

Introduction

Classification schemes of psychiatric disorders in childhood and adolescence have to take into account three particular features. Firstly, the individual is continually changing and growing. Sound knowledge is therefore required of the normal range of development and its limits. For example, some fears may be normal in a 5-year-old but abnormal in an 8-year-old. Once identified, it is helpful to decide if abnormalities are due to *delay in* or *deviance from* the usual pattern of development. The implications of each differ, and should be classified differently. Secondly, the majority of childhood mental health problems arise from an excess of behaviours exhibited by many young people, such as aggression or dieting. They are seldom due to qualitatively distinct phenomena of the kind more often seen in adult conditions, such as hearing voices or hanging oneself. Consequently choosing a cut-off point to make a categorical entity from a dimensional construct is more often used in child psychiatry. This is inevitably an arbitrary process (albeit informed by empirical criteria), which may lead to loss of information, and may be held to be labelling the child unnecessarily. That dimensions can be interchanged with categories does not necessarily mean they are unhelpful—after all, day and night are useful terms yet the boundary between them is continuous and arbitrary. Psychiatrists in particular may be criticized for 'medicalizing' a child's difficulties by talking about disorders or diagnoses, whereas other professionals and parents may prefer to see them as understandable variations in child development, and prefer to call them 'emotional and behavioural difficulties'. However, diagnoses are a quick way to convey a lot of information that dimensions may not. Thus to only say a child is at an extreme of an antisocial behaviour dimension does not necessarily convey the association with specific reading retardation and ADHD, which could be (and often are) consequently missed.

Thirdly, children's difficulties nearly always arise in the context of relationships within the family. More often than in adulthood, some or all, of the problem may appear to be the result of the functioning of the family, rather than in the individual child who may merely be reacting to the situation. For example, a child who is disobedient and shouts in class may simply be behaving the way his parents do at home. A classification system will be stronger if it can take family functioning into account—it will have a greater chance of capturing clinically important causal and therapeutic considerations.

A valid and useful classification system will need to take into account these features and be based on a thorough understanding of normal development and how it can go wrong, rather than merely include static descriptions of presumed pathological states. The term of *developmental psychopathology* was coined in the early 1980s to denote the scientific study of how abnormalities can be understood in terms of processes underpinning human development.[1,2] There are now journals and books incorporating the term into their titles.[3,4] Many disciplines are relevant, from embryology and genetics to social learning theory and criminology. Developmental psychopathology, besides studying the impact of pathogenic influences on pathways through life, such as the way the

monoamine oxidase-A (MAOA) genotype interacts with an abusive upbringing to cause conduct disorder,[5] or how specific reading retardation leads to low self-esteem, also investigates protective mechanisms, such as the ameliorating effect of high IQ on the propensity to juvenile offending, or the beneficial effect of a trusting and caring relationship with an adult on the impact of childhood abuse. This chapter aims to show how findings from developmental psychopathology have informed current classification systems, and what challenges remain.

General issues

Change over time

Because mental processes and behaviour change as a child develops, it is not always clear whether the same diagnoses should be applied across the age range. Thus a highly aggressive toddler may throw himself screaming onto the floor in daily tantrums, whereas a highly aggressive teenager may assault old ladies and rob them. Do they suffer from the same disorder? ICD 10 holds that they do—both meet criteria for conduct disorder, which is defined in terms of antisocial behaviour that is excessive for the individual's age, and that violates societal norms and the rights of others. DSM IV-R on the other hand has two separate diagnoses, oppositional-defiant disorder for the younger case, and conduct disorder for the older. However, as both diagnoses have similar correlates and there is a strong continuity from one to the other, the validity of the division is questionable. Yet current adult psychiatric schemes have no diagnosis at all to apply to antisocial behaviour, unless it is part of a personality disorder.

The extent to which adult criteria should be applied to children requires good empirical data. In the case of obsessive–compulsive disorder, the phenomenology is remarkably similar in childhood, so there is no problem. However, for depression the picture is rather different. Currently, ICD 10 and DSM IV-R have few emotional disorder categories specific to childhood, and they are mostly subtypes of anxiety. Mood disorders are diagnosed according to adult criteria, with the consequence that surveys of depression find prevalence rates close to zero under 8 years of age. Yet there are miserable children who cry frequently, say they are unhappy, look sad, and are withdrawn.[6] However, they usually sleep and eat reasonably well, and their mood fluctuates during the day, with spells when they sometimes appear more cheerful. Should they not be allowed a diagnosis? ICD 9 had a category for 'disturbance of emotions specific to childhood and adolescence, with misery and unhappiness', and such children suffer impairment.[7] Follow-up studies of prepubertal children referred with this picture showed a moderately increased risk of adult type depression later on, whereas adolescents with depressive symptoms had a higher risk of adult depression.[8] Genetic studies show that symptoms of depression in prepubertal children are predominantly due to environmental influences, whereas after puberty genetic influences become more important.[9] Finally, tricyclics are not effective in childhood but are effective in adults. This example shows that misery in younger children has some phenomenological features and external correlates in common with adult depression but also several differences, so the current approach which makes a comprehensive yet parsimonious classification system for all ages loses validity.

In contrast, there is continuing reluctance to diagnose personality disorders in childhood. This may be because they are often seen as a life sentence of a noxious, untreatable condition, in distinction to the general hope that there is opportunity for 'growing out of' conditions in childhood, or treatment for them. However, with perhaps the most destructive personality type, dissocial, there is growing evidence that the combination of antisocial behaviour and callous-unemotional traits is well established by the age of seven. Moreover, this combination of childhood characteristics has a far higher heritability than antisocial behaviour without callous-unemotional traits.[10]

Validity

Categories need to be distinct not only in terms of the phenomena used to define them, but, crucially, also in terms of external criteria. Even if categories can be reliably distinguished, if external criteria are the same, then one is likely to be dealing with two variants of the same condition. An analogy would be the difference between black and white cats.

Typical validating criteria in child psychiatry derived from developmental psychopathology are:

1 *Epidemiological data*, such as age of onset and sex ratio. Forty years ago 'childhood psychosis' was a unitary classification, but work showing the clear difference in age of onset helped validate the distinction between autism and schizophrenia, which seldom co-occur. Disruptive disorders occur four times more commonly in boys, whereas emotional disorders are commoner in girls.

2 *Long-term course*. Most childhood disorders show reasonable *homotypic continuity*, that is they stay the same. Some show *heterotypic continuity*, so that for example, some cases of childhood hyperactivity end up as antisocial adults. This does not necessarily invalidate the category, but requires explanation.

3 *Genetic findings*. If individuals with distinct categorical diagnoses have relatives with different disorders, this helps validate the distinction. This has confirmed the validity of several diagnostic categories, but not all. For example, it has not held for the many specific subtypes of anxiety disorder in ICD 10, whose validity is questionable. Genetic studies can also clarify the scope of symptom clusters. For example, family studies of autism have revealed a broader phenotype in relatives of probands,[11] so that new disorders may need to be considered, which encompass only one of the original three constituent domains of classical autism, namely social relatedness, communication problems, and repetitive and stereotyped behaviours.

   The hunt is now on for specific genes associated with particular psychiatric disorders. Thus dopamine receptor and transporter genes are reliably associated with Attention Deficit Hyperactivity Disorder,[12] but unless (i) the gene always leads to the disorder and (ii) all cases of the disorder are caused by the gene, particular genotypes are unlikely to be used to validate diagnostic categories.

4 *Psychosocial risk factors*. The association between institutional upbringing with many changes of carer and reactive attachment disorder is so strong that it has been made a requirement for diagnosis in ICD 10. Conduct disorders are strongly associated with discords at home, whereas autistic disorders are not. However, most psychosocial risk factors are less specific in their associations, and so are only modestly helpful as validating criteria.

5 *Neuropsychological tests*. The hyperkinetic syndrome is clearly distinguishable from conduct disorder on tests of attention such

as the continuous performance task. Recently, there has been considerable progress in showing that one of the core deficits in autism is failure on 'theory of mind' tests of ability, to see another person's point of view, which non-autistic children, with comparable levels of intellectual disability, can do.

6 *Medical investigations*. There have been many failed attempts in this field, including biochemical markers of adolescent depression and endocrine markers of aggression. However, the advent of functional neuroimaging is allowing exciting relatively non-invasive pictures of children's brains to be built up, and reliable findings are beginning to emerge, for example in ADHD.[13] In future these may well be helpful validators for classification.

Reliability

This is a prerequisite for validity, and most categories have reasonable inter-rater and test–retest values, once investigators are trained up. Where there are many overlapping categories, as in current definitions of the many varieties of anxiety disorders, or personality disorders, inter-rater reliability falls.[14]

Effect of informant and instrument

Traditionally information is obtained from parents and the child, and is then combined by the clinician on a case-by-case basis. However, the need for consistent diagnostic rules that is imposed by a 'menu-driven' approach can prove difficult, since the weight given to a particular informant may best vary according to condition. Thus, if a parent says a child has symptoms of conduct disorder but the child denies it, the parent is more likely to be right and the child may be covering up or ashamed. However, if the parent says the child is not depressed but the mental state examination of the child reveals otherwise, it is the parent who may be ignorant of their child's true state. Such difficulties reduce the validity of interviews which use invariant combination rules. Further, in genetic studies, the heritability of a condition may vary greatly according to which informant is believed. Thus in the Virginia Twin study, conduct disorder was 69 per cent heritable according to the information derived from the mother interview, 36 per cent using information from the child, and only 27 per cent using information from the father.[15] Studies such as these underline the need for clinically sensitive ways of combining information, and the use of multi-informant, multi-method ascertainment of information. Statistical techniques such as latent variable analysis may help reduce measurement error, but may build in unwarranted assumptions which distort the raw data.

Structured interviews, which accept the respondent's reply, do not require lengthy training or clinically informed investigators, and so are popular in epidemiological surveys. However, the quality of information differs little from that obtained by questionnaire,[16] and often has a high false-positive and false-negative rate in comparison to semi-structured interviews. Direct observation, although expensive, often provides the most reliable and valid information for assessment of disruptive disorders.

Comorbidity

There are many artefactual reasons for comorbidity appearing high, such as Berkson's bias[17] in clinical samples (where not all cases get referred, the chance of referral will be related to the *combined* likelihood of referral for each condition separately), or overlapping criteria, or artificial subdivision of syndromes. However, even after taking these possible sources of error into account, comorbidity is marked for child psychiatric disorders. In a meta-analysis of community samples,[18] the odds ratio for anxiety with either Attention Deficit Hyperactivity Disorder (ADHD) or conduct disorder is 3, for anxiety and depression 8, and ADHD and conduct disorder 10. Rates are even higher in clinical samples. True comorbidity may arise through several mechanisms:[19] (i) shared risk factors (e.g. early deprivation may lead to oppositional-defiant disorder and an attachment disorder), (ii) overlap between risk factors (thus a depressed mother may pass on a genetic liability to depression in her son and provide inconsistent discipline which predisposes him to conduct disorder), (iii) one disorder creating an increased risk for the other (e.g. conduct disorder leading on to drug dependency), or (iv) the comorbid pattern constitutes a meaningful syndrome (e.g. depressive conduct disorder, described below under combined categories).

Some classification schemes

A simple scheme with three main groups of disorders

A simple but well researched, valid way of grouping child disorders 'lumps' them into three groups, which are helpful to hold in mind when considering specific diagnoses:

Emotional disorders including anxiety, depression, phobias, somatization, and obsessive–compulsive disorder; **disruptive disorders** including conduct disorder and hyperactivity; and **developmental disorders** including intellectual disability, the autistic spectrum, language and reading delays, and enuresis and encopresis.

Comorbidity within each grouping is very common, but only occurs across groups in a minority of cases. External criteria validating the differences between these groups are given in Table 9.1.1.1.

Table 9.1.1.1 Validating criteria for main diagnostic groupings

	Emotional disorders	Disruptive disorders	Developmental disorders
Age of onset	over 8	under 8	under 3
Sex ratio	commoner in girls after puberty	commoner in boys	commoner in boys
Family size	normal	large	normal
Family history	anxiety and depression increased	criminality increased	related disorders may be increased
Socio-economic status	normal	lower	normal
IQ	normal	lower range of normal	normal or low, sometimes very low
Specific delays	absent	present in a third	common
Neurological signs	absent	uncommon	common
Cause	mixed, sometimes mainly genetic	mixed, sometimes mainly environmental	often mainly genetic

Current schemes: ICD 10 and DSM IV-R

The DSM IV-R and ICD 10 committees worked closely together and strove to have names and criteria that are as close as possible. However, there are some general differences.

(a) 'Picture-fitting' versus 'menu-driven' approaches

Firstly, as in adulthood, ICD 10 has one set of 'clinical descriptions and diagnostic guidelines' and a separate set of 'diagnostic criteria for research'. The former comprises general descriptions of disorders requiring a qualitative matching of case characteristics with the scheme, a 'picture-fitting' approach which is similar to the way clinicians practise. The latter comprises lists of symptoms with explicit criteria detailing the number and permutation required for diagnosis, a 'menu-driven' approach. DSM IV-R has only the latter. It has advantages in increased reliability, but is relatively cumbersome so that many clinicians do not bother to apply the criteria rigorously. Even for the simpler DSM III criteria, a study found that whilst trained researchers achieved kappa values of 0.83, 0.80, and 0.74 for attention deficit disorder, conduct disorder, and emotional disorder, the comparable figures for United States clinicians in regular practice were 0.30, 0.27, and 0.27, which are seriously low.[20]

A further disadvantage of the 'menu-driven' approach arises in cases where although the clinician believes a diagnosis is present because of the severity of symptoms, their number is insufficient to meet criteria. For example, consider the following youth: he repeatedly mugs old ladies, sets fires frequently, often argues, is often spiteful or vindictive, has unusually severe tantrums, and has no friends or job because of his behaviour. According to ICD 10 research diagnostic criteria (or DSM IV-R criteria) he has no diagnosis, as he has two but not three symptoms of conduct disorder, and three but not four symptoms of oppositional-defiant disorder. However, according to ICD 10 'diagnostic guidelines' he easily meets the requirements for conduct disorder since 'any category, if marked, is sufficient'.

(b) Multiple diagnoses

A second difference between ICD 10 and DSM IV-R is in multiple diagnoses. ICD 10 encourages the selection of one diagnosis that closest fits the picture, assuming that differences are due to a variation upon the typical theme. DSM IV-R (and the closely linked ICD 10 research criteria) encourage selection of as many diagnoses as criteria are met. Problems arise with this approach when symptoms are common to two disorders, for example irritability contributes to affective disorders and to conduct disorders, so double coding is more likely. Since comorbidity is very common in clinical practice, multiple coding is frequent using a 'menu-driven' approach so that it begins to approach a dimensional system and to lose the advantages of categorization.

The pros and cons of each approach will vary according to whether extra information is conveyed by the second diagnosis. Where there is good evidence of the validity of common comorbid conditions, ICD 10 has combined categories. Thus the external validating characteristics of 'depressive conduct disorder' are similar to those of pure conduct disorder, with no increase of affective disorders in individuals, nor in their relatives, followed up to adulthood. Double coding would convey erroneous information about the depressive aspect. 'Hyperkinetic conduct disorder', on the other hand, is characterized by more severe neuropsychological deficits than occur in either condition alone, and by worse psychosocial outcome in adulthood. Double coding would not convey the poor prognosis.

(c) Multiaxial framework

The ICD 10 has a multiaxial framework for psychiatric disorders in childhood and adolescence[21] which will be described here. DSM IV-R uses a somewhat different multiaxial framework, which is applicable for disorders arising at all ages. It will not be described here except as a contrast to ICD 10. Each axis except the last (psychosocial impairment) is coded independently of the apparent causal contribution to the psychiatric syndrome. This avoids tricky decisions about causality and allows conditions to be recognized and clinical needs addressed.

(i) Axis one: clinical psychiatric syndromes

Criteria for particular diagnoses are described in the relevant chapters of this text.

(ii) Axis two: specific disorders of development

These include speech and language, reading, spelling, and motor development. In DSM IV-R they are included in Axis one. However, having a separate axis helps to ensure that they are not overlooked. This can easily happen, for example, in children with conduct disorder, where the antisocial behaviour tends to command attention, while in fact one-third of the children also have specific reading retardation (dyslexia), which if untreated worsens the prognosis.[22] It very desirable to administer standardized psychometric tests in order to characterize specific disorders of development.

(iii) Axis three: intellectual level

The categories are no intellectual disability (IQ 70 or over), mild intellectual disability (50–69), moderate intellectual disability (IQ 35–49), severe intellectual disability (IQ 20–34), and profound intellectual disability (IQ under 20). In DSM IV-R personality disorders are also included on the axis.

Subtyping intellectual disability gives a good example of substantial differences which arise when categories are imposed on top of a dimensional construct. If all children with an IQ below 50 are taken together (often together also called severe), and compared with those having an IQ of 50–70 (mild), major differences emerge on independent validating criteria, as shown in Table 9.1.1.2.

From the table it will be seen that there are major differences between the categories on fronts as varied as brain pathology and life expectancy. There is no particular psychiatric pattern arising in children with intellectual disability, rather the incidence of all disorders is raised, so that in those with IQ under 50, fully one half have a psychiatric disorder.[23]

(iv) Axis four: associated medical conditions

All medical conditions should be coded. A few have specific associations with psychiatric disorders, for example tuberous sclerosis predisposes to autism, Cornelia de Lange syndrome to self-injury; Down syndrome on the other hand protects against autism but often leads to presenile dementia. Even where there is no specific disorder, congenital syndromes are often characterized by a particular pattern of behaviour. The study of these *behavioural phenotypes* is a discipline in its own right.

(v) Axis five: associated abnormal psychosocial conditions

These include a range of pyschosocial hazards, from abnormal intrafamilial relationships such as physical or sexual abuse, to mental disorders in other family members, distorted intrafamilial communication patterns, abnormal upbringing, e.g. in an institution,

Table 9.1.1.2 Characteristics of children with severe versus mild intellectual disability

	Severe retardation	Mild retardation
Definition	IQ under 50	IQ 50–70
Social functioning	Invariably marked impairment	Many have minor or no impairment
Cause	Organic pathology in majority	Usually no organic cause evident
Family history	Parents and siblings usually of normal intelligence	Parents and siblings often at lower levels of intelligence
Background	Fairly equal distribution across SES levels Neglect at home unlikely	Much commoner at lower SES levels Neglect at home more likely
Appearance	Dysmorphic features often evident	Normal appearance
Medical complications	Physical handicap common (e.g. cerebral palsy) Major health problems frequent Life expectancy shortened Fertility low	Physical handicap uncommon Health in normal range Life expectancy normal Fertility little impaired
Psychiatric complications	Severe and pervasive disorders such as hyperactivity, autism, and self-injury especially common Presentation of disorders often altered, mental state may be difficult to determine	Disorders similar in type to those found in children without MR, but occur more frequently Form of disorders and mental state examination similar to children without retardation

acute life events, and chronic interpersonal stress arising from difficulties at school. Each is coded dimensionally on a three point scale. As the number of psychosocial adversities goes up, the rate of psychiatric disorders increases.[24] Conduct disorder is particularly associated with poor immediate psychosocial environments. As with other axes, abnormalities are coded irrespective of apparent cause. This is particularly relevant since while perhaps 20 years ago the mechanism was thought to be directly environmental, in the last 10 years good evidence has been collected to show that some environmental characteristics of the home are genetically mediated.[25] For example, the association between lack of books in the home and poor child reading is partly mediated through parents with lower IQ buying fewer books.

(vi) Axis six: global social functioning

Here a judgement is made on a nine point dimensional scale ranging from superior social functioning to profound and pervasive social disability. Unlike other axes, ratings of disability are not independent, but have to be judged as due to a psychiatric or developmental disorder on axes one to three. Thus impairment arising from adverse circumstances cannot be coded—it must arise from intraindividual factors. This rule therefore excludes recognition of psychosocial interventions which aid functioning, from reduction of parental Expressed Emotion to changing schools. DSM IV-R studies often use the Children's Global Assessment scale,[26] an adaptation of the Global Assessment of Functioning (GAF) used in adults. An advantage of the CGAS is that it is rated without impairment having to be caused by psychiatric disorder. A disadvantage is that psychiatric symptoms, rather than impairment alone, contribute to the rating.

(d) Should impairment of social function be part of psychiatric diagnosis?

In general, ICD 10 and DSM IV-R do not require impairment of social functioning to be present in order to make a diagnosis. There are exceptions, thus in DSM IV-R, oppositional-defiant disorder *does* require impairment. With many qualitatively distinct adult disorders, having no impairment criterion makes sense, so that a person experiencing the delusions and hallucinations characteristic of schizophrenia, but able to go to work and form relationships while on neuroleptics still has schizophrenia. But should a child who says he is afraid of dogs and crosses to the other side of the pavement on seeing one, but otherwise functions well, be deemed to suffer from a phobia? If impairment criteria are not applied, very high rates of disorder are obtained in epidemiological surveys. This lacks credibility with the general public, who may then dismiss all psychiatric problems in children, and is unrealistic for clinicians and health planners, who would not see most of the identified individuals as cases needing treatment. For example, a large epidemiological survey[27] found that using DSM III criteria, 50 per cent of children and adolescents had a diagnosis. However, when an impairment criterion was added, the figure came down to 18 per cent. This would appear to be a much more realistic figure. However, it could be argued that social impairment is too constraining, and for example would exclude an adolescent who is fairly depressed but able to function. The term *impact* can be used to include subjective distress as well as impairment, and is gaining in popularity among many child psychiatrists.[28]

Falling through the cracks: children with social impairment but no diagnosis

Diagnostic systems have to be practically useful above all. If they are overinclusive, the risk is that there are too many categories, which have poor reliability and high overlap. If on the other hand they are too exclusive, the risk is that there will be many individuals suffering from problems which are not encompassed by the scheme. In one thorough survey, 9.4 per cent had no diagnosis but significant impairment.[29] Across a variety of 'caseness' measures, the individuals were as disturbed as those with a diagnosis. Many of the difficulties were around relationships with parents and siblings, and arguably, such children who have symptoms associated with psychosocial impairment should be regarded as suffering from a psychiatric disorder.

Conclusion

Classification of child psychiatric conditions has advanced enormously in the last 20 years. There is a much stronger empirical basis to support current schemes, which are grounded in the many scientific disciplines that contribute to developmental psychopathology. Nonetheless there are considerable obstacles to overcome if DSM V and ICD 11 are to be major steps forward.

Further information

To access the journal *Development and Psychopathology*, visit http://journals.cambridge.org/action/displayJournal?jid=DPP

Cicchetti, D. and Cohen, D. (eds.) (2003). *Developmental psychopathology* (2nd edn). Wiley, New York.

Taylor, E. and Rutter, M. (2008). Classification. In *Rutter's child and adolescent psychiatry* (5th edn) (eds. M. Rutter, D. Bishop, D. Pine, *et al.*). Blackwell, Oxford.

Goodman, R. and Scott, S. (2005). *Child psychiatry* (2nd edn). Blackwell, Oxford.

To access the official website discussing issues around DSM IV criteria, visit http://www.dsmivtr.org/index.cfm and regarding ICD 10 visit http://www.who.int/classifications/icd/en/

References

1. Cicchetti, D. (1984). The emergence of developmental psychopathology. *Child Development*, **55**, 1–7.
2. Rutter, M. (1988). Epidemiological approaches to developmental psychopathology. *Archives of General Psychiatry*, **45**, 486–95.
3. To access the journal *Development and Psychopathology*, visit http://journals.cambridge.org/action/displayJournal?jid=DPP
4. Cicchetti, D. and Cohen, D. (eds.) (2003). *Developmental psychopathology* (2nd edn). Wiley, New York.
5. Kim-Cohen, J., Caspi, A., Taylor, A., *et al.* (*2006*). MAOA, maltreatment, and gene-environment interaction predicting children's mental health: new evidence and a meta-analysis. *Molecular Psychiatry,* **11**, 903–13.
6. Puura, K., Tamminen, T., Almqvist, F., *et al.* (1997). Should depression in young children be diagnosed with different criteria? *European Child & Adolescent Psychiatry*, **6**, 12–19.
7. Costello, E.J., Mustillo, S., Erkanli, A., *et al.* (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry*, **60**, 837–44.
8. Harrington, R., Fudge, H., Rutter, M., *et al.* (1990). Adult outcomes of childhood and adolescent depression—I. Psychiatric status. *Archives of General Psychiatry*, **47**, 465–73.
9. Scourfield, J., Rice, F., Thapar, A., *et al.* (2003). Depressive symptoms in children and adolescents: changing aetiological influences with development. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **44**, 968–76.
10. Viding, E., Blair, R.J.R., Moffitt, T.E., *et al.* (*2005).* Evidence for substantial genetic risk for psychopathy in 7-year-olds. *Journal of Child Psychology and Psychiatry, and Allied Disciplines,* **46**, 592–7.
11. Rutter, M. (2000). Genetic studies of autism: from the 1970s into the millennium. *Journal of Abnormal Child Psychology*, **28**, 3–14.
12. Thapar, A., O'Donovan, M., and Owen, M.J. (2005). The genetics of attention deficit hyperactivity disorder. *Human Molecular Genetics*, **14**, R275–82.
13. Rubia, K., Smith, A.B., Brammer, M.J., *et al.* (2005). Abnormal brain activation during inhibition and error detection in medication-naive adolescents with ADHD. *The American Journal of Psychiatry*, **162**, 1067–75.
14. Thomsen, P., Jorgensen, J., and Nedergaard, N. (1992). ICD-10: a field study in a Danish child psychiatric unit. *European Psychiatry*, **7**, 287–91.
15. Eaves, L., Silberg, J., Meyer, J., *et al.* (1997). Genetics and developmental psychopathology: 2. The main effects of genes and environment on behavioral problems in the Virginia twin study of adolescent behavioral development. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **38**, 965–80.
16. Boyle, M., Offord, D., Racine, Y., *et al.* (1997). Adequacy of interviews vs checklists for classifying childhood psychiatric disorder based on parent reports. *Archives of General Psychiatry*, **54**, 793–9.
17. Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometrics*, **2**, 47–53.
18. Angold, A., Costello, E., and Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **40**, 57–87.
19. Rutter, M. (1997). Comorbidity: concepts, claims and choices. *Criminal Behaviour and Mental Health*, **7**, 265–85.
20. Prendergast, M., Taylor, E., Rapoport, J., *et al.* (1988). The diagnosis of childhood hyperactivity: a US-UK cross-national study of DSM-III and ICD-9. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **29**, 289–300.
21. World Health Organization. (1996). *Multiaxial classification of child and adolescent psychiatric disorders*. Cambridge University Press, Cambridge.
22. Trzesniewski, K., Moffitt, T., Caspi, A., *et al.* (2006). Revisiting the association between reading achievement and antisocial behaviour: new evidence of an environmental explanation from a twin study. *Child Development,* **77**, 72–88.
23. Stromme, P. and Diseth, T. (2000). Prevalence of psychiatric diagnoses in children with intellectual disability: data from a population-based study. *Developmental Medicine and Child Neurology*, **42**, 266–70.
24. Appleyard, K., Egeland, B., van Dulmen, M., *et al.* (2005). When more is not better: the role of cumulative risk in child behavior outcomes. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **46**, 235–45.
25. Braungart, J., Fulker, D., and Plomin, R. (1992). Genetic influence of the home environment during infancy: a sibling adoption study of the HOME. *Developmental Psychology*, **28**, 1048–55.
26. Shaffer, D., Gould, M., Brasic, J., *et al.* (1983). A children's global assessment scale (CGAS). *Archives of General Psychiatry*, **40**, 1228–31.
27. Bird, H., Yager, T., Staghezza, B., *et al.* (1990). Impairment in the epidemiological measurement of childhood psychopathology in the community. *Journal of the American Academy of Child and Adolescent Psychiatry*, **29**, 796–803.
28. Goodman, R. and Scott, S. (2005). *Child psychiatry* (2nd edn). Blackwell, Oxford.
29. Angold, A., Costello, E., Farmer, E., *et al.* (1999). Impaired but undiagnosed. *Journal of the American Academy of Child and Adolescent Psychiatry*, **3**, 129–37.

9.1.2 Epidemiology of psychiatric disorder in childhood and adolescence

E. Jane Costello and Adrian Angold

Epidemiology is the study of patterns of disease in human populations.[1] Patterns are non-random distributions, and patterns of disease distribution occur in both time and space. Whenever we observe a non-random distribution, we have the opportunity to identify causal factors that influence who gets a disease and who does not. For example, we observe that depression rises rapidly after puberty in girls, but not to the same extent in boys.[2] This non-random distribution in time suggests that there may be something about puberty in girls that is causally related to depression.[3] An example of disease distribution in space can be seen in the Methods for the Epidemiology of Child and Adolescent Mental Disorders (MECA) study of five sites in the United States and Puerto Rico.[4] Although the prevalence of psychiatric disorders was fairly similar across sites, the likelihood that a psychiatric diagnosis was accompanied by significant functional impairment was much higher in children at the mainland sites than in Puerto Rico. This offers the opportunity to study between-site differences that might result in differences in the level of impairment caused by psychiatric disorders. The task of epidemiology is to understand these observed patterns in time and space, and to use this understanding as a basis for the prevention and control of disease.

Epidemiological medicine has both similarities to and differences from clinical medicine. Like clinical medicine, epidemiology is an action-oriented discipline, whose goal is intervention to prevent and control disease. Scientific knowledge about the cause and course of disease is another common goal. Epidemiology also reflects clinical medicine in using two methods of attack on disease: *tactical* methods, concerned with the practical and administrative problems of disease control at the day-to-day level, and *strategic* methods, concerned with finding out what causes disease so that new weapons of prevention and control can be engineered.[5,6] Thus, for example, in their tactical or public health role epidemiologists can be found reporting on the prevalence of adolescent drug abuse, the social burden (including cost) that drug abuse creates, and the best ways to control its spread, while others working at the strategic level might be exploring the science underlying environmental constraints on gene expression.

Epidemiology diverges from clinical medicine to the extent that it concentrates on understanding and controlling disease processes in the context of the *population at risk*, whereas the primary focus of clinical medicine is the *individual* patient. This does not mean that epidemiology is not concerned with the individual; on the contrary, it is very much concerned with understanding the individual's illness and the causes of that illness. The difference lies in the frame of reference. Put crudely, clinical medicine asks: 'What is wrong with this person *and how should I treat him or her*?' Epidemiology asks: 'What is wrong with this person and *what is it about him or her that has resulted in this illness*?' Why is this child depressed, but not her brother? If her mother is also depressed, is the child's depression a cause, a consequence, or an unrelated, chance co-occurrence? Such questions immediately set the individual child within a frame of reference of other children, or other family members, or other people of the same sex or race or social class.

Sampling, or selecting the population within which to count cases, is of central importance in epidemiology. Counting cases is an important first step towards measuring the social burden caused by a disease, and the effectiveness of prevention. For most diseases, however, simply counting the number of individuals presenting for treatment will produce estimates that are seriously biased by referral practices, ability to pay, and other factors. This is a big problem in child psychiatry because parents, teachers, and pediatricians all serve as 'gatekeepers' to treatment.[7] Community-based data are needed to measure the extent of need, and the unmet need, for prevention or treatment. Methods for assessing psychiatric disorders in the general population are discussed in another chapter. However, it is worth noting that methods for assessing disorder, whether they take the form of interviews, questionnaires, or neuropsychiatric tests, can only be as good as the taxonomy they are designed to operationalize. Current instruments mainly use scoring algorithms that turn the responses into diagnoses based on the DSM-IV or ICD-10 taxonomies. If these taxonomies do not mirror the 'reality' of psychiatric disorder then the results of using interviews or questionnaires based on them will in turn be faulty.

Estimating the burden of child and adolescent psychiatric disorders

In a world of scarce health care resources, it is important to understand the size of the burden to the community caused by these disorders. Burden, in terms of numbers affected, impact on the individual, and cost to the community, is a crucial factor in the battle for resources for treatment and prevention.

Attempts to reduce the burden of mental illness must, of necessity, pay attention to the early years. It is becoming increasingly clear that most psychiatric disorders have their onset before adulthood, and that many should be regarded as chronic or relapsing disorders. For example, the National Comorbidity Survey Replication, a representative population sample of over 9000 adults aged 18 and over in the United States,[8] found that, of the 46.4 per cent of all participants reporting one or more psychiatric disorders during there lifetime, half reported onset by age 12, and three-quarters by age 24.[9] Since we can expect a lot of forgetting of early episodes by older participants,[10] it is likely that onset in childhood is even more common than this.

If the burden of mental illness begins to be felt in childhood, it is important to know the extent of the problem so that we can begin to plan for treatment and prevention. Unfortunately, the data on which to build such estimates are very sparse. We have to rely on a national prevalence study of psychiatric disorders in the United Kingdom, and another of a large area of Brazil, together with a few national or large community surveys using symptoms scales, and a handful of diagnosis-based studies in smaller community samples, some of them longitudinal. Questionnaire-based surveys are not very useful for measuring prevalence, because they tend to define 'caseness' in terms of a certain percentage of the sample with high scores; a method that predefines prevalence.

In the past decade the United Kingdom has carried out a national prevalence study,[11,12] conducted by the Office for National Statistics, with funding from the Department of Education and other agencies. The primary purpose was to produce prevalence estimates of conduct, emotional, and hyperkinetic disorders, as well as pervasive developmental disorder, eating disorders, and tic disorders, using both ICD-10 and DSM-IV criteria. The second aim was 'to determine the *impact* or *burden* of children's mental health. *Impact* covers the consequences for the child; *burden* reflects the consequences for others'.[13] (p. 185). Third, the study measured service use. A stratified random sampling plan for England, Scotland, and Wales produced a sample of 10 438 children aged 5 to 15. Parent and child were interviewed using the Development and Well-Being Assessment (DAWBA),[14] a computer-assisted lay interview that uses a 'best-estimate' approach to diagnosis, in which responses recorded by lay interviewers are evaluated by clinicians. The first interview wave, conducted in 1999,[13] was followed by a questionnaire mailed 18 months later to all 'cases' with a diagnosis at Time 1, and a one-in-three random sample of non-cases. A second interview of all those completing questionnaires at Time 2, and all others who were cases at Time 1, was completed in 2002.[15] By weighting the responses to account for the various selection factors and for non-response, Meltzer and colleagues developed estimates of prevalence (i.e. the presence of a disorder at the Time 1 interview), of incidence (new cases between the two interviews), and of persistence.

The UK study found that almost one child in 10 (9.5 per cent) aged 5 to 15 had a psychiatric disorder based on the ICD-10 classification system. Prevalence was higher in adolescents (11.2 per cent at 11 to 15) than in children (8.2 per cent at 5 to 15), and in boys (11.4 per cent than girls 7.6 per cent). Conduct disorders were the most common (5.3 per cent), followed by anxiety disorders (3.8 per cent). Depression was rare in both sexes and all age groups

(0.9 per cent over all), as were hyperkinetic disorders (1.4 per cent). Seven per cent of previously unaffected children developed a psychiatric disorder in the 3 years between the interviews. Four per cent developed a new emotional disorder (anxiety and/or depression), and 5 per cent a behavioural and/or hyperkinetic disorder. More girls developed emotional disorders, and more boys developed behavioural disorders. Persistence, measured as the presence of the same diagnosis the years apart, was higher for behavioural disorders (43 per cent) than for emotional disorders (about one in four).

Factors affecting prevalence estimates

It is not a simple matter to compare the British prevalence rates with those from other countries, because there are few large studies, and the age ranges do not overlap. A study of youth age 7 to 14 in south-eastern Brazil, which used the same diagnostic interview but the DSM-IV taxonomy, found an overall prevalence of 12.7 per cent. Although prevalence estimates were slightly different from those reported by the UK study, the relative ordering was the same. Behavioural disorders were again the most common (7 per cent), followed by anxiety disorders (5.2 per cent) and ADHD (1.8 per cent). Once again, depression was rare (1.0 per cent). Other studies from around the world[16] usually generate prevalence rates of around 20 per cent. This puts the British and Brazilian studies at the low end of the range. However, there are many factors other than the 'true' rate of psychiatric disorder (if there is any such thing) that affect a published prevalence rate. The most important of these are:

1 *The time frame of the diagnostic measure*. Questions can be asked about symptoms occurring 'now', 'in the past month', 'in the past 3, 6, or 12 months', or 'ever'. Clearly, if recall is accurate the latter questions will elicit more symptoms than the former. Unfortunately, recall is not always accurate. Prevalence rates are higher from interviews with longer time frames, but not as much higher as would be consistent with accurate recall. For example, The National Comorbidity Study Replication, based on a nationally representative sample of adults in the United States, found that the lifetime prevalence of any disorder was 46.4 per cent, while the 12-month prevalence was 26.2 per cent. This means that 26.2 per cent /46.4 per cent = 56.5 per cent of all cases across the lifespan were present in the past 12 months. This could be explained in several ways: (i) there was an epidemic of psychiatric disorders in the 12 months before the survey; (ii) over half of all psychiatric diseases are chronic; once they occur they remain active for the rest of life; (iii) many early episodes are forgotten, and people report the onset of the most recent episode as the first occurrence of the disorder. In the absence of any evidence for (i), some combination of (ii) and (iii) seems the most likely explanation. We have evidence that the reliability with which children and adults recall the first occurrence of a symptom falls dramatically after 3 months,[17] and recommend concentrating on symptoms occurring in the past 3 months if a fairly reliable estimate is sought.

   In general, when comparing prevalence rates from different reports it is important to bear in mind the time frame. In a comparison of reported rates of child and adolescent depression published since the 1970s, we found that the time frame of the interview accounted for most of the variance, compared with taxonomy (DSM-III, DSM-IIIR, DSM-IV, ICD-9, ICD-10), diagnostic interview, or birth cohort.[2]

2 *The number and nature of the informants*. For several decades now clinicians and epidemiologists alike have recommended collecting information about a child from a range of informants: the parents, siblings, teachers, and peers, as well as the child. Most diagnostic instruments, whether questionnaires or interviews, exist in forms for diverse informants, with scoring algorithms that allow a diagnosis to be made on the basis of one informant or more. In the latter case, most follow the rule that clinicians generally observe, of counting a symptom as present if reported by any informant, rather than expecting agreement among informants, which rarely occurs.[18] Rates of psychiatric disorder will vary with the number of informants, and also depending on which informants report on which diagnoses. For example, across repeated assessments of 1420 children and adolescents, only 26 per cent of those with a diagnosis from the child interview also had one from the parent interview, and only 22 per cent of those with a diagnosis based on the parent interview had one from the child interview. This was statistically a highly significant level of agreement (OR 5.4, 95 per cent; CI 3.6, 7.9; $p < .0001$), but nevertheless only 13.5 per cent of cases were reported by both informants. Readers of epidemiological studies need to decide for themselves how much the number and type of informant matters in judging the accuracy of a prevalence estimate of a specific disorder. For example, parents often do not know much about their children's drug use, while young children themselves generally have little insight into their own hyperactivity, and teachers seldom notice children's depression. Prevalence rates based solely on these informants would be likely to be quite low.

3 *The age and sex of the subjects*. The prevalence rates of different disorders vary markedly by age, sex, and age-by-sex across childhood and adolescence. For example, a meta-analysis of 26 studies of child and adolescent depression[2] estimated the prevalence of adolescent depression (5.6 per cent) as twice that of childhood depression (2.9 per cent), and that of adolescent girls (5.9 per cent) as significantly higher than that of adolescent boys (4.6 per cent). Figure 9.1.2.1 shows prevalence rates of any psychiatric disorder (dotted lines) from a representative population sample of 1420 youth assessed regularly between ages 9 and 21. It is clear that prevalence, even when measured over time in the same subjects, varies markedly with age. This is because some of the common disorders of childhood, such as functional enuresis and encopresis, ADHD, and separation anxiety, diminish as children grow up, but then later on the problems of adolescence and young adulthood, such as drug abuse and depression, take their place. Between about 11 and 14, when the disorders of childhood have faded and those of adulthood not yet appeared, relatively few children have disorders. Prevalence rates will also differ depending on the distribution of males and females in the sample. Boys are significantly more likely to have developmental disorders, enuresis and encopresis, and ADHD in the early years, and drug abuse in the later years. Although girls are more vulnerable to depression after puberty,[19] this does not have a large effect on the overall prevalence of psychiatric disorder.

4 *The inclusion of measures of functional impairment*. Most DSM-IV and ICD-10 diagnoses require that to be clinically significant, symptoms must have a harmful effect on patients' ability to

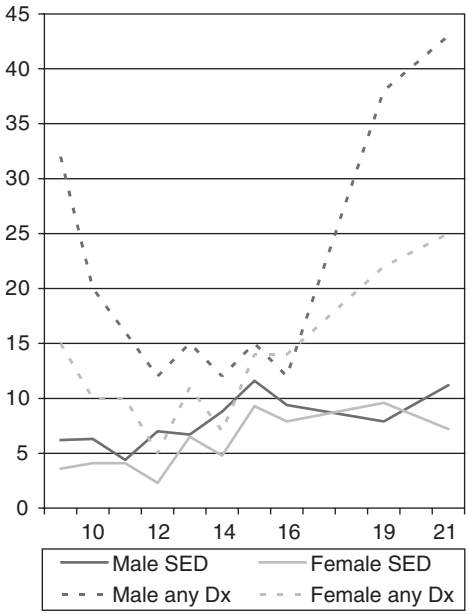


Fig. 9.1.2.1 (solid lines) shows the effect of applying a functioning criterion to diagnoses. It has the effect of flattening the U-shaped curve and revealing a doubling of psychiatric disorder with age, from around 5 per cent at age 9–10 to 10 per cent at age 21. This gradual increase is seen in both boys and girls.

function in their normal environments.[20] There is a wide range of measures of functional impairment,[20] and many diagnostic interviews employ measures of impairment as part of their diagnostic algorithms.

The inclusion or exclusion of a measure of impairment can make a dramatic difference to prevalence estimates. For example, version 2.3 of the Diagnostic Schedule for Children, a widely used interview for youth and parents, uses two measures of impairment. First, if a symptom is reported further questions are asked about whether it affects the child's functioning. Second, the interviewer scores the child on the Global Assessment Scale,[21] which rates the child's overall level of function from 0 to 100. Table 9.1.2.1 shows the impact on the prevalence of any anxiety disorder of including or excluding either or both of these measures of impairment, using data from a multi-site epidemiological study.[21] When no measure of impairment was used almost 40 per cent of subjects received an anxiety diagnosis. With both criteria applied at their most rigourous level, the prevalence of anxiety was cut to 3.2 per cent.

Future directions in the epidemiology of child and adolescent psychiatric disorders

Up to this point, the role of epidemiology has been mainly a descriptive one, addressing the basic questions: how many? who? where? when? However, child psychiatry is changing, and epidemiology will change as well. The goal is now to understand how risk exposure and vulnerability change over the life course, and how the requirements of 'normal' development shape the types of psychopathology that emerge if these requirements are not met. The term 'developmental epidemiology', first coined by Kellam in the 1970s,[22] is useful to describe what epidemiology is doing these days.

In this section we describe some rapidly growing research areas that will contribute to the next generation of studies, and will contribute to the shift from 'child psychiatric epidemiology' to 'developmental epidemiology'. We discuss the future under five headings: longitudinal research, genetic epidemiology, life course epidemiology, intergenerational epidemiology, and prevention science.

Longitudinal research

Although there have been many longitudinal developmental studies, some of them beginning at birth (or even before), longitudinal studies of psychiatric disorders had to await the development of appropriate technology; specifically, data collection methods that validly and reliably translated the psychiatric taxonomy into instruments that could be used repeatedly with the same subjects. Several of these have become available in the past 20 years.[23]

There are now several research groups that have used their longitudinal data to look at continuities and discontinuities in mental illness from childhood into adolescence and beyond. Some of the longitudinal studies have followed their subjects into adulthood.[16] These are beginning to show indications of continuity of disorder across childhood and adolescence,[24] and between temperamental characteristics in early childhood and the onset of psychiatric disorders in late adolescence and young adulthood.[25]

Genetic epidemiology

There have been two revolutions in genetic epidemiology in the past two decades that will have a tremendous impact on psychiatry in the next decade.

(a) Psychiatric–behavioural genetics

The first revolution occurred when the methods of psychiatric epidemiology were applied to behavioural genetics. Psychiatric interviews

Table 9.1.2.1 Effect of different rules for defining impairment on the per cent prevalence of any anxiety disorder (parent or child interview) using the DISC 2.3

	Diagnosis without diagnosis-specific impairment criteria				Diagnosis with diagnosis-specific impairment criteria			
	Criteria only	CGAS <= 70 (mild)	CGAS <= 60 (moderate)	CGAS <= 50 (severe)	Criteria only	CGAS <= 70 (mild)	CGAS <= 60 (moderate)	CGAS <= 50 (severe)
Any anxiety diagnosis	39.5	18.5	9.6	4.3	20.5	13.0	7.2	3.2

(Reproduced from D. Shaffer *et al.* The NIMH diagnostic interview schedule for children version 2.3 (DISC 2.3): description, acceptability, prevalence rates, and performance in the MECA study, *Journal of the American Academy of Child and Adolescent Psychiatry*, **35**, 865–77, copyright 1996, American Acadamy of Child and Adolescent Psychiatry, Lippincott Willams & Wilkins.)

DISC = Diagnostic Interview Schedule for Children. CGAS = Children's Global Assessment Scale.

like those described earlier were used in studies with genetically informative designs, such as twin, adoption, family, and migrant studies. For the first time, researchers examined categorical disorders such as depression, in ways that approximate clinical diagnosis. Furthermore, behavioural geneticists began to take seriously, issues of sampling, so that they could talk about the contribution of genes to disease in the population as a whole, rather than in highly selected families or groups. There have also been some longitudinal studies looking at how genes can have different effects at different developmental stages.[26]

(b) Molecular genetics

The second genetic revolution occurred when it became feasible to apply the methods of molecular genetics to epidemiologic samples. This development opens up the opportunity to use not only twin or adoption studies but a wide range of singleton samples to test theories about candidate genes for specific symptoms. Even more exciting is the new opportunity to use the treasure house of data from longitudinal studies to test for gene–environment interactions. Such studies can answer questions about which genes interact with which environmental factors, and at what developmental stage.[27,28]

Life course epidemiology

Life course epidemiology is the study of long-term effects on chronic disease risk, of physical and social exposures, during gestation, childhood, adolescence, young adulthood, and later adult life. It includes studies of the biological, behavioural, and psychosocial pathways that operate across an individual's life course, as well as across generations, to influence the development of chronic diseases.[29]

Life course epidemiology has developed a special concern with 'the "embodiment" of social phenomena into the biological'[30] encapsulated in the concept of 'health inequalities'. This concern arose historically from work showing that mortality from many diseases is spread unequally across the population and that these differences in risk can be linked to social inequalities that often go back to infancy or even to the parental generation. This body of work has had enormous significance for international thinking about social policy and is having a direct effect on the allocation of public resources in the United Kingdom and elsewhere.

Intergenerational epidemiology

A life course approach to epidemiology intertwines biological and social transmission of risk across generations, recognizing that geographical and secular characteristics may be unique to one cohort of individuals.[31,32]

Experiences of the previous generation can operate at many different levels of generality. They may be specific to the mother–child dyad (e.g. the effect of drug use during pregnancy), or may affect everyone living in a certain neighborhood (e.g. poverty, or exposure to an environmental toxin). All mothers and children maybe affected by a particular event, such as a period of famine or disease, or children may be affected by their mother's developmental stage (e.g. children of teen mothers or elderly mothers). Models for intergenerational research have recently appeared[33] and statistical methods have become more tractable.

Prevention science

Prevention science uses theory about the causes of disease to generate interventions, which when tested provide information not only about the effectiveness of the intervention, but also about the aetiology of the disease. Epidemiology traditionally divides prevention into thee categories, depending on the mean level of risk in the population of concern. Programmes available to all, like clean water, car seat belts, and parental leave programmes, are examples of *primary* or *universal* prevention. For example, the 'Just Say No' drug abstinence programme was introduced as a primary prevention for all children in school, designed to stop drug use before it began. Unfortunately, the results were neutral if not negative.[34] On the other hand, primary prevention with both children[35] and *families*[36] can be both effective, and suggest aetiologic pathways that could be explored in further research.

Secondary intervention *programmes* are based on high-risk children, schools, or communities. Many of them are both theory-driven and scientifically sound. A good example of a secondary intervention that yields insights for epidemiology is the 'Fast Track' programme for aggressive children in grade school. This was based on clearly articulated theory about cognitive difficulties that could interact with environmental risk to produce aggressive behaviour in socially ambiguous situations.[37] Hostile attributional bias was indeed found to be a partial mediator of the effect of the intervention on reductions in aggressive behaviour.

Once children have developed clinically defined psychiatric disorders, interventions tend at present to focus on clinical treatment rather than tertiary *prevention*. Tertiary prevention programmes are rare. One example of proven effectiveness is Multisystemic Therapy.[38] Given the early onset of most psychiatric disorders, this is clearly a vitally important area for future work.

Conclusions

This chapter has covered a lot of ground; from the first stirrings of understanding about childhood psychiatric disorders to the possibility of using molecular genetics to identify gene–environment interactions that can generate psychiatric disorder. There are fuzzy boundaries between epidemiology and developmental psychopathology, life course epidemiology, genetic epidemiology, services research, and clinical psychiatry. It will be important to keep these boundaries pervious, to share a common language where possible, and to learn and use one another's methods.

Further information

Meltzer, H., Gatward, R., Goodman, R., *et al.* (1999). *The mental health of children and adolescents in Great Britain*. Office for National Statistics, London.

Meltzer, H., Gatward, R., Corbin, T., *et al.* (2003). *Persistence, onset, risk factors and outcomes of childhood mental disorders*. Office for National Statistics, London.

References

1. Kleinbaum, D.G., Kupper, L.L., and Morgenstern, H. (1982). *Epidemiologic research: principles and quantitative methods*. Van Nostrand Reinhold, New York.

2. Costello, E.J., Erkanli, A., and Angold, A. (2006). Is there an epidemic of child or adolescent depression? *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **47**, 1263–71.
3. Angold, A., Costello, E.J., and Worthman, C.M. (1998). Puberty and depression: The roles of age, pubertal status, and pubertal timing. *Psychological Medicine*, **28**, 51–61.
4. Shaffer, D., Fisher, P.W., Dulcan, M., *et al.* (1996). The NIMH diagnostic interview schedule for children (disc 2.3): Description, acceptability, prevalences, and performance in the meca study. *Journal of the American Academy of Child and Adolescent Psychiatry*, **35**(7), 865–77.
5. Earls, F. (1980). Prevalence of behavior problems in 3-year-old children: a cross- national replication. *Archives of General Psychiatry*, **37**, 1153–7.
6. Susser, M. (1973). *Causal thinking in the health sciences: concepts and strategies in epidemiology*. Oxford University Press, New York.
7. Horwitz, S.M., Leaf, P.J., and Leventhal, J.M. (1998). Identification of psychosocial problems in pediatric primary care. *Archives of Pediatrics & Adolescent Medicine*, **152**, 367–71.
8. Kessler, R.C., Berglund, P., Demler, O., *et al.* (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, **62**(6), 593–602.
9. Insel, T.R. and Fenton, W.S. (2005). Psychiatric epidemiology: it's not just about counting anymore. *Archives of General Psychiatry*, **62**, 590–2.
10. Giuffra, L.A. and Risch, N. (1994). Diminished recall and the cohort effect of major depression: a simulation study. *Psychological Medicine*, **24**, 375–83.
11. Ford, T., Goodman, R., and Meltzer, H. (2003). The British child and adolescent mental health survey 1999: the prevalence of DSM-IV disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, **42**, 1203–11.
12. Meltzer, H., Gatward, R., Goodman, R., *et al.* (2003). Mental health of children and adolescents in Great Britain. *International Review of Psychiatry*, **15**(1–2), 185–7.
13. Meltzer, H., Gatward R., Goodman, R., *et al.* (1999). *The mental health of children and adolescents in Great Britain*. Office for National Statistics, London.
14. Goodman, R., Ford, T., Richards, H., *et al.* (2000). The development and well-being assessment: Description and initial validation of an integrated assessment of child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, **41**, 645–56.
15. Meltzer, H., Gatward, R., Corbin, T., *et al.* (2003). *Persistence, onset, risk factors and outcomes of childhood mental disorders*. Office for National Statistics, London.
16. Costello, E.J. and Angold, A. (2006). Developmental epidemiology. In *Theory and method* (eds. D. Cicchetti and D. Cohen), pp. 41–75. Wiley, Hoboken.
17. Angold, A., Erkanli, A., Costello, E.J., *et al.* (1996). Precision, reliability and accuracy in the dating of symptom onsets in child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, **37**, 657–64.
18. Achenbach, T.M., McConaughy, S.H., and Howell, C.T. (1987). Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, **101**, 213–32.
19. Angold, A. and Costello, E.J. (2006). Puberty and depression. In *Child and adolescent psychiatric clinics of north America* (eds. G. Zalsman and D.A. Brent), pp. 919–37. Saunders, Philadelphia.
20. Canino, G., Costello, E.J., and Angold, A. (1999). Assessing functional impairment and social adaptation for child mental health services research: a review of measures. *Journal of Mental Health Services Research*, **1**, 93–108.
21. Shaffer, D., Gould, M.S., Brasic, J., *et al.* (1983). A children's global assessment scale (CGAS). *Archives of General Psychiatry*, **40**, 1228–31.
22. Kellam, S.G., Ensminger, M.E., and Turner, R.J. (1977). Family structure and the mental health of children. *Archives of General Psychiatry*, **34**, 1012–22.
23. Angold, A. and Fisher, P.W. (1999). Interviewer-based interviews. In *Diagnostic assessment in child and adolescent psychopathology* (eds. D. Shaffer, C. Lucas, and J. Richters), pp. 34–64. Guilford Press, New York.
24. Costello, E.J., Mustillo, S., Erkanli, A., *et al.* (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry*, **60**, 837–44.
25. Caspi, A., Henry, B., McGee, R.O., *et al.* (1995). Temperamental origins of child and adolescent behavior problems: From age three to fifteen. *Child Development*, **66**(1), 55–68.
26. Rutter, M. (2002). The interplay of nature, nuture, and developmental influences. *Archive of General Psychiatry*, **59**, 996–1000.
27. Caspi, A., McClay, J., Moffitt, T.E., *et al.* (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, **297**, 851–4.
28. Foley, D.L., *et al.* (2004). Childhood adversity, monoamine oxidase A genotype, and risk for conduct disorder. *Archives of General Psychiatry*, **61**, 738–44.
29. Ben-Shlomo, Y. and Kuh, D. (2002). A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology*, **31**, 285–93.
30. Krieger, N. (2001). Theories for social epidemiology in the 21st century: an ecosocial perspective. *International Journal of Epidemiology*, **30**, 668–77.
31. Cairns, R.B., Elder, G.H., and Costello, E.J. (1996). *Developmental science*, Vol. 20. Cambridge University Press, New York.
32. Stein, Z., Susser, M., Saenger, G., *et al.* (1975). *Famine and human development: the dutch hunger winter of 1944–45*. Oxford University Press, New York.
33. Rossi, A.S. (1989). A life-course approach to gender, aging, and intergenerational relations. In *Social structure and aging* (eds. K.W. Schaie and C. Schoder), pp. 207–36. Erlbaum, Hillsdale, NJ.
34. Lynam, D.R., Milich, R., Zimmerman, R., *et al.* (1999). Project dare: No effects at 10-year follow-up. *Journal of Consulting and Clinical Psychology*, **67**(4), 590–3.
35. Kellam, S.G., Koretz, D., and Moscicki, E. (1999). Core elements of developmental epidemiologically based prevention research. *American Journal of Community Psychology*, **27**, 463–82.
36. Costello, E.J., Compton, S.N., Keeler, G., *et al.* (2003). Relationships between poverty and psychopathology: a natural experiment. *The Journal of the American Medical Association*, **290**, 2023–9.
37. Dodge, K.A., Pettit, G.S., and Bates, J.E. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development*, **65**, 649–65.
38. Henggeler, S.W., Schoenwald, S.K., Pickrel, *et al.* (1994). *Treatment manual for family preservation using multisystemic therapy*. South Carolina Health and Human Services Finance Commission, Columbia, SC.

9.1.3 Assessment in child and adolescent psychiatry

Jeff Bostic and Andrés Martin

The goals of assessment of a child/adolescent are to (1) detect psychopathology and its impacts on the child's functioning in family, school, and peer domains, (2) allow appropriate intervention targets to be identified and prioritized; and (3) identify relevant variables, including family or school factors that may influence treatment adherence.

Distinctive aspects of the psychiatric assessment in children

1 Parents (or other adults) ordinarily initiate and pursue the evaluation of the child for diverse reasons. Adult expectations for the child sometimes exceed the child's abilities, or the adult's own parenting or teaching style may be a poor fit with this child. Some adults may seek treatment to alter the child to remedy this poor fit.

2 Children may not be receptive to changing their behaviour. Children may attribute problems to others and be unable to accept their contribution to an identified problem. The psychiatric assessment of children requires attention to what the child wishes would change.

3 Young children may not trust unfamiliar adults (including clinicians), and adolescents may perceive the clinician as another adult imposing expectations or judgements. Multiple informants[1] are often needed to identify the child's functioning in school, home, and peer domains, to identify the child's areas of strength on which the clinician can build, and to identify others (peers or adults) able to introduce or reinforce more adaptive skills or behaviours.

4 Most DSM-IV-TR diagnoses were defined amongst adult samples.[2] Efforts to consider where a particular child fits on the depressed mood, anxiety, and aggression axes, for example, requires attention to developmental differences in symptom expression.

5 The ability of the clinician to forge alliances with the child, the parent, and outside entities is essential. A breach in any of these relationships can impede treatment. Parental permission should be obtained to contact and collaborate with relevant parties.

Content of the clinical interview

Reason for referral

Who initiated this referral, their motivations, and what changes they seek is vital. Expectations of various parties may collide and must be reconciled for effective treatments to be implemented. For example, the school may seek changes in parental discipline, while parents may expect the evaluation to yield additional school services.

History of problem(s)

Parents often experience intense pain while recounting the deterioration or anguish of their child. Clinicians should provide parents an opportunity to describe the evolution of the problem, attending to the context in which symptoms emerged and occur, changes in frequency and intensity of symptoms, and their current progression. The clinician should inquire directly about the *functions* of problem behaviours, including secondary gains (e.g. tantrums diminish chore requirements, etc.). The clinician should clarify whether symptoms are specific to one functional domain or whether they pervade multiple areas of the child's functioning at home, school, and with peers.

Past problems

Significant past symptoms impairing the child should be identified. It is especially important to understand whether symptoms have been persistent since early childhood, are intermittent, or represent deterioration from a previously better level of functioning.

Comorbid problems

Clinicians should inquire about disorders often seen in tandem. For example, bipolar disorder in children is often associated with previous attention deficit hyperactivity disorder.[3] Screening instruments (such as those selectively available free of cost at websites such as www.schoolpsychiatry.org) can be useful to provide comprehensive information about less conspicuous symptoms.

Substance use history

Clinicians should inquire about the child's exposure to and use of tobacco, alcohol, and illicit substances. Children may perceive that substances alleviate their distress (e.g. anxiety, depression) and 'self-medicate.' Clarifying impacts of substances on symptoms may yield intervention points attractive to the child.

Previous treatment(s)

Chronological assessment of past treatments may reveal strategies adaptable to the current problem. Past treatment history may suggest treatment modalities (in)tolerable to this patient (and family). Medication trials, counselling, hospitalizations, or alternative treatments should be explored.

Developmental history

Parents may vary in their recollection of their child's attainment of developmental milestones. Review of earlier videotapes of the child may improve the reliability and completeness of reports regarding the sequence of the child's growth.

The child's development regulating *sleep, eating,* and *toileting* should be investigated. Attained skills may suddenly be lost, sometimes signalling the importance of emotional events at particular times. Eating behaviour has become complicated as both hunger and obesity increase risks of psychopathology.[4,5]

Psychomotor development includes standing, walking, running, throwing, hopping, and playing sports or musical instruments. How the child fares at sports may clarify psychomotor skills. Fine motor and gross motor skills may not be congruent.

Cognitive development refers to the child's acquisition of thinking skills. Specific inquiry concerning speech development, reading, writing, and math skill progression may reveal global or specific difficulties.

Interpersonal development refers to how the child interacts with others, particularly family members and other children and adults.

Stability of relationships, numbers of friends, types of activities shared, and expectations of peers often reveal sources of difficulty or maladaptive patterns.

Emotional development and *temperament* reveal the child's capacity to recognize his or her own mood state and to self-soothe or regulate negative affect. Prevailing moods can be described by parents, who may also detail past suicidality, irritability, specific fears and anxieties, and conditions associated with the child's happiness and pleasure.

The child's *moral development* indicates whether conscience or moral values are too lax, too harsh, overly focused on particular areas, or uneven and out of proportion to daily events. The child's ability to recognize impacts of decisions on others, and to acknowledge and correct mistakes provides clarity about the child's strengths and limitations. The child's religious and cultural/ethical views and practices also shape this area, and may guide treatment interventions.

Trauma may impact or even arrest development. Investigation of actual events (such as documented abuse), but also of events perceived traumatic by the child and family may shed important light on the child's behaviours and patterns of relating to others. Events surrounding the trauma, disclosures to others, and reactions of adults are also important for the clinician to recognize and address.

Harmful behaviour, towards self or others, may reveal important developmental progressions that warrant intervention. Head-banging may reveal sensory disturbances, thoughts or comments about death may reveal suicidality, and self-harmful acts such as self-mutilation or cutting may reveal primitive coping mechanisms.[6] Harmful acts towards animals or people may indicate needs for monitoring while other diagnostic or treatment interventions occur.

Family history

Few psychiatric disorders appear transmitted exclusively genetically. Many parents fear that their other child may be destined to suffer psychopathology when a family member manifests a disorder, so clarification of contributions to expression of disorders can reduce unwarranted fear, guilt, and distress. Please refer to Chapter 6.3.8 for more information on assessment of family functioning.

Divorce, separation, and single-parent family circumstances may stress all family members. Even when parents part amicably, children may attempt to reunite family members. Children may exhibit symptoms even years after separations as they enter different developmental phases.[7]

Adoption may be a positive event for the child, and adoption warrant tactful attention by the clinician, including age at adoption of the child and biological parents, the involvement with biological parents, the child's understanding of the adoption, and how the adoption is discussed at home.

Medical history

Pregnancy complications, birth difficulties, hospital stays, and medical illnesses requiring treatments (e.g. asthma, diabetes) should be investigated, as they increase the child's risk for psychopathology.[8] Inquiry into emergency room visits or surgeries can shed light on the child's fears, or parental over/underprotectiveness. Allergies should be ascertained, as well as responses or side effects to medications, including naturopathic or homeopathic agents.

Child strengths/weaknesses

Interests, hobbies, and talents of the child should be obtained from the child and parents. Parents may have aspirations the child does not share, or the child may have fantasies beyond apparent abilities. In most cases, though, the child will have some identifiable interests or abilities that serve as potential points of connection with peers and adults (including clinicians).

The child's media diet

Children are exposed to television, music, videos, electronic games, cell phones, e-mail and instant messaging, personal digital assistants, etc. It is important to clarify which media the child uses, how much time each day is spent with these various media, and what consequences these media have on the child (e.g. in response to watching action TV show the child becomes more violent versus has developed interest in Asian food through watching cooking programmes).[9] The degree of parental awareness and appropriate limit-setting regarding TV, video games, and instant messaging may warrant intervention.

Mental status examination (MSE)

The MSE must be adjusted for children (see Fig. 9.1.3.1). The MSE includes a clinical description of the child's appearance, mood, sensorium, intelligence, and thought content and process. Much of the MSE takes place implicitly as the clinician interacts and observes the child during the child and family interviews.

Structure of the clinical interview

Preparatory phase of the child interview

Unlike regular pediatric check-ups, the psychiatric evaluation usually occurs because of prominent symptoms often perceived as embarrassing by the parents or the child. A phone call before the interview by the clinician or staff can clarify the structure of the interview, the collaboration anticipated to devise solutions, and the opportunity for parents to provide any confidential information to the clinician.

The parent interview

The parent interview can be complicated by parental ambivalence about having a child evaluated by a psychiatrist, fears of loss of control or criticism, or parental shame or embarrassment about perceived parenting faults. The clinician should remain sensitive throughout the interview to parent vulnerabilities. Techniques to help parents overcome such obstacles during the interview are summarized in Fig. 9.1.3.2.

The developmentally sensitive clinical interview of the child

The interview process and wording of questions must be tailored to fit with the child's understanding.[10] The child may not understand terms necessary to answer questions accurately. The child may also provide misleading answers to shield other family members, to protect against acknowledging some perceived failing, or to address circumstances if the child fears it might entail placement out of the home. Please refer to Chapter 9.1.1 for more specific information for obtaining reliable information during the child interview.

Category	Components	What to Assess
Appearance	Physical Appearance	Gender; ethnicity; age (actual and apparent); cleanliness and grooming, hair/clothing style, presence of physical anomalies, indicators of self-care and parental attentiveness
	Manner of Relating to Clinician and Parents	Ease of separation from parent, guardedness, defiance, eagerness to please, flirtatiousness
	Activity Level	Psychomotor retarded to agitated, sustained or episodic, goal-oriented or erratic; coordination, unusual postures or motor patterns (e.g., tics, stereotypies, compulsions, catatonia, akathisia, dystonia, tremors)
	Speech	Fluency (including stuttering, cluttering, speech impediments), rate, volume, prosody
Mood	Current Affect	Predominant emotion and range (constricted to labile) during the interview, and appropriateness to content (e.g., giggles while talks about sibling's illness); intensity; lability
	Persisting Mood	Predominant emotion over days/weeks; whether current affect unusual or consistent with mood; whether mood reactive to situations or same across range of situations
	Coping Mechanisms and Regulation of Affect	How child manages conflict or distress, age-appropriateness of responses to and dependency on parents; sexual interests, impulses, aggression; control or modulation of urges (finding alternative or socially appropriate means of satisfying urges); how deals with frustration or when anxious
Sensorium	Orientation	Self (name), place (town, State), time (awareness of morning, day of week, month, year varies by age), situation (why at this appointment)
Intellect/ Cognition	Attention	Need for repeating, how long sustained on activity, degree to which child shifts from activity to activity, distractibility (to outside noises, etc.)
	Memory	Immediate (repeat numbers, names back), short-term (recall 3 objects at 2 and 5 minutes), long-term (recall events of past week)
	Intelligence; Fund of Knowledge	Age-appropriate recognition of letters, vocabulary, reading, counting, computational skills; age-appropriate knowledge of geography, history, culture (celebrities, sports, movies, etc.); concrete to abstract thinking, ability to classify and categorize
	Judgment	Best assessed after rapport established, as initially minimization or denial more common); what would do if found stamped envelope next to mailbox, fire started in theater, say if saw man with big feet
	Insight	Ability to see alternative explanations, others' points of view; locus of control (internal v. external); defense mechanisms
Thought	Process: *Coherence*	Logical, goal-directed, circumstantial or tangential (consider age-appropriateness), looseness of associations, word salad (incoherent, clanging, neologisms)
	Process: *Speed*	Mutism, poverty of thought (long latency, thought blocking), poverty of content (perseveration), racing thoughts, flight of ideas
	Perceptions	Altered bodily experiences (depersonalization, derealization), misperception of stimulus (illusion), no stimulus (hallucination: auditory [psychosis > PTSD > organic causes], visual [substance use, delirium], olfactory (neurological, seizure disorder] gustatory [from medicine side effects])
	Content	Obsessions (ego-dystonic), delusions (ego-syntonic), thoughts of harm to self or others (magical thinking, or fears at night often age appropriate)

Fig. 9.1.3.1 The mental status examination in children.

1 Forming a clinical alliance with parents
- (a) Facilitating Narrative History
  Open-ended questions allow parents control, and can be followed with narrow questions to fill in needed details. Using the parent's own words can help parents feel heard.
- (b) Finding Common Themes/Patterns
  Inquiry into problems or conflicts the child has with other adults, peers, or unfamiliar others may illuminate patterns of the child's behavior that play out in a variety of settings, decreasing parents' anxiety that they alone provoke the child's problem.
- (c) Finding Good Intentions Gone Awry
  Parents may feel ashamed of past parenting efforts done in desperation.
  Acknowledging the parent's good intention leading to a misguided effort can diminish self-reproach. For example, a parent's harsh response often belies a fear about the child's future behavior, so identifying the fear and then examining alternative responses can be productive).
- (d) Partnering with Parents (Clinician as "partner" in decision-making process)
  Clinicians increasingly serve as partners, outlining several appropriate treatments, risks, and side effects, and helping parents to choose and invest in preferred treatments. If parents propose treatments the clinician regards as unhealthy or unproven, the clinician can identify potential risks of such treatments to minimize risks to the child.
- (e) Clarifying Expectations of the Evaluation
  Parents sometimes have unrealistic fantasies about what the evaluation will accomplish. Inquiring early about what the parent hopes will be accomplished by this evaluation can reveal such expectations and fantasies, which the clinician can realistically address. For example, parents may believe the evaluation can definitively prove the child had been abused by someone. At the other extreme, parents may fear that the clinician will tell them that their child will never be normal, will require institutionalization, or ultimately harm others.

2 Eliciting Sensitive Information
- (a) Providing the Parent Opportunities to Convey Sensitive Information
  Apprising parents of times and methods to convey information can provide appropriate mechanisms for sharing of information.
- (b) Revisiting Sensitive Information at Safer Points
  If parents resist disclosing information, the clinician should not force answers (as they are more likely to be inaccurate or incomplete), but rather proceed to less distressing information.
- (c) Explaining the Purpose of Sensitive Information
  Some parents may need to understand the underlying reasons for inquiring about personal information. For example, the clinician may need to explain the need to inquire about relatives to clarify genetic contributions to the child's difficulties.
- (d) Describing How Sensitive Information Will Be Reported
  Parents are sometimes fearful that details of embarrassing past parental personal problems may be included in reports to be seen by others. Parents may fear that marital conflict information might be used to alter custody arrangements, or symptoms in a report that could jeopardize their child's future educational or occupational pursuits. Clarifying that general information will be provided ("history of substance abuse on maternal side") rather than specifics and that parents will be able to review reports whose release they authorize can diminish resistance to sharing sensitive information.

3 Handling Discrepant Reports
- (a) Contextualizing Points of View
  Differences between observers' descriptions of a child's behavior have several potential sources. For example, teachers sometimes report very different presentations than parents. Examining what precipitates the child's problem, and how it expresses itself in different environments may allow clinicians to borrow effective strategies across environments without "blaming" adults.
- (b) Aligning Different Perspectives
  When parents or adults exhibit conflict during the psychiatric evaluation, the clinician may continue to refocus adults to the child's needs. For example, the clinician may encourage "middle ground" approaches to increase consistency between environments.

Fig. 9.1.3.2 Parent interview techniques.

The child's understanding of the psychiatric interview

The child and parent are usually seen together at the beginning of the child psychiatric interview to put the child at ease. Once comfortable, the child usually can tolerate the parents leaving the room. Transitional objects (books, electronic devices from home) may ease these transitions. Inquiring about what the child believes parents, teachers, or other adults want to be different as a result of this interview often elucidates what the child recognizes about others' perspectives, and also facilitates the child projecting thoughts or fantasies about this evaluation.

Adolescents sometimes fear parents will skew the interview by telling their 'version' first to get the clinician to side against the adolescent.[11] Meeting briefly with the parent and adolescent to clarify objectives, and then meeting with the adolescent alone at length may enhance an alliance with an adolescent. During this initial segment the clinician can clarify the plan to meet alone with parents after meeting with the adolescent to review birth history, developmental milestones, and family.

Adolescents may resist answering questions or participating. Clinicians can identify the adolescent's priorities and side with those that are reasonable, or identify what the adolescent needs to do to satisfy parents so that the adolescent no longer needs to see a psychiatrist. Clinicians may also decrease resistance by inquiring first about the adolescent's interests, strengths, musical preferences,

rather than focusing on their 'problems,' as adolescents are developmentally struggling with their identity, and may resist fitting into the 'psychiatric patient' category.

Developmentally sensitive techniques for the psychiatric interview

Four categories of techniques are commonly employed in these interviews. *Engagement* techniques are often required to put the child at ease so that the child will provide accurate clinical information. *Projective* techniques allow the child to reveal underlying themes or issues which cannot be verbalized directly. *Direct questioning* techniques clarify particular points needed to distinguish disorders, contributions to the child's problems, and intervention options. *Interactive* techniques clarify how the child relates to, as well as accepts or integrates input from, others.

Techniques to engage the child

Child psychiatrists often provide toys or objects for patients in the waiting room and office. Toy figures, puppets, and 'relationship-oriented' toys may ease the child into the interview. Generic toy figures are usually preferable, since they are more likely to evoke the child's specific themes and concerns rather than 'scripts' based on TV shows or movies. Tasks framed as 'games' or active (e.g. drawing a house or family) often help the child transition into the psychiatric interview. By allowing the child to direct the content, the interviewer can follow the sequence of the child's concerns, note themes that emerge, and observe the points at which a child avoids or shifts to a new topic.

With *adolescents*, efforts to indicate familiarity with contemporary adolescent tastes (music, movies, terms, etc.) can be perceived ingenuine by the adolescent. Instead, clinicians may inquire about current interests, musical preferences, and current adolescent values from a curious, 'help me understand it' perspective, rather than from one of 'trying to be hip.' Manipulable items (squeeze balls, modelling clay, finger cuffs, cards, etc.) may allow adolescents a socially acceptable option for keeping their hands busy so that the interview feels less like an interrogation.

Projective techniques

Projective techniques may help the child express concerns indirectly, so that anxiety about significant fears, telling family secrets, or betraying loyalties is minimized. Common projective techniques include having the child draw a picture of him- or herself or family doing something. For pictures of the child, body details including sizes of appendages or body parts and articulation (fingers, toes), relative size of the figure to the page, and frequent erasures can all reveal underlying issues of anxiety, perceived agency to address difficulties, or needs to control the environment. Depictions of the self as non-human, grotesque, imbued with super powers, or of the opposite gender may provide clues about the child's self-image and underlying wishes. The relative size and placement or omission of family members in a family drawing may illuminate the child's feelings about family relationships. Aggressive or sexual themes may be revealed in drawings.

Verbal projective techniques can similarly yield important information. Asking what animal or character (TV/movie star, cartoon, superhero) the child would most like to be, or whom the child would take along to a deserted island, or asking what the child would do with three magic wishes often allow underlying issues to emerge. Wishes may reveal basic needs, such as food or a safe place to live, or longings for parents to reunite or for the return of a departed friend. Wishes sometimes reveal specific desires, such as 'not to have tics anymore,' or 'never to get teased.' Very general or altruistic wishes, such as 'world peace' or 'to live in a big house with lots of money' warrant further exploration, such as 'Are there particular fights you would especially like to stop?' or 'Who else would live there?' and 'What would you do first with lots of money?'

Projective techniques may help *adolescents* to reveal and share emotionally significant concerns with the clinician. Inquiries into favourite, or most disliked, movies, television characters, political or historical figures, musicians or artists, or sports figures, all allow elaboration of the teenager's ideas in displacement. Adolescents less distrustful of the clinician may readily speak about their own social longings or anxieties regarding friends at school. Adolescent resistances are often revealed by reluctance to divulge names of friends, or even questions about why the clinician needs to know this information. If resistance is detected, questions about what the adolescent most admires about a character, or what the adolescent imagines this character would do in given situations may reveal the adolescent's perceptions. Asking the adolescent about the different cliques or groups at school and his or her relationship to them provides useful information about the teen's self-image. Similarly, questions about what the adolescent sees as fair or would most like to change about school or the world often reveals underlying concerns and issues.

Direct questioning

Direct questioning can specify symptoms or events, clarify how the child sees the world and functions within it, and follow-up on material from other parts of the evaluation. Asking the child to describe friends ('Tell me about your best friend.'), siblings, or parents, is preferable to 'Do you get along with your brother?' Open-ended questions such as 'What sorts of things make you mad/afraid/happy?' and 'What do you daydream about?' are similarly preferable to 'Do you get mad?' or 'Do you ever daydream?'

Anchoring direct questions to major events may help children provide more accurate answers. For example, 'Did that happen before or after your birthday?' or 'How has that (problem) been since school ended?' improve respondent accuracy.

Substance abuse, sexuality, and risky behaviours are often assessed through direct questions. The clinician can use simple questions, such as 'Substance use?' that allow significant latitude, and then focus in further, contingent on the child's responses. For example, the clinician may hear 'No, I don't do any of that anymore,' which could then be followed by 'What led to that decision?' and then proceed back to when and what substances were used. Similarly, sexuality can be assessed by gentle direct questions that do not prematurely close off response options, such as 'Have you had romantic feelings towards another? How did that go?' (rather than 'Have you had a boyfriend yet?'). Adolescents may fear the interviewer will be disapproving, so questions like 'romantic feelings towards anyone' are preferable to 'are there any girls you like?'[12] Finally, direct inquiries into risky behaviours (stealing, vandalism, assaults, gambling, etc.) often require general questions such as 'Have you done anything that you now look back on and think was dangerous?' before proceeding to specific questions (e.g. 'Have you ever stolen anything? Have you ever been beaten up?

Beat up someone else?'). Suicidal risk behaviours may be minimized or trivialized, so additional questions to examine fantasies about impacts of the suicide on family and friends or value contradictions may be needed to clarify suicidality risks.[13]

Interactive techniques

Throughout, the clinician observes how the child relates to another person and what feelings or reactions this child elicits. How the child reacts to a new person, sustains interactions, and terminates the interview often reveal patterns important in the child's larger social life. The clinician can evaluate more complex social interactions during transitions ('It's time to put these toys up in the box.') and during games. Short games (tic tac toe) are useful since the clinician can quickly detect the child's response to winning, tying, and losing.

Adolescents employ more complex patterns, often specific to a subgroup to which they now belong, so clarifying what clothing symbols represent, meanings of confusing terms, and values espoused by subgroups can clarify how the adolescent relates to others. The clinician should observe provocative comments, often used to titrate space between the clinician and the adolescent, or to reject others first.

Concluding the interview

Ending collaboratively increases the likelihood that the child will feel positive about subsequent encounters with clinicians, including treatment. Questions such as 'Are there other things that would be important for me to know about what you're like or how things have been for you?' or 'What else have I not asked about that is important?' facilitate this process.

The child may be curious about what the clinician will say and to whom. The interview is one piece of a larger evaluation, so the clinician may need to clarify that other testing, conversations with others, or additional meetings may be needed. Discussing findings (including treatment recommendations) with parents is usually advisable since parents may disagree with the clinician's conclusions or resist suggested interventions (e.g. medication, school placement, etc.).

Confidentiality is one of the most challenging issues surrounding child psychiatric interviews, especially with adolescents. Describing to the adolescent what will be told to specific others is helpful, as well as what information will not be revealed (e.g. specific details about substance abuse or sexual behaviours). Parents and the child should be told explicitly that confidentiality does *not* extend to situations that pose a clear danger to the child or others. In cases where dangerous content emerges (e.g. the child describes obtaining bullets to frighten a peer), the clinician should clarify with the child *how* they will tell appropriate others, preferably together.

Neuropsychological testing

Patients may have subtle or complicated difficulties processing certain types of information. Consultation with a pediatric neuropsychologist may clarify appropriate tests to address persisting diagnostic questions. Clinicians should recognize that young patients may not be 'interested' in testing tasks, so scores should be interpreted cautiously, with input from the person who did the testing, when the clinician discusses findings with families.

Laboratory evaluation in the child psychiatric evaluation

Few definitive clinical tests identify specific child psychiatric disorders. Laboratory testing remains useful when symptoms and physical findings suggest a particular disorder. Collaboration with the primary pediatric care provider may guide decisions about possible further medical consultations (e.g. audiometric, genetic, neurological, speech, etc.) or diagnostic tests (e.g. blood tests, neuroimaging, sleep studies).

Testing in specific childhood disorders

Laboratory testing yields findings that alter the working diagnosis in approximately 1 per cent of cases, and the yield for laboratory abnormalities, without the presence of other supportive physical findings, remains less than 5 per cent.[14] Laboratory tests commonly considered are summarized in Table 9.1.3.1. Specialized technologies, such as positron emission tomography (PET), single photon emission computerized tomography (SPECT), functional

Table 9.1.3.1 Laboratory tests to consider in childhood psychiatric disorders

Lab test	Disorder					
	MR/PDD	Mood	Psychosis	OCD tics	Substance abuse	Eating disorders
Chromosomal testing	X		X			
Wood's (UV) lamp	X					
Monospot		X				
Thyroid		X	X			X
Lyme titre		X				
CBC	X	X	X		X	X
Serum chemistry	X	X	X		X	X
Lead level	X					
Throat culture antistreptolysin O antibody (ASO), antideoxyribonuclease B titres				X		
Urine drug screen		X	X		X	
Cerebrospinal fluid analysis		X	X			
Neuroimaging			X			
EEG	X					

MRI (fMRI), and brain electrical activity mapping (BEAM) remain attractive research tools at this time in child psychiatry.

Further information

Recommended Websites:

www.aacap.org: the home site for American child psychiatry; includes current practice parameters for various psychiatric disorders.

www.schoolpsychiatry.org: rating scales, school interventions for psychiatric symptoms.

Baron, I.S. (2004). *Neuropsychological evaluation of the child*. Oxford, New York.

Martin, A. and Volkmar, F.R. (eds.) (2007). *Lewis's child and adolescent psychiatry: a comprehensive textbook* (4th edn). Lippincott, Williams, & Wilkins, Philadelphia.

Mash, E.J. and Barkley, R.A. (eds.) (2007). *Assessment of childhood disorders* (4th edn). Guilford, New York.

Rutter, M. and Taylor, E.A. (eds.) (2002). *Child and adolescent psychiatry* (4th edn). Blackwell, Oxford.

References

1. Ferdinand, R.F., Hoogerheide, K.N., van der Ende, J., *et al.* (2003). The role of the clinician: three-year predictive value of parents', teachers' and clinicians' judgment of childhood psychopathology. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **44**(6), 867–76.
2. Patel, V., Flisher, A.J., Hetrick, S., *et al.* (2007). Mental health of young people: a global public-health challenge. *Lancet*, **369**(9569), 1302–13.
3. Masi, G.P.G., Millepiedi, S., Mucci, M., *et al.* (2006). Developmental differences according to age at onset in juvenile bipolar disorder. *Journal of Child and Adolescent Psychopharmacology*, **16**(6), 679–85.
4. Weinreb, L., Wehler, C., Perloff, J., *et al.* (2002). Hunger: its impact on children's health and mental health. *Pediatrics*, **110**(4), e41.
5. Vila, G., Zipper, E., Dabbas, M., *et al.* (2004). Mental disorders in obese children and adolescents. *Psychosomatic Medicine*, **66**(3), 387–94.
6. King, R.A., Ruchkin, V.V., and Schwab-Stone, M. (2003). Suicide and the continuum of adolescent self destructiveness: is there a connection? In *Suicide in children and adolescents* (eds. R.A. King and A. Apter). Cambridge University Press, Cambridge.
7. Wallerstein, J.S. and Blakeslee, S. (2000). *The unexpected legacy of divorce: a 25 year landmark study*. Hyperion, New York.
8. Indredavik, M.S., Vik, T., Heyerdahl, S., *et al.* (2004). Psychiatric symptoms and disorders in adolescents with low birth weight. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, **89**(5), F445–50.
9. Pataki, C., Bostic, J.Q., and Schlozman S. (2005). The functional assessment of media in child and adolescent psychiatric treatment. *Child and Adolescent Psychiatric Clinics of North America*, **14**(3), 555–70.
10. Lewis, M. (ed.) (1991). *Psychiatric assessment of infants, children, and adolescents*. Williams and Wilkins, Baltimore.
11. King, R.A. and Schowalter, J.E. (2004). The clinical interview of the adolescent. In *Textbook of child and adolescent psychiatry* (3rd edn) (eds. J.A. Wiener and M.K. Dulcan), pp. 113–16. American Psychiatric Publishing, Washington, DC.
12. King, R.A. (1997). Practice parameters for the psychiatric assessment of children and adolescents. American Academy of Child and Adolescent Psychiatry. *Journal of the American Academy of Child and Adolescent Psychiatry*, **36**(Suppl. 10), 4S–20S.
13. Galvin, M.R., Fletcher, J., and Stilwell, B.M. (2006). Assessing the meaning of suicidal risk behavior in adolescents: three exercises for clinicians. *Journal of the American Academy of Child and Adolescent Psychiatry*, **45**(6), 745–8.
14. Challman, T.D., Barbaresi, W.J., Katusic, S.K., *et al.* (2003). The yield of the medical evaluation of children with pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, **33**(2), 187–92.

9.1.4 Prevention of mental disorder in childhood and other public health issues

Rhoshel Lenroot

Introduction

Over the last two decades advances in psychiatric classification systems and screening tools have allowed the global and national burden of mental disorder to be described with the first large-scale epidemiologic studies. The World Health Organization's *World Health Report 2001* estimated that over 450 million individuals suffer from mental disorders, and that psychiatric disorders ranked as 5 of the top 10 causes of disability in the global population.[1] Studies specifically of psychiatric disorders in children report that between 3 per cent and 18 per cent of children have a clinically significant psychiatric disorder, a number far exceeding those with access to treatment.[2] A recent study which included data on age of onset found that 50 per cent of psychiatric disorders had their onset by age 14, and 75 per cent by age 24.[3] Treatment on this scale is unlikely to ever be feasible, even if available methods were more effective and less risky than those currently available. Preventing mental health disorders from occurring is an alternative to decrease the extent of this public health problem. However, if a key characteristic of prevention is acting prior to onset of a disorder, the early age of onset for most mental disorders indicates intervention must occur during long before adulthood.

Neuroscience has contributed evidence that longitudinal trajectories of brain development are affected by a combination of genetic and environmental factors. Neuroimaging studies have shown dynamic changes in brain structure and function continuing through childhood and adolescence, and geneticists have found that gene expression is highly dependent on environmental conditions. These findings imply that the brain is still highly plastic during childhood and adolescence. This may confer greater vulnerability to long-term effects of insults from trauma, substance abuse, or other adverse influences than in adulthood, but also the potential for lifelong beneficial effects from early positive interventions.

Growing interest in the possibilities afforded by research into prevention in children's mental health stimulated a series of large-scale reports and initiatives beginning in the early 1990s.[4–7] Advances in epidemiology, developmental psychopathology, and prevention science have converged to provide a framework to guide and evaluate prevention programmes. This chapter will discuss basic principles of public health and preventive medicine with application to mental health disorders in children and adolescents.

Public health and prevention: history and basic concepts

The goal of public health is the prevention of disease and promotion of health in communities. The World Health Organization has defined *health* as 'a state of physical, mental and social well-being and not merely the absence of disease or infirmity',[8] and *mental health* as 'a state of well-being in which the individual realizes his or

her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community'.[1] Public health differs from clinical medicine in that it addresses health-related matters on the level of populations rather than individuals. Public health activities include assessment of the health status and risk factors within a community through epidemiology, and population-focused interventions such as supporting the practice of preventive medicine, health education and behavioural modifications, creating and enforcing measures to maintain a healthy environment, and working to increase support for public health initiatives within the political sphere. In countries without universal access to health care public health offices may also act as providers of medical treatment for individuals without other means of access.[9]

Communities have acted to support the health of their members throughout history.[10] Common concerns for most societies have included control of epidemics, public sanitation, and promotion of personal hygiene, although the forms of public health interventions have varied depending on societal values, conceptions of the causes of ill health, and available resources. The health risks posed by the large-scale urban poverty and overcrowding associated with the industrial revolution helped to stimulate the growth of modern public health organizations, whose concerns eventually broadened to include issues such as workplace safety and regulation of the production of foods and medicines. Public health interventions changed to reflect advances in the understanding of disease processes, for example moving from general notions of the value of sanitation to focusing on specific infectious agents. Measures such as widespread vaccination and regulation of sanitary conditions have been so effective in developed countries that the focus of public health in these areas has shifted to chronic disorders such as heart disease and hypertension. Although emotional and behavioural issues have always been a concern of communities, systematic intervention to prevent mental disorders has lagged behind other disorders, in large part because of the lack of consensus regarding the nature of these problems or even how to classify them. A key factor in the advances in public mental health of the past several decades has been progress in epidemiology of mental health disorders.[9]

Epidemiology in public mental health

Incidence and prevalence

Epidemiology provides information about the *incidence* of a condition, meaning the number of new cases, which arise during a certain period of time, and its *prevalence*, meaning how many individuals have the condition during a certain period. The goal of prevention is to decrease a condition's incidence, i.e. prevent new cases from occurring, while successful treatment results in the decrease of the prevalence. Mental health disorders have presented challenges to epidemiology on several levels. In order to determine how many cases of a certain condition exist within the population, it is necessary to know how to define a case, but this is far from straightforward in the realm of mental health. Classification of medical disorders tends to evolve from symptom-based to mechanism-based as the links between a specific pathophysiology and the observed signs and symptoms are established. The lack of knowledge about the mechanisms producing cognitive and behavioural symptoms means that classification of mental disorders still relies upon descriptions of constellations of symptoms. *The International Classification of Disease version 10*(*ICD-10*),[11] and its United States counterpart *the Diagnostic and Statistical Manual TR-IV* (*DSM-TR-IV*),[12] are the results of iterative attempts by experts in the field to create meaningful classifications of psychiatric disorders based upon such observations in conjunction with applicable considerations of length and severity of illness, age of onset, and risk factors. This work has provided the standardized terminology that made possible the first large-scale epidemiologic descriptions of mental disorders. However, problematic issues pertinent to epidemiology remain, including questions regarding the relative merits of categorical versus dimensional classification systems; how to interpret the high rate of comorbidities for several disorders; and how best to account for individuals who have subthreshold symptoms, including how to determine the starting point of a disorder. It is not uncommon for individuals who have come to meet criteria for a mental health disorder such as schizophrenia or depression to have had a preceding period of subthreshold 'prodromal' symptoms, but healthy individuals also have occasional subthreshold symptoms that resolve without intervention. Unfortunately this differentiation often cannot be determined except retrospectively, despite the fact that there may be different implications for epidemiologic and preventive efforts.

The question of how symptoms change over time gains additional relevance when attempting to describe the epidemiology of mental health disorders in children and adolescents. As described in more detail elsewhere in this volume,[13] the science of developmental epidemiology has arisen as a response to the recognition that mental disorders may manifest in different ways over the lifespan, and that certain types of symptoms at one age may indicate that an individual is at high risk for developing a different disorder at a later stage of maturation. Risk factors may also have differing impact depending on an individual's developmental stage. Function may appear impaired if children are developing slowly in comparison with their peers, and it must be decided when this is normal variation and when it should be considered pathological. An additional layer of complexity in epidemiology in paediatric populations is the incorporation of information from additional informants such as parents and/or teachers, and determining how to evaluate the relationship of symptoms to particular contexts.

Risk factors

Epidemiology is also used to assess for the presence of risk factors. *Fixed risk factors* are those that cannot be altered, such as genotype. *Malleable risk factors* are susceptible to intervention, such as exposure to lead-based paint or domestic violence. *Causative risk factors* are those with known relationships to a particular outcome, and are of particular interest to prevention because they represent potential points of intervention. *Protective factors* instead decrease the risk of an adverse outcome. *Resilience* is a term used to describe an individual's ability to do well despite exposure to a typically high-risk situation.

Effective intervention to decrease risk factors or increase protective factors requires determining how these factors relate to each other and to the targeted health issues. The ultimate goal is a chain of causative steps leading from risk factor to outcome, but epidemiological data itself may provide sufficient guidance for action. One of the most famous examples of this was John Snow's identification of tainted drinking water from a particular well as the root

of a cholera epidemic in London, which he did based solely on epidemiological observations. Removal of the pump handle stopped the epidemic and proved that exposure was a causative risk, decades before the bacteria itself was identified. We are currently in a similar situation to Snow in regards to connecting risk factors to mechanisms for many mental health disorders, with the additional complication that mental health disorders are typically associated with combinations of a large number of individually modest potential risk factors.

Risk factors can be classified in terms of how they relate to each other and to the specified outcome,[14] and thus what type of intervention if any is appropriate. *Mediating* risk factors are those which explain how or why another factor affects the outcome; for example, the phenylketonuria enzyme *mediates* the effects of the phenylketonuria gene on IQ.[14] Although all causal factors are mediators, the reverse is not true, and experimental conditions are generally necessary to demonstrate that a particular mediator plays a causal role. A *moderating* risk factor instead specifies under what conditions or for whom another risk factor will affect outcome. Moderating risk factors describe populations that have differing responses to a given exposure, and may also represent potential sites of intervention to prevent an adverse outcome by reducing vulnerability or increasing resilience. A *proxy* risk factor, also called a *pseudocorrelation*, is one that itself does not strongly predict outcome but is highly correlated to a risk factor that does. *Overlapping* risk factors are those that arise from the same underlying construct and are observed to equally predict outcome, be highly correlated with each other and not stand in a specific temporal relationship; these can often be combined into a single factor. *Independent* risk factors conversely are unrelated to each other; they both predict outcome but without correlation or temporal precedence.

Theoretical models in prevention

The identification of risk factors and their interpretation evolves together with theoretical models for the causes and treatments of health problems. The fundamental model used throughout public health and epidemiology is that of *host-agent-environment*, in which the *host* is the person affected or at risk, the *agent* is the direct cause of disease, and the *environment* includes external factors which affects the host's vulnerability to the agent and the vector by which the agent reaches the host. While this model was first developed for infectious disease, it has been expanded to include other types of chronic non-infectious disorders.[10] Examples of pathogenic agents in the latter case include nutrition, chemicals, and genes; host factors include age, sex, and lifestyle; while social or economic issues are among those potentially affecting the environment. Another dimension that has gained increased attention in psychopathology is the actual transaction between the individual and environment—for example, the features of the way a child and parent interact. Intervening to remove risk factors from multiple domains simultaneously can potentially provide the most effective outcome.

Incorporating development into this model adds many challenges. The fields of developmental science and developmental psychopathology arose to create as a framework for the integration of information from developmental epidemiology, neuroscience, genetics, psychology, psychiatry, sociology, and other disciplines in order to better understand the complex interplay of factors affecting the health of an individual throughout their lifespan.[15] Major contributions from work in this area have been establishing the importance of interactions between genes and environment in determining the trajectory of development, rather than attributing mental health outcome to being due entirely to one factor or the other, and the dialectical nature of the relationship between the developing individual and their environment. The recognition of the importance of context in development has led to an elaboration of the different overlapping systems, or *ecologies*[16] that a child resides within and which present unique risks and opportunities for intervention.

Risk factors may be generalized, such as malnutrition and poverty, or more disease-specific, such as exposure to a particular toxin. Many risk factors will tend to occur together, and often risk factors have a non-linear relationship to outcomes; i.e. one or two may not significantly affect outcome, while as the number goes above a certain level risk increases sharply for a number of disorders. An additional complexity in developmental psychopathology is the presence of multicausality and multifinality. Multiple risks or disease processes may produce similar behavioural phenomena, while specific risk factors may be associated with a wide range of clinical presentations. Tracing causal paths and determining what are the factors that are mediating and moderating the relationships between risks and outcomes depends upon the ability to follow the impact of specific interventions over time.

The prevention research cycle and evidence-based prevention

Although direct experimentation on human subjects to establish causality among the risk factors affecting developmental trajectories is not in itself ethically feasible, suitably designed longitudinal controlled trials of preventive interventions can address the same goals.[17] Recognition of the value of considering prevention research as an iterative process led to the formulation of the *preventive research cycle*.[4] The steps in the cycle are: (i) identification of the problem or disorder and the size of its impact on a community; (ii) review of relevant information, particularly regarding relevant risk and protective factors available data from existing preventive research programmes; (iii) design, conduct, and analysis of pilot studies, including replication at multiple sites; (iv) implementation of larger-scale trials which will provide additional information about which populations may be more or less appropriate, and how the intervention does when scaled up in size; and (v) large-scale implementation and ongoing evaluation.

The randomized clinical trial, in which individuals or discrete communities are randomly assigned to receive either the intervention under investigation, a different intervention, or no intervention at all, continues to be a gold standard for determining whether a prevention programme itself is responsible for observed changes and thus to establish causality. Only a randomized clinical trial can determine if the intervention actually results in prevention, i.e. evidence that new cases did not develop that otherwise would have. Some trials, particularly those for populations in which some symptoms may already be present, result in decrease of those subsyndromal symptoms. While this is not without value, it is more strictly considered treatment than prevention.

The ability to make a convincing case for the value of a preventive intervention is particularly important because investing in prevention is asking an individual or community to devote resources towards a problem that has not yet occurred. Standards for evidence-based preventions have been explicitly identified to help with design and evaluation of studies, including criteria for when there is sufficient grounds to move along the research cycle from pilot studies to large-scale field trials and final dissemination.[18] The recommendations provide guidance for appropriate statistical methodology and design, and emphasize the need for replication in independent samples, adequate provision of training materials for non-research personnel as the scope of the project grows, and ongoing data collection after dissemination to inform communities and researchers about the impact of the intervention and direct the next iteration. They also differentiate between *effectiveness*, defined as showing a positive result in pilot studies under highly controlled circumstances, and *efficacy*, indicating a programme is also able to produce results in the less-optimal conditions associated with larger-scale trials.

Types of preventive interventions

Once it has been determined that a particular problem is present, and pertinent risk and protective factors have been identified, it is necessary to determine what type of intervention is most likely to be effective. Two broad distinctions are applicable to any intervention. The first, as implied by the host-agent-environment model, is whether to address the individual, their environment, or both. The second distinction concerns which portions of the population potentially at risk are to be addressed.[19]

Primary, secondary, and tertiary prevention

The first widely used public health prevention categories were proposed by the Chronic Disease Commission in 1957, who classified prevention as being primary, secondary, or tertiary.[20] *Primary prevention* is aimed at the normal population and defined as efforts aimed at decreasing the incidence of new cases, such as preventing access to contaminated water supplies as illustrated by the case of John Snow. *Secondary prevention* is targeted towards individuals who already show early signs of disease or disorder, with the aim of decreasing the prevalence of already established cases. *Tertiary prevention* attempts to minimize the degree of morbidity associated with an established illness, through decreasing its duration or associated disability. Such definitions were a crucial step in designing interventions to focus on a specific population and problem and take into account specific characteristics of that situation. However, classifying prevention by the disease stage may require a greater understanding of how risks related to disorders than is possible for many conditions.

Universal, selected, and indicated prevention

An alternative classification system based upon risk–benefit considerations for preventive interventions was introduced by Gordon in 1983[21] and disseminated through the seminal *Institute of Medicine* report in 1994.[4] Gordon proposed that the benefit of a prevention programme could be assessed by comparing an individual's risk of developing a disorder with the risk or cost of the associated intervention. In his system, prevention is classified as *universal*, *selected*, or *indicated*, depending on the degree of identified risk. *Universal prevention* is applied to everyone in a defined population, and the associated interventions are optimally low risk, low cost, and may be administered by individuals who possess relatively little specialized training. However, universal prevention spends resources on a large number of individuals who would not have become ill in any case. *Selective intervention* is aimed at individuals at above-average risk for a disorder, and anticipates a commensurately higher cost and intensity of intervention. Finally, *indicated prevention* is for individuals who are showing early signs of a disorder or exhibit biological markers indicating risk; the acceptable cost and risk here would again be higher to reflect the increased need of the individual.

There are areas of similarity between the two systems, which has led to some confusion. The populations and goals of primary and universal prevention are comparable, but selected and indicated groups indicate individuals at increasing levels of risk but who do not yet meet criteria for a disorder, whereas secondary and tertiary address issues related to different stages of having a disorder.

Comparison of prevention and health promotion

An alternate conceptualization of how to proactively intervene to improve outcomes is *health promotion*, defined as measures to increase likelihood of wellness as a positive quality rather than limiting efforts to decreasing risks for a negative outcome.[5] From a practical standpoint it overlaps largely with universal prevention, but the theoretical foundations and targeted outcomes differ. Although few would disagree with the potential benefits of promoting health in the community, health promotion has not always been included within the scope of prevention policy due to concerns that it may dilute efforts towards risk prevention that are characterized by more clearly definable and measurable outcomes. Others have argued that in view of the complex pathways of developmental psychopathology, a less-specific approach is more consonant with our existing knowledge, and may actually be more effective over the long run to create resilience against a broader range of disorders.[22]

Effective preventive interventions for children's mental health

The degree of implementation of preventive measures for mental health disorders in children and adolescents depends largely on how convincingly specific risk factors can be demonstrated which are malleable to politically and economically feasible actions. For example, realization of the adverse effects of prenatal alcohol exposure on neurodevelopment resulted in widespread public education efforts. Other toxins such as lead-based paint have also been the focus of education and regulations to decrease children's exposure. Vaccination programmes have significantly reduced mental disorders associated with infectious diseases such as rubella, and programmes have been put in place to decrease risks from accidents through measures such as use of car seats and bicycle helmets.

For conditions where the core risk factors are less clearly defined progress has been slower, but enough data has accrued over the past two decades of systematic prevention trials to be able to begin to assess the effectiveness of preventive interventions in this context. The scope of the current chapter does not allow a detailed description, and the reader is referred to relevant chapters for

specific disorders within this text, as well as general reviews and meta-analyses available elsewhere.[4,5,16,19]

Health promotion

A recent report by the World Health Organization summarized globally relevant general risk factors and the state of evidence for specific interventions to promote mental health and resilience.[6] Social, environmental, and economic factors have a major impact on mental health. Increasing attention is also being paid to *social capital*, a concept which broadly refers to aspects of social organization and community norms that facilitate the ability of individuals to work together for mutual benefit. General measures for health promotion include improving nutrition, housing, access to education and economic security, as well as strengthening community networks, reducing exposure to violence, decreasing substance abuse, and intervention to help with recovery from disasters. Risks for children's mental health from the proximal family environment include adverse maternal behaviour during pregnancy, such as substance abuse, child abuse, parental mental illness, and domestic violence, and may be addressed with measures such as home-visiting programmes for pregnant women and new mothers and pre-school programmes.

While the benefits of such general health promotion activities seems highly plausible, rigorous evidence of their effects on mental health allowing quantification for cost-benefit questions is difficult to come by and currently patchy, especially for larger-scale interventions. Reasons for this include the length of time necessary to see the results of these interventions, which may be much longer than the policy environment which fostered them; the large samples necessary, which may range from difficult to impossible to randomize appropriately, and the lack of funding for this type of information-gathering. Here in particular naturalistic 'experiments' may be of aid, in which populations exposed to changes in risk factors or social policies are closely monitored for the impact on health outcomes.

Prevention

The practical implementation of universal preventive measures overlaps largely with health promotion, despite the differences in their theoretical background and aims. Universal prevention programmes have the significant advantage of not conferring stigma upon participants, but their benefits are difficult to quantify due to the generally small effect sizes and consequent large samples necessary,[23] and they by definition devote significant resources to individuals who likely would not have had problems regardless. Universal prevention programmes with evidence of benefit have been developed for issues such as conduct disorder, anxiety, and depression. These have been primarily school-based, focusing on classroom behavioural management, social skills training, and cognitive strategies to help children learn prosocial behaviours and cope with stressful situations. Some programmes adopt a multimodal approach which includes parents. Universal approaches to decrease substance abuse have had mixed results. Educational techniques have shown clear success in increasing the knowledge base regarding risks of substance abuse, but impact on actual usage has been harder to demonstrate outside of more comprehensive programmes targeting multiple types of risk.[16]

Preventive measures for selected populations become more specific to individual disorders. Children at risk for conduct disorder often come from impoverished environments with high rates of exposure to violence, substance abuse, and weak family and community structures. Secondary preventions in these settings accordingly rely more strongly on multimodal interventions which include the family. Children with a depressed parent are at increased risk for depression, and secondary interventions in this case may include treatment of the depressed parent in addition to cognitive therapies for the child. Stress related to difficult transitions such as parental death, divorce, or unemployment are also significant risk factors for children, and have been effectively addressed with courses of cognitive group therapy.

Indicated prevention measures, for children with early symptoms or biological markers of a disorder, have the narrowest scope and generally the clearest evidence of effectiveness. A seminal study in the prevention of depression was performed by Clarke and colleagues,[24] who showed that cognitive therapy in adolescents with subsyndromal symptoms of depression and a depressed parent could reduce the incidence of new cases of depression compared to a control group. Schizophrenia has become a target of preventive medicine through studies showing that treatment of adolescents with early symptoms of psychosis may delay onset of a full psychotic break.[25] Multimodal interventions have also been shown to be effective for children and adolescents already showing signs of increased aggressive or antisocial behaviour.

Common themes in 'best-practice' mental health prevention programmes have included the need for multimodal approaches which simultaneously address both the child and components of the environment, and the increased durability of improved outcomes when interventions are maintained for significant lengths of time. Many programmes focus on reduction of proximal risk factors rather than mental health disorders themselves as a more feasible outcome measure, although when possible it is optimal to incorporate both. When preventive research began to stratify interventions into universal, selected, and indicated, it was originally predicted that the lowest-risk individuals would benefit the most from universal-level interventions. It was instead found that higher-risk children actually showed the greatest response, supporting the development of tiered systems in which children who did not benefit adequately from universal measures could also be referred for secondary and indicated levels.

In general, meta-analyses have found that preventive programmes in mental health for selected and indicated populations have small to moderate effect sizes, similar to those seen in other areas of medicine. Anxiety and depression have shown the most consistent responses. Universal programmes have not shown significant effectiveness in meta-analyses of controlled trials, which is understandable given the necessary sample sizes, but has led to debate regarding the justification of their claim on scarce resources. Another concern is that most research has been carried out at the level of pilot studies, with much less available from large-scale trials or fully disseminated programmes. What information is available shows a tendency for a fall off in effectiveness when moving to larger-scale implementation. This suggests a need to spend more attention from the earliest stages on issues relating to dissemination such as adapting programmes for existing community infrastructure. Early collaboration with the members of the targeted community also helps to ensure relevance of a programme and consequent participation; for example parenting classes, while potentially valuable, may not attract individuals preoccupied with issues,

such as safety or securing transportation. Finally, the issue of how to transfer preventive programmes into different settings requires much more extensive attention. Most prevention research has been done in a few of the more affluent nations, primarily the United States, United Kingdom, Canada, Australia, and countries in northern Europe. Little is known about how to transfer programmes or which programmes may be suitable for other less-affluent areas.

Conclusion

Enormous progress has been made in recognizing the scope of mental health problems for children around the world, and in developing the theoretical framework needed to address decreasing this burden in a systematic fashion. Technological advances in neuroimaging, genetics, and computational biology are providing the tools to start describing the biological processes underlying the complex course of development, and have renewed appreciation of the role of the environment in determining how a genetic heritage is expressed.

However, rapid technological change is also altering the environment of children and their families at an unprecedented rate, and what kinds of challenges to public health these changes may present is not yet fully understood. What is becoming clear is that as technological advances increase the range of available health care treatments, along with the potential cost, the choices for societies between spending limited resources on treatment or prevention will have to become increasingly deliberate.

A substantial body of work has demonstrated that prevention in mental health can be effective, but those who would benefit the most from preventive interventions are often not those with the political or economic resources to make them a priority. While the potential interventions to prevent mental health disorders in children are constrained by the knowledge and resources available, what is actually done depends upon the social and political values of individual communities and nations.[9] It is to be hoped that as our understanding of these disorders grows, public policies to prevent the development of mental health disorders in children will become as commonplace a responsibility for modern societies as the provision of clean drinking water.

Further information

UK. National Health Service Guide for Child and Adolescent Mental Health: http://www.bma.org.uk/ap.nsf/Content/Childadolescentmentalhealth

World Health Organization webpage for mental disorders: http://www.who.int/topics/mental_disorders

U.S. substance abuse and mental health services administration: clinical preventive services in substance abuse and mental health update: from Science to services http://www.samhsa.gov/publications/allpubs/SMA04-3906/i.asp

Society for prevention research: http://www.preventionresearch.org

References

1. WHO. (2001). *The world health report 2001: mental health: new understanding, new hope*. World Health Organization, Geneva.
2. Costello, E.J., Egger, H., and Angold, A. (2005). 10-year research update review: the epidemiology of child and adolescent psychiatric disorders: I. Methods and public health burden. *Journal of the American Academy of Child and Adolescent Psychiatry*, **44**(10), 972–86.
3. Kessler, R.C., Berglund, P., Demler, O., *et al.* (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **62**(6), 593–602.
4. Mrazek, P.B. and Haggerty, R.J. (1994). *Institute of Medicine (U.S.). Committee on prevention of mental disorders, United States. Congress. Reducing risks for mental disorders: frontiers for preventive intervention research*. National Academy Press, Washington, DC.
5. WHO. (2004). *Prevention of mental disorders: effective interventions and policy options summary report*. World Health Organization, Geneva.
6. WHO. (2004). *Promoting mental health: concepts, emerging evidence, practice*. World Health Organization, Geneva.
7. HMSO. (1992). *The health of the nation: a summary of the strategy for health in England*. HMSO, London.
8. WHO. (2001). *Basic documents*. World Health Organization, Geneva.
9. Detels, R. (2002). *Oxford textbook of public health* (4th edn). Oxford University Press, Oxford, New York.
10. Tulchinsky, T.H. and Varavikova, E. (2000). *The new public health: an introduction for the 21st century*, In (eds. K. Rhoshel, and M.D. Lenroot). *Child Psychiatry Branch*, National Institute of Mental Health, pp. 5–54. Academic Press, San Diego.
11. World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*. World Health Organization, Geneva.
12. American Psychiatric Association. (2000). *Task force on DSM-IV. Diagnostic and statistical manual of mental disorders: DSM-IV-TR* (4th edn). American Psychiatric Association, Washington, DC.
13. Costello, F.J and Angold, A. (2009). *Epidemiology of psychiatric disorder in Chilood and adolecence*. In *New Oxford Textbook od Psychiatry* (2nd edn.) (eds. M.G. Gelder, N.C. Andreasen, J.J. López-I bor Jr, and J.R. Geddes)
14. Kraemer, H.C., Stice, E., Kazdin, A., *et al.* (2001). How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *The American Journal of Psychiatry*, **158**(6), 848–56.
15. Rutter, M. and Sroufe, L.A. (2000). Developmental psychopathology: concepts and challenges. *Development and Psychopathology*, **12**(3), 265–96.
16. Greenberg, M.T., Domitrovich, C., and Bumbarger, B. (2001). The prevention of mental disorders in school-aged children: current state of the field. *Prevention & Treatment*, **4**(1), 1–62.
17. Coie, J.D., Watt, N.F., West, S.G., *et al.* (1993). The science of prevention. A conceptual framework and some directions for a national research program. *The American Psychologist*, **48**(10), 1013–22.
18. Flay, B.R., Biglan, A., Boruch, R.F., *et al.* (2005). Standards of evidence: criteria for efficacy, effectiveness and dissemination. *Prevention Science: The Official Journal of the Society for Prevention Research*, **6**(3), 151–75.
19. Durlak, J.A. and Wells, A.M. (1997). Primary prevention mental health programs for children and adolescents: a meta-analytic review. *American Journal of Community Psychology*, **25**(2), 115–52.
20. Commission on Chronic Illness. (1957). *Chronic illness in the United States*. Harvard University Press, Cambridge, MA.
21. Gordon, R.S., Jr. (1983). An operational classification of disease prevention. *Public Health Reports*, **98**(2), 107–9.
22. Cowen, E.L. and Durlak, J.A. (2000). Social policy and prevention in mental health. *Development and Psychopathology*, **12**(4), 815–34.
23. Cuijpers, P. (2003). Examining the effects of prevention programs on the incidence of new cases of mental disorders: the lack of statistical power. *The American Journal of Psychiatry*, **160**(8), 1385–91.
24. Clarke, G.N., Hornbrook, M., Lynch, F., *et al.* (2001). A randomized trial of a group cognitive intervention for preventing depression in adolescent offspring of depressed parents. *Archives of General Psychiatry*, **58**(12), 1127–34.
25. McGorry, P.D., Yung, A.R., Phillips, L.J., *et al.* (2002). Randomized controlled trial of interventions designed to reduce the risk of progression to first-episode psychosis in a clinical sample with subthreshold symptoms. *Archives of General Psychiatry*, **59**(10), 921–8.

9.2

Clinical syndromes

Contents

Note *Substance abuse is considered in Part 4, Section 4.12. Aspects relevant to young people are considered within the chapters of this section.*

9.2.1 Neuropsychiatric disorders

James C. Harris

The developmental perspective

Developmental neuropsychiatry addresses the neurobiological basis of behaviour in infants, children, and adolescents with neurodevelopmental disorders and in those with brain damage occurring during the developmental period. As a field, it includes the aetiology, diagnosis, and treatment of behavioural, emotional, interpersonal, and psychiatric disorders.[1, 2] The parent's response, adjustment to, and involvement in treatment is a critical element in outcome.

The developmental neuropsychiatrist utilizes a developmental perspective that focuses on the developing person who is active, socially oriented, and emerging rather than passively responding to the environment. The adaptive plasticity of the developing nervous system to change is emphasized, and the essential role of environmental experience in brain development is acknowledged. When working with the affected child, an effort is made to provide the supports needed to facilitate the mastery of age-appropriate developmental tasks always keeping in mind the child's individual capacities and strengths.

Scope of developmental neuropsychiatry

The scope of developmental neuropsychiatry is broad[2] and includes the following.

1 Neurodevelopmental disorders that are described in other chapters of this book, including attention-deficit and hyperactivity disorders (Chapter 9.2.4), pervasive developmental disorders and childhood-onset schizophrenia (Chapter 9.2.3), obsessive–compulsive disorder and Tourette's syndrome (Chapter 9.2.8), and specific developmental disorders (Chapter 9.2.2).

2 Neurogenetic disorders, both cytogenetic and metabolic, with behavioural phenotypes, several of which are also reviewed in Chapter 10.4.

3 Teratogenic exposure from both organic and inorganic toxins. In these instances, behavioural dysfunction may result from gestational substance abuse with alcohol and other substances or exposure to inorganic metals.

4 Endocrinopathies.

5 Traumatic brain injury.

6 Other neurological disorders (e.g. epilepsy).

Clinical features

Neurodevelopmental disorders

Developmental psychopathology applies developmental concepts to the study of neurodevelopmental disorders. The relationship of disordered to non-disordered behaviour is considered, as are the early origins of maladaptive behaviours that may not appear in clinical form until adolescence or adulthood. Knowledge of normal development is utilized to study children whose development is atypical, in order to understand the natural history of their disorder and establish the developmental trajectory of that particular condition. Conversely, the investigation of such deviant behaviour associated with a particular disorder is considered in regard to our understanding of normal development. For example, attention-deficit hyperactivity disorder has been investigated as a disorder of executive functions of the prefrontal cortex, and autistic disorder as a disorder of social cognition and communication. In both instances, new knowledge about brain functions has been derived from these formulations. Among the neurodevelopmental disorders, the age of recognition varies, multiple causes are involved, and many transformations in behaviour may occur in determining their complex course. The goal is to understand the mechanisms and processes through which risk factors lead to the emergence of a disorder. Disordered behaviour is not viewed as a static condition, but is considered as part of a dynamic transactional engagement. Behaviour and development are viewed within a social context, and the transactional nature of interactions is considered from infancy through adulthood to understand these processes.

Attention-deficit hyperactivity disorder, pervasive developmental disorders, obsessive–compulsive disorder, Tourette's syndrome, and childhood-onset schizophrenia are developmental neuropsychiatric disorders under active investigation and each is reviewed in the respective chapters. Their developmental psychopathology is investigated by addressing the origins and course of individual patterns of behavioural maladaptation in each of these disorders and determining their genetic bases, thought to be complex, and involving more than one gene. Information derived from genetics, developmental psychology, clinical psychology, psychiatry, sociology, physiological sciences, neurosciences, and epidemiology is included in the description of each of these disorders.

The interrelationship of the various child neuropsychiatric disorders is an important consideration. Disorders may be risk factors for other conditions, so that attention-deficit disorder may be a risk factor for conduct disorder. In this instance, the child's behaviour affects the adult and the transactional interactions between child and adult may result in further disruptive behaviours. Moreover, there may be a developmental basis for disorders whose full presentation is not evident until later in life, as is the case with schizophrenia—generally considered to be a disorder of late adolescence or early adult life, but with origins in the developmental period.[3] Some disorders may have co-occurring diagnoses that influence their outcome, as in Tourette's syndrome, where co-occurring conditions may determine the behavioural presentation. In Tourette's syndrome, obsessive–compulsive symptoms may be an aspect of 'pure' Tourette's syndrome, while co-occurring disruptive behaviour may be secondary to co-occurring attention-deficit disorder. Social and behavioural dysfunction in children with Tourette's syndrome is largely ADHD-specific. Children with TS alone have a different social-emotional profile.[4–6] Compulsive behaviours may not only interfere with the normal routines for the affected child but also become particularly problematic for their impact on other family members.

Neurogenetic syndromes with behavioural phenotypes

Particular patterns of behaviour, temperament, and psychopathology may be associated with specific chromosomal and genetic disorders.[2,5,7–9] The term 'behavioural phenotype' was introduced by Nyhan in 1972[7,10] to describe patterns of unusual behaviour that are so characteristic that they suggest a specific neurogenetic disorder. Nyhan described stereotypical patterns of behaviour occurring in syndromic fashion in sizeable numbers of affected individuals with a given syndrome, and observed that these patterns seemed self-programmed. In these children, he proposed that it is reasonable to hypothesize that their behaviours are associated with an abnormal neuroanatomy and that such stereotypical patterns of unusual behaviour could reflect the presence of structural deficits in the central nervous system. Recent developments in the neurosciences provide a means to investigate the biological bases of behavioural phenotypes. Behavioural assessments, neuropsychological testing, and neuroimaging procedures, carried out in well-characterized genetic syndromes, are being utilized to understand pathways from genes to cognition and complex behaviours in these conditions.

Comprehensive study of children with different neurogenetic disorders may increase our appreciation for the relative contribution of genetic variables in the pathogenesis of specific, affective, and behavioural disorders. Behavioural phenotypes have been studied most extensively in Down syndrome (language),[11] fragile X syndrome (gaze aversion, hyperkinesia, autistic-like behaviour),[12] Williams syndrome (sociability, hyperverbal behaviour, and visuospatial deficits),[13,14] Lesch–Nyhan syndrome (compulsive self-injury and aggression),[15–17] and Prader–Willi syndrome (hyperphagia, obsessive–compulsive behaviour).[13,18,19] The number of identifiable behavioural phenotypes is growing with careful observations of behaviours in neurogenetic disorders.[8,9] Besides behaviours, particular temperamental features have also been considered in these disorders. However, when studying temperament, the appropriate measures must be chosen. For example, when Down syndrome, proposed to be linked to a particular temperament, was studied using temperamental clusters of easy temperament, slow to warm-up, and difficult temperament, Gunn *et al.*[20] demonstrated both easy and difficult temperament in children with Down syndrome; therefore, a typical temperamental pattern among these three categories was not demonstrated. However, when a more comprehensive assessment was carried out in other syndromes[21] (that included the personality factors of extraversion,

agreeableness, conscientiousness, emotional stability, and openness, along with motor activity and irritability), specific personality phenotypes were identified. These were differentially related to parental behaviours and family context in Prader–Willi, fragile X, and Williams syndromes. Moreover, isolated special abilities, as in calculation and in music,[22] are recognizable that might be considered as phenotypes and linked to the proposed modular organization of the central nervous system. Finally, physical and behavioural phenotypes are not only identified in neurogenetic syndromes but also in those caused by environmental events, such as intrauterine exposure to alcohol: namely, the foetal alcohol syndrome. Because alcoholism is a familial disorder, there may vulnerability to its effects resulting in a severe presentation in some individuals and less severe presentation in others.[23]

Both traditional Mendelian laws of inheritance (Lesch–Nyhan syndrome) and non-traditional inheritance have been identified in conditions with behavioural phenotypes. Among the non-traditional forms of inheritance are triplet repeat amplification (fragile X syndrome), microdeletion or contiguous gene deletion (Williams syndrome), imprinting (Prader–Willi syndrome), transcriptional derepression (Rett's syndrome), and excessive gene dosage (Down syndrome). A key finding is the recognition that mutations of single genes can lead to complex behavioural symptoms, especially if the affected protein is essential for the expression or processing of multiple 'downstream' genes.

Behavioural phenotypes are also discussed in relation to intellectual disability in Chapter 10.5.1.

Neurobehavioural teratology

Neurobehavioural teratology investigates abnormal development of the nervous system and of cognition and complex behaviour that results from prenatal environmental insults. Neurobehavioural research addresses the prevalence of cognitive behavioural disorders in exposed individuals and the consequences of the brain insult on other developing brain systems, to identify risks for functional or behavioural deficits. Investigators focus on cognitive behavioural deficits and their underlying anatomy and embryology. Assessment emphasizes not only IQ but also neuropsychological profiles, because learning disability or difficulty in visuomotor integration may be evident in children who function in the low to average range of general mental ability.

The natural history of intrauterine drug exposure on motor, cognitive, emotional, and social behaviour is an area of growing concern. Multiple drug exposures during pregnancy are common among substance-abusing mothers. Of syndromes associated with intrauterine substance abuse, alcohol abuse has been studied the most extensively. Subsequently, retinoids, anticonvulsants (lithium, tegretol, and valproic acid), and the selective serotonin-reuptake inhibitors have also been studied. Other teratogens do not lead to major malformations of the nervous systems but they do compromise its integrity (for example, lead, heroin, methadone), and are associated with neurotoxic damage or effects on neurochemical systems.

The greatest period of vulnerability to drugs in a human pregnancy is during the period of embryogenesis (days 14 to 60). During embryogenesis, many neurobehavioural teratogens (for instance, retinoids and ethanol) produce syndromes with abnormalities that involve craniofacial, neural, and major organ systems. Behavioural abnormalities without detectable physical abnormality can occur when the insult occurs during the foetal period. The extent of malformation is stage-specific and dose-dependent, with outcomes ranging from death with malformation, malformation and survival, effects on growth, and cognitive–neuropsychological or behaviour disorder. The same exposure to alcohol needed to produce cognitive behavioural change in the foetal period would generally cause malformation if it occurred during embryogenesis. The term 'developmental toxicology' is sometimes used if the insult occurs in the postnatal period.

There may be a genetic vulnerability that influences the extent of expression of response to environmental toxins in an individual. A common family of regulatory genes is involved in the formation of structures of the face, head, hindbrain, parts of the heart, and thymus gland, all of which share a common origin from neural crest cells (anterior neural tube). These regulatory genes, known as HOX genes, provide rules for assembling various structures and for determining particular anatomical segments.[24] Homozygous HOXA1 mutations have been shown to disrupt human brainstem, inner ear, cardiovascular and cognitive development. Because the retinoid family is involved in controlling these HOX genes,[25] a similar pattern is produced by excessive retinoid administration, as in hypervitaminosis of vitamin A (retinol). Moreover, the enzyme alcohol dehydrogenase functions in the metabolism of both retinol and ethanol so that intoxicating levels of ethanol can competitively inhibit the metabolism of retinol and impact brain development. Thus, both genetic and teratogenetic agents may produce similar developmental abnormalities. Understanding these mechanisms helps to understand how an abnormal facial appearance may suggest an abnormal brain.

Foetal alcohol spectrum disorder syndrome

Foetal alcohol syndrome is one of the most commonly recognized causes of intellectual disability; one that is preventable if recommended guidelines regarding alcohol use are followed by mothers.[26]

Clinical features

Children with the full foetal alcohol syndrome demonstrate prenatal and postnatal growth deficiency, microcephaly, infantile irritability, mild to moderate intellectual disability, and a characteristic facial appearance.[23] The extent of the abnormality depends on the time of maximal exposure to alcohol and the dose. Approximately half of those affected have co-ordination problems, are hypotonic, and have attention deficits. Between 20 and 50 per cent have other birth defects, including eye and ear anomalies and cardiac anomalies. Those children who do not show growth retardation or congenital anomalies may show more subtle changes, such as attention problems, disruptive behaviour, reduced speed of information processing, motor clumsiness, speech disorders, fine motor impairment, and learning problems, especially in mathematics.[23,27] These findings have been documented in a prospective longitudinal study of the effects of prenatal alcohol exposure on a birth cohort of 500 offspring who were selected from 1529 consecutive pregnant women in prenatal care in community hospitals.[28] Dose-dependent effects are most clear from the neurobehavioural status of subjects when regular neurodevelopmental evaluations are carried out from birth to age 14 years. The more subtle abnormalities are referred to as 'foetal alcohol effects', or alcohol-related neurodevelopmental disorder. The full range of disabilities is described as foetal alcohol spectrum disorder.[27]

Subjects with average to above-average IQ may demonstrate neuropsychological deficits in complex attention, verbal learning, and executive functioning. Disruptive behaviour, attention-deficit disorder, anxiety disorder, and communication disorder have been described[29,31,33] in children with foetal alcohol syndrome and foetal alcohol spectrum disorder who test in the low normal range and in the moderate to severe range of intellectual disability.

Behavioural phenotype

The behavioural phenotype is characterized by problems in cognitive functioning, academic problems in arithmetic, difficulty with abstractions, understanding cause and effect, and generalizing from one situation to another. Thus, inattention, poor concentration, impaired judgement, memory deficits, and problems in abstract reasoning are characteristic. Behavioural problems related to impulsivity and hyperactivity makes them vulnerable to later diagnoses of oppositional defiant and conduct disorder.[27,30]

Natural history

Foetal alcohol spectrum disorder is not only a childhood disorder; the cognitive and behavioural effects and psychosocial problems may persist throughout adolescence into adulthood.[28,33] Although the facial features are not as distinctive after puberty and the growth deficiency is not as apparent as in the younger child, the central nervous system effects do persist throughout life. Approximately 50 per cent of those affected function as intellectually disabled persons. Moreover, adaptive behavioural problems in communication skills and in socialization are apparent in those with foetal alcohol spectrum disorder whose intelligence test scores are in the normal range.

Poor judgement, attention problems, distractibility, difficulty in recognizing common social cues, and problems in modulating mood continue as characteristic features. Family environmental problems often continue as risk factors for behavioural problems if there is a lack of stability in family life. In one follow-up study[34] that used structured interviews with non-intellectually disabled affected subjects, the most common diagnoses were alcohol or drug dependence, mood disorders, and personality disorders (especially passive aggressive or antisocial). Further follow-up is needed to investigate the mechanisms involved in these psychiatric presentations, and particularly in determining the pathways leading to alcoholism.

Epidemiology

Foetal alcohol syndrome is a common cause of neuropsychiatric disorders, with a worldwide incidence of approximately 1.9 in 1000 live births. When foetal alcohol syndrome and alcohol-related neurodevelopmental disorder are considered together, the combined rate in one study conducted in the United States was 9.1 in 1000.[35] Despite its frequency and severity, the syndrome may go unrecognized because physicians may not systematically enquire about alcohol use and may not recognize the spectrum of the effects of prenatal alcohol exposure on neurodevelopment.

Aetiology

The amount and pattern of alcohol consumption and the trimester of use during pregnancy, especially if during critical periods of brain development, are major factors in determining outcome. Binge drinking patterns with high blood concentrations are especially deleterious. Rapid changes in alcohol concentrations in the blood and central nervous system cause apoptotic damage (cell degeneration) in developing neurons and other cells in rat models.[36] Microcephaly is commonly reported in foetal alcohol syndrome and suggests an underdevelopment of the brain. Neuropathological studies demonstrate the underdevelopment or absence of the corpus callosum and enlarged lateral ventricles. Dendritic changes have been observed in animals with prenatal exposure to alcohol; these changes were correlated with decreased learning ability. Magnetic resonance imaging studies have documented brain abnormalities in foetal alcohol syndrome, particularly in midline frontal structures such as the corpus collosum.[37,38] Research to identify specific polymorphisms contributing to foetal alcohol spectrum disorder is at an early stage. Polymorphisms of only one of the genes for the alcohol dehydrogenase enzyme family, the ADH1B, have been demonstrated to contribute to vulnerability.[39]

Treatment

(a) Evidence

Mothers of children with foetal alcohol syndrome who drank more alcohol and drank excessively early in gestation have more severe clinical features. Alcohol use in late pregnancy is primarily associated with prematurity and infants who are small for gestational age, rather than with the full foetal alcohol syndrome. Because of these risks, treatment must begin with prevention.[32] There is no clearly agreed safe dose of alcohol for pregnant women. Because there is no known safe amount of alcohol consumption during pregnancy, it is recommended that women who are pregnant or who are planning a pregnancy abstain from drinking alcohol. Special efforts for educating women of child-bearing age are required that highlight the harmful effects of alcohol; identified children must be referred for early educational services.

(b) Management

A comprehensive treatment programme begins with parental acknowledgement of the aetiology of foetal alcohol syndrome or spectrum disorder and treatment for the parent, as indicated, for alcohol misuse and abuse. Parental counselling should include discussion of the physical and behavioural phenotype. The family should be advised about the need for special educational programmes and assisted in behavioural management. Family therapy is often required to help family members cope with the developmental disorder. Appropriate educational and behavioural treatment resources are needed to address the social deficits, particularly in those cases where disruptive behaviour, attention-deficit disorder and mood disorders are identified.

Foetal alcohol syndrome is also considered as a cause of intellectual disability in Chapter 10.4.

Gestational substance abuse

Opiates

Exposure to heroin and methadone ranges from a neonatal withdrawal syndrome to less predictable long-term outcomes.[40] An impoverished environment may have disproportionate adverse effects on methadone-exposed children when compared to unexposed children.[41] Methadone effects have been associated with

increased body tension, poor motor co-ordination, and delay in motor skills acquisition.[42] However, the effects on mental development are less clear, but they do affect the child-rearing environment. Since methadone exposure produces an increased vulnerability to the effects of poor parent–infant relationships, these relationships require careful monitoring.

Cocaine

Cocaine is a central nervous system stimulant that inhibits nerve conduction in the peripheral nervous system. Cocaine is metabolized primarily through the plasma cholinesterase system, with the primary metabolic product being benzoylecgonine. Since cocaine rapidly crosses the placenta by simple diffusion, foetal peak blood levels are reached as quickly as 3 min.[43] Having crossed the placenta in the foetus, cocaine has the same direct actions on the foetal cardiovascular system as seen in the maternal system. These cardiac changes involve the direct effects of cocaine, as well as indirect effects such as foetal hypoxia. Cocaine may lead to placental dysfunction (via vasocontracture effects), structural changes (via vascular compromise), and neurobehavioural abnormalities (via postsynaptic junction neurotoxicity).

Infant gestational age, birth weight, head circumference, and length have been found to be decreased in affected infants, and low birth weight is a frequent finding in studies of the offspring of cocaine-using women. In addition to abnormal growth patterns, congenital anomalies involving the genitourinary tract, heart, and central nervous system as well as limb-reduction abnormalities have been reported. A potential mechanism for all these anomalies appears to be interruption of the intrauterine blood supply, with subsequent disruption of embryonic development. Although approximately 25 to 30 per cent of infants exposed to cocaine *in utero* may have physical difficulties, overall neurobehavioural problems may be more common, and most apparent in early infancy and childhood.[40, 44] In one large study, at age 4 years, prenatal cocaine exposure was not associated with lower full-scale, verbal, or performance IQ scores but was associated with an increased risk for specific cognitive impairments. A better home environment was associated with IQ scores for cocaine-exposed children that are similar to scores in non-exposed children. Although irritability and problems in state regulation are reported in infants and impulsive behaviour in pre-school children, these behaviours diminish over time with behavioural and psychosocial interventions.[45] Child abuse is closely linked to substance abuse.

Treatment

These findings suggest that careful attention be paid to the postnatal home-rearing environment of children who are exposed to drugs *in utero*. Overall, the treatment programme must take into account physical and psychological change secondary to intrauterine drug use as well as the postnatal nurturing environment. Both substance use and psychiatric disorder in the parents must be considered, because parents with attention-deficit disorder and mood disorders may themselves self-medicate with cocaine. Without early intervention, special school programmes, behavioural management programmes, and a structured day programme will be necessary. Ongoing parent training is also required.[40, 42]

Endocrinopathies

Congenital hypothyroidism

Congenital hypothyroidism is associated with intellectual disablity and may be associated with decreased motor activity at birth, hoarse cry, and difficulty with feeding. It is rarely diagnosed at birth from clinical assessment alone, but it is recognized from newborn screening tests with confirmation by measurement in blood samples. Symptoms of hypothyroidism may not be clearly detected until the second month of life. The overall prevalence is 1 in 4000 live births. Neurological and learning disorders associated with untreated congenital hypothyroidism include attention-deficits, hearing loss, speech defects, ataxia, and abnormal muscle tone.[46] Rapid diagnosis in infancy is essential to prevent these complications. Without treatment, severe neurological dysfunction ensues. With initiation of oral thyroid hormone treatment (levothyroxine in a single daily dose of 8 to 10 μg/kg per day) in the first 6 weeks of life, IQ is in the normal range. If treatment is delayed until 3 to 6 months, IQ drops to an average of 75, and, if initiated after 6 weeks, to an IQ of 55 or less. Rearing environment is important in long-term outcome. Despite early treatment there still may be enduring cognitive and motor deficits in young adults.[47]

Traumatic brain injury

Traumatic brain injury is defined as physical damage or impairment in function of the brain as a consequence of the application of acute mechanical force. Other causes of brain injury result from birth trauma, poisoning, or asphyxia. Traumatic brain injury is a major cause of death and disability among children, adolescents, and young adults, and is one of the most common causes of chronic brain syndromes in childhood. Traumatic head injury is common and becoming increasingly more so.

Clinical features

(a) Cognitive and behavioural

The most common long-term outcomes of traumatic brain injury are cognitive and behavioural changes. Immediately after emerging from a coma, the child will be unable to form new memories. The time, from the accident to the time when new memories emerge, is referred to as post-traumatic amnesia. The length of coma and the duration of post-traumatic amnesia are especially important in regard to the extent of cognitive recovery. Moreover, there is a strong inverse relationship between subsequent IQ and duration of coma. The persistence of cognitive deficits is correlated with the duration of post-traumatic amnesia; the more persistent deficits follow more than 3 weeks of post-traumatic amnesia. Persistent verbal memory impairment is reported as long as 10 years after injury in up to one-quarter of those studied. Psychiatric symptoms in adults occurs more often following focal frontal-lobe traumatic brain injury than injury to other cerebral areas. In children, Rutter[48,49] reported behavioural disinhibition after severe closed traumatic brain injury characterized by over-talkativeness, ignoring social conventions, impulsiveness, and poor personal hygiene.

(b) Psychiatric

Psychiatric outcomes can be divided into those that occur during the early phases of recovery and those that occur later. The earliest

psychiatric sequelae are found before the termination of post-traumatic amnesia. During this time, behavioural and affective symptoms are linked to the neurological presentation. The most common psychiatric diagnosis is delirium. Symptoms include short attention span, agitation, hallucinations, and disturbances in the sleep–wake cycle.

Subsequent occurrence of post-traumatic psychiatric symptoms is linked to the severity of the injury, its location, the child's behavioural and emotional features prior to the accident, and the psychosocial interactions of the family members during the recovery phases. The more severe the traumatic brain injury, the greater the likelihood of psychiatric sequelae. All children in one prospective study of severely injured children who had premorbid psychiatric conditions showed post-traumatic psychiatric disorders.[50,51] Moreover, over half the children in this group who had no premorbid symptoms prior to the accident had developed psychiatric symptoms during a 28-month, follow-up period. The greatest premorbid risks for psychiatric disorder were previous difficulties with impulse control and disruptive behaviour. In addition, a prior history of family dysfunction increased the risk for later symptomatology. The range of disorders includes attention-deficit hyperactivity disorder,[52] disruptive behaviour[53] post-traumatic mood disorders (both depressive and manic symptoms), post-traumatic stress disorder,[54] and family dysfunction.[55, 56] Transient psychotic features may occur. Hallucinations tend to be less bizarre and more concrete than the typical hallucination in schizophrenia. Moreover, head injury in childhood may accelerate the expression of schizophrenia in families where there is strong genetic predisposition[57] Injuries involving focal frontal-lobe dysfunction are associated with impulsive aggression and behavioural dyscontrol,[58] most often following focal orbitofrontal injury. The rate of actual aggression is less than often assumed. When forensic issues are considered regarding violent behaviour each case should be evaluated individually taking into account the type of head injury and other risk factors especially a history of physical abuse.

Classification—types of traumatic brain injury

Neurological damage associated with head trauma can be produced in several ways. Traumatic brain injury is classified as open or closed; these types differ in the pattern of injury and neurobehavioural outcome. Open refers to penetration of the skull, as in a depressed skull fracture or bullet wound, the extent depending on the regions damaged by contusion or cerebral oedema. Closed head injury results from acceleration and deceleration of the brain within the hard skull; this often leads to contusion of the brain from a sudden impact and may result in subarachnoid haemorrhage. Different parts of the brain have different densities, and therefore shearing stresses that develop during rapid brain movement cause injury. Furthermore, compression of blood vessels against the falx cerebri or tentorium may result in infarction of the areas, which these blood vessels supply. Penetrating traumatic brain injury causes specific and direct loss of neural tissue.

Epidemiology

It is estimated that 185 children per 100 000 from infancy to 14 years of age and 295 per 100 000 adolescents and young adults aged between 15 and 24 are hospitalized each year for traumatic brain injury.[59] The risk is highest among the 15- to 19-year-olds where the rate is 550 per 100 000.[60] The incidence in paediatric populations is similar to that in adults. In the United Kingdom the rates for those under 16 years is approximately 45 000, with about 300 deaths each year.[61] A mortality rate of 10 per 100 000 makes head trauma a major cause of death in children, but the death rate is still less than that in adults. There is no difference in the death rate between boys and girls before the age of 5 years, but after this age males are four times more likely to die than females. Approximately 90 per cent of head injuries are mild.[61] Falls and transport injuries make-up the majority of cases. Inflicted traumatic brain injury from physical abuse is a growing concern especially of repeated injuries. A US statewide population based study found the incidence of inflicted traumatic brain injury to be 17 per 100 000 person-years in the first 2 years of life with the highest incidence in infants during their first year (30/100 00). The rate was higher in boys than girls.[62]

Aetiology

The causes of traumatic brain injury are different depending on the age of the child. The incidence is twice as high in males as in females, and children who live in poor psychosocial circumstances are at greater risk. Traumatic brain injury from child abuse occurs in infancy: in the pre-school years the most common cause is falls; in early elementary school, it is pedestrian accidents. From 10 to 14 years of age there is an increase in sports and bicycle accidents, but by 15 years motor vehicle accidents and violent assault are the most common. Risk factors include poverty, single-parent homes, congested living arrangements, and a parental history of psychiatry disorder.

A common complication of traumatic brain injury is cerebral oedema, but there are other complications such as infection and haematoma formation both inside and outside the brain. These complications result in neurological deficits that may be extensive. Furthermore, compensatory mechanisms that are involved in recovery from head trauma may alter brain function. A child who has suffered a traumatic brain injury is likely to experience both neurological and psychiatric difficulties depending on the brain regions involved. Multiple mechanisms lead to psychological symptom formation—both psychosocial and physiological factors are involved.

Course and prognosis

The level of consciousness, degree of somatic injury, extent and duration of post-traumatic amnesia, severity of head injury, and degree of neurocognitive dysfunction in the early post-trauma period are important in determining outcome. Children who experience severe traumatic brain injury usually follow a predictable postoperative course.[48] As previously noted, landmarks for recovery are associated with the time of emergence from coma and the time of emergence from post-traumatic amnesia. The emergence from coma is most often defined as the point at which the patient is able to follow simple verbal commands. Concurrently, visual tracking of objects in the environment may be observed.

Post-traumatic amnesia ends when the child is able to form new memories. The frequency of post-traumatic amnesia is probably related to concurrent injury to the temporal lobes associated with the head trauma. However, older memories may be recalled that do not involve the temporal lobe. The hippocampus has a central role

in the formation of new memories. Besides recovery from post-traumatic amnesia, another form of memory loss-retrograde amnesia-for events that took place before the accident, typically becomes shorter and shorter during the recovery process. It is important to remember that children with severe head trauma will rarely have specific memories of the accident itself. Overall, the most important milestones in recovery for future outcome are the length of coma and the duration of post-traumatic amnesia.

Treatment

(a) Evidence

Most mild head injury and post concussive problems will resolve without treatment. When there are ongoing symptoms, parents and teachers must adjust expectations depending on the extent of injury. Complete recovery of all brain functions following severe brain injury is rarely accomplished. Still, if recovery is defined as a reduction in impairments in behavioural and physiological functions over time then changes do occur so that, typically, there is recovery of function together with a fair amount of substitution of function. Mechanisms include resolution of brain swelling (oedema), resolution of damage to other brain regions damaged through shock (diaschisis), changes in the structure of the nervous system (plasticity), and regrowth of neural tissue (regeneration). The extent of recovery depends on the severity of the injury, the number of times injured, the age at the time of injury, premorbid cognitive status, extent to which loss functions can be subsumed under other systems, integrity of other parts of the brain, individual brain structures, motivation, emotional considerations, and the quality of rehabilitation programme.[63, 64]

Although children and adolescents tend to have a better outcome after severe traumatic brain injury than those over the age of 21, the adult brain has greater plasticity than previously considered. Despite this general rule, children who are younger than 7 years may have a worse outcome since they may be at increased risk of child abuse as a cause of injury which may be particularly traumatic. Furthermore, younger children may have a worse outcome based on the global effects of trauma on the developing brain. The duration of recovery of significant neuropsychological, behavioural, and emotional deficits may last several years following injury. These higher cognitive deficits lead to the major disability observed with traumatic brain injury.

(b) Management

Partial recovery of function can and does occur over time, not only in children but also in adults. Intervention through retraining and the use of cognitive memory aids is targeted to improve areas of cognitive functioning such as memory, attention, language, and perception.[63] Even though partial recovery does occur after various types of brain injury, there is variability in the extent of recovery.

Possibilities for prevention

The most important primary injury prevention activities focus on teaching safe behaviour, the use of seat belts in cars, and wearing helmets when riding horses, bicycles, or motorcycles. Once an injury has occurred both anticipatory guidance, which teaches the family and child what to expect, and preventive intervention strategies are necessary. Early and focused rehabilitation procedures coupled with medication for associated psychiatric disorder, behaviour management, supportive therapy for families, and appropriate school programmes are necessary to prevent behavioural and psychiatric complications.

Epilepsy

Epilepsy refers to recurrent seizures that are idiopathic (of unknown aetiology) or due to congenital or acquired brain lesions. Epilepsy is the symptomatic expression of brain pathology or disordered brain function and is not a disease in itself. The symptom complex is episodic and associated with an excessive self-limiting neuronal discharge. The seizure is a frightening experience for parents and they require support and guidance. Epilepsy impacts the whole family and can create problems for all family members.[65]

Clinical features and clinical course

Complex partial seizures involving the temporal and frontal lobe is the most common condition where complex neurological and psychiatric symptoms are seen in the same person. Complex symptoms include behavioural automatisms, perceptual alterations, changes in affect and memory, distorted thinking, and hallucinations[66] Forms occurring in infancy and childhood include temporal lobe epilepsy, frontal-lobe epilepsy, infantile spasms, Lennox–Gastaut syndrome, Landau–Kleffner syndrome, and benign focal epilepsy.[67]

Children with partial seizures and electroencephalographic evidence of frontal involvement have more severe formal thought disorders and deficits in communication discourse than those with temporal involvement. Because these seizures are rare in children, reports of symptoms are primarily found in case reports. For example, Saygi *et al.*[68] and Stores *et al.*[69] have described sexual disinhibition, pressured and tangential speech, screaming, aggression, disorganized behaviour, and nightmares in affected children.

Frontal-lobe epilepsy should be considered if there are episodes of brief sudden unresponsiveness without loss of consciousness. These episodes occur with continued understanding of spoken language and clonic or tonic motor phenomena involving the face and arms bilaterally. Laughing, crying, pedalling movements, and sexual automatisms may also suggest this diagnosis. A normal electroencephalograph does not rule out the diagnosis. Left frontal hypometabolism on positron-emission tomography scanning or reduced cerebral blood flow to the frontal area, although not diagnostic, support this diagnosis.

The Lennox–Gastaut syndrome is characterized by early onset of intractable seizures and bilateral slow spike-wave complexes on the EEG.[70] The onset is typically between the ages of 1 and 7 years. The seizure pattern includes tonic, generalized tonic–clonic, atypical absence, atonic, and myoclonic seizures. Approximately half of children with the Lennox–Gastaut syndrome test as intellectually disabled. Marked language delay, overactivity, and irritability are characteristic. However, these behavioural symptoms may improve with seizure control. Ultimately, the diagnosis is based on the characteristic EEG finding of interictal slow spike-wave discharges in children with the early onset of poorly controlled seizures and a developmental disorder. In some instances, there is prolonged minor status epilepticus. Such episodes may last for several weeks during which the child engages in a variety of everyday activities but is socially unresponsive, aggressive, less articulate, and has

minor twitching of the face and hands. This presentation must be differentiated from a psychiatric disorder.

Classification

Classification of epileptic seizures utilizes both clinical and electroencephalographic features.[71] The current classification divides seizures into two categories: partial and generalized. Partial seizures involve one cerebral hemisphere, in part or totally. They begin focally, although they may become generalized. Consciousness is preserved but cognitive functions may be transiently impaired: for example, speech may be impaired if the dominant hemisphere is affected. Partial seizures are further subdivided into those with simple or complex symptomatology. In children, simple complex seizures are most often simple motor or sensory phenomena. Complex partial seizures usually begin in temporal or frontal-lobe structures. It is this group that is particularly important to psychiatrists.

Diagnosis and differential diagnosis

Epilepsy is a clinical rather than a laboratory diagnosis, and diagnostic errors most commonly occur due to inadequate history and physical examination. The accuracy of diagnosis has improved with the establishment of a universally agreed upon classification. In some instances there may be confusion between sleep arousal disorders and epilepsy.[72]

The differential diagnosis includes complex partial seizures of temporal lobe origin and pseudoseizures. Frontal-lobe complex partial seizures differ from those of temporal lobe origin in that the amnesia of frontal-lobe seizures is more pronounced than the extent of loss of consciousness. Moreover, frontal-lobe involvement is associated with unilateral or bilateral tonic posturing and pedalling movements, partial and not complete loss of consciousness, and eye and head deviation to the contralateral side. In complex partial seizures of temporal lobe origin, oroalimentary and repetitive hand automatisms, and looking around are characteristic. Lastly, sensory, gustatory, or olfactory hallucinations in frontal-lobe epilepsy must be differentiated from psychotic disorders such as schizophrenia and manic psychosis.

The distinction between true seizures and pseudoseizures can be difficult. Children with pseudoseizures commonly also have true seizures. Emotional dysphoria can precipitate true seizures and many children with chronic seizures have psychiatric diagnoses. Frontal-lobe seizures may be confused with pseudoseizures. Frontal complex partial seizures differ from pseudoseizures in that pseudoseizures have a gradual onset and longer duration, while frontal-lobe seizures start slowly and last less than 1 min. Pseudoseizures include thrusting or rolling movements rather than the rhythmic flexion and extension clonic movements seen in frontal-lobe epilepsy. Still, it may be difficult to distinguish pseudoseizures[73] and video and electroencephalograph monitoring with depth electrodes may be necessary to definitively diagnose frontal-lobe epilepsy.

Other features differentiating pseudoseizures are as follows:

1 The seizure occurs when the child is observed, but not when alone.

2 The seizures are gradual rather than of sudden onset.

3 Uncontrolled flailing occurs, rather than true tonic–clonic movements.

4 The seizure is accompanied by histrionics, with screaming and shouting.

5 Painful stimuli are avoided during an attack;

6 There is a sudden cessation of the seizure, with immediate return to an alert and responsive state.

7 There is absence of paroxysmal discharge during an attack on electroencephalography.[73]

Epidemiology

The incidence of epilepsy ranges from 0.7 to 1.1 per cent of the general population. It is the most common of the neurological diseases diagnosed in children.[67] Approximately 50 per cent of all cases of epilepsy begin during the childhood years and about 5 per cent of children will experience repeated epileptic seizures without a known extracerebral cause. In addition, about 3 per cent of children will have febrile convulsions that are usually benign and accompany a febrile illness. The great majority of these children, approximately 98 per cent, do not go on to develop true epilepsy. Other causes include hypoglycaemic seizures in children with diabetes. In some instances, as in tuberous sclerosis complex, cognitive impairment and autistic regression with onset in the first year of life, are linked to epilepsy; and in others there is a late onset of language disorder as in the Laudau–Kleffner syndrome.

The British Child and Adolescent Mental Health Survey[74] canvassed over 10 000 children and adolescents and identified 0.7 per cent of 5–15 year olds with a diagnosis of epilepsy. Among them there was an increased prevalence of emotional, behavioural, and relationship problems within families and among peers.

Aetiology

Advances in understanding epilepsy in childhood have come from the newer medical technologies. Recognition of a typical spike-wave pattern has led to the identification of benign focal epilepsy. CT scanning and high-resolution magnetic resonance imaging (**MRI**) have led to the recognition of mesial temporal sclerosis, tuberous sclerosis, neuroblast migrational disorders, and small temporal lobe tumours. Positron-emission tomography scanning can demonstrate lesions undetected by MRI, such as focal lesions in patients with hypsarrthymia. Advances in surgical procedures have decreased the risks associated with callosotomies and hemispherectomies used for catastrophic seizures. New understanding about neurotransmitters involved in the production and inhibition of seizures has led to advances in seizure medications.

Epileptic seizures are the result of an imbalance between inhibitory [(γ-aminobutyric acid (**GABA**) and excitatory (glutamate)] neurotransmitter systems. Neuronal hyperexcitability leading to seizures may result from decreased inhibition or increased excitation.[75] Epilepsy has its highest incidence in childhood, suggesting that the immature brain is more vulnerable to seizures than the mature brain—a finding that is borne out by animal studies. Decreased inhibition or increased excitation may result in neuronal excitability and seizures.[75] The specific mechanisms responsible for this imbalance remain uncertain. However, it is known that the binding of GABA to $GABA_A$ receptors opens a chloride channel (ionophore) leading to a flux of chloride ions and consequent membrane hyperpolarization: it is also known that there are fewer $GABA_A$ high-affinity receptors in immature animals. Similarly, there are maturational differences in the development of major

ionotrophic receptors in excitatory systems and in the activation of *N*-methyl-d-aspartate receptors. In younger animals this results in larger excitatory postsynaptic potentials. It remains a puzzle why certain seizure types are age-specific in their onset.[76]

Epilepsy syndromes may have a genetic basis.[77] Gene localization for five epilepsy syndromes with Mendelian inheritance are recognized, and localization has been suggested in three epilepsies with complex inheritance. Those epilepsies with a single gene inheritance include symptomatic epilepsies with associated diffuse brain dysfunction and idiopathic epilepsies, where the seizures are the primary brain abnormality. Idiopathic single gene epilepsies include benign, familial neonatal convulsions. To date four autosomal-dominant forms of epilepsy have been described. Most genes discovered to be involved in human epilepsies encode subunits of ion channels, both voltage-gated and ligand-gated.[77] Molecular genetic studies are expected to lead to the discovery of other epilepsy genes. Investigation of animal models of epilepsy is continuing.

The aetiology of temporal lobe seizures includes mesial temporal sclerosis, tumours, and cortical dysplasia. The younger the child, the less frequent is mesial temporal sclerosis. Other factors linked to aetiology are proposed: temporal lobe hypoperfusion and hypometabolism in Landau–Kleffner syndrome, and diffuse cortical and subcortical hypoperfusion in Lennox–Gastaut syndrome.

Course and prognosis

Early-onset epilepsies are associated with cognitive, behavioural, and communication disorders. Moreover, there is evidence that both clinical and subclinical epilepsy may result in developmental deviance, which has led to earlier and more aggressive treatment to try to prevent these impairments. Psychosocial factors are important in impairment. One prospective study evaluated 220 adults with childhood-onset seizures[78] up to age 35. The majority of subjects were free of seizures as adults, but were at risk for social and educational problems. When compared with random control subjects, those with epilepsy demonstrated correlations between neurological and cognitive impairment and social deficits. Those with epilepsy only (100 subjects) had a fourfold risk of psychiatric disorder. The authors reported social adjustment problems, competence problems, and reduction in marriage rate and fertility. Psychotic disorders occur significantly more frequently in people with epilepsy than in the general population with prevalence rates ranging from 2 to 8 per cent; the prevalence varying with the type of seizure disorder.[79]

Management

Cognitive and behavioural findings suggest the importance of early intervention to prevent negative outcomes. The behavioural and psychiatric problems should be treated with the same approach used in children who are neurologically intact and include educational, family, and pharmacological approaches. The indications and choice of psychiatric drugs is similar; epilepsy is not a strong contraindication for the use of neuroleptic or antidepressants, even though some of these medications may increase the frequency of seizures. Dexedrine may be the treatment of choice for hyperkinetic behaviour because it may increase the seizure threshold. Although caution is needed in those with more severe neurological involvement, there is no strong evidence for an increased risk for neuroleptic-induced tardive dyskinesia. When there are behavioural problems one must consider the behavioural effects associated with anticonvulsant medications.[80] In some instances, reducing the dose or changing the medication may be helpful and this should be discussed with the referring physician.

The major drugs used for treatment include carbamazepine, valproic acid, gabapentin, vigabatrin, and topiramate. These medications are used for the various forms of epilepsy described above including temporal lobe seizures and Lennox–Gastaut syndrome. Lamotrigine is also used, but with caution because severe dermatological side-effects may occur. In some instances, temporal lobectomy has been successful in the control of behavioural dysfunction and illogical thinking when performed in children with intractable temporal lobe seizures. In tuberous sclerosis complex the seizure medication vigabatrin may be helpful (and more so than corticosteroids) for infantile spasms.[81]

Possibilities for prevention

A developmental perspective is indicated as there is increasing evidence of there being a developmental period during which a structure or function can be developed most completely.[82] For example, in tuberous sclerosis complex the cognitive impairment and autistic regression may be approached by way of early drug therapy and, in some instances, by the surgical removal of tubers.[83] Thus a developmental understanding of epilepsy is now crucial in treatment planning. Research is continuing to clarify why, in some instances, epilepsy may have a severe developmental impact and in other instances be more benign. With greater understanding of genetic mechanisms appropriate family counselling will be needed, and perhaps, new drug treatments may emerge. An important treatment goal is to prevent adverse psychosocial outcome by correct diagnosis, early intervention for seizures, continual assessment for cognitive and behavioural disorders, appropriate schooling, as well as effective family support, guidance, and therapy. Careful prospective follow-up studies are needed to demonstrate which interventions are most appropriate to specific types of epilepsy.

Further information

Harris, J. (ed.) (1998). *Developmental neuropsychiatry: the fundamentals.* Oxford University Press, New York.

Harris, J. (ed.) (1998). *Developmental neuropsychiatry: assessment, diagnosis and treatment of the developmental disorders.* Oxford University Press, New York.

References

1. Harris, J. (1998). *Developmental neuropsychiatry: the fundamentals.* Oxford University Press, New York.
2. Harris, J. (1998). *Developmental neuropsychiatry: assessment, diagnosis and treatment of the developmental disorders.* Oxford University Press, New York.
3. Rapoport, J.L., Addington, A.M., Frangou, S., *et al.* (2005). The neurodevelopmental model of schizophrenia: update 2005. *Molecular Psychiatry*, **10**, 434–49.
   Harris, J. (2006). *Intellectual disability: understanding its development, causes, classification, evaluation, and treatment.* Oxford University Press, New York.
4. Spencer, T., Biederman, J., Harding, M., *et al.* (1998). Disentangling the overlap between Tourette's disorder and ADHD. *Journal of Child Psychology and Psychiatry*, **39**, 1037–44.
5. Carter, A.S., O'Donnell, D.A., Schultz, R.T., *et al.* (2000). Social and emotional adjustment in children affected with Gilles de la Tourette's

syndrome: associations with ADHD and family functioning. Attention deficit hyperactivity disorder. *Journal of Child Psychology and Psychiatry*, **41**, 215–23.
6. Rizzo, R., Curatolo, P., Gulisano, M., *et al.* (2007). Disentangling the effects of Tourette syndrome and attention deficit hyperactivity disorder on cognitive and behavioral phenotypes. *Brain and Development*, **29**, 413–20.
7. Flint, J. (1998). Behavioral phenotypes: conceptual and methodological issues. *American Journal of Medical Genetics*, **81**, 235–40.
8. Skuse, D.H. (2000). Behavioural phenotypes: what do they teach us? *Archives of Diseases of Childhood*, **82**, 222–5.
9. O'Brien, G. and Yule, W. (eds.) (1995). *Behavioural phenotypes. Clinics in developmental medicine*, No. 138. MacKeith, London.
10. Nyhan, W. (1972). Behavioral phenotypes in organic genetic disease. Presidential address to the society for pediatric research, May 1, 1971. *Pediatric Research*, **6**, 1–9.
11. Chapman, R.S. and Hesketh, L.J. (2000). Behavioral phenotype of individuals with Down syndrome. *Intellectual Disability and Developmental Disability Research Reviews*, **6**, 84–9.
12. Hagerman, R.J. (1999). Clinical and molecular aspects of fragile X syndrome. In *Neurodevelopmental disorders* (ed. H. Tager-Flusberg), pp. 27–42. MIT Press, Cambridge, MA.
13. Cassidy, S.B. and Morris, C.A. (2002). Behavioral phenotypes in genetic syndromes: genetic clues to human behavior. *Advances in Pediatrics*, **49**, 59–86.
14. Howlin, P., Davies, M., and Udwin, O. (1998). Syndrome specific characteristics in Williams syndrome: to what extent do behavioral patterns extend into adult life? *Journal of Applied Research in Developmental Disabilities*, **32**, 129–41.
15. Lesch, M. and Nyhan, W.L. (1964). A familial disorder of uric acid metabolism and central nervous system function. *The American Journal of Medicine*, **36**, 561–70.
16. Harris, J.C. (1998). Lesch–Nyhan disease. In *Developmental neuropsychiatry: assessment, diagnosis and treatment of the developmental disorders* (ed. J.C. Harris), pp. 306–19. Oxford University Press, New York.
17. Schretlen, D.J., Ward, J., Meyer, S.M., *et al.* (2005). Behavioral aspects of Lesch-Nyhan disease and its variants. *Developmental Medicine and Child Neurology*, **47**, 673–7.
18. Holland, A.J., Whittington, J.E., Butler, J., *et al.* (2003). Behavioural phenotypes associated with specific genetic disorders: evidence from a population-based study of people with Prader-Willi syndrome. *Psychological Medicine*, **33**, 141–53.
19. Dykens, E.M. and Kasari, C. (1997). Maladaptative behavior in children with Prader–Willi syndrome, Down syndrome and non-specific intellectual disability. *American Journal of Intellectual disability*, **102**, 228–37.
20. Gunn, P., Berry, P., and Andrews, R.J. (1981). The temperament of Down's syndrome in infants: a research note. *Journal of Child Psychology and Psychiatry*, **22**, 189–94.
21. van Lieshout, C.F.M., DeMeyer, R.E., Curfs, L.M.G., *et al.* (1998). Family contexts, parental behaviour, and personality profiles of children and adolescents with Prader–Willi, fragile-X, or Williams syndrome. *Journal of Child Psychology and Psychiatry*, **39**, 699–710.
22. Miller, L.K. (1999). The savant syndrome: intellectual impairment and exceptional skill. *Psychological Bulletin*, **125**, 31–46.
23. Hoyme, H.E., May, P.A., Kalberg, W.O., *et al.* (2005). A practical clinical approach to diagnosis of fetal alcohol spectrum disorders: clarification of the 1996 institute of medicine criteria. *Pediatrics*, **115**, 39–47.
24. Tischfield, M.A., Bosley, T.M., Salih, M.A., *et al.* (2005). Homozygous HOXA1 mutations disrupt human brainstem, inner ear, cardiovascular and cognitive development. *Nature Genetics*, **37**, 1035–7.
25. Glover, J.C., Renaud, J.S., and Rijli, F.M. (2006). Retinoic acid and hindbrain patterning. *Journal of Neurobiology*, **66**, 705–25.
26. American Academy of Pediatrics. (2000). Committee on substance abuse and committee on children with disabilities. Fetal alcohol syndrome and alcohol-related neurodevelopmental disorders. *Pediatrics*, **106**(2 Pt 1), 358–61.
27. Nash, K., Rovet, J., Greenbaum, R., *et al.* (2006). Identifying the behavioural phenotype in fetal alcohol spectrum disorder: sensitivity, specificity and screening potential. *Archives of Womens Mental Health*, **9**, 181–6.
28. Streissguth, A.P., Barr, H.M., Sampson, P.D., *et al.* (1994). Prenatal alcohol and offspring development: the first fourteen years. *Drug and Alcohol Dependence*, **36**, 89–99.
29. Olson, H.C., Streissguth, A.P., Sampson, P.D., *et al.* (1997). Association of prenatal alcohol exposure with behavioral and learning problems in early adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry*, **36**, 1187–94.
30. Steinhausen, H.C., Willms, J., Metzke, C.W., *et al.* (2003). Behavioural phenotype in foetal alcohol syndrome and foetal alcohol effects. *Developmental Medicine and Child Neurology*, **45**, 179–82.
31. O'Malley, K.D. and Nanson, J. (2002). Clinical implications of a link between fetal alcohol spectrum disorder and attention-deficit hyperactivity disorder. *Canadian Journal of Psychiatry*, **47**, 349–54.
32. Mukherjee, R.A., Hollins, S., and Turk, J. (2006). Fetal alcohol spectrum disorder: an overview. *Journal of the Royal Society of Medicine*, **99**, 298–302.
33. Streissguth, A.P. and O'Malley, K. (2000). Neuropsychiatric implications and long term consequences of foetal alcohol spectrum disorders. *Seminars in clinical neuropsychiatry*, **5**, 177–90.
34. Barr, H.M., Bookstein, F.L., O'Malley, K.D., *et al.* (2006). Binge drinking during pregnancy as a predictor of psychiatric disorders on the structured clinical interview for DSM-IV in young adult offspring. *The American Journal of Psychiatry*, **163**, 1061–5.
35. Sampson, P.D., Streissguth, A.P., Bookstein, F.L., *et al.* (1997). Incidence of fetal alcohol syndrome and prevalence of alcohol-related neurodevelopmental disorder. *Teratology*, **56**, 317–26.
36. Ikonomidou, C., Bittagau, P., Ishimaru, M.J., *et al.* (2000). Ethanol induced apoptotic neurodegeneration and fetal alcohol syndrome. *Science*, **287**, 1056–60.
37. Bookstein, F.L., Streissguth, A.P., Sampson, P.D., *et al.* (2002). Corpus callosum shape and neuropsychological deficits in adult males with heavy fetal alcohol exposure. *Neuroimage*, **15**, 233–51.
38. Spadoni, A.D., McGee, C.L., Fryer, S.L., *et al.* (2007). Neuroimaging and fetal alcohol spectrum disorders. *Neuroscience and Biobehavioural Reviews*, **31**, 239–45.
39. Warren, K.R. and Li, T.K. (2005). Genetic polymorphisms: impact on the risk of fetal alcohol spectrum disorders. *Birth Defects Research A Clinical and Molecular Teratology*, **73**, 195–203.
40. Chiriboga, C.A. (2003). Fetal alcohol and drug effects. *The Neurologist*, **9**, 267–79.
41. Hans, S.L. (1989). Developmental consequences of prenatal exposure to methadone. *Annals of the New York Academy of Sciences*, **562**, 195–207.
42. Rosen, T.S. and Johnson, H.L. (1985). Long-term effects of prenatal methadone maintenance. *NIDA Research Monograph*, **59**, 73–83.
43. Stewart, D.J., Inaba, T., and Lucassen, M. (1979). Cocaine metabolism: cocaine and norcocaine hydrolysis by liver and serum esterases. *Clinical Pharmacology and Therapeutics*, **25**, 464–8.
44. Chasnoff, I.J. (1992). Cocaine, pregnancy, and the growing child. *Current Problems in Pediatrics*, **22**, 302–21.
45. Singer, L.T., Minnes, S., Short, E., *et al.* (2004). Cognitive outcomes of preschool children with prenatal cocaine exposure. *The Journal of the American Medical Association*, **291**, 2448–56.
46. New England Congenital Hypothyroidism Collaborative. (1990). Elementary school performance of children with congenital hypothyroidism. *The Journal of Pediatrics*, **116**, 27–32.
47. Oerbeck, B., Sundet, K., Kase, B.F., *et al.* (2003). Congenital hypothyroidism: influence of disease severity and L-thyroxine treatment on intellectual, motor, and school-associated outcomes in young adults. *Pediatrics*, **112**, 923–30.

48. Rutter, M. (1981). Psychological sequelae of brain damage in children. *The American Journal of Psychiatry*, **138**, 1533–44.
49. Brown, G., Chadwick, O., Shaffer, D., *et al.* (1981). A prospective study of children with head injuries. III: psychiatric sequelae. *Psychological Medicine*, **11**, 63–78.
50. Max, J.E., Koele, S.L., Smith, W.L. Jr, *et al.* (1998). Psychiatric disorders in children and adolescents after severe traumatic brain injury: a controlled study. *American Academy of Child and Adolescent Psychiatry*, **37**, 832–40.
51. van Reekum, R., Bolago, I., Finlayson, M.A., *et al.* (1996). Psychiatric disorders after traumatic brain injury. *Brain Injury*, **10**, 319–27.
52. Max, J.E., Schachar, R.J., Levin, H.S., *et al.* (2005). Predictors of secondary attention-deficit/hyperactivity disorder in children and adolescents 6 to 24 months after traumatic brain injury. *Journal of the American Academy of Child and Adolescent Psychiatry*, **44**, 1041–9.
53. Max, J.E., Lindgren, S.D., Knutson, C., *et al.* (1998). Child and adolescent traumatic brain injury: correlates of disruptive behaviour disorders. *Brain Injury*, **12**, 41–52.
54. Max, J.E., Castillo, C.S., and Robin, D.A. (1998). Posttraumatic stress symptomatology after childhood traumatic brain injury. *The Journal of Nervous and Mental Disease*, **186**, 589–96.
55. Max, J.E., Castillo, C.S., and Robin, D.A. (1998). Predictors of family functioning after traumatic brain injury in children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, **37**, 83–90.
56. Wade, S.L., Taylor, H.G., Drotar, D., *et al.* (1998). Family burden and adaptation during the initial year after traumatic brain injury in children. *Pediatrics*, **102**, 110–16.
57. AbdelMalik, P., Husted, J., Chow, E.W., *et al.* (2003). Childhood head injury and expression of schizophrenia in multiply affected families. *Archives of General Psychiatry*, **60**, 231–6.
58. Brower, M.C. and Price, B.H. (2001). Neuropsychiatry of frontal lobe dysfunction in violent and criminal behaviour: a critical review. *Journal of Neurology, Neurosurgery, and Psychiatry*, **7**, 720–6.
59. Kraus, J.F. and Nourjah, P. (1988). The epidemiology of uncomplicated brain injury. *The Journal of Trauma*, **28**, 1637–43.
60. Kraus, J.F., Fife, D., and Conroy, C. (1987). Pediatric brain injuries: the nature, clinical course, and early outcomes in a defined United States population. *Pediatrics*, **79**, 501–7.
61. Middleton, J.A. (2001). Practitioner review: psychological sequelae of head injury in children and adolescents. *Journal of Child Psychology and Psychiatry*, **42**, 165–80.
62. Keenan, H.T., Hooper, S.R., Wetherington, C.E., *et al.* (2007). Neurodevelopmental consequences of early traumatic brain injury in 3-year-old children. *Pediatrics*, **119**, e616–23.
63. Limond, J. and Leeke, R. (2005). Practitioner review: cognitive rehabilitation for children with acquired brain injury. *Journal of Child Psychology and Psychiatry*, **46**, 339–532.
64. Rosen, C.D. and Gerring, J.P. (1986). *Head injury: educational reintegration*. College Hill Press, Boston, MA.
65. Ellis, N., Upton, D., and Thompson, P. (2000). Epilepsy and the family: a review of current literature. *Seizure*, **9**, 22–30.
66. Aicardi, J. (1992). *Diseases of the nervous system in children. Clinics in developmental medicine*, No. 115/118. MacKeith Press, London.
67. Guerrini, R. (2006). Epilepsy in children. *Lancet*, **367**, 499–524.
68. Saygi, S., Katz, A., Marks, D.A., *et al.* (1992). Frontal lobe partial seizures and psychogenic seizures: comparison of clinical and ictal characteristics. *Neurology*, **42**, 1274–7.
69. Stores, G., Zaiwalla, Z., and Bergel, N. (1991). Frontal lobe complex partial seizures in children: a form of epilepsy at particular risk of misdiagnosis. *Developmental Medicine and Child Neurology*, **33**, 998–1009.
70. Stephani, U. (2006). The natural history of myoclonic astatic epilepsy (Doose syndrome) and Lennox-Gastaut syndrome. *Epilepsia*, **47**(Suppl. 2), 53–5.
71. Commission on Classification and Terminology of the International League Against Epilepsy. (1989). Proposal for revised classification of epilepsies and epileptic syndromes. *Epilepsia*, **30**, 389.
72. Stores, G. (1991). Confusions concerning sleep disorders and the epilepsies in children and adolescents. *The British Journal of Psychiatry*, **158**, 1–7.
73. Stores, G. (1999). Practitioner review: recognition of pseudoseizures in children and adolescents. *Journal of Child Psychology and Psychiatry*, **40**, 851–7.
74. Davies, S., Heyman, I., and Goodman, R. (2003). A population survey of mental health problems in children with epilepsy. *Developmental Medicine and Child Neurology*, **45**, 292–5.
75. Scott, R.C. and Neville, G.R. (1998). Developmental perspectives on epilepsy. *Current Opinion in Neurology*, **11**, 115–18.
76. Holmes, G.L. (1997). Epilepsy in the developing brain: lessons from the laboratory and clinic. *Epilepsia*, **38**, 12–30.
77. Berkovic, S.F., Mulley, J.C., Scheffer, I.E., *et al.* (2006). Human epilepsies: interaction of genetic and acquired factors. *Trends in Neuroscience*, **29**, 391–7.
78. Jalava, M., Sillanpaa, M., Camfield, C., *et al.* (1997). Social adjustment and competence 35 years after onset of childhood epilepsy: a prospective controlled study. *Epilepsia*, **38**, 708–15.
79. Kanner, A.M. and Dunn, D.W. (2004). Diagnosis and management of depression and psychosis in children and adolescents with epilepsy. *Journal of Child Neurology*, **19**(Suppl. 1), S65–72.
80. Mula, M., Monaco, F., and Trimble, M.R. (2004). Use of psychotropic drugs in patients with epilepsy: interactions and seizure risk. *Expert Review of Neurotherapeutics*, **4**, 953–64.
81. Chiron, C., Dumas, C., Jambaque, I., *et al.* (1997). Randomized trial comparing vigabatrin and hydrocortisone in infantile spasms due to tuberous sclerosis. *Epilepsy Research*, **26**, 389–95.
82. Neville, B. (1998). The reemergence of critical periods of development. *Current Opinion in Neurology*, **11**, 89–90.
83. Bebin, E., Kelly, P., and Gomez, M. (1993). Surgical treatment for epilepsy in cerebral tuberous sclerosis. *Epilepsia*, **34**, 651–7.

9.2.2 Specific developmental disorders in childhood and adolescence

Helmut Remschmidt and Gerd Schulte-Körne

Introduction

The term 'specific developmental disorders' includes a variety of severe and persistent difficulties in spoken language, spelling, reading, arithmetic, and motor function. Skills are substantially below the expected level in terms of chronological age, measured intelligence, and age-appropriate education and cannot be explained by any obvious neurological disorder or any specific adverse psychosocial or family circumstances. As the deficits are quite substantial, analogies were initially made to neurological concepts and disorders such as word-blindness, alexia, aphasia, and apraxia, thus giving rise to the notion that neurological deficits are the aetiological basis of these disorders. Since this could not be demonstrated, the next step was to define the disorders in a more functional way,

Table 9.2.2.1 Specific developmental disorders: a comparison of ICD-10 and DSM-IV NOS, not otherwise specified

ICD-10	DSM-IV
Specific developmental disorders of speech and language (F80)	Communication disorders
Specific speech articulation disorder (F80.0)	Expressive language disorder (315.31)
Expressive language disorder (F80.1)	Mixed receptive–expressive language disorder (315.31)
Receptive language disorder (F80.2)	Phonological disorder (315.39)
Acquired aphasia with epilepsy (Landau–Kleffner syndrome) (F80.3)	Stuttering (307.0)
Other developmental disorders of speech and language (F80.8)	Communication disorder NOS (307.9)
Specific developmental disorders of scholastic skills (F81)	Learning disorders
Specific reading disorder (F81.0)	Reading disorder (315.00)
Specific spelling disorder (F81.1)	Mathematics disorder (315.1)
Specific disorder of arithmetical skills (F81.2)	Disorder of written expression (315.2)
Specific disorder of scholastic skills (F81.3)	Learning disorder NOS (315.9)
Other developmental disorders of scholastic skills (F81.8)	
Specific developmental disorder of motor function (F82)	Motor skills disorder
	Developmental co-ordination disorder (315.4)
Mixed specific developmental disorders (F83)	

taking into account not only psychometric testing but also psychosocial risk factors and the quality of schooling and education.

Today, numerous findings support the validity of the diagnostic concept of specific developmental disorders. These disorders and pervasive developmental disorders have the following features in common (ICD-10)[1]:

- An onset that invariably appears during infancy or childhood.
- An impairment or delay in the development of functions that are strongly related to biological maturation of the central nervous system.
- A steady course that does not involve the remissions and relapses that tend to be characteristic of many mental disorders.

Thus the term 'specific developmental disorders' reflects the fact that the deficits are circumscribed and relatively isolated against the background of an otherwise undisturbed psychological functioning.

Classification

In the multiaxial classification of child and adolescent psychiatric disorders,[2] specific developmental disorders are classified on the second axis named 'Specific disorders of psychological development', whereas pervasive developmental disorders are classified on the first axis (clinical psychiatric syndromes).

Based on the history and course of the disorders, two types can be distinguished:

- Disorders in which a phase of previously normal development has occurred prior to manifestation of the disorder. This, for example, applies to the Landau–Kleffner syndrome.
- An additional condition in which the abnormality was present from birth. This is especially true for autism. Autism is classified in the category 'pervasive developmental disorders', which are discussed elsewhere.

In DSM-IV,[3] nomenclature is somewhat different, but generally includes disorders identical or similar to those in ICD-10. In DSM-IV, 'communication disorders' correspond to 'specific developmental disorders of speech and language'. However, they also include stuttering, which is not included in the corresponding category of ICD-10.

'Learning disorders' (DSM-IV) is the category that corresponds to specific developmental disorders of scholastic skills, and 'motor skills disorder' to 'specific disorders of motor function'.

Table 9.2.2.1 shows the terminology used in both classification systems. The headlines of the two systems correspond; however, the subcategories show some differences.

Figure 9.2.2.1 shows a decision tree which includes the three main areas of dysfunction and addresses diagnosis and differential diagnosis.

Specific developmental disorders of speech and language

The main characteristic of these disorders is a disturbance of language acquisition from the early stages of development. The disturbance, however, is not directly attributable to neurological or speech mechanism abnormalities, sensory impairments, intellectual disability, or environmental factors.[1]

There are three main problems in distinguishing these disorders from the normal state and other conditions:

1 **Differentiation from the normal state**: the disorders must be distinguished from normal speech and language development, bearing in mind the great variations seen in the normal pattern. To make the diagnosis, the disorder must clearly be clinically significant, which can be determined by four main criteria: severity, course, pattern, and associated problems.

2 **Differentiation from intellectual disability (mental retardation)**: the degree of speech and language dysfunction must always be considered with respect to the child's cognitive level.

3 **Differentiation from disorders due to sensory impairment or impairments of the central nervous systems**: speech and language disorders resulting from severe deafness, specific

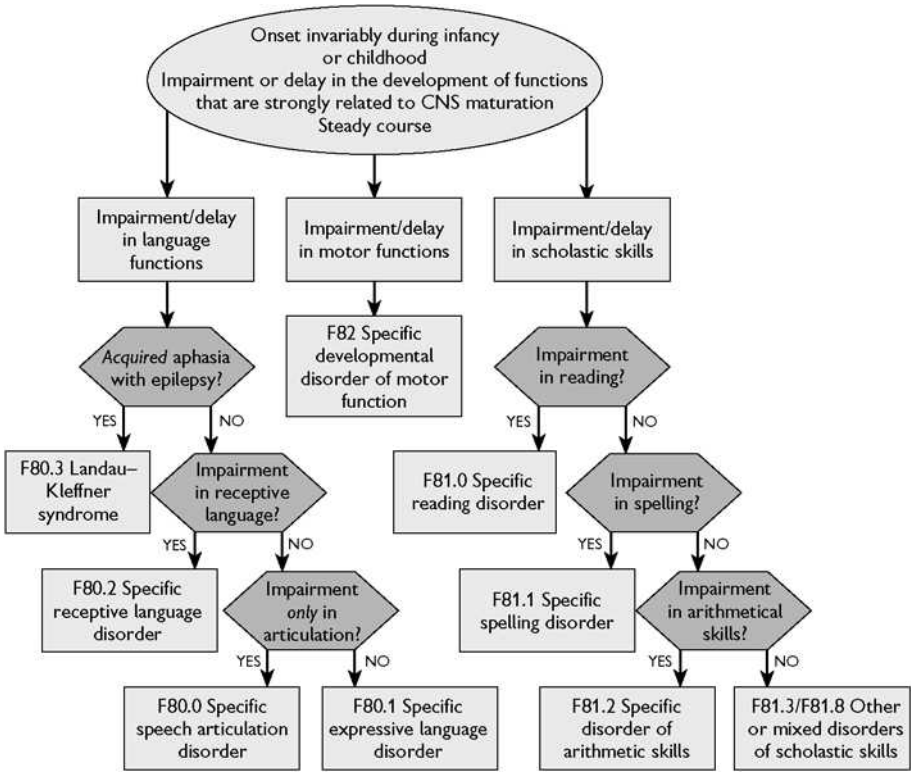


Fig. 9.2.2.1 Specific developmental disorders. CNS, central nervous system.

neurological impairments, or structural brain abnormalities are not classified in the category of specific developmental disorders of speech and language.

Specific speech articulation disorder

(a) Clinical features

The main feature of the disorder is the child's failure to use speech sounds appropriate for his or her mental age, while other language skills are within the normal range. Difficulties include errors in sound production and use, especially substitution of one sound for another. Difficulties in speech sound production usually interfere either with academic achievement or social communication. There are several degrees of severity reaching from mild or no impairment of speech intelligibility to completely incomprehensible speech. Sound substitutions are considered less severe than sound omissions. The sounds most frequently misarticulated are those acquired later during speech development (l, r, s, z, th, ch). However, consonants and vowels that range early in the development sequence may be affected in younger children.

It is very important to relate the misarticulations to normal development. At the age of 4 years, errors in speech and sound production are very frequent, but children are usually understood even by strangers. At 6 to 7 years of age, most speech sounds can be adequately reproduced, and by the age of 11 to 12, children should be capable of almost all speech sounds.

(b) Classification

In ICD-10, speech articulation disorder (F80.0) is classified in the category 'Specific developmental disorders of speech and language'. The counterpart in DSM-IV is the category 'Phonological disorder' (315.39), classified in the category 'Communication disorders' (see Table 9.2.2.1).

(c) Diagnosis and differential diagnosis

The leading feature, the age-appropriate misarticulation of speech sounds with the result that others have difficulties in understanding the child, usually allows one to diagnose the disorder. There are three types of symptoms that can be observed: substitutions, omissions, and distortions of speech sounds. The diagnosis should only be made if the severity of misarticulation is outside the limits of the normal variation for the child's mental age. Further requirements are that non-verbal intelligence and expressive and receptive language skills should be within the normal range.

Differential diagnoses include intellectual disability, hearing impairment, or other sensory deficits or severe environmental deprivation.

(d) Epidemiology

Moderate to severe developmental articulation disorders can be found in 2 to 3 per cent of 6- and 7-year-old children, with less severe disorders even more frequent. The frequency of occurrence falls to 0.5 per cent by the time children are 17 years old (DSM-IV).

(e) Aetiology

As it has been demonstrated that the disorder runs in families, it is assumed that genetic factors are important for its manifestation.

(f) Course and prognosis

The prognosis is favourable if no other associated features such as hearing impairment, neurological conditions, cognitive impairments, or psychosocial problems are present. However, the course

varies depending on the severity and the above-mentioned associated features.

(g) Treatment

Treatment is necessary and appropriate if the child is handicapped in his or her everyday life and cannot be understood by parents, siblings, or other persons. The focus of the therapy depends on whether speech articulation disorder is an isolated phenomenon or if other impairments or dysfunctions are present (e.g. developmental disorder of motor functions). If it is an isolated phenomenon, functional speech therapy can be carried out on the principle that mispronounced sounds should not be repeated when correcting them, but substituted by the correctly pronounced sound.[4] It is advisable to carry out this kind of therapy before the child enters school. If other disorders are present, a comprehensive therapeutic programme that includes speech therapy needs to be developed.

Expressive language disorder

(a) Clinical features and classification

The main feature of this disorder is that the child's ability to use expressive spoken language is reduced below the mental age appropriate level, while language comprehension ranges within normal limits. Abnormalities in articulation may co-occur.

In ICD-10, the following symptoms are considered important for diagnosis (ICD-10, p. 237)[1]:

- delay of the development of expressive language (e.g. absence of single words by the age of 2 years, failure to generate simple two-word sentences by 3 years)
- restricted vocabulary development
- overuse of a small set of general words
- difficulties in selecting appropriate words and word substitutions
- short utterance length and immature sentence structure
- syntactical errors, especially omissions of word endings or prefix
- misuse of or failure to use grammatical features such as prepositions, pronouns, articles, and verb and noun inflexions.

The DSM-IV criteria requires measures of expressive language development being substantially below those obtained from standardized measures of both non-verbal intellectual capacity and receptive language development, interference with academic or occupational achievement, and the exclusion of mixed receptive–expressive disorder and pervasive developmental disorders.

(b) Diagnosis and differential diagnosis

The diagnosis is made by clinical observation, with special emphasis on expressive language functions and the use of individually administered standardized tests of expressive language. The differential diagnosis should rule out mixed receptive–expressive language disorder (DSM-IV), characterized by an impairment of receptive language functions. Autistic disorder may also involve expressed language impairment, but autism can be distinguished by characteristic communication impairments. Finally, intellectual disability and sensory impairments (e.g. hearing impairment or other sensory deficits) need to be ruled out, as well as severe environmental deprivation. The diagnosis is confirmed using intelligence tests, audiometric tests, neurological investigations, and a careful history. Finally, acquired aphasia needs to be ruled out. This can be done by assessing any medical condition that may have caused the disorder.

(c) Epidemiology

In the absence of thorough epidemiological studies, estimates suggest that between approximately 3 and 5 per cent of children may be affected by expressive language disorder of the developmental type. The acquired type seems to be less common.

(d) Aetiology

DSM-IV distinguishes two types of expressive language disorders: the developmental type and the acquired type. In the developmental type, impairment of expressive language begins at a very early age and is not associated with neurological factors, while the acquired type occurs after a period of normal development and is caused by neurological or general medical conditions (e.g. head trauma, encephalitis). It is assumed that the developmental type is caused by genetic factors that influence language development.

(e) Course and prognosis

The course depends on the type of disorder (developmental or acquired type) and severity. Usually, the disorder can be diagnosed by the age of 3 years, while milder forms are often only detected later. According to DSM-IV,[3] approximately half of the children appear to outgrow the developmental type of expressive language disorder, while the other half have persistent difficulties. The outcome of the acquired type depends on the severity and location of the brain pathology.

(f) Treatment

As causal treatment is not possible, treatment measures are based on general principles that have been found to be useful and effective in clinical practice.

1 The first step is to explain clearly to parents the nature of the disorder and the fact that several other disturbances manifested by the child may be a result of the child's communication deficit.

2 The best time to commence speech therapy depends upon the severity of the disorder, the child's cognitive and motivational structure, and other disorders that might be present. Instead of treating children too early (e.g. before the age of 3 years), offering advice and guidance to the parents is extremely important.

3 Treatment itself concentrates on teaching language skills using techniques such as imitation and modelling. The therapist should focus interventions selectively on the areas of difficulty, thus increasing the child's phonological repertoire. Non-verbal communication techniques may be used if verbal communication is substantially impaired. But the therapist should always make sure that non-verbal communication does not dominate the verbal one.

4 In therapeutic programmes, everyday situations are now preferred to very structured programmes. This is because many therapists found that therapeutic progress during sessions was not transferred to everyday life situations. During structured treatment sessions the children are taught to give correct answers to questions that have nothing to do with their situation in everyday life, and it is now thought that structured language training may prevent them using language according to their needs.[5]

5 Alternative communication, such as sign language, should only be used if the child suffers from severe auditory comprehension deficits. The use of a sign language, however, is no longer regarded as an obstacle to the improvement of expressive language skills.[6]

Receptive language disorder

(a) Clinical features and classification

This disorder is characterized by the child's inability or reduced ability to understand language in a way appropriate for his or her mental age. As expressive language production depends on language comprehension; expressive language is also profoundly disturbed and abnormalities in word-sound production can be observed.

The diagnostic guidelines of ICD-10 include the following features:

- failure to respond to familiar names (in the absence of non-verbal clues) by the first birthday
- inability to identify at least a few common objects by 18 months
- failure to follow simple, routine instructions by the age of 2 years
- inability to understand grammatical structures (e.g. questions, comparatives)
- lack of understanding of the more subtle aspects of language (tone of voice, gestures, etc.).

Owing to the disturbances in both receptive and expressive functions, the disorder is called 'receptive–expressive language disorder' in DSM-IV. The diagnostic criteria require scores of both receptive and expressive language development substantially below those obtained from standardized measures of non-verbal intellectual capacity, interference with academic or occupational achievement, and exclusion of pervasive developmental disorders.

(b) Diagnosis and differential diagnosis

Diagnosis is based on three factors: a careful history taken from the child's parents, a thorough clinical investigation including neurological assessment and detailed speech and language assessment, and standardized tests measuring expressive and receptive language functions.

Differential diagnosis should rule out expressive language disorder (which is the case in the presence of language comprehension), specific speech articulation disorder, in which the receptive and expressive language functions are unimpaired, autism (which can be distinguished by the typical communication disturbance), intellectual disability, sensory deficits, and severe environmental deprivation. These disorders can be excluded by intelligence tests, audiometric tests, neurological investigations, and taking a history.

(c) Epidemiology

Owing to the absence of epidemiological studies, the frequency with which the disorder occurs can only be estimated. According to estimations, the disorder occurs in up to 3 per cent of school-age children and is probably less common than expressive language disorder.

(d) Aetiology

As in other developmental language disorders, there is evidence that genetic factors play the most important role in aetiology.[7] The frequent association of disturbed language acquisition with adverse psychosocial factors in the family does not contradict a primarily genetic cause, as many children who grow up under these circumstances show entirely normal developmental patterns of speech and language skills.[8]

(e) Course and prognosis

The long-term prognosis is poor. Only half the patients in the sample studied by Rutter *et al.*[9] had normal conversational skills when they were in their twenties, and there was a decline in non-verbal IQ from childhood to adulthood. The course again depends on the type (developmental or acquired) and severity of the disorder. The disorder is usually detected before the age of 4 years, but earlier in severe cases. The prognosis is poorer than in expressive language disorder. As far as the acquired type is concerned, the prognosis varies depending on severity, location of brain pathology, the child's age, and the level of language development prior to the disorder.

(f) Treatment

Treatment is generally undertaken along the same lines as in expressive language disorders. However, owing to the nature of the disorder, all factors that facilitate language comprehension should especially be encouraged. Non-verbal forms of communication such as sign language can be helpful.

Acquired aphasia with epilepsy (Landau–Kleffner syndrome)

(a) Clinical features and classification

The Landau–Kleffner syndrome is a rare disorder characterized by receptive and expressive language impairment and epileptic seizures, but retained general intelligence, and manifestation after a period of normal development, including language development. The onset of the disorder typically occurs between 3 and 7 years of age and is accompanied by paroxysmal electroencephalographic abnormalities, mainly bilateral spikes in the posterior temporal and parietal regions and epileptic seizures in about 80 per cent of cases.[10] The loss of language may occur gradually over a period of months or abruptly within a few days or weeks.

Aphasia usually starts with receptive language problems occurring together with the characteristic electroencephalographic changes, followed by expressive language difficulties. Usually, the first sign is the impairment of receptive language functions with difficulties in auditory comprehension. During the manifestation period, symptomatology is variable: some children become mute, others express jargon-like sounds, and some produce misarticulations and have difficulties in word fluency. During the manifestation period, emotional and behavioural symptoms are common; they can be regarded as a reaction to the loss of language functions and appear as anxiety reactions, acting-out behaviour, and aggression.

In DSM-IV, the condition is classified in the category mixed receptive–expressive language disorder, acquired type.

(b) Diagnosis and differential diagnosis

The diagnosis can be based upon a detailed history of the child's development, assessment of language functions, careful neurological assessment, and by electroencephalography. The differential diagnosis includes other types of acquired aphasia without epileptic seizures and electroencephalographic abnormalities, and

disintegrative disorders of childhood such as dementia infantilis Heller (Heller's syndrome) (see Chapter 9.2.2).

(c) Epidemiology

The prevalence of the disorder is unknown.

(d) Aetiology

The aetiology is unknown. There is, so far, no indication of any genetic cause. Clinical characteristics suggest that an encephalitis might be considered a causal mechanism. The electroencephalographic changes and the seizures are thought to cause the language and the behavioural and emotional problems.

(e) Treatment

So far, three different approaches to treatment have been undertaken:

1 Anticonvulsant treatment (mainly with carbamazepine): the frequency of seizures can be reduced to some extent with this medication, but paroxysmal electroencephalographic changes are not substantially influenced.

2 Corticosteroids have also been administered,[11] but the benefits remain unclear.

3 Finally, surgical treatment by bilateral temporal transection has been attempted.[12,13]

Specific developmental disorders of scholastic skills

'Specific developmental disorders of scholastic skills' (ICD-10) or 'Learning disorders' (DSM-IV) include disorders characterized by one or more significant impairments in acquisition of reading, spelling, or arithmetical skills. ICD-10 suggests that the category 'Mixed disorder of scholastic skills' (F81.3) be used as an ill-defined, but necessary, category in which arithmetical and reading or spelling skills are significantly impaired, although not because of general intellectual disability or inadequate schooling.

The disorders classified in the category 'Specific developmental disorders of scholastic skills' (**SDDSS**) resemble specific disorders of speech and language. As in these latter disorders, normal patterns of skill acquisition are disturbed and detectable at an age when these functions are required. The disorders are not due to a lack of opportunity to learn or a consequence of brain trauma or disease, but represent a specific type of dysfunction in cognitive processing. The dysfunction affects specific skills, which can be distinguished from the cognitive functions that are usually in the normal range. As in other specific developmental disorders, the condition is more common in boys than in girls.

ICD-10 notes five difficulties regarding diagnosis and differential diagnosis:

1 differentiation of the disorder from normal variations in scholastic achievement (this problem applies to all specific developmental disorders and was discussed in relation to specific developmental disorders of speech and language);

2 consideration of the normal developmental course;

3 interference with learning and teaching;

4 underlying abnormalities in cognitive processing;

5 uncertainties over the best way of subdifferentiating SDDSS.

Based on these considerations, the following diagnostic guidelines for all SDDSS have been suggested (ICD-10):

- **Clinically significant degree of impairment:** this is judged on the basis of severity (e.g. occurrence in less than 3 per cent of school children), developmental precursors (e.g. speech or language disorder in preschool years), and associated problems (e.g. inattention).
- **Specific impairment not explained solely by intellectual disability or by lesser impairments in general intelligence:** for this requirement to be met, individually administered and standardized IQ scholastic achievement tests are obligatory to demonstrate that the child's level of achievement is substantially below the expected level compared to a child of the same mental age and IQ.
- **Developmental nature of the impairment:** this must be demonstrated by the presence of the disorder during the early years of schooling and by exclusion of impairment acquired later. The child's history of school progress is decisive in this respect.
- **Absence of external factors that could explain the impairment:** SDDSS are thought to be mainly based on factors intrinsic to the child's development, and not due to inadequate schooling or any other environmental factors such as absence from school or educational discontinuities. However, such conditions may occur, making the diagnostic process difficult.
- **Exclusion of visual and hearing impairments:** by definition, SDDSS do not occur as a result of impairment of sensory function, such as visual or hearing impairment.

The main differential diagnostic task is distinguishing SDDSS from neurological disorders (e.g. alexia, aphasia, agraphia, apraxia) or impairments that also could influence the development of scholastic skills (e.g. emotional disorder). In cases of normal child development prior to the manifestation of a defined neurological disorder, differential diagnosis is not difficult. However, if minor neurological signs (soft signs) were diagnosed previously, independent of any defined disorder, and the findings persist, it may be difficult to distinguish recent symptoms from previous ones. In such cases, associated disorders or symptoms should be classified separately in the appropriate neurological section of the classification systems.

Specific reading disorder

The ICD-10 classification system distinguishes between 'Specific reading disorder' and 'Specific spelling disorder'. In DSM-IV, 'Specific reading disorder' is distinguished from 'Disorder of Written Expression'. The latter is not identical with the ICD-10 category 'Specific spelling disorder', insofar as that disorder excludes children whose sole problem is one of handwriting.

(a) Clinical features and classification

The main feature of this disorder is a specific and significant impairment in the development of reading skills, which is not solely accounted for by mental age, visual acuity problems, or inadequate schooling.[1]

Other functions may also be affected:

- word recognition
- reading comprehension skills

- oral reading skills
- performance of tasks requiring reading

In many cases, spelling difficulties continue into adolescence and persist in adulthood, even when reading skills improve considerably. The history of children with specific reading disorder frequently reveals a specific developmental disorder of speech and language. Symptoms of these disorders may still be present at elementary school when the specific reading disorder is first diagnosed. Additional frequently associated problems include poor school attendance and problems with social adjustment.

The DSM-IV criteria for reading disorder state that reading achievement as measured by standardized tests should be substantially below the expected level and that the disturbance should interfere with academic achievement or activities of daily living that require reading skills.

(b) Diagnosis and differential diagnosis

The diagnosis is made on the basis of the ICD-10 and/or DSM-IV criteria, which are similar. The ICD-10 diagnostic guidelines require the following:

- A reading performance below the level that is expected on the basis of age, general intelligence, and school placement. For clinical purposes, usually 1.5 standard deviations below the expected level is regarded as a requirement for the diagnosis. For research purposes, two standard deviations are used.
- Performance to be assessed by individually administered standardized tests of reading accuracy, latency, and comprehension.
- Errors demonstrated in oral reading skills and deficits in reading comprehension. Errors in oral reading include:
  - (a) omissions, substitutions, distortions, or additions of words or parts of words;
  - (b) slow reading rate;
  - (c) false starts, long hesitations, or 'loss of place' in text, and inaccurate phrasing;
  - (d) reversals of words in sentences or of letters within words.

Deficits in reading comprehension include:

- (a) an inability to recall facts that have been read;
- (b) inability to draw conclusions or interferences from material that has been read;
- (c) use of general knowledge as background information, rather than of information from a particular story, to answer questions about a story that has been read.

(c) Comorbidity and associated features

It is important to analyse the features of the disorder with a longitudinal perspective. Thus, several associated disorders can be observed: emotional problems during the early school years; hyperactivity and conduct disorders in later childhood and adolescence. Additional frequent features include low self-esteem, adjustment problems at school, and problems in peer relationships. In about 40 per cent of children with reading and/or spelling disorder, other disorders of clinical relevance are present. After finishing school, this rate decreases to 30 per cent, which includes a high proportion of antisocial behaviour and delinquency.[14]

(d) Epidemiology

Specific reading and spelling disorder occur in about 4 per cent of 8- to 10-year-old children, when defined as two standard deviations below non-verbal IQ.[15,16] By using a wider definition, the rates are approximately 7 to 8 per cent, with a predominance of boys (2:1).

(e) Aetiology

Currently, four main aetiological factors have been discussed:

1 genetic influences;

2 deficits in central information processing;

3 general psychosocial factors;

4 specific learning conditions.

Familial clustering in dyslexia was recognized a few years after the first description of the disorder by Hinselwood in 1895. A child with an affected parent has a risk of between 40–60 per cent of developing dyslexia.[17] This risk is increased when other family members are also affected. There is estimated to be a 3 to 10-fold increase in the relative risk for a sibling (λ_s), with an increase in λ_s being observed when stricter criteria are applied.[18]

Twin studies have confirmed that genetic factors are substantially responsible for the familial clustering of dyslexia. It is generally accepted that the proportion of inherited factors involved in the development of dyslexia is between 40–80 per cent, the highest estimates being reported for the phenotype dimensions word reading (up to 58 per cent) and spelling (70 per cent).[19] Whereas shared environmental effects are low for word reading, for with reading, and spelling correlated traits, for example, phonological awareness, shared environmental is substantially higher at about 14 per cent. Based on genome-wide linkage analyses nine candidate gene regions (DYX1-9) could be identified. Most replicated are two regions, 6p22, and 15q21. More recently, four candidate genes, DCDC2, KIAA0319, ROBO1, and DYX1C1 were identified by systematic association analyses. All these genes play a functional role in neuronal migration making them promising candidate genes for dyslexia. However, a functional relevant mutation has not been identified yet.[19]

The hypothesis of deficits in central information processing is based on behavioural and neuroimaging studies that identified cognitive processes that are impaired in dyslexics individuals. These are impaired visual processing, auditory processing, speech perception, phonologic processing, orthographic processing, and motor coordination.[20] The mainstream hypothesis is a phonological processing deficit supported by longitudinal, intervention, and brain imaging studies. Among the latter, the importance of left hemispheric specialization has been widely discussed in the literature, either suggesting maturational lag of the left hemisphere or a structural deficit in white and grey matter.[20]

General psychosocial factors may also play a role in the manifestation of specific reading and spelling disorder. However, the influence seems to be rather marginal. The same applies, more or less, to the special learning condition, because severe deficits in schooling are excluded by definition as a main cause of these disorders. However, given a genetic predisposition for specific reading or spelling disorder, poor learning conditions at school and at home may contribute to the manifestation of these disorders. In summary, the different factors responsible for the manifestation of specific

reading and specific spelling disorder can be best understood in terms of a vulnerability model, in which several genetic predispositions, on the one hand, and general psychosocial factors and special learning conditions, on the other, interact with one another.

(f) Course and prognosis

More than 40 longitudinal follow-up studies have shown similar results:

- There is a high persistence of reading and spelling problems, phonological difficulties, and slowness in word recognition.[23,24]
- Retarded readers make poorer progress than backward readers.[25]
- On the other hand, considerable progress is possible in oral reading and reading comprehension.[26]

As far as schooling is concerned, a substantial proportion of children with specific reading disorder remain behind the school level of their age group. However, there is a difference regarding social background. Children from middle-class homes more frequently show a positive educational outcome compared with children from socially disadvantaged homes.[27]

An epidemiological study in Germany[16] showed that only 3 per cent of children with specific reading disorder were able to attend high school and more than 50 per cent remained at the lowest normal school level. At the age of 18, the rate of unemployment was three times as high as in a normal control group.

(g) Treatment

As mentioned above, treatment is difficult because the disorder tends to persist. Nevertheless, the following principles have been found to be useful:

1 Treatment should start as early as possible in order to avoid a sense of failure and low self-esteem.

2 The treatment should focus on individual instruction and teaching sessions in basic phonetic and other skills such as reading, spelling, and writing. This needs to be done in an age-appropriate way based on the principles of learning theory and starting at a very low level to avoid disappointment and a sense of failure. There are several programmes used in different countries based on these principles, sometimes using computers.

3 Although very popular methods based on training basic perceptual skills like figure-ground discrimination, tone discrimination, temporal auditory processing have not been proven by empirical studies.

4 Even when feelings of failure and, consequently, low self-esteem are present, the instruction in basic skills is the appropriate approach. The child's psychological and learning situation deserves special attention. Psychotherapeutic measures alone are not successful.

5 Parental support is extremely important. Therefore, the parents should not only be educated in detail about the disorder, but also encouraged to listen to their children reading from school books. This has been shown to be a successful approach.[28]

6 There is no specific medication to improve reading and spelling skills, but there is some indication that stimulants may be helpful for poor readers who simultaneously suffer from attention-deficit hyperactivity disorder.

Specific spelling disorder

(a) Clinical features

In ICD-10, the main characteristic of this disorder is a specific and significant impairment in the development of spelling skills in the absence of a history of specific reading disorder, which is not solely accounted for by low mental age, visual acuity problems, or inadequate schooling. The children have difficulties in spelling orally and writing words correctly. For this diagnosis, the following criteria are required (ICD-10):

- The spelling performance of the child should be significantly below the expected level regarding age, general intelligence, and school placement. This has to be assessed individually by administration of a standardized spelling test.
- The reading skills should be within the normal range and there should be no history of preceding reading difficulties.
- The spelling difficulties should not be due to grossly inadequate teaching, to sensory deficits, or to neurological disorders or dysfunctions. They should not be acquired, either as a result of neuropsychiatric or any other disorders.

In DSM-IV, there is no category that corresponds exactly to the ICD-10 category 'Specific spelling disorder'. The DSM-IV category that most closely resembles 'Specific spelling disorder' is 'Disorder of written expression', as defined by measured writing skills substantially below the expected level, interference with academic achievement, and, in the case of a sensory deficit, writing skills greater than the difficulties usually associated with sensory deficits.

It is uncertain whether a pure spelling disorder as described in ICD-10 actually exists or whether spelling skills usually overlap with other functions that constitute scholastic skills. It is, however, possible to assess the different functions separately, as several studies show.

(b) Diagnosis and differential diagnosis

The diagnosis is made according to the criteria mentioned above and by the administration of standardized spelling tests.

Specific disorders of arithmetical skills

(a) Clinical features

The main clinical feature of this disorder (also called dyscalculia) is a specific impairment in arithmetical skills that cannot be explained on the basis of general intellectual disability or inadequate schooling. Dyscalculia is a difficulty in learning and remembering arithmetic facts and executing calculation procedures, with immature problem-solving strategies, long solution times, and high error rates.[29] A number of different skills may be impaired, as understanding or naming mathematical terms, operations, or concepts, and decoding written problems into mathematical symbols.

The impairment affects basic computational skills of addition, subtraction, multiplication, and division, whereas other functions such as reading and writing or motor skills are within the normal range (except in mixed disorder of scholastic skills). The arithmetical difficulties vary, but in most cases include the following features (ICD-10):

- difficulties in understanding the concepts underlying arithmetical operations
- difficulties or lack of understanding of mathematical terms or signs
- difficulties in recognizing numerical symbols

- difficulties in carrying out arithmetical manipulations
- difficulties in aligning numbers or symbols when performing calculations
- poor spatial organization of arithmetical calculations
- reduced ability to learn multiplication tables satisfactorily.

The diagnosis is made according to ICD-10 (Specific disorders of arithmetical skills) or DSM-IV criteria (Mathematics Disorder) and the diagnostic guidelines. The ICD-10 guidelines require the following criteria:

- The child's arithmetical performance should be significantly below the expected level on the basis of age, general intelligence, and school placement, assessed by an individually administered standardized arithmetical test.
- Reading and spelling skills should be within the normal range expected for the child's mental age, also tested by an individually administered standardized test.
- The difficulties in arithmetical skills should not be mainly due to grossly inadequate teaching or the direct effects or defects of visual, hearing, or neurological functions and should not be acquired as sequelae of neurological, psychiatric, or other disorders.

The DSM-IV criteria follow the same principles as with other specific developmental disorders. Required criteria include mathematical ability substantially below the expected level, interference with academic achievement, and in the case of a sensory deficit, the difficulties in mathematical ability greater than the difficulties usually associated with sensory deficits.

(b) Diagnosis and differential diagnosis

The diagnosis is made according to the above-mentioned criteria in the ICD-10 and DSM-IV systems. Differential diagnosis must rule out acquired arithmetical disorder (acalculia), arithmetical difficulties associated with a reading or spelling disorder, and arithmetical difficulties as a result of inadequate teaching.

(c) Epidemiology

It is estimated that between 3 and 6 per cent of all schoolchildren suffer from a specific arithmetical disorder.[30] The sex ratio is approximately equal, in some studies girls are found to be more often affected than boys.[31] An important correlate of maths disorder is dyslexia. It is estimated that 40 per cent of dyslexics also have maths disorder.[31]

(d) Aetiology

Dyscalculia has not been studied with the same intensity as dyslexia. Therefore, knowledge about aetiology, course, and outcome is limited.

Numerical abilities, including arithmetic, are mediated by areas in the parietal lobe. In functional imaging studies performed during mental calculation tasks, a pattern of bilateral activation in the prefrontal, premotor, and parietal cortices has been observed.[32] Neuropsychological evidence indicates that numerical processing is localized to the parietal lobes bilaterally, in particular the intra-parietal sulcus[33] and is independent of other abilities. A reduced volume of grey cortical matter in dyscalculic individuals was found in the sulcus intraparietalis of the left hemisphere.[34,35]

There have been a few studies into genetic contributions to mathematical cognitive ability. Most have studied the possible genetic aetiology of mathematics disorder, at least partly because of its comorbidity with reading disorder.[36]

Results from twin studies were consistent with a genetic basis for mathematics disorder whether combined with reading disability or not and estimates of high heritability of mathematical ability were obtained in a sample of twins with normal intelligence ascertained for reading disability and family samples.[37,38]

So far developmental dyscalculia is likely to be the result of the failure of these brain areas to develop normally and can best be defined as a deficit in the representation or processing of specifically numerical information.[39]

A substantial proportion of children with specific disorder of arithmetical skills have associated emotional problems and difficulties in social interactions. It is not quite clear whether these difficulties are secondary complications of the specific difficulties of arithmetical skills.

(e) Course and prognosis

As a matter of fact, specific disorder of arithmetical skills is only diagnosed at the end of the first year or the beginning of the second year of elementary school, because of the necessity of these skills at that time. Especially in cases when the disorder is associated with a high IQ, children may initially compensate for the difficulties with the result that the dysfunction is discovered only in the third year of elementary school or later. According to some studies,[40,41] children with specific disorder of arithmetical skills show poor visuospatial abilities and also have difficulties in complex and motor tasks. Share *et al.*[42] compared these results with their own. Deficits of right hemisphere functioning were found, but only in boys. These results suggest that boys and girls with specific disorder of arithmetical skills should be studied separately. A further interesting result is the association of the disorder with anxiety.[43] This association is more pronounced in those children in whom arithmetical skills are substantially impaired compared with relatively good reading and spelling skills.

(f) Treatment

The treatment of specific disorders of arithmetical skills follows the same general lines as treatment of specific reading disorder. All treatment components have to focus on the training of skills that are impaired in a way that keeps the child motivated. Treatment strategies should focus on the mathematical disability itself or the mathematics anxiety with which the disorder is frequently associated. In many cases, both facets need to be included in the treatment programme. When treating the mathematical disability according to the child's individual profile of impairment, four aspects should be emphasized: semantic memory, procedural difficulties, visuospatial difficulties, and difficulties with mathematical problem-solving.[44] Mathematics anxiety requires a more psychotherapeutic approach using relaxation techniques, with the aim of reducing anxiety prior to and during maths lessons in order to avoid a sense of failure.

Mixed disorders of scholastic skills

In ICD-10, this is specified as an ill-defined and inadequately conceptualized, but necessary residual, category of disorders in which both arithmetical and reading or spelling skills can be significantly

impaired, and in which the disorder cannot be explained in terms of general intellectual disability or inadequate schooling. This category covers disorders that meet the criteria of 'Specific disorder of arithmetical skills' (F81.2) and either 'Specific reading disorder' (F81.0) or 'Specific spelling disorder' (F81.1). As has been explained earlier, in the case of a mixed disorder of scholastic skills, it is specific arithmetical disorder that seems to dominate both in severity and with respect to associated psychopathological features.

Specific developmental disorder of motor function

Clinical features and classification

Many children to whom this category applies, were previously diagnosed as having 'minimal brain dysfunction'. This term is no longer used. The essential clinical features of the disorder include the following (ICD-10):

- An impairment of motor coordination that is significantly below the expected level on the basis of age and general intelligence assessed by an individually administered and standardized test.
- The difficulties in coordination should already have been present early in development.
- They should not be acquired and not be a direct result of deficits of vision or hearing or of any neurological disorder.

Variability of fine or gross motor coordination is great. The milestones of motor development are usually delayed. In many cases, there is an association with speech difficulties, especially articulation. Parents usually report that the child was slow in learning to sit, run, hop, climb stairs, and ride a bicycle. Many children also have difficulties in learning to tie shoelaces, fasten and unfasten buttons, and throw or catch balls. Some children may be generally clumsy in fine and gross movements others tend to have their main difficulty with fine movements and coordination. In many cases, drawing skills are also impaired and the child's difficulties are particularly obvious during ball games, which require a considerable amount of gross motor coordination.

The DSM-IV criteria are similar, and emphasize a substantial backlog of motor coordination, significant interference with academic achievement, and require general medical conditions and pervasive developmental disorders to be ruled out.

There is growing evidence that specific developmental disorders of motor function are a quite heterogeneous group that needs to be subclassified.[45]

Diagnosis and differential diagnosis

Diagnosis is made according to the criteria and guidelines in ICD-10 and DSM-IV. Differential diagnosis should rule out specific neurological disorders, which can be done with a careful history and neurological examination, pervasive developmental disorders, which can be distinguished by the criteria of these disorders, or attention-deficit hyperactivity disorder. The latter can be distinguished by their pronounced distractibility and impulsivity rather than impairment of motor coordination.

Epidemiology

According to an estimation in DSM-IV about 6 per cent of 5- to 11-year-old children suffer from the disorder.

Aetiology

There are two main factors said to be responsible for the aetiology, genetic influences and brain damage. As far as genetic influences are concerned, there are no valid studies that confirm this assumption. Regarding brain damage, the question arises whether early brain damage can result in a specific impairment of motor functions, while other functions are within normal limits.[46–48] As this appears very unlikely, comorbidity with other disorders should be considered the norm and specificity regarded as the exception.[49]

Course and prognosis

The few follow-up studies have shown that children who suffer from the disorder between 5 and 10 years of age show a high persistence of motor problems in adolescence. Almost all children who were identified as having had motor difficulties at elementary school, had similar problems as teenagers.[50–52]

Treatment

Treatment measures should focus on two main facets: the difficulties and impairments in the different motor functions, and the associated social and emotional difficulties. The first facet requires an active functional treatment of motor functions focusing on the child's individual difficulties. Programmes are available for this kind of treatment, which is usually carried out by physiotherapists and occupational therapists.[53]

In treating the second facet, therapists are confronted with the child's insecurity, which is a direct result of motor difficulties, poor body scheme, avoidance reactions of peers and classmates, and frequent feelings of inferiority and a low self-esteem.[54] These associated problems need to be addressed in psychotherapy, in addition to programmes that focus solely on training motor functions.

Further information

Butterworth, B. (2004). *Dyscalculia guidance helping pupils with specific learning difficulties in Maths*. Nelson Publishing Company, London.

Snowling, M. and Stackhouse, J. (2006). *Dyslexia, speech and language: a practitioner's handbook* Whurr Publishers, West Sussex.

Bishop, D. (2000). *Speech and language impairments in children: causes, characteristics, intervention and outcome*. Taylor and Francis, Philadelphia, PA.

References

1. World Health Organization. (1992). *International statistical classification of diseases and related health problems*, 10th revision. WHO, Geneva.
2. World Health Organization. (1996). *Multiaxial classification of child and adolescent psychiatric disorders*. Cambridge University Press, Cambridge.
3. American Psychiatric Association. (1994). *Diagnostic and statistical classification of diseases and related health problems* (4th edn). American Psychiatric Association, Washington, DC.
4. Böhme, G. (1980). *Therapie der Sprach-, Sprech- und Stimmstörungen*. Fischer, Stuttgart.
5. Webster, A. and McConnell, C. (1987). *Children with speech and language difficulties*. Cassell, London.
6. Bishop, D.V.M. (1994). Developmental disorders of speech and language. In *Child and adolescent psychiatry* (eds. M. Rutter, E. Taylor, and L. Hersov), pp. 546–68. Blackwell, London.
7. Bishop, D.V.M. (1987). The causes of specific developmental language disorder ('developmental dysphasia'). *Journal of Child Psychology and Psychiatry*, **28**, 1–8.

8. Paul, R. (1992). Language and speech disorders. In *Developmental disorders: diagnostic criteria and clinical assessment* (eds. S.R. Hooper, G.W. Hynd, and R.E. Mattison), pp. 209–38. Erlbaum, Hillsdale, NJ.
9. Rutter, M., Mawhood, L., and Howlin, P. (1992). Language delay and social development. In *Specific speech and language disorders in children* (eds. P. Fletcher and D. Hall), pp. 63–78. Whurr, London.
10. Aicardi, J. (ed.) (1992). *Diseases of the nervous system in childhood. Clinics in developmental medicine, No. 115/118*. McKeith Press, London.
11. Lerman, P. and Lerman-Sagie, T. (1989). Early steroid therapy in Landau–Kleffner syndrome. *Advances in Epileptology*, **17**, 330–2.
12. Morrell, F., Whisler, W.W., and Bleck, T.P. (1989). Multiple subpial transsection: a new approach to the surgical treatment of focal epilepsy. *Journal of Neurosurgery*, **70**, 231–9.
13. Nass, R., Heier, L., and Walker, R. (1993). Landau–Kleffner syndrome: temporal lobe tumor resection results in good outcome. *Paediatric Neurology*, **9**, 303–5.
14. Esser, G. and Schmidt, M.H. (1993). Die langfristige Entwicklung von Kindern mit Lese-Rechtschreibschwäche. *Zeitschrift für Klinische Psychologie*, **22**, 100–16.
15. Rutter, M., Tizard, J., and Whitmore, K. (1970). *Education, health and behaviour*. Longmans, London.
16. Esser, G. and Schmidt, M.H. (1994). Teilleistungsstörung und depression. *Kindheit und Entwicklung*, **3**, 157.
17. Schulte-Körne, G. (2001). Annotation: genetics of reading and spelling disorder. *Journal of Child Psychology and Psychiatry*, **42**, 985–97.
18. Ziegler, A., König, I.R., Deimel, W., *et al.* (2005). Developmental dyslexia—recurrence risk estimates from a German bi-center study using the single proband sib pair design. *Human Heredity*, **5**, 136–43.
19. Schumacher, J., Hoffmann, P., Schmäl, C., *et al.* (2007). Genetics of dyslexia: the evolving landscape. *Journal of Medical Genetics*, **44**, 289–97.
20. Démonet, J.F., Taylor, M.J., and Chaix, Y. (2004). Developmental dyslexia. *Lancet*, **363**, 1451–60.
21. Satz, P. and Sparrow, S. (1970). Specific developmental dyslexia: a theoretical formulation. In *Specific reading disability: theory and method* (eds. D.J. Bakker and P. Satz), pp. 41–60. Rotterdam University Press.
22. Masland, R.L. (1975). Neurological bases of correlates of language disabilities: diagnostic implication. *Acta Symbolica*, **6**, 1–34.
23. Finucci, J.M. (1986). Follow-up studies of developmental dyslexia and other learning disabilities. In *Genetics and learning disabilities* (ed. S.D. Smith), pp. 97–121. Taylor and Francis, Philadelphia, PA.
24. Spreen, O. (1988). Prognosis of learning disability. *Journal of Consulting and Clinical Psychology*, **56**, 836–42.
25. Rutter, M., Tizard, J., Yule, W., *et al.* (1976). Isle–of–Wight studies 1964–1974. *Psychological Medicine*, **6**, 313–32.
26. Bruck, M. (1990). Word recognition skills in adults with childhood diagnosis of dyslexia. *Developmental Psychology*, **26**, 439–54.
27. Clay, M. (1985). *The early detection of reading difficulties* (3rd edn). Heinemann, Tadworth, Surrey.
28. Hewison, J. (1988). Long-term effectiveness of parental involvement in reading: a follow-up to the Haringey Reading Project. *British Journal of Educational Psychology*, **58**, 184–90.
29. Geary, D.C. (1993). Mathematical disabilities: cognitive, neuropsychological, and genetic componenents. *Psychological Bulletin*, **114**, 345–62.
30. Gross-Tsur, V., Manor, O., and Shalev, R.S. (1996). Developmental dyscalculia: prevalence and demographic features. *Developmental Medicine and Child Neurology*, **38**, 25–33.
31. Lewis, C., Hitch, G.W., and Walker, P. (1994). The prevalence of specific arithmetic difficulties and specific reading difficulties in 9 to 10 year–old boys and girls. *Journal of Child Psychology and Psychiatry*, **35**, 283–92.
32. Dehaene, S., Spelke, E., Pinel, P., *et al.* (1999). Source of mathematical thinking: behavioral and brainimaging evidence. *Science*, **284**, 970–4.
33. Dehaene, S., Piazza, M., Pinel, P., *et al.* (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, **20**, 487–506.
34. Isaacs, E.B., Edmonds, C.J., Lucas, A., *et al.* (2001). Calculation difficulties in children of very low birthweight: a neural correlate. *Brain*, **124**, 1701–7.
35. Levy, L.M., Reis, I.L., and Grafman, J. (1999). Metabolic abnormalities detected by 1H-MRS in dyscalculia and dysgraphia. *Neurology*, **53**, 639–41.
36. Light, J.G. and DeFries, J.C. (1995). Comorbidity of reading and mathematics disabilities—genetic and environmental etiologies. *Journal of Learning Disability*, **28**, 96–106.
37. Plomin, R. and Kovas, Y. (2005). Generalist genes and learning disabilities. *Psychological Bulletin*, **131**, 592–617.
38. Schulte-Körne, G., Ziegler, A., Deimel, W., *et al.* (2006). Interrelationship and familiality of dyslexia related quantitative measures. *Annals of Human genetics*, **70**, 1–16.
39. Landerl, K., Bevan, A., and Buttherworth, B. (2004). Developmental dyscalculia and basic numerical capacities: a study of 8-9-year-old students. *Cognition*, **93**, 99–125.
40. Rourke, B.P. and Finlayson, M.A.J. (1978). Neurophysiological significance of variations in pattern of academic performance: verbal and visuo-spatial abilities. *Journal of Abnormal Child Psychology*, **6**, 121–33.
41. Rourke, B.P. and Fuerst, D.R. (1991). *Learning disabilities and psychosocial functioning: a neuropsychological perspective*. Guilford Press, New York.
42. Share, D.L., Moffitt, T.E., and Silva, P.A. (1988). Factors associated with arithmetic-and-reading disability and specific arithmetic disability. *Journal of Learning Disabilities*, **21**, 313–20.
43. Lansdown, R. (1978). Retardation in mathematics: consideration of multifactorial determination. *Journal of Child Psychology and Psychiatry*, **19**, 181–5.
44. Geary, D.C. (1994). *Children's mathematical development. Research and practical applications*. American Psychological Association, Washington, DC.
45. Hoare, D. (1994). Subtypes of developmental coordination disorder. *Adapted Physical Activity Quarterly*, **11**, 158–69.
46. Caron, D. and Rutter, M. (1991). Co-morbidity in child psychopathology: concepts, issues, and research strategies. *Journal of Child Psychology and Psychiatry*, **32**, 1063–80.
47. Henderson, S.E., Barnett, A.L., and Henderson, L. (1994). Visuo–spatial difficulties and clumsiness. On the interpretation of conjoint deficits. *Journal of Child Psychology and Psychiatry*, **35**, 961–9.
48. Powell, R.P. and Bishop, D. (1992). Clumsiness and perceptual problems in children with specific language impairment. *Developmental Medicine and Child Neurology*, **34**, 755–65.
49. Henderson, S.E. and Barnett, A.L. (1988). Developmental movement problems. In *Perspectives on the classification of specific developmental disorders* (eds. J. Rispens, T.A. van Yperen, and W. Yule), pp. 209–30. Kluwer Academic, Dordrecht.
50. Knuckey, N.W. and Gubbay, S.S. (1983). Clumsy children: a prognostic study. *Australian Paediatric Journal*, **19**, 9–13.
51. Losse, A., Henderson, S.E., Elliman, D., *et al.* (1991). Clumsiness in children—do they grow out of it? A ten-year follow-up study. *Developmental Medicine and Child Neurology*, **33**, 55–68.
52. Cantell, M.H., Smyth, M.M., and Ahonen, T.P. (1994). Clumsiness in adolescents: educational, motor a nd social outcomes of motor delay detected at five years. *Special Issue of Adapted Physical Activity Quarterly*, **11**, 115–29.
53. Schoemaker, M.M., Hijlkema, M.G.J., and Kalverboer, A.F. (1994). Physiotherapy for clumsy children: an evaluation study. *Developmental Medicine and Child Neurology*, **36**, 143–55.
54. Kalverboer, K.F. (1998). On the relevance of specific classifications of disorders with particular focus on DCD, developmental coordination disorder. In *Perspectives on the classification of specific developmental disorders* (eds. J. Rispens, T.A. van Yperen, and W. Yule), pp. 265–78. Kluwer Academic, Dordrecht.

9.2.3 Autism and the pervasive developmental disorders

Fred R. Volkmar and Ami Klin

Introduction

The pervasive developmental disorders (PDDs) are characterized by patterns of deviance and delay in social-communicative development in the first years of life, which are associated with restricted patterns of interest or behaviour. The prototypic PDD is childhood autism; other conditions included in the PDD class in ICD-10[1] include Rett's syndrome, childhood disintegrative disorder, Asperger's syndrome, and atypical autism. Except for one additional category in ICD-10 (hyperkinetic stereotyped movement disorder), the disorders included in ICD-10 and DSM-IV[2] are essentially identical. In this chapter each of these conditions will be reviewed in terms of their clinical features, definition, epidemiology, course, and aetiology; final sections of the chapter address aspects of treatment and prevention for the group of disorders as a whole (Box 9.2.3.1).

Childhood autism

Autism was first recognized by Kanner[3] in his report of 11 children with 'autistic disturbances of affective contact'. He emphasized two essential diagnostic features: autism and difficulties with change; but he also noted atypical language (when language was present at all). He used Bleuler's term 'autistic' to convey the children's social isolation. Although children with autism had undoubtedly been previously observed,[4] it was Kanner's particular genius to so precisely describe the condition. False leads for research arose since the term autism introduced an unintended confusion with schizophrenia. Also, Kanner assumed that the children had normal intellectual potential. Subsequently, it became clear that autism and schizophrenia were distinct and that autism was often associated with intellectual disability.[5] Although describing autism as inborn, Kanner mentioned that parents were very well educated and successful. In turn this led to the idea, common during the 1950s, that autism might somehow result from deviant patterns of care by unusually successful parents. A large body of evidence shows that this is most certainly not the case.[6] It is clear that families of children with autism come from all social classes and circumstances[7] and that the original impression had been the result of referral bias.

Clinical features

Social deficits of a particular type remain a hallmark of autism. The nature of this deficit varies, somewhat, over time but remains a source of great disability to the affected individual throughout life.[8] In younger and more impaired individuals there may be little interest in social interaction; less impaired individuals may come to a passive acceptance of social interaction while older and more cognitively able individuals are more likely to seem highly eccentric and one-sided.[9] Difficulties are observed in the use of eye contact or other non-verbal social cues, in social emotional reciprocity and empathy, in activities involving shared interests with others, and with peer relationships (see Table 9.2.3.1). As Rutter[10] suggested, these problems do not simply reflect cognitive impairment, which is present in about 60 per cent of individuals affected. Abnormalities in communication (not only in language) are also observed. In a substantial minority (perhaps 30 per cent) of cases, the child never acquires the capacity for communicative speech; among individuals who do talk, various unusual features of language are observed such as echolalia, idiosyncratic language, deficits in prosody, and pronoun reversal.[11] Deficits in pragmatic language are particularly striking. As with the social disturbance, the deficits observed are not solely due to intellectual disabilities. Various unusual behaviours are subsumed under the term 'resistance to change'. These behaviours include literal resistance to change ('insistence on sameness'), stereotyped and repetitive motor mannerisms, strict adherence to non-functional routines, and interest in non-functional parts of objects. Various other features may be observed, such as unusual sensitivities to aspects of the environment and attachments to unusual objects.

Box 9.2.3.1 The pervasive developmental disorders

- Childhood autism/autistic disorder
- Rett's disorder
- Childhood disintegrative disorder
- Asperger's disorder
- PDD-NOS/atypical autism

Definition

In the 1950s and 1960s there was disagreement about autism, e.g. was it a form of schizophrenia or psychosis? Gradually evidence began to accumulate that suggested the role of central nervous system dysfunction in pathogenesis—for example, high risk for developing seizures. Differences in clinical features, onset and course, and family history also supported the distinctiveness of autism apart from childhood schizophrenia.[12,13] By 1978, there was a substantial body of work on the validity of autism. Rutter synthesized this in his influential definition of autism,[10] which required the presence of patterns of delay and deviance in the areas of social and language development that were not simply the result of developmental delay along with the group of unusual behaviours subsumed under Kanner's term 'insistence on sameness'. Onset before 30 months of age was required. In ICD-9[14] infantile autism was still termed a psychotic condition but by 1980 the highly influential DSM-III[15] appeared and recognized autism as a condition apart from schizophrenia, including it in a new class of disorders—the pervasive developmental disorders. The latter term has been the topic of some debate, although a better term has yet to be proposed and, in any case, the term PDD has now come into general usage in both DSM and ICD.[16,17]

The name chosen in DSM-III ('Infantile autism') was consistent with Kanner's original report but reflected a lack of developmental orientation; these concerns were addressed in DSM-III-R,[18] which provided a detailed, and more developmentally oriented, set of diagnostic guidelines. Unfortunately, this definition proved overly inclusive and it became apparent that additional work would be needed. Further impetus was given to this effort by pending changes in ICD-10[1] and DSM-IV. Accordingly, an international field trial was undertaken.[19] Based on the results of this field trial

Table 9.2.3.1 ICD-10 criteria for childhood autism (F84.0)

A. Abnormal or impaired development is evident before the age of 3 years in at least one of the following areas:
1 receptive or expressive language as used in social communication;
2 the development of selective social attachments or of reciprocal social interaction
3 functional or symbolic play
B. A total of at least six symptoms from (1), (2), and (3) must be present, with at least two from (1) and least one from each of (2) and (3)
1 Qualitative impairment in social interaction is manifested in at least two of the following areas:
(a) failure adequately to use eye-to-eye gaze, facial expression, body postures, and gestures to regulate social interaction
(b) failure to develop (in a manner appropriate to mental age, and despite ample opportunities) peer relationships that involve a mutual sharing of interests, activities, and emotions
(c) lack of socio-emotional reciprocity as shown by an impaired or deviant response to other people's emotions; or lack of modulation of behaviour according to social context; or a weak integration of social, emotional, and communicative behaviours
(d) lack of spontaneous seeking to share enjoyment, interests, or achievements with other people (e.g. a lack of showing, bringing, or pointing out to other people objects of interest to the individual)
2 Qualitative abnormalities in communication in at least one of the following areas:
(a) delay in or total lack of development of spoken language that is not accompanied by an attempt to compensate through the use of gestures or mime as an alternative mode of communication (often preceded by a lack of communicative babbling)
(b) relative failure to initiate or sustain conversational interchange (at whatever level of language skill is present), in which there is reciprocal responsiveness to the communications of the other person
(c) stereotyped and repetitive use of language or idiosyncratic use of words or phrases
(d) lack of varied spontaneous make-believe play or (when young) social imitative play
3 Restricted, repetitive, and stereotyped patterns of behaviour, interests, and activities are manifested in at least one of the following:
(a) an encompassing preoccupation with one or more stereotyped and restricted patterns of interest that are abnormal in content or focus; or one or more interests that are abnormal in their intensity and circumscribed in nature though not in their content or focus
(b) apparently compulsive adherence to specific non-functional routines or rituals
(c) stereotyped and repetitive motor mannerisms that involve either hand or finger flapping or twisting or complex whole-body movements:
(d) preoccupations with part-objects or non-functional elements of play materials (such as their odour, the feel of their surface, or the noise or vibration they generate)
C. The clinical picture is not attributable to the other varieties of pervasive developmental disorders; specific developmental disorder of receptive language (F80.2) with secondary socio-emotional problems, reactive attachment disorder (F94.1), or disinhibited attachment disorder (F94.2); mental retardation (F70–F72) with some associated emotional or behavioural disorders; schizophrenia (F20.-) of unusually early onset; and Rett's syndrome (F84.12).

Taken from Disorders of psychological development ***(criteria for research)***, pp. 154–5. © World Health Organization, www.who.int

autism is defined (see Table 9.2.3.1) on the basis of characteristic problems in three areas: social interaction, communication and play, and restricted patterns of interest. By definition, autism must be present by the age of 3 years. ICD-10 provides for various ways in which a diagnosis of atypical autism can be made—for example, because of failure to meet age of onset or behavioural criteria. Data on this system suggest that it has good agreement with the diagnoses of experienced clinicians, avoids the problem of the overdiagnosis of autism in the most mentally handicapped persons, and has reasonably good reliability.

Epidemiology and demographics

Over 40 epidemiological studies of autism and related conditions have been conducted with prevalence estimates ranging from 0.7 per 10 000 to 72.6 per 10 000.[20] In his recent 2005 review, Fombonne notes that prevalence rates are negatively correlated with sample size and there is an apparent trend for increased rates over time. Various considerations (including changes in definition) complicate the interpretation of increased rates. In this review, Fombonne suggests that a reasonable estimate of prevalence is 13 per 10 000 (Box 9.2.3.2). Although higher rates of 1 in 150 are reported if broad definitions are used.

Box 9.2.3.2 Epidemiology of autism and related conditions

Autism. . . . . .13 per 10 000
Rett's disorder. . . . . 1 per 10 000 girls
Childhood disintegrative disorder . . . . . .1.9 per 100 000
Asperger's syndrome . . . . . .4 per 10 000
PDD-NOS/atypical autism . . . . . .1 per 150

Source: Cohen, D.J. and Volkmar, F.R. (eds.) (2005) ***Handbook of autism and pervasive developmental disorders*** (3rd edn.) Wiley, New York.

A number of studies, including both epidemiological and clinically referred samples, report higher rates of autism in boys than in girls (typically 3.5:1 or 4:1). This ratio varies with IQ level, i.e. females with autism who have IQs in the normal range are probably

20 times less common than males.[21] The explanation for this sex difference remains unclear. It is possible that the degree of insult required to produce autism in females must be greater than for males, but other hypothesis have been raised. Ethnic and cultural differences have been little studied.[20,22]

Course and prognosis

Although childhood autism is a chronic disability, there is some suggestion that with early intervention and remediation outcome is improving.[23] For example, the number of individuals with either a 'good' or 'fair' outcome is now about 50 per cent—a noteworthy increase since 1980.[24] However, even in the highest functioning individuals marked social problems persist.

Changes in the degree of social relatedness, communication, and self-help skills are observed with increases in developmental level. Seizure disorders are observed in up to 25 per cent of individuals and may have their onset at any point, but adolescence and early childhood are particularly common times.[25] Factors that predict long-term outcome include the presence of some communicative speech by the age of 5 or 6 years, and non-verbal intellectual level.

Aetiology

Early interest centred on the possibility that experiential factors might somehow cause autism, but a host of research findings suggests that this is not the case. Rather, a fundamental disturbance in the central nervous system is implicated.

(a) Psychological factors

Disabilities affecting attentional mechanisms, arousal, sensory deficits, memory management, and complex information processing, among others, have been proposed as 'primary' deficits underlying the social impairment in autism. Although each of these helps us understand some aspects of the condition, none has, as yet, provided a more comprehensive account of the condition as a whole.[26] Among the most influential recent theories attempting to do that is the hypothesis that posits a lack of a central drive for coherence in children with autism, with the consequent focus on dissociated fragments of their environment rather than integrated 'wholes', leading to a fragmentary and overly concrete experience of the world.[27] Another cognitive account of autism posits that the commonly found difficulties in abstracting rules, inhibiting irrelevant responses, shifting attention, and profiting from feedback as well as in maintaining relevant information 'on-line'—the so-called 'executive functions'—underlie the social, communicative, and behavioural disabilities in autism.[28] Although both these theories—'weak central coherence' and 'executive dysfunction'—provide insightful new views of well-known clinical features, neither phenomena can be seen as specific to autism relative to other developmental disorders.

Probably the most influential current cognitive hypothesis focuses on mechanisms directly impacting on social understanding. This view, the 'theory of mind' hypothesis, posits that autism is caused by the child's inability to attribute mental states such as beliefs and intentions to others. Devoid of this ability, individuals with autism are thought to be unable to infer the thoughts and motivations of others, thus failing to predict their behaviour and adjust their own actions accordingly, which results in a lack of reciprocity in communication and social contact.[29] Although more than 50 studies have documented such deficits in autism, there are still many limitations to this hypothesis. For example, more able individuals with autism do exhibit 'theory of mind' skills—and yet may be totally unable to utilize this capacity in their spontaneous social adjustment. Such phenomena suggest that factors other than a cognitive understanding of mental phenomena are required for a person to meet the demands of everyday social life. For example, the 'enactive mind' hypothesis focuses on early emerging and highly conserved mechanisms of socialization that precede the advent of mentalizing abilities, and which culminate in the development of joint attention and perspective taking skills.[30] Of great interest in the past few years has been the confluence of experimental psychological paradigms and functional neuroimaging studies focusing on the same constructs. This new trend is leading to new insights into brain systems subserving basic social mechanisms such as gaze behaviour, face processing, social-affective responses, and thinking about other people's intentions and beliefs,[31] all of which are greatly compromised in autism.

(b) Biological factors

The importance of biological factors in the pathogenesis of autism is suggested by several lines of evidence. Autism has been associated with a host of medical conditions; but the absence of population data and rigorous diagnostic assessment makes such associations difficult to interpret. For example, early reports suggested an association with congenital rubella, but this now seems questionable given the diagnostic dilemmas and the observation that 'autistic-like' features diminish over time.[32] Gillberg[33] argues that medical conditions may be associated with autism in as many as one-third of the cases, but Rutter and colleagues[34] suggest that a more reasonable figure would be roughly 10 per cent of all cases. The strongest associations are with fragile X syndrome and tuberous sclerosis—both conditions having a strong genetic component (Box 9.2.3.3).

Box 9.2.3.3 Medical conditions associated with autism

- Seizure disorder (epilepsy)
- Fragile X syndrome
- Tuberous sclerosis

Fragile X syndrome is an X-linked intellectual disability syndrome involving a mutation characterized by a triplet repeat of cytosine–guanine–guanine (CGG) that may amplify with succeeding generations. It is associated with a characteristic facial appearance, enlarged testicles, intellectual disability, and some autistic features. Early reports suggesting high rates of fragile X in autism have now been modified; fragile X affects perhaps 1 to 2 per cent of all individuals with autism.[20]

The autosomal dominant disorder tuberous sclerosis is characterized by abnormal tissue growth, or benign tumours (hamartomas), in the brain and in other organs. The condition, which may affect 1 in 10 000 individuals, is variably expressed; the phenotype ranges from minor skin problems or seizures to severe intellectual disability with intractable seizures. The rate of this condition in autism (0.4–2.8 per cent) is significantly increased.[20]

Autism is a strongly genetic condition. Studies of monozygotic and dizygotic twins revealed much higher levels of concordance for monozygotic relative to dizygotic twin pairs, but also elevated

rates of concordance in dizygotic twins relative to population rates.[35] General studies suggest that the recurrence risk of autism in siblings is in the order of between 2 and 10 per cent—which is a substantial increase in risk over population rates. There also appear to be higher rates of social and language problems and rigid patterns of behaviour in siblings and close relatives, raising the possibility that what is inherited is a broader phenotype reminiscent of autism but which also may reflect a more general predisposition to developmental difficulties. Recent work also suggests elevated rates of anxiety and mood disorders in family members. Specific modes of inheritance remain unclear. It now appears that several interacting genes are probably involved in the pathogenesis of autism. Efforts are underway to identify susceptibility genes and trace their impact on brain development. Although several studies have shown increased rates of pre-, peri-, and neonatal complications in children with autism, it is possible that some of these difficulties may reflect a genetic vulnerability in the child or that there may be an interaction of genetic and perinatal factors.[35] A recent report has noted the presence of placental abnormalities in pregnancies of children with autism.

Attempts have been made to identify neuropathological and neuroanatomical correlates of autism. Areas of interest have included the cortical areas responsible for language and social interaction (frontal and temporal lobes) as well as the neostriatum and cerebellum.[36] The report of reduced cerebellar size in the neocerebellar vermal lobules VI and VII has not proven readily replicable. Some individuals with autism have enlarged brains and head sizes, with some evidence that abnormal growth occurs in the first 2 years of life.[37] Neuropathological studies have suggested possible cellular changes in areas of the brain such as the hippocampus and amygdala and changes in the cytoarchitecture of the brain, e.g. in the arrangement of cortical 'minicolumns'.

Various neurotransmitter systems have been studied. Probably the most robust finding has been the observation that about one-third of the children with autism have increased peripheral levels of serotonin. This finding is not specific to autism and its significance remains unclear.[38] Studies of other neurotransmitters such as dopamine produced inconsistent findings. The possible involvement of dopamine is suggested by the high levels of stereotyped behaviours in autism—behaviours, which can be induced in animals by the administration of agents (stimulants) that affect levels of dopamine in the brain. Agents such as the neuroleptics, which block dopamine receptors, are effective in reducing the stereotyped and hyperactive behaviours of many autistic children. Another hypothesis has centred on the possible role of endogenous opioids, in that overproduction of such compounds might lead to social withdrawal and unusual sensitivities and behaviours. This has led to the administration of opioid antagonists such as naltrexone in autism; unfortunately results have been disappointing. Studies of the immune system in autism have been relatively uncommon and findings inconsistent.

Rett's syndrome

In 1966, Andreas Rett described an unusual syndrome in girls characterized by a history of initial normal development, subsequent head growth deceleration, and the development of specific clinical findings such as breathing difficulties, movement problems, and some features suggestive of autism.[39] His findings were replicated and extended by Hagberg and colleagues.[40] As more extensive information became available it became clear that the more 'autistic-like' phase of Rett's syndrome was relatively brief, but this was a major rationale for its inclusion in the PDD class.[41]

Clinical features

Early pre- and perinatal histories are generally unremarkable in Rett's syndrome, as is very early development. Usually within the first or second year of life development begins to slow or actually regress and various motor problems—including characteristic hand-washing stereotypies start to develop.[42] A significant loss of developmental skills occurs and head growth decelerates.[40] The potential for misdiagnosis of autism is greatest during this time; during school age developmental regression often stabilizes and children are more socially responsive. As individuals with Rett's syndrome approach adolescence they are frequently subject to increased spasticity, scoliosis, loss of ambulation, bruxism, hyperventilation, areophagia, apnoea, and seizures.[43] Although debated,[44] the inclusion of Rett's syndrome in the PDD class reflects an awareness of the confusion with autism and the importance of including the condition somewhere.[41]

Definition

Various diagnostic criteria for Rett's syndrome have been developed.[45] As presently defined in ICD-10 (see Table 9.2.3.2) the condition develops after some months of normal development; head circumference is normal at birth but begins to decelerate between 5 and 48 months. Characteristic midline hand-wringing or hand-washing stereotypies develop and purposeful hand movements are lost.

Epidemiology and demographics

Prevalence estimates range from 1 per 12 000 to 1 per 23 000 females; cases are observed in all racial and ethnic groups.[46,47] Rett's syndrome may account for one-quarter to one-third of progressive developmental disabilities among females.[46] A small number of cases have now been reported in males.

Course and prognosis

Various clinical stages of the condition have been identified.[42] Early developmental losses may be subtle initially but become more marked with time. Purposeful hand movements are often then lost, as are speech skills. Ataxia and gait abnormalities are noted in those

Table 9.2.3.2 ICD-10 criteria for Rett's syndrome (F84.2)

A. There is an apparently normal prenatal and perinatal period and apparently normal psychomotor development through the first 5 months and normal head circumference at birth.
B. There is a deceleration of head growth between 5 months and 4 years and loss of acquired purposeful hand skills between 5 and 30 months of age that is associated with concurrent communication dysfunction and impaired social interactions and the appearance of poorly co-ordinated/unstable gait and/or trunk movements.
C. There is severe impairment of expressive and receptive language, together with severe psychomotor retardation.
D. There are stereotyped midline hand movements (such as hand-wringing or 'hand-washing') with an onset at or after the time when purposeful hand movements are lost.

Taken from Disorders of psychological development (***criteria for research***), pp. 154–5.

patients who had acquired the ability to walk, unusual breathing patterns (hyperventilation and/or apnoea), and seizures may also be observed. While social and communication skills may improve in middle childhood motor problems are more pronounced. During the final phase (roughly after age 10) progressive scoliosis and spasticity are observed, while cognitive function stabilizes and seizure activity may diminish. There is a dearth of information on adults with the condition. There does appear to be an increased risk of sudden death due to seizures and/or respiratory difficulties.[48]

Aetiology

Rett[39] originally speculated that the condition might be associated with high peripheral ammonia levels, but this was incorrect. Recent work identified mutations in the MeCP2 gene as the cause of Rett syndrome in most cases; this gene has a major role in regulating various genes during brain development.[49]

Childhood disintegrative disorder

Shortly after the turn of the twentieth century a special educator working in Vienna, Theodor Heller,[50] reported children who had a period of several years of normal development prior to a marked regression with loss of skills in multiple areas and minimal recovery. He initially termed the condition dementia infantilis; subsequently it has been referred to as Heller's syndrome, disintegrative psychosis, or childhood disintegrative disorder.[51] Once it develops the condition is indistinguishable from autism,[52] but it is accorded separate diagnostic status since it appears distinctive in terms of onset and course.

Clinical features

In this condition an 'autistic-like' clinical picture develops after a prolonged period of normal development.[50] More than 100 cases have now been reported; the condition is rare but probably underdiagnosed.[51] Essential clinical features include a relatively long period of normal development followed by a marked developmental regression and development of various 'autistic-like' features; onset is typically between the ages of 3 and 4 years.

Onset can be relatively abrupt or more gradual, and a premonitory phase of non-specific anxiety or agitation may be observed. Parents often note associations between the onset of the condition and various psychosocial or medical events but such associations are probably correlational rather than causative.[53] Once established it resembles autism.[52] Given the behavioural similarity to autism, it might be argued either that the condition does not warrant separate diagnostic recognition or that the condition is inevitably the result of an association with some progressive medical condition. However, the pattern of onset is quite distinctive, the outcome appears to be even worse than in autism, and usually even with very intensive medical evaluations no general medical condition accounting for the regression is identified.[52]

Definition

Criteria for the disorder are listed in Table 9.2.3.3. By definition the disorder cannot coexist with autism or another other explicitly defined pervasive developmental disorder, schizophrenia, elective mutism, or the syndrome of acquired aphasia with epilepsy.

Epidemiology and demographics

Epidemiological data are quite limited. Some limited data suggest a prevalence rate of 1 in 100 000. Initially cases seemed to be equally distributed between males and females, but it now seems likely that cases of Rett's syndrome may have been included in early case series. Recent reviews have noted a preponderance of males.[20]

Table 9.2.3.3 ICD-10 criteria for other childhood disintegrative disorder (F84.3)

A. Development is apparently normal up to the age of at least 2 years. The presence of normal age-appropriate skills in communication, social relationships, play, and adaptive behaviour at age 2 years or later is required for diagnosis.
B. There is a definite loss of previously acquired skills at about the time of onset of the disorder. The diagnosis requires a clinically significant loss of skills (not just a failure to use them in certain situations) in at least two of the following areas: 1 expressive or receptive language 2 play 3 social skills or adaptive behaviour 4 bowel or bladder control 5 motor skills.
C. Qualitatively abnormal social functioning is manifested in at least two of the following areas: 1 abnormalities in reciprocal social interaction (of the type defined for autism) 2 qualitative abnormalities in communication (of the type defined for autism) 3 restricted, repetitive, and stereotyped patterns of behaviour, interests, and activities, including motor stereotypes and mannerisms 4 a general loss of interest in objects and in the environment.
D. The disorder is not attributable to the other varieties of pervasive developmental disorder; acquired aphasia with epilepsy (F80.6); elective mutism (F94.0); Rett's syndrome (F84.2); or schizophrenia (F20.-).

Taken from Disorders of psychological development (criteria for research), pp. 154–5

Course and prognosis

In about 75 per cent of cases the deterioration reaches a plateau, in that there is no further loss of skills, but subsequent gains tend to be minimal. Thus the child who previously was normally socially related, spoke in full sentences, and was toilet-trained becomes indifferent to social interaction, loses all expressive language and toilet skills, and remains mute and relatively low functioning.[52] In the remainder of cases there is more limited recovery, for example the child regains the capacity to speak but only in single words or only echoes language. If a progressive metabolic or neuropathological process is present, the developmental progression may continue until death ensues; such cases often have a later onset.[54] Life expectancy otherwise appears to be normal. In a very small number of cases significant recovery has been noted.

Aetiology

Various lines of evidence suggest the importance of neurobiological factors in pathogenesis. Occasionally medical conditions such as the neurolipidoses, metachromatic leukodystrophy, Addison–Schilder's disease, and subacute sclerosing panencephalitis are associated with the condition. Rates of electroencephalographic abnormalities and seizure disorders are of roughly the same frequency as seen in autism.

Asperger's syndrome

The condition known as Asperger's syndrome was described by a paediatrician with interest in intellectual disability, Hans Asperger,[55] who reported on four boys with marked social problems, unusual perseverative interests, and motor clumsiness but with seemingly good verbal and cognitive abilities. Like Kanner, Asperger used the word autism (autistic psychopathy) to describe this condition. His concept, however, had points of difference, as well as similarity, to autism. For example, verbal abilities tended to be an area of strength, concerns typically did not arise until later in the preschool period, and there was a tendency for the condition to run in families—particularly in fathers. Lorna Wing's[56] report of Asperger's work and publication of a series of cases brought wider attention to the diagnostic concept. The validity of this condition, particularly apart from higher-functioning autism, remains the topic of much debate. A major complication has been the marked differences in definition of the conditions and its potential overlap with other diagnostic concepts (e.g. schizoid personality,[57] nonverbal learning disabilities,[58] semantic–pragmatic disorder,[59] and right hemisphere learning problems).[60] As a result, the literature on this condition is difficult to interpret, although areas of potential differences from autism have been identified, such as neuropsychological profiles[61] and family history.[62]

Clinical features

This condition is characterized by impairments in social interaction and restricted interests and behaviours as seen in autism. However, the child's early development is marked by lack of any clinically significant delay in spoken or receptive language, cognitive development, self-help skills, and curiosity about the environment. Consistent with Asperger's[55] original report, all-absorbing and intense circumscribed interests as well as motor clumsiness are typical, but are not required for diagnosis. The validity of this condition, apart from high-functioning autism and PDD not otherwise specified (PDD-NOS) is controversial. Available research is difficult to interpret given the markedly different ways in which the diagnostic concept has been used. Differences are more likely to be noted relative to autism if a rather stringent diagnostic approach is used. Evidence for external validity of the condition relative to autism includes differences in neuropsychological profiles, patterns of comorbidity, and family history.

Persons with Asperger's syndrome often exhibit a somewhat eccentric social style rather than the more passive or aloof style noted in autism; for example, they may engage others in very one-sided conversations about their area of special interest. They maybe overly reliant on rigid rules for social interaction and may fail to 'see the forest for the trees' in social matters (e.g. an appreciation of exactly when the usual rules do not apply is as important as when they do). Their social oddity and lack of flexibility is a source of much disability.

While early speech-communication skills are apparently normal, certain aspects of communication become more deviant over time. Prosody may be poor, rate of speech may also be unusual, or it may have a somewhat disorganized, tangential, and circumstantial quality. The issue of whether such persons are at increased risk for thought disorder and psychosis remains unresolved, but some part of this impression probably reflects communication problems.

It is rather typical for patients to amass considerable factual information about their topic of interest, which they pursue with great intensity; Asperger originally observed that family life may revolve around the topic of special interest. He also suggested that motor clumsiness was present and, although not required for the diagnosis, there is often a history of motor delay and persistent motor awkwardness—for instance, the child may talk before he walks, have trouble fastening fastener, catching a ball, learning to ride a bicycle, engaging in cursive handwriting, and may also display stiffened gate.

Differences in neuropsychological profiles have been reported.[63] A stringent diagnostic approach may suggest areas of relative strengths (auditory and verbal skills and rote learning) and weakness (visuomotor and visuoperceptual skills); this pattern differs from that observed in higher-functioning individuals with autism[64] and the heritability of social difficulties may be even greater in Asperger's syndrome than in autism.[62]

Interest in the condition has revolved around the possibility that it might represent a transition between autism and other disorders such as schizophrenia. Associated conditions have included depression, anxiety and other mood problems, violence, and other psychotic conditions.[63] Unfortunately, almost all of this literature rests on case reports; controlled studies are needed.

Definition

The ICD-10 criteria for Asperger's syndrome are given in Table 9.2.3.4. As presently defined the social deficit is the same as in autism. In contrast to autism, however, early language, cognitive, and other skills develop typically early in life. By definition, the case does not meet the criteria for childhood autism. Miller and Ozonoff[65] note that several aspects of the current definition can be problematic; Asperger consistently felt that the syndrome he described differed from Kanner's autism (1979). Thus the current definition almost certainly will be refined (or discarded) in future editions of ICD and DSM.

Table 9.2.3.4 ICD-10 criteria for Asperger's syndrome (F84.4)

A. There is no clinically significant general delay in spoken or receptive language or cognitive development. Diagnosis requires that single words should have developed by 2 years of age or earlier and that communicative phrases be used by 3 years of age or earlier. Self-help skills, adaptive behaviour, and curiosity about the environment during the first 3 years should be at a level consistent with normal intellectual development. However, motor milestones may be somewhat delayed and motor clumsiness is usual (although not a necessary diagnostic feature). Isolated special skills, often related to abnormal preoccupations, are common, but are not required for the diagnosis.
B. There are qualitative abnormalities in reciprocal social interaction (criteria as for autism).
C. The individual exhibits an unusual intense circumscribed interest or restricted, repetitive, and stereotyped patterns of behaviour interests and activities (criteria as for autism; however, it would be less usual for these to include either motor mannerisms or preoccupations with part-objects or non-functional elements of play materials).
D. The disorder is not attributable to other varieties of pervasive developmental disorder; simple schizophrenia; schizotypal disorder; obsessive-compulsive disorder; anankastic personality disorder; reactive and disinhibited attachment disorders of childhood.

Taken from Disorders of psychological development (criteria for research), pp. 154–5

Epidemiology and demographics

Estimates of prevalence vary markedly depending on the stringency of the definition used. A stringent approach to diagnosis would suggest a rate in the order of 1 in 2000 or so, but a much less stringent approach may yield one in several hundred.[66]

Asperger[55] originally reported the condition only in boys, although Wing reported on girls with the condition. There does appear to be a male predominance in the order of 20 to 1; similar sex ratios are observed in autism not associated with intellectual disability.

Course and prognosis

Asperger's original[55] impression was of favourable long-term prognosis.[67] Many individuals can attend regular school with some additional support; unfortunately such children may be seen as eccentric and are often prime targets for being victimized. Better verbal skills can mislead educators about the child's vulnerability in other areas, and difficulties academically may be misattributed to wilful non-compliance. There is the impression that these individuals are capable of greater degrees of personal and occupational self-sufficiency than those with autism, but definitive data are lacking. It does appear that the social difficulties persist into adulthood.[68]

Aetiology

Although the cause of Asperger's syndrome remains unknown, the report of high rates of the condition in family members and the reports of occasional familial associations with autism suggests the potential importance of genetic factors. Neurobiological information on the condition is limited. The potential association of the condition with specific neuropsychological profiles is of some interest.

Atypical autism/PDD not otherwise specified

Recent editions of both ICD and DSM have included a 'subthreshold' category (termed either atypical autism or pervasive developmental disorder not otherwise specified—PDD-NOS). In some ways, this notion has historical links to earlier diagnostic concepts.[23] In practice, the term atypical autism in ICD-10 and the term PDD-NOS in DSM-IV refer to what is a residual diagnostic category. ICD-10 provides the possibility for various forms of special coding—for example, failure to meet the onset criteria for autism, failure to meet developmental/behavioural criteria, or failure to meet both.

Research on this diagnostic category has been less advanced than that for other disorders—no doubt reflecting the problems intrinsic to 'subthreshold' disorders. Several attempts have recently been made to identify potential subgroups within the rather heterogeneous disorder, but none has yet achieved general acceptance. Various attempts have been made to identify specific subgroups within this broader category.[69] It is sometimes the case that social and/or communicative skills are relatively preserved but the child exhibits unusual sensitivities, affective responses, and thought processes. It is likely that the term presently encompasses a number of conditions, which may be identified in the future.

Differential diagnosis

Autism and related PDDs must be differentiated from each other, from the specific developmental disorders (e.g. of language), from intellectual disability not associated with PDD, and from other conditions. In intellectual disability not associated with PDD, social and communicative skills are typically on a par with the child's overall intellectual ability. Diagnostic differentiation can be most challenging in persons with severe and profound intellectual disability where assessment is more difficult and stereotyped movements common. Occasionally language and other specific developmental disorders may be confused with autism/PDD. Usually, however, the child's social abilities are preserved and the child is very communicative non-verbally. Schizophrenia rarely has its onset in childhood and almost never before the age of 5 years.

On occasion, selective mutism or social anxiety disorder may be confused with a PDD (particularly PDD-NOS/atypical autism). However, in selective mutism the child can speak in some situations. Similarly, children with anxiety in social situations will usually not exhibit the other symptoms characteristic of autism/PDD. The unusual behaviour and interests of children with obsessive–compulsive disorder may be taken to suggest autism or PDD, but social and language-communication skills are preserved.

In considering the differential diagnosis of conditions that present with regression it is important to review carefully previous diagnostic evaluations. Occasionally, progressive neuropathological conditions may have their onset in childhood. In the Landau–Kleffner syndrome (acquired aphasia with epilepsy) social skills should be relatively preserved even in the face of an extensive aphasia.

Sometimes children who have experienced marked neglect may present with social difficulties, initially suggesting autism or PDD. However, in reactive attachment disorder the history of severe neglect is observed and, as the name of the condition implies, social deficits should remit substantially if an appropriate and nurturing environment is provided.

Treatment

Over the past decade a relatively substantial body of research on treatment of autism has appeared. Recent summaries of this work are available.[70] Much of this work relates to behavioural and educational interventions although a body of well-controlled studies of psychopharmacological agents has appeared as well.[71]

Children with autism and other PDDs generally require an intensive and highly structured intervention programme. More able children may be able to tolerate regular classroom situations, with appropriate support, but more impaired children often need higher levels of teacher supervision and a more intensive classroom setting.[72] For lower-functioning children areas of priority include the ability to tolerate adult guidance and intrusion, to follow routines, to develop communicative abilities, and move from associative to more conceptual learning strategies.[73] The classroom setting can be important, as children with PDD can be readily distracted by extraneous stimuli. The tendency of such children to rely on routines can be used effectively to help promote more systematic learning. Generalization of skills learned is particularly important since the child may have difficulties in applying skills learned in new settings. Speech and communication are a critically important aspect of any intervention programme.[11] Techniques to foster communication through non-verbal means such as sign language, picture-exchange, visual schedules, and other augmentative methods can be very helpful to non-verbal children.[74] The use of such methods does not preclude, and in fact may foster, the use of spoken communication.

Behaviour modification techniques are helpful in increasing the frequency of desired behaviours while simultaneously diminishing problem behaviours. Typically, a functional analysis of the target behaviour is initially performed, and then a plan developed for prompting or decreasing the behaviour.[73] While there is general agreement that children with autism/PDD profit from a behaviourally based intervention, there is more controversy over the degree to which progress can be made; for instance, there have been some claims for dramatic improvement and even 'cures' of autism.

Neuroleptics have been the most intensively studied psychopharmacological agents in this population, and there have been several well-designed double blind, placebo-controlled trials.[71] The main mechanism of action appears to be dopamine-receptor blockade. The agents may reduce maladaptive behaviours such as stereotypies,[75] but side effects include sedation, irritability, and movement problems (including tardive dyskinesia). Recently, interest has centred on the newer atypical neuroleptics.[76]

Management

Goals for treatment include promoting learning and reducing behaviours that interfere with learning. Treatment is best based on a comprehensive view of the child and his or her strengths and areas of need. A structured and individualized intervention programme is needed. Various professionals such as speech pathologists, special educators, occupational and physical therapists may be involved.[72,73] Goals for intervention will vary depending on developmental level, life circumstance, and clinical context—for example, vocational factors will be more important during adolescence, and for individuals with conditions like Rett's syndrome the efforts of other professionals (orthopaedists and respiratory therapists) may be needed.

For higher functioning and older individuals the acquisition and generalization of social skills are particularly important. The use of rehearsal and social scripts may be indicated. Teaching must be explicit and can include modelling and rehearsal within individual instruction and small group settings, with the use of naturalistic settings to encourage generalization whenever possible. For higher functioning students, including those with Asperger's syndrome, this can include explicit analyses of challenging social situations, videotaping for self-observation, role playing, and the use of individualized social stories.[77]

There is no evidence that unstructured psychotherapy is useful in autism and related conditions. Structured and supportive psychotherapy may be appropriate for some carefully selected, higher-functioning individuals, particularly if it focuses on explicit problem-solving strategies for frequently troublesome situations.

Pharmacotherapy interventions are not curative, but they may provide considerable help with specific problematic symptoms.[71] The best evidence relates to the atypical neuroleptics but data on other agents are less substantive. The balance of potential benefits and risks should be considered, and informed consent obtained from parents or, whenever possible, the affected child.

Mood stabilizers and antidepressants have sometimes been used, given the increase in affective lability, anxiety, and depression in individuals with autism and PDDs. However, the response to antidepressants has been somewhat variable. Lithium and other mood stabilizers are sometimes used clinically, particularly if there is a strong family history of bipolar disorder, but there have been few controlled studies of these agents.

Selective serotonin-reuptake inhibitors (SSRIs) were of initial interest in autism, given the repeated reports of high group levels of peripheral serotonin in this population as well as of the high levels of repetitive behaviours observed in this group (i.e. reminiscent of those seen in obsessive–compulsive disorder). Several reports have suggested that, at least in adults, these agents may be helpful in lowering the levels of obsessive–compulsive-like behaviours, although activation is sometimes observed; Studies of children have been less frequent, but there is some suggestion that they may also respond.[71]

Various other agents have been used in autism, including anxiolytics, β-blockers, clonidine, and naltrexone. Unfortunately, it is difficult to draw firm conclusions from the limited data available, but at present the clinical efficacy of these agents is not well established in this population. Surprisingly few studies have systematically evaluated the use of stimulants in autism. However, the observation that these agents induce stereotyped behaviours in animals would suggest their potential for increasing levels of stereotyped behaviours.

Prevention

At present information on the prevention of autism and related PDDs is clearly quite limited. Apart from the association with two strongly genetic conditions (fragile X syndrome and tuberous sclerosis) no specific biological markers for autism have yet been found although some promising leads for early screening or children at risk (i.e. siblings) have appeared. It is likely that early diagnosis may change dramatically in the not so distant future as potential susceptibility genes are identified or are combined with innovative approaches to screening—this will have potentially major importance since there is some suggestion that early intervention may significantly improve outcome.

Further information

www.autism.fm. Regularly updated website with links to other sites.

Quackwatch www.quackwath.com. Information about non-conventional treatments.

Jacobson, J.W., Foxx, R.M., and Mulick, J.A. (2005). *Controversial therapies for developmental disabilities: fad, fashion and science in professional practice*. Lawrence Erlbaum Associates, Mahwah, NJ. A comprehensive guide to nonconventional treatments.

Center for disease control—autism related information:

http://www.cdc.gov/ncbddd/dd/ddautism.htm This website provides basic information for physicians (including early warning signs of autism) in both English and Spanish.

Federal autism research networks website: www.autismresearchnet.work.org provides links to US federally funded autism research projects.

References

1. World Health Organization. (1992). *International statistical classification of diseases and related health problems* (10th revision). WHO, Geneva.
2. American Psychiatric Association. (1994). *Diagnostic and statistical classification of diseases and related health problems* (4th edn). American Psychiatric Association, Washington, DC.

3. Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, **2**, 217–50.
4. Treffert, D. (1989). *Extraordinary people*. Bantam, New York.
5. Rutter, M. (1970). Autistic children: infancy to adulthood. *Seminars in Psychiatry*, **2**, 435–50.
6. Rutter, M., Bailey, A., Simonoff, E., *et al.* (1997). Genetic influences in autism. In *Handbook of autism and pervasive developmental disorders* (2nd edn) (eds. D.J. Cohen and F.R. Volkmar), pp. 370–87. Wiley, New York.
7. Schopler, E., Andrews, C.E., and Strupp, K. (1979). Do autistic children come from upper-middle-class parents? *Journal of Autism and Developmental Disorders*, **9**, 139–52.
8. Carter, A.S., Davis, N.O., Klin, A., *et al.* (2005). Social development in autism. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 312–34. Wiley, New York.
9. Wing, L. (1997). Syndromes of autism and atypical development. In *Handbook of autism and pervasive developmental disorders* (2nd edn) (eds. D.J. Cohen and F.R. Volkmar), pp. 148–72. Wiley, New York.
10. Rutter, M. (1978). Diagnosis and definition of childhood autism. *Journal of Autism and Childhood Schizophrenia*, **8**, 139–61.
11. Tager-Flusberg, H., Paul, R., and Lord, C. (2005). Language and communication in autism. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 335–64. Wiley, New York.
12. Kolvin, I. (1971). Studies in the childhood psychoses. I. Diagnostic criteria and classification. *The British Journal of Psychiatry*, **118**, 381–4.
13. Rutter, M. (1972). Childhood schizophrenia reconsidered. *Journal of Autism and Childhood Schizophrenia*, **2**, 315–37.
14. World Health Organization. (1977). *International statistical classification of diseases and related health problems* (9th revision). WHO, Geneva.
15. American Psychiatric Association. (1980). *Diagnostic and statistical classification of diseases and related health problems* (3rd edn). American Psychiatric Association, Washington, DC.
16. Volkmar, F.R. and Cohen, D.J. (1991). Debate and argument: the utility of the term pervasive developmental disorder. *Journal of Child Psychology and Psychiatry*, **32**, 1171–2.
17. Rutter, M. and Schopler, E. (1992). Classification of pervasive developmental disorders: some concepts and practical considerations. *Journal of Autism and Developmental Disorders*, **22**, 459–82.
18. American Psychiatric Association. (1987). *Diagnostic and statistical classification of diseases and related health problems* (3rd edn, revised). American Psychiatric Association, Washington, DC.
19. Volkmar, F.R., Klin, A., Siegel, B., *et al.* (1994). Field trial for autistic disorder in DSM-IV. *The American Journal of Psychiatry*, **151**, 1361–7.
20. Fombonne, E. (2005). Epidemiological studies of pervasive developmental disorders. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 42–69. Wiley, New York.
21. Volkmar, F.R., Szatmari, P., and Sparrow, S.S. (1993). Sex differences in pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, **23**, 579–91.
22. Schopler, E. (2005). Cross-cultural program priorities and reclassification of outcome research methods. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 1174–91. Wiley, New York.
23. Howlin, P. (2005). Outcomes in autism spectrum disorders. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 201–21. Wiley, New York.
24. Howlin, P. and Goode, S. (1998). Outcome in autism and related conditions. In *Autism and pervasive developmental disorders* (ed. F.R. Volkmar), pp. 209–41. Cambridge University Press, New York.
25. Volkmar, F.R. and Nelson, D.S. (1990). Seizure disorders in autism. *Journal of the American Academy of Child and Adolescent Psychiatry*, **29**, 127–9.
26. Volkmar, F.R., Lord, C., Bailey, A., *et al.* (2004). Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry*, **45**, 1–36.
27. Happé, F. (2005). The weak central coherence account of autism. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 640–9. Wiley, New York.
28. Ozonoff, S., South, M., and Provencal, S. (2005). Executive functions. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 606–27. Wiley, New York.
29. Baron-Cohen, S. (1995). *Mind blindness*. MIT Press, Cambridge, MA.
30. Klin, A., Jones, W., Schultz, R.T., *et al.* (2003). The enactive mind—from actions to cognition: lessons from autism. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **358**, 345–60.
31. Schultz, R.T. (2005). Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area. *International Journal of Developmental Neuroscience*, **23**, 125–41.
32. Chess, S. (1977). Follow-up report on autism in congenital rubella. *Journal of Autism and Childhood Schizophrenia*, **7**, 69–81.
33. Gillberg, C.L. (1992). The Emanuel Miller Memorial Lecture 1991. Autism and autistic-like conditions: subclasses among disorders of empathy. *Journal of Child Psychology and Psychiatry*, **33**, 813–42.
34. Rutter, M., Bailey, A., Bolton, P., *et al.* (1994). Autism and known medical conditions: myth and substance. *Journal of Child Psychology and Psychiatry*, **35**, 311–22.
35. Rutter, M. (2005). Genetic influences and autism. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 425–52. Wiley, New York.
36. Minshew, N.J., Sweeney, J.A., Bauman, M.L., *et al.* (2005). Neurologic aspects of autism. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 473–514. Wiley, New York.
37. Hazlett, H.C., Poe, M., Gerig, G., *et al.* (2005). Magnetic resonance imaging and head circumference study of brain size in autism: birth through age 2 years. *Archives of General Psychiatry*, **62**, 1366–76.
38. Anderson, G.M. and Hoshino, Y. (2005). Neurochemical studies of autism. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 453–72. Wiley, New York.
39. Rett, A. (1966). Uber ein eigenartiges hirntophisces Syndröm bei hyperammonie intramuscular Kindersalter. *Wein Medizinische Wochenschrift*, **118**, 723–6.
40. Hagberg, B., Aicardi, J., Dias, K., *et al.* (1983). A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome, report of 35 cases. *Annals of Neurology*, **14**, 471–9.
41. Rutter, M. (1994). Debate and argument: there are connections between brain and mind and it is important that Rett syndrome be classified somewhere. *Journal of Child Psychology and Psychiatry*, **35**, 379–81.
42. van Acker, R., Loncola, J.A., and van Acker, E.Y. (2005). Rett syndrome: a pervasive developmental disorder. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 126–64. Wiley, New York.
43. Trevathan, E. and Adams, M.J. (1988). The epidemiology and public health significance of Rett syndrome. *Journal of Child Neurology*, **3**(Suppl.), S17–20.

44. Gillberg, C. (1994). Debate and argument: having Rett syndrome in the ICD-10 PDD category does not make sense. *Journal of Child Psychology and Psychiatry*, **35**, 377–8.
45. Rett Syndrome Diagnostic Criteria Work Group. (1986). Diagnostic criteria for Rett syndrome. *Annals of Neurology*, **23**, 125–8.
46. Hagberg, B. (1985). Rett's syndrome: prevalence and impact on progressive severe mental retardation in girls. *Acta Paediatrica Scandinavica*, **74**, 405–8.
47. Kozinetz, C.A., Skender, M.L., MacNaughton, N., *et al.* (1993). Epidemiology of Rett syndrome: a population-based registry. *Pediatrics*, **91**, 445–50.
48. Hagberg, B. (1989). Rett syndrome: clinical peculiarities, diagnostic approach, and possible cause. *Pediatric Neurology*, **5**, 75–83.
49. Kaufmann, W.E., Johnston, M.V, and Blue, M.E. (2005). MeCP2 expression and function during brain development: implications for Rett syndrome's pathogenesis and clinical evolution. *Brain & Development*, **28**(Suppl. 1), S77–87.
50. Heller, T. (1908). Dementia Infantilis. Zeitschrift fur die Erforschung und Behandlung des Jugenlichen. *Schwachsinns*, **2**, 141–65.
51. Volkmar, F.R. and Rutter, M. (1995). Childhood disintegrative disorder: results of the DSM-IV autism field trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, **34**, 1092–5.
52. Volkmar, F.R., Koenig, K., and State, M. (2005). Childhood disintegrative disorder. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 70–87. Wiley, New York.
53. Rutter, M. (1985). Infantile autism and other pervasive developmental disorders. In *Child and adolescent psychiatry—modern approaches* (eds. M. Rutter and L. Hersov), pp. 545–66. Blackwell, London.
54. Corbett, J. (1987). Development, disintegration and dementia. *Journal of Mental Deficiency Research*, **31**, 349–56.
55. Asperger, H. (1944). Die 'autistichen Psychopathen' intramuscular Kindersalter. *Archive für Psychiatrie und Nervenkrankheiten*, **117**, 76–136. [Translated by U. Frith in U. Frith (ed.) (1991). *Autism and Asperger syndrome*, pp. 37–92. Cambridge University Press, New York.]
56. Wing, L. (1981). Asperger's syndrome: a clinical account. *Psychological Medicine*, **11**, 115–29.
57. Wolff, S. and Barlow, A. (1979). Schizoid personality in childhood: a comparative study of schizoid, autistic and normal children. *Journal of Child Psychology and Psychiatry*, **20**, 29–46.
58. Rourke, B. (1989). *Nonverbal learning disabilities: the syndrome and the model*. Guilford Press, New York.
59. Bishop, D.V. (1989). Autism, Asperger's syndrome and semantic-pragmatic disorder: where are the boundaries? *British Journal of Disorders of Communication*, **24**, 107–21.
60. Ellis, H.D., Ellis, D.M., Fraser, W., *et al.* (1994). A preliminary study of right hemisphere cognitive deficits and impaired social judgments among young people with Asperger syndrome. *European Child and Adolescent Psychiatry*, **3**, 255–66.
61. Klin, A., Volkmar, F.R., Sparrow, S.S., *et al.* (1995). Validity and neuropsychological characterization of Asperger syndrome: convergence with nonverbal learning disabilities syndrome. *Journal of Child Psychology and Psychiatry*, **36**, 1127–40.
62. Klin, A., Pauls, D., Schultz, R., *et al.* (2005a). Three diagnostic approaches to Asperger syndrome: implications for research. *Journal of Autism and Developmental Disorders*, **35**, 221–34.
63. Klin, A., McPartland, J., and Volkmar, F.R. (2005b). Asperger syndrome. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 88–125. Wiley & Sons, New York.
64. Lincoln, A., Courchesne, E., Allen, M., *et al.* (1988). Neurobiology of Asperger syndrome: seven case studies and quantitative magnetic resonance imaging findings. In *Asperger syndrome or high functioning autism*? (eds. E. Schopler, G.B. Mesibov, and L.J. Kunc), pp. 145–66. Plenum, New York.
65. Miller, J.N. and Ozonoff, S. (1997). Did Asperger's cases have Asperger disorder? A research note. *Journal of Child Psychology and Psychiatry*, **38**, 247–51.
66. Fombonne, E. (2003). The prevalence of autism. *The Journal of the American Medical Association*, **289**, 87–9.
67. Asperger, H. (1979). Problems of infantile autism. *Communication*, **13**, 45–52.
68. Tsatsanis, K. (2003). Outcome research in Asperger syndrome and autism. *Child and Adolescent Psychiatric Clinics of North America*, **12**, 47–64.
69. Towbin, K. (2005). Pervasive developmental disorder not otherwise specified. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 165–200. Wiley, New York.
70. National Research Council. (2001). *Educating children with autism*. National Academy Press, Washington, DC.
71. Scahill, L. and Martin, A. (2005). Psychopharmacology. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 1102–21. Wiley, New York.
72. Handleman, J.S., Harris, S.L., and Martins, M.P. (2005). Helping children with autism enter the mainstream. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 1029–42. Wiley, New York.
73. Schreibman, L. and Ingersoll, B. (2005). Behavioral interventions to promote learning in individuals with autism. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 882–96. Wiley, New York.
74. Prizant, B.M. and Wetherby, A.M. (2005). Critical issues in enhancing communication abilities for persons with autism spectrum disorders. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 925–45. Wiley, New York.
75. Research Units on Pediatric Psychopharmacology Autism Network. (2002). A double blind, placebo-controlled trial of risperidone in children with autistic disorder. *The New England Journal of Medicine*, **347**, 314–21.
76. McDougle, C.J., Scahill, L., Aman, M.G., *et al.* (2005). Risperidone for the core symptom domains of autism: results from the study by the autism network of the research units on pediatric psychopharmacology. *The American Journal of Psychiatry*, **162**, 1142–8.
77. Marans, W., Rubin, E., and Laurent, A. (2005). Addressing social communication skills in individuals with high-functioning autism and Asperger syndrome: critical priorities in educational programming. In *Handbook of autism and pervasive developmental disorders* (3rd edn) (eds. F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen), pp. 977–1002. Wiley, New York.

9.2.4 Attention deficit and hyperkinetic disorders in childhood and adolescence

Eric Taylor

Introduction

The concept of ADHD arose from neurological formulations, but does not entail them, and the modern definition simply describes a set of behavioural traits. The historical evolution of the concept was described by Schachar.[1] It began with the idea that some behavioural problems in children arose, not from social and familial adversity, but from subtle changes in brain development. The term 'minimal brain dysfunction (MBD)' was often applied, and covered not only disorganized and disruptive behaviour but other developmental problems (such as dyspraxias and language delays) presumed to have an unknown physical cause. MBD, however, stopped being a useful description when studies of children with definite and more-than-minimal brain damage made it plain that they showed a very wide range of psychological impairment, not a characteristic pattern (see Harris, this volume); and therefore it was invalid to infer the presence of brain disorder from the nature of the psychological presentation.

The successor to the concept of MBD was attention deficit and hyperactivity: defined, observable behaviour traits without assumption of cause. 'Attention Deficit/Hyperactivity Disorder' (ADHD) in DSM-IV, and 'Hyperkinetic Disorder' in ICD-10, describe a constellation of *overactivity, impulsivity* and *inattentiveness*. These core problems often coexist with other difficulties of learning, behaviour or mental life, and the coexistent problems may dominate the presentation. This coexistence, to the psychopathologist, emphasizes the multifaceted nature of the disorder; to the sociologist, a doubt about whether it should be seen as a disorder at all; to the developmentalist, the shifting and context-dependent nature of childhood traits. For clinicians, ADHD symptoms usually need to be disentangled from a complex web of problems. It is worthwhile to do so because of the strong developmental impact of ADHD and the existence of effective treatments. Public controversy continues, but professional practice in most countries makes ADHD one of the most commonly diagnosed problems of child mental health.

Clinical features

Overactivity

The idea of overactivity refers simply to an excess of movement. It is not totally dependent on context and cannot be reduced to non-compliance: physical measures of activity level have indicated that it is higher in children with ADHD than in controls, even during sleep.[2] It is, however, partly dependent upon context: it is often inhibited by a novel environment, creating a pitfall for the inexperienced diagnostician who may exclude it incorrectly because it is not manifested during observation at a first clinic visit. It may not be evident in situations where high activity is expected, such as the games field. The key situations where it is evident are familiar to the child and where calm is expected, such as visiting family friends, attending church, mealtimes, homework and—often the most troublesome—at school, during class.

(a) Impulsivenes

Impulsiveness means action without reflection—often described as a failure to 'stop and think'. The term covers premature, unprepared and poorly timed behaviours—such as interrupting others, and giving too little time to appreciate what is involved in a school task or a social situation.

(b) Inattentiveness

Inattentiveness means disorganized and forgetful behaviour: short-sequence activities, changing before they are completed, with a lack of attention to detail and a failure to correct mistakes. All these are behavioural observations, not psychological constructs. At a cognitive level, 'Attention deficit' is a rather poor descriptor; the performance of affected children does not fade with time on a task any more than that of ordinary people and the presence of irrelevant information ('distractors') does not worsen their performance disproportionately to that of other people.[3] There are cognitive changes of a different kind (see 'Aetiology' below); but the diagnosis of inattentiveness depends on descriptions and observations of behaviour rather than on tests of performance.

Many other behavioural changes characterize some children with ADHD. They are, for instance, often irritable and their emotions can flash very rapidly when provoked. They may sleep badly (and this in turn can contribute to poor concentration). They can be aggressive to other people and non-compliant to authority. They can also be charming, humourous, inquisitive and intuitive. None of these, however, are either constant in ADHD or confined to those with ADHD. They are worth noting, but they do not make the diagnosis.

Classification

Attention Deficit/Hyperactivity Disorder in DSM-IV is defined as the presence of a number, above a cut-off, of behaviours considered to reflect the cardinal features described above to a degree that is developmentally inappropriate and gives rise to some impairment in more than one setting (e.g. school and home).[4] Overactive and impulsive behaviours are considered together as a single construct of 'hyperactivity-impulsiveness', and for convenience the combined dimension will be referred to in this chapter as 'hyperactivity'. Examples of inattentive behaviour are added together and form a separate dimension. There are therefore three subtypes: hyperactive-impulsive, inattentive and combined.

The same problems characterize the ICD-10 definition of 'Hyperkinetic Disorder' (HD),[5] but with added requirements: especially, that all three cardinal features are present, pervasively across home, school and other situations. HD is therefore, in effect, a subtype of ADHD.[6,7] The subtypes of hyperkinetic disorder are based on the presence or absence of conduct disorder—and indeed the presence of conduct disorder is an important factor with which to reckon in the course.

Diagnosis

Description of the symptoms makes them sound easy to recognize, and indeed the problems are usually very salient, disruptive to

other people, and common causes of referral to health and special education services. Nevertheless, there are pitfalls in the diagnosis, making it necessary for a specialist assessment to be undertaken before the diagnosis is given.

Ambiguous criteria

The behavioural problems, described in outline above, are translated into detailed criteria in DSM-IV and ICD-10, and some of them can be ambiguous. For example, 'does not follow instructions' is a DSM item intended to imply that instructions are forgotten or not attended to; but the behaviour can also be shown for reasons of wilfulness and therefore a part of oppositional/defiant disorder. Careful description or witnessing the behaviours complained of is necessary.

Confusion of cardinal and associated features

Many behavioural problems—such as temper tantrums, sleeplessness, aggression and disobedience to adult authority—are common in children with ADHD, and may be the key reasons for presentation. It is easy to make the mistake of diagnosing ADHD when only disruptive behaviour is present. Much of the confusion comes from the way impulsiveness is operationalized in the diagnostic schemes. Behaviours, such as calling out in class and interrupting others, can indeed come from a difficulty in holding oneself back; but they can also represent deliberate flouting of the rules, and in a London survey of 6- to 8-year-old boys they were as common in non-hyperactive but defiant children as in the hyperactive.[8] Direct observation can usually make the distinction—watching either the children tackling tasks requiring them to stop and think in the clinic, or their natural behaviour in the classroom. Inattentive behaviour also helps to make the diagnosis of ADHD and is less confounded by oppositionality.

(a) Reliance on non-expert judgements

The behaviours of ADHD are continuously distributed in the population (see 'Epidemiology'). The level that is considered normal or acceptable will vary from one culture to another and from one rater to another. To be diagnosed, they should be excessive not only for the child's age but also for the developmental level; and this demands considerable familiarity with the usual range of variation. The diagnostician will acquire this in the course of training and experience; experienced teachers will be excellent judges; but inexperienced or overstressed parents may identify the problems at a low level of hyperactive behaviour, or suppose that an abnormal level is only to be expected in childhood. It is usually helpful to obtain a detailed behavioural account rather than rely on an overall judgment of 'overactivity' or 'failure to concentrate'. Contradiction between sources may occur, and leads to arguments between parents and teachers. This may be due to different expectations, the emotional relationship of raters with the child, or children behaving very differently in contexts that vary in the demands placed on the children. The clinician needs to understand the full context of the way involved adults describe the child.

(b) Problems of recognition in the presence of coexistent problems

It is commonplace for children whose problems meet the criteria for ADHD to show other patterns of disturbance as well. This is often, confusingly, called 'comorbidity' – confusing because it assumes that the other pattern is a distinct disorder, which is only one of the explanations for coexistent problems. Clinicians need to understand the relationships for two reasons; so that they do not make or miss the diagnosis of ADHD; and so that they can make good strategies for treating ADHD in the presence of other disorders and other disorders in the presence of ADHD (see 'Treating Complex cases', below).

(c) Conduct and oppositional disorders

The commonest association and the best researched is with conduct and oppositional disorders. Nearly half the children with hyperactive behaviour in a community survey showed high levels of defiant and aggressive conduct as well; but the associations of the two problems were different, with hyperactivity (but not conduct problems) being associated with delays in motor and language development.[8] Genetic research indicates higher heritability for hyperactivity than for defiant and aggressive behaviour; but there are some genetic influences that are common to both.[9] Hyperactivity is more responsive to stimulant medication than are less hyperactive forms of conduct problem.[10]

When both ADHD and conduct problems are present, then the combined diagnosis ('hyperkinetic conduct disorder' in ICD-10) shows the associations of both disorders. ADHD is therefore not to be diagnosed by the absence of conduct disorder features, but by the clear presence of the core problems of inattentivess and disorganization.

(d) Tourette disorder and multiple tics

A different kind of differential is presented by children with Tourette disorder. Their motor restlessness may indeed represent the coexistence of ADHD, but can result directly from tics. If a child's tics are very frequent and there are a large number of them, then their repetitive and stereotyped nature may not be apparent and they may be seen simply as restless fidgetiness. Again, direct observation of the pattern of overactivity is the key. When there is doubt, videorecording the child and subsequent slow-motion review may make repetitive patterns evident.

(e) Autism spectrum disorders

Children with autism have clear and characteristic impairments of language, communication, and social development. Spectrum disorders, however, can raise diagnostic challenges. Children with ADHD alone often show language delays (usually of an expressive nature with over-simple utterances, by contrast with the receptive difficulties and idiosyncratic patterns of autism). Their attention difficulties may make them unresponsive to the overtures of others in a way that can simulate the social obliviousness of people in the spectrum of autism, and they are often friendless—not because of lack of interest in others but because of the capacity of hyperactive behaviour to irritate other people. Indeed, attention problems can extend to perseverativeness on certain activities such as video games that may be mistaken for the restricted interests of autism. All these factors can lead to ADHD being mistaken for autism, but the reverse can happen too. There are other reasons for overactive behaviour in autism. First, stereotyped patterns of driven overactivity can be seen: they are not disorganized or impulsive and are often made worse by change and novelty (which usually reduce the overactivity of ADHD). Second, episodic bursts of extreme activity can be seen and may be best regarded and treated as catatonic. Third, akithisia may result from neuroleptic medication, or irritable

restlessness from anticonvulsants, and it will be necessary to establish a clear history that ADHD has been a persistent trait.

(f) Attachment disorders

Reactive attachment disorder (RAD) may share with ADHD a disinhibited style of relating to other people (an unreserved but shallow making of social contact). Children with RAD, however, tend to be controlling rather than disorganized, and vigilant rather than inattentive; and inattention and impulsiveness are not cardinal features of RAD; so it is not difficult to recognize both patterns when present in an individual child. The confusion in practice often comes from theoretical misconceptions. Those caring for neglected or abandoned children may consider that the diagnosis of ADHD cannot be accurate because the cause of the children's problems is clearly to be found in their early deprivation. The causal pathway may indeed be that of neglect (though genetic inheritance and fetal exposure to toxins also need considering); but ADHD is a descriptive category, not an explanatory one. If the pattern of ADHD is present it still needs recognizing—not least because the cause of the ADHD behaviour does not seem to determine the response to stimulant medication, and children who have encountered neglect or abnormal early attachment may still have their ADHD problems reduced by medication.[10]

(g) Bipolar disorders

Both ADHD and manic conditions are characterized by overactivity, overtalkativeness, a sensation of whirling thoughts, and often by irritable mood. The distinction is made by the presence in bipolar disorder of episodicity, euphoria, and grandiosity. A suggestion that these distinguishing features are not in fact present in childhood bipolar disorder has naturally led to great overlap between the expanded childhood bipolar diagnosis and ADHD with poor emotional regulation, and further research will be needed to clarify whether there is a distinction.

In all these differential diagnoses, the principle is to establish that the child shows not only overactive behaviour, but the specific pattern of ADHD. Experienced judgement may be required, and the practice of diagnosing on the basis of questionnaire scores alone risks overidentification.

In adult life, there are still more possibilities for misdiagnosis. The commonest reasons for uncertainty are in distinguishing from atypical bipolar disorder and the effects of substance misuse. 'Personality disorder' is sometimes applied; and indeed ADHD shares with personality disorders a long-standing trait quality, but can also be a more precise way of describing the difficulties presented. Differentiation from the normal range of variation can be difficult in the absence of clear standards. The task of the diagnostician is harder when adults are presenting for the first time if only self-report is available; the self-description of hyperactivity may be a form of self-depreciation.

Methods of recognition

(a) Rating scales and informant interviews

Questionnaire ratings by parents or teachers are very useful for screening purposes, and in group studies they give a fairly good discrimination between people with a clinical diagnosis of ADHD and controls from the ordinary population.[11] Many are available[12,13] and the most famous are those from Conners, which yield several different scoring systems; and derivatives such as the Iowa Conners, the SWAN and SNAP scales.[14] They do however leave a fair number of individuals misclassified, and are not suitable as the sole means of establishing a diagnosis. A detailed interview with parents establishes what actual behaviours are the basis for ratings, allows professional judgement to be included, and remains the most informative single method.

(b) Psychiatric interview

Interview with the child is valuable for the observation of attention and social interaction that it yields, and for understanding a child's view of their predicament. Children, however, are not good witnesses about their own concentration and impulse control, and even affected adults are not good at describing themselves in these terms. The experience of ADHD is usually one of suffering the reactions evoked from other people, or an experience of repeated failure. Adults often describe an experience of whirling and interrupted thoughts (in the absence of manic features); and some children will say the same, especially if treatment has enabled them to make a comparison with another way of being.

(c) Investigating underlying causes

Assessment needs not only to distinguish ADHD from related disorders, but to consider whether the ADHD pattern may result from remediable causes. The anamnestic history is by far the most productive investigation. It should include whether hearing problems have been excluded by previous testing (and, if not, an expert assessment should be arranged), and any injuries or diseases are potentially damaging the brain. The strengths and weaknesses of the family environment need to be assessed; they may dictate the choices of treatment. Physical examination should be sufficient to detect congenital anomalies, skin lesions, and motor abnormalities that can be the pointers to a neurological cause. Psychometric assessment is desirable whenever there are problems at school, both to generate an idea of developmental level against which the 'developmental inappropriateness' of behavioural symptoms can be judged, and to detect barriers to learning that may be the reason for inattentiveness. Special physical investigations are not routinely necessary. EEG often yields evidence of immaturity, but this does not advance assessment much and is not routinely indicated. It is valuable in the investigation of epilepsy and in the rare cases when deterioration of function suggests the possibility of a degenerative disorder. Blood tests should be planned only on the basis of history and examination, but may include tests of thyroid function, lead (in high-lead areas) chromosomal integrity (including fragile-X probe) when there is other evidence of developmental delay, and specific DNA tests when there is clinical suspicion of a phenotype such as that of Williams syndrome.

Epidemiology

Prevalence estimates vary widely; but most of the variation between studies comes from differences in definition[6] A community survey in London of more than 2000 6–8-year-old boys found a continuum of severity on rating scales: at each successively higher level of hyperactive behaviour there were successively fewer number of children[8] The genetic evidence also supports a continuum: in a population-based twin study, the influences on hyperactive behaviour were similar over the whole range of variation.[15] Estimates of prevalence are therefore critically dependent upon the cut-off point chosen.

Two major influences on the cut-off are the diagnostic criteria applied and the cultural attitudes of raters. Attention Deficit/Hyperactivity Disorder has a rate in the school age population usually given at about 5 per cent, but varies from 2.4 to 9 per cent[6,16] probably depending on how rigorously 'impairment' is defined. The ICD-10 diagnosis of Hyperkinetic disorder yields rates around 1 to 2 per cent of the school age population.[6,17] Sex differences are marked: population surveys suggest that 2–3 boys are affected for every girl.[18]

The frequency of hyperactive behaviour in the population, at least as indexed by rating scales in surveys, has not been increasing over the last two decades.[19] By contrast, there have been large increases in the frequency with which hyperactivity as a medical condition has in practice been recognized—most obviously evidenced by a great increase in the rates of stimulant prescription between 1995 and 2005 in the UK (Wong *et al*, in submission) and a continuing increase in the USA.[20] The studies suggest that stimulant medication is given for about 3 children per 1000 in the UK (i.e. about 12 per cent of those in the community meeting ADHD criteria with impairment) and about 40 per 1000 in the USA. It is likely that health service organization plays a part in determining recognition. In the USA survey, a diagnosis of ADHD was more likely to have been made for children whose families Carried Health Insurance.[20] In a UK survey, children with high hyperactivity as rated by teachers and parents seldom received a diagnosis, with the main filter coming at the level of recognition by primary health care services.[21]

In adult life, those who were hyperactive as children still have an elevated rate of hyperactivity and related social impairment (reviewed systematically by Faraone *et al.*).[22] Indeed, a cross-sectional population survey of adults described a surprisingly high prevalence rate of about 4 per cent, with a high rate of co-existent psychological morbidity.[23] More evidence is needed on the extent of the adult problem. It is however clear that a substantial number of adults, who were not diagnosed in childhood, may be affected, and an increasing number are presenting for the first time to adult services.

Aetiology

Genetic inheritance

Genetic influences are strong: Twin studies suggest a heritability around 80 per cent, making it one of the psychological disorders most strongly influenced by genetic inheritance,[24,25] and adoptive family studies concur in emphasizing the strength of association with biological relatives.[26] Indeed, several DNA variants in genes coding for relevant proteins have now been identified and replicated.[27] In particular, the genes coding for the dopamine D4 and D5 receptors, the dopamine transporter, SNAP25 (affecting synaptosomal protein), the serotonin 1b receptor and the serotonin transporter have all been associated with ADHD by more than one group of investigators. Several kinds of caution are, however, needed in interpreting these findings. The odds ratios are all quite small (between 1.1 and 1.5), no polymorphism so far found is either necessary or sufficient; it is possible that there are subtypes of ADHD with different genetic influences[28]

Current research continues to seek more associated genes, especially by genome scans and positional cloning and to emphasize the likely importance of gene-environment interactions. Individual studies have reported that the risk alleles for genes in the dopamine system magnify the effects on the foetus of maternal smoking and alcohol consumption during pregnancy,[29,30] and catechol o-methyl transferase (COMT) of low birth weight[31]

Environment

Environmental influences are reviewed by Taylor & Warner Rogers.[32] There are associations with several kinds of adversity in fetal and early postnatal life; and genetic factors may influence the exposure to some hazards (e.g. to lead, via playing in contaminated areas) as well as their impact. Many of the insults have generalized effects on brain development and can also lead to low IQ.

Prenatal

The prenatal factors implicated include smoking and drinking in pregnancy,[33] cocaine,[34] maternal stress during pregnancy,[35] anticonvulsant use[36] and the factors causing very low birth weight.[37] For some of these, there is experimental evidence for a harmful effect in animals. Smoking, for example, has high biological plausibility: the substances inhaled have an effect in animal models, and there is a dose-response relationship in human studies.[38] It is important to recognize these risk factors in assessing a referred child, because one may be able to prevent a subsequent child from suffering the same injury. Interpretation of a positive history, however, is not straightforward, because of the likely effects of genetic influences as well. There is no doubt of the existence of the fetal alcohol syndrome, nor that it can include ADHD symptoms, but the effect of lesser degrees of exposure is uncertain. Apparent associations could be magnified by gene-environment correlations. Maternal drinking may be influenced by the same genes that influence ADHD; the genes and the pregnancy toxin may be handed down together. Knopic *et al.*[39] investigated this by studying the offspring of mothers who were identical twins yet differed in whether they had a history of alcohol abuse: ADHD was common in both groups: the suggestion was that the genes were more important than the presumed exposure to alcohol.

Postnatal

In postnatal life, the best defined risks are at extreme levels of misfortune. Head injury and brain disease have to be severe before they have a causative effect; minor injury is often a result of hyperactivity rather than a cause.[32] Children who experienced extreme deprivation in the orphanages of Romania showed increased rates of pervasive and persistent overactivity and inattention in later childhood, even though they had been adopted into English families before the age of 4 years.[40] Minor degrees of psychological adversity have not been shown to cause ADHD (though they may well be associated with coexistent conduct disorder). Indeed, the twin studies that show genetic influences can also be used to distinguish between the environment that all children in the family share (such as a chaotic family life style or the use of television), and the environmental influences that affect one child but not another; only the latter play a part.

Diet is often blamed for hyperactivity. There is some truth in it, but the effects seem to be modest. The main evidence comes from therapeutic trials (see under 'Treatment') which indicate that a range of foodstuffs can be harmful for individual children – including cow's milk, wheat flour, eggs, and artificial colourings and additives. Individual idiosyncrasies seem more important than a damaging

effect of the substances on everyone. Experimental trial, however, giving colourings including tartrazine to an unselected population of preschool children, suggests that the substances have a small but measurable adverse effect on behaviour across the whole range and so ought to be seen as mildly toxic.[41]

Pathogenesis

The effect of these aetiological influences on the developing brain is being clarified by the neuroimaging possibilities being created by magnetic resonance and other non-invasive techniques. Several brain areas are smaller in ADHD than controls.[42] The difference persists through adolescence into adult life and is more marked in those who have never received medication than those who have. The areas most affected—frontal, striatal and cerebellar—are involved in self-organizational abilities that fail in those with ADHD.

At a neuropsychological level, there have been extensive comparisons between young people referred for, and diagnosed with ADHD in the USA and age-matched controls without psychopathology. 'Executive function'—which has become a broad and ill-defined term for psychological processes by which people modify their responsiveness to stimuli or the organization of their responses—has received special attention and is reviewed by Willcutt *et al.*[43] In general summary, many such functions show significant differences between ADHD and controls, but the effect sizes are modest and do not suggest that research has yet hit on either a fundamental deficit or on a means of diagnosis to replace behavioural description.

Motor inhibition and cognitive inhibition have received particular attention, deriving from the behavioural observation that children with ADHD can be described as 'disinhibited', and from an influential suggestion by Barkley[44] that failures of inhibitory process could underlie the other cognitive deficits – such as inefficient planning ahead, and poor self-control by internal language. There is not much doubt that experiments reliably produce poor performance in ADHD on tests of suppressing motor responses.[45] Indeed, functional neuroimaging has found that people with ADHD, as a group, show less activation of brain structures involved in response suppression, even when they are performing at a satisfactory level on a simple test.[46] There is more uncertainty about whether this form of impulsiveness does indeed derive from deficits in inhibition or from other kinds of psychological alteration, such as reluctance to put effort into planning responses of any kind, or to be patient during a period of waiting. This last idea, 'delay aversion', has been elaborated and tested[47] and suggests that some children with hyperactive behaviour are still capable of delaying a response when appropriate provided that the length of time they have to wait for the reward is controlled. A head-to-head comparison of inhibition failure action (in a 'stop' test) and delay aversion (in a test of delaying gratification) has been carried out, with the result that either test on its own produced a moderate distinction between ADHD and controls, but combining the two resulted in a much better discrimination, with sensitivity and specificity around 80 per cent.[48]

The clinical applicability of the extensive research investment in psychological testing is rather small. The tests have for the most part lacked either standardization or establishment of test–retest reliability; the interpretation of an individual child's score, accordingly, lacks quantitative support. There are a few tests of related abilities that have normative values with age standardization (e.g. the Tests of Everyday Attention for Children: TEACH). Their place in practice is not to make a diagnosis of ADHD, but to suggest which of several possible cognitive weaknesses apply in the individual child. In principle, useful advice for education could follow from such testing; but evaluations—of the uptake by teachers, of the advice or the impact on the child—are lacking.

Course and prognosis

First 3 years

A 'difficult temperament' in early childhood includes overactivity and poor self-regulation, and can have a harmful effect on parent–child relationships; but the concept of inattentiveness is hard to apply at this age and the diagnosis would be insecure.

Age 3–6

ADHD behaviours are clearly recognizable by this age, and there is a strong likelihood of persistence into the school years.[49] Parent training is an effective intervention (see 'Treatment') and should be available for parents with children at risk, without waiting for formal diagnosis.

Age 7–11

School and peer demands make ADHD behaviours impairing; the tolerance of families and the culture at large help to determine whether ADHD is seen as a problem; and this is a very common age for referral and diagnosis. Hyperactivity (as opposed to inattentiveness alone) becomes important in generating aggressive and antisocial behaviour and delinquency.[50] The extent to which there is a poor social outcome may depend upon genetic influences,[51,52] on environmental influences such as a hostile home atmosphere,[53] and on gene–environment interactions (a COMT gene polymorphism together with a low birth weight predicted the development of antisocial symptoms in those with ADHD).[31]

Age 12–18

During adolescence, there is a maturing in the abilities of self-control, and some children with ADHD will lose their problems; but the demands for self-control rise as well, and so the children are still more impulsive and inattentive than their peers and four times as likely to merit a psychiatric diagnosis.[54] Indeed, about half of cases diagnosed in childhood will retain the full diagnosis in adolescence.[55]

Those who continue to show hyperactivity are at risk for other problems, notably aggressive and antisocial behaviour and delinquency,[50] and motor traffic accidents.[56]

Adult life

By adult life, most will no longer meet full diagnostic criteria for ADHD; but, equally, most will retain some functional impairment related to hyperactivity.[22] This should imply a falling prevalence, but survey of adults has found high rates (about 4 per cent).[23] Some part of this discrepancy may derive from adults developing impairment for the first time; they may have had ADHD symptoms as children, but the symptoms were not impairing, and have only become impairing when adult life imposes responsibility and high expectations.

The implications for practice are that from childhood to early adult life, and perhaps longer, severe levels of hyperactivity and inattentiveness should be seen as potentially chronic disability; and that intervention should not target only the core symptoms but also the surrounding tangle of adverse personal relationships and educational failure.

Treatment evaluations

Medication

There have been many trials of central nervous stimulants (especially methylphenidate, with some work on dexamfetamine and pemoline) and atomoxetine. A systematic review was undertaken by NICE (National Institute for Clinical Excellence).[57] Sixty-five trials met quality criteria and were assessed. Quantitative review indicated heterogeneity among the trials, so a meta-analysis was not attempted; but there was no doubt about the superiority of methylphenidate, atomoxetine, and dexamfetamine to placebo. Economic analyses were undertaken and were not very robust, but suggested that all three gave acceptable cost per Quality-Adjusted Life Year. All three should be in clinical use, with the decision regarding which product to use to be based on comorbidity, adverse effects, compliance, potential for drug diversion, and individual preferences with differences in cost as a secondary consideration.

Several proprietary preparations of methylphenidate have appeared that offer an extended release through the day; they differ in the physics of their delivery systems and therefore in their speed of onset and duration of action. Banaschewski *et al.*[58] made a systematic review of trials on them and on atomoxetine, which also has a sustained effect through the day. They indicated that the effect size of extended-release methylphenidate preparations was comparable to that of immediate-release—around 0.8–1.1 SD; but, not surprisingly, the effect of an 8-hour preparation was somewhat smaller than that of a 12 preparation on parent ratings, though similar on teachers' ratings of child behaviour. The effect size of atomoxetine was around 0.6 SD.

Most studies have been carried out on children and adolescents of school age. In children under 6 years, the limited trial evidence suggests that stimulants are more effective than placebo in reducing hyperactivity and the level of stress in family relationships[59] The safety of the drugs in this age group is uncertain. For adults, enough randomized controlled trials have appeared for stimulants and atomoxetine that meta-analysis has been possible, with the conclusion that they are more effective than placebo.[58,60]

Psychological evaluations

Behaviour therapy has received several trials, but no satisfactory systematic review has yet appeared. Miller *et al.*[61] attempted one, but decided to exclude most of the trials because they did not meet the quality criteria that were imposed. Nevertheless, reasonably good effect sizes have been reported in randomized trials comparisons for the comparisons of behaviour modification programmes (usually delivered on an individual family basis) with no treatment or treatment as usual.[62,63] Group programs of parent training—which typically include supportive education in behavioural management—are also effective, perhaps particularly for preschool children.[64,65] Cognitive therapy, by contrast, has been disappointing in trials.[66]

Elimination diets

Several trials of eliminating foods that seem to be incriminated for an individual child, followed by double-blind administration of those foods in experimental design, have found that the identified foods can worsen that child's behaviour more than a placebo.[67] The implication, as for food effects on disorders such as eczema, is of idiosyncratic intolerances so that each child needs investigating individually. This is troublesome for families, and perhaps only applicable to younger children whose diet is still under parental control.

Drug vs. psychosocial intervention

There has been controversy over the relative merits of medication and behaviour therapy. In the USA, the debate has been sharpened by a perceived over-prescription of drugs and led to a large-scale random-allocation non-blind trial.[68] The trial compared rather idealized versions of: medication (with very careful and systematic monitoring of dose and response), behaviourally oriented psychosocial therapy (delivered with high intensity and a combination of approaches to teachers, parents and the young people themselves), both interventions given together, and a 'treatment as usual' policy of referring back to community agencies (which usually resulted in medication). At the prime outcome point—14 months after randomization—the outcome for those given the research style of medication was better than those given behavioural treatment only and considerably better than those given treatment as usual, even when that included medication. Adding medication to behaviour therapy improved the outcome for the primary measures of hyperactive behaviour; adding behaviour therapy to medication did not—but did yield better control of aggression at home, improvement in the overall sense of satisfaction of parents, lower medication dosage, and a higher rate of very good outcomes. These improvements in the combination treatment were real, but very expensive to achieve, and it remains uncertain whether such benefits could be matched by behaviour therapy delivered under the constraints of ordinary practice. The marked superiority of careful medication to other forms of intervention did not persist at later follow-up points. At 2 and 3 years after the start of the trial, those who had been allocated to all arms of the trial showed rather similar outcomes. None were untreated, and all groups showed less hyperactivity than at the beginning of the trial, so the finding should not lead to therapeutic nihilism. The likely reasons for the waning of the medication effect are that the drug loses its effect, stops being taken, or depends upon careful and skilled adjustment of dosage in the longer term.

Management

Psychoeducation

Unlike most psychiatric conditions, a diagnosis of ADHD is often sought by parents and welcomed by them. The image, of being a physically caused neurological disease, is often perceived as a relief from the stigma of mental disorder. On the other hand, the media controversy over whether it is a 'real' disorder, and over the use of controlled drugs, leaves some parents confused and fearful.

Assessment on the principles above will have led to an individual formulation of the nature and causes of the impairment. Extended explanation is worthwhile in the longer term. An over-simple

description in terms of a chemical deficiency in the brain may seem a useful starting point but can lead to unrealistic expectations for treatment and frustration with the doctor or, worse, with the child. A model of chronic disability is in keeping with the evidence from longitudinal studies; but needs to be modulated by the good outcome for some children, the improvement for most, and the ability of warm and encouraging parenting to reduce the risks of antisocial behaviour in later childhood and adolescence.[53]

Children's understanding of their problems is also worth a good deal of effort. Little research has so far addressed the issue, but it is important to their ability to cope. They need to know that their problem is understood, that treatments are available, that they can influence their outcome by their own actions, and that the people around them understand all this and can be encouraging. Positive role models are useful: some successful sports stars, performers, politicians and business people have outed themselves as having, and sometimes using, ADHD. Explanations need to be repeated as the young people mature and expect a fuller and more interactive discussion.

Explanation is often needed by teachers as well. They may need to revise their expectations of the level of challenge with which the child can cope; and for some frustration can lead to antagonism towards the child's family. If they already see ADHD as a neurological disease, then the frequent observations of changeability in the children, and of ability to cope sometimes with difficult tasks, may make them reject a neurological cause—and with it the diagnosis and the validity of drug treatment. They may need to know that physical and psychological factors can both enter into the child's presentation and that the effect of medication does not depend on the aetiology.[10]

After explanation comes basic advice on helping the children's development. The first steps with parents are to establish whether there is already a framework of frequent warm interactions and effective ways of giving instructions and following up children's actions with consistent patterns of reward or loss of reward. If this does not already exist, then a parent training group is often helpful. Both a supportive atmosphere and the teaching of skills in behaviour modification seem to be necessary. The target behaviours for modification are often the ones most troublesome to parents—disobedience and aggression—rather than restlessness or inattentiveness specifically.

Liaison with schools should include advice on the severity of the problem and the intensity and nature of extra help that will be required. Teachers will often be able to share good practice in classroom management. One of the principles is to maintain good stimulus control, for instance by having the affected child at the front of the class under the teacher's eye. Another is to find opportunities for the children to let off physical energy (they can sometimes be used as messengers between classrooms) and to learn in short chunks. Variety and interest in the material to be learned or understood is useful. Transitions between activities in the classroom are often the time for children to become disorganized, and the child with ADHD should be the first to change activity with the teacher's supervision. Individual attention is probably the most effective resource in the classroom, but it is also very demanding: a classroom assistant may help to achieve it. Star charts for younger children and token economy systems for older ones are often recommended, but usually depend upon the system used for the rest of the class.

Specific interventions

When straightforward advice is not enough, then the two best-evaluated treatment approaches are *behaviour modification* and *medication management*. The choice of which to start with will depend on several conclusions from the assessment: the severity of the problem (with more severe problems responding preferentially to medication rather than behaviour therapy[7] the availability of treatment; the willingness and ability of parents (or teachers) to engage in psychological intervention; the urgency of the problem (with medication affording a more rapid change); and the wishes of the family. Whichever approach is taken first, the other should be available without undue delay if the response is below expectation.

(a) Behaviour modification

The principles of behaviour modification do not differ from those used in other kinds of behaviour problem (e.g. 62). Target behaviours should be clearly specified and monitored; the antecedents and consequences of the behaviours should be understood and modified as appropriate; clear schemes of reward and punishment should be established, understood by the child, and applied consistently. There are, in addition, some modifications to suggest for the specific needs that come from the nature of ADHD. The rapid delay-of-reward gradient calls for contingencies to be applied with particular attention to speed. For example, a kitchen timer can be set for an appropriate length of activity depending upon the individual child (for instance, 5 min application to homework, or 30 min spent free of aggression to siblings). When the timer sounds, an obvious reward (such as a token) is given within a very few seconds. The reward may swiftly lose its reinforcing quality with repetition, so frequent changes in the reward (or the backup to a token) are needed. Impairment in error correction may make it all the more necessary to be explicit and swift in explaining to the child which of their behaviours has earned the reward, or its loss. Response cost (such as loss of tokens) is usually advocated in conjunction with the reward scheme.

(b) Medication

Prescription of medicines can be guided by published schemes (e.g.[13]). Specialist assessment is highly desirable when problems are at the level that warrants medication—not because the treatment is specially risky, but because it is important that remediable causes and associated conditions are not overlooked. The first choice of medicine is usually methylphenidate. If immediate-release is chosen, then one usually begins with doses, three times a day about 5 mg to 10 mg, depending on the child's weight. If there are no adverse effects, then the dose is increased upwards (probably weekly) until there is a good response, or adverse effects become troublesome, or the ceiling of 0.7 mg/kg/dose is reached—whichever comes first. If an extended-release preparation is chosen, then a similar policy is followed of starting at a low level (e.g. 10–20 mg as a single dose) and titrating in the light of response.

The choice of immediate—or extended—release preparation should be discussed with the family. School children often have a strong preference for a single tablet to be taken in the morning before school, so as to avoid stigmatization. Schools should also be part of the decision making, because of the organizational problems for them of maintaining secure storage and accurate administration. On the other hand, advantages of immediate-release include lower cost and the possibility of accurate control of the profile of action through the day.

Individual variation in drug response is considerable, so good monitoring is a key to achieving good effects. A simple rating scale such as the abbreviated Conners is suitable: a short scale is more likely to be completed than a long one. The wide variety of presentations means that key problems for the individual child may not be included on a standard scale. An individualized scale can therefore be constructed as part of the assessment and used as the prime outcome measure. Ratings by teachers are particularly important, but communication problems can mean that their voice is not heard. If the dose is set only by the level seen as optimal by the parents, then there is a danger of over-treatment. The child's behaviour will be seen at home in the mornings and evenings of schooldays, i.e. at times when the blood level of medication is lower than during school hours. The best dose for mornings and evenings may then lead a child being over-controlled and unspontaneous during school hours. Internet feedback from class teacher can be quick and accessible, but care is needed to maintain confidentiality. Telephone monitoring is useful, especially to allow frequent adjustments in the initial phase of setting dosage, but cannot replace individual contact.

Physically, blood pressure and height and weight need regular checks; mentally, the examiner should be alert to the possibilities that agitation, depression, loss of spontaneity and perseveration can appear as a result of medication and not only as part of the condition.

Under some circumstances, atomoxetine is the medication of first choice. It is not a controlled drug, and does not maintain an illicit market, so it may be preferred if there is a substance-misusing family member. The media controversy over the use of stimulants has entailed that atomoxetine may be acceptable to some families who reject 'Ritalin'. It may also be preferred in the presence of Tourette disorder and perhaps of high levels of anxiety. Children who have failed to respond to a stimulant may nevertheless show a good response to atomoxetine. The balance of adverse effects is somewhat different and atomoxetine may therefore be preferred when, for example, insomnia has resulted from stimulants or is a major problem in itself. The action of atomoxetine may take some weeks to appear, and close titration is not recommended. Rather, a test dose around 0.5 mg/kg is given (in case adverse effects appear even on a small dose), and is followed after a week by an increase to 1.2 mg/kg.

Treatment in comorbid conditions

In general, and as considered in 'Diagnosis' above, the cluster of ADHD symptoms is similar whether or not co-existent disorders are present. The principles are for the most part the same as when treating uncomplicated ADHD. In the most frequent combination—of ADHD and conduct disorder—stimulant treatment can reduce antisocial problems as well as the core of ADHD[57] and is often worth trying even before conduct disorder is addressed.

(a) Anxiety

In the combination of ADHD and anxiety states, there is some trial evidence that the superiority of stimulant to placebo is less than in ADHD without anxiety.[10,69] There may need to be particular attention to monitoring both problems in establishing the correct dose level, and atomoxetine will sometimes be chosen. The reasons for anxiety should be sought and corrected.

(b) Pervasive developmental disorders

When an autism spectrum disorder is also present, then treatment of ADHD with stimulants is possible, but particular attention needs to be given to the possibility of exacerbating social withdrawal and repetitive patterns of behaviour, and monitoring these should be given a priority as high as detecting the desired effects. The RUPP Autism Network[70] treated 72 cases in a design with a 1-week test period, 4 weeks randomized crossover, and 8 weeks of continued treatment for those who responded well. Methylphenidate produced a better reduction of hyperactive behaviour than did placebo. The most satisfactory dose level was a modest 0.25 mg/kg.

(c) Substance misuse

In the presence of substance misuse, many clinicians are wary of prescribing the potentially misusable stimulants. There may be too much hesitation. People with ADHD taking stimulants show lower rates of substance misuse than those who do not take prescribed medication.[71]

(d) Epilepsy

The presence of epilepsy raises extra needs in assessment before treatment. In poorly controlled epilepsy, ADHD symptoms may be the direct result of very frequent small seizures ('absence status' at the extreme) or of very frequent seizures at night, so ambulant and sleeping EEGs are useful. In less extreme cases, brief lapses of attention can be the result of minor seizures causing transient cognitive impairment; simultaneous recording of EEG with behaviour observation and/or psychological test performance is the best way of getting the answer. Anticonvulsant medications can also cause disturbances of attention and irritability; the clues come from high blood levels of anticonvulsants, low folate levels, polypharmacy, and a temporal relationship between drug changes and hyperactivity or inattention. Once these are excluded, then treatment of ADHD can proceed as usual. Methylphenidate, it has been copied from textbook to textbook, can worsen epilepsy. I have not been able to find empirical evidence for this, do not find that it matches with clinical experience and regard methylphenidate as safe in controlled epilepsy. Atomoxetine has had numerous reports of seizures following its use, but very few of first seizures. In uncontrolled epilepsy, and especially where there is a risk for status epilepticus, I prefer to use dexamfetamine.

Review after a satisfactory response

When a child has responded well to the first treatment chosen, a specialist's review of the case is in order.

Is the improvement sufficient? Impulsiveness and inattentiveness may not have disappeared entirely; but the goal should be that they are no longer impairing. If the problems are still more than minor, then the other main intervention should be explored.

What has happened to any co-existent disorders: are they satisfactorily resolved or is further treatment or referral indicated?

Has the improvement led to different understanding of the child? Old habits of reacting to the children or setting expectations for them may need to be modified. At home, parents may well have found that they set disciplinary sanctions at a high level when the need was to have an impact on an inattentive person. Those sanctions may be too severe, for a person who is now more responsive to reward and punishment and lead to distress or discouragement.

At school, there may be opportunities for normalizing the curriculum. Social skills learning, which may have been abandoned in the past because of the child's failure to profit, may now be well worth another try.

Does the child understand the nature of the improvement? This, like the nature of the disability, will need discussing in different ways as the child matures. The initial reaction may well be one simply of relief at being out of trouble. If, however, medicines are seen as a tablet 'to make me good', this attitude may lead to a rejection of medicine in adolescence when rejecting other aspects of adult authority. There are usually many decisions to make about medication—whether to take it during the school holidays and at weekends, whether to vary the dose in line with environmental demands, and whether to continue taking it even though indulging in alcohol or cannabis. There is every reason to involve the child as an active agent in these decisions and to help them to learn from the consequences.

For how long should the treatment continue? Scientific study has not given a secure answer to this question, and individual decision making is needed. Periodic spells off medication—perhaps for a fortnight every two years—is a good way of deciding whether it is still needed. These are also good times for the patient to review why he is taking it, and perhaps to seek the reactions of others to his state off medication.

Are there satisfactory follow-up arrangements? Shared care between primary care and the specialist service is the ideal, with physical monitoring (perhaps 6-weekly) and minor dose adjustments carried out in primary care, and psychological monitoring and strategy decisions about therapy (perhaps 6-monthly) carried out by the specialist team.

Review in refractory cases

A case can be considered refractory when the problems are still impairing after the exhibition of methylphenidate (or dexamfetamine), atomoxetine and behavioural therapy.

There are several reasons for failure at the first line, of which failure to follow treatment is the commonest. If behaviour modification was the first line tried, and the child was too hyperactive for it to be adequately delivered, then it may well be worth another try in combination with medication, even if the medication alone was not obviously successful. If medication was not taken, then the reason may have been stigma and careful discussion may allow a more successful attempt. Public controversy about ADHD and medication has been intense for decades and has not been resolved by increasing knowledge. It is right for there to be strong debate, not least because the issues raised—of whether it is legitimate to make changes in one's learning and social abilities through physical methods—arise in many other areas of public concern. Unfortunately, however, some journalism accuses parents and teachers of bad faith, in pretending a physical cause to disguise failings in parental childrearing or inadequacies in schools. This is understandably disquieting for children and those around them. Involvement with a user group can help to maintain a positive attitude to overcoming disability, and counter some of the noxious attitudes expressed in some of the media.

Another reason for failure of medication is the appearance of adverse effects, either precluding the treatment or limiting the dose to subtherapeutic levels. Symptomatic treatment of adverse reactions is often possible.

Appetite loss can result from stimulants (and less commonly from atomoxetine). This will often disappear towards the end of the day, as the medication is cleared from the body, so increased intake in the evenings (or at weekends or other holidays from stimulant medication) is often enough to prevent faltering in growth.

Insomnia is a frequent complaint, but should be carefully recorded at baseline as it often precedes medication. The commonest problem to result from stimulants is a delay in settling and falling asleep. Sleep hygiene measures can help, for instance, a planned deceleration of activity towards bedtime (perhaps in a place other than the bedroom, to avoid conditioning the bedroom environment to wakefulness); a reduction in light intensity (including sitting farther away from the television or computer screen); and prescription of melatonin shortly before settling. A switch from stimulants to atomoxetine may be needed.

Tics can be worsened or produced by stimulants. Some people find that mild tics are a price worth paying for the beneficial actions, or not even notice them; but they can be disfiguring and stigmatizing. A switch to atomoxetine will often be the first action; or the combination of methylphenidate with clonidine may be useful in both reducing tic severity and reducing the necessary dose of methylphenidate.

Treatment can also fail because the initial assessment was incomplete and set the wrong targets for therapy. Another disorder may have masqueraded as ADHD (see 'Diagnosis') and a failure to respond should raise the index of suspicion for the presence of autistic or hypomanic overactivity. (The converse, however, does not hold; a response to methylphenidate does not make the diagnosis of ADHD, because qualitatively similar changes can be seen when ordinary children receive a stimulant, vide Rapoport *et al.*)[72] A variant of this comes when ADHD was present, but not the main problem, and the main problem perceived by parents or teachers is not drug-sensitive.

When the above reasons for failure to respond to treatment have been considered and dealt with, then the psychiatrist should consider some of the wide variety of unlicensed medications that have been shown effective in randomized controlled trials.

Some noradrenergic agents (clonidine, guanfacine), may act to stimulate presynaptic autoreceptors and may downregulate noradrenergic activation. They can be useful when there is a great deal of agitation as part of the symptom pattern (e.g. in autism). They do not, however, improve the cognitive aspects of inattentiveness.

Modafinil has effects in reducing hyperactivity and may enhance cognition as well. A licence has been applied for, but was interrupted by the possible emergence of skin disorders as a complication.

Tricyclic antidepressants (e.g. imipramine, protriptyline) and some other antidepressants (bupropion, but not SSRIs) are more effective than placebo in reducing hyperactivity; their effect often wanes after a few weeks or months, but they can be useful for short periods, e.g. to allow a period off stimulants in a child with a growth problem.

Monoamine oxidase inhibitors also reduce hyperactivity; they are in general unsafe for use in children, partly because of the difficulty in maintaining dietary restrictions, but the reversible MAOIs such as meclobomide are somewhat safer and could be considered. Nicotine patches can be considered for their combination of cognitive and behavioural effects, but quite often produce nausea and local irritation.

Risperidone and other atypical neuroleptics are often used, especially in intellectually impaired populations, but have not been evaluated for the treatment of ADHD. Their power, and the reason they are prescribed, is in the symptomatic reduction of severely aggressive and agitated behaviour rather than the improvement of attentive and reflective behaviour. Clarity of indication is therefore important, and their benefits should be set against the many hazards.

Some efficacious drugs have been contraindicated because of rare but severe, adverse effects: pemoline after reports of liver failure, desipramine because of cardiac toxicity. In general, the use of unlicensed or unevaluated drugs should be embarked on only by prescribers with specialist experience, who obtain carefully informed consent and monitor appropriately. The most effective medication protocol yet evaluated was that of the MTA approach (see above), in which stimulants were sufficient for about 90 per cent of cases and there was little recourse to the second line of drugs.

Treatment in adult life

A key problem in current knowledge is that of making an accurate diagnosis when adults present for the first time. Adults may be mistaken in identifying themselves (see above); but their recall of their childhoods is a reasonably reliable predictor of their parents' ratings.[73] Self-report scales have emerged[74, 75] but are not yet fully validated. The account of somebody who knows the patient well—perhaps a spouse or a partner—is very desirable, but does of course need interpreting in the light of their own interests in a diagnosis.

Psychosocial treatment for adults is not yet well evaluated. In principle, adults ought to have greater capacities for cognitive and other self-instructional approaches. Young and Bramham[76] provide a useful guide. Simply the giving of a diagnosis comes as a relief to some who have puzzled over the reasons for their failures, and can liberate problem-solving approaches.

Treatment with stimulant drugs and atomoxetine has been evaluated by several randomized controlled trials in adults (reviewed by Faraone *et al.* and Banaschewski *et al.*).[58,60] The drugs are efficacious. Only atomoxetine has a licence in Europe, and that only when treatment was started in childhood; but it does not seem reasonable to withhold a therapy because it was unavailable to the person earlier. Their use follows similar principles to those described above for children, and Asherson[77] provides a guide.

In conclusion, this chapter has presented a picture of ADHD and its severe form, hyperkinetic disorder, as disabilities that change with development and are often accompanied by other problems that can mask it or themselves be masked by it. They are rewarding challenges for diagnosis and treatment in adulthood as well as during childhood and adolescence.

Further information

The National Institute of Mental Health (NIMH) Website: http://www.nimh.nih.gov/health/topics/attention-deficit-hyperactivity-disorder-adhd/index.shtml

The National Attention Deficit Disorder Information and Support Service (ADDISS) Website: www.addiss.co.uk

Eric, T. (ed.) (2007). *People with Hyperactivity: Understanding and Managing Their Problems.* Mac Keith Press.

References

1. Schachar, R. (1986). Hyperkinetic syndrome: Historical development of the concept. In *The Overactive Child* (ed. E. Taylor), Clinics in Developmental Medicine No. 97. MacKeith Press/Blackwells, London.
2. Porrino, L.J., Rapoport, J.L., Behar, D., *et al.* (1983). A naturalistic assessment of the motor activity of hyperactive boys. I. Comparison with normal controls. *Archives of General Psychiatry,* **40**, 681–7.
3. Sergeant, J.A. (2005). Modeling attention-deficit/hyperactivity disorder: A critical appraisal of the cognitive-energetic model. *Biological Psychiatry,* **57**, 1248–55.
4. American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders version IV text revision*. American Psychiatric Association, Washington, DC.
5. World Health Organization (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva, World Health.
6. Swanson, J.M., Sergeant, J., Taylor, E., *et al.* (1998). Attention Deficit Hyperactivity Disorder and Hyperkinetic Disorder. *The Lancet,* **351**, 429–33.
7. Santosh, P.J., Taylor, E., Swanson, J., *et al.* (2005). Refining the diagnoses of inattention and overactivity syndromes: A reanalysis of the Multimodal Treatment study of attention deficit hyperactivity disorder (ADHD) based on ICD-10 criteria for hyperkinetic disorder. *Clinical Neuroscience Research,* **5**, 307–14.
8. Taylor, E., Sandberg, S., Thorley, G. *et al.* (1991). *The epidemiology of childhood hyperactivity. Maudsley Monograph No. 33*. Oxford University Press, Oxford.
9. Nadder, T.S., Rutter, M., Silberg, J.L., *et al.* (2002) Genetic effects on the variation and covariation of attention deficithyperactivity disorder (ADHD) and oppositional-defiant disorder/conduct disorder (Odd/CD) symptomatologies across informant and occasion of measurement. *Psychological Medicine*, **32**(1), 39–53.
10. Taylor, E. A., Schachar, R., Thorley, G., *et al.* (1987). Which boys respond to stimulant medication? A controlled trial of methylphenidate in boys with disruptive behaviour. *Psychological Medicine,* **17**, 121–43.
11. Green, M., Wong, M., Atkins, D., *et al.* (1999). *Technical Review Number 3: Diagnosis of Attention-Deficit/Hyperactivity Disorder*. Rockville, MD: Agency for Health Care Policy and Research, US Department of Health and Human Services, AHCPR Publication 99-0050.
12. Stein, M.T., and Perrin, J.M. (2003). Diagnosis and treatment of ADHD in school-age children in primary care settings: a synopsis of the AAP Practice Guidelines. *Pediatrics in Review,* **24**, 92–8.
13. Taylor, E., Döpfner, M., Sergeant, J., *et al.* (2004). European clinical guidelines for hyperkinetic disorder - first upgrade. *European Child and Adolescent Psychiatry,* **13**(Suppl 1), 17–30.
14. Conners, C.K. (1997) Conners' ADHD/DSM-IV Scales (CADS). www.mhs.com
15. Gjone, H., Stevenson, J., and Sundet, J.M. (1996). Genetic influence on parent-reported attention-related problems in a Norwegian general population twin sample. *Journal of the American Academy of Child and Adolescent Psychiatry,* **35**, 588–96.
16. Ford, T., Goodman, R., Meltzer, H. (2003). The British Child and Adolescent Mental Health Survey 1999: the prevalence of DSM-IV disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, **42**(10), 1203–11.
17. Meltzer, H., and Gatward, R. (with Goodman, R., and Ford, T.) (2000). *Mental health of children and adolesdents in Great Britain*, London, The Stationery Office.
18. Heptinstall, E., and Taylor, E. (2002). Sex differences and their significance. In *Hyperactivity and Attention Disorders of Childhood* (ed. S. Sandberg). Cambridge University Press, Cambridge.
19. Collishaw, S., Maughan, B., Goodman, R., *et al.* (2004). Time trends in adolescent mental health. *Journal of Child Psychology and Psychiatry,* **45**, 1350–62.

20. CDC (2003). *Prevalence of Diagnosis and Medication Treatment for Attention-Deficit/Hyperactivity Disorder*. Center for Disease Control http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5434a2.htm
21. Sayal, K., Taylor, E., Beecham, J., *et al.* (2002). Pathways to care in children at risk of attention-deficit hyperactivity disorder. *British Journal of Psychiatry,* **181**, 43–8.
22. Faraone, S.V., Biederman, J., and Mick, E. (2006). The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. *Psychological Medicine,* **36**, 159–65.
23. Kessler, R.C., Adler, L., Barkley, R., *et al.* (2006). Prevalence of adult ADHD in the United States: results from the National Comorbidity Survey Replication (NCS-R). *American Journal of Psychiatry,* **163**, 716–23.
24. Asherson, P., Kuntsi, J., and Taylor, E. (2005). Unravelling the complexity of attention-deficit hyperactivity disorder: a behavioural genomic approach. *British Journal of Psychiatry,* **187**, 103–5.
25. Thapar, A., O'Donovan, M., and Owen, M.J. (2005a). The genetics of attention deficit hyperactivity disorder. *Human Molecular Genetics,* **14**, R275–R282.
26. Sprich, S., Biederman, J., Crawford, M.H., *et al.* (2000). Adoptive and biological families of children and adolescents with ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry,* **39**, 1432–37.
27. Faraone, S.V., Perlis, R.H., Doyle, A.E., *et al.* (2005). Molecular genetics of attentiondeficit/hyperactivity disorder. *Biological Psychiatry*, **57**, 1313–23.
28. Todd, R.D., Sitdhiraksa, N., Reich, W., *et al.* (2002). Discrimination of DSM-IV and latent class attention-deficit/hyperactivity disorder subtypes by educational and cognitive performance in a population based sample of child and adolescent twins. *Journal of the American Academy of Child and Adolescent Psychiatry*, **41**, 820–8.
29. Kahn, R.S., Khoury, J., Nichols, W.C., *et al.* (2003). Role of dopamine transporter genotype and maternal prenatal smoking in childhood hyperactive-impulsive, inattentive, and oppositional behaviors. *Journal of Pediatrics*, **143**, 104–10.
30. Brookes, K., Mill, J., Guindalini, C., *et al.* (2006) A common haplotype of the dopamine transporter gene associated with attention deficit/hyperactivity disorder and interacting with maternal use of alcohol during pregnancy. *Archives of General Psychiatry*, **63**, 74–81.
31. Thapar, A., Langley, K., Fowler, T., *et al.* (2005b). Catechol O-methyltransferase gene variant and birth weight predict early-onset antisocial behavior in children with attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*, **62**, 1275–78.
32. Taylor, E., and Warner-Rogers, J. (2005). Practitioner review: Early adversity and developmental disorders. *Journal of Child Psychology and Psychiatry,* **46**, 451–67.
33. Linnet, K.M., Dalsgaard, S., Obel, C., *et al.* (2003). Maternal lifestyle factors in pregnancy risk of attention deficit hyperactivity disorder and associated behaviors: Review of the current evidence. *American Journal of Psychiatry,* **160**, 1028–40.
34. Linares, T.J., Singer, L.T., Kirchner, H.L., *et al.* (2006). Mental health outcomes of cocaine-exposed children at 6 years of age. *Journal of Pediatric Psychology,* **31**, 85–97.
35. O'Connor, T.G., Heron, J., Golding, J., *et al.* (2003). *Maternal antenatal anxiety and behavioural/emotional problems in children: A test of a programming hypothesis.*
36. Steinhausen, H.C., Losche, G., Koch, S., *et al.* (1994). The psychological-development of children of epileptic parents. I. Study design and comparative findings. *Acta Paediatrica,* **83**, 955–60.
37. Bhutta, A.T., Cleves, M.A., Casey, P.H., *et al.* (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm - A meta-analysis. *Journal of the American Medical Association,* **288**, 728–37.
38. Thapar, A., Fowler, T., Rice, F., *et al.* (2003). Maternal smoking during pregnancy and attention deficit hyperactivity disorder symptoms in offspring. *American Journal of Psychiatry,* **160**, 1985–89.
39. Knopik, V.S., Heath, A.C., Jacob, T., *et al.* (2006). Maternal alcohol use disorder and offspring ADHD: disentangling genetic and environmental effects using a children-of-twins design. *Psychological Medicine,* **36**, 1461–71.
40. Kreppner, J.M., O'Connor, T.G., and Rutter, M. (2001). Can inattention/overactivity be an institutional deprivation syndrome? *Journal of Abnormal Child Psychology,* **29**, 513–28.
41. Bateman, B., Warner, J.O., Hutchinson, E., *et al.* (2004) The effects of a double blind, placebo controlled, artificial food colourings and benzoate preservative challenge on hyperactivity in a general population sample of preschool children. *Archives of disease in childhood*, **89**(6), 506–11.
42. Castellanos, F.X., Lee, P.P., Sharp, W., *et al.* (2002). Developmental trajectories of brain volume abnormalities in children and adolescents with attention-deficit/hyperactivity disorder. *Journal of the American Medical Association,* **288**, 1740–48.
43. Willcutt, E.G., Doyle, A.E., Nigg, J.T., *et al.* (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry,* **57**, 1336–46.
44. Barkley, R.A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin,* **121**, 65–94.
45. Castellanos, X., Sonuga-Barke, E.J.S., Tannock, R., *et al.* (2006). Characterising Cognition in ADHD: Beyond Executive Dysfunction. *Trends in Cognitive Science,* **10**, 117–23.
46. Rubia, K., Smith, A.B., Brammer, M.J., *et al.* (2005). Abnormal brain activation during inhibition and error detection in medication-naive adolescents with ADHD. *American Journal of Psychiatry,* **162**, 1067–75.
47. Sonuga-Barke, E.J. (2005). Causal models of attention-deficit/hyperactivity disorder: from common simple deficits to multiple developmental pathways. *Biological Psychiatry,* **57**, 1231–8.
48. Solanto, M.V., Abikoff, H., Sonuga-Barke, E.J.S., *et al.* (2001). The ecological validity of measures related to impulsiveness in AD/HD. *Journal of Abnormal Child Psychology*, **29**, 215–28.
49. Lahey, B.B., Pelham, W.E., Loney, J., *et al.* (2004). Three-year predictive validity of DSM-IV attention deficit hyperactivity disorder in children diagnosed at 4–6 years of age. *American Journal of Psychiatry,* **161**, 2014–20.
50. Farrington, D.P. (1995). The Twelfth Jack Tizard Memorial Lecture. The development of offending and antisocial behaviour from childhood: key findings from the Cambridge Study in Delinquent Development. *Journal of Child Psychology & Psychiatry,* **36**, 929–64.
51. Faraone, S.V. (2004). Genetics of adult attention-deficit/hyperactivity disorder. *Psychiatric Clinics of North America,* **27**, 303–21.
52. Mill, J., Caspi, A., Williams, B.S., *et al.* (2006). Prediction of heterogeneity in intelligence and adult prognosis by genetic polymorphisms in the dopamine system among children with attention-deficit/hyperactivity disorder: Evidence from 2 birth cohorts. *Archives of General Psychiatry,* **63**, 462–9.
53. Rutter, M., Maughan, B., Meyer, J., *et al.* (1997). Heterogeneity of antisocial behavior: Causes, continuities, and consequences. In *Motivation and delinquency (ed.* D. W. Osgood), (pp. 45–118). Lincoln, NE: University of Nebraska Press; and *Nebraska Symposium on Motivation,* **44**, 45–118.
54. Taylor, E., Chadwick, O., Heptinstall, E., *et al.* (1996). Hyperactivity and conduct problems as risk factors for adolescent development. *Journal of the American Academy of Child and Adolescent Psychiatry,* **35**, 1213–26.
55. Klein, R.G., and Mannuzza, S. (1991). Long-term outcome of hyperactive children: a review. *Journal of the American Academy of Child and Adolescent Psychiatry,* **30**, 383–7.
56. Barkley, R.A., Fischer, M., Edelbrock, C.S., *et al.* (1990). The adolescent outcome of hyperactive children diagnosed by research criteria: I. An 8-year prospective follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry,* **29**, 546–57.
57. NICE (National Institute for Health and Clinical Excellence). (2005). *Methylphenidate, atomoxetine and dexamfetamine for the treatment of*

attention def cit hyperactivity disorder in children and adolescents. http://www.nice.org.uk/page.aspx?o=TA013
58. Banaschewski, T., Coghill, D., Santosh, P., *et al.* (2006). Long-acting medications for the hyperkinetic disorders: A systematic review and European treatment guideline. *European Child and Adolescent Psychiatry*. [Epub ahead of print]
59. Kratochvil, C.J., Egger, H., Greenhill, L.L., *et al.* (2006). Pharmacological management of preschool ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry,* **45**, 115–18.
60. Faraone, S.V., Spencer, T., Aleardi, M., *et al.* (2004). Meta-analysis of the efficacy of methylphenidate for treating adult attention-deficit/hyperactivity disorder. *Journal of Clinical Psychopharmacology*, **24**, 24–9.
61. Miller, A., Lee, S.K., Raina, P., *et al.* (1998). *A review of therapies for attention-deficit/hyperactivity disorder*. Ottawa, ON: Canadian Coordinating Office for Health Technology Assessment.
62. Barkley, R.A., Edwards, G., Laneri, M., *et al.* (2001). The efficacy of problem-solving communication training alone, behaviour management training alone, and their combination for parent-adolescent conflict in teenagers with ADHD and ODD. *Journal of consulting and clinical psychology,* **69**(6), 926–41.
63. Pelham, W.E., Gnagy, E.M., Greiner, A.R., *et al.* (2000). Behavioral versus behavioral and pharmacological treatment in ADHD children attending a summer treatment program. *Journal of abnormal child psychology*, **28**(6), 507–25.
64. Sonuga-Barke, E.J.S., Daley, D., Thompson, M., *et al.* (2001). Parent-based therapies for preschool attentiondeficit/hyperactivity disorder: A randomized, controlled trial with a community sample. *Journal of the American Academy of Child and Adolescent Psychiatry,* **40**, 402–8.
65. Bor, W., Sanders, M.R., and Markie-Dadds, C. (2002). The effects of the Triple P-Positive Parenting Program on preschool children with co-occurring disruptive behavior and attentional/hyperactive difficulties. *Journal of Abnormal Child Psychology,* **30**, 571–87.
66. Gittelman-Klein, R., and Abikoff, H. (1989) The role of psychostimulants and psychosocial treatments in hyperkinesis. In *Attention def cit Disorder: Clinical and basic research* (eds. T. Sagvolden and T. Archer), Lawrence Erlbaum, Hillsdale NJ, pp. 167–180.
67. Schab, D.W., and Trinh, N.H.T. (2004). Do artificial food colors promote hyperactivity in children with hyperactive syndromes? A meta-analysis of double-blind placebo-controlled trials. *Journal of Developmental and Behavioral Pediatrics,* **25**, 423–34.
68. MTA Cooperative Group. (1999). A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. Multimodal Treatment Study of Children with ADHD. *Archives of General Psychiatry,* **56**, 1073–86.
69. Buitelaar, J.K., Van der Gaag, R.J., Swaab-Barneveld, H., *et al.* (1995). Prediction of clinical response to methylphenidate in children with attention-deficit hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry,* **34**, 1025–32.
70. Research Units on Pediatric Psychopharmacology Autism Network. (2005). Randomized, controlled, crossover trial of methylphenidate in pervasive developmental disorders with hyperactivity. *Archives of General Psychiatry,* **62**, 1266–74.
71. Wilens, T.E., Faraone, S.V., Biederman, J., *et al.* (2003). Does stimulant therapy of attention-deficit/hyperactivity disorder beget later substance abuse? A meta-analytic review of the literature. *Pediatrics,* **111**, 179–85.
72. Rapoport, J.L., Buchsbaum, M.S., Weingartner, H., *et al.* (1980). Dextroamphetamine. Its cognitive and behavioral effects in normal and hyperactive boys and normal men. *Archives of General Psychiatry,* **37**, 933.
73. Ward, M.F., Wender, P.H., and Reimherr, F.W. (1993). The Wender Utah Rating Scale: An aid in the retrospective diagnosis of childhood attention deficit hyperactivity disorder. *American Journal of Psychiatry,* **150**, 885–90. Erratum in, **150**, 1280.
74. Conners, C.K., Erhardt, D., Sparrow, E., *et al.* (1998). In *CAARS Adult ADHD Rating Scales*. New York, NY, Multi Health Systems Inc.
75. Brown, T.E. (1996). In *Brown Attention-Deficit Disorder Scales*, San Antonio, The Psychological Corporation.
76. Young, S., and Bramham, J. (2007). *ADHD in Adults: A Psychological Guide to Practice*. Chichester, UK, John Wiley & Sons.
77. Asherson, P. (2005). Clinical assessment and treatment of attention deficit hyperactivity disorder in adults. *Expert Review of Neurotherapeutics*, **5**, 525–39.

9.2.5 Conduct disorders in childhood and adolescence

Stephen Scott

Introduction

The term conduct disorder refers to a persistent pattern of antisocial behaviour in which the individual repeatedly breaks social rules and carries out aggressive acts which upset other people. It is the commonest psychiatric disorder of childhood across the world, and the commonest reason for referral to child and adolescent mental health services in Western countries. Antisocial behaviour has the highest continuity into adulthood of all measured human traits except intelligence. A high proportion of children and adolescents with conduct disorder grow up to be antisocial adults with impoverished and destructive lifestyles; a significant minority will develop antisocial personality disorder (psychopathy). The disorder in adolescence is becoming more frequent in Western countries and places a large personal and economic burden on individuals and society.

Relation to other disorders

Conduct disorder is one of the two *disruptive disorders* of childhood, (also known as *externalizing disorders*); the other is the hyperkinetic syndrome (ICD 10), a more severe form of attention-deficit hyperactivity disorder (ADHD, DSM IV-R). Conduct disorder and the hyperkinetic syndrome are distinct disorders but often co-occur. As discussed in Chapter 9.1.1 on classification, disruptive disorders can be distinguished on a number of criteria from the other main grouping of child psychiatric conditions, the *emotional disorders* (also known as *internalizing disorders*). For example, unlike emotional disorders, disruptive disorders are commoner in boys, the socially disadvantaged, children from large families, and where there is parental discord.

Juvenile delinquency is a legal term referring to an act by a young person who has been convicted of an offence which would be deemed a crime if committed by an adult. Most, but not all, recurrent juvenile offenders have conduct disorder. In this chapter the term conduct disorder is used as defined by ICD 10 diagnostic criteria; the term conduct problems will be used for less severe antisocial behaviour.

Social problem or medical diagnosis?

Infringement of the rights of other people is a requirement for the diagnosis of conduct disorder. Since the manifestations include a

failure to obey social rules despite apparently intact mental state and social capacities, many have seen the disorder as principally socially determined. They therefore believe the responsibility for its cause and elimination lies with people who can influence the socialization process, such as parents, schoolteachers, social service departments, and politicians. Due to the impossibility of their seeing all cases, there is some debate within child and adolescent psychiatry as to whether doctors and mental health professionals should be involved in any but the most complex presentations.[1] Some have argued that involvement of medical personnel carries the risk of their becoming agents of social control through the misapplication of diagnostic labels, which may lead to abuses of the kind seen in some totalitarian regimes.

However, advances in the last decade have shown there are substantial genetic and biological contributions to conduct disorder, and in some cases the symptoms may be responsive to medication. Work in the last 25 years mainly from the field of child and adolescent mental health has clarified many of the mechanisms contributing to the development and persistence of antisocial behaviour, and has led to the development of effective treatments. As yet these are not being widely used with the children and adolescents who need them. Therefore psychiatrists need to be able to contribute to the planning and delivery of an appropriate service.

Clinical features

Aggressive and defiant behaviour is an important part of normal child and adolescent development which ensures physical and social survival. Indeed, parents may express concern if a child is too acquiescent and unassertive. The level of aggressive and defiant behaviour varies considerably amongst children, and it is probably most usefully seen as a continuously distributed trait. Empirical studies do not suggest a level at which symptoms become qualitatively different, nor is there a single cut-off point at which they become impairing for the child or a clear problem for others. There is no hump towards the end of the distribution curve of severity to suggest a categorically distinct group who might on these grounds warrant a diagnosis of conduct disorder.

Picking a particular level of antisocial behaviour to call conduct disorder is therefore necessarily arbitrary. For all children, the expression of any particular behaviour also varies according to child age, so that for example physical hitting is at a maximum at around 2 years of age but declines to a low level over the next few years. Therefore any judgement about the significance of the level of antisocial behaviour has to be made in the context of the child's age. Before deciding that the behaviour is abnormal or a significant problem, a number of other clinical features have to be considered:

- Level: severity and frequency of antisocial acts, compared with children of the same age and gender
- Pattern: the variety of antisocial acts, and the setting in which they are carried out
- Persistence: duration over time
- Impact: distress and social impairment of child; disruption and damage caused to others.

Change in clinical features with age

The type of behaviour seen will depend on the age and gender of the individual.

Younger children, say from 3 to 7 years of age, usually present with general defiance of adults wishes, disobedience of instructions, angry outbursts with temper tantrums, physical aggression to people especially siblings and peers, destruction of property, arguing, blaming others for things that have gone wrong, and a tendency to annoy and provoke others.

In *middle childhood*, say from 8 to 11, the above features are often present but as the child grows older, stronger, and spends more time out of the home, other behaviours are seen. They include: swearing, lying about what they have been doing, stealing of others belongings outside the home, persistent breaking of rules, physical fights, bullying of other children, cruelty to animals, and setting of fires.

In *adolescence*, say from 12 to 17, more antisocial behaviours are often added: cruelty and hurting of other people, assault, robbery using force, vandalism, breaking and entering houses, stealing from cars, driving and taking away cars without permission, running away from home, truanting from school, extensive use of narcotic drugs.

Not all children who start with the type of behaviours listed in early childhood progress on to the later, more severe forms. Only about half continue from those in early childhood to those in middle childhood[2]; likewise only about a further half of those with the behaviours in middle childhood progress to show the behaviours listed for adolescence. However, the early onset group are important as they are far more likely to display the most severe symptoms in adolescence, and to persist in their antisocial tendencies into adulthood. Indeed over 90 per cent of severe, recurrent adolescent offenders showed marked antisocial behaviour in early childhood. In contrast, there is a large group who only start to be antisocial in adolescence, but whose behaviours are less extreme and who tend to desist by the time they are adults.

Girls

Severe antisocial behaviour is less common in girls who are less likely to be physically aggressive and engage in criminal behaviour, but more likely to show spitefulness, emotional bullying (such as excluding children from groups, spreading rumours so others are rejected by their peers), frequent unprotected sex leading to sexually transmitted diseases and pregnancy, drug abuse, and running away from home.

Pattern and setting

Prognosis is determined by the frequency and intensity of antisocial behaviours, the variety of types, the number of settings in which they occur (e.g. home, school, and in public), and their persistence. For general populations of children, the correlation between parent and teacher ratings on the same measures is only 0.2 to 0.3, so that there are many children who are perceived to be mildly or moderately antisocial at home but well behaved at school, and vice versa. However, for more severe antisocial behaviour, there are usually manifestations both at home and at school.

Impact

At home the child often is subject to high levels of criticism and hostility, and sometimes made a scapegoat for a catalogue of family misfortunes. Frequent punishments and physical abuse are not uncommon. The whole family atmosphere is often soured and siblings also affected. Maternal depression is often present, and families who are unable to cope may, as a last resort, give up the child to be cared for by the local authority. At school, teachers may

take a range of measures to attempt to control the child and protect the other pupils, including sending the child out of the class, sometimes culminating in permanent exclusion from the school. This may lead to reduced opportunity to learn subjects on the curriculum and poor examination results. The child typically has few if any friends, who get fed up with their aggressive behaviour. This often leads to exclusion from many group activities, games, and trips, so restricting the child's quality of life and experiences. On leaving school the lack of social skills, low level of qualifications, and presence of a police record make it harder to gain employment.

Classification

The ICD-10 classification has a category for conduct disorders, F91. The *Clinical descriptions and diagnostic guidelines*[3] state:

> Examples of the behaviours on which the diagnosis is based include the following: excessive levels of fighting or bullying; cruelty to animals or other people; severe destructiveness to property; firesetting; stealing; repeated lying; truancy from school and running away from home; unusually frequent and severe temper tantrums; defiant provocative behaviour; and persistent severe disobedience. Any one of these categories, if marked, is sufficient for the diagnosis, but isolated dissocial acts are not. (p. 267)

An enduring pattern of behaviour should be present, but no time frame is given and there is no impairment or impact criterion stated.

The ICD-10 *Diagnostic criteria for research*[4] differ, requiring symptoms to have been present for at least 6 months, and the introductory rubric indicates that impact upon others (in terms of violation of their basic rights), but not impairment of the child, can contribute to the diagnosis. The research criteria take a menu-driven approach whereby a certain number of symptoms have to be present. 15 behaviours are listed to consider for the diagnosis of **Conduct Disorder**, which usually but not exclusively apply to older children and teenagers. They can be grouped into four classes:

(a) Aggression to people and animals

- often lies or breaks promises to obtain goods or favours or to avoid obligations
- frequently initiates physical fights (this does not include fights with siblings)
- has used a weapon that can cause serious physical harm to others (e.g. bat, brick, broken bottle, knife, gun)
- often stays out after dark despite parenting prohibition (beginning before 13 years of age)
- exhibits physical cruelty to other people (e.g. ties up, cuts, or burns a victim), and
- exhibits physical cruelty to animals.

(b) Destruction of property

- deliberately destroys the property of others (other than by firesetting) and
- deliberately sets fires with a risk or intention of causing serious damage).

(c) Deceitfulness or theft

- steals objects of non-trivial value without confronting the victim, either within the home or outside (e.g. shoplifting, burglary, forgery).

(d) Serious violations of rules

- is frequently truant from school, beginning before 13 years of age
- has run away from parental or parental surrogate home at least twice or has run away once for more than a single night (this does not include leaving to avoid physical or sexual abuse)
- commits a crime involving confrontation with the victim (including purse-snatching, extortion, mugging)
- forces another person into sexual activity
- frequently bullies others (e.g. deliberate infliction of pain or hurt, including persistent intimidation, tormenting, or molestation), and
- breaks into someone else's house, building, or car.

To make a diagnosis, three symptoms from this list have to be present, one for at least 6 months. There is no impairment criterion. There are three subtypes: *conduct disorder confined to the family context* (F91.0), *unsocialized conduct disorder* (F91.1, where the young person has no friends and is rejected by peers), and *socialized conduct disorder* (F91.2, where peer relationships are normal). It is recommended that age of onset be specified, with *childhood onset type* manifesting before age 10, and *adolescent onset type* after. Severity should be categorized as *mild*, *moderate*, or *severe* according to number of symptoms *or* impact on others, e.g. causing severe physical injury, vandalism, theft.

For younger children, say up to 9 or 10 years old, there is a list of eight symptoms for the subtype known as **Oppositional Defiant Disorder** (F91.3):

1 has unusually frequent or severe temper tantrums for his or her developmental level
2 often argues with adults
3 often actively refuses adults' requests or defies rules
4 often, apparently deliberately, does things that annoy other people
5 often blames others for his or her own mistakes or misbehaviour
6 is often touchy or easily annoyed by others
7 is often angry or resentful
8 is often spiteful or resentful.

To make a diagnosis of the oppositional defiant type of conduct disorder, four symptoms from *either* this list *or* the main conduct disorder 15 symptom list have to be present, but no more than two from the latter. Unlike the main variant, there is an impairment criterion: the symptoms must be amaladaptive and inconsistent with the developmental level (p. 161).

Where there are sufficient symptoms of a comorbid disorder to meet diagnostic criteria, the ICD-10 system discourages the application of a second diagnosis, and instead offers a single, combined category. There are two major kinds: mixed disorders of conduct

and emotions, of which **Depressive Conduct Disorder** (F92.0) is the best researched; and **Hyperkinetic Conduct Disorder** (F90.1). There is modest evidence to suggest these combined conditions may differ somewhat from their constituent elements.

The DSM IV-R system[5] follows the ICD-10 research criteria very closely and does not have separate clinical guidelines. The same 15 behaviours are given for the diagnosis of conduct disorder 312.8, with almost identical wording. As for ICD-10, three symptoms need to be present for diagnosis. Severity and childhood or adolescent onset are specified in the same way. However, unlike ICD-10, there is no division into socialized/unsocialized, or family context only types, and there *is* a requirement for the behaviour to cause a clinically significant impairment in social, academic, or social functioning. Comorbidity in DSM IV-R is handled by giving as many separate diagnoses as necessary, rather than by having single, combined categories.

In DSM IV-R, oppositional defiant disorder is classified as a separate disorder on its own, and not as a subtype of conduct disorder. Diagnosis requires four symptoms from a list of eight behaviours which are the same as for ICD-10, but unlike ICD-10, all four have to be from the oppositional list, and none may come from the main conduct disorder list. It is doubtful whether oppositional defiant disorder differs substantially from conduct disorder in older children in any associated characteristics, and the value of designating it as a separate disorder is arguable. In this article, the term conduct disorder will henceforth be used as it is in ICD-10, to refer to all variant including oppositional defiant disorder.

Differential diagnosis

Making a diagnosis of conduct disorder is usually straightforward but comorbid conditions are often missed. The differential diagnosis may include:

1 *Hyperkinetic syndrome/Attention-deficit hyperactivity disorder*. These are the names given by ICD-10 and DSM IV-R respectively for similar conditions, except that the former is more severe. For convenience the term *hyperactivity* will be used here. It is characterized by impulsivity, inattention, and motor overactivity. Any of these three sets of symptoms can be misconstrued as antisocial, particularly impulsivity which is also present in conduct disorder. However, none of the symptoms of conduct disorder are a part of hyperactivity so excluding conduct disorder should not be difficult. A frequently made error however, is to miss comorbid hyperactivity when conduct disorder is definitely present. Standardized questionnaires are very helpful here, such as the Strengths and Difficulties Questionnaire, which is brief, and just as effective at detecting hyperactivity as much longer alternatives.[6]

2 *Adjustment reaction to an external stressor*. This can be diagnosed when onset occurs soon after exposure to an identifiable psychosocial stressor such as divorce, bereavement, trauma, abuse, or adoption. The onset should be within 1 month for ICD-10, and 3 months for DSM IV-R, and symptoms should not persist for more than 6 months after the cessation of the stress or its sequelae.

3 *Mood disorders*. Depression can present with irritability and oppositional symptoms but unlike typical conduct disorder mood is usually clearly low and there are vegetative features; also more severe conduct problems are absent. Early manic depressive disorder can be harder to distinguish, as there is often considerable defiance and irritability combined with disregard for rules, and behaviour which violates the rights of others. Low self-esteem is the norm in conduct disorder, as is a lack of friends or constructive pastimes. Therefore it is easy to overlook more pronounced depressive symptoms. Systematic surveys reveal that around a third of children with conduct disorder have depressive or other emotional symptoms severe enough to warrant a diagnosis.

4 *Autistic spectrum disorders*. These are often accompanied by marked tantrums or destructiveness, which may be the reason for seeking a referral. Enquiring about other symptoms of autistic spectrum disorders should reveal their presence.

5 *Dissocial/antisocial personality disorder*. In ICD-10 it is suggested a person should be 17 or older before dissocial personality is considered. Since at age 18 most diagnoses specific to childhood and adolescence no longer apply, in practice there is seldom difficulty. In DSM IV-R conduct disorder can be diagnosed over 18 so there is potential overlap. A difference in emphasis is the severity and pervasiveness of the symptoms of those with personality disorder, whereby all the individual's relationships are affected by the behaviour pattern, and the individual's beliefs about his antisocial behaviour are characterized by callousness and lack of remorse.

6 *Subcultural deviance*. Some youths are antisocial and commit crimes but are not particularly aggressive or defiant. They are well adjusted within a deviant peer culture that approves of recreational drug use, shoplifting, etc. In some localities a third or more teenage males fit this description and would meet ICD-10 diagnostic guidelines for socialized conduct disorder. Some clinicians are unhappy to label such a large proportion of the population with a psychiatric disorder. Using DSM IV-R criteria would preclude the diagnosis for most youths like this due to the requirement for significant impairment.

Multiaxial assessment

ICD-10 recommends that multiaxial assessment be carried out for children and adolescents, while DSM IV-R suggests it for all ages. In both systems axis one is used for psychiatric disorders which have been discussed above. The last three axes in both systems cover general medical conditions, psychosocial problems, and level of social functioning respectively; these topics will be alluded to below under aetiology. In the middle are two axes in ICD-10, which cover specific (Axis two) and general (Axis three) learning disabilities respectively; and one in DSM IV-R (Axis two) which covers personality disorders *and* general learning disabilities.

Both specific and general learning disabilities are essential to assess in individuals with conduct problems. Fully a third of children with conduct disorder also have specific reading retardation[7] defined as having a reading level two standard deviations below that predicted by the person's IQ. While this may in part be due to lack of adequate schooling, there is good evidence that the cognitive deficits often precede the behavioural problems. General learning disability (mental retardation) is often missed in children with conduct disorder unless IQ testing is carried out. The rate of conduct disorder rise several-fold as IQ gets below 70.

Epidemiology

Between 5 per cent and 10 per cent of children and adolescents have significant persistent oppositional, disruptive, or aggressive behaviour problems.[8,9] With respect to historical period, a modest rise in diagnosable conduct disorder over the second half of the twentieth century has also been observed comparing assessments of three successive birth cohorts in Britain.[10] There is a marked social class gradient.[9] With respect to ethnicity, youth self-reports of antisocial behaviours, and crime victim survey reports of perpetrators' ethnicity show an excess of offenders of black African ancestry. Importantly, Hispanic Americans in the United States of America and British Asians in the United Kingdom do not tend to show an excess of offending compared to their white counterparts.

Sex differences in prevalence

The sex ratio is approximately 2:5 males for each female overall, with males further exceeding females in the frequency and severity of behaviours. On balance, research suggests that the causes of conduct problems are the same for the sexes, but males have more conduct disorder because they experience more of its individual-level risk factors (e.g. hyperactivity, neurodevelopmental delays). However, recent years have seen increasing concern amongst clinicians about treating antisocial behaviour amongst girls.[11]

Developmental subtypes

Life-course persistent versus adolescence-limited

There has been considerable attention paid to the distinction between conduct problems that are first seen in early childhood versus those that start in adolescence[2] and these two subtypes are encoded in the DSM-IV. Early onset is a strong predictor of persistence through childhood, and early onset delinquency is more likely to persist into adult life. Those with early onset differ from those with later onset in that they have lower IQ, more attentional and impulsivity problems, poorer scores on neuropsychological tests, greater peer difficulties and they are more likely to come from adverse family circumstances.[2] Those with later onset become delinquent predominantly as a result of social influences such as association with other delinquent youths. Findings from the follow-up of the Dunedin cohort support relatively poorer adult outcomes for the early onset group in domains of violence, mental health, substance abuse, work, and family life.[2] However the 'adolescence-limited' group were not without adult difficulties. As adults they still engaged in self-reported offending, and they also had problems with alcohol and drugs. Thus, the age-of-onset subtype distinction has strong predictive validity, but adolescent onset antisocial behaviours may have more long-lasting consequences than previously supposed, and so both conduct problems warrant clinical attention.

Aetiology

Individual-level characteristics

(a) Identified genotypes

The search for specific genetic polymorphisms is a very new scientific initiative, and little has yet been accomplished. The most-studied candidate gene in relation to conduct problems is the MAOA promoter polymorphism. The gene encodes the MAOA enzyme, which metabolizes neurotransmitters linked to aggressive behaviour. Replicated studies show that maltreatment history and genotype interact to predict antisocial outcome.[12]

(b) Perinatal complications and temperament

Recent large-scale general population studies have found associations between life-course persistent type conduct problems and perinatal complications, minor physical anomalies, and low birth weight.[13] Most studies support a biosocial model in which obstetric complications might confer vulnerability to other co-ocurring risks such as hostile or inconsistent parenting. Smoking in pregnancy is a statistical risk predictor of offspring conduct problems,[13] but a causal link between smoking and conduct problems has not been established. Several prospective studies have shown associations between irritable temperament and conduct problems.[14]

(c) Neurotransmitters

In general the findings with children have not been consistent.[15] For example, in the Pittsburgh Youth cohort, boys with long-standing conduct problems showed downward changes in urinary adrenaline level following a stressful challenge task, whereas prosocial boys showed upward responses. However other studies have failed to find an association between conduct disorder and measures of noradrenaline in children.[15] It should be borne in mind that neurotransmitters in the brain are only indirectly measured, most measures of neurotransmitter levels are crude indicators of activity, and little is known about neurotransmitters in the juvenile brain.

(d) Verbal deficits and autonomic reactivity

Children with conduct problems have been shown consistently to have increased rates of deficits in language-based verbal skills.[16] The association holds after controlling for potential confounds such as race, socio-economic status, academic attainment, and test motivation. Children who cannot reason or assert themselves verbally may attempt to gain control of social exchanges using aggression; there are likely also to be indirect effects in which low verbal IQ contributes to academic difficulties which in turn mean that the child's experience of school becomes unrewarding, rather than a source of self-esteem and support.

A low resting pulse rate or slow heart rate has been found consistently to be associated with antisocial behaviour, and a meta-analysis of 40 studies suggested it is the best replicated biological correlate of antisocial behaviour.[17] Other psychophysiological indicators show that antisocial and psychopathic boys are also slowest to show a skin-conductance response to aversive stimuli.[17] The explanation for the link between slow autonomic activity and antisocial behaviour remains unclear.

(e) Information-processing and social cognition

Dodge proposed the leading information-processing model for the genesis of aggressive behaviours within social interactions.[18] The model hypothesises that children who are prone to aggression focus on threatening aspects of others' actions, interpret hostile intent in the neutral actions of others, and are more likely to select and to favour aggressive solution to social challenges. Several studies have demonstrated that aggressive children make such errors of social cognition.[18]

Risks outside the family

(a) Risks in the neighbourhood

It has long been assumed that bad neighbourhoods have the effect of encouraging children to develop conduct problems. Many parents

strive to secure the best neighbourhood and school for their child that they can afford. Although it is obvious that some local areas have higher crime rates than others, it has been difficult to document any direct link between neighbourhood characteristics and child behaviour, for a number of reasons. For example, neighbourhood characteristics were conceptualized in overly simple structural-demographic terms such as percentage of non-white residents or percentage of single-parent households. Moreover, research designs could not rule out the alternative possibility that families whose members are antisocial tend to selectively move into bad neighbourhoods. A new generation of neighbourhood research is addressing these challenges, and suggests that the neighbourhood factors that are important include social processes such as 'collective efficacy' and 'social control', do influence young children's conduct problems, probably by supporting parents in their efforts to rear children.

(b) Peer influences

Children with conduct problems have poorer peer relationships than non-disordered children in that they tend to associate with children with similar antisocial behaviours, they have discordant interactions with other children, and experience rejection by non-deviant peers. Three processes have been identified, namely that children's antisocial behaviours lead them to have peer problems, deviant peer relationships lead to antisocial behaviours, and thirdly some common factor leads to both.[19]

Risks within the family

(a) Concentration of crime in families

Fewer than 10 per cent of the families in any community account for more than 50 per cent of that community's criminal offenses, which reflects the coincidence of genetic and environmental risks. There is now solid evidence from twin and adoption studies that conduct problems assessed both dimensionally and categorically are substantially heritable.[20] However, knowing that conduct problems are under some genetic influence is less useful clinically than knowing that this genetic influence appears to be reduced, or enhanced, depending on interaction with circumstances in the child's environment. Several genetically sensitive studies have allowed interactions between family genetic liability and rearing environment to be examined. Both adoption and twin studies have reported an interaction between antisocial behaviour in the biological parent and adverse conditions in the adoptive home that predicted the adopted child's antisocial outcome, so that the genetic risk was modified by the rearing environment.

(b) Family poverty

There is an association between severe poverty and early childhood conduct problems. Early theories proposed direct effects of poverty related to strains arising from the gap between aspirations and realities, and from lacking opportunity to acquire social status and prestige. Subsequent research has indicated that the association between low income and childhood conduct problems is indirect, mediated via family processes such as marital discord and parenting deficits.

(c) Parent–child attachment

Parent–child relationships provide the setting for the development of later social functioning, and disruption of these attachment relationships, for example through institutional care, is associated with subsequent difficulties in relating. Thus, conduct problems might be expected to arise from infant attachment difficulties. One study found that ambivalent and controlling attachment predicted externalizing behaviours after controlling for baseline externalizing problems[21]; disorganized child attachment patterns seem to be especially associated with conduct problems. Although it seems obvious that poor parent–child relations in general predict conduct problems, it has yet to be established whether attachment difficulties as measured by observational paradigms have an independent causal role in the development of behaviour problems; attachment classifications could be markers for other relevant family risks.

(d) Discipline and parenting

Patterns of parenting associated with conduct problems were delineated by Patterson in his seminal work *Coercive Family Process*.[22] Parents of antisocial children were found to be more inconsistent in their use of rules, to issue more, and unclear, commands, to be more likely to respond to their children on the basis of mood rather than the characteristics of the child's behaviour, to be less likely to monitor their children's whereabouts, and to be unresponsive to their children's prosocial behaviour. Patterson proposed a specific mechanism for the promotion of oppositional and aggressive behaviours in children. A parent responds to mild oppositional behaviour by a child with a prohibition to which the child responds by escalating his behaviour, and mutual escalation continues until the parent backs off thus negatively reinforcing the child's behaviour. The parent's inconsistent behaviour increases the likelihood of the child showing further oppositional or aggressive behaviour. In addition to specific tests of Patterson's reinforcement model there is ample evidence that conduct problems are associated with hostile, critical, punitive, and coercive parenting.[23]

In considering the role of coercive processes in the origins or maintenance of conduct problems, we need to consider possible alternative explanations, (i) that the associations reflect familial genetic liability towards children's psychopathology and parents' coercive discipline, (ii) that they represent effects of children's behaviours on parents, and (iii) that coercive parenting may be a correlate of other features of the parent/child relationship or family functioning that influence child behaviours. There is considerable evidence that children's difficult behaviours do indeed evoke parental negativity. The fact that children's behaviours can evoke negative parenting does not however mean that negative parenting has no impact on children's behaviour. The E-risk longitudinal twin study of British families examined the effects of fathers' parenting on young children's aggression.[24] As expected, a prosocial father's *absence* predicted more aggression by his children. But in contrast, an antisocial father's *presence* predicted more aggression by his children, and his harmful effect was exacerbated the more time each week he spent taking care of the children.

(e) Exposure to adult marital conflict and domestic violence

It is likely that family processes other than parenting skills and quality of parent–child attachment relationships have a role. Many studies have shown that children exposed to domestic violence between adults are subsequently more likely to themselves become aggressive. Cummings and Davies[25] proposed that marital conflict influences children's behaviour because of its effect on their regulation of emotion. For example a child may respond to frightening emotion arising from marital conflict by down-regulating his own emotion through denial of the situation. This in turn may

lead to inaccurate appraisal of other social situations and ineffective problem-solving. Repeated exposure to family conflict is thought to lower childrens' thresholds for psychological dysregulation, resulting in greater behavioural reactivity to stress.[25] Children's aggression may also be increased by marital discord because children are likely to imitate aggressive behaviour modelled by their parents. Through parental aggression children may learn that aggression is a normative part of family relationships, that it is an effective way of controlling others, and that aggression is sanctioned, not punished.

(f) Maltreatment

Physical punishment is widely used, and parents of children with conduct problems frequently resort to it out of desperation. Overall, associations between physical abuse and conduct problems are well established.[15] In the Christchurch longitudinal study, child sexual abuse predicted conduct problems, after controlling for other childhood adversities.[26] Links with conduct problems are not however straightforward. The risk for conduct problems does not apply equally to all forms of physical punishment. The E-risk longitudinal twin study was able to compare the effects of corporal punishment (smacking, spanking) versus injurious physical maltreatment using twin-specific reports of both experiences.[27] Results showed that children's genetic endowment accounted for virtually all of the association between their corporal punishment and their conduct problems. This indicated a 'child effect', in which children's bad conduct provokes their parents to use more corporal punishment, rather than the reverse. Findings about injurious physical maltreatment were the opposite. There was no child effect provoking maltreatment and moreover, significant effects of maltreatment on child aggression remained after controlling for any genetic transmission of liability to aggression from antisocial parents.

From risk predictor to causation

Associations have been documented between conduct problems and a wide range of risk factors. A variable is called a 'risk factor' if it has a documented predictive relation with antisocial outcomes, whether or not the association is causal. The causal status of most of these risk factors is unknown; we know what statistically predicts conduct-problem outcomes, but not how or why. Establishing a causal role for a risk factor is by no means straightforward, particularly as it is unethical to experimentally expose healthy children to risk factors to observe whether those factors can generate new conduct problems. There is no one solution to the problem, although the use of genetically sensitive designs and the study of within-individual change in natural experiments and treatment studies have considerable methodological advantages for suggesting causal influences on conduct problems.

Course and prognosis

Of those with early onset conduct disorder (before eight) about half persist with serious problems into adulthood. Of those with adolescent onset, the great majority (over 85 per cent) desist in their antisocial behaviour by their early twenties.

Many of the factors which predict poor outcome are associated with early onset (Table 9.2.5.1).

To detect protective factors, children who do well despite adverse risk factors have been studied.

Table 9.2.5.1 Factors predicting poor outcome

Onset	**Early onset of severe problems, before 8 years of age**
Phenomenology	Antisocial acts which are severe, frequent, and varied
Comorbidity	Hyperactivity and attention problems
Intelligence	Lower IQ
Family history	Parental criminality; parental alcoholism
Parenting	Harsh, inconsistent parenting, with high criticism, low warmth, low involvement, and low supervision
Wider environment	Low-income family in poor neighbourhood with ineffective schools

These so-called 'resilient' children, however, have been shown to have lower levels of risk factors, for example a boy with antisocial behaviour and low IQ living in a rough neighbourhood but living with supportive, concerned parents. Protective factors are mostly the opposite end of the spectrum of the same risk factor, thus good parenting, high IQ are protective. Nonetheless there are factors which are associated with resilience independent of known adverse influences. These include a good relationship with at least one adult, who does not necessarily have to be the parent; a sense of pride and self-esteem; and skills or competencies.

Adult outcome

Studies of groups of children with early onset conduct disorder indicate a wide range of problems not only confined to antisocial acts, as shown in Table 9.2.5.2.

What is clear is that not only are there substantially increased rates of antisocial acts, but that the general psychosocial functioning of children with conduct disorder grown up is strikingly poor. For most of the characteristics shown in Table 9.2.5.2, the increase

Table 9.2.5.2 Adult outcome

Antisocial behaviour	**More violent and non-violent crimes, e.g. mugging, grievous bodily harm, theft, car crimes, fraud**
Psychiatric problems	Increased rates of antisocial personality, alcohol and drug abuse, anxiety, depression and somatic complaints, episodes of deliberate self-harm and completed suicide, time in psychiatric hospitals
Education and training	Poorer examination results, more truancy and early school leaving, fewer vocational qualifications
Work	More unemployment, jobs held for shorter time, jobs low status and income, increased claiming of benefits and welfare
Social network	Few if any significant friends, low involvement with relatives, neighbours, clubs, and organizations
Intimate relationships	Increased rate of short-lived, violent cohabiting relationships; partners often also antisocial
Children	Increased rates of child abuse, conduct problems in offspring, children taken into care
Health	More medical problems, earlier death

compared to controls is at least double for community cases who were never referred, and three to four times for referred children.[18]

Pathways

The path from childhood conduct disorder to poor adult outcome is neither inevitable nor linear. Different sets of influences impinge as the individual grows up and shape the life-course. Many of these can accentuate problems. Thus a toddler with an irritable temperament and short attention span may not learn good social skills if he is raised in a family lacking them, and where he can only get his way by behaving antisocially and grasping for what he needs. At school he may fall in with a deviant crowd of peers, where violence and other antisocial acts are talked up and give him a sense of esteem. His generally poor academic ability and difficult behaviour in class may lead him to truant increasingly, which in turn makes him fall further behind. He may then leave school with no qualifications and fail to find a job, and resort to drugs. To fund his drug habit he may resort to crime, and once convicted, find it even harder to get a job. From this example, it can be seen that adverse experiences do not only arise passively and independently of the young person's behaviour; rather, the behaviour predisposes them to end up in risky and damaging environments. Consequently, the number of adverse life events experienced is greatly increased.[28] The path from early hyperactivity into later conduct disorder is also not inevitable. In the presence of a warm supportive family atmosphere it is far less likely than if the parents are highly critical and hostile.

Other influences can however steer the individual away from and antisocial path. For example, the fascinating follow-up of delinquent boys to age 70 by Laub and Sampson[29] showed that the following led to desistence: being separated from a deviant peer group; marrying to a non-deviant partner; moving away from a poor neighbourhood; military service which imparted skills.

Treatment

Evidence-based treatments

Proven treatments include those which singly or in combination address (i) Parenting skills, (ii) Family functioning, (iii) Child interpersonal skills, (iv) Difficulties at school, (v) Peer group influences, and (vi) Medication for coexistent hyperactivity.

(a) Parenting skills

Parent management training aims to improve parenting skills. There are scores of randomized controlled trials showing that it is effective for children up to about 10 years old.[30] They address the parenting practices identified in research as contributing to conduct problems. A more detailed account is given by Scott.[30] Typically, they include five elements:

(i) Promoting play and a positive relationship

In order to cut into the cycle of defiant behaviour and recriminations, it is important to instil some positive experiences for both sides and begin to mend the relationship. Teaching parents the techniques of how to play in a constructive and non-hostile way with their children helps them recognize their needs and respond sensitively. The children in turn begin to like and respect their parents more, and become more secure in the relationship.

(ii) Praise and rewards for sociable behaviour

Parents are helped to reformulate difficult behaviour in terms of the positive behaviour they wish to see, so that they encourage wanted behaviour rather than criticize unwanted behaviour. For example, instead of shouting at the child not to run, they would praise him whenever he walks quietly; then he will do it more often. Through hundreds of such prosaic daily interactions, child behaviour can be substantially modified. Yet some parents find it hard to praise, and fail to recognize positive behaviour when it happens, with the result that it become less frequent.

(iii) Clear rules and clear commands

Rules need to be explicit and constant; commands need to be firm and brief. Thus shouting at a child to stop being naughty doesn't tell him what he *should* do, whereas for example telling him to play quietly gives a clear instruction which makes compliance easier.

(iv) Consistent and calm consequences for unwanted behaviour

Disobedience and aggression need to be responded to firmly and calmly, but for example putting the child in a room for a few minutes. This method of timeout from positive reinforcement sounds simple but requires considerable skill to administer effectively. More minor annoying behaviours such as whining and shouting often respond to being ignored, but again parents often find this hard to achieve in practice.

(v) Reorganizing the child's day to prevent trouble

There are often trouble spots in the day which will respond to fairly simple measures. For example, putting siblings in different rooms to prevent fights on getting home from school; banning TV in the morning until the child is dressed; and so on.

Treatment can be given individually to the parent and child which enables live feedback in light of the parent's progress and the child's response. Alternatively, group treatments with parents alone have been shown to be equally effective.[31] Trials show that parent management training is effective in reducing child antisocial behaviour the short-term, with moderate to large effect sizes of 0.5 to 0.8 standard deviations, and there is little loss of effect at 1 or 3 year follow-up.[32]

(b) Family functioning

Functional Family Therapy, Multisystemic Therapy, and *Treatment Foster Care* aim to change a range of difficulties which impede effective functioning of teenagers with conduct disorder. Functional family therapy addresses family processes which need to be present such as improved communication between parent and young person, reducing interparental inconsistency, tightening up on supervision and monitoring, and negotiating rules and the sanctions to be applied for breaking them. Functional family therapy has been shown to reduce reoffending rates by around 50 per cent.[33] Other varieties of family therapy have not been subjected to controlled trials for young people with conduct disorder or delinquency, so cannot be evaluated for their efficacy.

In multisystemic therapy,[34] the young person's and family's needs are assessed in their own context at home and in their relations with other systems such as school and peers. Following the assessment, proven methods of intervention are used to address difficulties and promote strengths. Multisystemic therapy differs from most types of family therapy such as the Milan or systemic approach as usually practised in a number of regards. Firstly, treatment is delivered in the situation where the young lives, e.g. at

home. Secondly, the therapist has a low caseload (4–6 families) and the team is available 24 h a day. Thirdly, the therapist is responsible for ensuring appointments are kept and for making change happen—families cannot be blamed for failing to attend or 'not being ready' to change. Fourthly, regular written feedback on progress towards goals from multiple sources is gathered by the therapist and acted upon. Fifthly, there is a manual for the therapeutic approach and adherence is checked weekly by the supervisor. Several randomized controlled trial attest to the effectiveness, with reoffending rates typically cut by half and time spent in psychiatric hospitalization reduced further.[34]

Treatment foster care is another way to improve the quality of encouragement and supervision that teenagers with conduct disorder receive. The young person lives with a foster family specially trained in effective techniques; sometimes it is ordered as an alternative to jail. Outcome studies show useful reductions in reoffending.[35]

(c) Anger management and child interpersonal skills

Most of the programmes to improve child interpersonal skills derive from cognitive behaviour therapy. A typical example is the *Coping Power* Programme.[36] This and other programmes have in common, in training the young person to:

i) slow down impulsive responses to challenging situations by stopping and thinking,

ii) recognize their own level of physiological arousal, and their own emotional state,

iii) recognize and define problems,

iv) develop several alternative responses,

v) choose the best alternative based on anticipation of consequences,

vi) reinforce himself for use of this approach.

Over the longer-term they aim to increase positive social behaviour by teaching the young person to:

i) learn skills to make and sustain friendships,

ii) develop social interaction skills such as turn-taking and sharing,

iii) express viewpoints in appropriate ways and listen to others.

Typically, given alone, treatment gains with interpersonal skills training are good within the treatment setting, but only generalize slightly to 'real-life' situations such as the school playground. However, when they are part of a more comprehensive programme which has those outside the young person reinforcing the approach, they add to outcome gains.[36]

(d) Difficulties at school

These can be divided into learning problems and disruptive behaviour. There are proven programmes to deal with specific learning problems such as specific reading retardation, such as. reading recovery. However, few of the programmes have been specifically evaluated for their ability to improve outcome in children with conduct disorder, although trials are in progress. Preschool education programmes for high risk populations have been shown to reduce arrest rates and improve employment in adulthood (see below).

There are several schemes for improving classroom behaviour, which vary from those which stress improved communication such as 'circle time', and those which work on behavioural principles or are part of a multimodal package. Many of these schemes have been shown to improve classroom behaviour, and some specifically target children with conduct disorder.[37]

(e) Peer group influences

A few interventions have aimed to reduce the bad influence of deviant peers. However, a number attempted this through group work with other conduct disordered youths, but outcome studies showed a *worsening* of antisocial behaviour. Current treatments therefore either see youths individually try to steer them away from deviant peers, or work in small groups (say 3–5 youths) where the therapist can control the content of sessions. Some interventions place youths with conduct disorder in groups with well-functioning youths, and this has led to favourable outcomes.[38]

(f) Medication for coexistent hyperactivity

Where there is comorbid hyperactivity in addition to conduct disorder, several studies attest to a large (effect size of 0.8 standard deviations or greater) reduction in both overt and covert antisocial behaviour,[39] both at home and at school. However, the impact on long-term outcome is unstudied.

Management

Engagement of the family is particularly important for this group of children and families as dropout from treatment is high, at around 30–40 per cent. Practical measures such as assisting with transport, providing childcare, holding sessions in the evening, or at other times to suit the family will all help. Many of the parents of children with conduct disorder may themselves have difficulty with authority and officialdom and be very sensitive to criticism. Therefore the approach is more likely to succeed if it is respectful of their point of view, does not offer overly prescriptive solutions, and does not directly criticize parenting style. Practical homework tasks increase changes, as do problem-solving telephone calls from the therapist between sessions.

Parenting interventions may need to go beyond skill development to address more distal factors which prevent change. For example, drug or alcohol abuse in either parent, maternal depression, and a violent relationship with the partner are all common. Assistance in claiming welfare and benefits and help with financial planning may reduce stress from debts.

A multimodal approach is likely to get larger changes. Therefore involving the school in treatment by visiting and offering strategies for managing the child in class is usually helpful, as is advocating for extra tuition where necessary. If the school seems unable to cope despite extra resources, consideration should be given to moving the child to a different school which specializes in the management of behavioural difficulties. Avoiding antisocial peers and building self-esteem may be helped by getting the child to attend after school clubs and holiday activities.

Where parents are not coping or a damaging abusive relationship is detected, it may be necessary to liaise with the social services department to arrange respite for the parents or a spell of foster care. It is important during this time to work with the family to increase their skills so the child can return to the family. Where there is permanent breakdown, long-term fostering, or adoption may be recommended.

Opportunities for prevention

Conduct disorder should offer good opportunities for prevention since:

1 it can be detected early reasonably well,

2 early intervention is more effective than later,

3 there are a number of effective interventions.

In the United States of America, a number of comprehensive interventions based on up to date empirical findings are being carried out. Perhaps the best known is Families and Schools Together.[40] Here the most antisocial 10 per cent of 5–6 year olds in schools in disadvantaged areas were selected, as judged by teacher and parent reports. They were then offered intervention which was given for a whole year in the first instance and comprised:

i) weekly parent training in groups with videotapes

ii) an interpersonal skills training programme for the whole class

iii) academic tutoring twice a week

iv) home visits from the parent trainer

v) a pairing programme with sociable peers from the class.

Almost 1000 children were randomized to receive this condition or controls, and the project has cost over $50 million. However, so far, preliminary reports of outcome have been limited with no improvement of antisocial behaviour at home on questionnaire measures and modest improvements in the classroom. There are a number of possible reasons for the smaller effects compared to those obtained in trials with clinically referred populations. The motivation of families may be less as they don't perceive they have a problem; starting levels of antisocial behaviour are lower, so there is not so far to go to reach normal levels; and keeping up the quality of the intervention across several sites is harder. It remains to be seen whether longer-term effects will be greater.

In the United States of America, preschool education programmes for disadvantaged children have shown good outcomes in small demonstration projects, but replication on a larger scale has generally proved rather disappointing. In the United Kingdom, the government stressed the importance of helping parents of children in the first 3 years of life and put substantial resources (£540 million) into *SureStart* centres in specifically targeted high risk neighbourhoods to support parenting. Early evaluation of outcome showed no change on 24 of 25 variables; maternal acceptance of the child was the only measured outcome to change, child antisocial behaviour did not.[41] Separate from conduct disorder prevention but related is crime prevention, which can include reducing the opportunities for antisocial behaviour by tighter policing, reducing access to drugs and guns, and so on.

Conclusion

Much is known about the risk factors leading to conduct disorder and effective treatments exist. The challenge is to make these available on a wide scale, and to develop approaches to prevention which are effective and can be put into practice at a community level.

Further information

To access the US Federal Government's site National Youth Violence Prevention Resource Center which has recent research findings, visit http://www.safeyouth.org/scripts/index.asp

To access the US Surgeon Generals' thorough report on youth violence, visit http://www.surgeongeneral.gov/library/youthviolence/

Bloomquist, M.L. and Schnell, S.V. (2002). *Helping children with aggression and conduct problems: best practices for intervention*. Guilford Press, New York.

Connor, D.F. (2002). *Aggression and antisocial behaviour in children and adolescents: research and treatment*. Guilford Press, New York.

Essau, C. (2003). *Conduct and oppositional defiant disorders: epidemiology, risk factors and treatment*. Lawrence Erlbaum Associates, Mahwah, New Jersey.

References

1. Vostanis, P., Meltzer, H., Goodman, R., *et al.* (2003). Service utilisation by children with conduct disorders: findings from the GB national study. *European Child & Adolescent Psychiatry,* **12**, 231–8.
2. Moffitt, T.E. (2006). Life-course-persistent versus adolescence-limited antisocial behaviour. In *Developmental psychopathology: risk, disorder, and adaptation,* Vol. 3 (2nd edn) (eds. D. Cicchetti and D.J. Cohen), pp. 570–98. John Wiley, Hoboken, NJ.
3. World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. World Health Organization, Geneva.
4 World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders—diagnostic criteria for research*. World Health Organization, Geneva.
5. American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders—DSM IV-R* (4th edn), text revision. American Psychiatric Association, Washington, DC.
6. Goodman, R. and Scott, S. (1999). Comparing the strengths and difficulties questionnaire and the child behaviour checklist: is small beautiful? *Journal of Abnormal Child Psychology*, **27**, 17–24.
7. Trzesniewski, K., Moffitt, T., Caspi, A., *et al.* (2006). Revisiting the association between reading achievement and antisocial behaviour: new evidence of an environmental explanation from a twin study. *Child Development,* **77**, 72–88.
8. Angold, A. and Costello, E.J. (2001). The epidemiology of disorders of conduct: nosological issues and comorbidity. In *Conduct disorders in childhood and adolescence* (eds. J. Hill and B. Maughan). Cambridge University Press, Cambridge.
9. Green, H., McGinnity, A., Meltzer, H., *et al.* (2005). *Mental health of children and young people in Great Britain*. TSO, London.
10. Collishaw, S., Maughan, B., Goodman, R., *et al.* (2004). Time trends in adolescent mental health. *Journal of Child Psychology and Psychiatry*, **45**, 1350–62.
11. Pullatz, M. and Bierman, K.L. (2004). *Aggression, antisocial behaviour, and violence among girls: a developmental perspective*. Guilford Press, New York.
12. Kim-Cohen, J., Caspi, A., Taylor, A., *et al.* (*2006*)**.** MAOA, maltreatment, and gene-environment interaction predicting children's mental health: new evidence and a meta-analysis. *Molecular Psychiatry,* **11**, 903–913.
13. Brennan, P.A., Grekin, E.R., and Mednick, S.A. (2003). Prenatal and perinatal influences on conduct disorder and serious delinquency. In *Causes of conduct disorder and delinquency* (eds. B. Lahey, T.E. Moffitt, and A.Caspi), pp. 319–44. Guilford Press, New York.
14. Keenan, K. and Shaw, D.S. (2003). Starting at the beginning: exploring the etiology of antisocial behaviour in the first years of life. In *Causes of conduct disorder and delinquency* (eds. B. Lahey, T.E. Moffitt, and A.Caspi), pp. 153–81. Guilford Press, New York.
15. Hill, J. (2002). Biological, psychological and social processes in the conduct disorders. *Journal of Child Psychology and Psychiatry*, **43**, 133–64.
16. Lynam, D.R. and Henry, W. (2001). The role of neuropsychological deficits in conduct disorders. In *Conduct disorders in childhood and adolescence* (eds. J. Hill and B. Maughan). Cambridge University Press, Cambridge.

17. Ortiz, J. and Raine, A. (2004). Heart rate level and antisocial behaviour in children and adolescents: a meta-analysis. *Journal of the American Academy of Child and Adolescent Psychiatry*, **43**, 154–62.
18. Dodge, K. (2006). Translational science in action: hostile attributional style and the development of aggressive behaviour problems. *Development and Psychopathology,* **18**, 791–814.
19. Coie, J.D. (2004). The impact of negative social experiences on the development of antisocial behaviour. In *Children's peer relations: from development to intervention* (eds. J.B. Kupersmidt and K.A. Dodge), pp. 243–67. American Psychological Association, Washington, DC.
20. Moffitt, T.E. (2005). Genetic and environmental influences on antisocial behaviours: evidence from behavioural-genetic research. *Advances in Genetics*, **55**, 41–104.
21. Moss, E., Smolla, N., Cyr, C., *et al.* (2006). Attachment and behaviour problems in middle childhood as reported by adult and child informants. *Development and Psychopathology*, **18**, 425–44.
22. Patterson, G.R. (1982). *Coercive family process*. Castalia Publishing Company, Eugene, OR.
23. Rutter, M., Giller, H., and Hagell, A. (1998). *Antisocial behaviour by young people*. Cambridge University Press, Cambridge.
24. Jaffee, S.R., Moffitt, T.E., Caspi, A., *et al.* (2003). Life with (or without) father: the benefits of living with two biological parents depend on the father's antisocial behavior. *Child Development*, **74**, 109–26.
25. Cummings, E.M. and Davies, P. (2002). Effects of marital conflict on children: recent advances and emerging themes in process-oriented research. *Journal of Child Psychology and Psychiatry*, **43**, 31–64.
26. Fergusson, D.M., Horwood, L.J., and Lynskey, M.T. (1996). Childhood sexual abuse and psychiatric disorder in young adulthood. II. Psychiatric outcomes of childhood sexual abuse. *Journal of the American Academy of Child and Adolescent Psychiatry*, **35**, 1365–74.
27. Jaffee, S.R., Caspi, A., Moffitt, T.E., *et al.* (2004). The limits of child effects: evidence for genetically mediated child effects on corporal punishment, but not on physical maltreatment. *Developmental Psychology*, **40**, 1047–58.
28. Champion, L., Goodall, G., and Rutter, M. (1995). Behavioural problems in childhood and stressors in early adult life: a 20 year follow-up of London school children. *Psychological Medicine*, **25**, 231–46.
29. Laub, J. and Sampson, R. (2003). *Shared beginnings, divergent lives: delinquent boys to age 70*. Harvard University Press, Cambridge, MA.
30. Scott, S. (2008). Parenting programs. Chapter in *Rutter's child and adolescent psychiatry* (5th edn) (eds. M. Rutter, D. Bishop, D. Pine, *et al.*). Blackwell, Oxford.
31. Scott, S., Spender, Q., Doolan, M., *et al.* (2001). Multicentre controlled trail of parenting groups for child antisocial behaviour in clinical practice. *British Medical Journal*, **323**, 194–7.
32. Scott, S. (2005). Do parenting programmes for severe child antisocial behaviour work over the longer term, and for whom? 1 year follow up of a multi-centre controlled trial. *Behavioural and Cognitive Psychotherapy*, **33**, 1–19.
33. Alexander, J.F., Holtzworth-Munroe, A., and Jameson, P.B. (1994). The process and outcome of marital and family therapy research: review and evaluation. In *Handbook of psychotherapy and behaviour change* (eds. A. E. Bergin and S. Garfield), pp. 595–630. Wiley, New York.
34. Curtis, N.M., Ronan, K.R., and Borduin, C.M. (2004). Multisystemic therapy: a meta-analysis of outcome studies. *Journal of Family Psychology*, **18**, 411–19.
35. Eddy, J.M., Whaley, R.B., and Chamberlain, P. (2004). The prevention of violent behavior by chronic and serious male juvenile offenders: a 2-year follow-up of a randomized clinical trial. *Journal of Emotional and Behavioral Disorders,* **12**, 2–8.
36. Lochman, J.E. and Wells, K.C. (2004). The coping power program for preadolescent aggressive boys and their parents: outcome effects at the 1-year follow-up. *Journal of Consulting and Clinical Psychology,* **72**, 571–8.
37. Durlak, J. (1995). *School-based prevention programs for children and adolescents*. Sage, Thousand Oaks, CA.
38. Feldman, R. (1992). *The St Louis experiment: effective treatment of antisocial youths in prosocial peer groups. Preventing antisocial behaviour* (eds. J. McCord and R. Tremblay), pp. 233–52. Guilford Press, New York.
39. Connor, D.F., Glatt, S.J., Lopez, I.D., *et al.* (2002). Psychopharmacology and aggression. I. A meta-analysis of stimulant effects on overt/covert aggression-related behaviors in ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry*, **41**, 253–61.
40. Conduct Problems Prevention Research Group. (1992). A developmental and clinical model for prevention of conduct disorder. *Development and Psychopathology*, **4**, 509–27.
41. Belsky, J., Melhuish, E., Barnes, J., Leyland, A., and Romanuik, H. (2006). Effects of sure start local programmes on children and families: early findings from a quasi-experimental, cross sectional study. *British Medical Journal*, **332**, 1476–9.

9.2.6 Anxiety disorders in childhood and adolescence

Daniel S. Pine

Introduction

The term 'fear' refers to the brain state evoked by dangerous stimuli that are avoided because they are capable of harming the organism. The term 'anxiety', in contrast, refers to the brain state evoked by 'threats', stimuli that signal the *possibility* of danger at some point in the near future. Fear and anxiety represent adaptive responses to overt dangers and threats, in that these responses typically reduce the potential for harm to the organisms. Anxiety *disorders* represent conditions where the level of fear is maladaptive either because it leads to clinically significant distress or impairment in function. These effects can result from the production of an anxiety response in a situation not perceived as dangerous by healthy people or by the production of an extreme anxiety response in a situation that healthy people would find mildly anxiety provoking.

The current chapter summarizes recent research on paediatric anxiety disorders. A focus on developmental aspects of anxiety is important since most clinically impairing forms of anxiety typically begin during childhood.[1] Moreover, childhood anxiety disorders show associations with a range of adult psychopathologies beyond anxiety, including most prominently various mood disorders. This fact has stimulated considerable debate concerning the degree to which childhood anxiety disorders reflect early manifestations of adult anxiety disorders. Separation anxiety disorder (SAD) represents the only specific anxiety disorder that primarily occurs in children and adolescents but not adults. Two other disorders frequently co-occur with SAD, social phobia (SOPH), and generalized anxiety disorder (GAD). The current chapter focuses specifically on these three conditions. The chapter also reviews in somewhat less detail data for specific phobia (SPH), a typically minimally impairing condition, and panic disorder (PD), a condition that occurs primarily in adults.[1] Other chapters review material for conditions that frequently co-occur with these five anxiety disorders. This includes major depression (see Chapter 9.2.7), obsessive–compulsive disorder (Chapter 9.2.8), and trauma-related disorders (Chapters 9.3.2). Material on SAD, SOPH, GAD, SPH,

and PD are reviewed in three sections. The first, most detailed, section reviews clinical features of these disorders, including typical presentations and diagnosis. The second somewhat briefer section reviews pathophysiology, and the final section briefly reviews therapeutics.

Clinical features

Clinical presentation

Children presenting with symptoms of anxiety typically manifest signs of various disorders. In fact, in the clinical setting, presentation with a 'pure' form of anxiety is relatively rare. This suggests that current classifications group children into categories that are unlikely to represent distinct pathophysiologies. Nevertheless, while the current nosology is likely to change as understandings of pathophysiology advance, current classification schemes remain quite useful in that they facilitate communication among individuals working with a child and allow clinicians to draw on research in therapeutics using a common diagnostic system. The current section describes clinical presentation of five specific anxiety disorders, as defined in the fourth edition of the *Diagnostic and Statistical Manual* (DSM-IV). These definitions are similar to those used in the 10th edition of the *International Classification of Disease* (ICD-10), although ICD-10 provides single diagnosis for children with multiple anxiety disorders, whereas and DSM-IV provides multiple diagnoses.

(a) Separation anxiety disorder

The key feature of SAD involves presentation of anxiety related to fear that harm will befall an attachment figure. In severe forms, SAD typically presents with avoidance of situations, such as school, where separation is required. The term 'school phobia' had been used on occasion for these presentations, but current approaches no longer use this term. Symptoms of SAD often are severe at night, leading many children to refuse to sleep alone or at friends' homes. Considerable research examines the relationship between childhood SAD and adult panic disorder.

(b) Social phobia

The key features of SOPH involve intense fear or anxiety in situations where the individual is scrutinized. This presents either as extreme form of pervasive shyness or as extreme fear in particular social situations, such as during class presentations. SOPH can also be classified as a 'generalized subtype', indicating that most social situations are feared. The condition can markedly interfere with function by leading children to avoid important academic exercises that must be performed in social settings or by markedly impacting on social development. This effect on social relationships has led to some controversy concerning the boundaries between SOPH and pervasive developmental disorders (PDD). Classically, this distinction can be made based on the presence of language dysfunction and stereotypic behaviour in PDD but not SOPH. Considerable research examines the relationship between late-childhood SOPH and early-childhood temperament.

(c) Generalized anxiety disorder

The key feature of GAD involves a pervasive sense of worry about various events or circumstances. For example, children with GAD frequently worry about their competence, as might manifest on school or athletic performances. These worries are associated with other symptoms, such as muscle tension or other somatic complaints, irritability, and trouble sleeping. Because children with SAD and SOPH also present with worry, clinicians face difficulties when attempting to determine if worries reflect aspects of these disorders or another problem. GAD is diagnosed only when worries cannot be accounted for by another diagnosis. Considerable research examines the relationship between GAD and major depressive disorder (MDD).

(d) Specific phobia

The key feature of SPH involves fear of a specific stimulus or object. SPH can manifest to a range of objects, such as potentially dangerous animals or natural scenarios, and SPH can be categorized into one of five types, based on the content of the fear. Children rarely present for clinical care when they suffer from SPH in the absence of another anxiety disorder, despite the fact that SPH does present relatively commonly in pure forms in the community. This suggests that SPH typically is associated with relatively mild degrees of distress and impairment, unless SPH is associated with another anxiety disorder.

(e) Panic disorder

The key feature of PD involves spontaneous panic attacks. The term 'panic attack' refers to crescendo paroxysms of severe anxiety that occur suddenly and are associated with somatic and cognitive sensations, such as rapid heart beat, shortness of breath, and a strong desire to flee. Panic attacks occur in many situations and with various clinical syndromes. The key feature of PD is that at least some of these attacks occur in the absence of any cue or trigger. As such, the patient cannot attribute the attack to fear of any specific circumstance. PD virtually never occurs prior to puberty, and the disorder is also very rare before adulthood.

Assessment

The assessment for anxiety involves input from multiple sources. Clearly obtaining information directly from the patient is vital. Children with anxiety disorders may be reluctant to report the precise nature of their fears. As a result, adults may be unaware of vital symptoms. On the other hand, children also often show reluctance to acknowledge their anxiety, either because they are unaware of their degree of incapacitation or because they are highly embarrassed about their symptoms. In this instance, adults provide vital information concerning specific objects or situations that might be feared by children or adolescents.

Various forms of standardized assessment are available for paediatric anxiety.[2] This includes rating scales that can be directly completed by parents, teachers, or children, as well as scales that are completed by clinicians based on their interview of the child and parent. Moreover, standardized observational batteries typically are used for the assessment of temperament, in very young children, that relate to anxiety disorders in older children. Temperament also can be measured by parent or self-report.[3] In general, while high scores on various rating scales does provide some indication regarding the presence of an anxiety disorder, structured psychiatric interviews, completed by a trained clinician, represents the gold standard for arriving at a diagnosis.

Prevalence and demographics

As a group, paediatric anxiety disorders probably represent the most common form of developmental psychopathology. It is

difficult to provide precise data concerning their overall prevalence, as the rate of anxiety disorders is highly variable across studies, most likely due to variations in assessment. Rates of anxiety disorders are unusually sensitive to even subtle changes in assessments of impairment.[4] In general, overall lifetime rates of paediatric anxiety probably fall in the 10–20 per cent range.[5] Rates of individual disorders vary with age. Thus, SAD represents the most common condition, with prevalence typically in the 5 per cent range, before puberty, whereas GAD and SOPH become more prevalent during adolescence, again with rates in the 5 per cent range. Rates of SPH are highly variable, depending on the stringency of impairment criteria, with some estimates surpassing 20 per cent. As noted above, PD is very rare before late adolescence.

In terms of demography, anxiety disorders show a strong female predominance. This gender difference manifests for all of the conditions examined here, and, unlike data for MDD, it emerges before puberty. While the overall rate of anxiety disorders changes relatively little from childhood to adolescence, the nature of disorders does change. Thus, SAD is most common in young children, whereas SOPH is most common in adolescence. Data concerning associations with social class appear somewhat mixed. While some inconsistent reports note higher rates among individuals in the relatively lower social strata, the data appear most consistent for SPH, with weaker or absent associations in other conditions.[1] Consistent with weak relationships, recent work suggests that abrupt changes in family economics do not lead to changes in rates of anxiety disorders, despite strong associations with changing rates of other disorders.[6]

Comorbidity

Data concerning comorbidity reveal distinct trends in the clinic relative to the community, most likely due to the effects of referral biases on data from the clinic. Thus, in the clinic, paediatric anxiety disorders have been linked to virtually every form of psychopathology. This includes mood disorders, behaviour disorders, attention deficit hyperactivity disorder, and substance use disorders. In the community, however, associations appear particularly strong with a more restricted group of conditions. The most common comorbidity represents associations with other anxiety disorders, with odds ratios typically appearing in the three-to-five range.[1] Associations between SOPH and GAD appear particularly strong in this work. Comorbidity with mood disorder, particularly MDD, is only slightly weaker than comorbidity among the anxiety disorders.[7] Other forms of psychopathology show far weaker associations.

Clinical course

Paediatric anxiety disorders predict an increased risk for a range of adverse psychiatric outcomes in adults. This includes most prominently risk for adult anxiety disorders and MDD.[1,8] In general, children and adolescent with one or another anxiety disorder face a two- to five-fold increased risk for adult anxiety or MDD. These relationships reflect the fact that most adults with various forms of mood or anxiety disorder show the initial signs of their problem during childhood or adolescence, manifest as a paediatric anxiety disorder. However, the overall magnitude of these longitudinal relationships between paediatric anxiety and any form of adult psychopathology appears somewhat weaker than longitudinal relationships for other developmental psychopathologies, such as the behaviour disorders.[9]

Relatively few studies consider the long-term outcome of the specific paediatric anxiety disorders. In the few studies that do examine this issue, the overall weight of the evidence suggests that risk for poor outcome is similar among all paediatric anxiety disorders.[5] However, some inconsistent data do note specific associations among individual child and adult disorders. For example, some evidence documents a particularly strong association between paediatric GAD and adult MDD,[1] though studies following adolescents into their 30s suggest a comparable risk for MDD in adolescent SOPH.[8] Similarly, some studies note an association between childhood SAD and adult PD, but the overall weight of the evidence does not provide strong support for this link.[10] Finally, some inconsistent evidence also suggests that the outcome of paediatric SPH appears relatively good, as compared with other anxiety disorders.

Pathophysiology

Neuroscience and fear

Work on the pathophysiology of anxiety disorders benefits from a wealth of research examining brain regions involved in fear and anxiety among rodents and non-human primates.[11] Data also document strong cross-species parallels in the effects of threats and danger on behaviour, physiology, and information processing. This suggests that neural circuits implicated in fear and anxiety among rodents or non-human primates are likely involved in fear and anxiety among humans.

Basic science work delineates a distributed neural circuit engaged by various forms of dangerous or threatening stimuli. This includes stimuli recognized as dangerous through learning, as classically studied in the 'fear conditioning' paradigm, whereby a neutral conditioned stimulus is paired with an aversive unconditioned stimulus.[12] This also includes stimuli innately recognized as dangerous, even in the absence of prior training.[13] Finally, work in immature rodents and non-human primates demonstrates strong developmental influences on the neural circuitry associated with both learned and innate fears.[11] Specifically, in immature relative to mature organisms, both genetic and environmental manipulations show the capacity to exert more robust, long-standing effects on anxiety and fear-related behaviours as well as the underlying neural circuitry mediating these behaviours.

Figure 9.2.6.1 illustrates the core components of the underlying circuitry associated with fear and anxiety. The amygdala represents a hub in the fear circuit. As shown in Fig. 9.2.6.1, this collection of nuclei lies within the medial temporal lobe, where it receives input from various sensory cortices as well as brain-stem monoamine systems, and where it sends output to the hypothalamus and other structures that orchestrate the organism's response to danger. While some debate continues concerning the precise role played by the amygdala in fear, the structure has been implicated in both learned and innate fears as well as various positive-valence emotions. Some data suggest that the amygdala plays a vital role in regulating attention when organisms learn to associate neutral stimuli with salient events.

Fear and anxiety represent complex states that reflect influences from other brain regions beyond the amygdala. Figure 9.2.6.1 shows the location of two particularly important regions. Thus, the hippocampus also plays a role in fear and anxiety, with data most clearly implicating this brain region in the representation of spatial

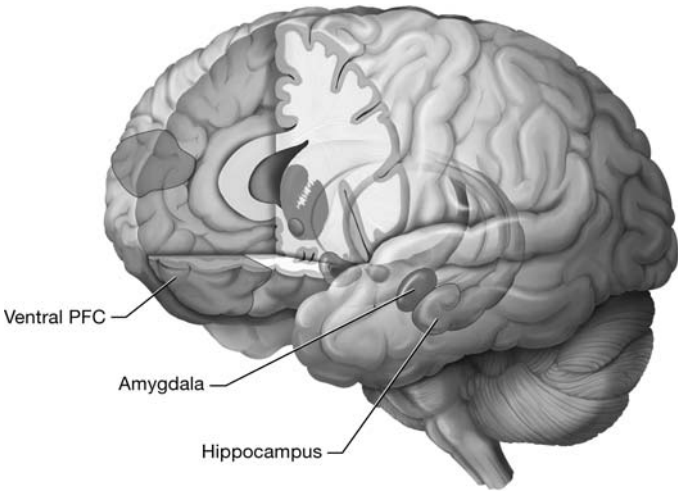


Fig. 9.2.6.1 Displays key anatomical components of brain circuitry engaged when various organisms encounter threats or dangers. Functional aspects of this circuitry show strong cross-species conservation, and Figure 9.2.6.1 depicts the location in the human brain of three particularly important neural structures: the ventral prefrontal cortex (PFC), amygdala, and hippocampus.

contexts associated with threat. As shown in Fig. 9.2.6.1, the hippocampus lies posterior to the amygdala in the medial temporal lobe. Various components of the prefrontal cortex (PFC) also are involved in fear and anxiety. As shown in Fig. 9.2.6.1, particularly strong associations occur with ventral PFC, including both lateral and medial expanses. PFC is thought to play a regulatory role for fear and anxiety responses, serving to delineate the temporal context where fear and anxiety are either necessary or not appropriate, given the organism's goals.

Familial aggregation and genetics

Most complex behaviours, including fear and anxiety, represent phenomena that result from influences of both genes and the environment. The associations of genes and the environment with anxiety can be examined directly with behavioural indicators of anxiety or psychiatric diagnosis. Alternatively, these associations can be examined with constructs beyond symptoms that show closer relationships to brain function. The term 'endophenotype' has been used to describe such underlying constructs linked to both disorders and their underlying risks.[14]

Family studies consistently demonstrate strong associations among various anxiety disorders in parents and anxiety disorders in their children. More than 10 studies show that children born to parents with PD, SOPH, or GAD face a two- to four-fold increased risk for anxiety disorders.[5] As with data from longitudinal studies, some non-specificity emerges in other family studies. Children born to parents with MDD even in the absence of anxiety face the same elevated risk for paediatric anxiety as children born to parents with anxiety disorders. These data are consistent with a wealth of data among adults documenting strong familial associations.[15]

These findings on familial aggregation might reflect the effects of either genes or the environment. Data from twin and adoption studies suggest that genes account for approximately 40 per cent of the risk for anxiety both among children and adults.[5,15] Much like for longitudinal and family aggregation studies, twin studies suggest pathophysiologic similarities in the genetics of anxiety and depression. In particular, GAD and MDD appear to share many of the same genes.[16] In terms of specific genes, the field has only begun to examine associations with specific paediatric anxiety disorders. While considerable enthusiasm pertains to research on serotonin-related genes, this enthusiasm emerges predominantly from studies in adults.[11]

The effects of genes on risk for paediatric anxiety are not thought to moderate overt symptomatic expressions. Rather, genes, either as main effects or through interactions with the environment, are hypothesized to produce disruptions in the underlying function of the neural circuit illustrated in Fig. 9.2.6.1. These disruptions are expected to produce perturbations in physiologic regulation and information processing functions, examples of endophenotypes for paediatric anxiety disorders. Work on endophenotypes in paediatric anxiety disorders generally focus on three related profiles.

First, considerable work examines variations in children's temperaments, as they relate to both parental histories of anxiety as well as children's risk for anxiety, manifest later in life. This work shows that children who react with fear and hesitation in novel social scenarios face a high risk for anxiety.[3] This temperamental classification is known as 'behavioural inhibition'. Some evidence suggests that these associations pertain particularly strongly to the association with later-life SOPH.

Second, other work examines associations with variations in physiologic or cognitive responses to various threats. While various forms of fear and anxiety produce robust changes in autonomic physiology, inconsistent data document strong associations between individual differences in anxiety among humans and the magnitude of these physiological responses. Particularly interest focuses on between-group differences in conditioned physiologic reactions, but findings in this area appear weak and inconsistent.[17] Some of the strongest findings emerge from research on the startle response. This defensive reflex shows strong cross-species similarities in the degree to which it can be modulated by the presence of a threat. In general, adult anxiety disorders are characterized by enhanced startle in some contexts, though data in paediatric anxiety disorders appear inconsistent. Moreover, children born to parents with either PD or MDD also show enhanced startle responses under some circumstances.[18,19] Other work focuses on biases in various cognitive processes, such as attention and memory, where associations with specific disorders generally appear stronger than for studies of physiology. Here some work suggests that perturbations in face processing predict both risk for anxiety and the presence specifically of paediatric SOPH.[20]

Finally, perturbations in respiration have been linked most convincingly to the diagnosis of PD.[21] Data among adults show these respiratory perturbations aggregate within families. Moreover, findings in children and adolescents show that respiratory perturbation is associated with SAD but not other paediatric anxiety disorders, such as SOPH. Given evidence of familial aggregation between parental PD and childhood SAD, these data suggest that respiratory perturbation may confer a familial risk for panic disorder. Nevertheless, data examining respiratory function in offspring of PD patients are not consistent with this possibility.[22]

Stress

Work in animal models demonstrates strong relationships between exposure to various forms of physical or emotional stress

and individual differences in anxiety or fear.[11] These associations appear particularly strong in juvenile organisms. For studies among children and adolescents, strong associations also emerge with various measures of stress including either stress that occurs within the family or in other social contexts.[23] Nevertheless, it remains unclear the degree to which these association reflect specific connections with anxiety, given that stress is associated with a range of other psychopathologies besides anxiety.[24] Moreover, considerable heterogeneity exists in terms of the relationship between stress and anxiety, given that some individuals exposed to extreme stress are resilient, whereas other individuals exposed to mild stress develop anxiety. At least some of these individual differences are though to reflect the influences of genes on underlying neural circuitry, such that the development of paediatric anxiety reflects influences of gene–environment interactions.[11]

Brain imaging

Through studies of brain imaging, it is now possible to examine associations between paediatric anxiety disorders and perturbations in brain structure or function. Relatively few brain-imaging studies have examined paediatric anxiety disorders, and the few studies that do focus on neural structures depicted in Fig. 9.2.6.1.

Without question, the amygdala stands as the most frequently investigated brain structure in paediatric anxiety. Two studies examine amygdala morphometry in paediatric anxiety disorders, with one of these studies reporting reduced volume and the other reporting enlarged volume, both focusing mostly on children with GAD.[25,26] Such inconsistencies are consistent with mophometry studies in adults, where inconsistent evidence of amygdala enlargement or reduction emerges across a range of studies. Two other studies use functional magnetic resonance imaging (fMRI) to examine amygdala function in paediatric anxiety disorders.[27,28] Here, the findings are more consistent, much like in a larger series of studies in adults. Specifically, both studies reported enhanced amygdala activation in paediatric GAD, consistent with data in adult SOPH, MDD, as well as post-traumatic distress disorder or behavioural inhibition. Finally, imaging work focused on other structures documents abnormalities with less consistency. This includes structural and functional studies of the PFC and hippocampus.[5]

Therapeutics

A range of approaches has been suggested as useful in the treatment of paediatric anxiety disorders. The current chapter restricts considerations to a review of modalities studied with the randomized controlled trial (RCT). Two modalities have been studied in sufficient detail to provide conclusions on efficacy: cognitive behavioural psychotherapy treatment (CBT) and selective serotonin reuptake inhibitors (SSRIs). The data for these two treatments are reviewed in most detail, whereas other treatments are mentioned only briefly.

CBT relies on the principles of extinction, whereby an individual with an anxiety disorder undergoes exposure to a feared object or situation while relying on cognitive techniques taught as part of the therapy. In general, CBT is easiest when a child presents with a relatively specific set of fears and worries that allow the therapist to work with the child to create a fear hierarchy. The child then gradually undergoes exposure to situations on this hierarchy that are increasingly anxiety provoking. More than 10 studies use an RCT design to examine the efficacy of CBT, and the overwhelming majority of these document strong efficacy.[5] Nevertheless, most of these studies compare CBT to a wait-list control condition, a condition that may actually be aversive. The few RCTs of CBT using more suitable control conditions generally find weaker advantages for CBT. These studies suggest that CBT is a viable treatment option for any of the anxiety disorders considered in this chapter.

Five RCTs examine the efficacy of one or another SSRI in paediatric anxiety disorders, all using placebo control.[5] As with the data for CBT, this work provides strong justification for using SSRIs, in that robust treatment effects emerge. Moreover, these studies rely on placebo, a more credible control than in the CBT studies relying on wait-list comparison. Thus, the strength of the evidence supporting efficacy is probably somewhat greater in SSRIs than it is for CBT. Nevertheless, serious concerns about the safety of SSRIs emerged in 2002, due to the suggestion that SSRIs were associated with an increased risk over placebo for suicidal thoughts or behaviour. This ultimately led the Food and Drug Administration to place a 'black box' warning on the use of SSRIs in children, a warning recently extended to adults aged 25 and younger. Given these concerns, CBT probably represents the most reasonable first-line treatment for paediatric anxiety disorders. Among children who either cannot complete a course of CBT or who fail to respond to CBT, SSRIs represent an eminently reasonable treatment.

A range of other treatments have been considered for paediatric anxiety disorders. These include both psychotherapies, such as dynamically oriented therapy, and medications, such as various non-SSRI antidepressants or benzodiazepines. Due to either the dearth of data on efficacy or concerns with safety, all of these treatments should be considered third-line options, after CBT and SSRIs.

Conclusions

The current chapter reviews data for paediatric anxiety disorders in three sections. The longest section reviews clinical characteristics of these disorders. This section describes the clinical features, demography, and outcome of paediatric anxiety disorders, the most common class of mental syndrome afflicting children and adolescence. The first section is followed by a shorter section focused on pathophysiology. Here, data on neural circuits implicated in anxiety are reviewed most comprehensively. Finally, treatment is briefly reviewed in the third section. More detailed considerations of therapeutics for a range of paediatric psychiatric disorders can be found in section 9.5. This includes a discussion of both psychotherapies and medication.

Further information

Costello, E.J., Egger, H.L., and Angold, A. (2005). The developmental epidemiology of anxiety disorders: phenomenology, prevalence, and comorbidity. *Child and Adolescent Psychiatric Clinics of North America*, **14**, 631–48, vii.

Kagan, J., Snidman, N., McManis, M., *et al.* (2001). Temperamental contributions to the affect family of anxiety. *The Psychiatric Clinics of North America*, **24**, 677–88.

Pine, D.S. (2007). Research review: a neuroscience framework for pediatric anxiety disorders. *Journal of Child Psychology and Psychiatry, and allied disciplines*, **48**, 631–48.

Pine, D.S. and Klein, R.G. (2008). Anxiety disorders. In *Rutter's Child and Adolescent Psychiatry* (5th edn) (eds. M. Rutter, D. Bishop, D.S. Pine, *et al.*). Blackwell Publishing, Oxford.

References

1. Pine, D.S., Cohen, P., Gurley D., *et al.* (1998). The risk for early-adulthood anxiety and depressive disorders in adolescents with anxiety and depressive disorders. *Archives of General Psychiatry*, **55**(1), 56–64.
2. Brooks, S.J. and Kutcher, S. (2003). Diagnosis and measurement of anxiety disorder in adolescents: a review of commonly used instruments. *Journal of Child and Adolescent Psychopharmacology*, **13**(3), 351–400.
3. Perez-Edgar, K. and Fox, N.A. (2005). Temperament and anxiety disorders. *Child and Adolescent Psychiatric Clinics of North America*, **14**(4), 681–706, viii.
4. Shaffer, D., Gould, M.S., Fisher, P., *et al.* (1996). Psychiatric diagnosis in child and adolescent suicide. *Archives General of Psychiatry*, **53**(4), 339–48.
5. Pine, D.S. and Klein, R.G. (2008). Anxiety disorders. In *Rutter's child and adolescent psychiatry* (5th edn) (eds. M. Rutter, D. Bishop, D.S. Pine, *et al.*). Blackwell Publishing, Oxford.
6. Costello, E.J., Compton, S.N., *et al.* (2003). Relationships between poverty and psychopathology: a natural experiment. *The Journal of the American Medical Association*, **290**(15), 2023–9.
7. Angold, A., Costello, E.J., Erkanli A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry,* **40**(1), 57–87.
8. Beesdo, K., Bittner, A., Pine, D.S., *et al.* (2007). Incidence of social anxiety disorder and the consistent risk for secondary depression in the first three decades of life. *Archives of General Psychiatry*, **64**(8), 903–12.
9. Pine, D.S. (2002). Treating children and adolescents with selective serotonin reuptake inhibitors: how long is appropriate? *Journal of Child and Adolescent Psychopharmacology*, **12**(3), 189–203.
10. Aschenbrand, S.G., Kendall, P.C., Webb, A., *et al.* (2003). Is childhood separation anxiety disorder a predictor of adult panic disorder and agoraphobia? A seven-year longitudinal study. *Journal of the American Academy of Child and Adolescent Psychiatry*, **42**(12), 1478–85.
11. Gross, C. and Hen, R. (2004). The developmental origins of anxiety. *Nature Reviews. Neuroscience*, **5**(7), 545–52.
12. LeDoux, J.E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, **23**, 155–84.
13. Davis, M. (1998). Are different parts of the extended amygdala involved in fear versus anxiety? *Biological Psychiatry*, **44**(12), 1239–47.
14. Gottesman, I.I. and Gould, T.D. (2003). The endophenotype concept in psychiatry: etymology and strategic intentions. *The American Journal of Psychiatry*, **160**(4), 636–45.
15. Hettema, J.M., Neale, M.C., Kendall K.S., *et al.* (2001). A review and meta-analysis of the genetic epidemiology of anxiety disorders. *The American Journal of Psychiatry*, **158**(10), 1568–78.
16. Silberg, J.L., Rutter, M., Eaves L. (2001). Genetic and environmental influences on the temporal association between earlier anxiety and later depression in girls. *Biological Psychiatry*, **49**(12), 1040–9.
17. Lissek, S., Powers, A.S., McClure, E.B., *et al.* (2005). Classical fear conditioning in the anxiety disorders: a meta-analysis. *Behaviour Research and Therapy*, **43**, 1391–424.
18. Merikangas, K.R., Avenevoli, S., Dierker L., (1999). Vulnerability factors among children at risk for anxiety disorders. *Biological Psychiatry*, **46**(11), 1523–35.
19. Grillon, C., Warner, V., Hille J., *et al.* (2005). Families at high and low risk for depression: a three-generation startle study. *Biological Psychiatry*, **57**(9), 953–60.
20. Pine, D.S., Klein, R.G., Roberson-Nay, R., *et al.* (2005a). Face emotion processing and risk for panic disorder in youth. *Journal of the American Academy of Child and Adolescent Psychiatry*, **44**, 664–72.
21. Klein, D.F. (1993). False suffocation alarms, spontaneous panics, and related conditions. An integrative hypothesis. *Archives of General Psychiatry*, **50**(4), 306–17.
22. Pine, D.S., Klein, R.G., Roberson-Nay, R., *et al.* (2005b). Response to 5% carbon dioxide in children and adolescents: relationship to panic disorder in parents and anxiety disorders in subjects. *Archives of General Psychiatry*, **62**(1), 73–80.
23. Pine, D.S. and Cohen, J.A. (2002). Trauma in children and adolescents: risk and treatment of psychiatric sequelae. *Biological Psychiatry*, **51**(7), 519–31.
24. Wood, J.J., McLeod, B.D., Sigman, M., *et al.* (2003). Parenting and childhood anxiety: theory, empirical findings, and future directions. *Journal of Child Psychology and Psychiatry, and allied disciplines*, **44**, 134–51.
25. De Bellis, M.D., Casey, B.J., Dahl R.E., *et al.* (2000). A pilot study of amygdala volumes in pediatric generalized anxiety disorder. *Biological Psychiatry*, **48**(1), 51–7.
26. Millham, M.P., Nugent, A.C., Drevets, W.C., *et al.* (2005). Selective reduction in amygdala volume in pediatric generalized anxiety disorder: a voxel-based morphometry investigation. *Biological Psychiatry*, **57**, 961–6.
27. Thomas, K.M., Drevets, W.C., Dahl, R.E., *et al.* (2001). Amygdala response to fearful faces in anxious and depressed children. *Archives of General Psychiatry*, **58**(11), 1057–63.
28. McClure, E.B., Monk, C.S., Nelson E.E., *et al.* (2007). Abnormal attention modulation of fear circuit function in pediatric generalized anxiety disorder. *Archives of General Psychiatry*, **64**(1), 97–106.

9.2.7 Paediatric mood disorders

David Brent and Boris Birmaher

In this chapter, we describe the nosology and epidemiology of paediatric unipolar and bipolar disorders, risk factors and predictors of course, and the evidence base for pharmacological and psychosocial treatments. We conclude this chapter by suggesting areas for future research.

Clinical picture

Mood disorders may be classified on three dimensions: (a) severity; (b) course; and (c) presence or absence of mania/hypomania.[1] Depressed children and adolescents may not describe their mood as sad, but instead as, 'grouchy', 'bored', 'having no fun', or 'empty'.[2] The most severe depressive condition is major depression, which requires at least 2 weeks of a depressed, sad, bored, or anhedonic mood for most of the time, and four additional depressive symptoms involving impairment in concentration, suicidal thoughts, difficulty making decisions, impaired sleep and appetite, guilt, and a decreased sense of self-worth (Box 9.2.7.1). Patients with depressive symptoms, but whose clinical picture is below the threshold for major depression (so-called minor depression or depression NOS) can still show significant impairment.[3] Dysthymic disorder is more chronic and intermittent than major depression, with periods of depression interspersed with normal mood, but with duration of at least 1 year (Box 9.2.7.2). Adjustment disorder with depressed

Box 9.2.7.1 Criteria for the diagnosis of a major depressive episode

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

   Note: Do not include symptoms that are clearly due to a general medical condition, or mood-incongruent delusions or hallucinations.

   1 Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g. feels sad or empty) or observation made by others (e.g. appears tearful). *Note*: In children and adolescents, can be irritable mood.

   2 Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation made by others).

   3 Significant weight loss when not dieting or weight gain (e.g. a change of more than 5 per cent of body weight in a month), or decrease or increase in appetite nearly every day. *Note*: In children, consider failure to make expected weight gains.

   4 Insomnia or hypersomnia nearly every day.

   5 Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).

   6 Fatigue or loss of energy nearly every day.

   7 Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional nearly every day (not merely self-reproach or guilt about being sick).

   8 Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).

   9 Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

B. The symptoms do not meet criteria for a Mixed Episode (see p. 365).[1]

C. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

D. The symptoms are not due to the direct physiological effects of a substance (e.g. a drug of abuse, a medication (or a general medical condition (e.g. hypothyroidism).

E. The symptoms are not better accounted for by bereavement, i.e. after the loss of a loved one, the symptoms persist for longer than 2 months or are characterized by marked functional impairment morbid preoccupation with worthlessness, suicidal ideation, psychotic symptoms, or psychomotor retardation.

(Modified from APA (2000), *Diagnostic and statistical manual of mental disorders* (4th edn), American Psychiatric Association Press, Washington, DC.)

Box 9.2.7.2 Criteria for the diagnosis of dysthymic disorder

A. Depressed mood for most of the day, for more days than not, as indicated either by subjective account or observation by others, for at least 2 years. *Note*: In children and adolescents, mood can be irritable and duration must be at least 1 year.

B. Presence, while depressed of two (or more) of the following:

   1 poor appetite or overeating

   2 insomnia or hypersomnia

   3 low energy or fatigue

   4 low self-esteem

   5 poor concentration or difficulty making decisions

   6 feelings of hopelessness

C. During the 2-year period (1 year for children or adolescents) of the disturbance, the person has never been without the symptoms in Criteria A and B for more than 2 months at a time.

D. No Major Depressive Episode (see p. 356)[1] has been present during the first 2 years of the disturbance (1 year for children and adolescents); that is the disturbance is not better accounted for by chronic Major Depressive Disorder, or Major Depressive Disorder, In Partial Remission. *Note*: There may have been a previous Major Depressive Episode provided there was a full remission (no significant signs or symptoms for 2 months) before development of the dysthymic disorder. In addition, after the initial 2 years (1 year in children or adolescents) of dysthymic disorder, there may be superimposed episodes of Major Depressive Disorder, in which case both diagnoses may be given when the criteria are met for a Major Depressive Episode.

E. There has never been a Manic Episode (see p. 362),[1] a Mixed Episode (see p. 365),[1] or a Hypomanic Episode (see p. 368),[1] and criteria have never been met for Cyclothymic Disorder.

F. The disturbance does not occur exclusively during the course of a chronic Psychotic Disorder, such as Schizophrenia or Delusional Disorder.

G. The symptoms are not due to the direct physiological effects of a substance (e.g. a drug of abuse, a medication) or a general medical condition (e.g. hypothyroidism).

H. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

(Modified from APA (2000), *Diagnostic and statistical manual of mental disorders* (4th edn), American Psychiatric Association Press, Washington, DC.)

mood is a milder and self-limited disturbance of mood that follows a significant life stressor (Box 9.2.7.3).

The presence of clinically significant manic or hypomanic symptomatology suggests bipolar spectrum disorder. The symptomatology of mania can be thought of as the mirror image of depression, with mood characterized by elation or grandiosity. Mania is associated

Box 9.2.7.3 Criteria for the diagnosis of adjustment disorder with depressed mood

A. The development of emotional or behavioural symptoms in response to an identifiable stressor(s) occurring within 3 months of the onset of the stressor(s).

B. These symptoms or behaviours are clinically significant as evidenced by either of the following:

1 marked distress that is in excess of what would be expected from exposure to the stressor

2 significant impairment in social or occupational (academic) functioning

C. The stress-related disturbance does not meet the criteria for another specific Axis I disorder and is not merely an exacerbation of a pre-existing Axis I or Axis II disorder.

D. The symptoms do not represent bereavement.

E. Once the stressor (or its consequences) has terminated, the symptoms do not persist for more than an additional 6 months.

(Modified from APA (2000), *Diagnostic and statistical manual of mental disorders* (4th edn), American Psychiatric Association Press, Washington, DC.)

with clear impairment, whereas hypomania, while associated with a change in functioning, does not always result in impairment per se. Bipolar individuals, especially in the paediatric age group, frequently do not show the classic distinct alternating manic and depressive periods found in adult bipolar patients. Instead they may either experience depression and manic symptoms simultaneously, so-called mixed episodes, or alternations of mania and depression that may occur within a month, a week, or even a day, e.g. rapid cycling.[4] Common symptoms of paediatric bipolar disorder are pressure of speech, increased energy, and decreased need for sleep. Risk-taking behaviour showing poor judgement (e.g. gambling, hypersexuality, excessive spending) and joking and excessive humour are very specific for paediatric bipolar disorder, but less common in paediatric samples. While irritability is a common symptom of paediatric bipolar disorder, it is very non-specific and is commonly found in many other conditions, such as depression, oppositional defiant disorder, and attention deficit disorder. The DSM-IV requires a relatively long duration of mania (7 days) and hypomania (4 days) in order to meet criteria. Many paediatric patients may show the same symptom pattern but have very rapid cycling and therefore, do not meet these criteria. If altered function is present, such patients should receive a diagnosis of Bipolar Disorder NOS. Bipolar NOS is a common diagnosis for children with manic symptoms because very often, paediatric bipolar illness does not fulfil the duration criteria for mania, in part due to the frequency of rapid cycling conform to the classic adult patterns of distinct patterns of mania and depression.[4] However, in children and adolescents, Bipolar disorder NOS does not appear to be different from Bipolar I or II with regard to impairment, rate of comorbid disorders, response to treatment, or family history of bipolar disorder, and many patients with BP-NOS upon longitudinal follow-up go on to develop BP-I or BP-II disorders.[4] Individuals who have had a history of full mania plus major depression receive a diagnosis of Bipolar I disorder, those with hypomania plus major depression receive a diagnosis of Bipolar II disorder, and those with hypomania and dysthymia receive a diagnosis of cyclothymic disorder (see Boxes 9.2.7.4 and 9.2.7.5).

While some in the field continue to raise questions about the validity of the diagnosis of paediatric bipolar disorder, the convergent evidence from longitudinal and high-risk studies is that there it is the essentially an earlier manifestation of the same illness as is found in adults, is highly familial, and shows a chronic and consistent course.[4,5]

Differential diagnosis

Attention deficit hyperactive disorder (ADHD) and disruptive disorders

Patients with ADHD, oppositional disorder, and conduct disorder are often irritable, show a low frustration tolerance, and can become demoralized due to school failure and peer rejection. However, in the absence of true depression, their mood will be restored as soon as the source of their frustration has been remedied. While both ADHD and depression are associated with poor concentration, the age of onset of ADHD is usually earlier than in mood disorders. Patients with ADHD have other accompanying difficulties such as hyperactivity and impulsivity that are part of the depressive picture. Conversely, depressed patients will show changes in sleep, energy level, appetite, mood, and self-worth that are not part of the picture of ADHD. The symptoms of ADHD, such as poor concentration, hyperactivity, and impulsivity can also be seen in bipolar disorder but patients with ADHD rarely have concomitant hypersexuality, grandiosity, and decreased need for sleep.[4,5] However, hypersexuality may also be seen in victims of sexual abuse, but in contrast with the hypersexuality of bipolar disorder, is not accompanied by clinically significant grandiosity, pressure of speech, increased energy, and diminished need for sleep. Often, the diagnostic difficulty is not simply distinguishing between disruptive and mood disorders, but in the proper attribution of shared symptoms in patients with comorbidity, as is very often the case. When patients have both mood disorder and ADHD, usually the ADHD antedates the mood disorder. A diagnosis of a mood disorder can only be made when the shared symptoms, such as impaired concentration become worse in association with depressed or manic mood.

Anxiety disorders

Patients with anxiety disorder may also become quite dysphoric, but when the anxiogenic situation is removed, normal mood frequently ensues. Anxiety is a frequent antecedent of paediatric depression and bipolar disorder.[5,6] Symptoms that are shared between disorders, such as difficulty with sleep, or impaired concentration, are attributed to the mood disorder only if they become worse with the onset of a depressed or manic mood state. Panic disorder is often comorbid with paediatric bipolar disorder.[5] However, it is important to distinguish between the symptoms of panic disorder, that are prominently somatic and associated with thoughts and feelings of dread, and rapid cycling and a mixed state, which are marked with mood instability and the presence of simultaneous, or rapidly alternating depressive and manic symptoms.

Box 9.2.7.4 Criteria for the diagnosis of bipolar disorder

A. Currently (or most recently) in a Manic Episode (see p. 362).[1]

B. There has previously been at least one Major Depressive Episode (see p. 356),[1] Manic Episode (see p. 362),[1] or Mixed Episode (see p. 365).[1]

C. The mood episodes in Criteria A and B are not better accounted for by Schizoaffective Disorder and are not superimposed on Schizophrenia, Schizophreniform Disorder, Delusional Disorder, or Psychotic Disorder Not Otherwise Specified.

Past or current history of a Manic Episode is characterized by:

A. A distinct period of abnormally and persistently elevated, expansive, or irritable mood, lasting at least 1 week (or any duration if hospitalization is necessary).

B. During the period of mood disturbance, three (or more of the following symptoms have persisted (four if the mood is only irritable) and have been present to a significant degree:

1 inflated self-esteem or grandiosity

2 decreased need for sleep (e.g. feels rested after only 3 h of sleep)

3 more talkative than usual or pressure to keep talking

4 flight of ideas or subjective experience that thoughts are racing

5 distractibility (i.e. attention to easily drawn to unimportant or irrelevant external stimuli)

6 increase in goal-directed activity (either socially, at work or school, or sexually) or psychomotor agitation

7 excessive involvement in pleasurable activities that have a high potential for painful consequences (e.g. engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments)

C. The symptoms do not meet criteria for a Mixed Episode (see p. 365).[1]

D. The mood disturbance is sufficiently severe to cause marked impairment in occupational functioning or in usual social activities or relationships with others, or to necessitate hospitalization to prevent harm to self or others, or there are psychotic features.

E. The symptoms are not due to the direct physiological effects of a substance (e.g. a drug of abuse, a medication, or other treatment) or a general medical condition (e.g. hyperthyroidism).

Note: Manic-like episodes that are clearly caused by somatic antidepressant treatment (e.g. medication, electroconvulsive therapy, light therapy, should not count toward a diagnosis of Bipolar I disorder.

(Modified from APA (2000), *Diagnostic and statistical manual of mental disorders* (4th edn), American Psychiatric Association Press, Washington, DC.)

Box 9.2.7.5 Criteria for the diagnosis of cyclothymic disorder

A. For at least 2 years, the presence of numerous periods with hypomanic symptoms (see p. 368)[1] and numerous periods with depressive symptoms that do not meet criteria for a Major Depressive Episode. *Note*: In children and adolescents, the duration must be at least 1 year.

B. During the above 2-year period (1 year in children and adolescents), the person has not been without the symptoms in Criterion A for more than 2 months at a time.

C. No Major Depressive Episode (p. 356),[1] Manic Episode (p. 362),[1] or Mixed Episode (see p. 365)[1] has been present during the first 2 years of the disturbance.

Note: After the initial 2 years (1 year in children and adolescents) of Cyclothymic Disorder, there may be superimposed Manic or Mixed Episodes (in which case both Bipolar I disorder and Cyclothymic Disorder may be diagnosed) or Major Depressive Episodes (in which case both Bipolar II disorder and Cyclothymic Disorder may be diagnosed).

D. The symptoms in Criterion A are not better accounted for by Schizoaffective Disorder and are not superimposed on Schizophrenia, Schizophreniform Disorder, Delusional Disorder, or Psychotic Disorder Not Otherwise Specified.

E. The symptoms are not due to the direct physiological effects of a substance (e.g. a drug of abuse, a medication) or a general medical condition (e.g. hyperthyroidism).

F. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

(Modified from APA (2000), *Diagnostic and statistical manual of mental disorders* (4th edn), American Psychiatric Association Press, Washington, DC.)

Substance abuse

The use of marijuana, alcohol, or opiates can mimic the symptoms of depression, such as difficulty with concentration, motivation, low energy, and dysphoria. Amphetamine and cocaine abuse can mimic mania. Depressed and bipolar patients are at greatly increased risk of abusing substances, so that the presence of substance abuse does not rule out a mood disorder or vice versa, but in fact, should raise the suspicion of possible comorbidity.

Eating disorder

Patients with a restricting eating disorder who are nutritionally compromised may show symptoms that overlap with depression, including decreased appetite, low energy, and sad mood. Often the sadness is found in patients with anorexia who are being forced to gain or maintain weight against their will. A diagnosis of depression, unless there is a clear historical precedent that antedates the eating disorder, should only be made when the nutritional status of the patient has been normalized. Bulimic disordered patients often have difficulties with impulse control that need to be differentiated from bipolar disorder.

Borderline personality disorder

Although there is evidence that borderline personality disorder can be reliably diagnosed in adolescents,[7] diagnostic convention requires that this diagnosis only be applied for adults. Still, there is general agreement that many adolescents, particularly those with mood disorders, have 'borderline features', such as mood lability, impulsivity, suicidal thoughts and behaviour, chaotic interpersonal relationships, and risky behaviour that has a high likelihood of resulting in personal harm. Others have argued that borderline personality disorder is really a form of bipolar spectrum disorder, although family studies have not confirmed this.[5] Instead, the high degree of overlap between personality disorder and bipolar disorder suggests that care be taken in not attributing symptoms that more appropriately are associated with a lifelong personality style to bipolar disorder. Conversely, in the presence of a clear and unremitting paediatric mood disorder, personality disorder should not be diagnosed.

Psychosis

Although rare in childhood, incipient schizophrenia can present with sad and detached mood, sleep disturbance, and social withdrawal. Psychotic symptoms that evolve in schizophrenia are more likely to be mood-incongruent. In contrast, psychosis in depression and bipolar disorder is more often, but not always, mood-congruent.[2,5] This is a diagnosis that often can only be made upon careful longitudinal follow-up. Since psychosis is often seen in youth with mood disorders, and schizophrenia is rare at this age group, any child or adolescents with psychosis needs to be carefully assessed for the presence of a mood disorder, particularly bipolar illness.

Comorbidity

Comorbidity is the rule, rather than the exception.[8] Anxiety disorder frequently antedates paediatric depression and bipolar disorder, with common precursors being social phobia and panic disorder, respectively. ADHD is frequently comorbid with both conditions. Substance abuse is often a complication of mood disorder, although this condition in turn lengthens episodes and increases the risk for recurrence.

Medical comorbidity

Medications used to treat epilepsy, inflammatory bowel disease, and rheumatic and allergic disease can have profound effects on mood. Corticosteroids can induce depression or mania. Phenobarbital is associated with depression, as is use of interferon.[9] Moreover, systemic aspects of the diseases themselves may increase the risk for depression, in epilepsy, asthma, diabetes, and thyroid illness. Oral contraceptives can also result in mood changes.

Descriptive epidemiology

The point prevalence of major depression in around 1–2 per cent in prepubertal samples, and between 3–8 per cent in adolescent samples.[10] The prevalence of bipolar disorder is around 1 per cent in paediatric populations, although the rate of 'soft' bipolar disorder, which has some, but not all of the core features of bipolar disorder, has been reported to be as high as 5 per cent in some adolescent samples.[11] The male to female ratio is around 1:1 for prepubertal depression, but increases to around 1:3 for depression after puberty. In contrast, the males and females have similar risk for bipolar disorder, regardless of pubertal status. The increased rate of depression after puberty is accounted for almost entirely by the increased risk in females, and may be related to changes in estradiol and testosterone associated with puberty.[12] Prepubertal major depression is an admixture of two subtypes: one is highly familial, with a high risk for recurrence and for eventual paediatric mania, and the second with comorbid with disruptive disorders, a low risk for depressive recurrence, an association with parental criminality, substance abuse, and family discord, and a course more similar to conduct disorder than to mood disorder.[13] A clear clinical syndrome of depression has been reported in children as young as aged 3, particularly in those young children with a family history of mood disorders.[14]

Course

Paediatric mood disorders tend to be both chronic and recurrent. While prepubertal depression comorbid with conduct disorder is likely not to be recurrent, studies of child-onset depression with a family history of depression show high rates of recurrence, with risks of recurrence of 40 per cent in 2 years, and over 70 per cent within 5 years.[2] The average length of a depressive episode is around 4–6 months in community samples, and 6–8 months in clinical samples.[15] The duration of dysthymic disorder is much longer, on average, around 5 years, according to one careful longitudinal study.[2] In patients with comorbid dysthymic disorder and depression, so-called double depression, the risk for prolonged episodes and recurrence are both very high.[2] Longer episodes are also predicted by comorbidity with substance abuse, conduct disorder, or anxiety disorder, family conflict, and parental depression.[10,16]

Paediatric bipolar disorder does not often present with 'classic' periods of alternating depression and mania. Instead, such patients frequently present with either a mixed state, e.g. simultaneous occurrence of depression and mania, or rapid cycling, with brief and alternating periods of depression and mania.[4] In comparing the course of paediatric and adult bipolar patients, paediatric bipolar patients have many more episodes per year, and spend less time in remission.[4] Consistent with these longitudinal observations are findings from adult pedigrees that age of onset in bipolar disorder appears to be familial, and that earlier age of onset is associated with higher rates of drug abuse, alcohol abuse, rapid cycling, and suicide attempts.[17] Much of the impairment in paediatric bipolar disorder is associated with depressive symptoms that often never completely remit. As noted above, the adult criteria requiring 1 week and 4 days for mania and hypomania, respectively, may be overly stringent, insofar as a fairly high proportion (25 per cent) of patients below those criteria, so-called bipolar NOS, go on in longitudinal follow-up to develop clear Bipolar I or II disorder and 20 per cent of those with BP-II go on to develop BP-I within 2 years of follow-up.[4] A longer period to recovery is predicted by longer duration of mood disorder, rapid cycling or mixed episode, psychosis, and lower SES.[4]

Children and adolescents who present with a unipolar depressive disorder are at increased risk for developing a bipolar disorder, both in comparison to children without a mood disorder, and to individuals whose mood disorder has its onset in adulthood. Young depressed patients with a family history of bipolar disorder, who present with psychotic symptoms, and/or pharmacologically

induce mania or hypomania are at increased risk for developing paediatric bipolar disorder.[5,18] According to one pharmacoepidemiological study, the younger the depressed patient, the higher the risk for pharmacologically induced mania, although there was no standardized, direct assessment of manic behaviour.[19] Additionally, one study suggests that paediatric bipolar patients with comorbid ADHD tolerate amphetamine as well as non-bipolar children with ADHD.[20]

Sequelae

The most dreaded consequent of paediatric mood disorders is suicide. A unipolar depression conveys a 10–60-fold increased risk of suicide; nearly 80 per cent of adolescent suicide attempters have some form of a mood disorder.[21] Suicide attempts may be even more frequent in paediatric bipolar disorder, with almost one-third of clinical samples showing a lifetime history of a suicide attempt. Studies in adults and adolescents that have assessed for bipolar disorder find that as many as 10–20 per cent of all suicides have some form of bipolar spectrum disorder.[21] Correlates of suicidal behaviour in both unipolar and bipolar disorder include earlier age of onset, history of abuse, comorbid disruptive and substance abuse disorders, hopelessness, mood lability, and chronic and unremitting course.[21]

Other sequelae of untreated depression include educational and occupational under-attainment, interpersonal difficulties, obesity, cardiovascular disease, and alcohol and substance abuse. The effect of depression on body mass index (BMI) appears to be independent of treatment effects.[22]

Aetiology

Both unipolar and bipolar disorders have a strong genetic component. The child of a unipolar depressed parent is at around 2–4 times the risk of the population to develop a depressive disorder; this is even higher in children of parents with earlier onset (<age 20) and recurrent depression.[23] Twin studies indicate that around 50 per cent of the variance in familial transmission of depressive symptoms is explained by heritable factors, and that the liability of depression and to anxiety may be co-transmitted.[6] There is some evidence that adolescent onset depression is more highly heritable than prepubertal depression.[24]

Bipolar disorder has an even stronger genetic component, with a bipolar parent conveying at least an 8-fold increased risk of bipolar spectrum disorder in children; twin studies suggest that heritable factors account for 70–80 per cent of the variance in familial transmission.[23] High-risk studies of both unipolar depression and bipolar disorder show the transmission of a wider phenotype, with increased rates of anxiety disorder being most prominent in both the offspring of depressed and bipolar parents.[5,25]

Genetic linkage studies are beginning to converge on specific regions associated with depression and anxiety, with other regions implicated in bipolar disorder.[23] Linkage studies for early onset, recurrent depression suggest that there may be sex-specific linkage sites.[23]

The serotonin transporter promoter gene has a 44 bp insertion/deletion, with the latter resulting in less vigorous transcription. Several studies have now shown an interaction between stressful life events and the less functional form of the transporter gene with regard to an increased risk for depression in adolescents and young adults.[26] brain-derived neurotrophic factor (BDNF) is a gene that appears to protect the hippocampus system from the neurotoxic effects of stress. A three-way interaction between maltreatment, the less functional 5HTTLPR allele, and the met allele of the BDNF gene has been found with regard to risk for depression in children and adolescents.[27] An association between the val66 form of BDNF and early onset depression has been reported.[28]

Genetic linkage studies have also identified regions of interest for bipolar disorder, although paediatric bipolar disorder per se has not been studied. In aggregate, linkage studies converge on chromosomal regions 6q and 8q.[23,29] Certain phenotypes of bipolar disorder have been reported, such as 'lithium responsiveness', 'alcoholism/suicidal behaviour', and 'psychosis', with distinct areas of familial aggregation and linkage.[17] Geller *et al.*[30] has reported in an association study of paediatric bipolar families that val66 form of BDNF is associated with the disease.

Structural changes in depression have been reported, with the most widely replicated result being changes in the anterior cingulated, an area associated with emotion regulation, in both bipolar and unipolar familial depression.[31]

Paediatric bipolar subjects, compared to healthy controls, show decreased grey matter in the dorsolateral prefrontal cortex (DLPFC), cingulate cortex, and amygdale.[32] Diffusion tensor imaging and magnetic resonance spectroscopy both point to alterations in axonal development and organization in the superior frontal (decreased white matter) and DLPFC regions (decreased levels of *n*-acetyl aspartate), respectively.[33,34]

Neurocognitive factors and emotion regulation

Depression

Depressed children and adolescents, relative to normal controls, show greater attention to, and distraction by sad stimuli, whereas normal controls are more likely to be distracted by happy stimuli.[35] Depressed adolescents are more vulnerable to the effects of rumination and sad mood induction[35] Depressed children and adolescents may show less activation of reward-related circuitry when participating in a reward paradigm.[36] These findings are consistent with earlier research showing that tendency to pessimism and rumination were risk factors for the onset of depressive symptoms, especially when confronted with stressful life events.[35] Functional neuroimaging studies find alterations in amygdala activation to threat and other cognitive tasks, although the direction is not consistent across studies.[37,38]

There is evidence that these tendencies may be present prior to the development of a mood disorder. The young children of mothers with a history of childhood onset depression (COD) show greater evidence of physiological distress (e.g. poor heart rate recovery after disappointment, resort to more passive waiting and less active distraction than normal controls, and show less efficient cognitive processing when confronted with an affectively laden cognitive task.[39,40] Positive reward anticipation, however, moderated the relationship between parent early onset depression and child internalizing symptoms.[41]

The extent to which negative affective bias, and difficulty with active coping is intrinsic versus learned is still unclear. Mothers with early onset depression are less responsive to their children's expression of distress, tend to endorse and promote fewer emotion regulation strategies for their children, whereas maternal accuracy

of recognition of their child's emotional state was protective against child psychopathology.[42–44] Taken together, these findings support a role in helping parents teach their children emotion regulation strategies such as distraction, emphasis on positive reward, and active coping as a means of preventing or treating depression.[45]

Bipolar disorder

There is growing consensus that paediatric bipolar disorder is associated with difficulty with attention, verbal and visuospacial memory, executive function, set-shifting, and recognition of facial expressions.[32] In some studies, these findings are present regardless of current mood state or medication[46] and some of these findings are present in the unaffected, high-risk offspring of adults with bipolar disorder.[47] Bias towards threat, and less activation of areas involved in emotion regulation in the face of frustration have also been reported.[32] Many of these findings are also reported in unipolar depression, and therefore, a direct comparison of these two conditions is needed.

Functional neuroimaging data are consistent with these findings, with greater activation of reward-related circuitry (e.g. caudate and thalamus) and greater activation of inhibitory areas (e.g. DLPFC, anterior cingulated) when performing working memory tasks and viewing negatively valenced pictures compared to healthy controls.[32]

Neuroendocrine/sleep

Neuroendocrine studies suggest that alterations in serotonergic and noradrenergic neurotransmission are associated with early onset unipolar depression.[48] There are no clear findings with regard to cortisol regulation and depression, although hypersecretion of cortisol close to the time of sleep has been reported to be related to adolescent depression.[49] While subjective sleep complaints are common in child and adolescent depression, polysomnographic studies have not consistently shown the decreased REM latency associated with depression that has been reported in adults.[50]

Environmental risk factors

Environmental factors can also influence the onset and expression of paediatric mood disorders, often interacting with a genetic diathesis. Early abuse and neglect is a profound risk factor for depression, especially in interaction with a positive family history for mood disorder and certain genetic polymorphisms.[26,27,51] Family discord, substance abuse, and criminality are associated with depression in children from families at low familial risk for depression.[52] Maternal-child conflict shows a birectional relationship over time with regard to both parental and child mood disorder.[53] A history of abuse is associated with an earlier age of onset and more prolonged and unremitting course in bipolar disorder.[54] Loss of a parent, close friend, or sibling is associated with an increased risk of depression in those children and adolescents with a pre-existing depressive diathesis.[55] Conversely, a positive connection to family, school, and a pro-social peer group, can, in cross sectional studies, protect against depression and other health risk behaviours.[56]

Assessment and monitoring

The properties of different assessment tools of mood disorder were recently reviewed.[57,58] The most common interview-based assessment for the severity of depression is the Children's Depression Rating Scale, Revised (CDRS-R), a 17-item rating scale. Commonly used self-rating measures are the Children's Depression Inventory for children and early adolescents, the Beck Depression Inventory (for adolescents and adults), the Reynolds Adolescent Depression Scale, the Center for Epidemiological Studies, Depression Scale (CES-D), and the Mood and Feelings Questionnaire (MFQ). The CDI shows treatment sensitivity, but does not distinguish well between anxiety and depression, and the CES-D does not have an item about suicidal ideation. The main advantage of the MFQ is that it has a short form for screening, has been validated in both community and clinical sites, has a parent form (so does the CDI), and can be used for both children and adolescents.

Two interview-based methods for monitoring the level of mania are the Young Mania Rating Scale and the Mania Rating Scale from the K-SADS.[4] The advantage of the latter is that it can be taken from one of the most commonly used diagnostic interviews for children and adolescents. A self-report for mania also shows promise, but has not yet been used in clinical trials.[5]

Treatment and prevention

Depression

Practice guidelines recommend that the initial treatment for mild-to-moderate depression be support, education, and one of two forms of psychotherapy (cognitive behaviour therapy [CBT] or interpersonal therapy [IPT]). For patients who eschew psychotherapy, or who live in a region where specific indicated forms of psychotherapy are not available, antidepressant medication is an appropriate approach. Guidelines agree that for more severe depression, combination of antidepressant medication and psychotherapy are ideal, although the data are mixed on this point (reviewed below).[59,60]

Because depression is a chronic and recurrent illness, a long-term approach should be taken to the management of child and adolescent depression. Therefore, after symptomatic relief, treatment should be continued for at least 6–12 months, since there is evidence that without psychotherapy booster sessions or continued medication treatment, there is a substantially increased risk of relapse or recurrence.[59]

There is strong and convergent evidence for the efficacy of CBT for adolescent depression in clinical samples, and for child depression in symptomatic volunteers,[59,61] with a relatively modest effect size ($d = 0.34$) relative to a waitlist condition or a comparison treatment. While some studies in clinical samples showed superiority of CBT to credible alternative treatments, the largest and most comprehensive study of the treatment of adolescent depression found that CBT was no better than placebo with regard to acute clinical response (43 per cent versus 35 per cent).[62] In that study, CBT was more efficacious than placebo in those with higher incomes (>$75 000) and with higher levels of cognitive distortion.[63] While the response rate for combination (of medication and CBT) treatment was not different than for medication alone (71 per cent versus 61 per cent), combination resulted in a more rapid response and greater likelihood of remission (37 per cent versus 20 per cent) than medication alone.[63] However, studies of combination treatment for depression have been inconsistent. The addition of CBT to antidepressant treatment, both in primary care, and to the

management of moderately to severely ill depressed patients failed to improve outcome over antidepressant treatment alone.[64,65] A more recent study of depressed adolescents who did not respond to an adequate initial trial of an antidepressant found that the combination of medication and CBT resulted in a higher rate of improvement among subjects than medication alone.[66]

Interpersonal therapy is a well-established treatment for adult depression, and more recently has been shown to be superior to waitlist control, clinical management, and treatment as usual.[61] This treatment has been demonstrated to be superior to treatment as usual in a community setting, as well.[59] Other forms of treatment that show promise are attachment-based therapy and family psychoeducation, but have not yet been replicated by other groups.[59]

A group CBT approach that has been used for the treatment of depression,[67] has been adapted for the prevention of depression. In adolescents with subsyndromal depression, the group CBT resulted in a lower risk of onset of major depression than in the treatment usual group.[67] This approach was extended to the adolescent offspring of depressed parents. Adolescents, in addition to having parents with a history of depression needed to have had a prior depressive episode or subsyndromal symptoms. In an initial clinical trial and one 4-site replication, the intervention resulted in a 2–5-fold lower risk for new-onset depression.[68] The presence of current depression in the caregiver moderated the effectiveness of the CBT intervention, with children of parents with current depression failing to show an effect from the intervention. Weissman *et al.*[69] recently demonstrated that treatment of maternal depression resulted in symptomatic improvement in their children, particularly with regard to internalizing symptoms. A family psychoeducational approach has been shown to improve communication and support with regard to depression in a family member, although there was no difference in the incidence of depression in this approach versus a comparison educational treatment.[70]

The selective serotonin reuptake inhibitors (SSRIs) form the mainstay of medication management of depression. Fluoxetine is the only medication that is approved by the Food and Drug Administration (FDA) and the Medicines and Health care Regulatory Agency (MHRA) for use in paediatric depression, because it has the strongest evidence of efficacy (http://www.fda.gov and http://www.mhra.gov.uk). Other medications for which there is some evidence of efficacy are citalopram, sertraline, and venlafaxine, although for each of these medications, there is some evidence that these agents are more efficacious for adolescent than for child-onset depression.[71] In contrast, fluoxetine shows similar efficacy for both children and adolescents. Tricyclic antidepressants have been shown to be ineffective for children and adolescents with depression.[72] One possible exception is clomipramine, that has been demonstrated, when given IV to reverse chronic and refractory depression.[73] While there are no controlled trials, buproprion is commonly used for paediatric depression, and in open trials shows evidence of efficacy.[74] One small controlled study suggests that omega-3 fatty acids may be efficacious for the relief of child depression.[75]

Children and adolescents metabolize several of the antidepressants more quickly (e.g. citalopram, sertraline) compared to adults, and so equal or higher doses may be required in order to achieve a similar effect.[76] There has been little work in pharmacogenetics in paediatric populations, although one study has replicated adult findings showing that the less functional form of the serotonin transporter gene is associated with a less vigorous response to an antidepressant.[77]

The use of SSRIs increased steadily over the past decade, but enthusiasm for their use on the part of both families and clinicians has been curtailed by recent reports of an association between antidepressant use and the occurrence of spontaneously-reported suicidal adverse events (i.e. new-onset or worsening suicidal ideation or an attempt), which resulted in the FDA issuing a black box warning about this side effect. The FDA conducted a meta-analysis that showed on average that around 4 per cent of the drug-treated and 2 per cent of those on placebo developed a suicidal adverse event.[78] In the subset of studies where suicidal ideation was measured systematically, there was no difference in suicidality by treatment condition, with a trend towards a protective effect in the medication group. A more recent meta-analysis, using random- rather than fixed-effects modelling and including more studies that were not available at the time of the FDAs analysis also found an increased risk, although the estimates of the risk difference for suicidal ideation and behaviour were 0.7 per cent rather than 2 per cent.[71] In this meta-analysis, the benefits of antidepressants were also assessed, and in the case of depression, around 11 times more individuals showed clinical improvement than developed these suicidal adverse events, suggesting a favourable risk-benefit ratio for the use of antidepressants, given careful clinical monitoring.

For patients who have been treated with psychotherapy or medication, addition of a complementary modality (e.g. psychotherapy or medication) is indicated. Family discord should be addressed by family therapy, and parental depression should be identified and referred for treatment. Failure to respond at that point suggests the need to try a second SSRI, followed by either venlafaxine or bupropion. A recent study comparing depressed adolescents who did not respond to an adequate trial with an SSRI found that a switch to second SSRI was as efficacious as a switch to venlafaxine.[66] Augmentation is indicated if a patient shows a partial but palpable response but still is symptomatic, whereas those who have not responded at all should be tapered and switched to another medication.[79] There have been no clinical trials of augmentation in paediatric depressed subjects, but placebo-controlled trials in adults support the use of augmentation of SSRI treatment with lithium, T3, and bupropion.[80] Also, for depression with a seasonal component, light therapy has been shown to be efficacious in psychiatric clinical trials.[81]

Paediatric bipolar disorder

Because paediatric bipolar disorder is rarer than unipolar depression and because of the previous controversy about diagnosis, there has only recently been increased attention paid to its treatment. Best practice guidelines are based upon downward extension of experience in adults, as well as a handful of clinical trials, but the field is changing very rapidly, and it is expected that these guidelines will change in parallel.[82] Paediatric bipolar disorder has intrinsic in it a paradox: the most functionally impairing aspect of the condition is depression, but treatment of depression may induce mania. Therefore, treatment of paediatric bipolar disorder must be viewed as the prevention of future episodes, with an emphasis on mood stabilization, and not just on the relief of acute symptoms, important as that may be in the short-run. This idea that one needs to take medication in order to stay well, as compared to achieve symptomatic relief, is one that is difficult for children

and adolescents to grasp, and therefore needs to be an important target of ongoing management.

Acute management of mania

Emergent mania represents a true emergency, as it can result in risky behaviour with irreversible consequences. One key to the control of mania is to restore sleep, since sleep deprivation increases mania in a vicious cycle. Most commonly, for the acute control of mania, practitioners use atypical neuroleptics such as risperdone, olanzapine, and quetiapine, which have been shown to be more efficacious than placebo in reducing manic symptoms.[82] Quetiapine has also been shown, in one study, to be superior to divalproex in reducing manic symptoms and achieving both response and remission, whereas lithium and divalproex were shown to have similar efficacy. An open trial comparing the efficacy of three mood stabilizers for achieving clinical response in paediatric bipolar disorder found that divalproex was somewhat more efficacious (ES = 0.58) compared to either lithium or carbamazepine (ESs = 0.38). Oxcarbazepine, a metabolite of carbamazepine, on the other hand, was found to be no better than placebo in achieving stabilization. Both for the management of psychotic symptoms and mania, the combination of a mood stabilizer such as lithium or divalproex and a neuroleptic appears to be more efficacious in producing remission than a mood stabilizer alone.[82]

Acute management of depression

For the treatment-naïve patient, the first step in the management of the depressed bipolar patient is treatment with a mood stabilizer.[5,82] Open trials show that divalproex and lithium are relatively efficacious as mood stabilizers, with carbamazepine being efficacious, but less so than either of the former two agents. In adults, atypical neuroleptics are also being established as mood stabilizers with potency equal to lithium and divalproex, although with very concerning side effects of rapid weight gain. Once a therapeutic blood level of divalproex or lithium has been attained patients often will experience a relief of their depression. Sometimes alteration in the dosage, either an increase or a decrease, can bring further symptomatic relief. If a patient is still experiencing significant depressive symptoms, then an antidepressant can be added, but very carefully. Some data from adult studies suggests that bupropion may result in fewer manic break throughs than other antidepressants but those data do not exist yet in children. Lamotrigine has been shown in adults to provide prophylaxis against future depressive episodes, but is not helpful for the treatment of acute depression; this has not yet been investigated in children, except in open trials.[82]

Medical management

The medications used for mood stabilization all have systemic and potentially very serious side effects. The atypical neuroleptics cause rapid weight gain; therefore weight, a lipid profile, fasting blood sugar, and waistline should be carefully monitored. One preliminary study suggests that concomitant treatment with metformin may attenuate weight gain.[83] Hypoprolactinaemia and galactorrhoea are also consequences of neuroleptic use. Lithium is associated with thyroid disease (usually hypothyroidism), which if undetected can affect mood and treatment reponse. Lithium also can impair kidney function, resulting in an inability to concentrate urine, reduced glomerular fitration rate, and proteinuria. Therefore, renal function should be assessed prior to treatment and annually with a creatinine clearance and a 24 h urine for protein should be obtained at baseline and annually. Divalproex can have toxic effects on the hemopoetic system and the liver, both of which must be carefully monitored. Lamotrigine is rarely associated (0.5 per cent) with Stevens–Johnson syndrome, a disease of mucous membranes that can be potentially life-threatening.

Psychotherapeutic management

Patient and parent education and support are essential, including the importance of medication adherence, keeping regular sleep habits, avoidance of caffeine and other substances, and ability to recognize subtle signs of a shift in mood. Some specific forms of psychotherapy that target family process and emotion regulation have been developed.[84] Family focused treatment (FFT) has been used successful for adult bipolar disorder and results in fewer depressive episodes and better overall functioning. Pilot studies indicate what appear to be similar effects for adolescent bipolar disorder. A family psychoeducational approach also shows promise for improving adherence and reducing the risk of relapse.[85] Two other approaches that have been piloted but not yet tested in randomized clinical trials are the application of CBT to paediatric bipolar disorder, which includes family education, emotion regulation, self-monitoring, and social skills training, and the adaptation of dialectic behaviour therapy (DBT) to adolescent bipolar disorder.[86,87]

Future directions

With regard to both conditions, it is important to try to understand neural circuitry and identify potential intermediate phenotypes, such as emotion dysregulation or impaired executive functioning, which are trait markers for risk for these disorders. The identification of intermediate phenotypes then opens up great opportunities for monitoring treatment response, identifying youth at risk for the disorder, and conducting genetic studies on less complex phenotypes that are more likely to yield definitive results.

While there have been great strides in the treatment of depression, it is difficult to predict who is going to respond to what treatment, and also, who is most likely to experience side effects such as increased suicidal ideation. Pharmacogenetics and monitoring of biomarkers may hold promise for improving matching of patient to treatment. The best approaches to continuation and maintenance have only been addressed for the simplest cases (e.g. uncomplicated episode with patient having a successful response to fluoxetine.[88]

Almost every aspect of treatment in paediatric bipolar disorder requires further study, including the best approach to the management of mania, depression, and mixed state, testing the role of various promising psychosocial approaches, and identifying pharmacogenetic and biomarker predictors of treatment response.

While major depression has been described in high-risk children as young as age 3,[14] it is less clear what are the earliest manifestations of paediatric bipolar disorder. Ongoing longitudinal and high-risk studies will help to clarify the answer to this question in coming years, and may then provide the basis for preventive interventions.

References

1. APA. (2000). *Diagnostic and statistical manual of mental disorders* (*DSM-IV-TR*) (4th edn). American Psychiatric Association, Washington, DC.
2. Kovacs, M. (1996). Presentation and course of major depressive disorder during childhood and later years of the life span. *Journal of the American Academy of Child and Adolescent Psychiatry*, **35**, 705–15.
3. Fergusson, D.M., Horwood, L.J., Ridder, E.M., *et al.* (2005). Subthreshold depression in adolescence and mental health outcomes in adulthood. *Archives of General Psychiatry*, **62**, 66–72.
4. Birmaher, B. and Axelson, D. (2006). Course and outcome of bipolar spectrum disorder in children and adolescents: a review of the existing literature. *Development and Psychopathology*, **18**, 1023–35.
5. Pavuluri, M.N., Birmaher, B., and Naylor, M.W. (2006). Pediatric bipolar disorder: a review of the past 10 years. *Journal of the American Academy of Child and Adolescent Psychiatry*, **44**, 846–71.
6. Kovacs, M. and Devlin, B. (1998). Internalizing disorders in childhood. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **39**, 47–63.
7. Becker, D.F., Grilo, C.M., Edell, W.S., *et al.* (2002). Diagnostic efficiency of borderline personality disorder criteria in hospitalized adolescents: comparison with hospitalized adults. *The American Journal of Psychiatry*, **159**, 2042–7.
8. Angold, A., Costello, E.J., and Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **40**, 57–87.
9. Birmaher, B., Dahl, R.E., Perel, J., *et al.* (1996). Corticotropin-releasing hormone challenge in prepubertal major depression. *Biological Psychiatry*, **39**, 267–77.
10. Lewinsohn, P.M., Rohde, P., and Seeley, J.R. (1998). Major depressive disorder in older adolescents: prevalence, risk factors, and clinical implications. *Clinical Psychology Review*, **18**, 765–94.
11. Lewinsohn, P.M., Klein, D.N., and Seeley, J.R. (1995). Bipolar disorders in a community sample of older adolescents: prevalence, phenomenology, comorbidity and course. *Journal of the American Academy of Child and Adolescent Psychiatry*, **34**, 454–63.
12. Angold, A., Costello, E.J., Erkanli, A., *et al.* (1999). Pubertal changes in hormone levels and depression in girls. *Psychological Medicine*, **29**, 1043–53.
13. Harrington, R., Rutter, M., and Fombonne, E. (1996). Developmental pathways in depression: multiple meanings, antecedents, and endpoints. *Development and Psychopathology*, **8**, 601–16.
14. Luby, J.L., Heffelfinger, A.K., Mrakotsky, C., *et al.* (2003). The clinical picture of depression in preschool children. *Journal of the American Academy of Child and Adolescent Psychiatry*, **42**, 340–8.
15. Birmaher, B., Arbelaez, C., and Brent, D. (2002). Course and outcome of child and adolescent major depressive disorder. *Child and Adolescent Psychiatric Clinics of North America*, **11**, 619–37.
16. Brent, D.A. and Birmaher, B. (2006). Treatment resistant depression in adolescents: recognition and management. *Child and Adolescent Psychiatric Clinics of North America*, **15**, 1015–34.
17. Potash, J.B. (2007). Carving chaos: genetics and the classification of mood and psychotic syndromes. *Harvard Review of Psychiatry*, **14**, 47–63.
18. Geller, B., Zimerman, B., Williams, M., *et al.* (2001). Bipolar disorder at prospective follow-up of adults who had prepubertal major depression disorder. *The American Journal of Psychiatry*, **158**, 125–7.
19. Martin, A., Young, C., Leckman, J.F., *et al.* (2004). Age effects on antidepressant-induced manic conversion. *Archives of Pediatrics & Adolescent Medicine*, **158**, 773–80.
20. Scheffer, R.E., Kowatch, R.A., Carmody, T., *et al.* (2005). Randomized, placebo-controlled trial of mixed amphetamine salts for symptoms of comorbid ADHD in pediatric bipolar disorder after mood stabilization with divalproex sodium. *The American Journal of Psychiatry*, **162**, 58–64.
21. Bridge, J.A., Goldstein, T.R., and Brent, D.A. (2006). Adolescent suicide and suicidal behavior. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **47**, 372–94.
22. Pine, D.S., Goldstein, R.B., Wolk, S., *et al.* (2001). The association between childhood depression and adulthood body mass index. *Pediatrics*, **107**, 1049–56.
23. Craddock, N. and Forty, L. (2006). Genetics of affective (mood) disorders. *European Journal of Human Genetics*, **14**, 660–8.
24. Silberg, J., Rutter, M., Meyer, J., *et al.* (1996). Genetic and environmental influences on the covariation between hyperactivity and conduct disturbance in juvenile twins. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **37**, 803–16.
25. Weissman, M.M., Wickramaratne, P., Nomura, Y., *et al.* (2005). Families at high and low risk for depression: a 3-generation study. *Archives of General Psychiatry*, **62**, 29–36.
26. Caspi, A., Sugden, K., Moffitt, T.E., *et al.* (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HT gene. *Science*, **301**, 386–9.
27. Kaufman, J., Yang, B.Z., Douglas-Palumberi, H., *et al.* (2006). Brain-derived neurotrophic factor-5-HTTLPR gene interactions and environmental modifiers of depression in children. *Biological Psychiatry*, **59**, 673–80.
28. Strauss, J., Barr, C.L., George, C.J., *et al.* (2005). Brain-derived neurotrophic factor variants are associated with childhood-onset mood disorder: confirmation in a Hungarian sample. *Molecular Psychiatry*, **10**, 861–7.
29. McQueen, M.B., Devlin, B., Faraone, S.V., *et al.* (2005). Combined analysis from eleven linkage studies of bipolar disorder provides strong evidence of susceptibility loci on chromosomes 6q and 8q. *American Journal of Human Genetics*, **77**, 582–95.
30. Geller, B., Badner, J.A., Tillman, R., *et al.* (2004). Linkage disequilibrium of the brain-derived neurotrophic factor Val66Met polymorphism in children with a prepubertal and early adolescent bipolar disorder phenotype. *The American Journal of Psychiatry*, **161**, 1698–700.
31. Todd, R.D. and Botteron, K.N. (2001). Family, genetic, and imaging studies of early-onset depression. *Child Adolescent Psychiatric Clinics of North America*, **10**, 375–90.
32. Dickstein, D.P. and Leibenluft, E. (2006). Emotion regulation in children and adolescents: boundaries between normalcy and bipolar disorder. *Development and Psychopathology*, **18**, 1105–31.
33. Adler, C.M., Adams, J., DelBello, M.P., *et al.* (2006). Evidence of white matter pathology in bipolar disorder adolescents experiencing their first episode of mania: a diffusion tensor imaging study. *The American Journal of Psychiatry*, **163**, 322–4.
34. Sauri, R.B., Stanley, J.A., Axelson, D., *et al.* (2005). Reduced NAA levels in the dorsolateral prefrontal cortex of young bipolar patients. *The American Journal of Psychiatry*, **162**, 2109–15.
35. Kyte, Z. and Goodyer, I. (2005). The neurobiology of social cognition and its relationship to unipolar depression. In *The cognitive neuroscience of social behaviour* (eds. A. Easton and N.J. Emery), pp. 257–90. Psychology Press, New York.
36. Forbes, E.E., May, C., Siegle, G.J., *et al.* (2006). Reward-related decision-making in pediatric major depressive disorder: an fMRI study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **47**, 1031–40.
37. Roberson-Nay, R., McClure, E.B., Monk, C.S., *et al.* (2006). Increased amygdala activity during successful memory encoding in adolescent major depressive disorder: an fMRI Study. *Biological Psychiatry*, **60**, 966–73.
38. Thomas, K.M., Drevets, W.C., Whalen, P.J., *et al.* (2001). Amygdala response to facial expressions in children and adults. *Biological Psychiatry*, **49**, 309–16.
39. Silk, J.S., Shaw, D.S., Skuban, E.M., *et al.* (2006). Emotion regulation strategies in offspring of childhood-onset depressed mothers. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **47**, 69–78.

40. Perez-Edgar, K., Fox, N.A., Cohn, J.F., *et al.* (2006). Behavioral and electrophysiological markers of selective attention in children of parents with a history of depression. *Biological Psychiatry*, **60**, 1131–8.
41. Silk, J.S., Shaw, D.S., Forbes, E.E., *et al.* (2006). Maternal depression and child internalizing: the moderating role of child emotion regulation. *Journal of Clinical Child and Adolescent Psychology*, **35**, 116–26.
42. Garber, J., Keiley, M.K., and Martin, N.C. (2002). Developmental trajectories of adolescents' depressive symptoms: predictors of change. *Journal of Consulting and Clinical Psychology*, **70**, 79–95.
43. Shaw, D.S., Schonberg, M., Sherrill, J., *et al.* (2006). Responsivity to offspring's expression of emotion among childhood-onset depressed mothers. *Journal of Clinical Child and Adolescent Psychology*, **35**, 490–503.
44. Sharp, C., Fonagy, P., and Goodyer, I.M. (2006). Imagining your child's mind: psychosocial adjustment and mothers' ability to predict their children's attributional response styles. *British Journal Developmental Psychology*, **24**, 197–214.
45. Kovacs, M., Sherril, J., George, C.J., *et al.* (2006). Contextual emotion-regulation therapy for childhood depression: description and pilot testing of a new intervention. *Journal of the American Academy of Child and Adolescent Psychiatry*, **45**, 892–903.
46. Pavuluri, M.N., Schenkel, L.S., Aryal, S., *et al.* (2006). Neurocognitive function in unmedicated manic and medicated euthymic pediatric bipolar patients. *The American Journal of Psychiatry*, **163**, 286–93.
47. Gotlib, I.H., Traill, S.K., Montoya, J.J., *et al.* (2005). Attention and memory biases in the offspring of parents with bipolar disorder: indications from a pilot study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **46**, 84–93.
48. Ryan, N.D. (1998). Psychoneuroendocrinology of children and adolescents. *The Psychiatric Clinics of North America*, **21**, 435–41.
49. Forbes, E.E., Williamson, D.E., Ryan, N.D., *et al.* (2006). Peri-sleep-onset cortisol levels in children and adolescents with affective disorders. *Biological Psychiatry*, **59**, 24–30.
50. Bertocci, M.A., Dahl, R.E., Williamson, D.E., *et al.* (2005). Subjective sleep complaints in pediatric depression: a controlled study and comparison with EEG measures of sleep and waking. *Journal of the American Academy of Child and Adolescent Psychiatry*, **44**, 1158–66.
51. Kaufman, J., Birmaher, B., Perel, J., *et al.* (1998). Serotonergic functioning in depressed abused children: clinical and familial correlates. *Biological Psychiatry*, **44**, 973–81.
52. Nomura, Y., Wickramaratne, P.J., Warner, V., *et al.* (2002). Family discord, parental depression, and psychopathology in offspring: ten-year follow-up. *Journal of the American Academy of Child and Adolescent Psychiatry*, **41**, 402–9.
53. Frye, A.A. and Garber, J. (2005). The relations among maternal depression, maternal criticism, and adolescents' externalizing and internalizing symptoms. *Journal of Abnormal Child Psychology*, **33**, 1–11.
54. Leverich, G.S. and Post, R.M. (2006). Course of bipolar illness after history of childhood trauma. *Lancet*, **367**, 1040–2.
55. Brent, D.A., Moritz, G., and Liotus, L. (1996). A test of the diathesis-stress model of adolescent depression in friends and acquaintances of adolescent suicide victims. In *Severe stress and mental disturbance in children* (ed. C.R. Pfeffer), pp. 347–60. American Psychiatric Press, Inc., Washington, DC.
56. Borowsky, I.W., Ireland, M., and Resnick, M.D. (2001). Adolescent suicide attempts: risks and protectors. *Pediatrics*, **107**, 485–93.
57. Myers, K. and Winters, N.C. (2002). Ten-year review of rating scales. II. Scales for internalizing disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, **41**, 634–59.
58. Dierker, L., Albano, A.M., Clarke, G.N., *et al.* (2001). Screening for anxiety and depression in early adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry*, **40**, 929–36.
59. Birmaher, B. and Brent, D. (in press). Work group on quality issues. Practice parameters for the assessment and treatment of children and adolescents with depressive disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*.
60. National Institute for Health and Clinical Excellence (NICE). (2005). *Depression in children and young people. Identification and management in primary, community, and secondary care*. NICE, London.
61. Weisz, J.R., McCarty, C.A., and Valeri, S.M. (2006). Effects of psychotherapy for depression in children and adolescents: a meta-analysis. *Psychological Bulletin*, **132**, 132–49.
62. March, J.S., Silva, S., Petrycki, S., *et al.* (2004). Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression. Treatment for Adolescent Depression Study (TADS) randomized controlled trial. *The Journal of the American Medical Association*, **292**, 807–20.
63. March, J., Silva, S., Vitiello, B., *et al.* (2006). The Treatment for Adolescents with Depression Study (TADS): methods and message at twelve weeks. *Journal of the American Academy of Child and Adolescent Psychiatry*, **45**, 1393–403.
64. Clarke, G., Debar, L., Lynch, F., *et al.* (2005). A randomized effectiveness trial of brief cognitive-behavior therapy for depressed adolescents receiving antidepressant medication. *Journal of the American Academy of Child and Adolescent Psychiatry*, **44**, 888–98.
65. Goodyer, I., Dubicka, B., Wilkinson, P., *et al.* (2007). Selective serotonin reuptake inhibitors (SSRIs) and routine specialist care with and without cognitive behaviour therapy in adolescents with major depression: randomised controlled trial. *British Medical Journal*, **335**, 106–7.
66. Brent, D.A., Emslie, G.J., Clarke, G.N., *et al.* (2007). Treatment of SSRI-Resistant Depression in Adolescents (TORDIA): a test of treatment strategies in depressed adolescents who have not responded to an adequate trial of a Selective Serotonin Reuptake Inhibitor (SSRI). Presented at NCDEU 47th Annual Meeting, Boca Raton, FL.
67. Rohde, P., Lewinsohn, P.M., Clarke, G.N., *et al.* (2005). The adolescent coping with depression course: a cognitive-behavioral approach to the treatment of adolescent depression. In *Psychosocial treatments for child and adolescent disorders: empirically based strategies for clinical practice* (eds. E.D. Hibbs and P.S. Jensen), pp. 219–37. Washington, DC.
68. Clarke, G.N., Hornbrook, M., Lynch, F., *et al.* (2001). A randomized trial of a group cognitive intervention for preventing depression in adolescent offspring of depressed parents. *Archives of General Psychiatry*, **58**, 1127–34.
69. Weissman, M.M., Pilowsky, D.J., Wickramaratne, P.J., *et al.* (2006). Remissions in maternal depression and child psychopathology: a STAR*D child report. *The Journal of the American Medical Association*, **295**, 1389–98.
70. Beardslee, W., Gladstone, T.R.G., Wright, E.J., *et al.* (2003). A family based approached to the prevention of depressive symptoms in children at risk: evidence of parental and child change. *Pediatrics*, **112**, e119–e31.
71. Bridge, J., Iyengar, S., Salary, C.B., *et al.* (2007). Clinical response and risk for reported suicidal ideation and suicide attempts in pediatric antidepressant treatment: a meta-analysis of randomized controlled trials. *The Journal of the American Medical Association*, **297**, 1683–96.
72. Hazell, P., O'Connell, D., Heathcote, D., *et al.* (1995). Efficacy of tricyclic drugs in treating child and adolescent depression: a meta-analysis. *British Medical Journal*, **310**, 897–901.
73. Sallee, F.R., Vrindavanam, N.S., Deas-Nesmith, D., *et al.* (1997). Pulse intravenous clomipramine for depressed adolescents: double-blind, controlled trial. *The American Journal of Psychiatry*, **154**, 668–73.
74. Daviss, W.B., Perel, J.M., Brent, D.A., *et al.* (2006). Acute antidepressant response and plasma levels of bupropion and metabolites in a

pediatric-aged sample: an exploratory study. *Therapeutic Drug Monitoring*, **28**, 190–8.
75. Nemets, H., Nemets, B., Apter, A., *et al.* (2006). Omega-3 treatment of childhood depression: a controlled, double-blind pilot study. *The American Journal of Psychiatry*, **163**, 1098–100.
76. Findling, R.L., McNamara, N.K., Stansbrey, R.J., *et al.* (2006). The relevance of pharmacokinetic studies in designing efficacy trials in juvenile major depression. *Journal of Child and Adolescent Psychopharmacology*, **16**, 131–45.
77. Kronenberg, S., Apter, A., Brent, D., *et al.* (in press). Serotonin transporter (5HTT) polymorphism and citalopram effectiveness and side effects in children with depression and anxiety disorders. *Journal of Child and Adolescent Psychopharmacology*.
78. Hammad, T.A., Laughren, T., and Racoosin, J. (2006). Suicidality in pediatric patients treated with antidepressant drugs. *Archives of General Psychiatry*, **63**, 332–9.
79. Hughes, C.W., Emslie, G.J., Crismon, L., *et al.* (1999). The Texas children's medication algorithm project: report of the Texas consensus conference panel on medication treatment of childhood major depressive disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, **38**, 1442–54.
80. Brent, D. and Birmaher, B. (2006). Treatment resistant depression in adolescents: recognition and management. *Child and Adolescent Psychiatric Clinics of North America*, **15**, 1015–34.
81. Swedo, S.E., Allen, A.J., Glod, C.A., *et al.* (1997). A controlled trial of light therapy for the treatment of pediatric seasonal affective disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, **36**, 816–21.
82. McClellan, J., Kowatch, R., and Findling, R.L. (2007). Work group on quality issues. Practice parameter for the assessment and treatment of children and adolescents with bipolar disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, **46**, 107–25.
83. Klein, D.J., Cottingham, E.M., Sorter, M., *et al.* (2006). A randomized, double-blind, placebo-controlled trial of metformin treatment of weight gain associated with initiation of atypical antipsychotic therapy in children and adolescents. *The American Journal of Psychiatry*, **163**, 2072–9.
84. Miklowitz, D.J. (2006). A review of evidenced-based psychosocial interventions for bipolar disorder. *The Journal of Clinical Psychiatry*, **67**, 28–33.
85. Fristad, M.A. (2006). Psychoeducational treatment for school-aged children with bipolar disorder. *Development and Psychopathology*, **18**, 1289–306.
86. Pavuluri, M.N., Graczyk, P., Henry, D., *et al.* (2004). Child- and family-focused cognitive behavioral therapy for pediatric bipolar disorder: development and preliminary results. *Journal of the American Academy of Child and Adolescent Psychiatry*, **43**, 528–37.
87. Goldstein, T.R., Axelson, D.A., Birmaher, B., *et al.* (2007). Dialectical behavior therapy for adolescent with bipolar disorder: a one year open trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, **46**, 820–30.
88. Emslie, G.J., Heiligenstein, J.H., Hoog, S.L., *et al.* (2004). Fluoxetine treatment for prevention of relapse of depression in children and adolescents: a double-blind, placebo-controlled study. *Journal of the American Academy of Child and Adolescent Psychiatry*, **43**, 1397–405.

9.2.8 Obsessive–compulsive disorder and tics in children and adolescents

Martine F. Flament and Philippe Robaey

Introduction

Although obsessive–compulsive disorder (**OCD**) has long been considered as a disorder of adulthood, the early child psychiatric literature contains famous descriptions of typical cases. At the beginning of the twentieth century, Janet reported on a 5-year-old with classical obsessive–compulsive (**OC**) symptoms, and Freud described in his adult patients obsessional behaviours dating back from childhood, while speculating on the strong constitutional influence in the choice of these symptoms. In the 1950's, Kanner noted the resemblance and sometimes the association between compulsive movements and tics, and Despert described the first large series of obsessive–compulsive children, noting the preponderance of males and the children's perception of the abnormality and undesirability of their behaviours.

Tics have been described since antiquity, but the first systematic reports are those of Itard, in 1825, and Gilles de la Tourette, in 1885. Both noted the association between tic disorders and OC symptoms, and speculated on the hereditary nature of the syndrome.

For the last two decades, there has been a tremendous growth of interest and research on OCD and tic disorders. Significant advances have occurred regarding the phenomenology, epidemiology, genetics, neurophysiology, pathogenesis, and treatment of both disorders. The frequent association of OCD and/or tic disorders with other neuropsychiatric disorders, as well as the increasing evidence coming from in-vivo neuroimaging studies, have led to a fascinating aspect of current neurobiological research—the possible localization of brain circuits mediating the abnormal behaviours. Of all paediatric psychiatric disorders, OCD and tic disorders now appear as model neurobiological disorders to investigate the role of genetic, neurobiological, and environmental mechanisms that interact to produce clinical syndromes of varying severity.

Clinical features

Obsessive–compulsive disorder

Obsessions are persistently recurring thoughts, impulses, or images that are experienced as intrusive, inappropriate, and distressing, and that are not simply excessive worries about realistic problems. Compulsions are repetitive behaviours or mental acts that a person feels driven to perform according to a rigidly applied rule, in order to reduce distress or to prevent some dreaded outcome. Obsessions and compulsions are egodystonic, i.e. there are considered by the subject himself as irrational or unrealistic, and are, at least partly, resisted. Children and adolescents with OCD may hide their symptoms, or will only allow them to appear at home, or in the presence of family members, suggesting partial voluntary control.

The clinical presentation of OCD during childhood and adolescence has been documented in various cultures, with clinical series

reported from the U.S., Japan, India, Israel, Denmark, and Spain. Typically, children and adolescents with OCD experience multiple obsessions and compulsions, whose content may change over time. The most frequent obsessions in young people include fear of dirt or germs, of danger to self or a loved one, symmetry or exactness, somatic, religious and sexual obsessions. The most common compulsions consist of washing rituals, repeating, checking, touching, counting, ordering, and hoarding. Generally, compulsions are carried out to dispel anxiety and/or in response to an obsession (e.g. to ward off fear of harm). However, some obsessions and rituals involve an internal sense that 'it does not feel right' until the thought or action is completed, and certain children with OCD may be unable to specify the dreaded event that the compulsive rituals are intended to prevent, beyond a vague premonition of something bad happening. Simple compulsions, such as repetitive touching or symmetrical ordering, may even lack any discernable ideational component, and may be phenomenologically indistinguishable from complex tics. Several symptom dimensions have been identified in OCD, which could suggest possible aetiologic heterogeneity. Based on the symptom categories of the Children's Yale-Brown Obsessive–Compulsive Scale (CY-BOCS, the most widely used symptomatic measure in paediatric OCD research,[1]) Stewart *et al.*[2] identified four distinct factors, using principal components analysis: (1) symmetry/ordering/repeating/checking; (2) contamination/cleaning/aggressive/somatic; (3) hoarding; and (4) sexual/religious symptoms. These symptom dimensions are congruent with those described in similar studies of adults with OCD, suggesting fairly consistent covariation of OCD symptoms through the developmental course.

Tics

Tics are sudden, rapid, non-rhythmic, stereotyped, repetitive movements (motor tics) or sounds (vocal tics). They may mimic simple or more complicated fragments of normal motor or vocal behaviours, which are misplaced in context. Tics vary greatly in nature, location, number, intensity, forcefulness, and frequency.[4,5] Common simple motor tics are neck jerking, eye blinking, elevation of shoulders, mouth movements. Common simple vocal tics include throat clearing, sniffing, sucking air, grunting, snorting, humming, or barking. Complex motor tics may combine simple tics, and involve facial movements, jumping, gyrating, touching, kicking, grooming behaviours, or echokinesis (repeating someone else's movement). They may appear to be purposive in character, as brushing hair back, or suddenly rotating on one foot to make a 360-degree turn. In a small fraction of cases, the complex motor tics are self-injuring behaviours, which may be potentially dangerous. Complex verbal tics include the repetition of what was just heard (echolalia) or said (palilalia), or socially inappropriate utterances (coprolalia), even disguised through sign language. The specific tic repertoire of an individual typically changes over time with no predictable course, but complex tics are rare in the absence of simple tics. Tics often occur in discrete unpredictable bouts over the course of a day, separated by tic-free intervals. The combination of bouts over different time scales explains why globally tics wax and wane over time.

Tics are suggestible, as indicated by their transient reappearance when they are recalled. They are also suppressible, as they can generally be willfully held back for brief periods of time. Tics are preceded by an inner tension, an urge to move or utter that may build-up during suppression. Suppressing tics requires mental effort and may accentuate inattention; conversely, attentional problems decrease the ability to suppress tics, and are associated with more severe tics. Various premonitory sensory urges have been reported to prompt the tics, together with feelings of inner conflicts over whether and when to yield to these urges. Sensory urges include focal tension, pressure, tickling, cold, warmth, paresthesias, and generalized inner tension or anxiety. They usually arise in the part of the body involved in the subsequent motor act, and completing the tic seems to yield a temporary relief of the urge. Also, various auditory and visual cues, highly selective for each individual, can elicit tics, and some patients are extremely sensitive to these external cues, as in echo phenomena. Excitement and fatigue typically worsen tics, which are often more frequent and forceful when the individual is alone. Activities requiring fine motor skills and attention improve tics. Although much diminished, tics can occur during sleep, unlike many other movement disorders.

Children and adolescents with tic disorders may present a broad array of associated behavioural difficulties, including OC symptoms, disinhibited speech or conduct, impulsivity, distractibility, and motor hyperactivity.[5] The presence of motor and/or phonic tics can be associated with difficulties in self-esteem, self-definition, family life, peer acceptance or relationships, and school performance.

Age of onset

The age at onset of OCD appears bimodal.[6] Prepubertal onset is associated with a male preponderance and an increased risk for tic disorders, including Tourette's disorder. A second peak of onset occurs at or after puberty. Overall, the mean age at onset of OCD in children and adolescents have ranged from 9 years in referred subjects[7] to 12.8 years in a community sample.[8]

The median onset age for simple motor tics is between 4 and 6 years.[9] Phonic or vocal tics usually appear several years after the onset of motor tics, in most cases between 8 and 15 years. Many young children are completely oblivious of their tics, or experience them as wholly involuntary. Premonitory urges typically show up several years after the onset of the tics, on average around 10 years of age. Suppressibility of tics developmentally precedes awareness of premonitory urges, but may get easier as awareness increases.[10]

Sex ratio

In community-based samples of adolescents with OCD, there are approximately equal numbers of males and females, while in most studies of referred children and youth with OCD, males outnumber females by 2:1 or 3:1.[11]

Most studies show that the prevalence of tics is higher among boys than girls, with a ratio of 6–8:1 in clinic-based samples, and about 2:1 in community-based studies.[12] The sex ratio generally increases with tic duration and severity. Thus, in one study, the ratio of boys to girls was 1.6:1 for motor tics present for 1-2 consecutive months, increasing to 7.5:1 when tics were present for 2 non consecutive months, or more than 3 months.[13]

Comorbidity

In referred children and adolescents with OCD, the frequency of a diagnosis of any tic disorder ranges from 17 per cent to 40 per cent, and that of Tourette's disorder from 11 per cent to 15 per cent.[14]

Conversely, one study found that 29 per cent of Tourette's disorder patients displayed OC behaviours.[15] In longitudinal studies, about 50 per cent of children and adolescents with Tourette's disorder develop OC symptoms or OCD by adulthood,[16] whereas, in a follow-up study of children and adolescents initially treated for OCD, nearly 60 per cent were found to have a lifetime history of tics that ranged from simple, mild, and transient tics to Tourette's disorder, for which the rate was 11 per cent.[17] On the basis of personal or family history of tics, a distinction has been proposed between 'tic-related OCD' and 'non-tic-related OCD', under the assumption that the two forms might differ in terms of clinical phenomenology, neurobiological concomitants, and responsiveness to pharmacological interventions.[18] Tic-related OCD appears to have an earlier onset, and to occur more frequently in boys than in girls. The need to touch or rub, blinking and staring rituals, worries over symmetry and exactness, a sense of incompleteness, and intrusive aggressive thoughts and images, are significantly more common in tic-related OCD, whereas contamination worries and cleaning compulsions are more frequent in patients with non-tic-related OCD.

The overall lifetime psychiatric comorbidity in children and adolescents with OCD is about 75 per cent, both in referred and in community cases. The most common conditions comorbid with OCD are affective disorders, with prevalence ranging across studies from 8 per cent to 73 per cent for mood disorders, and from 13 per cent to 70 per cent for anxiety disorders.[19] While occurring less frequently in non-referred subjects, a high rate of disruptive behaviour disorders—attention deficit/hyperactivity disorder (ADHD) and oppositional defiant disorder—has been reported in subjects seen in paediatric OCD clinics. In girls, OCD can be comorbid to anorexia nervosa.

Less than 10 per cent of clinically referred children and adolescents with Tourette's disorder do not have another morbid condition. About 55 per cent also have ADHD, and more than a third have anger control problems.[9] Rage attacks in response to minimal provocation, lasting from a few minutes to an hour and usually followed by remorse, as well as an increased vulnerability for drug abuse, depression, and antisocial behaviour, are primarily observed when comorbid ADHD is present. Globally, comorbidity of Tourette's disorder increases in adolescence, but more markedly for OCD and anxiety disorders in those without ADHD. Individuals with Tourette's disorder have consistently shown difficulties with fine motor control and visual motor integration, as well as impairment in procedural or habit-based learning. Sleep is often disturbed, with increased short lasting motor activity, especially in non-REM sleep, compared to healthy controls.

Classification

Both DSM-IV and ICD-10 define OCD, regardless of age, by obsessions and/or compulsions, which are described, at some point during the course of the disorder, as excessive or unreasonable (criterion B), and are severe enough to cause marked distress or to interfere significantly with the person's normal routine, or usual social activities or relationships. The specific content of the obsessions or compulsions cannot be restricted to another Axis I diagnosis, such as an eating disorder, a mood disorder, or schizophrenia. The DSM-IV adds that the disturbance is not due to the direct physiological effects of a substance or a general medical condition. The ICD-10 allows subclassification of forms with predominant obsessions, predominant compulsions, or mixed symptoms. In DSM-IV, the only difference in diagnostic criteria between children and adults appears in criterion B; although most children and adolescents actually acknowledge the senselessness of their symptoms, the requirement that insight is preserved is waived for children.

In both DSM-IV and ICD-10, tic disorders are divided into four categories, according to duration of the symptoms, and presence of vocal tics in addition to motor tics: Tourette's disorder, chronic motor or vocal tic disorder, transient tic disorder, and tic disorder not otherwise specified (NOS). Transient tic disorder is defined by single or multiple motor and/or vocal tics that occur many times a day, nearly everyday for at least 4 weeks, but for no longer than 12 consecutive months. In chronic motor or vocal tic disorder, either motor or vocal tics, but not both, have been present at some time during the illness. In Tourette's disorder, both multiple motor and one or more vocal tic have to be present, although not necessarily concurrently. Both Tourette's disorder and chronic motor or vocal tic disorder have a duration of more than 1 year, with no tic-free period of more than 3 months. All tic disorders must have onset before age 18 years. In all, the disturbance causes marked distress or significant impairment in social, occupational, or other important areas of functioning, and is not due to the direct physiological effects of a substance (e.g. stimulants), or a general medical condition. The ICD-10 recognizes that there is an immense variation in the severity of tics. At the one extreme of the continuum, the presence of transient tics, at some time during childhood, is near-normal. At the other extreme, Tourette's disorder is an uncommon, chronic, and incapacitating disorder.

Diagnosis and differential diagnosis

In ICD-10, it is stated that OCD cannot be diagnosed if the patient meets Tourette's disorder criteria, while both diagnoses may be given simultaneously in DSM-IV. Unlike tics, compulsions are aimed at neutralizing the anxiety resulting from an obsession, and/or they are performed according to rules that must be applied rigidly. However, both compulsive rituals and complex tics may be preceded by premonitory urges, which persist until the action is completed. In individuals with both Tourette's disorder and OCD, these symptoms are sometimes so closely intertwined that efforts to distinguish them would be futile.

From a developmental perspective, pathological OC behaviours and thoughts differ from normal childhood rituals, mainly by their emotional context and their use of maladaptive versus adaptive cognitive and behavioural strategies.[20] Developmental childhood rituals are part of learning new skills, and accompanied by expressions of positive affect and interest. They are most intense in 4- to 8-year-olds, stress rules about daily life, help the child master anxiety, and enhance the socializing process. In contrast, perseverative behaviours in OCD are not goal-oriented, they are accompanied by a burdened, anxious affect, provoke frustration, are incapacitating and painful, and promote social isolation and regressive behaviour.

OCD must be distinguished from other anxiety disorders and, in some cases, from autism or schizophrenia. In phobias, subjects are preoccupied by their fears only when confronted to the phobogen stimuli, and, in separation anxiety disorder, fear of harm to

parents or loved ones are part of persistent worries and behaviours which are not criticized by the child. Stereotyped movements and ritualistic behaviours are frequent in intellectual disability and autism, but they convey no particular intentionality, and the child does not try to resist them. In schizophrenia, there are erroneous belief systems in several areas, but the subject does not criticize them and does not consider the subsequent behaviours to be abnormal.

Tics should be differentiated from other types of abnormal movements which can occur in numerous congenital or acquired neurological and neuropsychiatric disorders (Sydenham's chorea, encephalitis, Huntington disease, tuberous sclerosis, neuroacanthocytosis, Wilson's disease, head trauma, mental retardation, autism). The term of secondary tics or Tourettism has been applied to these disorders, and the abnormal movements can be choreiform movements, dystonic movements, myoclonic movements, spasms, or stereotypies. Some medications such as central nervous system stimulants (methylphenidate, amphetamine, pemoline, cocaine), antihistaminic and anticholinergic drugs, antiepileptics (carbamazepine, phenytoin), antipsychotics, and opioids may also produce or exacerbate tics.[21] The distinction between tic disorders and other disorders with abnormal movements is based on anamnesis, family history, observation, and neurological examination, which is usually normal in tic disorders. Specific diagnostic tests may be required to confirm neurological or exogenous causes.

Epidemiology

Tics might be one of the most common behavioural problems in childhood, but estimates of the prevalence of tic disorders greatly vary because of differences in the methods used (e.g. parental report versus direct observation), differences in the populations surveyed (e.g. age and sex distribution), and the transient nature of tics. Surveys among school-age children indicate that up to 18 per cent of boys, and 11 per cent of girls manifest frequent 'tics, twitches, mannerisms or habit spasms'. Race and socio-economic status do not seem to influence the frequency of tics. There are virtually no general population studies of transient tic disorder or chronic motor or vocal tic disorder. For Tourette's disorder, most population-based surveys yield prevalence estimates in the range of 5–10 per 10,000, with children being more likely to be identified than adults, and males more than females.[22, 13] In a study amongst all inductees into the Israeli Defence Force over 1 year, the point prevalence of Tourette's disorder was 4.9 per 10 000 males and 3.1 per 10 000 females, and the prevalence of OCD was elevated in those with Tourette's disorder (41.7 per cent vs. 3.4 per cent in others).[23] One longitudinal study assessed the presence of tics and OCD in an epidemiological sample of individuals followed from childhood to adulthood.[24] The prevalence of tics was 17.7 per cent at age 1–10 years, decreasing to 2–3 per cent in adolescence; childhood tics were associated with increased rates of OCD in adolescence; in adolescents with tics, the presence of comorbid OCD predicted persistence of tics into early adulthood.

There has been only one survey on the prevalence of OCD in children (5- to 15-year old), indicating an overall prevalence of OCD at 0.25 per cent, with an exponential increase as a function of age, from 0.026 per cent in 5–7 year olds to 0.63 per cent in 13–15 year olds.[25] In adolescents, epidemiological studies using strict diagnostic criteria and structured clinical interviews have been conducted in several parts of the world, estimating the prevalence of juvenile OCD between 1 and 4 per cent. In the largest study to date (N=5 596 high-school students), the lifetime prevalence of OCD in adolescents was estimated to 1.9 (±0.7) per cent, and none of the identified cases had been previously diagnosed.[8] In a later study,[26] the point prevalence of OCD was 3.6 (±0.7) per cent, decreasing to 1.8 per cent when excluding those individuals with only obsessions; among the OCD cases, there was a significant elevation of tic disorders (Tourette's disorder 5 per cent, chronic multiple tics 10 per cent, transient tics 10 per cent). In two longitudinal studies following cohorts of children in the community up to the age of 18 years, the prevalence for OCD ranged from 1.2 to 4 per cent.[27, 28] Thus, it appears that OCD might be as frequent in adolescents as it is in adults (see Chapter 4.8).

Aetiology

Psychological factors

Psychological theories of OCD have encompassed psychoanalytic as well as more general non-psychodynamic etiological approaches, focusing alternatively on volitional, intellectual, and/or emotional impairment. Freud's famous patient, the Rat Man, has been seen as a paradigm of a psychologically determined illness, illustrating the central role of anal sadistic concerns with control, ambivalence, magical thinking, and the salience of defenses such as reaction formation, intellectualization, isolation, and undoing. Freud also provided fascinating speculations on the similarity between OC phenomena, children's games, and religious rites. Later, Anna Freud stated that 'obsessional outcomes are promoted by a constitutional increase in the intensity of the anal-sadistic tendencies probably as the result of inheritance combined with parental handling'. However, despite the beautifully described dynamics of obsessional symptoms, most illustrative of unconscious processes, the psychoanalysts have also pointed out the extreme difficulty in treating OCD with classical analytic treatment.

Even though psychological factors are insufficient to cause Tourette's disorder, tic behaviours have long been identified as stress-sensitive, and temporally associated with important events in the lives of children. In a prospective study over 2 years, children and adolescents with Tourette's disorder and/or OCD experienced significantly more psychosocial stressors than did healthy controls, and the level of psychosocial stress was a significant predictor of future tic and OC symptoms severity.[29]

Biochemical factors

Although a variety of biological aetiologies have been proposed in OCD since the 19th century, modern neurobiological theories began with the clinical studies showing that clomipramine and other serotonin reuptake inhibitors (**SRIs**) had a unique efficacy in treating the disorder. This inspired a 'serotoninergic hypothesis' of OCD (see Chapter 4.8). In children, the involvement of the serotonin system in the pathophysiology of OCD is supported by one study in which improvement of OC symptoms during clomipramine treatment was closely correlated with pretreatment platelet serotonin concentration,[30] and reports of decreased density of the platelet serotonin transporter in children and adolescents with OCD but not in those with Tourette's disorder.[31] However, the

delayed and incomplete action of serotonergic drugs, suggesting multiple effects on other neurotransmitters as well, and numerous biochemical studies of OCD patients and controls have not yet indicated a single biochemical abnormality as a primary etiological mechanism in OCD.

In Tourette's disorder, multiple neurochemical systems have been implicated by pharmacological and metabolic studies, but a primary disturbance in the dopaminergic system is supported by the tic suppressing effect of dopamine receptor antagonists (see below). Post-mortem studies have shown an increase in the number of presynaptic dopamine transporter sites in the striatum and the frontal cortex of individuals with Tourette's disorder. PET/SPECT studies have demonstrated greater binding to dopamine transporter sites in both the caudate and putamen nuclei,[32] increased dopamine release by psychostimulants in the putamen,[33] and an association between density of dopamine receptors in the caudate and severity of tics.[34] The 'tonic-phasic hypothesis' proposes both a hyperresponsive spike-dependent (phasic) dopaminergic system (possibly related to an alteration in afferent cortical inputs), and a reduction in tonic dopamine levels (possibly secondary to an overactive dopamine transporter system), that would upregulate pre- and postsynaptic dopamine receptors and further increase the phasic-tonic unbalance. There is also some evidence for the role of serotonin in tic disorders, notably a study showing that reduced serotonin transporter binding correlated with vocal tics and OC symptoms.[35]

Genetic factors

In both OCD and Tourette's disorder, twin and family studies provide strong evidence that genetic factors are involved in the vertical transmission of vulnerability within families. The average concordance rate in monozygotic twins is 65 per cent for OCD,[36] and 53 per cent for Tourette's disorder.[37] Family studies have consistently found higher rates of OCD and tic disorders in probands with paediatric OCD, as well as higher rates of tic disorders and OCD in those with tic disorders. Thus, Lenane *et al.*[38] investigating 147 first-degree relatives of children and adolescents with OCD found that 44 per cent of the families had a positive history of tics in at least one first-, second-, or third-degree relative. Conversely, Pauls *et al.*[39] reported that the prevalence rates of OCD and tic disorders were significantly greater among the first-degree relatives of 100 probands with OCD (10.3 per cent and 4.6 per cent, respectively) than among relatives of psychiatrically unaffected subjects (1.9 per cent and 1.0 per cent). These findings suggested that Tourette's disorder and some forms of OCD could be variant expressions of the same underlying genetic factors.

Results from two genome-wide scans have been reported,[40,41] with the strongest linkage peaks being on chromosome 2 for Tourette's disorder ($p=0.00004$), and chromosome 3 for OCD ($p=0.0002$); there were also regions that showed moderate evidence for linkage to both disorders. In early-onset OCD, family-based evidence for association at several serotonin system genes (SCL6A4, HTR1B, HTR2A) and brain-derived neurotropic factor (BDNF) has been reported in some studies, and the association seemed stronger in subjects with tic disorders associated with OCD.[42] There is also preliminary evidence for an association between OCD and two glutamate genes, the glutamate transporter gene (SLC1A1), and a glutamate receptor gene (GRIN2B).[43] A complementary approach involving examining rare cases of cytogenetic abnormalities co-segregating with Tourette's and related disorders has pointed to regions on chromosomes 3p, 7q, 8q, 9p, and 18q.[44]

Thus, as suggested by earlier family studies, OCD and Tourette's disorder might have both shared and distinct susceptibility genes involved in their etiology. It is likely that epigenetic and non genetic factors may also contribute to phenotypal heterogeneity. A range of prenatal and perinatal events have been suggested as risk factors for increased tic severity, including lower birth weight, in utero exposure to caffeine, alcohol or tobacco, and maternal stress.[45] As the basal ganglia are especially sensitive to hypoxia, it is possible that factors associated with transient hypoxia could increase the risk for Tourette's disorder in those with a genetic vulnerability.

Dysfunction of frontal–subcortical circuits

It has been known for a long time that OC symptoms could be associated with neurological disorders of motor control, including Tourette's disorder, Huntington's disease, Parkinson's disease, as well as traumatic or infectious lesions of the basal ganglia.[46] Conversely, in both adults and children with typical OCD, an increased frequency of soft neurological signs has been reported.[47] Since the era of neuroimaging, numerous studies have consistently found that the ventral prefrontal cortical (VPFC) regions, such as orbital prefrontal cortex and anterior cingulate cortex, the striatum, the basal ganglia and the thalamus were basic brain structures involved in the pathophysiology of OCD. These studies have generally identified abnormally high metabolic activity and/or blood flow in the orbital cortex and the head of the striatal caudate nucleus in untreated OCD subjects at rest, compared to various control populations.[48,49] Furthermore, the same two regions, as well as the thalamus to which each projects, have shown further increase in activity during OC symptom provocation. In several functional neuroimaging studies of patients with childhood onset OCD, measures of VPFC and striatal activity correlated positively with OCD symptom severity and treatment response.[50,51] Some studies have also indicated that the anatomy of the caudate, putamen and globus pallidus could differ between paediatric OCD patients and controls,[52] especially in cases of paediatric autoimmune neuropsychiatric disorders associated with streptococcal infections (PANDAS).[53] Although most studies have implicated the VPFC in the pathogenesis of OCD, recent investigation suggests a role for the dorsolateral prefrontal cortex (DLPFC) as well. Thus, one study found a significant increase in N-acetyl-aspartate (NAA), a neuronal marker of activity, in the left DLPFC of unmedicated paediatric OCD patients compared to controls.[54]

Recent MRI studies found that the volume of the caudate nucleus is decreased in both children and adults with Tourette's disorder, whereas the volume of putamen and globus pallidus nuclei are primarily reduced in adults with the disorder.[55] This is consistent with a study comparing monozygotic twins discordant for tic expression, in which caudate nuclei volumes were smaller in the more severely affected co-twin.[56] In addition, subjects with Tourette's disorder were found to have larger volumes in dorsal prefrontal and parieto-occipital regions.[57] Although no association was found between tic severity and the volumes of the basal ganglia, ratings of worst-ever tic severity were associated with larger orbito-frontal and parieto-occipital regions. In one recent study, cortical and subcortical hyperintensities that are considered as a subclinical manifestation of small-vessel disease, were significantly more abundant in children and adolescents with Tourette's

disorder, OCD or ADHD than in healthy controls.[58] These results support a primary disturbance of the cortico-striato-pallidal-thalamo-cortical circuit, especially the projection into or out of the striatum. The small reduction of the caudate (about 5 per cent) may represent a marker for Tourette's disorder, and larger prefrontal cortex would likely result from the ability to suppress tics. Although tics are highly heritable, non genetic factors appear to contribute to these brain differences.

Autoimmune factors

For the last decade, clinical and research interest has grown in an autoimmune model of OCD and/or tic disorders, which could apply to a subgroup of subjects whose disorder begins abruptly during childhood. An association was first reported between acute onset OCD and Sydenham's chorea, a childhood movement disorder associated with rheumatic fever, which is thought to result from an antineuronal antibody-mediated response to group A beta-haemolytic streptococcus (GABHS), directed at portions of the basal ganglia.[59] OCD, or some of its symptoms, have been reported in 70 per cent of Sydenham's chorea cases.[60, 61] Furthermore, in the absence of the neurological symptoms of Sydenham's chorea, post-streptococcal cases of childhood-onset OCD, tics and/or other neuropsychiatric syndromes have been described under the acronym of paediatric autoimmune neuropsychiatric disorders associated with streptococcal infections (PANDAS). Swedo *et al.*[6] defined this novel group of patients using five diagnostic criteria: presence of OCD and/or tic disorder, prepubertal onset, episodic course of symptom severity, abrupt onset or dramatic exacerbations of symptoms temporally associated with GABHS infections (as evidenced by positive throat culture and/or elevated anti-GABHS titers), and association with neurological abnormalities (motoric hyperactivity or adventitious movements, such as choreiform movements or tics). An antigen labelled D8/17, on the surface of peripheral blood mononuclear cells has been shown to be a marker for the genetic tendency to generate abnormal antibodies to GABHS. Two independent groups of researchers have found a greater expression of the D8/17 antigen in the B lymphocytes of patients with childhood-onset OCD or Tourette's disorder compared with healthy controls, indicating that the presence of the D8/17 antigen may serve as a marker of susceptibility for OCD or tics.[62, 63]

Course and prognosis

Several follow-up studies of subjects treated for OCD during childhood or adolescence have looked at the outcome of the disorder in early adulthood.[11] All studies demonstrate the continuity of the diagnosis of OCD from childhood to adulthood: when subjects are still symptomatic, the main diagnosis is almost invariably OCD, although comorbid disorders are frequent, especially mood and/or anxiety disorders. Spontaneous course is most often marked by a waxing and waning severity of the disorder, whereas remissions under treatment can be followed by relapses, even after long periods of time. In the early studies in which subjects had received no or non-specific treatment, the recovery rate was poor (13–30 per cent). By the time patients had access to specific treatment with SRIs and/or cognitive behavioural therapy (CBT), recovery rates increased to 55–65 per cent, although many of the symptom-free subjects at follow-up were still taking medication. A meta-analysis analyzed 16 studies that followed paediatric OCD patients between 1 and 15 years.[64] The overall remission rate (not fulfilling criteria for subthreshold or full OCD) was 40 per cent, with pooled mean persistence rates of 41 per cent for full OCD. Poor prognostic factors included a poor initial treatment response, and comorbid psychiatric illness.

In the majority of cases, tics are transient (present for less than 12 months), or wax and wane in severity with periods of exacerbation of an average duration of 9 weeks. The course of worst-ever tic severity usually falls between 7 and 14 years of age, which also includes the period when tics are most variable (10–12 years). By the end of the second decade, there is usually a steady decline in tic severity. However, adults who are able to suppress tics may be left with distracting urges, and a significant minority (15–30 per cent) continues to have severe tics into adulthood.[65] Despite substantial problems in childhood, the majority of patients with Tourette's disorder grow up to become well socially integrated and economically independent adults. However, as much as 25 per cent have persistent mental health problems. In those, tic severity fluctuates, and psychiatric co-morbidities (ADHD, other disruptive behaviour problems, OCD, mood and anxiety disorders, learning problems) are often the main determinants of global outcome. Poorer prognoses are also associated with comorbid developmental disorders, chronic physical illness, unstable or unsupportive family environment, social difficulties, or exposure to psychoactive drugs such as cocaine.[5]

Treatment

Evidence

The treatment of paediatric OCD has changed dramatically over the past 20 years, with two modalities being empirically shown to ameliorate the core symptoms of the disorder: CBT and pharmacological treatment with SRIs. In Tourette's disorder, D_2 dopamine antagonists have been used with relative success since the 1960s, and CBT techniques are being increasingly scrutinized.

Cognitive behavioural treatment

The cognitive behavioural model of OCD posits that compulsions function to reduce fear, and are subsequently reinforced by fear reduction, which prevents normal habituation and realistic appraisal of the threat value of feared stimuli. Techniques incorporating exposure and ritual prevention are designed to break this cycle by exposing the individual to feared situations, while simultaneously reducing compulsive behaviours.[66] Discussion of obsessive thoughts, and other irrational beliefs, is often part of the exposure exercises but these informal cognitive techniques are used to support exposure rather than to replace it. The CBT of youth with OCD generally involves a three-stage approach, consisting of information gathering, therapist-assisted graded exposure with response prevention, and homework assignments.[67] Anxiety management training plays an adjunctive role. For children with predominantly internalizing symptoms, treatment also includes relaxation and cognitive training. Families need to be involved, to varying extents according to individual situations. CBT is usually implemented with 13 to 20 weekly individual or family sessions, and homework assignments. Partial responders or nonresponders may require more frequent sessions, and out-of-office therapist-assisted training.

Table 9.2.8.1 Controlled studies of pharmacological or psychological treatment in paediatric OCD

Study, Year	N (age), Study duration	Drug (daily dose), Study design	Outcome	% Improvement from baseline on active treatment for OC symptoms
Flament *et al.*, 1985	19 (6–18 yr), 5 wk	CMI (mean 141 mg) Crossover vs. PBO	CMI > PBO at 3–5 wk	22–44 %
Leonard *et al.*, 1989	47 (7–19 yr), 5 wk	CMI (mean 150 mg) Crossover vs. DES	CMI > DES at 3–5 wk	19–44 %
DeVeaugh-Geiss *et al.*, 1992	60 (10–17 yr), 8 wk	CMI (75–200 mg) Parallel vs. PBO	CMI > PBO at 3–8 wk	34–37 %
Riddle *et al.*, 1992	14 (8–15 yr), 8 wk	FLX (20 mg) Crossover vs. PBO	FLX > PBO at 8 wk	33–44 %
March *et al.*, 1998	187 (6–17 yr), 12 wk	SER (mean 167 mg) Parallel vs. PBO	SER > PBO at 3–12 wk	21–28 %
Riddle *et al.*, 2001	120 (8–17 yr), 10 wk	FLV (50–200 mg) Parallel vs. PBO	FLV > PBO at 1–10 wk	21–25 %
Geller *et al.*, 2001	103 (7–17 yr), 13 wk	FLX (mean 40 mg) Paralell vs. PBO	FLX > PBO at 7–13 wk	25–49 %
Liebowitz *et al.*, 2002	43 (8–17 yr), 16 wk	FLX (mean 64 mg) Parallel vs. PBO	FLX > PBO at 16 wk	42 %
Geller *et al.*, 2004	203 (7–17 yr), 10 wk	PAR (mean 30 mg) Parallel vs. PBO	PAR > PBO at 2–10 wk	47–65 %
Barrett *et al.*, 2004; 2005	77 (7–17 yr), 14 wk	Randomized parallel study ICBFT vs. GCBFT vs. control condition	ICBFT and GCBFT > control condition at 14 wk	65 % for ICBFT 60 % for GCBFT

CMI: clomipramine; DES: desipramine; FLV: fluvoxamine; FLX: fluoxetine; PAR: paroxetine; PBO: placebo; SER: sertraline

ICBT: individual cognitive behavioural treatment, GCBT: group cognitive behavioural treatment

A number of open trials, and four controlled studies (see Tables 9.2.8.1 and 9.2.8.2), have documented the beneficial effects of CBT, alone or in combination with pharmacotherapy, for children and adolescents with OCD, with improvement measured on the CY-BOCS scores ranging from 25 per cent to 67 per cent. In the first controlled study by Barrett *et al.*,[68] 88 per cent of youth treated with individual cognitive behavioural family treatment (CBFT), and 76 per cent of those treated with group CBFT showed clinically significant improvement, as compared to no improvement for any patients in the waitlist condition. Treatment gains were maintained at 12- to 18-month follow-up, with a total of 70 per cent of participants in individual therapy, and 84 per cent in

Table 9.2.8.2 Comparative studies of cognitive behavioural treatment and pharmacological treatment in paediatric OCD

Study, Year	N (age), Study duration	Drug (daily dose), Study design	Outcome	% Improvement / Remission for OC symptoms
De Haan *et al.*, 1998	22 (8–18yr), 12 wk	CMI CMI (25–200 mg) vs. BT	BT > CMI	Improvement from baseline BT: 59.9 % CMI: 33.4 %
POTS Team 2004	112 (7–17yr), 12 wk	SER (mean 150 mg) vs. PBO vs. CBT vs. Combi (CBT+SER) 3-site study, Parallel design	Combi > CBT = SER > PBO effect size: 1.4 (Combi), 0.97 (CBT), 0.67 (SER)	Improvement from baseline Combi: 53 %; CBT: 46 %; SER: 30 %; Placebo: 15 % Remission post-treatment (CY-BOCS<10) Combi: 53.6 %; CBT: 39.3 % SER: 21.4 %; PBO: 3.6 %
Asbahr *et al.*, 2005	40 (9–17 yr), 12 wk F/U: 9 mo	SER (mean 137 mg) vs. GCBT	Significant improvement on CY-BOCS with both GCBT and SER Relapse during F/y: 50% (SER), 5 % (GCBT)	NR

BT: behavioural treatment; CBT: cognitive behavioural treatment; CMI: clomipramine; Combi: combination treatment; F/U: follow-up; GCBT: group cognitive behavioural treatment; NR: not reported; PBO: placebo; POTS: paediatric obsessive–compulsive disorder treatment study; SER: sertraline

group therapy diagnosis-free at follow-up, and no significant difference between the two treatment modalities.[69] More evidence for the efficacy of CBT in the treatment of paediatric OCD comes from three studies that have compared CBT to pharmacotherapy and/or their combination (see below).

Behavioural techniques play an important role in the treatment of tics, although generally as adjunctive to medication. In a model of operant conditioning, ticking relieves unpleasant premonitory sensations, which reinforces the maintenance of tics. Habit reversal (HR) is based on awareness training regarding the premonitory urges, followed by training a competing response (a movement that involves the same muscle group as the tic) after the first sensation that a tic is about to occur. The response must be temporally contingent on each occurrence of the urge, but using a muscle group related or unrelated to the tic may not be crucial for tic suppression. Relaxation, self-monitoring, contingency training for positive reinforcement of not ticking, and social support are used as ancillary components of HR. An extension of HR is exposure and response prevention (ERP): after the urge, the patient suppresses the tic voluntarily, which should lead to its extinction. Both techniques have been showed to yield relatively large effect size (1.06 to 1.42) in reducing tic severity at post-treatment[70] and long-term follow-up.[71] A recent review of the literature[72] concluded that the use of HR to treat tics can currently be classified as a 'well established' treatment, and that of ERP as a 'probably efficacious' treatment. Contrary to initial fears, these behavioural techniques have no negative consequences, such as substitution of the targeted tics, or post-suppression rebound or worsening due to increased awareness of premonitory urges.

Psychopharmacological treatment

In the past 25 years, a number of randomized, controlled clinical trials (summarized in Table 9.2.8.1) have been conducted in children and adolescents with OCD demonstrating, as in adults, the selective and unique efficacy of the SRIs in the short-term and long-term treatment of the disorder.[67,73–79] Results have consistently shown that: the antiobsessional action of the SRIs is independent of the presence of depressive symptoms at baseline; their antiobsessional action takes longer to appear than their antidepressant action; the therapeutic response occurs gradually over a few weeks to a few months; final response is most often partial, with a mean reduction of OC symptoms from baseline to post-treatment ranging from 19 per cent to 44 per cent across measures and across studies.[80] Geller *et al.*[81] conducted a meta-analysis of 12 randomized, controlled medication trials in children and adolescents with OCD (total N=1 044), demonstrating that all serotonergic medications were highly significantly superior to placebo, with consistent findings across studies but a modest overall effect (the pooled standard mean difference between active drug and placebo was only 0.46). Clomipramine was statistically superior to the specific serotonin reuptake inhibitors (SSRI), but temporal trends might, at least, partly explain this apparent superiority: the clomipramine trials were conducted earlier in time when no other treatment was available, while the patient population included in subsequent controlled trials have changed over the years with increased availability of pharmacological alternatives. No head-to-head paediatric studies of clomipramine versus an SSRI have been conducted. The recommended daily dosages for SSRIs in the treatment of paediatric OCD are shown on Table 9.2.8.3.

Table 9.2.8.3 Recommended daily dosages of serotonin reuptake inhibitors for the treatment of paediatric OCD

Medication	Starting Dose[a]	Initial Targeted Dose[b,c]	Maximal Dose[b]
Citalopram	10 mg	40 mg	60 mg
Escitalopram	5 mg	20 mg	20 mg
Fluoxetine	10 mg	40 mg	80 mg
Fluvoxamine	25 mg	200 mg	300 mg
Paroxetine	10 mg	40 mg	60 mg
Sertraline	25 mg	100 mg	200 mg
Clomipramine	10 mg	150 mg	250 mg

[a] these doses should be given for about one week, that is about the time necessary to achieve steady state for these drugs, with the exception of fluoxetine. This would ensure that no agitation or increased anxiety is triggered by the medication.

[b] for subjects weighing at least 50 kg; for smaller individuals, a weight-proportional regimen should be used

[c] according to side effects and response

A few studies, summarized in Table 9.2.8.2, have compared pharmacological treatment to CBT or their combination for children and adolescents with OCD.[82–84] In the U.S. 12-week paediatric OCD treatment study (POTS; N=112), the combined treatment with CBT and sertraline had the best rate of clinical remission (53.6 per cent vs 39.3 per cent on CBT alone, 21.4 per cent on sertraline alone, and 3.6 per cent on placebo); the remission rate for the combined treatment was not statistically different from that in the CBT only condition.[83] In Asbahr *et al.*[84] study, both group CBT and sertraline induced a significant improvement in OC symptoms, but after a 9-month post-treatment follow-up period, subjects in the group CBT condition had a significantly lower rate of symptom relapse. A few case reports have also indicated that the addition of CBT to pharmacotherapy can allow successful withdrawal from medication.

The use of clomipramine can entail anticholinergic side effects (dry mouth, dizziness, headache, tremor, fatigue, constipation, sweating, dyspepsia, sexual dysfunction), which may not abate over time and even increase with ascending titration.[85] Clomipramine can cause tachycardia and prolongation of the QT and QTc intervals, and ECG monitoring is recommended.[86] Risks of toxicity also include seizures, and rare cases of sudden death have been reported in children taking tricyclic antidepressants.[87] Although less frequent and less disturbing that the secondary effects of clomipramine, the most commonly described adverse effects of the SSRIs include gastro-intestinal (nausea, constipation, abdominal pain), and central nervous system complaints (headache, tremor, drowsiness, akathisia, insomnia, disinhibition, agitation).[88] The possible induction of mania can also be of concern. Although not commonly reported in clinical trials, the eventual occurrence of sexual side effects should be reviewed with adolescents, since these may impact adherence to treatment. Recent reports of possible growth suppression associated with the SSRIs suggest that monitoring of height may also be advisable.[89] A recent concern, highly visible in the media, has been a possibly increased risk for suicidal thoughts, self-harm and/or harm to others, in youth treated with the SSRIs. However, no individual OCD study has documented a significantly increased risk for suicidal ideation or behaviour on

a SSRI compared to placebo. In pooled analyses of the controlled studies conducted in youth with OCD and other anxiety disorders, behavioural side effects variously labelled as activation, akathisia, disinhibition, impulsivity, and hyperactivity have appeared, but there was no evidence for a significant increase in the relative risk of suicidal thoughts or behaviours.[90] In a recent review of 27 trials of antidepressants in participants younger than 19 years (including six trials for treatment of OCD), there was an increased risk difference of suicidal ideation/suicide attempts across trials and across indications for drug versus placebo, but no completed suicide, and the benefits of antidepressants appeared much greater than the risk.[91] In any case, rigorous clinical monitoring for suicidal ideation and other potential indicators for suicidal behaviour remains advised in youth treated with the SSRIs.

The mainstay of treatment for Tourette's disorder has been traditional antipsychotics, i.e. the potent dopamine (D2) postsynaptic blockers haloperidol and pimozide.[92] The usual starting dose is 0.25 mg/day of haloperidol or 1 mg/day of pimozide. Increments (0.5 mg haloperidol or 1 mg pimozide) may be added at 7 to 14 days intervals, up to 1–4 mg/day for haloperidol and 2–8 mg/day for pimozide. Atypical antipsychotics have also been used for the treatment of tics, and differences in efficacy appear to be related to their relative potency of dopamine blockade. Risperidone has been shown to be superior to placebo,[93,94] and equally effective to pimozide,[95,96] at doses ranging from 1 to 3 mg/day (starting dose, 0.25–0.50 mg). The specific D2 receptor-blocking agents, tiapride and sulpiride, have been commonly used for the treatment of tics in Europe in doses ranging from 15–500 mg/day and 200–1 000 mg/day, respectively, but they are not available in the U.S. The use of traditional antipsychotics is limited by a range of side effects, both in the short term (parkinsonism, dystonia, dyskinesia, and akathisia) and in the long term (tardive dyskinesia). The newer antipsychotics appear to have a lower frequency of neurological side effects in the short term, and a lower relative risk of tardive dyskinesia, but weight gain, hyperlipidemia and diabetes are of growing concern. Among the antipsychotics used in the treatment of tics, pimozide is the most likely to be associated with prolonged QTc interval, although this is a rare occurrence at therapeutic doses. An ECG is recommended before starting treatment, during the dose-adjustment phase, and annually during ongoing treatment. Patients should also be informed that the risk for cardiac conduction abnormalities may increase when pimozide is combined with drugs that inhibit cytochrome P450 3A4 isoenzyme (e.g. macrolide antibiotics, SSRIs, etc.).[97]

Clonidine is an antihypertensive agent (α-2-adrenergic agonist) that has been shown effective for treatment of tic disorders, presumably via acute and chronic downstream effects on dopamine. Clinical trials indicate an average 25–35 per cent reduction in symptoms over 8 to 12 weeks. Clonidine seems especially useful in improving attention problems and ameliorating complex motor tics. Treatment must be started at a low dose (0.05 mg in the morning), and slowly increased to 0.15–0.30 mg per day, given in several doses throughout the day. The major side effects are sedation, hypotension, dizziness, and a decrease of salivatory flow; blood pressure and pulse should be measured at baseline and monitored during dose adjustment, and patients and families should be educated about the potential for rebound increases in blood pressure, tics, and anxiety upon abrupt discontinuation.[98] Guanfacine is another α-adrenergic antihypertensive that has entered into clinical practice. Given the added disability attributable to ADHD in children and adolescents with tic disorders, a treatment combining an α-2 agonist and a stimulant may produce better outcomes than either alone. Recent studies suggest that the acute onset or worsening of tic symptoms among patients receiving stimulants may be simply an expression of the spontaneous time course of tics and comorbid ADHD.[99] Atomoxetine is a selective norepinephrine reuptake inhibitor that reduces significantly ADHD symptoms, and may also improve tics.

Management

As described above, OCD and tic disorders are frequently chronic, and most treatments, notably medication, are suspensive but not curative. Therefore, when defining a treatment plan, clinicians should be aware that they embark on a long-lasting task.

CBT is generally favoured as the initial treatment of choice for OCD, especially in milder cases without significant comorbidity, whereas presence of comorbid depression, anxiety, disruptive behaviour, or insufficient cognitive or emotional ability to cooperate in CBT, are indications for including an SRI in the initial treatment. However, youth who have OCD and comorbid conditions may not be as responsive to SSRIs for OCD, as shown in Geller *et al.*[100] study, in which the response rate was 75 per cent in the non-comorbid OCD group, but significantly lower when OCD was comorbid with ADHD (56 per cent), tic disorder (53 per cent), or oppositional defiant disorder (39 per cent); comorbid OCD may also be more vulnerable to relapse with SSRI discontinuation. Although the SSRIs are indicated for OCD, depression, and anxiety disorders, which make them an ideal first drug teatment when OCD is comorbid with an affective disorder, monitoring for the emergence of manic symptoms is required. The treatment of OCD comorbid with ADHD using stimulants may present a challenge because theoretical concerns exist that stimulants may increase obsessional symptoms. However, it is common clinical practice to combine a SSRI (or CMI) with a psychostimulant.[101]

Similar to adult patients, at least one third of young people with OCD prove refractory to treatment, and many 'responders' exhibit only partial response.[11] For children and adolescents who do not seem to benefit from SRI treatment, the first steps are (i) to reevaluate the diagnosis and associated features, and (ii) to review the adequacy of the dose, duration, and compliance with medication. Then, for many youth with a partial response to pharmacotherapy, further improvement may be obtained by adding a concurrent CBT intervention. Furthermore, if a first SSRI trial fails to produce an adequate response, pharmacological algorithms generally recommend switching to a second SSRI. Considering that these drugs all inhibit the 5-HT transporter, the first drug can be discontinued abruptly, and the second initiated at a dose in the middle of its therapeutic range. Because of its longer half-life, if fluoxetine has been the first treatment, it could be stopped abruptly, but the second SSRI must be titrated slowly. If two or three successive trials of SSRIs have failed, clomipramine is generally considered as the next option. Given that it has a half-life in the same range as most of the other SRIs, no time should be wasted in the substitution, unless a switch from fluoxetine is carried out.

Very few drug augmentation or combination strategies have been tested for youth with treatment-resistant OCD. The addition of risperidone to various SRI agents (clomipramine, sertraline, fluoxetine, paroxetine) has been reported in a series of OCD adolescents

with no comorbid tic disorder, with only modest benefits.[102] The combination of an SSRI with clomipramine takes advantage of the pharmacokinetic and pharmacodynamic interactions of these medications, but it is important to monitor for adverse effects, particularly cardiovascular side effects, and the possible emergence of a serotonin toxic syndrome. Combining two SSRIs is a common clinical practice, despite the risk for drug interaction, since all SSRIs inhibit cytochrome P-450, and combinations may result in increased blood levels of each SSRI.

If short-term treatment with a SSRI often leaves OCD youth with residual symptoms, over an extended period of SSRI treatment, they may experience greater improvement. Three long-term (1–2 years) open studies have documented continued improvement after the acute phase, but at a much slower rate, with fluvoxamine, sertraline, and citalopram. Similarly, in two continuation studies of clomipramine (4–12 months), treatment continued to be effective and well tolerated. There are hardly any data on the doses of SRIs that should be used in treatment prolongation, versus those used in the acute treatment phase. It would thus be prudent to maintain the regimen that produced the maximal improvement in the acute treatment. The most recent guidelines for adult patients recommend a minimum treatment of one to two years, followed by a gradual taper to, first, avoid discontinuation phenomena and, second, monitor patients for a possible deterioration.[103] In youth, it is particularly recommended to choose a period free of stress (e.g. summer vacation) for tapering medication, and to provide alternative psychological support (CBT, education on relapse risk and management) for the period of discontinuation.

Although tics are a common childhood problem, only a small minority of cases find their way to clinics. Given the waxing and waning nature of tic disorders, usual therapeutic practice will initially focus on careful clinical observation, along with educational and supportive interventions, and pharmacological treatments are held in reserve. The decision about whether and how to treat will depend on the primary diagnosis, and the degree of interference with the child's development and functioning. Most simple tics occurring in the absence of severe functioning impairment respond to a simple explanation of the mechanisms. In case of comorbidity with a mood and anxiety disorder, it is not uncommon to see improvement in tic severity after successful treatment of the affective disorder with a SSRI.[104] When tics are responsible of functional impairment, the decision to use medication follows careful assessment and identification of target symptoms, that are interfering in the patient's quality of life. The selection of medication is based on a balance of risks and benefits, and in order not to expose the subject to excessive unwanted side effects, pharmacological treatment should not aim at complete disappearance of tics. For tics of moderate severity, clonidine or guanfacine may be considered as the first line treatment given their safety margin, and an expected 30 per cent decrease in tic severity may be sufficient. For tics in the marked or severe range, however, more potent medications, such as antipsychotics, that decrease tics severity by 35–60 per cent should be considered, despite the increased risk of adverse effects. In addition, patients with tic disorders and their families should be cautioned about both licit and illicit drug use, since sympathomimetic agents ranging from decongestants through speed and cocaine, markedly exacerbate tics.

For children and adolescents with OCD and comorbid tic disorder, the SSRIs alone might have little anti-obsessional effect, and there are reports suggesting that fluvoxamine and fluoxetine may exacerbate or even induce tics in some patients. The adult literature, and a few case reports of children and youth with OCD and comorbid tic disorder suggest that combined treatment with a SRI and a low dose of risperidone or another atypical antipsychotic is a reasonable option.[105] Using data from the POTS study, March *et al.*[106] found that tic disorders appeared to adversely impact the outcome of medication management of paediatric OCD, in contrast to CBT outcomes which were not differentially impacted.

Therapeutically, the finding of possible autoimmune cases of OCD and tic disorders raises the clinical possibility that immunosuppressant treatments might be effective, and a few experimental studies have been conducted. A double-blind, placebo-controlled study in resistant cases of PANDAS supported the efficacy of both plasma exchange and intravenous immunoglobulin compared to a sham condition at one month, and the effects were maintained after 1 year for 82 per cent of subjects with follow-up assessment.[107] In another study antibiotic prophylaxis with penicillin or azithromicin was administered for 12 months to a group of children with PANDAS; compared to the year prior to entry in the study, significant decreases were observed in the number of streptococcal infections, and the number of neuropsychiatric exacerbations with both prophylactic treatments.[108] There is preliminary evidence for the efficacy of CBT in OCD cases of the PANDAS phenotype.[109] However, it is still unknown what percentage of children with OCD may be part of the PANDAS subgroup, and neither immunosuppressant nor antibiotic treatments are to be used for such cases out of the context of board-approved research protocols. Nevertheless, children with abrupt onset or exacerbation of OC and/or tic symptoms require careful consideration of medical illnesses including upper respiratory tract infections during the preceding months, to be promptly treated if present. A throat culture and antistreptolysin O or antistreptococcal DNAase B titers may be considered to assist in diagnosing a GABHS infection.[101]

Even if they are less efficacious alone, other treatment modalities should not be neglected. The effectiveness of psychotherapy per se—apart from behavioural and cognitive interventions—on OCD and tic disorders has not been demonstrated, but the symptoms may have a profound impact on the life of subjects affected, and traditional psychotherapeutic approaches may be useful to help children and adolescents address the intrapsychic conflicts that may affect or result from their illness. Some families become extensively involved in participating in compulsive rituals or reassuring obsessional worries, others become mired in gruelling angry struggles with their symptomatic child. Work with families on how to manage the child's symptoms, cope with the stress and family disruption that often accompanies OCD and tic disorders, and participate effectively in behavioural or pharmacological treatment is crucial. Most cognitive behavioural approaches of paediatric OCD include the involvement of a parent in some therapeutic sessions, and it is noteworthy that the Barrett *et al.*[68] study, which actively involves the family in the child's treatment, reports improvement in the highest range of outcomes among CBT studies. In children with tic disorders, oppositional, defiant, and disruptive behaviours are common, and parenting skill training may be an important adjunctive to treatment. For cases of very incapacitating OCD, there is some empirical evidence that milieu therapy in an inpatient setting may be a useful resource. Finally, the growing availability, in many countries, of family support and

advocacy groups for patients with OCD and tic disorders may be most useful to alleviate the discouragement and incomprehension created by these disorders, and give access to appropriate treatment resources.

Possibilities for prevention

At present, there is no known preventive strategy individually targeted at either OCD or tic disorders. However, early intervention and comprehensive treatment, as long as needed, is certainly the best way to prevent severe incapacitation, and achieve complete recovery in some cases. Even when response to successive treatment efforts is less than optimal, the improvement in function and quality of life may be considerable. In families with one or several cases of OCD or tic disorders, clinicians should be attentive to the onset of similar symptoms in children and siblings, and treat cases of newly occurring disorder early and vigorously. Although this might concern only a fraction of patients, evidence of the onset or exacerbation of OCD or tics associated with streptococcal exposure warrants standard antibiotic treatment and ongoing monitoring for recurrent infection.

Conclusion

Paediatric OCD is the disorder, in child psychiatry, whose clinical picture most closely resembles its adult counterpart. Despite a relative diversity, the symptom pool is remarkably finite, and very similar to that seen in older individuals. Prevalence, comorbidity, and response to behavioural and drug treatment also appear similar across the lifespan. For tic disorders, there is continuity between child and adult presentations, but the disease is much more prone to resolve spontaneously, or to be less disruptive in adulthood. Both OCD and tics occur more often in males than in females, and are likely to be linked to an array of neurobiological abnormalities, many of which remain to be understood.

Invaluable benefits can now be obtained from available behavioural and pharmacological treatments, but complete remission remains uncertain and long-term management may be required. Thus, the treatment of OCD and tics in children and adolescents remains a clinical challenge. It requires careful assessment of the targeted symptoms and, in many cases, comorbidity; attention to the quality of the child's functioning at home and with peers; use of specific CBT interventions, which are not readily available (or accessible) in all communities; patience and caution in the choice and adjustment of medication; and vigilance in watching potential side effects. Given the possible chronicity of OCD and/or tic disorders, and their changing patterns in severity and impact over the childhood and adolescent years, optimal treatment generally requires a long-term ongoing relationship with the child and family.

Current conceptualizations of OCD and tic disorders have been shaped by advances in systems neuroscience and functional in vivo neuroimaging. Continued success in these areas should lead to the targetting of specific brain circuits for more intensive research. This should include testing novel pharmacological agents, tracking treatment response using neuroimaging techniques, and possibly investigating circuit-based therapies using deep-brain stimulation for refractory cases. The identification of the PANDAS subgroup of patients, with an abrupt onset and dramatic exacerbations, certainly brings new insights into the pathophysiology of OCD and tic disorders, and may lead to new assessment and treatment strategies. The increasing evidence for susceptibility genes in OCD and tic disorders will also doubtless point to new therapeutic directions. Furthermore, it is likely that many of the empirical findings used in research on paediatric OCD and tic disorders will be relevant to a better understanding of both normal development, and other disorders of childhood onset.

Further information

Obsessive Compulsive Foundation. Available at: http://www.ocfoundation.org

Tourette Syndrome Association. Available at: http://www.tsa-usa.org/index.html

Tourette syndrome online. Available at: http://www.tourette-syndrome.com/default.htm

References

1. Scahill, L., Riddle, M.A., McSwiggin-Hardin, *et al.* (1997). Children's Yale-Brown Obsessive-Compulsive Scale: reliability and validity. *Journal of the American Academy of Child and Adolescent Psychiatry,* **36**, 844–52.
2. Stewart, S.E., Rosario, M.C., Brown, T.A., *et al.* (2007). Principal component analysis of obsessive-compulsive disorder symptoms in children and adolescents. *Biological Psychiatry,* **61**(3), 285–91.
3. Comings, D.E., and Comings, B.G. (1985). Tourette syndrome: clinical and psychological aspects of 250 cases. *American Journal of Human Genetics,* **37**, 435–50.
4. Leckman, J.F., Bloch, M.H., Scahill, L. *et al.* (2006). Tourette Syndrome: The Self Under Siege. *Journal of Child Neurology,* **21**, 642–49.
5. Leckman, J.F., and Cohen, D.J. (1994). Tic disorders. In *Child and adolescent psychiatry. Modern approaches* (3rd edn) (ed. M. Rutter, E. Taylor and L. Hersov), pp. 455–66. Blackwell Science, Oxford.
6. Swedo, S., Leonard, H.L., Mittleman, B., *et al.* (1998). Pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections (PANDAS): clinical description of the first 50 cases. *American Journal of Psychiatry,* **155**, 264–71.
7. Riddle, M.A., Scahill, L., King, R., *et al.* (1990). Obsessive compulsive disorder in children and adolescents: phenomenology and family history. *Journal of the American Academy of Child and Adolescent Psychiatry,* **29**, 766–72.
8. Flament, M.F., Whitaker, A., Rapoport, J.L., *et al.* (1988). Obsessive compulsive disorder in adolescence: an epidemiological study. *Journal of the American Academy of Child and Adolescent Psychiatry,* **27**, 764–71.
9. Burd, L., Freeman, R., Klug, M., *et al.* (2006). Variables associated with increased tic severity in 5,500 participants with Tourette Syndrome. *Journal of Developmental & Physical Disabilities,* **18**(1), 13–24.
10. Banaschewski, T., Woerner, W. and Rothenberger, A. (2003). Premonitory sensory phenomena and suppressibility of tics in Tourette syndrome: developmental aspects in children and adolescents. *1: Dev Med Child Neurol*, **45**(10), 700–703.
11. Flament, M.F. and Cohen, D. (2000). Child and adolescent obsessive compulsive disorder. In *Obsessive-Compulsive Disorder: WPA Series Evidence and Experience in Psychiatry – Volume Four* (eds. M. Maj, N. Sartorius), John Wiley & Sons, Chichester, England, pp. 147–83.
12. Burd, L., Kerbeshian, J., Wikenheiser, M., *et al.* (1986). Related Articles, Links A prevalence study of Gilles de la Tourette syndrome in North Dakota school-age children. *Journal of the American Academy of Child Psychiatry*, **25**(4), 552–3.
13. Snider, L.A., Seligman, L.D., Ketchen, B.R., *et al.* (2002). Tics and problem behaviors in schoolchildren: prevalence, characterization, and associations. *Pediatrics,* **110**(2 Pt 1), 331–6.

14. Geller, D.A., Biederman, J., Griffin, S., *et al.* (1996). Comorbidity of obsessive–compulsive disorder with disruptive behaviour disorders. *Journal of the American Academy of Child and Adolescent Psychiatry,* **35**, 1637–46.
15. Kano, Y., Ohta, M., and Nagai, Y. (1998). Clinical characteristics of Tourette syndrome. *Psychiatry and Clinical Neuroscience,* **52**, 51–7.
16. Leckman, J.F. (1993). Tourette's syndrome. In *Obsessive compulsive related disorders* (ed. E. Hollander), pp. 113–37. American Psychiatric Press, Washington, DC.
17. Leonard, H.L., Lenane, M.C., Swedo, S.E., *et al.* (1992). Tics and Tourette's disorder. A 2-to 7-year follow-up study of 54 obsessive–compulsive children. *American Journal of Psychiatry,* **149**, 1244–51.
18. Leckman, J.F., McDougle, C.J., Pauls, D.L., *et al.* (2000). Tic–related vs non-tic related obsessive compulsive disorder. In *Obsessive compulsive disorders: contemporary issues in treatment* (ed. W.K. Goodman, J.D. Maser, and M.V. Rudorfer), pp. 43–68. Erlbaum, Hillsdale, NJ.
19. Geller, D.A., Biederman, J., Jones, J. *et al.* (1998). Is juvenile obsessive-compulsive disorder a developmental subtype of the disorder? A review of the pediatric literature. *Journal of the American Academy of Child and Adolescent Psychiatry,* **37**(4), 420–27.
20. Pollock, R. and Carter, A. (1999). The familiar and developmental content of obsessive-compulsive disorder. *Child Adoles Psychiatry,* **33**, 2–15.
21. Kumar, R., and Lang, A. (1997). Secondary tic disorders. *Neurologic Clinics of North America,* **15**, 309–30.
22. Scahill, L., Tanner, C. and Dure, L. (2001). The epidemiology of tics and Tourette syndrome in children and adolescent, in Tourette Syndrome and Associated Disorders. *Advances in Neurology Seriews,* **85**, 261–71.
23. Apter, A., Pauls, D., Zohar, A.H., *et al.* (1993). An epidemiologic study of Gilles de la Tourette in Israel. *Archives of General Psychiatry,* **50**, 734–38.
24. Peterson, B.S., Pine, D.S., Cohen, P. *et al.* (2001). Prospective, longitudinal study of tic, obsessive compulsive, and attention-deficit/ hyper activity disorders in an epidemiological sample. *Journal of the American Academy of Child and Adolescent Psychiatry,* **40**, 685–95.
25. Heyman, I., Fombonne, E., Simmons, H., *et al.* (2001). Prevelence of obsessive-compulsive disorder in the British Nationwide Survey of Child Mental Health. *British Journal of Psychiatry,* **179**, 324–9.
26. Zohar, A.H., Ratzosin, G., Pauls, D.L., *et al.* (1992). An epidemiological study of obsessive-compulsive disorder and related disorders in Israeli adolescent. *Journal of the American Academy of Child and Adolescent Psychiatry,* **31**, 1057–61.
27. Reinherz, H.Z., Giaconia, R.M., Lefkowitz, E.S., *et al.* (1993). Prevalence of psychiatric disorders in a community population of older adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry,* **32**, 369–77.
28. Douglass, H.M., Moffitt, T.E., Reuven, D., *et al.* (1995). Obsessive–compulsive disorder in a birth cohort of 18-year-olds: prevalence and predictors. *Journal of the American Academy of Child and Adolescent Psychiatry,* **34**, 1424–31.
29. Lin, H., Katsovich, L., Ghebremichael, M. *et al.* (2007). Psychosocial stress predicts future symptom severities in children and adolescents with Tourette syndrome and/or obsessive-compulsive disorder. *JF. 1: Journal of Child Psychology and Psychiatry, and Allied Disciplines,* **48**(2), 157–66.
30. Flament, M.F., Rapoport, J.L., Murphy, D.L., *et al.* (1987). Biochemical changes during clomipramine treatment of childhood obsessive compulsive disorder. *Archives of General Psychiatry,* **44**, 219–25.
31. Sallee, F.R., Richman, H., Beach, K., *et al.* (1996). Platelet serotonin transporter in children and adolescents with obsessive–compulsive disorder or Tourette's syndrome. *Journal of the American Academy of Child and Adolescent Psychiatry,* **35**, 1647–56.
32. Serra-Mestres, J., Ring, H.A., Costa, D.C. *et al.* (2004). Related Articles, Links Dopamine transporter binding in Gilles de la Tourette syndrome: a [123I]FP-CIT/SPECT study. *Acta Psychiatr Scand*, **109**(2), 140–6.
33. Singer, H.S., Szymanski, S., Giuliano, J. *et al.* (2002). Elevated intrasynaptic dopamine release in Tourette's syndrome measured by PET. *1: American Journal of Psychiatry*, **159**(8), 1329–36.
34. Wong, D.F., Singer, H.S., Brandt, J. *et al.* (1997). D2-like dopamine receptor density in Tourette syndrome measured by PET. *1: Journal of Nuclear Medicine*, **38**(8), 1243–7.
35. Heinz, A., Knable, M.B., Wolf, S.S. *et al.* (1998). Tourette's syndrome: [I-123]beta-CIT SPECT correlates of vocal tic severity. *1: Neurology,* **51**(4), 1069–74.
36. Rasmussen, S.A., and Tsuang, M.T. (1984). The epidemiology of obsessivecompulsive disorder. *Journal of Clinical Psychiatry,* 45, 450–7.
37. Price, R.A., Kidd, K.K., Cohen, D.J., *et al.* (1985). A twin study of Tourette syndrome. *Archives of General Psychiatry,* **42**, 815–20.
38. Lenane, M. (1989). Families in obsessive-compulsive disorder. Rapoport, J. L. Obsessive-compulsive disorder in children and adolescents. Washington DC: American Psychiatric Press, pp. 237–49.
39. Pauls, D., Alsobrook, J., Goodman, W., *et al.* (1995). A family study of obsessive–compulsive disorder. *American Journal of Psychiatry,* **152**, 76–84.
40. Tourette Syndrome International Consortium for Genetics (1999). A complete genome screen in sib-pairs affected with Gilles de la Tourette syndrome. *American Journal of Human Genetics,* **65**, 1428–36.
41. Merette, C., Brassard, A., Potvin, A., *et al.* (2000). Significant linkage for Tourette syndrome in a large French Canadian family. *American Journal of Human Genetics,* **67**, 1008–13.
42. Dickel, D.E., Veenstra-VanderWeele, J., Chiu Bivens, N., *et al.* (2007). Association studies of serotonin system candidate genes in early-onset obsessive-compulsive disorder. *Biological Psychiatry,* **61**, 322–9.
43. Arnold, P.D., Sicard, T., Burroughs, E., *et al.* (2006). Glutamate transporter gene SLC1A1 associated with obsessivecompulsive disorder. *Archives of General Psychiatry*, **63**(7), 769–76.
44. State, M., Greally, J.M., Cuker, *et al.* (2003). Epigenetic abmormalities associated with a chromosome 18(q21–q22) inversion and a Gilles de la Tourette syndrome phenotype. *Proceedings of the National Academy of Sciences of the United States of America,* **100**, 4684–9.
45. Mathews, C.A., Bimson, B., Lowe, T.L., *et al.* (2006). Association between maternal smoking and increased symptom severity in Tourette's syndrome. *Reus VI. 1: American Journal of Psychiatry,* **63**(6), 1066–73.
46. Rapoport, J.L. (1989). The neurobiology of obsessive compulsive disorder. *Journal of the American Medical Association,* **260**, 2888–90.
47. Flament, M.F., Koby, E., Rapoport, J.L., *et al.* (1990). Childhood obsessive–compulsive disorder: a prospective follow–up study. *Journal of Child Psychology and Psychiatry,* **31**, 363–80.
48. Baxter, L.R. (1998). Functional imaging of brain systems mediating obsessive-compulsive disorder. In *Neurobiology of Mental Illness* (eds. *E.J. Charney, E.J. Nestler, B.S. Bunney),* 534–47.
49. Evans, D.W., Lewis, M.D. and Iobst, E., (2004). The role of the orbitofrontal cortex in normally developing compulsive-like behaviors and obsessive-compulsive disorder. *Brain and cognition,* **55**, 220–34.
50. Baxter, L.R., Jr., Schwatrz, J.M., Bergman, K.S., *et al.* (1992). Caudate glucose metabolic rate changes with both drug and behavior therapy for obsessive-compulsive disorder. *Archives of General Psychiatry,* **49**, 681–9.
51. Swedo, S.E., Pietrini, P., Leonard, H.L., *et al.* (1992) Cerebral glucose metabolism in childhood on-set obsessive compulsive disorder. Revisualization during pharmacotherapy. *Archives of General Psychiatry,* **49**, 690–4.
52. Szeszko, P.R., MacMillan, S., McMeniman, M., *et al.* (2004). Brain structural abnormalities in psychotropic drug-naïve pediatric patients with obsessive-compulsive disorder. *Journal of the American Academy of Child and Adolescent Psychiatry,* **161**, 1049–56.
53. Giedd, J.N., Rapoport, J.L., Garvey, M.A., *et al.* (2000). MRI assessment of children with obsessive-compulsive disorder or tics associated with

streptococcal infection. *Journal of the American Academy of Child and Adolescent Psychiatry,* **157**, 281–3.
54. Russell, A., Cortese, B., Lorch, E., *et al.* (2003). Localized functional neurochemical marker abnormalities in dorsolateral prefrontal cortex in pediatric obsessive-compulsive disorder. *Journal of child and adolescent psychopharmacology,* **13**, S31–S38.
55. Peterson, B.S., Thomas, P., Kane, M.J., *et al.* (2003). Basal Ganglia volumes in patients with Gilles de la Tourette syndrome. *Archives of General Psychiatry,* **60**(4), 415–24.
56. Hyde, T., Aaronson, B., Randolph, C., *et al.* (1992). Relationship of birth weight to the phenotypic expression of Gilles de la Tourette's syndrome in monozygotic twins. *Neurology,* **42**, 652.
57. Peterson, B.S., Staib, L., Scahill, L. *et al.* (2001b). Regional brain and ventricular volumes in Tourette syndrome. *Archives of General Psychiatry,* **58**(5), 427–40.
58. Amat, J.A., Bronen, R.A., Saluja, S. *et al.* (2006). Increased number of subcortical hyperintensities on MRI in children and adolescents with Tourette's syndrome, obsessive-compulsive disorder, and attention deficit hyperactivity disorder. *American Journal of Psychiatry,* **163**(6), 1106–08.
59. Rapoport, J.L., and Fiske, A. (1998). The new biology of obsessive compulsive disorder: implications for evolutionary psychology. *Perspectives in Biology and Medicine,* **41**, 159–75.
60. Swedo, S.E., Leonard, H.L., Schapiro, M.B., *et al.* (1993). Sydenham's chorea: physical and psychological symptoms of St Vitus Dance. *Pediatrics,* **91**, 706–13.
61. Asbarh, F.R., Negrao, A.B., Gentil, V., *et al.* (1998). Obsessive-compulsive and related symptoms in children and adolesdents with rheumatic fever with and without chorea: a prospective 6-month study. *American Journal of Psychiatry,* **155**, 1122–4.
62. Murphy, T., Goodman, W., Fudge, M., *et al.* (1997). B lymphocyte antigen D8/17: a peripheral marker for childhood onset obsessive-compulsive disorder and Tourette's syndrome? *American Journal of Psychiatry,* **154**, 402–7.
63. Swedo, S.E., Leonard, H., Mittleman, B., *et al.* (1997). Identification of children with pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections by a marker associated with rheumatic fever. *American Journal of Psychiatry,* **154**, 110–12.
64. Stewart, S.E., Gellar, D.A., Jenkie, M., *et al.* (2004). Long-term outcome of pediatric obsessive-compulsive disorder: a meta-analysis and qualitative review of the literature. *Acta Psychiatrica Scandinavica*, **110**(1), 4–13.
65. Bruun, R.D., and Budman, C.L., (1993). The natural history of Gilles de la Tourette's syndrome. In *Handbook of Tourette's syndrome and related tic and behavioural disorders* (ed. R. Kurlan). Dekker, New York.
66. March, J.S., Franklin, M., Nelson, A., *et al.* (2001). Cognitive-behavioral psychotherapy for pediatric obsessive-compulsive disorder. *Journal of clinical child psychology*, **30**(1), 8–18.
67. March, J.S. and Mulle, K. (1998). OCD in Children and Adolescent. *A Cognitive-Behavioral Treatment Manual.* New York, Guilford Press.
68. Barrett, P.M., Healy-Farrell, L. and March, J.S. (2004). Cognitive-behavioral family treatment of childhood obsessive-compulsive disorder: a controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry,* **43**(1), 46–62.
69. Barrett, P., Farrell, L., Dadds, M., *et al.* (2005). Cognitive-behavioral family treatment of childhood obsessive-compulsive disorder: long-tern follow-up and predictors of outcome. **44**(10), 1005–14.
70. Verdellen, C., Keijsers, G., Cath, D., *et al.* (2004). Hoogduin Exposure with response prevention versus habit reversal in Tourettes's syndrome: a controlled study, *Behaviour Research and Therapy,* **42**, 501–11.
71. O'Connor, K.P., Brault, S., Robillard, J., *et al.* (2001). Stip Evaluation of a cognitive-behavioural program for the management of chronic tic and habit disorders. *Behaviour Research and Therapy,* **39**, 667–81.
72. Cook, C.R. and Blacher, J. (2007). Evidence-Based Psychosocial Treatments for Tic Disorders. *Clinical Psychology: Science and Practice,* **14**(3), 252–67.
73. Flament, M.F., Rapoport, J.L., and Kilts, C. (1985). A Controlled Trial of Clomipramine in Childhood Obsessive Compulsive Disorder. *Psychopharmacology Bulletin*, **21**(1), 150–152.
74. Leonard, H.L., Swedo, S.E., Rapoport, J.L., *et al.* (1989). Treatment of Obsessive-Compulsive Disorder With Clomipramine and Desipramine in Children and Adolescents. A Double-Blind Crossover Comparison. *Archives of General Psychiatry,* **46**(12), 1088–92.
75. DeVeaugh-Geiss, J., Moroz, G., Biederman, J., *et al.* (1992). Clomipramine Hydrochloride in Childhood and Adolescent Obsessive-Compulsive Disorder–a Multicenter Trial. *Journal of the American Academy of Child and Adolescent Psychiatry,* **31**(1), 45–9.
76. Riddle, M.A., Scahill, L., King, R.A., *et al.* (1992). Double-Blind, Crossover Trial of Fluoxetine and Placebo in Children and Adolescents With Obsessive-Compulsive Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry,* **31**(6), 1062–69.
77. Riddle, M.A., Reeve, E.A., Yaryura-Tobias, J.A., *et al.* (2001). Fluvoxamine for Children and Adolescents With Obsessive-Compulsive Disorder: a Randomized, Controlled, Multicenter Trial. *Journal of the American Academy of Child and Adolescent Psychiatry,* **40**(2), 222–9.
78. Liebowitz, M.R., Turner, S.M., Piacentini, J., *et al.* (2002). Fluoxetine in Children and Adolescents With OCD: a Placebo-Controlled Trial. *Journal of the American Academy of Child and Adolescent Psychiatry,* **41**(12), 1431–38.
79. Geller, D.A., Wagner, K.D., Emsile, G., *et al.* (2004). Paroxetine Treatment in Children and Adolescents With Obsessive-Compulsive Disorder: a Randomized, Multicenter, Double-Blind, Placebo-Controlled Trial. *Journal of the American Academy of Child and Adolescent Psychiatry,* **43**(11), 1387–96.
80. Flament, M.F., Geller, D. and Blier, P. (2007). Specifi cities of treatment in pediatric obsessive-compulsive disorder. *CNS Spectrum,* **12**(2), 43–57.
81. Geller, D.A., Biederman, J., Stewart, S.E., *et al.* (2003a). Which SSRI? A meta-analysis of pharmacotherapy trials in pediatric obsessivecompulsive disorder. *American Journal of Psychiatry,* **160**(11), 1919–28.
82. De Haan, E., Hoogduin, K.A., Buitleaar, J.K., *et al.* (1998). Behavior Therapy Versus Clomipramine for the Treatment of Obsessive-Compulsive Disorder in Children and Adolescents *Journal of the American Academy of Child & Adolescent Psychiatry,* **37**(10), 1022–9.
83. Pediatric OCD Treatment Study (POTS). (2004) Cognitive behavior therapy, setraline, and their combination for children and adolescents with obsessive-compulsive disorder: the Pediatric OCD Treatment Study (POTS) randomized controlled trial. *JAMA*, **292**(16), 1969–76.
84. Ashbar, F.R., Castillo, A.R., Ito, L.M., *et al.* (2005). Group cognitive-behavioral therapy versus setraline for the treatment of children and adolescents with obsessivecompulsive disorder. *Journal of the American Academy of Child and Adolescent Psychiatry,* **44**(11), 1128–36.
85. Blier, P., Habib, R. and Flament, M.F. (2006). Pharmacotherapies in the management of obsessive-compulsive disorder. *Canadian Journal of Psychiatry,* **51**(7), 417–30.
86. Leonard, H.L., Meyer, M.C., Swedo, S.E. *et al.* (1995). Electrocardiographic changes during desiprime and clomipramine treatment in children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry,* **34**(11), 1460–8.
87. Varley, C.K. (2001). Sudden death related to selected tricyclic antidepressants in children: epidemiology, mechanisms and clinical implications. *Pediatric Drugs,* **3**(8), 613–27.
88. Hammerness, P.G., Vivas, F.M. and Gellar, D.A. (2006). Selective serotonin reuptake in pediatric psychopharmacology: a review of the evidence. *Journal of pediatrics*, **148**(2), 158–65.
89. Nilsson, O., Jakobsen, A.M., Bernhardt, P., *et al.* (2004). Importance of vesicle proteins in the diagnosis and treatment of neuroendocrine tumors. *Annals of the New York Academy of Sciences,* **1014**, 240–83.

90. Hammad, T.A., Laughren, T. and Racoosin, J. (2006). Suicidality in pediatric patients treated with antidepressant drugs. *Archives of General Psychiatry,* **63**(3), 332–9.
91. Bridge, J.A., Iyengar, S., Salary, C.B., *et al.* (2007). Clinical response and risk for reported suicidal ideation and suicide attempts in pediatric antidepressant treatment: a meta-analysis of randominzed controlled trials. *JAMA*, **297**(15), 1683–96.
92. Shapiro, E., Shapiro, A.K., Fulop, G., *et al.* (1989). Controled study of haloperidol, pimozide, and placebo for the treatment of Gilles de la Tourette's syndrome. *Archives of General Psychiatry,* **46**, 722–30.
93. Dion, Y., Annable, L., Sandor, P., *et al.* (2002). Risperidone in the treatment of tourette syndrome: a double-blind, placebo-controlled trial. *Journal of Clinical Psychopharmacology,* **22**, 31–9.
94. Scahill, L., Leckman, J.F., Schultz, R.T., *et al.* (2003). A placebo-controlled trial of risperidone in Tourette syndrome. *Neurology,* **60**, 1130–5.
95. Bruggemann, R., van der Linden, C., Buitelarr, J.K., *et al.* (2001). Risperidone versus pimozide in Tourette's syndrome: a comparative double-blind parrallel group study. *Journal of Clinical Psychaitry,* **62**, 50–56.
96. Gilbert, D.L., Batterson, J.R., Sethuraman, G., *et al.* (2004). Tic reduction with risperidone versus pimozide in a randomized, doubleblind, crossover trial. *Journal of the American Academy of Child and Adolescent Psychiatry,* **43**, 206–14.
97. Desta, Z., Kerbusch, T. and Flockhart, D.A. (1999). Effect of clarithromycin on the pharmacodynamics of pimozide in healthy poor and extensive metabolizers of cytochrome P450 2D6 (CYP2D6). *Clinical pharmacology and therapeutics,* **65**, 10–20.
98. American Academy of Child and Adolescent Psychiatry (1997). Practice parameters for the assessment and treatment of children, adolescents, and adults with attention-deficit hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry,* **36**, 85S–121S.
99. Tourette Syndrome Study Group (2002). Treatment of ADHD in children with tics: a randomized controlled trial. *Neurology,* **58**, 527–36.
100. Geller, D.A., Biederman, J., Stewart, S.E. *et al.* (2003b). Impact of Comorbidity on Treatment Response to Paroxetine in Pediatric Obsessive-Compulsive Disorder: Is the Use of Exclusion Criteria Empirically Supported in Randomized Clinical Trials? *Journal of Child and Adolescent sychopharmacology,* **13** (Supplement 1), S19–S29.
101. American Academy for Child Adolescent Psychiatry (1998). Practice parameters for the assessment and treatment of children and adolescents with obsessive-compulsive disorder. *Journal of the American Academy of Child and Adolescent Psychiatry,* **37**(10), 27S–45S.
102. Thomsen, P.H. (2004). Risperidone augmentation in the treatment of severe adolescent OCD in SSRI-refractory cases: a case series. *Annals of Clinical Psychiatry,* **16**(4), 201–7.
103. Baldwin, D.S., Anderson, I.M., Nutt, D.J., *et al.* (2005). Evidence based guidelines for the pharmacological treatment of anxiety disorders: recommendations from the British Association for Psychopharmacology. *Journal of psychopharmacology*, **19**, 567–96.
104. Reinblatt, S., Walkup, J.T. (2005). Psychopharmacological treatment of pediatric anxiety disorders. *Child and Adolescent Psychiatric Clinics of North America*, **14**, 877–908.
105. Fitzgerlad, K.D., MacMaster, F.P., Paulson, L.D., *et al.* (1999). Neurobiology of childhood obsessive-compulsive disorder. *Child and Adolescent Psychiatric Clinics of North America,* **8**(3), 533–75.
106. March, J.S., Franklin, M.E., Leonard, H., *et al.* (2007). Tics moderate treatment outcome with sertraline but not cognitive-behavior therapy in pediatric obsessive-compulsive disorder. *Biological Psychiatry,* **61**(3), 344–7.
107. Perlmutter, S.J., Leitmen, S.F., Garvey, M.A., *et al.* (1999). Therapeutic plasma exchange and intravenous immunoglobulin for obsessivecompulsive disorders in childhood. *Lancet,* **354**, 1153–8.
108. Snider, L.A., Lougee, L., Slattery, M., *et al.* (2005). Antibiotic prophylaxis with azithromycin or penicillin for childhood-onset neuropsychiatric disorders. *Biological Psychiatry,* **57**(7), 788–92.
109. Storch, E.A., Murphy, T.K., Geffken, G.R., *et al.* (2006). Cognitive-behavioral therapy for PANDAS-related obsessive-compulsive disorder: findings from a preliminary waitlist controlled open trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 45(10), 1171–8.

9.2.9 Sleep disorders in children and adolescents

Gregory Stores

Introduction

It was argued in Chapter 4.14.1 that sleep disorders medicine should be viewed as an integral part of psychiatry, whatever the age group of patients, because of the various close connections between sleep disturbance and psychological disorders seen in clinical practice. This is certainly the case regarding child and adolescent psychiatry in view of the high rates of psychiatric disorder of which sleep disturbance is often a part, and also the frequent occurrence of sleep disorders in young people with potentially serious developmental effects of a psychological and sometimes physical nature. The temptation to view children's sleep disorders as merely transitory problems, mainly in infancy, encountered by many parents and of no lasting or serious significance, should be resisted. This may be true for some families but is frequently not the case in others.

The following account summarizes sleep disorders in childhood and adolescence. *Familiarity is assumed with the earlier accounts of sleep disorders in adults (4.14.1), including the introduction to that section which covers basic aspects of sleep and other fundamental issues.*

Sleep and sleep disorders in children compared to adults

In spite of the fact that much has now been discovered about the special characteristics of sleep disorders occurring at an early age, very little of this information has found its way into the training of paediatricians, child and adolescent psychiatrists, psychologists, or other professionals involved in the care of children. This must mean that many treatment and preventive possibilities are missed.

The solution does not lie simply in extending the practice of adult sleep disorders medicine to children. Children are not miniature adults and they need special approaches reflecting their many differences from older patients. These differences extend from basic features of sleep to various aspects of sleep disorders.

Sleep physiology

Profound changes take place during childhood in basic sleep physiology although many are complete by about 6–12 months of age. In general, there is a progression towards differentiation and organization of conventionally defined sleep states, shorter sleep time, less napping, less slow wave sleep (**SWS**), and longer sleep cycles.

Specific aspects of particular clinical importance are as follows:

- Typical sleep duration (including naps) at different ages as shown in Table 9.2.9.1.

Table 9.2.9.1 Average sleep requirements at different ages

Term birth	17 h
1 year	14 h
2 years	13 h
4 years	12 h
10 years	10 h
Adolescence	9 h plus*

*Many adolescents are thought to obtain far less sleep than this.

- The body clock controlling (amongst other processes) the circadian sleep–wake cycle has become established by about 6 months.
- Rapid eye movement (**REM**) sleep is prominent in early infancy, perhaps reflecting its role in brain maturation and early learning, and possibly explaining why sleep is fragile at this stage.
- In comparison, by early childhood **SWS** is especially pronounced. This predisposes children of that age to arousal disorders (e.g. sleepwalking) which arise from **SWS**.
- Between about 5 years and puberty, overnight sleep is especially sound and alertness is maximal during the day. Various conditions causing excessive daytime sleepiness in adults (e.g. narcolepsy) may not have this effect in children because of this increased alertness. However, overnight sleep may become extended.
- In contrast, adolescence is characterized by an increase in daytime sleepiness. The amount of **SWS** decreases, the sleep phase is physiologically delayed and, with the onset of puberty, there is no longer the decrease in physiological sleep requirements seen progressively at earlier ages. The combination of these factors and strong influences to stay up late (especially at weekends but perhaps also during the week) for social and recreational purposes frequently causes unsatisfactory sleep–wake patterns.

Parental influences

The influence of parents is seen throughout children's sleep disorders medicine.

- Especially in the case of young children, parents' perceptions usually determine whether there is a sleep problem. The same sleep pattern or behaviour may be a problem to one family but not another. Factors influencing parental attitudes include their expectations, family and cultural practices (e.g. regarding parents and children sleeping together), and their own emotional state. Sometimes parents can be reassured that what they think is a serious problem about their child's sleep is, in fact, within the normal range. The view taken of the situation might be the result of parental psychiatric illness needing attention in its own right; children of mothers with an affective illness have been shown to have an increased rate and severity of sleep problems although the nature of the connection is debatable.
- Conversely, parents may not seek help for their child's sleep when they ought to do. They may be unaware of the problem, indifferent, or they may mistakenly believe that the child's sleep problem is inevitable and untreatable. This mistaken view is sometimes expressed by parents of children with a learning disability (intellectual disability) whose sleep problems can be particularly severe yet amenable to treatment.
- Parental practices are commonly the reason why a child's sleep problem develops or is maintained. Early child-rearing practices determine sleep–wake patterns which can be delayed or disrupted by over-conscientious night-time feeding in infancy, failure to set limits on bedtime activities, or inconsistency (see later). Sleep disorders of physical origin may be complicated in these ways and exacerbated. It follows that treatment of many sleep disorders relies heavily on correcting parenting practices.
- Sometimes parents are not motivated to improve their child's sleep for reasons that may be difficult to influence. For example, a child's presence in the parental bed may be welcome by one partner as a means of distancing himself or herself from the other at night. Families of handicapped children may lose their extra state financial allowance if their child's sleep problems are successfully treated.
- The child's basic attitude to sleeping is also influenced by its parents. Wider cultural factors are important but, within westernized societies, the child's attitudes to going to sleep and being separated from its parents at night are strongly influenced by their ability to settle the child without being anxious about the separation. Children depend on their parents to provide positive attitudes to sleeping and to avoid negative associations such as disputes, punishment, and rejection.
- Especially in the early years, most children need their parents' help in coping with night-time separation from them, and the potentially frightening experience of the dark or their own thoughts and fantasies. Infants need the comfort of physical contact. Toddlers are helped by bedtime routines and comforting 'transitional objects', and encouragement to become 'self-soothing' so that they can fall asleep without their parents' presence and attention (see later). Parents' ability to provide such help depends on their personality and sensitivity and mental state and perhaps their cognitions about their child's sleep based partly on their own experiences in childhood. Hopefully, older children and adolescents become increasingly independent.

Effects on parenting and the family

The effects of a child's persistent sleep disturbance on family life, including its possible influence on parenting skills, is another important dimension.

- Mothers of children with a learning disability and severe sleep problems are reported to be more irritable, concerned about their own health, and less affectionate towards their children, with less control and increased use of punishment compared with mothers of such children without sleep problems. Similarly, associations have also been suggested between sleeplessness in toddlers in the general population and family problems, including marital discord and possibly physical abuse of the child.
- Family tensions are likely to increase when diagnosis of the child's sleep disorder is delayed or inaccurate, or when effective treatment is not provided.
- Some reports have suggested that successful treatment of the child's sleep problems generally leads to improvement in the

mother's mental state, confidence in her own parenting ability, her relationship with the child, and also the child's behaviour. Wider aspects of family function, including effects on siblings, have received little attention.

Developmental effects of sleep disturbance

- These parental and wider family issues indicate ways in which a child's psychological and social development can be affected by persistent sleep disturbance. In addition, children can be distressed by their experience of sleep disorder phenomena. Examples include night-time fears (which may be intense) alarming hypnagogic imagery, or sleepwalking and sleep (night) terrors which can be embarrassing, especially if they occur away from home. Excessive daytime sleepiness often leads to educational problems and can produce extreme reactions such as the denial, aggression or depression described in narcolepsy, or accidents and substance abuse in adolescence.
- In addition to these largely indirect ways in which a child's sleep disorder may have psychological effects, sleep disturbance can produce direct effects on mood, behaviour, and cognitive function. The developmental consequences might become severe if not arrested at an early age.
- Adolescents appear to be at particular risk of sleep loss and its possible psychological consequences, i.e. depressed mood, anxiety, behaviour problems, alcohol abuse, and even attempted suicide, as well as lower academic performance. The causal relationship between sleep loss and these problems, however, have yet to be fully established. The same is true of the outcome of attempts to correct this sleep loss by various means.
- Even impairment of physical growth is associated with sleep disturbance. Failure to thrive is a recognized possible consequence of early onset obstructive sleep apnoea (**OSA**) and possibly other severe and persistent sleep disturbance, perhaps as a result of reduced slow wave sleep (**SWS**) with which the production of growth hormone is closely linked.
- Other possible physical consequences of sleep disruption includes impaired immunity and endocrine disorders.

Patterns of occurrence of sleep disorders

- Some sleep behaviours which are developmentally usual in children are abnormal in adults and require investigation. Examples are bedwetting and repeated napping. Certain sleep disorders are seen exclusively in children (e.g. sleeplessness caused by infantile colic). Others, such as settling problems and confusion arousals, occur primarily in children (see later).
- Sleeplessness caused mainly by child-rearing practices is particularly common in early childhood. That attributable to the delayed sleep phase syndrome (see later) is considered to be particularly common in adolescence.
- Many of the parasomnias (such as headbanging, sleepwalking, or sleep terrors) are more common in childhood where, generally, they represent a temporary developmental phase without pathological significance. The same behaviours in adults might be more likely to be manifestations of psychological problems requiring exploration.
- Some sleep disorders thought to be confined to adulthood are now recognized in children. While much attention has been paid to OSA in adults, it is now thought that at least 2 per cent of children have this condition to some degree. Restless leg syndrome (**RLS**) and periodic limb movements in sleep (**PLMS**) are now known to occur not uncommonly in children. The RLS may explain some cases of 'growing pains'. PLMS has been implicated as a cause of poor quality sleep resulting (as in other forms of sleep disturbance) with daytime attention deficit hyperactivity disorder (**ADHD**) type of symptoms. Narcolepsy starts by the age of 15 years in at least one-third of cases. Even REM sleep behaviour disorder (once thought to be confined to elderly males), or something similar, has been reported in children and adolescents.

Manifestations of sleep disorders

- The clinical features of basically the same sleep disorder can be very different in children compared with older people. The overall behavioural effects of excessive sleepiness in adults are a reduction of physical and mental activity. In contrast, its effects in young children can be increased activity with irritability, tantrums, or other behavioural difficulties. Some examples of ADHD are thought to be the result of sleep disorders (OSA, PLMS, or circadian sleep–wake rhythm disorder) with improvement in the difficult behaviour following treatment of the sleep disorder.
- OSA illustrates the important differences between children and adults, not only in the clinical manifestations of a particular sleep disorder but also in the underlying cause and treatment needs. Similarly the many manifestations of narcolepsy in childhood may be very far removed from the classical narcolepsy syndrome in adults, at least in its fully developed form. The same sleep disorder may also show different physiological features according to age. Diagnostic criteria (e.g. for OSA and narcolepsy) derived from polysomnographic (PSG) studies in adults do not necessarily apply in children and may well need modification.

Misinterpretation of children's sleep disorders

Chapter 4.14.1 contains an account of the fundamental issue that, especially if clinicians are unfamiliar with the manifestations and consequences of the many sleep disorders now documented in the second edition of the International Classification of Sleep Disorders (ICSD-2), there is a serious risk that these disorders will be misconstrued as something else (or even overlooked completely). The examples given include a number of particular relevance to practice in child psychiatry and paediatrics.[1]

Treatment and prognosis

- Because of the aetiological differences discussed earlier, especially parental involvement, treatment often needs to be very different in children compared to adults. Appropriate behavioural approaches usually entail alterations to parenting practices designed to be acceptable and feasible in each individual family. Other forms of treatment, including chronobiological measures (such as adjustment of sleep schedules from the delayed sleep phase syndrome in adolescence) usually require considerable parental involvement. The same is true of the general sleep

hygiene principles described in Chapter 4.14.2. Explanation and (where appropriate) reassurance for the child and parents is an essential part of any treatment and may be effective in their own right without the need for more specific measures. As in adults, medication has a limited part to play overall.

- An optimistic point of view can be taken of the treatment of most children's sleep disorders because children's sleep is usually more amenable to change than that of adults where the factors underlying the sleep problem may well have become well established and complicated, as in many cases of chronic insomnia. However, treatment needs to be chosen carefully and implemented properly, and parents' confidence in the recommended measures, and their willingness and ability to play their part in treatment, are an important determinant of success or failure. In some instances, it is not possible to implement a treatment programme for the child until parents themselves have been helped (e.g. by treatment for a depressive illness) or problems in the family as a whole have been resolved.

Assessment

The various means by which sleep disorders might generally be detected and assessed are described in Chapter 4.14.1. These subjective and objective approaches need to be modified for use with children because of the involvement of parents, developmental factors, and the differences between children and adults regarding clinical manifestations and diagnostic criteria.

The detection of sleep problems can be improved by routinely asking basic screening questions as part of the history-taking in any child:

- Does the child have difficulty getting to sleep or staying asleep?
- Is there excessive sleepiness during the day?
- Are there episodes of abnormal behaviour or experiences at night?

Positive answers to any of these questions call for a detailed sleep history.

Sleep history and general review

This is the cornerstone of sleep assessment. Unfortunately, history-taking schedules are usually perfunctory in the attention they pay to sleep and its possible disorders. Parents and also the child (if old enough) should be interviewed and the reasons for any disparities considered. Sometimes sibs or teachers can provide important additional information. The main aspects that should be covered are as follows:

- Current sleep problems and their evolution.
- Past treatments and their effects.
- Review of the child's current 24 h sleep–wake cycle (see Table 9.2.9.2) in order to determine in particular
  (a) duration of sleep
  (b) quality of sleep (continuous or disrupted)
  (c) timing of sleep
  (d) features suggestive of specific sleep disorders (e.g. breathing difficulty or jerking limbs).
- Sleep environment and arrangements.
- Development of the child's sleep patterns and problems.

Table 9.2.9.2 Review of child's 24 h sleep–wake pattern (modified according to child's age)

Evening
Time of evening meal
Other evening activities
Going to bed
Preparation for bed, by whom
Time of going to bed
Reluctance to go at required time, parents' reactions
Fears, rituals
Wanting to sleep with someone, other comforts
Time taken to fall asleep, other experiences during that period
When asleep
Wakings, frequency, causes ability to return to sleep
Episodic events, exact nature, timing, frequency
Other behaviours during sleep, e.g. snoring, restlessness, bedwetting
Parents' reaction to night-time events
Waking
Wakes spontaneously or needs to be woken up
Time of final waking
Total duration of sleep period
Longest period of uninterrupted sleep
On waking: preoccupations, mood, feeling of being refreshed, other experiences
Difficulty getting out of bed, time of getting out of bed
Daytime
Sleepiness, naps
Lethargy
Mood
Overactivity
Concentration and performance
Other unusual episodes

- General review of possible sleep symptoms.
- Family history of sleep disorder or other conditions.

A **sleep questionnaire** completed by parents before the interview can provide a useful outline account of these and other aspects.[2]

Additional parts of the overall review of children with a sleep problem that are important in order to identify possible contributory factors are as follows:

- Developmental history including developmental delays, illnesses, or significant events at school or within the family.
- Review of physical health.
- Physical examination.
- Assessment of behaviour and emotional state.
- Family history and circumstances.

Following the initial consultation, a **sleep diary**, kept over a period of 2 or more weeks, can be particularly useful. This provides a more complete and balanced view than that obtained especially from fraught parents likely to give a distorted or unbalanced retrospective account.

Special investigations (see also Chapter 4.14.1)

These depend on the nature of the sleep problem:

- Indications for PSG are essentially the same as for adults.

- The use in children of multiple sleep latency tests (**MSLT**) as an objective measure of sleepiness is hampered by the absence of good normative data at different ages. However, in school-aged children, about 16–18 min to fall asleep is considered normal; less than this might indicate significant daytime sleepiness which is also indicated by falling asleep in three or more of the naps. Nevertheless, in the presence of sleep disorders usually characterized by excessive sleepiness, MSLT results can be normal in late childhood because of the naturally enhanced daytime wakefulness at that age.
- **Actigraphy**, which provides information unobtrusively on basic sleep–wake patterns, is well established for children of all ages.
- **Other possible measures** include toxic screening and the special tests mentioned earlier in Chapter 4.14.1.

Children at special risk of sleep disturbance

The prevalence of children's sleep disorders is not known with any accuracy, even for those which are severe and persistent. A number of methodological problems make it difficult to collect accurate figures and no really vigorous attempt has yet been made to overcome them. It seems that 20–30 per cent of children from infancy to adolescence have sleep problems that are considered significant by them or their parents.

The occurrence of sleep problems exceeds this overall rate considerably in various categories within the general population as a whole, and also in certain clinical subgroups. Determining the exact nature and cause of the sleep problems in these high risk groups (and also in any other affected children), with a view to successful treatment, is important because of the possible adverse effects on the child in a family that have just been discussed. Behavioural sleep problems probably predominate throughout these high risk groups but sleep disorders of a different nature may well be encountered instead or as well as those of behavioural origin.

Children in the general population

- Reference was made earlier to the fact that children in general appear to be particularly prone to different types of sleep disorders at certain ages of development i.e. early infancy, early childhood, and adolescence.
- Adverse psycho-social circumstances are also associated with increased risk of childhood sleep problems. The importance of such factors as the degree of organization in family life, parental concern, child-rearing practices, and the mental health of parents was stressed earlier. High rates of various sleep problems, as well as other psychological difficulties, have been reported in homeless children.

Children with psychiatric disorders

High rates of sleep disturbance have been described in child psychiatric groups in general compared with other children, and in specific psychiatric disorders.

- Various sleep problems, including panic attacks, have been described in **anxious children** in general including those with **panic disorders**. Similarly, many types of sleep problem (including nightmares and other disturbed nocturnal episodes, excessive daytime sleepiness, and bedwetting) have been reported to be particularly frequent in **traumatized children** including those who have suffered burn injury, abuse, or road traffic accidents. Treatment of the sleep disturbance has appeared to improve their emotional state but further research is needed to asses the therapeutic contribution of specific treatment for the sleep disorder as part of the overall care of the traumatized children.
- Difficulty in sleeping is the main complaint in children and in adolescents with severe **depressive disorders** but many complain of excessive sleepiness, possibly because of difficulty getting to sleep and/or poor quality sleep.
- Parental reports of sleep problems in children with **ADHD** are very common. Parental impressions can be distorted but preliminary objective evidence also suggests that persistent sleep disturbance is common and sometimes important as the primary cause (or a significant contributory factor) rather than simply a consequence of ADHD. It was mentioned earlier that ADHD symptoms have sometimes been attributed to definitive sleep disorders in which sleep quality is impaired, with improvement in ADHD symptoms following treatment of the sleep disorder. Preliminary studies of sleep physiology or other objective aspects of sleep in children have also produced evidence of sleep abnormalities. Even where ADHD is attributable to other factors, sleep disruption is likely to worsen a child's behaviour, meriting treatment in its own right wherever possible.
- **Other psychiatric disorders** in which different types of sleep disturbance is reported to be prominent are autism (including circadian sleep–wake rhythm disorders and Asperger's syndrome, tic disorders including Tourette syndrome (sleeplessness and parasomnias), and obsessive–compulsive disorders (poor quality sleep).

  Sleep complaints are also prominent in the chronic fatigue syndrome. As mentioned earlier, disruption by frequent awakenings (not obviously attributable to daytime inactivity) has been described in teenagers with this condition suggesting that daytime symptoms might be at least partly attributable to poor quality sleep. Occasionally Munchausen's syndrome by proxy have come to light in the form of parental complaints of a sleep disturbance. Reports of the sleep of conduct disordered children are in keeping with the expectation that their sleep is disturbed because of their adverse or disorganized home and social circumstances, and general way of life.
- Apart from psychiatric disorders themselves, **medications** used in their treatment may affect sleep. Stimulant medication for ADHD appears to cause sleeping difficulties in some children but some children with ADHD may settle to sleep more readily even if their medication is given later in the day because this improves their bedtime behaviour. See Chapter 4.14.1 for other possible medication effects on sleep.

Children with a learning disability or other neurological disorder

Particularly high rates of sleep disturbance has been consistently reported in children with a **learning disability**. The disturbance is often severe, poorly managed and, therefore, persistent.

- Sleep problems in this group are often behavioural in origin (largely attributed to parenting practices), arising from often

understandable over-permissiveness, inconsistency, or parents' inability to set limits on their child's behaviour because of their own emotional state or excessive demands on their time. Other physical sleep problems include some chronic physical conditions. OSA features prominently in various specific learning difficulty conditions such as Down syndrome, the mucopolysaccharidosis and fragile X syndrome. Epilepsy can also play an important role as well as other co-morbid conditions.

- Sleep problems are also widely reported in children with **neurodegenerative disorders**, e.g. Rett's syndrome and other neurological disorders such as head injury. Again, behavioural factors might be partly the reason, although interference with sleep mechanisms also seems likely, at least in the advanced stages of the disease.

Children with other chronic physical illness

Acute physical illnesses disturb sleep but only for the duration of the illness in most cases. By comparison, chronic illnesses are commonly complicated by long-standing sleep disturbance caused in various ways (see Chapter 4.14.1 for medical causes of sleep disorder some of which apply to paediatric cases).

Main sleep problems: sleeplessness

The second edition of the International Classification of Sleep Disorders was outlined in Chapter 4.14.1. The following selective account of sleep disorders in children and adolescents is organized according to the three main types of sleep complaint: sleeplessness, excessive sleepiness, and the parasomnias. Childhood psychiatric and medical conditions in which sleep disturbance is a prominent feature has already been mentioned. Emphasis is placed on the differential diagnosis of sleep complaints and also on points of particular relevance to psychiatric practice.

The breakdown of problems and disorders according to age should not be interpreted too strictly as there is overlap between the different age groups. In addition to specific treatments mentioned for particular sleep disorders, the promotion of adequate sleep, regular sleep habits, and the other sleep hygiene principles referred to in Chapter 4.14.1 are important. Evidence for the effectiveness of psychological treatments for sleeplessness in children is reviewed in detail elsewhere.[3]

Infants

Ways of preventing or dealing with babies' sleep problems are rarely taught to parents or prospective parents, with the result that many suffer needless sleep loss and distress because the child does not sleep well. It is important to encourage good sleep habits from the start to avoid bad sleep habits later on. There are certain general guidelines for achieving this, admitting that babies vary temperamentally in their response to recommendations and parents vary in their ability to adhere to them. The main basic principles are as follows:

- Establishing a consistent 24 h routine, including a bedtime routine that provides cues that is timed to go to sleep.
- Not prolonging night-time feeding beyond the age (about 6 months) when the baby's body clock has developed enough to confine feeding to daytime.
- Teaching the baby to fall asleep alone so that when he or she wakes in the night (a natural occurrence at all ages) it will be possible to fall asleep again without requiring parental attention ('**self-soothing**').
- Establishing a clear difference in the infant's experience between day and night to help to develop his/her body clock which controls sleep and wakefulness.
- Ensuring the environment is conducive to sleep.

Safety measures to reduce the risk of the infant coming to harm at night from suffocation, or other breathing problems associated with sudden infant death syndrome (**SIDS**), should also be part of parent education about sleep.[4] Main recommendations are: having the infant sleep on his/her back and on a firm mattress that will not obstruct breathing, ensuring that his or her face cannot be covered during the night, ensuring the bedroom is smoke free, and avoiding co-sleeping if either parent has consumed alcohol or has taken medication or other substances with a sedative affect. Also, the baby should not be overheated at night.

Toddlers and pre-school children

About 30 per cent of children of this age present a problem of recurrently not going to bed at the required time, and/or waking repeatedly at night and demanding their parents' attention including coming into their bed. Medical factors must be excluded but the usual explanations are behavioural especially:

- Anxiety about separating from parents at night
- Unhelpful associations with going to bed, e.g. stimulating activities within the bedroom, threats, or recriminations
- Inadequate limit setting on bedtime or night-time behaviour
- Failure to require self-soothing ways of coping with night waking.

Behavioural methods of treating these problems can be very effective, even in severe and long-standing cases, including children with developmental disorders such as learning disability or autism, providing the treatment programme is implemented properly. The main methods used include graded changes and desensitization rather than leaving the child to cry (a quickly effective measure but one which is unacceptable to many parents).

School-age children

Some of the causes of sleeplessness in pre-school children still apply in older children but other factors become more relevant with increasing age:

- **Night-time fears** are common from very early childhood onwards although, in keeping with cognitive development, the content of the fears changes from aspects of the immediate environment (e.g. shadows or noises) through imaginary objects (ghosts, monsters) or the dark, to more realistic and specific fears concerning the child's own health. Such fears are usually transient and require only reassurance and comfort until they cease.

  In some children the fears are so intense and persistent that they reach phobic proportions and need special attention. The cause of the fear should be investigated. The night-time fear might be one aspect of an anxiety state, including post-traumatic stress disorder in which case the child might also suffer from nightmares.

The content of the fear or nightmare might be revealing, suggesting abuse, for example. Other sleep disturbances (e.g. alarming hypnagogic hallucinations) may be the cause of the night-time fears. The child's reluctance to go to bed because he or she is genuinely afraid must be distinguished from pretending to be afraid as a delaying tactic.

Behavioural treatment is said to be very effective in cases of severe night-time fears. The child with night-time fears should be helped by positive associations with bedtime and by not going to bed so early that he or she lies awake in a fearful state.

- Even without night-time fears, a child will be unable to settle to sleep if **bedtime is too early**. Like some adults and even other species, children often have an evening period of intense wakefulness and activity before they begin to relax in preparation for sleep. A child is physiologically unable to sleep if put to bed in this 'forbidden zone'. Instead the sequence of events leading up to bedtime should be arranged so that the child goes to bed when 'sleepy tired'.
- **Worry and anxiety** about daytime matters such as school progress may cause difficulty in getting to sleep or staying asleep. The original source of concern may no longer exist but the difficulty falling asleep may persist because the child has developed the habit of lying awake in bed in an agitated state ('conditioned insomnia'). Sympathetic discussion of the child's worries, attention to the source of concern if possible, and ways of helping the child to relax at night, are generally thought to help. More specific psychiatric measures will be needed if the child has an anxiety or depressive disorder, or if there is evidence of serious problems within the family.
- '**Childhood onset insomnia**' or **'idiopathic insomnia'** refers to a lifelong difficulty sleeping not attributable to environmental, emotional, or medical factors and therefore of constitutional origin. The condition is usually diagnosed retrospectively in adult life.
- **Early morning waking** (i.e. when a child habitually wakes very early, does not return to sleep, and is noisy or demands attention) can be very distressing to parents, and disruptive to the whole family. In pre-school children, early waking may be the result of excessive or otherwise inappropriate napping, but at a later age the problem may be part of the advanced sleep phase syndrome. In this disorder the child's bedtime and sleep onset is so early that his or her sleep requirements have been met well before other members of the family wake in the morning. Gradual resetting of the child's sleep onset time is required.

  In older children and adolescents early morning wakening may be part of an anxiety or depressive disorder. Otherwise, the child may have been woken too early by noise or other environmental factors which intrude into his or her sleep.

Adolescents

High rates of insomnia have been consistently reported in adolescents. The change from the highly efficient sleep of pre-pubertal children to less satisfactory sleep in adolescence was mentioned earlier, including the biological influences in this change. The psychological and social demands and stresses of adolescence further conspire to disrupt sleep patterns. Worries, anxiety, and depression are commonly quoted reasons for not being able to sleep at this age. Nicotine, alcohol, and caffeine-containing drinks, as well as illicit drug use, are additional possible influences.

Difficulty getting off to sleep is often part of the delayed sleep phase syndrome which is reported to be particularly common in adolescence. In this condition (which will be considered further in relation to excessive sleepiness as this is often the major complaint) there is a physiological inability to go to sleep until much later than the required time because of a shift in the sleep phase. The adolescent's reluctance to go to bed earlier (or the bedtime struggles of parents with younger children with this disorder) are often misinterpreted as 'difficult' behaviour. Instead of recriminations and attempts to set limits, the timing of the sleep phase needs to be reset by so-called chronotherapeutic means.

Main sleep problems: excessive daytime sleepiness

Despite the evidence that it is a common problem,[5] sleepiness remains neglected in child and adolescent psychiatry and in paediatrics:

- Part of the explanation is that sleepiness is not usually viewed as a medical problem by parents, teachers, and children themselves, the symptoms being misperceived as laziness or disinterest. Otherwise, they may be interpreted as depression or even limited intelligence.
- Another difficulty is that excessive sleepiness can take various forms including prolonged overnight sleep or inappropriate periods of sleep during the day.
- Extreme degrees of sleepiness will cause a reduction of activity at any age, but lesser degrees in children may produce irritability, over-activity, restlessness, poor concentration, impulsiveness, or aggression. Explanations of such behaviours other than sleep loss or disturbance (such as ADHD) are more likely to be considered, as mentioned earlier.
- A high level of daytime alertness in older pre-pubertal children may be sufficient to offset a tendency to sleep, providing a different clinical picture of sleepiness but not seen at a younger or older age.

Because of these problems of recognition, the prevalence of excessive sleep in children and adolescents is not known. Clearly, it is not rare in view of the range of underlying conditions causing it, many of which are individually quite common.

It is important to establish that the problem really is excessive sleepiness. 'Tiredness' is an ambiguous term: ideally, sleepiness should be distinguished from fatigue or lethargy, without necessarily the need to sleep, for which different explanations are likely including physical illnesses, such as anaemia or endocrine disorders in which other signs are usually present. Occasionally, excessive sleepiness with long periods in bed or at home is simulated in order to escape from a difficult situation. Detection of such cases requires very careful clinical evaluation and assessment and possibly PSG.

Excessive sleepiness is mainly a problem in older children and (especially) adolescents. Many teenagers complain about excessive sleepiness but it has been claimed that very many more than those who seek help are likely to be suffering from chronic sleep deprivation

or '**sleep debt**'. As mentioned earlier, the adverse effects of this are thought to be wide-ranging from underperformance at school, college or work, to road traffic accidents and other mishaps, as well as antisocial behaviour.[6] Sometimes the situation is complicated by the use of stimulants to stay awake, and alcohol or sedative drugs to get to sleep.

The differential diagnosis of excessive sleepiness can be considered in terms of three main categories of cause: insufficient sleep, disturbed sleep, and an increased need for sleep (Table 9.2.9.3).

Insufficient sleep

The combination of late-night social activities or staying up late for study, and having to get up early for school or college reduces the number of hours many adolescents sleep to below that needed for satisfactory daytime functioning. Difficulty getting off to sleep at night and recurrent waking makes the problem worse (or may be sufficient in themselves to reduce or seriously impair sleep). The result is considerable difficulty getting up in the morning, irritability, emotional liability, lethargy, tiredness, or actually falling asleep during the day.

Correction of the problem of late-night social activities requires a change of lifestyle or other measures which may be difficult to achieve without there being strong motivation to do so. The ideal solution is an agreed, co-operative effort on the part of both the young person and parents.

More specific measures will be needed if the habit of going to bed late has developed into a disturbance of the circadian sleep–wake cycle. This may take the form of irregular sleep–wake schedules or, more usually, the **delayed sleep phase syndrome (DSPS)** which deserves special mention because it is common, especially in adolescents, and potentially very disruptive.

The time at which children fall asleep may become delayed during a period of illness, or because of protracted bedtime disputes about going to bed. In adolescents the problem arises from habitually staying up late for social or other reasons, especially at weekends or during holidays. After a time (the length of which varies with the individual) the sleep phase becomes physiologically delayed with the result that it becomes impossible to go to sleep earlier by choice, in spite of feeling tired and having been awake for a long time. Entreaties to go to bed at a sensible time and get up on time for school are likely to be ineffective.

The diagnostic features of DSPS are persistently severe difficulty getting to sleep, uninterrupted sound sleep, great difficulty getting up for school or work, and sleepiness and under-functioning especially during the first part of the day, giving way to alertness in the evening and early hours. The abnormal sleep pattern is maintained by sleeping in very late when able to do so at weekends and during holidays.

Treatment consists of gradually and consistently changing the sleep phase to an appropriate time. This can be achieved by slowly advancing the sleep phase (e.g. by 15 min a day) where the phased delay is about 3 h or less. More severe forms of the disorder require progressive sleep phase delay in 3 h steps ('round the clock'). Additional measures to maintain the improved sleep schedule include early morning exposure to bright light and firm agreement with the adolescent to maintain the new pattern of social activities and sleep. The place of melatonin remains unclear in view of the many uncertainties about its use and potential hazards, including possible adverse affects on reproductive physiology.

Achieving and maintaining an improved sleep–wake schedule by these means may not be easy. The difficulties are compounded if there is a vested interest in maintaining the abnormal sleep pattern, for example to avoid school ('**motivated sleep phase delay**'). The presence of psychological problems, including depression, may well make successful treatment less likely. 'Conditioned insomnia' may appear similar to DSPS but its origins and treatment are different.

Table 9.2.9.3 Differential diagnosis of excessive sleepiness in older children and adolescents

Insufficient sleep
Late-night activities combined with getting up early
Insomnia
Erratic sleep–wake patterns
Delayed sleep-phase syndrome
Disturbed sleep at night
Sleep-related upper airway obstruction
Recreational drugs (caffeine, alcohol, nicotine)
Illicit drugs (including withdrawal)
Medical and psychiatric disorders
Other sleep disorders (frequent parasomnias, periodic limb movements in sleep)
Increased need for sleep
Narcolepsy
Idiopathic central nervous system hypersomnia
Depression
Substance abuse
Neurological disease
Kleine–Levin syndrome (intermittent sleepiness)
Menstruation-related hypersomnia (intermittent sleepiness)

Disturbed nocturnal sleep

Daytime sleepiness, despite apparent normal time asleep at night, suggests that the restorative quality of sleep is impaired. Poor quality sleep can result from frequent awakenings or less obvious arousals including brief subclinical interruptions (or 'fragmentation') of sleep. Sleep may be disturbed by the following:

- Excessive caffeine, alcohol, or nicotine (combinations are particularly hazardous), and illicit drug use and withdrawal.
- Medical and psychiatric disorders in childhood, and some of their treatments.
- Other sleep disorders: frequent parasomnias are likely to be obvious but periodic limb movements in sleep (now considered to be more important in childhood than previously thought) are much more subtle in their effects on sleep continuity.
- Sleep-related respiratory problems, including OSA. This condition merits some emphasis because of its widespread occurrence: at least 2 per cent of children in the general population are affected with a peak onset at 2 to 6 years. The prevalence is much higher than this in various learning disability syndromes such as Down syndrome, as mentioned earlier.

There are important differences between OSA in children and OSA in adults. The typical adult with OSA is an obese middle-aged male who snores very loudly, under-functions during the day and usually responds to continuous positive airway pressure (CPAP) treatment at night.

- In contrast, most children with OSA are not obese, the usual causes are large tonsils and adenoids (the removal of which can be beneficial), the sex ratio is equal and, whereas adults have prolonged obstructive apnoeas, children often have partial airway obstruction with hypoventilation, actual apnoea events being less frequent and shorter. As already mentioned, the result may be over-activity and other disruptive behaviour rather than obvious sleepiness during the day.

Clinical assessment is the cornerstone of recognizing OSA.

- Night-time signs suggesting the diagnosis are combinations of snoring (although only about one in five children who snore most nights have this condition), other noises suggesting breathing difficulties during sleep, paradoxical chest–abdomen respiratory movements, unusual sleeping positions including neck extension, very restless sleep, profuse sweating, nocturnal enuresis, and sudden distressing awakenings during the obstructive event. There is also a higher incidence of other parasomnias but for reasons that are unclear.
- Daytime features include mouth breathing and adenoidal facies, headache and bad mood on wakening, and the behaviour problems already mentioned.
- Physical examination may reveal the anatomical cause of the obstruction (usually enlarged tonsils and adenoids). Radiological studies are required in the more complicated cases. PSG with respiratory measures is needed to assess the severity of the obstruction and the effects on blood gases during sleep which may be greater than suspected from clinical findings.

Treatment is essential to counter or prevent physical and psychological complications. This usually consists of adenotonsillectomy or, much less commonly, other measures such as CPAP depending on the individual case.

Disorders involving an increased tendency to sleep

This occurs where prolonged or otherwise excessive sleep is an intrinsic part of the condition, rather than a consequence.

(a) Narcolepsy

Narcolepsy is a neurological disorder mainly affecting REM sleep physiology. It is not the rarity once supposed; prevalence in the United States has been estimated at 4–9 per 10 000.

Onset has occurred by adolescence in a high number of cases but the diagnosis is often not made for several years. The reasons for this are that symptoms may be subtle in their early stages, concealed, misinterpreted as laziness or psychological disorder such as depression or conversion disorder, or overshadowed by the child's extreme emotional reaction to the condition. Especially because of its many manifestations in childhood, narcolepsy is a good illustration of how a sleep disorder can be misconstrued as another type of clinical condition, especially when familiarity with the field of children's sleep disorders is limited.[7]

The clinical presentation of early narcolepsy is very variable and some time usually elapses before the classic combination of daytime sleep attacks, overnight sleep disruption, cataplexy, hypnagogic hallucinations, and sleep paralysis develop if, indeed, it does at all. In young patients the first sign may consist of no more than prolonged overnight sleep.

It is appropriate to consider narcolepsy in any young person who is excessively sleepy during the day without an obvious explanation, but repeated clinical and PSG assessment may be required at intervals before a definite diagnosis can be made, including its distinction from other forms of sleepiness such as the group of conditions known as **idiopathic CNS hypersomnia**. The demonstration of low CSF levels of the neuropeptide hypocretin (orexin) is now considered diagnostic of narcolepsy. The PSG features of narcolepsy in adults, which are described elsewhere (Chapter 4.14.3) do not necessarily apply in the childhood stage of the condition.

Narcolepsy is a persistent and disturbing condition for which careful treatment with medication together with ancillary measures, as well as much support and advice about education, career, and psycho-social matters, are required. Other aspects of diagnosis and management are discussed in Chapter 4.14.3.

(b) Kleine–Levin syndrome

This also usually begins in the teenage years with periods of excessive sleepiness alternating with periods of normality. The sleepy episodes are associated (in its classical form) with overeating, hypersexuality, and other disturbed behaviours which are often bizarre and out of character. It is also frequently mistaken for a psychological disorder or other medical conditions. The condition should be distinguished from other causes of intermittent sleepiness in young people such as substance abuse, major depressive disorder (in which the sleepiness is much less marked), menstruation-related hypersomnia, and certain neurological disorders.

Main sleep problems: parasomnias

This category of sleep disorders is described in Chapter 4.14.4 in relation to adult psychiatry. The present account emphasizes aspects of the parasomnias of particular importance in childhood and adolescence when, collectively, they are more common than in adult life. Frequently, parasomnias cause parents much concern and they appear to be the subject of considerable diagnostic confusion and delay. As in older patients, different types of parasomnia may co-exist. A detailed account of parasomnias in young patients is available elsewhere.[8]

The following general points about childhood parasomnias have clear implications for clinical practice.

- Precise diagnosis is important as different parasomnias may well need contrasting types of treatment. Accurate diagnosis depends principally on a detailed account of the subjective and objective sequence of events from the onset of the episode to its resolution, as well as the circumstances in which the episode occurs, including its timing. Audio–visual recording (including the use of home video systems) can be very informative. Only occasionally is PSG required, although this can be instructive where clinical evaluation is inconclusive and, sometimes, where there is the possibility that another type of sleep disorder co-exists.
- The more dramatic forms of parasomnia seem to be a particular cause of diagnostic confusion and imprecision, and also quite possibly unnecessary concern about their psychological significance as most are benign.
- Especially when the range and manifestations of sleep disorders is not appreciated, parasomnias (and other sleep disorders) may well be misinterpreted as other physical or psychological conditions.

- A child may have more than one kind of parasomnia or, indeed, more than one sleep disorder (e.g. arousal disorders associated with obstructive sleep apnoea).
- As many childhood primary parasomnias remit spontaneously within a few years, children and parents can often be reassured about the future, although protective measures (e.g. in severe headbanging or sleepwalking) may be required in the meantime.
- Specific treatment, including medication, is needed in only a minority of cases of primary parasomnia but is likely to be required for the underlying disorder in many of the secondary parasomnias.
- Research information on this point is limited, but a primary parasomnia might be symptomatic of a psychological problem if it is very frequent, unusually late in onset or persistent, or associated with a traumatic experience.
- Parasomnias may lead to psychological complications if the child is frightened, embarrassed, or otherwise upset by the experience, or because of the reactions of other people to the episodes.

Primary parasomnias

- Sleep-related **rhythmic movement disorders** such as headbanging, occur in many young children, almost always remitting spontaneously by 3 to 4 years of age. Although alarming to parents, they are usually of no psychological significance (unlike daytime headbanging associated with severe neurodevelopmental disorder). However, protective measures, such as padding the cot-sides, may be needed.
- **Hypnagogic** (sleep onset) and **hypnopompic** (on waking) hallucinations are common and may be frightening to the child.
- Parents are often distressed to witness **confusional arousals, agitated sleepwalking**, or **sleep terrors** which are a form of '**partial arousal disorder**' common in young children (see Chapter 4.14.4 for an account of arousal disorders). The degree of agitation and confused behaviour may be extreme, suggesting that the child is suffering in some way. In fact, in arousal disorders the child remains asleep and unaware of the events. Understandable attempts to arouse the child and provide comfort should be discouraged, as this may cause real distress. Although violence during sleep is described mainly in sleepwalking adults, such behaviour can occur in children.
- The term 'nightmare' is sometimes used misleadingly for any form of dramatic parasomnia. **True nightmares** (frightening dreams) are common. If frequent and associated with intense bedtime fears, they may indicate an anxiety disorder and their content may suggest a cause.
- **Nocturnal enuresis** is very common, affecting about 5 per cent of 7-year-olds at least once a week. Delayed maturation often seems to be the explanation, but physical or psychological factors may be involved, especially where previous bladder control is lost. Behavioural treatment can be very effective.

Secondary parasomnias

Nocturnal epileptic seizures are not uncommon in children and must be distinguished from primary parasomnias because of their different significance, and also the investigation and treatment they require. Seizures which are behavioural in manifestation are the most likely to be misdiagnosed as non-epileptic, for example benign centro-temporal (Rolandic) epilepsy of childhood and nocturnal frontal lobe seizures both of which are closely related to sleep.

Other parasomnias, which are part of medical or psychiatric disorders and which may be encountered in patients of any age, include nocturnal asthmatic attacks with accompanying distress, those associated with OSA or gastro-oesophageal reflux, panic attacks, nocturnal disturbance that is part of the post-traumatic stress disorder, and dissociative states. Simulated parasomnias, shown by PSG to be enacted during wakefulness, can sometimes occur in children.

Further information

Stores, G. (2008). *Sleep problems in children and adolescents: the facts*. Oxford University Press, Oxford.

Stores, G. and Wiggs, L. (2001). *Sleep disturbance in children and adolescents with disorders of development: its significance and management*. Clinics in Developmental Medicine No. 155. Mac Keith Press, London.

References

1. Stores, G. (2007). Clinical diagnosis and misdiagnosis of sleep disorders. *Journal of Neurology, Neurosurgery and Psychiatry*, **78**, 1293–7.
2. Owens, J.A., Spirito, A., and McQuinn, M. (2000). The children's sleep habits questionnaire. Psychometric properties of a survey instrument for school-aged children. *Sleep*, **23**, 1043–51.
3. Kuhn, B.R. and Elliot, A.J. (2003). Treatment efficacy in pediatric sleep medicine. *Journal of Psychosomatic Research*, **54**, 587–97.
4. Fleming, P., and Blair, P.S. (2007). Sudden infant death syndrome. *Sleep Medicine Clinics*, **2**, 463–76.
5. Fallone, G., Owens, J.A., and Deane, J. (2002). Sleepiness in children: clinical implications. *Sleep Medicine Reviews*, **6**, 287–306.
6. Dahl, R.E. and Lewin, D.S. (2002). Pathways to adolescent health: sleep regulation and behavior. *Journal of Adolescent Health*, **31**, 175–84.
7. Stores, G. (2006). The protean manifestations of childhood narcolepsy and their misinterpretation. *Developmental Medicine and Child Neurology*, **48**, 307–10.
8. Stores, G. (2007). Parasomnias of childhood and adolescence. *Sleep Medicine Clinics*, **2**, 405–17.

9.2.10 Suicide and attempted suicide in children and adolescents

David Shaffer, Cynthia R. Pfeffer, and Jennifer Gutstein*

Introduction

Suicidal behaviour is a matter of great concern for clinicians who deal with the mental health problems of children and adolescents. The incidence of suicide attempts reaches a peak during the mid-adolescent years, and mortality from suicide, which increases steadily

through the teens, is, in many countries, one of the leading causes of death at that age.

Historical review

Until the late 1950s, knowledge about youth suicide was drawn from unrepresentative case reviews, reviews of the demography of suicide drawn from death certificate data, and speculation about dynamics. The late 1950s saw the first systematic psychological autopsy study among adults that demonstrated the importance of psychiatric disorder as a proximal cause of most suicides.[1] This was followed by similar studies on children and adolescents,[2–5] confirming the association in adolescence. Starting in the mid-1960s, the incidence of suicide in young males began to rise in many countries.[6] The rate of increase eventually stabilized in the late 1980s and, in many countries, is now showing signs of falling.[7] These changes stimulated efforts to develop methods of preventing youth suicide.[8–11].

A good deal is now known about which teenagers commit suicide, less about who attempts it, and very little about the optimal management of suicidal adolescents. The number of randomized controlled trials designed to assess different forms of treatment is exceedingly small, and many suggestions for clinical management are based on anecdotal accounts rather than on findings from well-designed experimental trials.

Clinical features

Completed suicide

Completed suicide occurs most commonly in older adolescents, and, although it can also occur in children as young as 6 years of age, it is excessively rare before puberty.[7] Psychological-autopsy studies have shown that about 90 per cent of adolescent suicides occur in individuals with a pre-existing psychiatric disorder, often present for several years.[3–5] In teenagers, the most common disorders are some form of mood disorder, substance and/or alcohol abuse, often comorbid with a mood disorder in boys over age 15, and anxiety disorders.[3,4] At a trait level, many suicide completers have been noted to be irritable, impulsive, volatile, and prone to outbursts of aggression. However, this pattern of behaviour is by no means universal, and anxious suicides have usually shown no evidence of prior behavioural, academic, or social disturbances.[3]

Although some adolescents—predominantly girls suffering from a major depressive disorder—appear to have thought about suicide for some time before death, most adolescent suicides appear to impulsively follow a recent stress event, such as getting into trouble at school or with a law-enforcement agency; a ruptured relationship with a boy- or girlfriend; or a fight among friends. In many instances, these stress events can be seen as a by-product of their underlying psychiatric disorder.[12]

It also appears that a completed suicide can be precipitated—in a presumably already suicidal youth-by exposure to news of another person's suicide, or by reading about or viewing a suicide portrayed in a romantic light in a book, magazine, or newspaper.[13]

About a third of completed suicides have made a previous known suicide attempt, more commonly girls and those who suffered from a mood disorder.[3] Completed suicide must be distinguished from autoerotic asphyxia, which is rare in teenagers.[3] Suicide pacts, common between middle-aged or elderly married couples and/or other family members, are similarly rare in adolescents, but are not unknown.[3]

Non-lethal suicidal behaviour

(a) Suicidal ideation

Suicidal ideation includes thoughts about wishing to kill oneself, making plans of when and where, and having thoughts about the impact of one's suicide on others. Such thoughts may occur without great significance among young children, who may not appreciate that suicide may result in irreversible death.[14] However, appreciation of the finality of death should not be a factor in judging the seriousness of suicidal ideation. Suicide threats made by young children and adolescents most often involve a threat to jump out of a window, to run into traffic, or to stab himself or herself.

(b) Attempted suicide

The most common profile of a teenaged attempter is a 15- to 17-year-old girl who has taken a small- or medium-sized overdose of an over-the-counter analgesic or medication taken by another family member. The behaviour is usually impulsive and occurs in the context of a dispute and humiliation with family or a boyfriend.[15] The clinical features most strongly associated with suicide attempts are irritability, agitation, threatening, violent, or psychotic behaviour, and a persistent wish to die.[16]

Groups in whom suicide attempts appear to be common include runaways,[3] children who have been exposed to physical and sexual abuse, and homosexual teenagers.[17] However, study-design issues make it unclear whether this is because of a high rate of psychopathology or substance abuse in these groups or because of some factor that specifically predisposes to suicidal behaviour.

A subset of non-fatal suicidal behaviour involving ingestion with a non-lethal intent is sometimes referred to as parasuicide. However, intent is difficult to gauge retrospectively, and not all teenagers are aware of the lethalness of an ingestion, so that this term carries with it a risk of complacency and is probably best avoided in teenagers.

Assessment

Suicide attempts

Assessment of a suicide attempt involves an evaluation of the short-term risk for suicide and attempt repetition, and an assessment of the underlying diagnosis or other promoting factors. If the child or teenager has been referred to as an ideator, it is important to determine whether they are contemplating or have secretly attempted suicide.

Repeated attempts, attempts by unusual methods (other than ingestions or superficial cutting), medically serious attempts, and attempts where the patient has taken active steps to prevent discovery all increase the risk for further attempts or death.[18,19] Children and adolescents systematically overestimate the lethality of different suicidal methods, so that a child with a significant degree of suicidal intent may fail to carry out a lethal act.[20–22]

The mental states leading to suicidal behaviour include anticipatory anxiety, pessimism, or hopelessness, as well as paranoid or other cognitive distortions arising from an underlying psychiatric diagnosis.[23] Inappropriate coping styles (e.g. impulsivity or catastrophizing) in response to external stress may also contribute to the behaviour. Motivating feelings may include the wish to effect a change in interpersonal relationships, to rejoin a dead

relative, to avoid an intolerable situation, to get revenge, or to gain attention.[22]

Classification of associated diagnoses

Suicide

Psychiatric diagnoses commonly associated with a suicide include depression, bipolar disorder, substance abuse, conduct disorder, and overanxious and panic disorders.[3–5] Although the rate of suicide in schizophrenics is high, because of the rarity of the condition, it accounts for very few suicides.

Suicide attempts

Recurring suicidal behaviour has been associated with hypomanic personality traits and cluster B personality disorders.[22] A history of impulsivity, mood lability, with rapid shifts from brief periods of depression, anxiety, and rage to euthymia and/or mania—associated with transient psychotic symptoms, including paranoid ideas and auditory or visual hallucinations-is associated with a risk for further suicide attempts and is compatible with the diagnosis of borderline personality disorder. Many of these symptoms are also features of bipolar mood disorder.

Epidemiology

Completed suicide

(a) Age

In the United States, the age-specific mortality rate from suicide for 10- to 14-year-olds was 1.6 per 100 000 in 1997.[24] This age group accounts for 7 per cent of the population but only 1 per cent of all suicides, and most of these occur in 12- to 14-year-olds.

The comparable figures for 15- to 19-year-olds are about six times higher. The suicide rates at this age in the United States and Canada in 1997 were 9.5 per 100 000 and 12.86 per 100 000 respectively.[7] The proportion of suicides that occur in this age group is about the same as its representation in the general population.

Suicide rates for 15- to 24-year-olds in some other English-speaking countries were 11.0 for males and 2.2 for females in the United Kingdom (1995), 16.0 in Australia (1995), and 26.1 in New Zealand (1997) (all per 100 000 population).[7]

(b) Gender

In the United States and most other countries, male suicides outnumber female suicides among 15- to 24-year-olds by a ratio of 4:1. In China and Cuba, the suicide rate is higher in females than in males.[7]

(c) Cultural and ethnic differences

Rates of suicide vary considerably in different cultural and national groups.[7] Possible reasons include variable access to lethal methods, different degrees of social support, integration, or group adherence, or the influence of religious beliefs or spirituality.[25–27] In some instances, the differences may be a function of geography rather than culture. Contagion within isolated groups may determine differences in rates.

(d) Secular changes

From 1964 to 1995, the suicide rate in the United States and Canada among 15- to 19-year-old males increased almost three-fold, and similar increases were reported in Australia, New Zealand, and the United Kingdom.[7] In most of these countries, there was little change in the female rate or in the rate amongst 10- to 14-year-olds. Fluctuations in the suicide rate appear to be real, rather than due to any methodological artefact (e.g. changes in reporting practices). The most plausible reason for the increase in suicidal behaviour among teenage boys is an increase in alcohol and substance use in the youth population.[3] The reasons offered for the recent decline in suicide rates include lowered substance- and alcohol-use rates among the young and more effective diagnosis and treatment.[28]

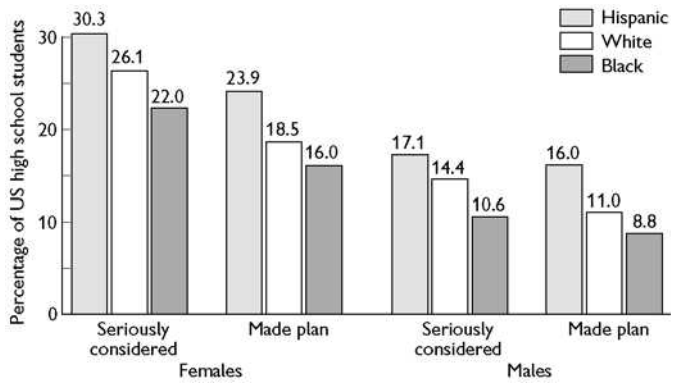


Fig. 9.2.10.1 Youth risk behaviour survey: prevalence of suicidal ideation in teenagers in the previous 12-month period (1997) broken down into gender and ethnicity. (Reproduced from Centers for disease control (1998). Attempted suicide among high-school students–US, 1997. Morbidity and Mortality weekly Report, **47**, 47–9, copyright 1998, centers for Disease Control and Prevention, US.)

Attempted suicide

There is a strong inverse relationship between attempted suicide and age. A large epidemiological survey of four suicide-related behaviours (ideation, plan, gesture, and attempt) in the United States has shown a significantly higher rate of all four behaviours in the youngest age group (15–24 years).[29] This study also compared rates of these four behaviours across two decades (1990–1992 and 2001–2003). It found that rates did not decrease, despite a dramatic increase in pharmacologic treatment.

Suicide attempts in adolescents are at least twice as common in females as males (see Figs 9.2.10.1 and 9.2.10.2). Considerable ethnic variation is seen in the United States, with, for unknown

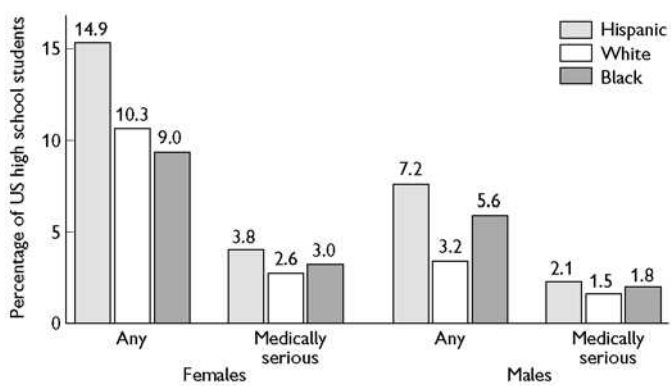


Fig. 9.2.10.2 Youth risk behaviour survey: prevalence of suicide attempts in teenagers in the previous 12-month period (1997) broken down into gender and ethnicity. (Reproduced from Centers for disease control (1998). Attempted suicide among high-school students–US, 1997. Morbidity and Mortality weekly Report, **47**, 47–9, copyright 1998, centers for Disease Control and Prevention, US.)

reasons, Hispanic high-school students having twice the rate of black or white teenagers.[30]

Aetiology

Completed suicide

(a) Psychiatric disorders

The most important risk factor for suicide is a psychiatric disorder.[3] Controlled studies of completed suicide suggest similar risk factors for boys and girls,[3,h31] but with marked differences in their relative importance[3–5] (Table 9.2.10.1). In girls, major depression is the most powerful risk factor, which, in some studies, increases the risk of suicide 12-fold; followed by a previous suicide attempt, which increases the risk approximately three-fold. In boys, a previous suicide attempt is the most potent predictor, increasing the rate over 30-fold. It is followed by depression (12-fold increase), disruptive behaviour (two-fold increase), and substance abuse (increasing the rate by just under two-fold).[3]

(b) Psychosocial stressors

Stressful life events often precede a suicide and/or suicide attempt.[12] They are rarely a sufficient cause in suicide, and their importance seems to lie in their action as a precipitant of stress in young people who are at risk by virtue of their psychiatric condition. Family discord, lack of family warmth, and a disturbed parent-child relationship are commonly associated with types of child and adolescent psychopathology, but these factors do not play a more important role in suicide.[12]

(c) Cognitive factors

Perceptions of hopelessness, negative views about one's own competence, poor self-esteem, a sense of responsibility for negative events, and the immutability of these distorted attributions may contribute to the 'hopelessness' repeatedly found to be associated with suicidality.[18,19]

(d) Biology

Biological factors, specifically dysregulation of the serotonergic system, are common in adult suicides.[32] Dysregulation is manifested by low levels of serotonin metabolites in central nervous system fluids, low concentrations of presynaptic serotonergic receptors, and dense concentrations of postsynaptic receptors. Such serotonin abnormalities have been localized to the ventrolateral prefrontal cortex and brainstem of suicide victims and attempters (in postmortem positron-emission tomographic studies as well as in *in vivo* biological challenges).[33] Serotonin may inhibit extreme fluctuations of mood and reactivity, and the vulnerability to suicide of individuals with these biological abnormalities may be mediated by impulsivity and emotional volatility. As the ventral prefrontal cortex plays a role in behavioural inhibition, it is conceivable that serotonin irregularities in this area make it more difficult for a suicidal individual to control his suicidal impulses.[33] The frequency with which these biological findings occur in adolescent suicide attempters is not yet clear, and studies to demonstrate the precise behavioural correlates of serotonin dysregulation profiles are still lacking. Nordstrom *et al.*[34] have suggested that knowing the biological status of suicide attempters may have a practical value, in that low 5-hydroxyindole acetic acid concentrations in cerebrospinal fluid examined shortly after a suicide attempt may differentiate between suicide attempters who will commit suicide or repeat the attempt within a year and those who will not. The biology of suicidal behaviour is considered more fully in chapter 4.15.3

Table 9.2.10.1 Psychiatric diagnoses in child and adolescent suicides

	Martunnen *et al.*[5]			**Shaffer *et al.*[6]**			**Brent *et al.*[4]**		
Country	Finland			USA			USA		
Area	National			Greater New York			Western Pennsylvania		
Period	1987–1988			1984–1986			1984–1994		
N	53			120			140		
Age	13–19			<20			13–19		
Percentage girls	17			21			15		
Control group	None			Matched community			Matched community		
Diagnostic system	DSM-IIIR			DSM-III			DSM-III		
	Males	**Females**	**All**	**Males**	**Females**	**All**	**Males**	**Females**	**All**
Any diagnosis (%)	93	100	94	90	92	91	82	81	82
Any mood disorder (%)	48	67	51	60	68	61	43	71	47
Substance abuse (%)	27	44	30	42	12	35	35	24	34
Conduct/antisocial/ disruptive disorder (%)	18	11	17	54	36	50	35	10	31
Any anxiety disorder (%)	2	11	4	27	28	27	13	24	14
Schizophrenia (%)	5	11	6	3	4	3	—	—	—
Past suicide attempt	27	67	34	28	50	33	37	62	41

(e) Imitation

Evidence has accumulated indicating that suicide in vulnerable teenagers can be precipitated by exposure to real or fictional accounts of suicide, such as intense media coverage of a real suicide or the fictional representation of a suicide in a popular film or television programme. The risk is especially high in the young, and lasts for approximately 2 weeks.[13] The phenomenon of suicide clusters is also presumed to be related to imitation.

"Cybersuicide," or a tendency for internet sites such as Bebo or MySpace" to encourage suicide pacts or to increase completed or attempted suicide in vulnerable adolescents, has become an increasing concern. This concern is based on multiple case reports from Japan, Wales, and the United States. These have captured substantial media attention. As yet no objective empirical data have been amassed to determine if there is a causal relationship between internet use and youth suicide.

One hypothetical model for how biological and social factors fit together is illustrated in Fig. 9.2.10.3.

In a longitudinal study of a large African-American community in the United States, Juon and Ensminger[35] found risk factors for suicidal behaviour in African-Americans to be very similar to those found in Caucasians (depression, substance use, and a number of family variables).

Course and prognosis

Natural history

Little is known about the natural history of suicidal behaviour, but early-onset suicidal behaviour in prepuberty predicts suicidal behaviour in adolescence[36,37] and an early-onset major depressive disorder is associated with suicidal behaviour in adolescence and adulthood.[38] Attempts to predict, at the time of the first attempt, which adolescents are likely to repeat their suicidal behaviour have been unsuccessful.[22]

Emergency treatment

Because of the need to respond to a suicidal crisis, it is desirable to offer treatment within a service-delivery system that includes inpatient and outpatient settings, acute crisis work, stabilization, extended management, and follow-up care and monitoring.

Outpatient treatment should be used when the child or adolescent is unlikely to act on suicidal impulses, when there is sufficient support at home, and when there is someone who can take action if the adolescent's behaviour or mood deteriorates. Children and adolescents should never be discharged from an emergency or care service without the child's or adolescent's caretaker having been interviewed (see Box 9.2.10.1) to ensure that firearms and/or lethal medications will be made inaccessible to the child. Unless this advice is given, parents will rarely, on their own initiative, take the necessary precautions. Before discharge, the clinician must have a good understanding of the amount of support that will be available for the child or adolescent if he or she is discharged home.

Acute psychiatric inpatient care should be reserved for patients for whom intensive surveillance and intervention are considered essential-as in the presence of active suicidal ideation and intent or when the youngster is unable to commit to not carrying out a suicidal act, when the youth is unpredictable, impulsive, agitated, or psychotic, or if there is a lack of support and supervision in the home. There is no evidence that exposure to other suicidal psychiatric inpatients increases the risk of suicidal behaviour. Determining when a patient is ready for discharge from the hospital or crisis centre will usually include an evaluation of the severity of existing suicidal ideation and intent. Implicit coercions (e.g. telling patients that discharge will be delayed until they can state that they are not suicidal) should be avoided.

Treatment compliance may be improved by offering definite, closely spaced, follow-up appointments, being flexible in arranging appointments if a crisis should arise, and reminding the family and patient by telephone or note about the next appointment. If an appointment is missed, the patient and parent should be contacted. Hopeless and depressed children and adolescents, who may be not be able to commit to a lengthy treatment process, may be better engaged by offering short-term treatment plans with defined intervention goals. While offering confidentiality for some issues,

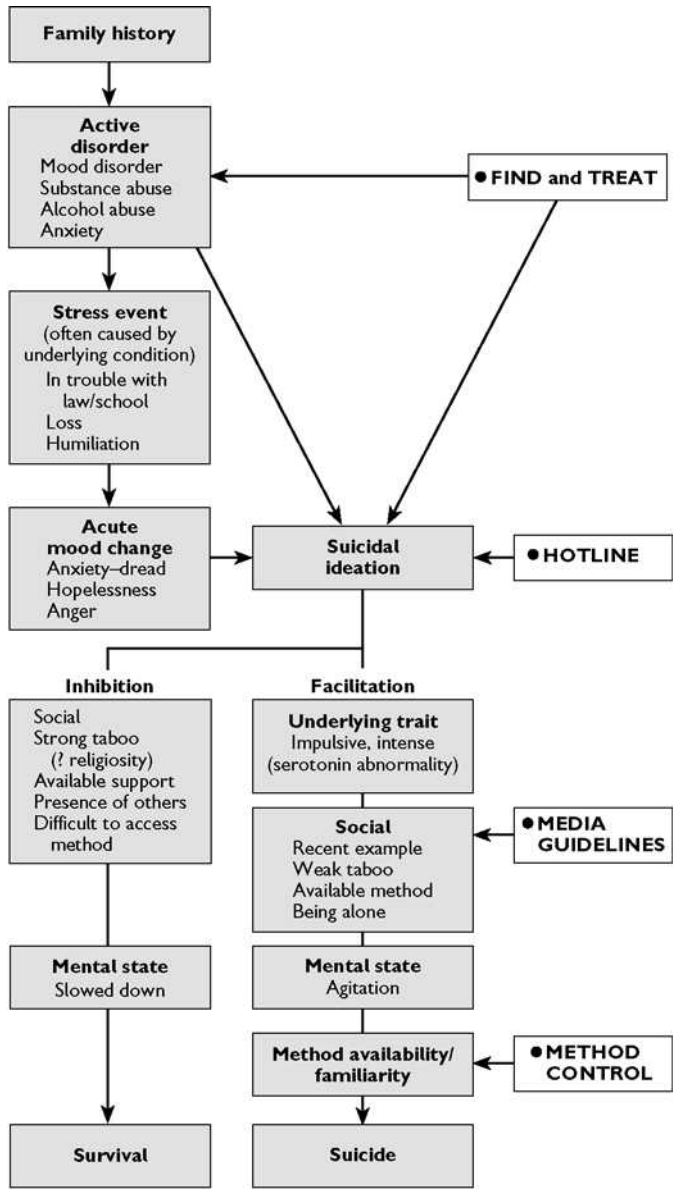


Fig. 9.2.10.3 How do suicides occur and how can they be prevented?

Box 9.2.10.1 Checklist before discharging an adolescent who has attempted suicide

Before discharging a patient from the emergency room or crisis centre, always:

- Check that *firearms* and lethal *medications* have been secured or removed
- Check that there is a *supportive person* at home
- Check that a *follow-up appointment* has been scheduled

it is essential that the clinician communicate to the patient that, if they feel that suicidal thinking or behaviour is imminent, such information will be shared with the parents.

Contracts

A written or verbal 'no-suicide' contract is commonly negotiated at the start of treatment in the hope that it will improve treatment compliance and reduce the likelihood of further suicidal behaviour.[39] In its usual form, the child or adolescent promises not to engage in suicidal behaviour without first informing the parents, therapist, or other responsible adult when he or she has thoughts of suicide or plans to commit suicide. No empirical studies have evaluated the efficacy of a contract, and contracts should be seen as no more than adjuncts to the management of patients with low intent. Even if the patient agrees to such a contract, suicide risk may persist. It should also be appreciated that a 'no-suicide' contract may lessen a patient's communication of stress and dysphoria, decrease the potential for developing a therapeutic alliance, and impair risk management. As mentioned above, coercive communications should be avoided, because these may encourage deceit and defiance.

Specific psychotherapies

Working with suicidal children and adolescents is best done by a clinician who is available, has skill and training in managing suicidal crises, relates to the patient in an honest and consistent way, and can convey a sense of optimism and activity. Given these personal attributes, the therapist may use various models of psychotherapy, although relatively few empirical studies have evaluated their efficacy.

(a) Cognitive behavioural therapy

Cognitive hbehavioural therapy is effective in depressed teenagers,[40] but its value for suicidal adolescents has not been demonstrated.[23] Brent *et al.*[23] modified the approach for depressed adolescents. The treatment comprized 12 to 16 once-weekly sessions, followed by a 6-month booster phase of monthly or bimonthly sessions. It included a psychoeducational manual about mood disorders, training to monitor and modify automatic thoughts, assumptions, and beliefs, training in more assertive and direct methods of communicating, and help in conceptualizing alternative solutions to problems. Meetings with parents were sometimes held to augment the treatment, and psychopharmacology was used adjunctively if depressed adolescents had not improved after 4 to 6 weeks of pharmacotherapy.

Brent's study provides no evidence of the efficacy of cognitive-behavioural therapy for teenagers who had made a suicide attempt who were not included in this study.

(b) Dialectical behavioural therapy

Dialectical behavioural therapy (**DBT**) is the only form of psychotherapy that has been shown in a randomized control trial to reduce suicidality in adults with borderline personality disorder.[41] This treatment is based on a biosocial theory in which suicidal behaviours are considered to be maladaptive solutions to painful negative emotions that also have affect-regulating qualities and elicit help from others.[41]

The treatment involves developing problem-oriented strategies to increase distress tolerance, emotion regulation, interpersonal effectiveness, and the use of both rational and emotional input to make more balanced decisions. It usually involves individual and group sessions over the course of a year, although an untested modification for adolescents (DBT-A) is designed to take 12 weeks.[42] It involves the participation of a relative who is charged to improve the home environment and to teach other relatives how to model and reinforce adaptive behaviours for the adolescents.

(c) Family therapy

As indicated above, family discord, poor communication, disagreements, lack of cohesive values and goals, and irregular routines and activities are common in suicidal children and adolescents who often feel isolated within the family. Family intervention aims to decrease such problems, improve family problem-solving and conflict resolution, and reduce blame directed at the suicidal child or adolescent. Family-based cognitive therapy aims to reframe the family's understanding of their problems, to alter the family's maladaptive problem-solving techniques, and to encourage positive family interactions. Psychoeducational approaches can help parents clarify their understanding of childhood and adolescent suicidal behaviour, identify changes in mental state that may herald a repetition, and reduce the extent of expressed emotion or anger.[31]

Psychopharmacological interventions

In meta-analyses of adult studies, lithium maintenance treatment greatly reduces (8.6-fold) the recurrence of suicide attempts in adults with bipolar or other major affective disorders. Further, when lithium is discontinued there is a seven-fold increase in the rate of suicide attempts and a nine-fold increase in the rates of suicide.[43] Other mood stabilizers, such as valproate and carbamazepine, are also widely used to treat bipolar disorders in children and adolescents; although their efficacy has yet to be empirically demonstrated. Depressed suicidal children and adolescents with a history of bipolar disorder should first be treated with a mood stabilizer before receiving an antidepressant.

Studies in depressed adults have found that the selective serotonin reuptake inhibitor (**SSRI**) antidepressants reduce suicidal ideation, and also reduce the frequency of suicide attempts in non-depressed patients with cluster B personality disorders with a past history of suicide-attempt behaviour.[44] In contrast to the highly lethal potential of tricyclic antidepressants when taken in overdoses, SSRIs have low lethal potential. In a controlled trial of the depot neuroleptic flupenthixol, Montgomery and Montgomery[45] noted a significant reduction in suicide-attempt behaviour in adults who had made numerous previous attempts. Similar studies have yet to be conducted for adolescents.

In the past decade, there has been much controversy over whether the SSRI antidepressants can induce suicidal ideation

and/or behaviour. A number of case reports appeared in 1990 describing patients who had developed suicidal preoccupations after starting treatment with fluoxetine. These reports were not supported by meta-analyses and re-analyses of large SSRI-treatment trials of depressed, bulimic, or anxious patients.[46,47] The conclusion was reached that suicidal ideation is a common feature of depression and that the prevalence in SSRI-treated depressives was no greater than expected.

However, one reanalysis of the data presented in certain of these studies suggested that new ideation was significantly more common in SSRI-treated depressed patients who had not previously reported suicidal ideation. Further, in a naturalistic challenge study, Rothschild and Locke[48] were able to reinduce suicidal ideas in a small series of patients who had first experienced ideation after starting treatment with fluoxetine. These patients had also experienced akathisia as a complication of fluoxetine treatment, and a relationship between suicidality and fluoxetine-induced akathisia has been noted by others.

Several meta-analyses have shed some additional light on this complex issue. A British meta-analysis of 702 clinical trials involving 87,650 adult patients documented a two-fold increase in suicide attempts in patients receiving SSRIs as compared to placebo.[49] An American meta-analysis of pediatric patients used data from 23 trials involving 4582 patients.[50] This study also found an increased risk of suicidality in patients taking SSRIs, after controlling for the risk associated with suffering from depression. Both British and American regulatory agencies now require warnings about suicidal risks associated with SSRIs. (See also chapter 9.5.5)

At this stage, the wisest course of action is for the practitioner to be particularly observant during the early stages of fluoxetine treatment of a depressed adolescent, to systematically enquire about suicidal ideation before and after treatment is started, and to be especially alert to the possibility of suicidality if SSRI treatment is associated with the onset of akathisia.

One must be careful about the risk of inducing suicidal ideation or behaviour through psychopharmacological activation or disinhibition. Clinicians should be cautious about prescribing medications that may reduce self-control, such as the benzodiazepines, and phenobarbitone (phenobarbital). These drugs also have a high lethal potential if taken in overdose. Montgomery[51] noted that benzodiazepines may disinhibit some individuals who then become aggressive and attempt suicide and that there are suggestions of similar effects from the antidepressants, maprotiline and amitriptyline, the amphetamines, and phenobarbitone. Amphetamines or other stimulant medication should only be prescribed when treating suicidal children and adolescents with attention-deficit hyperactivity disorder.

Possibilities for prevention

Community-based suicide prevention

The principal public health approaches to suicide prevention have been as follows:

- crisis hotlines
- method control
- media counselling to minimize imitative suicide
- indirect case-finding by educating potential gatekeepers, teachers, parents, and peers to identify the 'warning signs' of an impending suicide
- direct case-finding among high-school or college students or among the patients of primary practitioners by screening for conditions that place teenagers at risk for suicide
- training professionals to improve the recognition and treatment of mood disorders.

(a) Crisis hotlines

Although crisis hotlines are available almost everywhere in the United States, research so far has been fairly limited and has failed to show that they impact on the incidence of suicide.[52] Possible reasons for this include the fact that actively suicidal individuals (males and individuals with an acute mental disturbance) do not call hotlines because they are acutely disturbed, preoccupied, or intent on not being deflected from their intended course of action. It also seems that the large majority of callers are females, whereas males are at the greatest risk for suicide, that crisis lines are often busy and there may be a long wait before a call is answered so that callers disconnect, and that the advice that individuals receive on calling a hotline may be stereotyped, inappropriate for an individual's needs, and perceived as unhelpful by the caller.

While each of these deficiencies is potentially modifiable, to date there have been no systematic attempts to do so. Research studies in this area have been sparse and are sorely needed.

(b) Method restriction

Method preference varies by gender and by nationality. In the United States, the most common method for committing suicide is by firearm, and it has been suggested that reducing firearm availability will reduce the incidence of suicide. However, in a natural experiment in the United Kingdom, when self-asphyxiation with coal gas became impossible after the introduction of natural gas, the decline in the suicide rate was marked but short-lived. There is, as yet, no good evidence that reducing access to firearms by gun-security laws has a significant impact on suicides attributable to such, although they do impact on accidental and homicidal deaths from firearms.[53]

(c) Media counseling

The United States Centers for Disease Control have issued sensible guidelines for reporters and editors, pointing to the risks of exaggerated or prominent coverage of youth suicide in general, and of the risks in focusing attention on an individual suicide.[11] These sensible guidelines should be known to child clinicians who are engaged in public-health practice, even though there is, as yet, no good evidence that their application is effective in reducing the suicide rate.

(d) Indirect case-finding through education

Controlled studies have failed to show that classes for high-school students about suicide increase students' help-seeking behaviour when they are troubled or depressed.[54] On the other hand, there is evidence that previously suicidal adolescents are perturbed by exposure to such classes.[51] Such educational programmes seem, therefore, to be both an ineffective mode of case-finding and to carry with them an unjustified risk of activating suicidal thoughts. Educational approaches in schools are dsicussed further in chapter 4.15.4

(e) Direct case-finding

If asked in a non-threatening way, adolescents will provide accurate information about their own suicidal thoughts and/or behaviours.[21] Therefore a sensible approach to suicide prevention is to systematically screen 15- to 19-year-olds (the age group at greatest risk) for previous suicide attempts, recent serious suicidal preoccupations, depression, or complications of substance or alcohol use. Youths identified in this way should be referred for evaluation and, if necessary, treatment.

(f) Training primary care physicians and gatekeepers in the recognition and treatment of mood disorders

Preliminary and, as yet, unreplicated studies in Sweden[55] suggest that education of primary practitioners to identify the characteristics of mood disorders better and to treat these effectively produced a significant reduction in suicide and suicide-attempt rates in women. Because the optimal treatment of adolescent depression is not as well understood as that of adult depression, this is an option that may prove to be useful, but further work is needed.

Further information

Bridge, J.A., Goldstein, T.R. and Brent, D.A. (2006). Adolescent suicide and suicidal behaviour. *Journal of Child Psychology and Psychiatry* **47,** 372–94.

Hawton, K., Rodham, K. and Evans, E.(2006). *By their own hand: deliberate self harm and suicidal ideas in adolescents*. London, Jessica Kingsley

Hawton, K., & Fortune, S. (2008). Suicidal behaviour and deliberate self-harm. In *Rutter's Child and Adolescent Psychiatry,* (eds. M. Rutter, D. Bishop, D. Pine, S. Scott, J. Stevenson, E. Taylor, and A. Thapar) 5th Edition. pp. 647–68. Blackwell, Oxford.

References

1. Robins, E.R., Murphy, G.E., Wilkinson, R.H., *et al.* (1959). Some clinical considerations in the prevention of suicide based on a study of 134 successful suicides. *American Journal of Public Health*, **49**, 888–99.
2. Shaffer, D. (1974). Suicide in childhood and early adolescence. *Journal of Child Psychology and Psychiatry*, **15**, 275–91.
3. Shaffer, D., Gould, M.S., Fisher, P., *et al.* (1996). Psychiatric diagnosis in child and adolescent suicide. *Archives of General Psychiatry*, **53**, 339–48.
4. Brent, D.A., Baugher, M., Bridge, J., *et al* (1999). Age- and sex-related risk factors for adolescent suicide. *Journal of the American Academy of Child and Adolescent Psychiatry*, **38**, 1497–1505.
5. Martunnen, M.J., Aro, H.M., Henriksson, M.M., *et al.* (1991). Mental disorders in adolescent suicide DSM-III-R axes I and II diagnoses in suicides among thirteen- to nineteen-year-olds in Finland. *Archives of General Psychiatry,* **48**, 834–9.
6. Shaffer, D. and Hicks, R. (1993). The epidemiology of child and adolescent suicide. In *The epidemiology of childhood disorders* (ed. B. Pless), pp. 339–68. Oxford University Press, New York.
7. World Health Organization (1996). *World health statistics annual*. WHO, Geneva.
8. Taylor, S.J., Kingdom, D., and Jenkins, R. (1997). How are nations trying to prevent suicide? An analysis of national suicide prevention strategies. Review. *Acta Psychiatrica Scandinavica,* **95**, 457–63.
9. Ministry of Youth Affairs, Ministry of Health, Te Puni Kokiri (Ministry of Maori Development) (1999). *In our hands: New Zealand youth suicide prevention strategy*. New Zealand Ministry of Health, Wellington.
10. Commonwealth Department of Health and Family Services, Mental Health Branch (1997). *Youth suicide in Australia: the national youth suicide prevention strategy*. Australian Government Publishing Service, Canberra.
11. Center for Diseases Control Prevention Guidelines, United States of America (1994). Programs for the prevention of suicide among adolescents and young adults. *Morbidity and Mortality Weekly Report*, **43**, 1–18.
12. Gould, M.S., Fisher, P., Parides, M., *et al.* (1996). Psychosocial risk factors of child and adolescent completed suicide. *Archives of General Psychiatry*, **53**, 1155–62.
13. Gould, M.S. and Shaffer, D. (1986). The impact of television movies: evidence of imitation. *New England Journal of Medicine*, **315**, 690–4.
14. Cuddy–Casey, M. and Orvaschel, H. (1997). Children's understanding of death in relation to child suicidality and homicidality. *Clinical Psychology Review*, **17**, 33–45.
15. Schmidtke, A., Bille–Brahe, U., DeLeo, D., *et al.* (1996). Attempted suicide in Europe: rates, trends and sociodemographic characteristics of suicide attempters during the period 1989–1992. Results of the WHO/EURO Multicentre Study on Parasuicide. *Acta Psychiatrica Scandinavica,* **93**, 327–38.
16. Fergusson, D.M. and Lynskey, M.T. (1995). Suicide attempts and suicidal ideation in a birth cohort of sixteen-year-old New Zealanders. *Journal of the American Academy of Child and Adolescent Psychiatry*, **34**, 1308–17.
17. Remafedi, G., French, S., Story, M., *et al.* (1998). The relationship between suicide risk and sexual orientation: results of a population-based study. *American Journal of Public Health,* **88**, 57–60.
18. Beck, A.T., Weissman, A., Lester, D., *et al.* (1974). The measurement of pessimism: the Hopelessness Scale. *Journal of Consulting and Clinical Psychology,* **42**, 861–5.
19. Beck, A.T., Kovacs, M., and Weissman, A. (1979). Assessment of suicide intention: the Scale for Suicide Ideation. *Journal of Consulting and Clinical Psychology,* **47**, 343–52.
20. Pfeffer, C.R., Conte, H.R., Plutchik, R., *et al.* (1979). Suicidal behavior in latency-age children: an empirical study. *Journal of the American Academy of Child Psychiatry,* **18**, 679–92.
21. Safer, D.J. (1997). Self-reported suicide attempts by adolescents. *Annals of Clinical Psychiatry*, **9**, 263–9.
22. Hawton, K. (1986). *Suicide and attempted suicide among children and adolescents*. Sage, Beverly Hills, CA.
23. Brent, D.A., Holder, D., Kolko, D., *et al.* (1997). A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive therapy. *Archives of General Psychiatry,* **54**, 877–85.
24. National Center for Health Statistics (2000). *Death rates for 72 selected causes by 5-year age groups, race, and sex, 1979–1997*, pp. 485–90. Centers for Disease Control, Atlanta, GA. (Website: http://www.cdc.gov/nchs/datawh/statab/unpubd/mortabs.htm)
25. Neeleman, J. and Wessely, S. (1999). Ethnic minority suicide: a small area geographical study in south London. *Psychological Medicine*, **29**, 429–36.
26. Neeleman, J., Wessely, S., and Lewis, G. (1998). Suicide acceptability in African- and white Americans: the role of religion. *Journal of Nervous and Mental Disease*, **186**, 12–16.
27. Neeleman, J., Halpern, D., Leon, D., *et al.* (1997). Tolerance of suicide, religion and suicide rates: an ecological and individual study in 19 Western countries. *Psychological Medicine*, **27**, 1165–71.
28. Carlsten, A., Allebeck, P., and Brandt, L. (1996). Are suicide rates in Sweden associated with changes in the prescribing of medicines? *Acta Psychiatrica Scandinavica,* **94**, 94–100.
29. Kessler, R., Berglund, P., Borges, G., *et al.* (2005) Trends in Suicide Ideation, Plans, Gestures, and Attempts in the United States, 1990-1992 to 2001-2003. *Journal of the American Medical Association*, **293**, 2487–95.

30. Centers for Disease Control (1998). Attempted suicide among high–school students–United States, 1997. *Morbidity and Mortality Weekly Report*, **47**, 47–9.
31. Brent, D.A., Poling, K., McKain, B., *et al.* (1993). A psychoeducational program for families of affectively ill children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, **32**, 770–4.
32. Mann, J.J. and Stoff, D.M. (1997). A synthesis of current findings regarding neurobiological correlates and treatment of suicidal behavior. *Annals of the New York Academy of Sciences,* **836**, 352–63.
33. Arango, V., Underwood, M.D., and Mann, J.J. (1997). Biologic alterations in the brainstem of suicides. *Psychiatric Clinics of North America,* **20**, 581–93.
34. Nordstrom, P., Samuelsson, M., Asberg, M., *et al.* (1994). CSF 5–HIAA predicts suicide risk after attempted suicide. *Suicide and Life Threatening Behavior*, **24**, 1–9.
35. Juon, H.S. and Ensminger, M.E. (1997). Childhood, adolescent, and young adult predictors of suicidal behaviors: a prospective study of African-Americans. *Journal of Child Psychology and Psychiatry and Allied Disciplines,* **38**, 553–63.
36. Pfeffer, C.R., Klerman, G.L., Hurt, S.W., *et al.* (1991). Suicidal children grow up: demographic and clinical risk factors for adolescent suicide attempts. *Journal of the American Academy of Child and Adolescent Psychiatry*, **30**, 609–16.
37. Pfeffer, C.R., Klerman, G.K., Hurt, S.W., *et al.* (1993). Suicidal children grow up: rates and psychosocial risk factors for suicide attempts during follow-up. *Journal of the American Academy of Child and Adolescent Psychiatry*, **32**, 106–13.
38. Harrington, R., Bredenkamp, D., Groothues, C., *et al.* (1994). Adult outcomes of childhood and adolescent depression. III: Links with suicidal behaviours. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, **35**, 1309–19.
39. Drye, R., Goulding, R., and Goulding, M.E. (1973). No-suicide decisions: patient monitoring of suicide risk. *American Journal of Psychiatry,* **130**, 171–4.
40. Clarke, G.N., Rohde, P., Lewinsohn, P.M., *et al.* (1999). Cognitive-behavioral treatment of adolescent depression: efficacy of acute group treatment and booster sessions. *Journal of the American Academy of Child and Adolescent Psychiatry*, **38**, 272–9.
41. Linehan, M.M. (1993). *Cognitive behavior therapy of borderline personality disorder*. Guilford, New York.
42. Miller, A.L., Rathus, J.H., Linehan, M.M., *et al.* (1997). Dialectical behavior therapy adapted for suicidal adolescents. *Journal of Practical Psychiatry and Behavioral Health*, **3**, 78–86.
43. Tondo, L., Jamison, K.R., and Baldessarini, R.J. (1997). Effect of lithium maintenance on suicidal behavior in major mood disorders. *Annals of the New York Academy of Sciences,* **836**, 339–51.
44. Letizia, C., Kapik, B., and Flanders, W.D. (1996). Suicidal risk during controlled clinical investigations of fluvoxamine. *Journal of Clinical Psychiatry*, **57**, 415–21.
45. Montgomery, S.A. and Montgomery, D. (1982). Pharmacological prevention of suicidal behaviour. *Journal of Affective Disorders*, **4**, 291–8.
46. Fava, M. and Rosenbaum, J.F. (1991). Suicidality and fluoxetine: is there a relationship? *Journal of Clinical Psychiatry,* **52**, 108–11.
47. Beasley, C.M., Jr., Dornseif, B.E., Bosomworth, J.C., *et al.* (1991). Fluoxetine and suicide: a meta–analysis of controlled trials of treatment for depression. *British Medical Journal,* **21**, 685–92.
48. Rothschild, A.J. and Locke, C.A. (1991). Reexposure to fluoxetine after serious suicide attempts by three patients: the role of akathisia. *Journal of Clinical Psychiatry*, **52**, 491–3.
49. Fergusson, D., Doucette, S., Glass, K.C., *et al.* (2005). Association between suicide attempts and selective serotonin reuptake inhibitors: systematic review of randomized controlled trials. *British Medical Journal*, **330**, 396–402.
50. Hammad, T.A., Laughren, T., Racoosin, J., *et al.* (2006). Suicidality in Pediatric Patients Treated with Antidepressant Drugs. *Archives of General Psychiatry*, **63**, 332–9.
51. Montgomery, S.A. (1997). Suicide and antidepressants. *Annals of the New York Academy of Sciences,* **29**, 329–38.
52. Shaffer, D., Garland, A., Gould, M., *et al.* (1988). Preventing teenage suicide: a critical review. *Journal of the American Academy of Child and Adolescent Psychiatry*, **27**, 675–87.
53. Cummings, P., Grossman, D.C., Rivara, F.P., *et al.* (1997). State gun safe storage laws and child mortality due to firearms. *Journal of the American Medical Association,* **278**, 1084–6.
54. Shaffer, D., Vieland, V., Garland, A., *et al.* (1990). Adolescent suicide attempters: response to suicide-prevention programs. *Journal of the American Medical Association,* **264**, 3151–5.
55. Rihmer, Z., Rutz, W., and Pihlgran, H. (1995). Depression and suicide on Gotland: an intensive study of all suicides before and after a depression–training programme for general practitioners. *Journal of Affective Disorders*, **35**, 147–52.

9.2.11 Children's speech and language difficulties

Judy Clegg

Introduction

Speech and language difficulties have a significant impact on the lives of children and their families. This chapter will give an overview of the types of speech and language difficulties children present with and how these are generally classified and diagnosed. Specific Language Impairment (SLI) and speech and language difficulties associated with child psychiatric disorder, specifically disorders of attention and selective mutism will be a focus. The life course of children with speech and language impairments will be described through childhood, adolescence, and adult life. Current management approaches will be presented and evaluated and strategies for effective communication considered.

Clinical features

Typical speech and language development

It is remarkable how quickly and easily most children progress through the typical stages of speech and language development to become competent communicators by the age of 5 years. Much is known about how children acquire speech and language and when these skills are achieved.[1] Children need to be competent communicators prior to starting school, as learning is dependent on adequate speech and language abilities. At school entry age, children are expected to be able to speak clearly, to understand and use complex grammatical structures, to use language for a range of communicative reasons from requesting to negotiating and predicting, to take part confidently in conversations with both children and adults and to have a knowledge of letter names and sounds and to read some single words. The acquisition of these speech and language skills will enable the child to access the

educational curriculum where learning is dependent on both verbal and written language. If children are not competent in these skills then they will experience significant difficulties in their learning from the start of their school career.

Features of speech and language difficulties

Speech and language development can be affected by hearing impairment, visual impairment, general learning disability, epilepsy, and specific syndromes of learning disability such as Down's syndrome, and Fragile X syndrome. In these examples, speech and language difficulties are usually attributed to and explained by an aetiological cause. However, speech and language difficulties do occur in the absence of an obvious identifiable cause and are therefore considered as a specific impairment, e.g. SLI.

Prevalence rates of speech and language difficulties vary and are dependent on the criteria used to define and classify them. Law *et al.*[2] report prevalence rates in children as high as 24.6 per cent whereas rates for SLI are much lower between 3 and 7 per cent.[3] Importantly, speech and language difficulties can persist over time and often have a negative impact on the child's education and general well-being.

(a) Speech difficulties

A speech difficulty reduces a child's intelligibility and may result in speech sounds being omitted, substituted with another sound or distorted. Speech difficulties can be evident when a child says single words, sentences, and participates in conversation. The physical articulation of speech sounds is affected by physiological and structural abnormalities, such as cleft lip/palate, and neurological impairments leading to dysarthria characterized by weakness and/or in-coordination of the speech musculature system. There is another group of children who have phonological speech difficulties. These children have an intact speech musculature system but have not managed to acquire all the speech sounds of their language and so can only use a limited range, which subsequently limits their intelligibility.

(b) Language difficulties

Language difficulties can involve problems in the development of both comprehension and production.

(i) Vocabulary difficulties

Restricted word knowledge and poor development of the understanding of word meanings result in small vocabularies. Some children have impoverished vocabularies but other children can have specific word finding or retrieval difficulties. Here, the child knows the word he wants to say but is unable to retrieve it accurately and quickly. This is usually evident by 'searching' behaviours where the child may substitute the word for a related word, use a filler word such as 'thingy' or 'stuff', gesture the word instead of saying it or say the first sound of the word but not the rest. For example,

> ICE SKATING: 'I can't do that thing . . . erm . . . you know . . . where you put sharp shoes on . . . I always fall over'.
> PLUM: 'well, I don't really like that one which smells like soil and is purple and juicy'

These problems may not only be due to lexical difficulties but also problems retrieving the right phonological sounds of the word. Cognitive impairments in information processing, specifically short-term and phonological working memory have been associated with problems in vocabulary learning.[4]

(ii) Syntax difficulties

Children often have difficulties in their understanding and use of syntax and as a consequence find it very difficult to not only understand language but also to construct sentences in order to use language to communicate effectively, for example giving a narrative where past events are described and future events predicted. Common problems are learning how to use inflections to mark different tenses and understanding as well as constructing complex sentences such as passives. The child in the following example has lots of syntax difficulties as well as word finding difficulties and it is clear how this affects his ability to convey verbal information. The correct forms the child is attempting are shown in brackets.

'They erm . . . was . . . erm . . . goed to make (made) some vegetable circles (pizzas) and rolls (they rolled) it (the dough) out because that's what you do first and he was reading the menu (recipe) as well and then they is erm . . . erm . . . erm . . . printing (cut) them out and then they put them in the oven because they'll taste crunchy (to cook) and then erm . . . then they took them out of the oven so they be . . . er . . . get . . . cool down (could cool down) and then you would take (ate) them'.

(iii) Social communication difficulties

Children with speech and language difficulties often show associated problems in social communication behaviours, also referred to as pragmatics. These can be both verbal and non-verbal and include difficulties with eye contact, initiation, turn taking, interaction, sharing, requesting, and responding. Higher level social communication abilities can also be affected such as inferring information, giving the listener adequate information and self-monitoring. Ultimately, these can all hinder effective communication between the child and others and also expose the child to negative social experiences, particularly with their peers. For some children, the social communication difficulties may be an intrinsic part of a developmental disorder where speech and language difficulties are evident, for example children with autistic spectrum disorders. In other children, it is important to note that these behaviours can develop as a secondary consequence of poor communication skills due to the speech and language difficulty.

Classification

Speech difficulties can occur in isolation without the presence of language difficulties. Language difficulties can also occur without the presence of speech difficulties but often speech and language difficulties co-occur together. Children can have difficulties with both language comprehension and language production.

Within child psychiatry, both the ICD-10[5] and DSM-IV[6] systems categorize developmental speech and language difficulties. However, there is little robust empirical evidence to support the subtyping of speech and language difficulties. Children are usually classified according to whether the speech and language difficulty is specific, i.e. cognitive development is age appropriate and if there are any co-morbid aetiological or functional explanations. Descriptions of the type of speech and language difficulty involve identifying how the speech and language system is disrupted, describing the levels of impairment, and how this is impacting on the child's communication and their access to learning.

Diagnosis and differential diagnosis

Descriptions of developmental speech and language disorders

Children's language is said to be 'delayed' when their language abilities are behind those expected for their chronological age and 'impaired' or 'disordered' when a language delay does not resolve and the child continues to experience significant and severe problems. Several established diagnoses of developmental speech and language disorders are described below:

(a) Cleft lip and palate

A cleft/lip palate results from the incomplete fusion of the hard or soft palate in the embryonic stages of development. A cleft palate can be accompanied by a cleft lip or either one can occur independently. In the United Kingdom, cleft lip/palate is repaired in the first few months of life. However, some children can be left with fistulas and velopharyngeal incompetency, which significantly affects speech development and intelligibility. Children with cleft lip/palate receive speech and language therapy from birth onwards. At birth the focus of attention is primarily on feeding and then the development of speech and language.

(b) Dysarthria

Dysarthria is a speech disorder due to neurological impairment which affects how the speech musculature system functions. Children with cerebral palsy often have dysarthria, which makes their speech slow, weak, and uncoordinated. There may be a mild, slight slurring of speech to profound dysarthria where a child cannot produce any intelligible sounds or words. Children with moderate and severe dysarthria have shallow breathing which is insufficient to sustain speech and/or a low-pitched voice, nasal speech, and a reduced range of vowels and consonants that can be produced accurately.

(c) Developmental phonological disorder

Unlike cleft lip/palate and dysarthria, phonological speech disorders involve the child's developing speech sound or phonological system. The child's speech is difficult to understand because the child makes speech sound errors which are either due to the speech sound system developing more slowly or in an atypical way and this is not a result of obvious structural, sensory, or neurological impairments. Often, there are systematic patterns of errors in the child's speech, for example the child always replaces the 's' sound with a 'd' sound. Auditory processing and discrimination skills have been implicated in the development and maintenance of this disorder. Over time, phonological disorders often resolve with speech and language therapy input. However, for some children they are severe and do persist into adult life.

(d) Childhood apraxia of speech (CAS)

This developmental speech disorder is characterized by both speech and non-speech behaviours. The speech sound errors are inconsistent and are accompanied with oral movement difficulties in drooling, feeding, and blowing. Reduced early verbal behaviours such as babbling are often evident. CAS often co-occurs with motor apraxia but for some children, only speech and oral movements are affected. There is some debate as to the existence of CAS as there is no obvious cause although both neuromotor planning and the organization of the child's phonological system have been implicated. CAS is often a label given to children where the speech disorder has persisted despite intervention and oral non-speech movements are affected. See Dodd[7] for a detailed review of children's speech disorders.

(e) Fluency disorders

Although classified under speech disorders, stuttering is not an articulatory or phonological difficulty. There are no structural abnormalities and the child usually has a typically developing phonological system. Core stuttering behaviours include part-word or whole-word repetitions, revisions, pauses, blocks, sound prolongations, and obvious struggling behaviours such as jerky head movements. Secondary behaviours result from the stuttering and generally help the individual to avoid stuttering. For example, circumlocution where the speaker substitutes a word he knows he will stutter on for an easier word and environmental control such as avoiding the use of the telephone or talking to certain people. Fluency disorders are often identified in young children before the age of 5 years although many children experience a period of normal non-fluency usually between the ages of 2 and 5 years, which is not severe and resolves spontaneously.

(f) Learning disability

Level of cognitive ability is the strongest predictor of language ability and therefore language development is certainly affected in learning disability. The sequence of language development is similar to that found in typical development but with mild to moderate to severe and profound delay. A child with a profound learning disability may never develop an intent to communicate whereas another child may have established an intent but no verbal language and uses some signs or symbols to communicate instead. For children with mild and moderate learning disability, language abilities plateau with no further improvement, usually in adolescence at a level below the child's chronological age.

It should be noted that specific patterns of speech and language development have been identified in specific syndromes of learning disability. Down's syndrome is characterized by superior vocabulary development to grammatical development and children with William's syndrome often appear as competent communicators but do have significant language learning problems. Speech and fluency problems are common in learning disability and vary according to the aetiology of the learning disability. For example, conductive hearing loss and articulatory speech problems occur where there is cranio-facial involvement.

(g) Acquired childhood aphasias

Acquired aphasias refer to a loss or deterioration in language ability after a period of typical language development. The child acquires language but then loses these language abilities, usually between 3 and 7 years of age. Causes of childhood aphasia include open and closed head injury, cerebrovascular lesions, cerebral infections, cerebral tumours, and epilepsy. Landau Kleffner (first described by Landau and Kleffner in 1957)[8] is an acquired aphasia where language deteriorates after a period of typical language development and the deterioration in language is usually, although not always accompanied with a seizure disorder. Receptive language is severely affected with expressive language problems as well, often word finding difficulties. See Lees[9] and Deonna[10] for a complete review.

Specific language impairment (SLI)

Specific language impairment (SLI) is a term used to describe language impairment (and additional speech impairment) where there is no identifiable medical, neurological, sensory, or functional cause and where cognitive ability measured by non-verbal intelligence (IQ) is within the normal range. Therefore, there is a discrepancy between language and cognitive ability with the exclusion of any obvious causes for the language impairment. Diagnosis of SLI according to exclusionary and discrepancy criteria is dependent on standardized language and cognitive psychometric assessments. However, there is continuing debate regarding which criteria to use to establish a meaningful discrepancy between language and cognition. ICD-10,[5] for example adopt a strict criteria of language skills at least two standard deviations below the level expected for the child's chronological age and language skills at least one standard deviation below the child's level of non-verbal IQ. More liberal criteria advocates a non-verbal IQ of 75 or above with language abilities often only one SD below the mean. Proponents of liberal criteria claim that more stringent criteria may fail to identify children who are at risk of poor long-term outcomes. However, liberal criterion may identify children who simply perform at the lower end of the normal distribution of language ability. It should be recognized that different criteria are used. Although the diagnosis of SLI stipulates good cognitive ability, some specific cognitive deficits in phonological memory, verbal, and visuo-spatial memory and symbolic play are evident and thought to underpin the language impairment.

SLI is considered to affect 3–7 per cent of all children.[3] The use of the exclusionary and discrepancy criteria to define SLI means that as a group, children with SLI are very heterogenous with impairments in many areas of language. Although useful, attempts to subtype SLI[11] have not yet proved clinically robust. However, children with SLI are considered to show disproportionate difficulties in vocabulary and syntax compared to other aspects of language.

(a) Aetiology of SLI

Research in SLI primarily focuses on trying to establishing a cause. SLI is a heritable disorder and much research is underway to try and establish the genetic basis.[12,13] SLI is of particular interest to researchers because of the unusual dissociation between cognitive and language ability and whether this dissociation is explained by innate modular theories of language acquisition or more general cognitive processing deficit theories. Some attempt has been made to identify genetic markers of SLI such as a phonological memory deficit[12] which stems from the research into general cognitive processing deficits as underlying SLI and a specific tense marking deficit[14] or a syntax representational deficit[15] which argues for the disruption of innate modular components of language.

(b) Diagnostic overlaps between SLI and autistic spectrum disorders (ASD)

(i) Pragmatic language impairment

Autism and autistic spectrum disorders (ASD) are discussed extensively in Chapter 9.2.2 of this text. Language and communication difficulties are central to both SLI and ASD. However, the fundamental difference between these disorders is the severity and pervasiveness of the social communication impairment. In SLI, social communication difficulties are considered secondary to the language impairment where children with speech and language difficulties will have problems in developing appropriate social communication skills. In ASD, the social communication impairment is an intrinsic part of the disorder and does not develop as a secondary consequence of a speech and language impairment. Due to the increase in the identification of ASD and the use of the autistic spectrum many more children with milder difficulties are being diagnosed with ASD. This has led to some researchers proposing that there are overlaps between SLI and ASD.

Semantic–pragmatic disorder was first described in the 1980s as a subtype of SLI[16,17] and was a label used to describe children with comprehension problems, echolalia, behaviour difficulties, and difficulty with non-literal language, semantics and pragmatics. At the time, these children were not considered as autistic. However, the increasing use of the autistic spectrum led to debates about whether semantic–pragmatic disorder exists as a separate category of SLI or whether it should be included on the autistic spectrum.[18,19]The crucial issue was whether the social impairment was intrinsic to the language disorder or a secondary consequence of the language disorder. To address this, researchers have attempted to show differences in pragmatic abilities between children with SLI, ASD, and typically developing children. For example, Bishop and Norbury[19] identified a subgroup of SLI children who show a profile of Pragmatic Language Impairment (PLI). These children showed inappropriate behaviours across aspects of social communication including initiating conversations, understanding subtle aspects of language such as humour and sarcasm, adapting their communication to different contexts, understanding and using non-verbal communication, and engaging in conversations about specific interests. Importantly, these children did not show the non-verbal repetitive behaviours typically characteristic of autism. Overall it is argued[19] that there are continuities between autism and specific language impairment but not all children with pragmatic impairments have autism. Therefore, pragmatic language impairment alone should not be used to make diagnoses of autistic spectrum disorders. It is recognized that there are conflicting opinions about the increasing evidence that indicates continuity between disorders that have traditionally been regarded as distinct from one another. However, assessment should consider whether a child's social communication difficulties are being compounded by language difficulties as amelioration of the language difficulties may improve the child's social communication.

(c) Associations between language and behaviour in child psychiatric disorders

Children with primary psychiatric disorders often have a history of developmental problems which can include speech and language delay. Children with primary speech and language disorders are at greater risk of developing behaviour difficulties than children without speech and language disorders. Various mechanisms have been put forward to try and explain this association. These include common antecedents such as low intelligence and deprivation,[20] environmental factors where language stimulation is negatively affected by poor parent–child interactions or the child's inability to attend to the language stimulus,[21] the psychosocial rejections and academic failure experienced by children with communication impairments affecting their self-confidence and self-esteem and therefore their subsequent emotional behaviour development,[22–24] and a neurodevelopmental abnormality or immaturity as a shared

underlying cause.[25,26] Although the simplicity of these mechanisms is appealing, identifying, and differentiating them is certainly complex.

More recently studies have shown that children with primary psychiatric disorders can have undetected speech and language disorders.[27,28] ADHD is one of the most commonly reported psychiatric disorders associated with speech and language difficulties.[29] In ADHD, language difficulties consist of both receptive and expressive problems[30] and pragmatic language difficulties[31,32] such as excessive talking and poor topic maintenance. Although studies have not identified a distinct profile of speech and language difficulties in ADHD, they should be considered in assessment and management.

It has been hypothesized that these undetected speech and language disorders somehow play a role in the development and maintenance of the psychiatric disorder and even that the psychiatric disorder is secondary to the undetected speech and language disorder (see mechanisms above). There is very limited evidence available to specify what the associations are and importantly, it may be that referral practices play a role where the psychiatric problem takes priority and the child is referred to mental health services first. In the United Kingdom SLT and mental health services are usually very separate and it is not common that SLTs work in mental health services. The identification of previously undetected speech and language disorders in the studies above are really the late identification of pre-existing difficulties. Nevertheless, management of childhood psychiatric disorders should consider if speech and language difficulties are a factor in the child's behaviour as they may have an impact on how the child is managed, particularly with respect to participation in verbal therapies and education.

(c) Selective mutism

This childhood disorder is described as the persistent refusal to talk in certain social situations despite being able to talk in other situations. The most common pattern is talking at home but not at school and the refusal to talk cannot be better accounted for by a communication disorder or difficulties in understanding and using spoken language. Pervasive developmental disorder or psychotic disorder should also be excluded. The mutism must last for more than a month (this cannot be the first month of school) and interfere significantly with educational progress, social communication with others, and occupational achievement. In the case of a bilingual child, it is suggested that the mutism should persist for at least 6 months and be present in both the first and second language before diagnosis.[33]

Selective mutism is rare and as a result only a limited number of studies reporting the epidemiology of this disorder are available. Prevalence figures estimate a prevalence of approximately 0.75 to 0.80 per cent [34,35] and it is slightly more common in girls[36] with an onset between the ages of 3 and 5 years.[34] Although there is no clear consensus to explain the cause(s) of the disorder, social phobia, and anxiety are certainly involved. Co-morbidity with behaviour problems, communication difficulties and developmental delay are also found which indicates that a multi-factorial aetiology is the best explanation. Data regarding the long-term outcomes is scarce but there are indications that with early intervention improvements are made but often children are still left feeling uncomfortable in some speaking situations.

(d) Intervention for selective mutism

Intervention approaches include pharmacology, cognitive behaviour therapy, family therapy, psychodynamic therapy, and speech and language therapy. Although some success with fluoxetine has been reported[37] this has not been widely replicated. A behavioural approach considers the disorder as learned behaviour and techniques including contingency management, shaping and stimulus fading, systematic desensitization, and self-modelling are advocated. Family therapy aims to identify whether there are difficulties in family relationships that are contributing to the mutism and attempts to work with the whole family to foster more positive relationships. Psychodynamic approaches involve techniques of play therapy and art therapy to identify the underlying reasons for the mutism and to help the child to express the possible unconscious conflicts he is experiencing. Although children with selective mutism are expected to have good speech and language skills, several studies have reported a high incidence of speech and language difficulties such as articulation and expressive and receptive language difficulties.[38,39] In these circumstances, speech and language therapy is used as a valuable adjunct to the other approaches. Speech and language therapy aims to facilitate the child's communication rather than resolving the underlying causes of the mutism. Therapists work with the child to desensitize him/her to communicating with others by considering the child's communication environment and the communication load of the tasks he is expected to engage in. A hierarchy of stages is followed from easy to hard speech tasks within easy to hard speaking situations. A multi-modal perspective incorporating a combination of the above approaches is advocated. The combination of family involvement with cognitive behavioural, speech and language, psychodynamic and family involvement meet the multi-factorial needs of this disorder. See Cohan *et al.*[40] for a detailed review of the efficacy of the different intervention approaches described. Although social anxiety is the predominating feature of selective mutism, the resulting lack of communication is challenging. Several strategies to facilitate communication with these children are presented and these can easily be incorporated into other intervention approaches:

- Check that there are no speech and language difficulties that may be contributing to the mutism. For some children, although early speech and language difficulties may have resolved the child may still feel under confident in their talking.
- One-to-one settings are most comfortable and try and include the familiar person, (usually the primary carer) who the child communicates regularly with in your interventions to start with. After a while the child may be able to manage this setting without the familiar person. This process may have to be repeated to encourage the child to talk in another setting.
- Encourage and accept non-verbal communication such as head nods, writing, drawing, and gesture as well as verbal communication. Non-verbal communication is easier than talking for most of these children.
- Follow a hierarchy of verbal communication from easy to hard, most children find whispering and talking quietly easier than loud talking.
- Consider the complexity of the task the child is expected to engage in. Questions, which are factual or only require a yes or

no response are much less confrontational than questions, which ask the child about their feelings or opinion.

Further management considerations are detailed in Johnson and Wintgens.[41]

Course and prognosis

Life course and outcomes

An interest in the long-term outcomes of children with speech and language difficulties has emerged fairly recently. Historically, it was considered that primary speech and language difficulties resolved over time with no implications for other areas of development. This is certainly not true and much more is now known about the developmental trajectories through childhood, adolescence, and into adult life. Generally, children with speech and language difficulties continue to show difficulties not only in communication but in cognition, behaviour, educational attainment, and psychosocial functioning. This section will focus on the life course and outcomes of children with profiles of primary speech and language difficulties where there is no cognitive deficit. Cognitive ability is a powerful predictor of development, and therefore the outcomes of children with speech and language difficulties associated with cognitive delay are usually attributed to level of IQ rather than the specific speech and language difficulties themselves.

The impact of a speech and language impairment over the lifespan

Speech and language impairment leads to impoverished communication skills, which certainly impact on other areas of development. The developmental trajectories of children with primary speech and language impairment show impaired receptive and expressive language development in later childhood, adolescence, and even adult life. Cognition, as measured by non-verbal IQ has been shown to fluctuate and even deteriorate over time, particularly in adolescence and later life. Research to date suggests that any deterioration is temporary and resolves but much more needs to be known about this.[42,43] Children with speech and language difficulties are more likely to experience emotional and behavioural problems (see earlier section). During the course of childhood, the risk of developing emotional and behavioural problems seems to increase with obvious negative implications for other areas of development and functioning. There is a possible association between SLD and the development of antisocial behaviour in early adult life. At the age of 19 years participants with SLD in a Canadian community sample did not show high levels of aggression but did have higher rates of arrests and convictions.[26]

The increase in social and behaviour problems may be the result of communication and interaction problems young people experience in conjunction with the increasing social and academic demands placed on them. However, it is hard to disentangle cause and effect when so many variables interact. Socio-economic status (SES), learning ability, and type of educational placement no doubt also make a contribution. Therefore, SES and IQ are probably implicated in the development of social and behavioural difficulties as well as the speech and language impairment. In adult life, severe mental health conditions such as schizophrenia, depression, and personality disorder have been linked with early histories of and persisting severe speech and language impairment.[44]

There are strong associations between speech, language, and literacy development and in fact, speech and language are fundamental in learning to read and write. In order to learn, children need to be competent communicators when they start school. Language is the medium through which children are expected to learn and literacy is dependent on identifying and discriminating speech and letter sounds. Children with persisting speech and language difficulties are at risk for literacy problems and subsequent low academic achievement. Children with speech and language difficulties find it much harder to learn to read and write than children without speech and language difficulties. Educational attainment is dependent on literacy and therefore these children are very disadvantaged. Studies measuring educational attainment using Standard Assessment Tests (SAT) have found that children with speech and language difficulties gain significantly lower SAT scores than controls.[45] In older individuals, attainment at GCSE and A level is also affected. Children with speech and language difficulties are often bullied by their peers and can be targets for victimization.[46] Studies suggest that this is due to their odd communication particularly difficulties with social communication behaviour. Unclear speech is also a significant factor. In adult life, high levels of social maladaption and poor psychosocial functioning were found in adults in their mid 30s with SLI. Employment, relationships, independent living, and health were all significantly and negatively affected.[44] See Clegg[47] for a full review.

Identification of children at risk

There are variations in outcomes and not all children with speech and language impairment are at risk of later negative outcomes. Issues of variability and risk and resilience are not yet fully understood. Risk factors for poor outcomes in later life are the severity of the initial impairment, involvement of receptive language, low IQ, and low SES. Resilience factors are the presence of pure speech difficulties only, high IQ, high SES, and access to specialist support. A critical age hypothesis[48] proposes that any speech and language difficulties that impact on a child's communication after 5 years should be considered as significant and prioritized for intervention.

Speech and language difficulties should not be underestimated in terms of their impact on children's lives. With respect to clinical management, the risk factors should be considered in identifying those children at particular risk. In the United Kingdom, there is specialist educational provision now available that provides post-16 years provision for some young adults to try and reduce negative outcomes, e.g. supporting individuals to gain further qualifications and to enter the work place. Information about long-term outcomes can certainly inform the individuals themselves, their families, and other professionals about prognoses and how best to support individuals to meet their learning and other needs effectively.

Management

Speech and language therapists (SLTs) work in education, health, and social care, voluntary organizations and independent practice. A SLT is an important member of the interdisciplinary team and will lead on the assessment, differential diagnosis, intervention with, and management of individuals with communication and swallowing disorders. In recent years due to the shift to inclusion, many paediatric SLTs now work in schools with educational

professionals as part of the school team. This enables the SLT to contribute to statements of special education needs, formulating independent learning plans (ILP), and delivering the curriculum to make it accessible for children with speech and language difficulties. SLTs also work in community services such as the government Sure Start programme, primary and secondary health care, e.g. acute hospital settings and community clinics, specialist health services, e.g. child development centres and education, e.g. from preschool, mainstream and special schools/resourced provision. The inclusion of SLTs in Child and Adolescent Mental Health Services (CAMHS) is increasing but this is not consistent across the United Kingdom. However, there is now recognition that the SLT can have a valuable role within the CAMHS[49,50] in terms of the identification of any communication difficulties and the subsequent impact on mental health and offering intervention that will help to ameliorate the communication difficulty and facilitate communication. Within CAMHS, a SLT may be employed in a specialist unit, child psychiatry outpatient service, or other. Differential diagnoses are usually made in collaboration with the team and a SLT will use a combination of formal and informal assessment. Formal assessment consists of using measures that compare the child's speech and language abilities with abilities that are expected for the child's chronological age. Various types of assessments are available criterion referenced, standardized, developmental scales, and observational. Intervention may be direct, i.e. individual or as part of a group or indirect, i.e. through working with the family or in the context of the child's classroom and school.

Identification of speech and language difficulties in children

The following checklist may be helpful when working with young school age children in identifying speech and language difficulties and initiating referral to speech and language therapy services.

(a) Speech

1 Is the child's speech difficult to understand?
2 Does the child miss out sounds from words?
3 Is the child less intelligible when speaking in sentences than in single words?
If yes, to one or more of the above then speech difficulties are evident and referral to speech and language therapy is advised.

(b) Comprehension

1 Can the child follow and engage in conversations with both children and adults?
2 Can the child carry out verbal instructions correctly or does he need lots of prompting?
3 Can the child understand a range of concepts such as time and space?
If no, to one or more of the above then comprehension difficulties are evident and referral to speech and language therapy is advised.

(c) Production

1 Does the child have a wide range of vocabulary?
2 Can the child use more complex sentences such as the 'the boy ran for the bus because he was late' rather than the boy was late and ran for the bus?
3 Is the child able to give a narrative of a past and future event?
4 Does the child frequently say non-specific words such as 'stuff' or 'thingy' or hesitate when talking?
If no to 1 to 3 and yes to 4 then production difficulties are evident and referral to speech and language therapy is advised.

(d) Social communication

1 Is the child willing to participate in conversations and does he enjoy this?
2 When the child talks, is it meaningful and relevant to the conversation and/or situation?
3 Does the child use appropriate verbal and non-verbal communication behaviours?
If no to one or more of the above the social communication difficulties are evident and referral to speech and language therapy is advised.

(e) Education and social activities

1 How is the child functioning at school?
2 Is the child able to access and participate in a range of social activities?
3 Are the speech and language difficulties long standing and persisting beyond the age of 5 years?
If the child is struggling to meet the demands of school and presents with persisting speech and language difficulties then this is a cause of concern for later development and outcomes. Referral to speech and language therapy is advised.

General principles for working with children with speech and language difficulties

It may be useful to consider these general principles when working with children with speech and language difficulties

1 Consider your own communication and adapt it to:
- Offer forced choice answers.
- Break up long instructions and sentences into short steps.
- Slowdown delivery and use pauses.
- Use short simple sentences with familiar vocabulary and avoid ambiguous language.
- Use visual strategies such as pictures, real objects, and symbols to support spoken language.
- Remember that children with spoken speech and language difficulties will also have written language difficulties.

2 It is always challenging when you are unable to understand what a child is saying to you. While there is no perfect solution, the following strategies will help:
- Reassure the child that you are interested in what they are trying to tell you.
- Be honest and say that you don't understand but also make it clear which parts you did and did not understand.
- Offer a choice of possible answers to the child, as this will reduce the number of choices available to you to guess from.
- Ask the child to show you something to help or can the child describe it to you or point to it or draw it.
- If these are still not successful, reassure the child that you are interested in what they are trying to tell you and that you will try again later.

Further information

Speake, J. (2004). *How to identify and support children with speech and language difficulties*. LDA, Cambridge, UK.

This is a useful text, which describes children's speech and language difficulties and offers practical strategies for assessment and management.

www.afasic.org.uk

Afasic is a UK charity that offers information and advice to children with speech and language difficulties and their families as well as professionals working in this area.

www.ican.org.uk

ICAN is a UK charity that funds specialist educational provision for children and adolescents with severe speech and language difficulties as well as offering information and advice.

www.rcslt.org

The Royal College of Speech and Language Therapists (RCSLT) is the professional body for practicing speech and language therapists in the UK. Information about speech and language therapy services is available on this website.

References

1. Buckley, B. (2003). *Children's communication skills. From birth to five years*. Routledge, London, UK.
2. Law, J., Boyle, J., Harris, F., *et al.* (2000). The relationship between the natural history and prevalence of primary speech and language delays: findings from a systematic review of the literature. *International Journal of Language & Communication Disorders*, **35**, 165–88.
3. Tomblin, J.B., Records, N., Buckwalter, P., *et al.* (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, **40**, 1245–60.
4. Gathercole, S., Service, E., Hitch, G., *et al.* (1999). Phonological short term memory and vocabulary development: further evidence on the nature of the relationship. *Applied Cognitive Psychology*, **13**, 65–77.
5. World Health Organization (WHO). (1993). *Mental disorders: a glossary and guide to their classification in accordance with the 10th revision of the international classification of diseases* (*ICD 10*). World Health Organization, Geneva.
6. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn) (DSM-IV). American Psychiatric Association, Washington, DC.
7. Dodd, B. (2005). *Differential diagnosis and treatment of children with speech disorders*. Whurr, London, UK.
8. Landau, W.M. and Kleffner, F. (1957). Syndrome of acquired aphasia and convulsive disorder in children. *Neurology*, **7**, 523–30.
9. Lees, J. (2005). *Children with acquired aphasias* (2nd edn). Whurr Publishers, London.
10. Deonna, T. (2004). Acquired epileptic aphasia (AEA) or Landau-Kleffner syndrome: from childhood to adulthood. In *Speech and language impairments in children. Causes, characteristics, intervention and outcome* (eds. D.V.M. Bishop and L.B. Leonard). Psychology Press, Hove, UK.
11. Rapin, I. and Allen, D. (1983). Developmental language disorders: nosologic considerations. In *Neuropsychology of language, reading, and spelling* (ed. U. Kirk), pp. 155–84. Academic Press, New York.
12. Bishop, D.V.M., Laws, G., Adams, C.V., *et al.* (2006). Distinct genetic influences on grammar and phonological short-term-memory deficits: evidence from 6-year-old twins. *Genes, Brain and Behaviour*, **5**, 158–69.
13. Marcus, G.F. and Fisher, S.E. (2003). FOXP2 in focus: what can genes tell us about speech and language? *Trends in Cognitive Sciences*, **7**, 257–62.
14. Rice, M. (2000). Grammatical symptoms of specific language impairment. In *Speech and language impairments in children. Causes, characteristics, intervention and outcome* (eds. D.V.M. Bishop and L.B. Leonard). Psychology Press, Hove, UK.
15. Van der Lely, H.K.J. (2005). Domain-specific cognitive systems: insight from grammatical-specific language impairment. *Trends in Cognitive Sciences*, **9**(2), 53–9.
16. Bishop, D.V.M. and Rosenbloom, L. (1987). Classification of childhood language disorders. In *Language development and disorders: clinics in developmental medicine* (eds. W. Yule and M. Rutter). Mackeith Press, London.
17. Rapin, I. and Allen, D. (1987). Developmental dysphasia and autism in pre-school children: characteristics and subtypes. *Proceedings of the first international symposium on specific speech and language disorders in children*. Association for All Speech Impaired Children, London.
18. Bishop, D.V.M. (2004). Pragmatic language impairment: a correlate of SLI, a distinct subgroup, or part of the autistic continuum? In *Speech and language impairments in children. Causes, characteristics, intervention and outcome* (eds. D.V.M. Bishop and L.B. Leonard). Psychology Press, Hove, UK.
19. Bishop, D.V.M. and Norbury, C.F. (2002). Exploring the borderlands of autistic disorder and specific language impairment: a study using standardised diagnostic instruments. *Journal of Child Psychology and Psychiatry*, **43**, 917–29.
20. Cantwell, D.P. and Baker, L. (1977). Psychiatric disorder in children with speech and language retardation. *Archives of General Psychiatry*, **34**, 583–91.
21. Prizant, B., Audet, L., Burke, G., *et al.* (1990). Communication disorders and emotional/behavioural disorders in children and adolescents. *Journal of Speech and hearing research*, **55**, 179–92.
22. Gertner, B.L., Rice, M.L., and Hadley, P.A. (1994). Influence of communicative competence on peer preferences in a preschool classroom. *Journal of Speech and Hearing Research*, **37**, 913–23.
23. Howlin, P. and Rutter, M. (1987). The consequences of language delay for other aspects of development. In *Language development and disorders* (eds. J. Yule and M. Rutter). Mackeith Press, Oxford, UK.
24. Redmond, S.M. and Rice, M.L. (1998). The socio-motional behaviours of children with SLI: social adaptation of social deviance? *Journal of Speech, Language, and Hearing Research*, **41**, 588–700.
25. Beitchman, J.H. (1985). Speech and language impairment and psychiatric risk: toward a model of neurodevelopmental immaturity. *Symposium on Child Psychiatry*, **8**, 721–35.
26. Brownlie, E.B., Beitchman, J.H., Escobar, M., *et al.* (2004). Early language impairment and young adult delinquent and aggressive behaviour. *Journal of Abnormal Child Psychology*, **32**, 453–67.
27. Cohen, N.J., Barwick, M.A., Horodezky, N.B., *et al.* (1998). Language achievement and cognitive processing in psychiatrically disturbed children with previously identified and unsuspected language impairments. *Journal of Child Psychology and Psychiatry*, **39**, 865–77.
28. Gilmour, J., Hill, B., Place, M., *et al.* (2004). Social communication deficits in conduct disorder: a clinical and community service. *Journal of Child Psychology and Psychiatry*, **45**, 967–78.
29. Tirosh, E. and Cohen, A. (1998). Language deficits in ADD: a prevalent co-morbidity. *Journal of Child Neurology*, **13**, 493–97.
30. Oram, J., Fine, J., Okamoto, C., *et al.* (1999). Assessing the language of children with attention deficit hyperactivity disorder. *American Journal of Speech-Language Pathology*, **8**, 72–80.
31. Kim, O.H. and Kaiser, A.P. (2000). Language characteristics of children with ADHD. *Communication Disorders Quarterly*, **21**, 154–65.
32. Tannock, R., Purvis, K.L., and Schachar, R. (1993). Narrative abilities in children with attention deficit hyperactiviy disorder. *Journal of Abnormal Child Psychology*, **21**, 103–17.

33. Toppleberg, C.O., Tabors, P., Coggins, A., *et al.* (2005). Differential diagnosis of selective mutism in bilingual children. *Journal of the American Academy of Child and Adolescent Psychiatry*, **44**, 592–5.
34. Cline, T. and Baldwin, S. (1994). *Selective mutism in children*. Singular, San Diego, CA.
35. Bergman, R.L., Piacentini, J., and McCracken, J.T. (2002). Prevalence and description of selective mutism in a school-based sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, **41**, 938–46.
36. Cunningham, C.E., McHolm, A., Boyle, M.H., *et al.* (2004). Behavioural and emotional adjustment, family functioning, academic performance, and social relationships in children with selective mutism. *Journal of Child Psychology and Psychiatry*, **45**, 1363–72.
37. Dummit, E., Klein, R.G., Tancer, N.K., *et al.* (1996). Fluoxetine treatment of children with selective mutism: an open trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, **35**, 615–21.
38. Steinhausen, H.C. and Juzi, C. (1996). Elective mutism: an analysis of 100 cases. *Journal of the American Academy of Child and Adolescent Psychiatry*, **35**, 847–56.
39. Baltaxe, C.A.M. (1994). *Communication issues in selective mutism*. Paper presented at the American Speech-Language-Hearing Association Convention, New Orleans.
40. Cohan, L., Chavira, D.A., and Stein, M.B. (2006). Practitioner review: psychosocial interventions for children with selective mutism: a critical evaluation of the literature from 1990–2005. *Journal of Child Psychology and Psychiatry*, **47**, 1085–97.
41. Johnson, M. and Wintgens, A. (2006). *Selective mutism resource manual*. Speechmark Publishing, Oxon, UK.
42. Botting, N. (2005). Non-verbal cognitive development and language impairment. *Journal of Child Psychology and Psychiatry*, **46**, 317–26.
43. Botting, N. (2006). The interplay between language and cognition in typical and atypical development. In *Language and social disadvantage: theory into practice* (eds. Clegg, J. and J. Ginsborg). Wiley, London.
44. Clegg, J., Hollis, C., Mawhood, L., *et al.* (2005). Developmental language disorders—a follow up in later adult life. Cognitive, language and psycho-social outcomes. *Journal of Child Psychology and Psychiatry*, **46**, 128–49.
45. Nathan, L., Stackhouse, J., Goulandris, N., *et al.* (2004). Educational consequences of developmental speech disorder: key stage I national curriculum assessment results in English and Mathematics. *The British Journal of Educational Psychology*, **74**, 173–86.
46. Conti-Ramsden, G. and Botting, N. (2004). Social difficulties and victimisation in children with SLI at 11 years of age. *Journal of Speech, Language and Hearing Research*, **47**, 145–61.
47. Clegg, J. (2006). Speech and language difficulties and later life chances. In *Language and social disadvantage: theory into practice* (eds. Clegg, J. and J. Ginsborg). Wiley, London.
48. Bishop, D.V.M. and Edmundson, A. (1987). Language impaired 4-year olds: distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders*, **52**, 156–73.
49. Clegg, J. and Hartshorne, M. (2004). Speech and language therapy in hyperactivity: a United Kingdom perspective in complex cases. *Seminars in Speech and Language*, **25**, 263–71.
50. Law, J. and Garret, Z. (2004). Speech and language therapy: its potential role in CAMHS. *Child & Adolescent Mental Health*, **9**, 50–5.

9.2.12 Gender identity disorder in children and adolescents

Richard Green

Variance in psychosexual development

Psychosexual development of sex-typed behaviours spans a broad mix of the elements that comprise 'masculinity' and 'femininity'. The possibility for variation is extensive. Among males, there are boys and men whose stereotypical masculinity may pose problems in mental health and criminality. They are not the focus here. Rather, here it is the marked deviation from the mean towards the 'non-masculine' or 'feminine' extreme. That pattern can also cause clinical concern and constitutes gender identity disorder (**GID**) as manifested in childhood. For females, conventional 'tomboyism' is not the focus here, but rather the extreme that can cause clinical concern and constitutes GID.

Epidemiology

No epidemiological studies exist of GID in children. Prevalence can be estimated only roughly from indirect sources. Two items on the Child Behaviour Checklist[1] are consistent with components of the diagnosis. They are 'behaves like opposite sex' and 'wishes to be of opposite sex'. Among 4- to 5-year old boys, not clinically referred for behavioural problems, about 1 per cent of parents answer in the affirmative that their child 'wishes to be the opposite sex'. For ages 6 to 7 it drops to near zero, but rises to 2 per cent at age 11. For girls, the highest rate was 5 per cent at ages 4 to 5, but less than 3 per cent for other ages. With respect to 'behaves like opposite sex', among the boys the rate was 5 per cent and among girls 11 per cent for all ages. However, these data do not indicate any longitudinal aspect of the reported behaviour, and do not detail the behaviour.[2]

An alternative source of estimation looks to the percentage of adults believed to be homosexually oriented. From this population the percentage of homosexual men and women who typically report childhood cross-gender behaviour is used for the estimate. If the rate of exclusive homosexuality is 3 to 4 per cent for men and 1.5 to 2 per cent for women,[3] with perhaps half of homosexual men and women recalling childhood cross-gender behaviour,[4,5] the estimate of childhood cross-gender behaviour is about 3 per cent for boys and under 1 per cent for girls. However, this estimate suffers from problems of retrospective recall and poor comparability between surveys of adults. Further, the recalled behaviour may not have constituted GID.

A disparate sex ratio is evident in referral rates with GID. Four to five boys to one girl are referred. One reason may be greater parental concern over cross-gender behaviour in boys and the greater stigmatizing peer group response to 'sissiness' than to 'tomboyism'. An alternative explanation is that, as with most atypical patterns of sexuality, there is a higher ratio of males to females reflecting a common intrinsic predisposition among males.

Clinical picture

Children with GID differ from other children, including those who merely are not conventionally masculine or feminine as boys or

girls. Their behaviours are typical of other-sex children. Not only do they express a wish to be the other sex, at least in earlier years before they may learn not to verbalize it, but also their dressing preferences, peer group preferences, toy preferences, game preferences, and perhaps their physical mannerisms are those of the other sex.[6]

The picture of GID in children as described in DSM-IVTR,[7] can manifest, in part, by the repeatedly stated desire to be of the other sex: in boys by a preference for dressing in girls' or women's clothing or simulating female attire from available materials, and in girls an insistence on wearing stereotypically masculine clothing with refusal to wear traditional girls' clothing. In role playing, as in make-believe play or imitating media characters, there is a strong preference by the child for other-sex roles. There is also a strong preference for toys generally identified with the other sex, such as Barbie dolls by boys. The peer group is composed primarily or exclusively of other-sex children. Pictures drawn are generally of other-sex figures. There may be cross-sex physical mannerisms. Concurrently, there is an avoidance of traditionally sex-typed activities. Criteria in ICD-10 are similar.[8]

The diagnosis of GID in girls can be more problematic than in boys. This is because 'tomboyism' is a more common part of paediatric psychosexual development than 'sissyness'. There is, however, a distinction between GID in girls and tomboyism. Typical tomboys do not insist that they want to be boys and will wear girls' clothes from time to time, will have both girls and boys as playmates, and will not work to present themselves as young boys.

Substantial cross-gender behaviours are generally manifest in the third or fourth year. Although they are believed by parents and perhaps professional advisors to be a passing phase, at least with those children, seen clinically they endure into school years. Most children are evaluated at about age 7 or 8, when parents become increasingly concerned that the 'passing phase' is not passing and negative reactions by the peer group are enhanced, causing the child social distress.[6]

Aetiology

Understanding the aetiology of GID considers typical influences on early psychosexual development in male and female children.

Early sex differences

Very early behavioural differences are evident between males and females. There may be recognition of the 'like me', 'not like me' dichotomy of one's sex. When boys and girls aged between 10 and 18 months were shown pictures of faces of infants of the same and other-sex, males looked at faces of males longer and females looked at faces of females longer. This 'like me', 'not like me' dichotomy is also interpretable in the study in which two male and two female 1-year-old children were placed at the four corners of a room and permitted, one at a time, to crawl to any other child. Children more often crawled to a child of the same sex.[9,10]

In early play patterns, boys and girls may differ. When 12-month-old children were observed with their fathers in a waiting room, boys were more likely to handle 'forbidden' objects such as trays and vases.[11] The 1-year-old's toy preferences may also differ, with girls preferring soft toys and dolls and boys preferring transportation toys and robots.[12,13]

Preference for mother or father appears to discriminate boys and girls early. When 2- to 3-year-old children were asked which parent in an adjoining room they would prefer to play a game with, or to build with using blocks, or make a sketch with, both boys and girls preferred their father. At 4 years, girls shifted to mother but boys stayed with father.[14]

When children aged between 2 and 3 years were observed in a free play setting, boys were more aggressive toward peers and showed more rough-and-tumble play. When paired in a test play situation with a boy, girls showed more passive behaviour, i.e. standing or sitting quietly and watching their partner play.[15]

The preference for a same-sex peer group emerges early. When 3.5- to 4.5-year-olds were shown pairs of photographs of boys and girls and asked to select the children with whom they would prefer playing, boys preferred boys and girls preferred girls.[16]

Children become aware of sex role stereotypes early. At 2 years of age, boys believe that boys like to play with cars and help their father, 3-year-olds believe that boys like to build things and that only boys like to play with trains. They also believe that girls like to play with dolls, help mother, and cook dinner. Girls are also seen as more likely to say 'I need help'.[17]

The peer group influences psychosexual development. In mixed-gender peer groups, boys more often receive positive responses for masculine activities than girls receive for feminine activities. Boys seem more responsive to peer pressure, in that they will discontinue feminine activities more rapidly than girls will discontinue masculine activities when they are the target of negative responses from either boys or girls.[18]

These findings suggest that if sex-typed attributes emerge in psychosexual development of typical children shortly after the basic dichotomization of 'like me', 'not like me', or male/female, then the first two components of gender identity are consolidated early: (a) the basic sense of male or female; (b) masculine or feminine gender role. The age at which they consolidate coincides with the emergence of significant cross-gender identity and behaviour as seen in the GID of childhood.

Parental influences

Fathers, when observed with 12-month-old children, were more likely to present their sons with trucks rather than dolls, whereas daughters were given both trucks and dolls equally. However, among those children who were given dolls, boys played with them less.[19]

Mothers of young children have been observed with infant actor/actress babies (Baby X experiments). Some of these stranger infants are cross-dressed or given cross-sex names. The perceived sex of the infant influences the mother's behaviour. Children believed to be boys, whether they were or not, were more likely to be encouraged to physical action. Infants believed to be girls were more likely to be given a doll, whereas male infants were more likely to be presented with a football.[20]

None of these findings of early sex differences have been systematically observed with children followed up years later to determine whether early variation from the more common patterns are associated with later variation in psychosexual development.

In our prospective research of several dozen cross-gender behaving boys and conventionally masculine boys,[21] more mother-son shared time was not found in the group of feminine (prehomosexual) versus masculine boys. There was substantial variability in the extent to which mothers and sons were emotionally close. However, with respect to father-son experiences, feminine boys

shared less time with their fathers in their first years when compared with the contrast group of conventionally masculine boys or with their masculine preheterosexual brothers. There was an inverse relationship between the extent of father-son shared time in the first years and later Kinsey score of sexual orientation. Less father-son time was associated with a higher (more homosexual) score.

Identifying the 'chicken and egg' here is problematic. Possibly, boys with a feminine identification who prefer feminine-type activities are less interesting to their fathers. This would lead to father-son distancing. In many of the families, this was the case. However, the finding that early father absence or father-son alienation was associated with cross-gender behaviour was not invariable. Further, there are families in which the mother and father relate comfortably with children and in which GID manifests itself.

Hormonal influences

Evidence for hormonal influences on psychosexual development derives primarily from studies of the intersexed. Girls with congenital virilizing adrenal hyperplasia who produce an excess of androgen beginning prenatally are more rough-and-tumble in childhood play behaviours and less interested in doll play. They are more likely to be considered tomboys.[22] However, less evidence exists for a deficiency in prenatal androgen for boys with cross-gender behaviours.[23–25] Prenatal sex hormone levels are important theoretically, in that to the extent they influence sex-typed behaviour, such as rough-and-tumble or doll play, they may influence peer group composition. They may influence the labelling of the child as 'sissy' or 'tomboy', and may place the child on an atypical developmental track.

Seminal studies of the intersexed in the 1950s indicated that the sex of assignment in the first 2 to 3 years of life was the critical variable in establishing the basic concept of sexual identity as male or female. This was irrespective of gonadal status, hormonal status, internal reproductive structures, and, to some extent, genital configuration.[26] These studies have been criticized on the ground that with the anatomically intersexed the prenatal endocrine status has not been normal.[27] Thus recent interest has focused on individuals believed to have had normal prenatal development but who shortly after birth were nevertheless reassigned to live in the other-sex role, as well as those with a prenatal abnormality.

In one widely publicized case, one male of a pair of monozygotic twins suffered penectomy through circumcision trauma in the first year of life and was reassigned to live as a girl alongside the boy co-twin at about 23 months of age. Although earlier reports indicated that the reassigned twin was adjusting successfully to life as a girl,[28] more recent follow-up revealed that the individual reverted to living as a male in late adolescence, had undergone phalloplasty, and married a female.[29] The other case involves a male infant who also underwent penectomy from circumcision trauma and was assigned to live as a girl, earlier, in the seventh month. That individual was reported to be living as a woman and is bisexual in orientation.[30] One explanation for the discrepancy in the two reports is that the first child was reassigned as a female later than the time during which basic identity of male or female may be set. Both reports, however, suggest a prenatal influence on sexual orientation.

Children born with cloacal exstrophy also provide evidence for prenatal factors influencing gender identity, irrespective of postnatal socialization. Prenatal sex steroid levels are thought to be normal. However, the genital area of these infants, if chromosomally male is so malformed, that there is little prospect of male genital reconstruction. Therefore, many are socialized as girls. Reports from the US reveal a high rate of rejection of living as girls and transition to living as boys. [31] However, an early report from the UK does not indicate female gender role rejection.[32]

The enzyme deficiency of 5-alpha reductase is another clinical example of competing influences of prenatal sex steroids and postnatal socialization. Without this enzyme testosterone is not converted to dihydrotestosterone, needed prenatally to virilize the genitalia. At birth these chromosomal males with intra-abdominal testes appear to be girls based on their external genitalia. Traditionally, they have been raised as girls. Then, at puberty, they do not feminize but rather their clitoris grows substantially to resemble a phallus and there is no gynecomastia. Most then adopt a male role.[33] Debate continues whether this facility to live as heterosexual men is the product of prenatal testosterone or the extensive body virilization and social pressures to live as men. Long-term study of children where the testes are removed before puberty will provide further information.

Longitudinal aspects of atypical early development

Beginning in the late 1960s, the author conducted a prospective study of several dozen boys with extensive cross-gender identification and behaviour.[6] Most of these boys would today be diagnosed with GID, although at the time the diagnosis had not yet entered into the diagnostic nomenclature. These boys were evaluated periodically and assessments continued until late adolescence or young adulthood for two-thirds. At that time, three-quarters of the boys were homosexual or bisexual. One was gender dysphoric. In contrast, a demographically matched group of boys with conventional boyhood behaviours was heterosexual at outcome.[34] More recent follow-up studies at another program reveal a higher minority percentage of cross-gender children remaining gender dysphoric but with the majority homosexually oriented.[35]

These prospective studies are consistent with retrospective reports by adult transsexual males and homosexual males. Many transsexuals recall extensive cross-gender identification and behaviours in childhood. Often, however, these are not documentable because of the length of time from onset to description and the difficulty of corroboration. Several studies have interviewed adult gay men and lesbian women with respect to gender-typed behaviours in childhood. Typically, more extensive cross-gender behaviours are reported than by groups of heterosexual men and women. These retrospective studies of men and women are consistent cross-culturally.[4,5]

Of theoretical and practical import is the overlap in childhood gender behaviours between retrospective reports given by transsexuals and homosexuals, and in the prospective study of cross-gendered boys. Because transsexualism and homosexuality in the adult male are quite different, the question is: Why should there be such an overlap?

One possibility is that the two groups are relatively similar in earlier years, but that different life circumstances promote more comfort for one group continuing to live as males. Treatment intervention to change cross-gender behaviour may be decisive. Transsexuals were rarely treated as children. Different prevalence

rates may also be key. Whereas the incidence of transsexualism may be one in 10 000 males,[36] the incidence of homosexuality may be 3 or 4 per cent.[3] Thus if there are overlapping behaviours between the two in early years, probability would predict that the vast majority of cross-gendered males will emerge as homosexual, rather than transsexual. However, this does not explain the behavioural overlap between prehomosexual boys and pre-transsexual boys who will later be sexually attracted only to females (the latter living as lesbian women after sex reassignment surgery).

Gender identity and mental disorder

GID of childhood was introduced into the DSM in 1980. Its inclusion derived from the prospective study of cross-gender behaving boys described above, with the present author also being a member of the nomenclature committee. The criteria for a set of behaviours being included in the DSM was that a condition be experienced subjectively as distressing and that it constitute a social disadvantage. GID of childhood met the criteria because of the distress the children experienced in consequence of being either male or female and the peer group stigmatization that flowed from their behaviours.

In the past decade, there has been increasing controversy over whether GID of childhood should remain in the list of disorders. In the same period in which GID of childhood was introduced, homosexuality was removed. As our prospective study revealed that a substantial majority of boys with GID matured into homosexual men, to some critics the inclusion of GID was seen as a backdoor through which homosexuality reentered the list of disorders. A response to this concern is that when the subjective distress of being male or female present in children with GID disappears the person no longer has a disorder, whether a heterosexual or homosexual adult. On the other hand, when the distress of being male or female persists, the diagnosis remains GID as seen in adolescence or adulthood, commonly termed transsexualism.

Inclusion of GID for children in the list of disorders is also seen by some critics as perpetuating sex stereotyping in society, and demanding that children conform to traditional masculine/feminine behaviours. A response to this is that the diagnosis is not made merely for gender non-conformity but only when the child is unhappy being male or female, and where the child's behaviours are so atypical that there is substantial adverse reaction from the peer group.

Initial assessment

In the initial assessment of children with suspected GID, the professional should attempt to engage both parents as well as the child. Frequently there is reluctance by one parent, usually the father, to attend. However, this is a family matter, and the clinician needs to gain impressions of the parent-child relationship, the parent-parent relationship, and the child's behaviour from both parents as well as the child.

Assessment is directed towards understanding whether the behaviours described represent a normal variant of psychosexual development. Does the child overtly express dissatisfaction being the sex to which he or she was born? Is there a marked skewing of gender-typed behaviours towards those of other-sexed children or is there some mix? How long has there been cross-gender behaviour? What have parental reactions been to it initially and more recently? What is the child's response when parents attempt to frustrate the cross-gender toy or clothing preferences, if such attempts have been made? What time availability is there for each parent with the child? What do they do together? Are there other persons, for example grandparents or teachers, who may be reinforcing cross-gender behaviours? What are the parental concerns, both in the short and long term, with respect to the significance of the behaviours? Are there other behavioural or medical problems in addition to the gender identity issue?

Treatment

Typically, three principal targets are set for intervention with GID in children. First, the children are unhappy being the sex to which they were born; second, they are experiencing substantial peer group alienation; third, there is conflict with one or both parents in consequence of their atypical behaviour. A principal intervention strategy is helping the child understand that the world of gender is not necessarily black and white, but that greys exist as well. Boys can understand that not all boys need to be good athletes or rough-and-tumblers, and that boys can be sensitive and creative. Girls do not have to be boys to participate in rough-and-tumble play and sports. Children do not need to conform arbitrarily to all sex-typed attributes to remain in their birth sex role. To the extent this can be internalized, the path to transsexualism may be blocked.

The peer group of children with GID can be expanded to include children of both sexes. Parents may have to make efforts, particularly with cross-gendered boys, to find boys of their son's age who will enjoy non-athletic non-rough-and-tumble companionship, perhaps engaging in board games or computer games together. Similarly, girls with GID who are very athletically motivated may find girls who are also athletically inclined, and not just boys to play with. Children who develop comfort in socializing with both boys and girls may experience enhancement of same-sex identification.

Cross-gendered boys are notably alienated from their fathers and intervention can promote their relationship by finding mutually enjoyable activities. This may serve as a source of same-sex identification in the child, will enhance the quality of the parent-child relationship, and will be a positive outcome irrespective of its influence on later sexual identity.

Very few adult transsexuals had entered into any treatment intervention to address GID during childhood. Children with GID referred for evaluation or treatment may, as a product of that concern by parents, and/or professional intervention, have that route to transsexualism diverted. However, there is no empirical support for intervention directed at emerging sexual orientation. There is no evidence that a specific type of 'treatment' in childhood has any effect on outcome on that dimension of gender identity.[34] Parents should understand that if they are concerned about the ultimate sexual orientation of their child, that is a long time ahead. For the immediate period the child is unhappy who he or she is, is experiencing conflict with the peer group, and may be having difficulties at home with at least one parent. These are concerns that should be addressed.

Cross-gender living by children

In recent years, some children with GID have been permitted to live as children of the other sex. Their parents consider that the

strong preference by their child for the dress, activities, and companionship of the other sex with aversion to conventional sex-typed activities, along with the stated preference for being the other sex, argues for the child expressing its gender needs. Complexities of this decision include integrating it into the child's school and neighborhood environment.

Typically, children with GID experience peer group stigma and domestic conflict in consequence of their gender identity. Reduction of conflicts could enhance self-esteem. This social experiment should provide information on whether the longer-term status of the children differs from the children with GID not permitted cross-gender living, most of whom mature into homosexual adults and a minority into transsexual adults.

Early adolescent gender identity disorder

GID continuing into adolescence merges with GID of adulthood. Management issues address the young teenager's continuing gender dysphoria and the consequent social problems. There may be peer group alienation. Depression may develop. School avoidance may develop. Awareness of sexual attraction to same-sex persons may be an additional source of conflict. Parents may be unaware of their teen's GID.

GID in adolescents presents medical, legal, and ethical dilemmas for clinicians. The somatic changes of puberty are very distressing to these young teenagers. And for those who will ultimately progress to adult GID or transsexualism, these changes may pose substantial obstacles to effective 'passing' in their desired gender role. The latter is especially true for males as the voice deepens, facial hair sprouts, and skeletal proportions masculinize. For females, menses are especially troublesome, though not visible, and breast development, especially when prominent, is very distressing.

Clinical recognition of these issues has led to an innovative program for some young teens with GID. This is a trial period of putting puberty on hold and possibly a later cross-sex hormone induced puberty. In the Netherlands, treatment may consist of administering a gonadotrophin releasing hormone agonist (GnRH analogue) at Tanner Stage 2-3 to block secretion of sex steroids that promote pubertal changes. This could be at age 12–14. Depending on the clinical picture of gender identity, at age 16, cross-sex steroids may be administered, or the analogue withdrawn so that endogenous puberty continues.[37] UK practice disfavors analogue treatment prior to Tanner Stage 4 or 5 after substantial pubertal changes. In the early experience of the Dutch treatment program, no patients who have commenced hormonal treatments prior to age 18 have regretted the decision to live as a person of the other sex. Although there is concern that a couple of years of gonadal steroid suppression could predispose to osteoporosis, this concern remains theoretical.

Legally, there is no age barrier to a minor consenting to a medical intervention in the United Kingdom, providing that there is sufficient understanding of the implications of the treatment.[38] At 16 years, adolescents are presumed competent to consent to medical treatment..

The psychiatric management dilemma here is predicting which gender dysphoric adolescents will mature into adult transsexuals, and which will be able to live in the gender role expected from birth, perhaps as homosexual adult men and women.

Further information

Zucker, K.J. (2008) Gender identity disorder. In *Child and Adolescent Psychiatry* (5th edn.) (eds. M. Rutter and E. Taylor). Blackwell, Oxford.

References

1. Achenbach, T. and Edelbrook, C. (1981). Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. *Monographs of the Society for Research in Child Development*, 46.
2. Zucker, K. and Green, R. (1996). Gender identity disorders in children and adolescents. In *Child and adolescent psychiatry* (2nd edn) (ed. M. Lewis), pp. 611–22. Williams and Wilkins, Baltimore, MD.
3. Diamond, M. (1993). Homosexuality and bisexuality in different populations. *Archives of Sexual Behavior*, **22**, 291–310.
4. Whitam, F. and Mathy, R. (1986). *Male homosexuality in four societies: Brazil, Guatemala, the Philippines and the United States*. Praeger, New York.
5. Whitan, F. and Mathy, R. (1991). Childhood cross-gender behavior of homosexual females in Brazil, Peru, the Philippines and the United States. *Archives of Sexual Behavior*, **20**, 151–70.
6. Green, R. (1974). *Sexual identity conflict in children and adults*. Duckworth, London; Basic Books, New York. reprinted by Penguin, Baltimore, MD, 1975.
7. American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th edn, *Text revision*). American Psychiatric Association, Washington, DC.
8. World Health Organization. (1992). *International Classification of Diseases and Related Health Problems*.10. Geneva.
9. Lewis, M. and Weinraub, M. (1974). Sex of parent × sex of child. In *Sex differences in behavior* (ed. R. Friedman, R. Richart, and R. Vande Wiele), pp. 165–89. Wiley, New York.
10. Michalson, L., Brooks, J., and Lewis, M. (1974). Peers, parents, people. Cited in Lewis, M. (1975). Early sex differences. *Archives of Sexual Behavior*, **4**, 329–35.
11. Maccoby, E. and Jacklin, C. (1980). Sex differences aggression. *Child Development*, **51**, 964–80.
12. Fagot, B. (1974). Sex differences in toddlers' behavior and parental reaction. *Developmental Psychology*, **10**, 554–8.
13. Jacklin, C., Maccoby, E., and Dick, A. (1973). Barrier behavior and toy preference. *Child Development*, **44**, 196–200.
14. Lynn, D. and Cross, A. (1974). Parent preference of preschool children. *Journal of Marriage and the Family*, **36**, 555–9.
15. Jacklin, C. and Maccoby, E. (1978). Social behavior at thirty-three months in same-sex and mixed-sex dyads. *Child Development*, **49**, 557–69.
16. Strayer, F. (1977). Peer attachment and affiliative subgroups. In *Ethological perspectives on pre-school social organization* (ed. F. Strayer). Research Memo 5, Department of Psychology, University of Quebec.
17. Kuhn, D., Nash, S., and Brucbern, L. (1978). Sex role concept of two and three-year-olds. *Child Development*, **49**, 445–51.
18. Fagot, B. (1977). Consequences of moderate cross–gender behavior in preschool children. *Child Development*, **49**, 459–65.
19. Snow, M., Jacklin, C., and Maccoby, E. (1983). Sex-of-child differences in father-child interaction at one year of age. *Child Development*, **54**, 227–32.
20. Sidorowicz, L. and Lunney, G. (1980). Baby X revisited. *Sex Roles*, **6**, 67–73.
21. Green, R. (1987). *The 'sissy boy syndrome' and the development of homosexuality*. Yale University Press, New Haven, CT.
22. Ehrhardt, A. and Baker, S. (1974). Fetal androgens, human central nervous system differentiation and behavior sex differences.

In *Sex differences in behavior* (eds. R. Friedman, R. Richart, and R. van de Wiele), pp. 33–51. Wiley, New York.
23. Yalom, I., Green, R., and Fisk, N. (1973). Prenatal exposure to female hormones: effect on psychosexual development in boys. *Archives of General Psychiatry*, **28**, 554–61.
24. Kester, P., Green, R., Finch, S., *et al.* (1980). Prenatal female hormone administration and psychosexual development in human males. *Psychoneuroendocrinology*, **5**, 269–85.
25. Hines, M. and Kaufman, F. (1994). Androgen and the development of human sex–typical behavior. *Child Development*, **65**, 1042–53.
26. Money, J., Hampson, J., and Hampson, J. (1955). An examination of some basic sexual concepts: the evidence of human hermaphroditism. *Bulletin of the Johns Hopkins Hospital*, **97**, 301–19.
27 Diamond, M. (1965). A critical evaluation of the ontogeny of human sexual behavior. *Quarterly Review of Biology*, **40**, 147–75.
28. Money, J. and Ehrhandt, A. (1972). *Man and woman, boy and girl*. Johns Hopkins Press, Baltimore, MD.
29. Diamond, M. and Sigmundson, H. (1997). Sex reassignment at birth. Long–term review and clinical implications. *Archives of Pediatric and Adolescent Medicine*, **15**, 298–304.
30. Bradley, S., Oliver, G., Chesnick, A., *et al.* (1998). Experiment of nature: ablatio penis at 2 months, sex reassignment at 7 months and a psychosexual follow–up in young adulthood. *Pediatrics*, **102**, 9–21.
31. Reiner, W. and Gearhart, J. (2004). Discordant gender identity in some genetic males with cloacal exstrophy assigned to female sex at birth. *New England Journal of Medicine,* **350**, 333–41.
32. Carmichael, P., Schober, J., Hines, M. *et al.* (2003). Abstract, International Academy of Sex Research, Bloomington, IN.
33. Imperato-McGinley, J., Guerrero, L., Gautier, T. *et al.* (1974). Steroid 5 alpha reductase deficiency in man, an inherited form of male psuedohermaphrodism. *Science,* **186**, 1213–1215.
34. Green, R. (1987). The 'Sissy Boy Syndrome' and the Development of Homosexuality. Yale University Press, New Haven.
35. Zucker, K. (2005). Gender identity disorder in children and adolescents. Annual Review of Clinical Psychology, **1**, 467–492.
36. Kestersen, P., Gooren, L., and Magers, J. (1996). An epidemiological and demographic study of transsexuals in The Netherlands. *Archives of Sexual Behavior*, **25**, 589–600.
37. Delemarre-van de Waal, H., Cohen-Kettenis, P. (2006). Clinical management ofgender identity disorder in adolescents. European Journal of Endocrinology. Online via www.eje.online.org.
38. Gillick v West Norfolk and Wisbeck Area Health Authority (1986). AC 112.

9.3

Situations affecting child mental health

Contents

9.3.1 The influence of family, school, and the environment

Barbara Maughan

Introduction

Like adult disorders, most child psychiatric problems are now regarded as multifactorially determined: both genetic and environmental factors play a role in their development. This chapter provides an overview of some of the key environmental elements in that equation. Subsequent chapters discuss risks for specific disorders; the focus here is on the more general issues that arise when considering the effect of environmental influences on the onset or persistence of psychopathology in childhood.

Environments and development

As in all aspects of child psychiatry, a developmental perspective is crucial when considering environmental risks. Some developmental periods may be especially sensitive for neurodevelopment, and show heightened effects of environmental insults. In addition, key sources of environmental influence change with age, and the meaning and impact of events will vary with the child's stage of cognitive, emotional, and social development. The family is the central source of early environmental influences, charged as it is in most societies with prime responsibility for the care, nurture, and socialization of the young. As children develop, so their social worlds expand; childcare and school settings take on increased importance, as do relationships with friends and peers. Throughout, each of these proximal contexts is shaped by influences from the wider culture and society. Any comprehensive assessment of a child's environment needs to take each of these types and levels of influence into account.

Nature-nurture interplay

At one time, causal associations between adverse experiences and childhood disorder were assumed to run in just one direction. Today, it is clear that the situation is vastly more complex. Children are not simply passive recipients of experience; they influence, as well as being influenced by, those around them, and they play an active role in constructing and interpreting their social worlds.[1] Even very young infants influence the nature of their interactions with caregivers, and children's capacities for shaping and selecting their experiences increase as they mature. The temperamentally difficult child is likely to evoke more negative responses from parents; when parents themselves are under stress, or find it hard to maintain consistency, troublesome child behaviours can play a key role in fuelling harsh or punitive responses. Delinquent adolescents may seek out delinquent peers, who further encourage their antisocial activities. Associations between environmental factors and disorder often involve complex reciprocal patterns of effects.

Some of the evocative effects of children's behaviour will reflect heritable traits.[2] The advent of behaviour–genetic studies in child psychiatry has provided important insights into environmental as

well as genetic risks. Genetic analyses have shown, for example, that many ostensibly 'environmental' factors include some element of genetic mediation.[3] Parents provide children not only with their environments but also with their genes, so that in biologically related families, nature and nurture are inevitably interwoven. Musical parents will encourage their children to enjoy music, buy them a violin, and may also pass on musical talents. In a similar way, antisocial parents may rear children in hostile and punitive environments, provide models of antisocial behaviour, and also pass on genes that predispose to disruptive behaviours. In all likelihood, genes and environments will often be *correlated* in this way.

Genetically informative studies have also highlighted other key mechanisms in ***gene-environment interplay***.[4] First, environments may *moderate* genetic influences, such that the heritability of some traits may vary systematically with qualities of the environment. Second, genetic factors may contribute to *differential sensitivity to environmental risks*. Research has consistently shown marked individual differences in children's responses to all but the most severe forms of psychosocial adversity. As yet, reasons for these differences are not well understood. They may reflect variations in the severity of exposure; individual differences in resilience or coping strategies, or in environmental sources of protection; or variations in vulnerability. Genetic predispositions clearly constitute one source of such vulnerability, and several examples of gene x environment interactions have now been documented. Finally, pre-clinical studies provide clear evidence that environments can influence *gene expression* through *epigenesis;* as yet, the extent to which processes of this kind apply in humans is unknown.

Risk variables and risk mechanisms

Identifying environmental factors that show links with children's adjustment is only the first step in understanding *how* they function to increase risk for disorder. A variety of different mechanisms has been proposed here. Some may run through the effects of stress on the biological substrate. Exposure to aggression and hostility may influence children's cognitive processing, leading to the development of negative cognitive sets and attributional biases. In a related way, disrupted early attachments are argued to affect the psychological structures needed for later relationship formation. Adverse experiences may lead to direct increases in negative emotionality, disruptive behaviours, and impulsiveness, or to negative interactional styles that impact on social relationships. And finally, stress may affect children's self-concepts, or compromise their coping skills in ways that increase the risks for disorder. Any given environmental risk may be associated with a number of risk mechanisms, and the processes involved in the persistence of disorder may differ from those involved in its onset.

Family influences

Pre- and early post-natal development

Some vulnerability to psychopathology is laid down in foetal development. The potential for adverse effects of maternal substance use on the developing foetus have been known for many years; much recent attention has focussed on associations between prenatal cigarette smoking and risk for externalizing disorders in offspring. In addition, current estimates suggest that as much as 15 per cent of the load of childhood emotional/behavioural problems may be attributable to exposure to maternal anxiety and stress in pregnancy. Though the mechanisms involved here remain to be elucidated, there is speculation that these effects may reflect foetal programming of stress response systems akin to those posited in studies of early life influences on risk for cardiovascular disease.

Post-natally, as children progress from the complete dependence of infancy to increasing independence, they need stable and secure family relationships to provide emotional warmth, responsiveness, and constructive discipline. The influential work of Bowlby[5] and others has shown that a child's need to be attached to others is a basic part of our biological heritage. Infants become increasingly socially responsive over the first 6 months of life. At 6 to 8 months of age they begin to form selective attachments to particular individuals; they seek proximity to these attachment figures if distressed or frightened, and protest if the person they are attached to leaves. In evolutionary terms, these behaviours function to provide protection for the infant, and to reduce anxiety and distress.

Almost all infants—even those neglected or maltreated by their carers—develop attachment relationships of this kind. Their quality varies, however, depending on characteristics of the parent, the child, and the mesh between the two. Infants who have received sensitive and responsive care tend to show ***secure*** attachment patterns; ***insecure*** attachments are more likely to develop when parents themselves are stressed or unsupported, and are unresponsive to their children. Two main types of insecure attachment have been identified: avoidant attachments (associated with rejecting or highly intrusive parental care) and resistant–ambivalent patterns (associated with inconsistent or unresponsive parenting). More recently, a third disorganized category has been described, in which infants show a variety of contradictory behaviours after brief separations, and often appear confused, depressed, or apprehensive. This seems especially associated with parental behaviours that are frightening, unpredictable, or abusive.

Attachment theorists argue that the quality of these early relationships may have long-term implications. Though not entirely resistant to change, infants' attachment patterns do tend to be stable over time. Some of this stability may reflect continuity in the quality of family care. In addition, attachment theory proposes that early attachment experiences are internalized in internal working models of self and others, which function as templates for future relationship formation. Children who have experienced responsive early care come to expect others to be caring and reliable; those who have been ignored or rejected develop less positive expectancies of others, of relationships, and of themselves. Later in development, new relationships may be created in line with these expectancies.

Although many aspects of these models await confirmation, securely attached infants are known to go on to be more sociable and co-operative in their social relationships, and to show more positive affect and self-esteem. Insecurely attached infants show less positive relationships, and are at some increased risk for psychopathology. Taken alone, attachment security in infancy is only a weak predictor of global functioning in early adulthood, suggesting that early attachment experiences work with and through other experiences—including peer relationships, later family experiences, and eventually mature intimate relationships—to contribute to later functioning. In addition, both ICD-10 and DSM-IV recognize two varieties of attachment disorders: non-attachment with emotional withdrawal, typically associated with abuse, and non-attachment with indiscriminate sociability, most usually observed when

children have been exposed to repeated changes of caretaker or institutional care. Although as many as 40 per cent of infants receive insecure attachment classifications, these more severe forms of attachment disorder are rare.

Family relationships and parenting

Many other aspects of family life and relationships, and of parenting styles and behaviours, have been examined for their impact on children's development. Research on families emphasizes the complexity of family relationships; each dyadic relationship is influenced by other relationships in the family, and normative transitions in family life—the birth of a sibling, or mother starting work—reverberate to affect all family members.[6] Relationships with parents and siblings change as children develop, and both these, and specific aspects of parenting, may impact on risks for disorder.

The implications of the most severely compromised parenting, involving abuse or neglect, are examined in Chapter 9.3.3, and family-based risks for individual childhood disorders are discussed in detail in the chapters dealing with each specific condition. In general, these reflect four broad themes:

- discordant, dysfunctional relationships between parents, or in the family system as a whole;
- hostile or rejecting parent-child relationships, or those markedly lacking in warmth;
- harsh or inconsistent discipline;
- ineffective monitoring and supervision.

Within this broad pattern, differential treatment of siblings is known to increase conflict between children, and may have important implications for psychopathology. In addition, outcomes are markedly poorer when children face multiple family-related risks.

Family life can also provide important sources of protective influences for children facing life events and other stressors. Cohesion and warmth within the family, the presence of one good relationship with a parent, close sibling relationships, and the nature of parental monitoring and supervision have all been found to show protective influences of this kind.

Parent and family characteristics

Psychopathology in parents is associated with increased risks of emotional and behavioural problems in children. Recent estimates suggest that as many as 60 per cent of the children of parents with major depression will develop psychiatric problems in childhood or adolescence, and their risks of affective disorder are increased fourfold. Psychosis, alcohol and drug abuse, and personality disorders in parents are also associated with increased risks of disorder in offspring, and parental criminality is a strong risk factor for conduct problems and delinquency.

In most instances, these links will reflect a complex interplay between genetic and environmental effects. Disorder in parents is frequently associated with disturbed marital relationships, and parental psychopathology may also impair parenting capacities. Depressed mothers, for example, are less sensitive and responsive to their infants, and attend less, and respond more negatively, to older children. Alcohol and drug abuse and major mental disorders in parents may impair parenting in more wide-ranging ways. When parents are antisocial, effects may also be mediated through the endorsement of antisocial attitudes and social learning.

Young maternal age is associated with increased risk for child and adolescent conduct problems. In part, these associations are likely to reflect the educational and social disadvantages that predict very early parenthood; in part, the poor social conditions and lack of support faced by many young mothers; and in part, less than optimal parenting styles. Delinquency is also associated with large family size. Once again, the more proximal risks involved are likely to be complex: parental supervision may be less effective in large families, and opportunities to 'learn' from delinquent siblings higher. Beyond this, family size shows few consistent links with childhood disorder. Only children are not at increased psychiatric risk, and they share with other first-borns some small advantages in terms of cognitive development. Birth order also appears to have few implications for behavioural adjustment, although youngest children show some increased rates of school refusal.

Changing family patterns

Recent decades have seen massive changes in the pattern of many children's family lives. The most obvious markers are the dramatic increases in rates of divorce, single parenthood, and step-family formation, along with major increases in maternal employment. In the years immediately after the Second World War, just 6 per cent of British couples divorced within 20 years of marriage. By the mid-1960s that figure had increased four-fold, and divorce rates continued to rise into the 1980s. For most children, parental divorce will be followed by a period in a single-parent household; for a substantial minority, further family transitions will mean that they become part of a step family. In the early years of the 21st century more than 10 per cent of UK families with dependent children were step-families, and approaching a quarter of children lived in single parent households.

Parental divorce

There is now extensive evidence that divorce is associated with negative consequences for children.[7] Psychological and behavioural distress are common, especially in the period immediately following divorce; more severe disturbance is not. Boys in particular are at increased risk for conduct problems. Educational attainments and motivation are often compromised, and subsequent relationships may also be affected. As they approach adulthood, children of divorce move into close relationships earlier than their peers, but also experience higher risks of relationship breakdowns.

Events both before and after the separation seem central in understanding these effects. Longitudinal studies, for example, have shown that children in divorcing families often show disturbed behaviour well before their parents separate. Exposure to the discord and conflict that frequently precede divorce thus seem to be key components of risk. After separation, problematic relationships between parents may continue, and the parents' own distress may compromise their capacity to respond sensitively and consistently to their children's needs. Many families face a sharp decline in economic circumstances after divorce, and for many children their parents' separation may involve house moves, school changes, and other disruptions to their established social networks. Each of this constellation of factors may contribute to subsequent outcomes.

Single parents and step families

Research on the effects of growing up in single-parent and step families illustrates the complexity of family-related influences.[8]

Overall, children in single-parent and step families show higher mean levels of emotional and behavioural problems than those in non-divorced two-parent families; they also have an increased probability of health problems and educational underachievement. But there are also marked differences within each family type and associations between the quality of mother–child relationships and children's adjustment is similar across family settings. In addition, single-parent and reconstituted families often differ from stable two-parent families in a plethora of other ways; in particular, they are much more likely to face economic pressures, poor social support, and higher levels of maternal depression. Once these variations and the degree of negativity in family relationships are taken into account, family type *per se* shows few consistent links with children's adjustment.

Peer influences

Beyond the family, relationships with peers are now recognized to provide a unique and essential contribution to children's social, emotional, and cognitive development.[9] By the end of the pre-school period most children have at least one reciprocated friendship. In childhood and adolescence, peers take on increasing importance; in middle childhood, more than 30 per cent of children's social interactions are with peers, and adolescents are estimated to spend more than twice as much time with peers than they do with parents or other adults. The functions of friendship change with development, expanding to encompass companionship and stimulation, help and sharing, social and emotional support and intimacy.

With friends and peers children acquire skills, attitudes, and experiences that contribute to many aspects of their adaptation. By the same token, children who have poor social skills, or who are rejected or neglected by peers, are at risk of a range of adverse outcomes including poor school performance, school drop-out, and psychiatric disorder. Social rejection may increase children's feelings of loneliness, reduce supports that can buffer against stressors, and also mean that isolated children miss out on important social learning experiences. Since many children with psychiatric disorders also show difficulties in relationships with peers, processes of this kind may well compound their problems. In adolescence, affiliations with behaviourally deviant peers have attracted particular interest as correlates of conduct disorder and delinquency. Here, reciprocal influences have been demonstrated: aggressive disruptive children are more likely to associate with deviant peers, but relationships with peers also show an independent effect on both the onset and persistence of delinquency.

Child care and schooling

By the late 1990s approaching 50 per cent of mothers in the UK returned to full- or part-time work before their infants reached one year of age. This major increase in early maternal employment has prompted extensive research on the impact of alternative childcare on children's development. Recent evidence[10] suggests that multiple features of early care need to be considered in assessing effects. Higher child-care ***quality*** (as indexed by features such as sensitive and responsive care-giving, and cognitive and language stimulation), is associated with improved performance on tests of cognitive, language, and early academic skills, and with more prosocial skills and fewer behaviour problems, By contrast, higher ***quantity*** of child care (as indexed by hours per week in any kind of non-maternal care), is associated with some increased risks of problem behaviours in both the preschool and early school years.

School life then brings its own demands and challenges. Starting and changing schools are significant, sometimes troublesome, events for children; although most young children adapt well, a significant minority show some disturbance when they start school, and both attainment levels and self-perceptions are affected for many young adolescents after the transition from primary to secondary school. Tests and examinations rank high on children's lists of fears, and levels of psychological distress are elevated at times of major examinations. Although fears of this kind are not generally severe, they do show links with clinically significant symptoms. Bullying is a further problem especially associated with the school context. Self-report surveys suggest that over 15 per cent of young children experience some bullying at school, mostly unknown to parents or teachers. Although rates fall with age, up to 5 per cent of adolescents continue to face bullying in secondary school. Persistently victimized children have identifiable characteristics, with histories of anxious insecure behaviours and social isolation often beginning before they started school; bullying then increases their risks of adjustment problems.

Like families, schools differ in their atmosphere and social climate, and these variations show an independent impact on children's academic progress and behaviour. In part, these variations reflect differences in initial pupil intakes. In addition, they show systematic links with organizational characteristics of schools. Schools with more positive outcomes are characterized by purposeful leadership, constructive classroom management techniques, an appropriate academic emphasis, and consistent but not oversevere sanctions. In relation to behavioural outcomes, the composition of pupil groupings may also be influential. Young children are more likely to become aggressive if placed in highly aggressive classes, and risks of delinquency are increased in secondary schools where intakes include large proportions of less able children. For some severely disadvantaged groups, however, schooling may offer an important source of positive experiences. Experimental studies of preschool programmes, for example, have shown important long-term gains in terms of reduced risks of delinquency and unemployment many years after participants left school.

Wider social and environmental influences

Poverty and social disadvantage

Poverty and social disadvantage are most strongly associated with deficits in children's cognitive skills and educational achievements.[11] In the behavioural domain, disruptive behaviours also show links with family poverty. Effects appear to be more marked for boys than for girls, and seem to be stronger in childhood than in adolescence. Intermittent hardship is associated with some increased risk for conduct problems, but the impact is most marked for children in families facing persistent economic stress. Most current evidence suggests that these effects are indirect. Poverty imposes stress on parents, and reduces the supports available to them; these in turn increase the risks of harsh or coercive parenting, and reduce parents' emotional availability to their children's needs. Some studies suggest that relative deprivation—the perception that one is disadvantaged by comparison with others—may be more important than income levels *per se*.

Neighbourhood and community contexts

Rates of childhood disorder vary in different neighbourhoods and communities. Urbanization is frequently associated with increased risks of disorder, and rates may be especially high in chronically disadvantaged inner-city neighbourhoods. In early childhood, many of these effects seem to be indirect; neighbourhood disadvantage increases stress on families, and these in turn largely account for associations with children's difficulties. In severely disadvantaged settings, however, even quite young children may be directly exposed to community violence, and in adolescence, neighbourhood influences may be mediated through associations with delinquent peers.

Multiple stressors

For many children, exposure to these differing types of adversity will covary. Stressed families frequently live in poor neighbourhoods, where schools are under pressure and peer groups deviant. Early epidemiological findings suggested that isolated single risks have relatively little impact on disorder, but that rates rise sharply when risk factors combine. More recently, studies have shown that child, sociocultural, parenting, and peer-related risks each add uniquely to the prediction of behaviour problems. In addition, the total number of risks a child faces explains further variance in outcomes.

Secular trends in disorder and psychosocial risks

Finally, it is important to consider how psychosocial risks may impact on overall levels of disorder. There is now clear evidence that rates of many adolescent disorders—including depression, suicide, alcohol and drug use, and delinquency—have risen since the Second World War.[12] Since it is implausible that changes in the gene pool could occur so rapidly, environmental risk factors must be implicated. Some of these may overlap with risks for individual differences in disorder, but others may be quite distinct. Based on an extensive review of available evidence, Rutter and Smith[12] concluded that a variety of factors are likely to be implicated:

1 increased rates of family breakdown, with their associated effects on the disruption of relationships and exposure to conflict and discord;

2 a change in the meaning of adolescence, with prolonged education and economic dependence on parents occurring alongside increased autonomy in other spheres;

3 a possibly increased disparity between young people's aspirations and the opportunities available to meet them;

4 increased alcohol consumption and illegal drug use;

5 changing social attitudes to acceptable behaviour, possibly enhanced by influences from the mass media.

Other specific factors may affect rates of juvenile crime. In particular, the increasing commercialization of youth culture, providing more goods to steal, may have coincided with diminished surveillance and increased situational opportunities for property crime.

Further information

Clarke-Stewart, A. and Dunn, J. (2006). *Families count: effects on child and adolescent development*. Cambridge University Press, Cambridge.

Rutter M. (2005). Environmentally mediated risks for psychopathology: Research strategies and findings. *Journal of the Amrican Academy of Child and Adolescent Psychiatry,* **44**, 3–18.

References

1. Lytton, H. (1990). Child and parent effects in boys' conduct disorder: a reinterpretation. *Developmental Psychology*, **5**, 683–97.
2. Scarr, S. and McCartney, K. (1983). How people make their own environments: a theory of genotype–environment effects. *Child Development*, **54**, 424–35.
3. Plomin, R. and Bergeman, C.S. (1991). The nature of nurture. Genetic influences on 'environmental' measures. *Behavioural and Brain Sciences*, **14**, 373–86.
4. Rutter, M., Moffitt, T., and Caspi, A. (2006). Gene–environment interplay and psychopathology: Multiple varieties but real effects. *Journal of Child Psychology and Psychiatry*, **47**, 226–61.
5. Bowlby, J. (1988). *A secure base: clinical applications of attachment theory*. Routledge, London.
6. Dunn, J. (1994). Family influences. In *Development through life: a handbook for clinicians* (eds. M. Rutter and D.F. Hay), pp. 112–31. Blackwell Science, Oxford.
7. Rodgers, B. and Pryor, J. (1998). *Divorce and separation: the outcomes for children*. Joseph Rowntree Foundation, York.
8. Dunn, J., Deater-Deckard, K., Pickering, K., *et al.* (1998). Children's adjustment and prosocial behaviour in step–, single-parent, and non-stepfamily settings: findings from a community study. *Journal of Child Psychology and Psychiatry*, **39**, 1083–95.
9. Rubin, K.H., Bukowski, W., and Parker, J.G. (1998). Peer interactions, relationships, and groups. In *Handbook of child psychology*, Vol. 3 *Social, emotional and personality development* (ed. N. Eisenberg), pp. 619–700. Wiley, New York.
10. NICHD Early Child Care Research Network (2006). Child-care effect sizes for the NICHD Study of Early Child Care and Youth Development. *American Psychologist,* **61**, 99–116.
11. Duncan, G.J. and Brooks-Gunn, J. (ed.) (1997). *Consequences of growing up poor*. Russell Sage, New York.
12. Rutter, M. and Smith, D.J. (1995). *Psychosocial disorders in young people: times trends and their causes*. Wiley, Chichester.

9.3.2 Child trauma

David Trickey and Dora Black

Children's reactions to traumatic events

This chapter will focus on the impact on children of traumatic events other than child abuse or neglect, which are covered in Chapter 9.3.3. According to the DSM-IV-TR definition of post-traumatic stress disorder (PTSD), traumatic events involve exposure to actual or threatened death or injury, or a threat to physical integrity. The child's response generally involves an intense reaction of fear, horror, or helplessness which may be exhibited through disorganized or agitated behaviour. Terr suggested separating traumatic events into type I traumas which are single sudden events and type II traumas which are long-standing or repeated events.[1]

If the traumatic event includes bereavement, the reactions may be complicated and readers should consult Chapter 9.3.7 to address the bereavement aspects of the event.

Following a traumatic event, children may react in a variety of ways (see Chapters 4.6.1 and 4.6.2 for the adult perspective on reactions to stressful and traumatic events). Many show some of the symptoms of post-traumatic stress disorder—re-experiencing the

event (e.g. through nightmares, flashbacks, intrusive thoughts, re-enactment, or repetitive play of the event), avoidance and numbing (e.g. avoidance of conversations, thoughts, people, places, and activities associated with the traumatic event, inability to remember a part of the event, withdrawal from previously enjoyed activities, feeling different from others, restriction of emotions, sense of foreshortened future), and physiological arousal (e.g. sleep disturbance, irritability, concentration problems, being excessively alert to further danger, and being more jumpy). In young children the nightmares may become general nightmares rather than trauma-specific. Other reactions to trauma in children are:

- becoming tearful and upset or depressed
- becoming clingy to carers or having separation anxiety
- becoming quiet and withdrawn
- becoming aggressive
- feeling guilty
- acquiring low self-esteem
- deliberately self-harming
- acquiring eating problems
- feeling as if they knew it was going to happen
- developing sleep disturbances such as night-terrors or sleepwalking
- dissociating or appearing 'spaced out'
- losing previously acquired developmental abilities or regression
- developing physical symptoms such as stomach aches and headaches
- acquiring difficulties remembering new information
- developing attachment problems
- acquiring new fears
- developing problems with alcohol or drugs.

Such problems may individually or in combination cause substantial difficulties at school and at home. The reactions of some children will diminish over time; however, for some they will persist, causing distress or impairment, warranting diagnosis, and/or intervention. Research predicting which children will be more likely to be distressed following a traumatic event suffers from a number of methodological flaws. However, factors which are often identified as constituting a risk for developing PTSD across a number of studies include: level of exposure, perceived level of threat and peri-traumatic fear, previous psychological problems, family difficulties, co-morbid diagnoses, subsequent life events, and lack of social support.

Diagnosis of PTSD

Both DSM-IV-TR and the ICD-10 diagnostic classification of PTSD are appropriate for use with adults and although children from 8 years old do display similar symptoms to adults[2] there are some developmental differences particularly in younger children.[3] Alternative diagnostic criteria for pre-school children have therefore been developed, which draw on reports by carers and include more behavioural symptoms such as loss of developmental skills, and development of new fears or anxiety.[4]

Other diagnoses

Careful assessment is required to make an accurate differential diagnosis. According to DSM-IV-TR, PTSD can only be diagnosed 1 month after the event, prior to that a diagnosis of acute stress disorder (ASD) may be appropriate. Whereas ICD-10 PTSD can be diagnosed within the first month, and the acute stress reaction is reserved to describe a disturbance that resolves rapidly. If the event is not of sufficient severity to meet the criteria for PTSD and the reaction does not last more than 6 months after the stressor has ceased, then a diagnosis of adjustment disorder may be appropriate. Recovery may take longer for children if their parents continue to suffer from symptoms of PTSD which may constitute a chronic source of stress which may in turn prolong the symptoms of the child. Further information on these diagnoses from an adult perspective can be found in Chapters 4.6.1–4.6.5.

Other diagnosable disorders may result from traumatic events, and may be present singularly or co-morbidly with PTSD; 60 per cent of children with PTSD have a co-morbid mental health diagnosis.[5] According to Fletcher's meta-analysis, common co-morbid diagnoses are: anxiety disorders, depression, alcohol and drug abuse in adolescents, and attention deficit hyperactivity disorder (ADHD).[2]

Assessment

As with other psychiatric disorders, the best assessment can be made by integrating information from a number of sources such as an interview of the parent/carer alone, an interview of the family together, information from school, and information from psychological measures (see below). Careful consideration should be given to which members of the family will be involved in any interviews so as to avoid exposing previously unaffected children to the traumatic details of the event. Sometimes children try to 'protect' their carers from distress by under-reporting their symptoms of trauma, it is therefore also essential to interview the child on their own where possible.[6]

In order to assess what elements of the child's current functioning and distress may be a result of the traumatic event, and those that may pre-date it, it is important to gain as full a picture as possible of their developmental history and their pre-morbid functioning. Reports from teachers and other professionals may be particularly useful in this respect.

On assessment, some account of the traumatic event is necessary so that the clinician can gain an understanding of what exactly was experienced. Furthermore, it is helpful to give the sense that the clinician can bear to hear a story which the child and family may have been avoiding to tell for some time. However, this must be balanced against the child's understandable avoidance of the memory. There is little point gaining a full account of the event during the assessment, if the child becomes so distressed that they do not return for treatment. Pynoos and Eth offer a structure for conducting such an initial assessment which begins with a projective drawing and storytelling. It then proceeds to discussion of the actual event and its impact, followed by closure.[7] If the assessment has included talking about the traumatic event the child and family may become very distressed, and it may be necessary to invest some time in winding down the session, so that the family does not leave overly distressed. This will increase the likelihood of them engaging in the treatment process, which is likely to involve thinking through the event—something which they often do not intuitively want to do.

Assessment includes asking about the motivation for the contact with the service. Unlike many other problems, families with traumatized children may not simply present at the service in order to effect a change in their child's symptoms. They may be asking for advice following a traumatic event to try and prevent problems from appearing later by ensuring that they are 'doing the right thing', or they may be seeking an assessment for the purpose of compensation or other legal purposes.

Psychological instruments

A number of questionnaires and diagnostic interviews are available to assist in the assessment of PTSD. Such instruments cannot replace a clinical interview, but may assist by strengthening clinical opinion and giving a reliable and valid indication of perceived symptom severity or frequency. This can give an indication of the impact of an event, can raise hypotheses which can be further investigated (e.g. screening for possible presence of PTSD which can then be further assessed) and if repeated after treatment can help to measure change. Ohan and colleagues reviewed those measures used in research studies that have adequate psychometric properties.[8] The best diagnostic instrument is the children's PTSD inventory. This is a structured diagnostic interview, the psychometric properties of which are good and have been published in peer-reviewed journals.[9,10] The Children's Revised Impact of Events scale[11] is a useful self-report questionnaire of 13 items appropriate for children aged 8 years and above. It has good screening properties[12] and is freely available from the Children and War Foundation's website (see below). Similarly Foa's Child PTSD Symptom scale (CPSS) is a 24-item self-report questionnaire designed for children aged 8 years and above. It is emerging as having good psychometric properties.[13] The CPSS has the advantage of specifically assessing the symptoms of PTSD from the DSM-IV-TR, and unlike most other self-report measures, it also addresses the important question of the impact of the symptoms on the child's functioning.

Evaluation of treatments

An increasing amount of evidence indicates that trauma-focussed cognitive behavioural therapy (TF-CBT) is the treatment of choice for PTSD.[14] Much of the evidence is from studies of children with PTSD as a result of child sexual abuse. Dalgleish and colleagues summarize the evidence thus; 'The current take-home message from this nascent literature therefore is that CBT appears to have well-established efficacy in treating a range of post-traumatic stress responses following sexual abuse, with preliminary evidence in favour of this form of intervention following other types of trauma'.[15] Consequently, TF-CBT is the only intervention recommended by the United Kingdom's National Institute for Health and Clinical Excellence (NICE) for the treatment of children and young people with PTSD.[6]

Family therapy and psychodynamic psychotherapy as stand alone therapies for children with PTSD are not currently supported by the same level of evidence as TF-CBT. However, there is evidence to illustrate the importance of involving parents or carers in treatment where appropriate.[16] Psychodynamic psychotherapy may have some contribution to the treatment of traumatized children[17] and clinical experience indicates this to be true particularly in children with co-morbid diagnoses or where the traumas are multiple and prolonged (type II) such as in war and civil conflicts, and have led to failures of basic care.

Eye Movement Desensitisation and Reprocessing (EMDR) has proven to be effective with adults and has increasingly been used with children.[18] It shows promise for use with children and young people, but it has not yet established the same level of evidence as TF-CBT.[6]

The available evidence does not currently support the use of medication to treat PTSD in children or young people,[6] although medication may be necessary to treat any co-morbid conditions such as depression, ADHD, or sleep problems.

Management

Trauma-focussed cognitive behavioural therapy (TF-CBT)

'Processing' of an event involves bringing the event to mind and thinking it through often by talking, writing, or playing. This enables the memory to be stored as an ordinary narrative memory which is under the child's control, rather than as the vivid sensory information of the original event which is prone to being involuntarily re-experienced. Furthermore, processing enables a more helpful meaning to be attributed to the event; so whilst it may well be thought of as 'terrifying' and 'unfortunate', it no longer colours the way that *everything* in the child's world, including him or herself is seen.

Many children will process difficult and traumatic experiences naturally, often with the help of those around them such as their carers. Some may take a little while to do so. However, bringing up the memory may initially trigger the fear or distress of the original event, so some children avoid thinking about it, which in turn prevents processing.[19] Similarly the adults around the child may try to protect either themselves or the child from further distress by not discussing the event and so opportunities to enable the child to process it may be missed through avoidance by proxy. Therefore some children will not manage to process the event without professional help.

TF-CBT for traumatized children and families involves enabling the child or young person to bring the traumatic event to mind within the safe environment of therapy. This is akin to *exposure* to the memory, which enables the memory to be processed. *Cognitive restructuring* involves enabling the child to alter their unhelpful negative view of themselves or the world, which is based on the event, to a more helpful and realistic one. Cognitive behavioural therapy for children is further described in Chapter 9.5.3, Stallard provides excellent resources in the form of a workbook,[20] and Cohen offers a comprehensive description of the treatment of trauma and traumatic grief.[21]

Younger children with PTSD may not be able to make use of TF-CBT in the same way that older children can, and their treatment is likely to involve much more family work and ensuring that they learn that their environments are safe through re-assurance and stability.[22]

Early interventions

There continues to be much debate about the value of early intervention (i.e. within the first month). Research to date does not support the use of single session debriefing,[23] however, there may be value in the provision of practical and emotional support in the immediate aftermath, together with some education of what reactions can be expected. Development of a culture where the child is permitted, or encouraged to talk about the event (e.g. within the family or within the school) may protect against the development of PTSD.[24] Early trauma-focussed cognitive behavioural therapy may be offered to older children with the most severe symptoms[6]

which may be completed in groups (e.g.[25]). There may be a role for EMDR in the early stages following a traumatic event, although this has yet to receive empirical support.

Further information

Dyregrov, A. (2006). *Grief in Children: a handbook for adults*, (2nd edn) London, Jessica Kingsley Publishers.

Salmon, K. and Bryant, B.A. (2002). Posttraumatic stress disorder in children. The influence of developmental factors. *Clinical Psychology Review*, **22**, 163–88.

The National Child Traumatic Stress Network. An American based network that provides excellent resources for parents, schools, the media, and professionals. www.nctsnet.org

The Children and War Foundation. This is a charity which aims to improve the care of children affected by war and disaster. Two professional groups, the Center for Crisis Psychology in Bergen, Norway and the Institute of Psychiatry in London, UK, have been instrumental in setting up this foundation. Copies of the Children's Revised Impact of Events scale (CRIES), the Depression Self-rating Scale for Children and the Revised Children's Manifest Anxiety scale (RCMAS) are available free from the website in various languages. www.childrenandwar.org

References

1. Terr, L.C. (1991). Childhood traumas: an outline and overview. *The American Journal of Psychiatry*, **148**(1), 10–20.
2. Fletcher, K.E. (1996). Childhood post-traumatic stress disorder. In *Child psychopathology* (eds. E.J. Marsh and R. Barkley), pp. 242–76. Guilford Press, New York.
3. Salmon, K. and Bryant, R.A. (2002). Posttraumatic stress disorder in children. The influence of developmental factors. *Clinical Psychology Review*, **22**(2), 163–88.
4. Scheeringa, M.S., Zeanah, C.H., Myers, L., *et al.* (2003). New findings on alternative criteria for PTSD in preschool children. *Journal of the American Academy of Child and Adolescent Psychiatry*, **42**(5), 561–70.
5. AACAP. (1998). Practice parameters for the assessment and treatment of children and adolescents with posttraumatic stress disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, **37**(Suppl. 10), 4S–26S.
6. NICE. (2005). *The management of PTSD in adults and children in primary and secondary care*. National Clinical Practice Guideline 26. Gaskell & BPS.
7. Pynoos, R.S. and Eth, S. (1986). Witness to violence: the child interview. *Journal of the American Academy of Child and Adolescent Psychiatry*, **25**(3), 306–19.
8. Ohan, J.L., Myers, K., and Collett, B.R. (2002). Ten-year review of rating scales. IV: scales assessing trauma and its effects. *Journal of the American Academy of Child and Adolescent Psychiatry*, **41**(12), 1401–22.
9. Yasik, A.E., Saigh, P.A., Oberfield, R.A., *et al.* (2001). The validity of the children's PTSD inventory. *Journal of Traumatic Stress*, **14**(1), 81–94.
10. Saigh, P.A., Yasik, A.E., Oberfield, R.A., *et al.* (2000). The children's PTSD inventory: development and reliability. *Journal of Traumatic Stress*, **13**(3), 369–80.
11. Smith, P., Perrin, S., Dyregrov, A., *et al.* (2003). Principal components analysis of the impact of event scale with children in war. *Personality and Individual Differences*, **34**, 315–22.
12. Perrin, S., Meiser-Stedman, R., and Smith, P. (2005). The children's revised impact of event scale (CRIES): validity as a screening instrument for PTSD. *Behavioural and Cognitive Psychotherapy*, **33**(4), 487.
13. Foa, E.B., Johnson, K.M., Feeny, N.C., *et al.* (2001). The child PTSD symptom scale: a preliminary examination of its psychometric properties. *Journal of Clinical Child Psychology*, **30**(3), 376–84.
14. Stallard, P. (2006). Psychological interventions for post-traumatic reactions in children and young people: a review of randomised controlled trials. *Clinical Psychology Review*, **26**(7), 895–911.
15. Dalgleish, T., Meiser, S.R., and Smith, P. (2005). Cognitive aspects of posttraumatic stress reactions and their treatment in children and adolescents: an empirical review and some recommendations. *Behavioural and Cognitive Psychotherapy*, **33**(4), 459.
16. Cohen, J.A. and Mannarino, A.P. (1998). Factors that mediate treatment outcome of sexually abused preschool children: six- and 12-month follow-up. *Journal of the American Academy of Child and Adolescent Psychiatry*, **37**(1), 44–51.
17. Trowell, J., Kolvin, I., Weeramanthri, T., *et al.* (2002). Psychotherapy for sexually abused girls: psychopathological outcome findings and patterns of change. *The British Journal of Psychiatry*, **180**(3), 234.
18. Tufnell, G. (2005). Eye movement desensitization and reprocessing in the treatment of pre-adolescent children with post-traumatic symptoms. *Clinical Child Psychology and Psychiatry*, **10**(4), 587.
19. Meiser-Stedman, R. (2002). Towards a cognitive-behavioral model of PTSD in children and adolescents. *Clinical Child and Family Psychology Review*, **5**(4), 217–32.
20. Stallard, P. (2002). *Think good—feel good: a cognitive behaviour therapy workbook for children and young people*. John Wiley and Sons Ltd., Chichester.
21. Cohen, J.A., Mannarino, A.P., and Deblinger, E. (2006). *Treating trauma and traumatic grief in children and adolescents*. Guilford Press, New York.
22. Scheeringa, M.S. (1999). Treatment for posttraumatic stress disorder in infants and toddlers. *Journal of Systemic Therapies*, **18**(2), 20–31.
23. Stallard, P., Velleman, R., Salter, E., *et al.* (2006). A randomised controlled trial to determine the effectiveness of an early psychological intervention with children involved in road traffic accidents. *Journal of Child Psychology and Psychiatry*, **47**(2), 127–34.
24. Dyregrov, A. (2001). Early intervention—a family perspective. *Advances*, **17**(3), 168–74.
25. Giannopolou, I., Dikaiakou, A., and Yule, W. (2006). Cognitive-behavioural group intervention for PTSD symptoms in children following the Athens 1999 earthquake: a pilot study. *Clinical Child Psychology and Psychiatry*, **11**, 543–53.

9.3.3 Child abuse and neglect

David P. H. Jones

Introduction

Child abuse and neglect (child maltreatment) is a combination of a consensus about what comprises unacceptable child rearing/care, together with what children have a right to be free from. This is made explicit in the United Nations Convention on the Rights of the Child,[1] which sets out basic rights and standards for judging children's welfare, including, but not limited to, maltreatment. It incorporates both maltreatment of children within families and that arising from wider social influences, including child labour and sexual exploitation, and children in war zones.[1]

Maltreatment affects the healthy and normal course of development. It causes deviation from an expected trajectory, preventing the developing child's negotiation of sequential tasks and disrupting normal transaction between different facets of development.[2] Therefore maltreatment is the very antithesis of adequate child care and rearing, posing a major public health threat.[3]

Adequate rearing of the young is such a fundamental activity that the state must be concerned with the overall welfare of children within its society; in family settings where they are normally brought up, and in schools, hospitals, and residential settings.

While the Convention provides a framework, several states have developed a children's ombudsman, with wide-ranging powers to oversee the status of children's welfare and to tackle obstacles to it.

There are laws within each society to regulate the care and welfare of children, specifying the consequences if children are maltreated. In England and Wales, the Children Acts 1989, and 2004 address the overall welfare of children, including those deemed in need of extra help and support, and provide a legislative structure for those children who are at risk of, or are actually being, significantly harmed (child maltreatment).

Countries vary in their response to child maltreatment. In the United States, any professional who has reason to suspect that a child is being maltreated is legally required to inform the local child welfare agency (mandatory reporting). Some countries in Europe (e.g. Belgium and Holland) have a system whereby child-maltreatment concerns are dealt with confidentially, through health and social care supportive systems, rather than through primarily legal methods. The United Kingdom lies between these extremes, but relatively closer to the United States model than to the 'confidential doctor' system. Whatever system is in place, it is clear from the scope of the problem of child maltreatment that multidisciplinary working is a core requirement.

A developmental-ecological model is the most useful conceptual framework, which draws together the various factors known to contribute or be associated with the predisposition, occurrence, course and effects of child maltreatment.[3,4] It incorporates individual and interpersonal factors, family influences, immediate neighbourhood ones, together with broader social influences on child rearing and care. However, these layers of increasing social complexity, which surround the individual child, are not static. In addition to transactions between factors, there are important influences historically, and subsequent to any maltreatment, which have an impact on outcome. This inclusive conceptual framework enables genetic and environmental factors to be integrated in a manner that can inform clinical assessment and intervention.

Types of maltreatment

Identification of different types of maltreatment may be necessary for social and legal purposes, but epidemiologically, co-occurrence of varieties of maltreatment is more usual than singularity.[3,5] Official registers often record predominant type or that perceived to be the most serious. This knowledge is one of several methodological problems that affect confidence in research findings. However separate types are retained here, for descriptive purposes, while encouraging the reader to consider likely overlap in individual cases.

Epidemiology

Accurate figures for incidence and prevalence are bedevilled by ascertainment and recording difficulties, including secrecy and shame which are often associated. These influences are illustrated by the wide gulf between incidence and prevalence rates.[3]

Incidence rates increase from reported cases to higher rates obtained from representative community samples. The incidence of significant violence to children varies between 50 to 90 per thousand across cultures in community samples, and dropping to the mid-20s per thousand for cases known to professionals working with children. Cases known to social welfare agencies departments only comprise a minority of these. However, officially reported maltreatment ranges from 2 to 12 per thousand in England, North America and Australia. Neglect is commonest (34 to 59 per cent of cases); physical abuse (15 to 28 per cent); sexual (10 to 28 per cent); and emotional (7 to 34 per cent).

Most prevalence figures for each of physical, emotional and neglect range between 5 and 10 per cent. The equivalent rate for contact sexual abuse is 10 per cent (15 per cent of girls; 5 per cent of boys). Children with a disability are three times more likely to be maltreated.[3]

Life-threatening maltreatment rates have remained relatively constant, currently 0.1 to 2.2/1000 children in industrialized countries, rising to two to three times this in low to mid-income countries. Children are at their most vulnerable during infancy and neonatal periods.

Child sexual abuse

Definition and clinical features

This is defined as sexual activities which involve a child and an adult, or a significantly older child. There are two elements: the sexual activities and the abusive condition.[6] Contact sexual activities include penetrative acts (e.g. penile, digital, or object penetration of the vagina, mouth, or anus) and non-penetrative acts (e.g. touching or sexual kissing of sexual parts of the child's body, or through the child touching sexual parts of the abuser's body). Non-contact sexual activities include exhibitionism, involving the child in making or consuming pornographic material, or encouraging two children to have sex together.

The abusive condition is founded on the premise that children cannot generally give consent to sex, because of their dependent condition. Consent can be difficult to assess in older children or if there is a small age gap between abuser and abused. Considering whether exploitation has occurred can aid this decision: it comprises misuse of authority or age differentials through deceit, unreasonable persuasion, coercion, or overt force.

Half the sexual abuse cases coming to the attention of welfare agencies involve penetration or orogenital contact. The proportion is less in community samples, because reported cases tend to be more serious in nature.

Abuse perpetrated by a caretaking adult normally consists of increasingly severe sexual contact over time, with parallel increases in coercion and threats to the child if the 'secret' is disclosed. As the physical acts and psychological climate worsen, so the child's reluctance to disclose the predicament deepens.

Diagnosis

The most common presentation is through a statement from the child.[7] Unless the child is responded to sympathetically at this point, they may be reluctant to reveal the full nature of their plight. More than half of those who are abused do not disclose the fact, especially if they are male.

Less commonly the child's behaviour can draw attention to abuse, particularly if the child shows sexual behaviour problems, either directed towards themselves or towards other children. However, behaviour and emotional difficulties are normally non-specific, occurring in about two-thirds of children. Older children and adolescents show behaviour difficulties which are unexpected for themselves or their peer group, including substance abuse, suicide attempts, running away from home, or becoming unpredictably

out of control. Not surprisingly, high rates of prior sexual abuse are noted among young people involved in prostitution.

Medical presentations do occur, for instance venereal diseases, evidence of acute assault, or an otherwise unexplained pregnancy.

Prior to investigation, one-third of reported cases are already known to child welfare agencies for other reasons. Children are more likely to disclose their predicament if they have first made a spontaneous statement to someone before being interviewed by professionals.

Child psychiatric services may assist social workers interviewing children and young people with a psychiatric disorder, or very young children. Other specialists should be enlisted for those with communication problems and learning difficulties. The aim of interviewing is to help a child describe their predicament whilst avoiding suggestion.[8] Child psychiatry also has a role to play in providing psychological treatments for symptomatic children and working with disturbed families.

Screening for the possibility of child sexual abuse increases recognition in both adult and child populations, revealing information that can be essential for psychiatric management. Adult services have a role to play in addressing psychiatric problems in family members, including treatment for paraphilias, often in conjunction with the probation service or other specialized provision.

Aetiological and background factors

(a) Characteristics of abused children

Sexual abuse affects children of both sexes and all ages. The most common age when children are abused is between the ages of 7 and 13 years, but up to one-quarter of reported cases comprises the under-fives. Race and socio-economic status are not major risk factors, but there are increased rates of sexual abuse among children living with parents who are emotionally unavailable, psychiatrically disturbed, violent, or who abuse alcohol or drugs.[9] Children from lower socio-economic groups are over-represented in child protection samples, but in adult retrospective surveys there is a weaker link with economic status. Children who have been in substitute care are at higher risk.

Girls are more than twice as likely to be victimized. Boys are less likely to be reported or discovered to have been abused during their childhood. Compared with girls, boys are more likely to be older when first victimized and to be abused by someone from outside the immediate family, and more likely to be abused by women or by offenders who are known to have abused other children. The risk of sexual abuse is almost doubled for children with a disability.[10]

(b) Characteristics of abusers

Most abusers are male, but up to 10 per cent of children are abused by a female, though this figure is higher when the victim is male. Of abusers, 70 to 90 per cent are known to the child, with family members comprising between a third and a half of those who abuse girls, and between 10 and 20 per cent of abusers of boys.

Up to one-third of children are abused by a person who is under 18 years of age. **Young abusers** are, on average, 14 years old, while their victims are 7 years old and usually known to them.[11] The abusers lack social skills and assertiveness, and show impulse-control problems, learning difficulties, and clinical depression. Their home environments are characterized by instability, family violence, and sexual problems in their parents. Parental loss or separation is common among adolescent abusers.

Between 20 and 50 per cent of abusers have a history of childhood sexual abuse themselves. Physical abuse histories are even more common, together with deprivation and periods of substitute care in childhood. These characteristics are common among other offenders for non-sexual abuse offences, and thus do not explain the aetiological pathways through which some young people and adults develop a pathway of sexual attraction or desire to sexually assault a child. Marshall and Barbaree[12] have drawn together psychological, biological, and social factors into an integrated theory of aetiology.

Abusers typically deny sexual abuse allegations. Even measures of penile tumescence in response to childhood imagery are unlikely to discriminate a denying abuser from a falsely accused man. Some psychological features are common among abusers but are unlikely to be definitive, prior to any admission of guilt.[13] The demarcation between intrafamilial and extrafamilial abusers is less sharp than originally thought, and mixed abusers are relatively common.

(c) Family aspects

Up to half of all cases are abused by someone outside the family. In the majority of these extrafamilial cases the abuser is known to the child and in a position of trust, either providing care or supervision, or involved in an educational or recreational activity with the child. Among within-family cases, the original stereotype-of a closed family with a controlling abusive father and mother who is collusive with her husband's abuse of her child-has been demonstrated to be inaccurate. Although such a pattern may be seen, a variety of family styles of functioning occur. However, investigators have found that families containing sexual abuse victims are less cohesive, more disorganized, and permit less healthy expression of emotion than comparison families.[14] These differences may pre-date the onset of sexual abuse or be a consequence of its occurrence.[9] Nonetheless, the observations are important for intervention purposes.

Support from non-abusive adult carers (usually mothers) in terms of belief, protection, and help for children to understand their victimization, is positively linked with the children's response to their experience.[15] This is important for assessment and intervention purposes, because there is a significant link between sexual abuse and markers of parent-child relationship difficulties, such as emotional unavailability, interparental conflict, parental mental health, and substance abuse problems.

(d) Course and prognosis

A wide range of psychological sequelae in childhood and adult life are associated with prior childhood sexual abuse (Table 9.3.3.1).[9,15] However, these are linked with the effects of both the quality of the family environment at the time of abuse, and the nature of subsequent life events.[16] In particular, factors such as family disharmony and violence, existence of other forms of abuse and neglect, and parental mental health difficulties, in addition to subsequent events, such as losses through death or separation, combined with the child's own method of coping with the abuse and ameliorative effects of positive school or social relationships, all contribute to outcome.

About one-third of children are symptom free. Approximately 10 per cent of children show worsening symptoms over time, including depression and post-traumatic symptoms. While effects on personality and social relationships can be disabling during development, other children are relatively unaffected.[16,17]

Table 9.3.3.1 Impairments and problems associated with childhood sexual abuse

	Childhood impairment	Adult impairment
Affective symptoms	Fears PTSD Depression	Anxiety PTSD Depression
Behaviour problems	Conduct disorder Sexualized behaviour Self-destructiveness Hyperactivity	Aggressive conduct Self-destructiveness Alcohol/substance abuse
Cognitive functioning	Educational problems Language difficulties	Educational underachievement
Personality and social adjustment	Self-esteem Attachment Peer relationships	Pregnancy under 19 years Sexual aggression Prostitution Parenting problems Somatization Personality disorder Revictimization Sexual problems

Physical abuse

Definition and clinical features

Physical abuse is the physical assault of a child by any person having custody, care, or charge of that child. It includes hitting, throwing, biting, inducing burns or scalds, poisoning, suffocating, and drowning.[3] In the United States and United Kingdom physical chastisement of children is commonplace, leading to problems of definition. In other parts of Europe and in some Eastern cultures physical chastisement is regarded as unacceptable. Legal definitions in the United States and Western Europe normally link physical acts to observable harm. However, for research and clinical purposes an endangerment-based definition is preferable, because of the widely different sequelae resulting from similar assaults.[18] Failure to prevent injury or suffering is preferably considered a manifestation of neglect. Other definition problems include the frequency or repetitiveness of the acts, their severity, and whether intent to harm should be included. In addition, developmental factors affect the recognition of abuse and possibly its definition also—a smack to the head of an 8-year-old, although unacceptable, will have significantly different consequences from that to an 8-month-old.

The distinction between accidental injury, non-accidental injury, and specific medical diseases is sometimes straightforward (e.g. particular types of fractures, burns, or bruising) but difficult diagnostic dilemmas do occur. It is important to resolve these dilemmas so that the way forward for psychiatric assessment and treatment can be clarified.[19]

The 'battered child syndrome' refers to young children with multiple bruises, skeletal injuries, and head injuries, often accompanied by neglect, malnutrition, and fearfulness, whose parents deny responsibility.[20]

Diagnosis and recognition

Physical abuse is detected through the observation of physical injuries without an alternative non-abusive explanation.[19] Less commonly, a direct account comes from a child or a witness, or through confession by a parent or carer. Usually, the diagnosis is based upon a discrepancy between the physical findings and the history provided. The history may be insufficient or simply improbable. When an explanation is forthcoming, trigger events or developmental challenges are common—for example, persistent crying in infancy, problems of toileting or feeding among toddlers, or issues of discipline in later childhood. In adolescents, conflict surrounding independence may coincide with parental midlife crises. Not all physical abuse can be related to loss of control, however, and the assessor has to consider planned or even sadistic activities, such as scalding, burning, or torture.

There may have been previous episodes of similar or lesser concern, for which adequate explanations were unavailable at the time. Delay in presenting the child for medical attention is not a reliable diagnostic feature; neither is the apparent absence of parental concern nor their unreasonable behaviour at presentation.

Aetiology and background factors

(a) Child characteristics

Physical abuse occurs at all ages, although biological sequelae are more severe in infancy. There is no association with ethnic group, but a strong one with low socio-economic status among the under-fives, becoming weaker throughout childhood and disappearing by adolescence.[21] Children with developmental disabilities have a raised risk.[10] Associations with low birth weight, prematurity, or physical ill health disappear once parental and social variables are controlled for. Boys under 5 years of age are more likely to be abused, whereas girls are at greater risk in childhood overall.

(b) Abuser characteristics

Young maternal age at the time of the child's birth is linked with abuse, but generally the effect of age is overshadowed by low socio-economic status and high social stress.[21] Physical abusers of young children are likely to be female, but male abusers predominate during adolescence. They are more likely to be single parents and to have large numbers of closely spaced children. Their educational level, but not necessarily their intelligence, is lower; they are, however, more likely to be unemployed. Most physical abuse is perpetrated by parents, but others who adopt a caretaking role become increasingly significant in the abuse of older children.

Abusive parents are more likely to have had a childhood history of abuse themselves. However, regarded prospectively, 70 per cent of abused children do not abuse their own children.[22] Non-repeaters are more likely to have enjoyed social support from a partner, had a positive relationship with an adult during childhood, and to have received psychological help during adolescence. In addition, they have a more balanced and coherent perspective about their childhood experiences than those who show intergenerational continuity of parenting problems. The quality of attachment relationships between parents and children shows continuity, rather than the specific type of abuse. Hence, physically abused children have an increased risk of perpetrating both physical and sexual abuse when they become parents themselves.

Frank psychiatric disorder is relatively infrequent among abusers, but studies of physical abuse fatalities underline their importance in a minority of cases.[23] Personality difficulties and disorders are more common, however. Hostile adults with poor impulse control, low self-esteem, antisocial and aggressive personalities,

with accompanying mood disorder are more likely to abuse. These abusers have disrupted social relationships and inadequate coping responses in a wide range of domains. They are frequently socially isolated, alienated, and have disharmonious relationships with neighbours and relatives. For these adults, potentially protective supportive relationships with friends and relatives are inhibited.[3,21]

Abusive parents have maladaptive ideas about their children. They tend to have high expectations for their children's development and behaviour, perceiving it to be deviant when objectively it is not. They are more likely to believe in the appropriateness of strict physical discipline, and to hold negative views and perceptions about their children. They show limited attention to their children, less positive affect, and respond with aversion, anger, or irritation to their children's bids for care or attention, as well as to their positive behaviours, when compared with non-maltreating parents. Physically, abusers show heightened arousal to both child stimuli and non-child-related stressors.[3,21]

(c) Family aspects

Families in which physical abuse occurs are more likely to support mutually abusive coercive communications and interactions than controls. Partner abuse and domestic violence is relatively more common, combined with pervasive hostility and decreased cohesion. Discussion, positive displays of affection, and encouragement of prosocial behaviours are less common than in non-maltreating families.[3,21]

The quality of attachment between child and parent is significantly linked with physical abuse, especially when combined with high levels of social stress, low socio-economic status, and negative parental family attitudes and behaviours. Although infant temperament can be associated with maltreatment it probably only does so if combined with other risk factors, such as parent–child attachment problems, parental attitudes, and family difficulties of the sort described above. Clinicians have long observed that individual children can be perceived negatively by parents, without objective evidence, particularly if the child represents a particular issue or problem for the parent.

Course and prognosis

Some physically abused children have neurological and other physical sequelae as a result of their injuries.[4] Educational difficulties are consistently found on follow-up. The children are less attentive to social cues and less skilful at managing personal problems and more likely to attribute a hostile motivation to their peers, compared with non-abused children, at the age of 5. Their capacity for empathic concern with the everyday problems of their peers becomes blunted. Not surprisingly therefore, chronic oppositional and aggressive behaviour is the most consistently documented childhood outcome. These children range from the socially withdrawn and avoidant, to those who demonstrate fear, anger, and aggression. These features are linked both to the physical abuse and the family context of pervasive aggression and conflict.[3]

The children's attachments to their caretakers are anxious and insecure. Children view themselves negatively, and show increased rates of both depression and anxiety throughout childhood. Long-term exposure results in a constellation of reactions characterized by pervasive denial by the child, an apparent repression or dissociation of memories, relative indifference to pain or distress, episodes of rage directed towards self or others, and an unremitting sadness. Male victims may develop a characteristic hypervigilance.[21]

The major health consequences of physical abuse in childhood have become clarified.[3] The causal connection between physical abuse and later psychological and physical health problems is underlined through clear links between early age of onset, and severity of maltreatment and subsequent severity of psychological and physical ill effects in teenage and adult years. Further, physical abuse cases embedded within violent families and associated with accompanying neglect have relatively worse outcomes, psychosocially and in physical health. Approximately 20 to 30 per cent of physically abused children develop conduct problems in teenage years, starting earlier and displaying more violence than their non-abused counterparts. They are at increased risk of running away from home, and are overrepresented among young homeless children in inner cities. Childhood physical abuse is associated with subsequent substance abuse problems, self-destructive behaviour and suicidality, depression, teenage pregnancy and poor physical health outcomes. Genetic factors mediate the association between physical abuse and later antisocial behaviour and, probably, affective disorder too.[3]

Child neglect

Definition and clinical features

Neglect refers to the underprovision of the child's basic needs, both physical and psychological. Most cases comprise omissions of care by parents and others in the parental role. However, institutional neglect also occurs, mainly in the form of collective caretaking failure—for example, residential children's homes in the United Kingdom, orphanages and nurseries in Eastern Europe, and neglect of care by educational establishments.[3]

Definition problems include whether neglect should include the apparent impact on the child and/or the degree to which it was intended.[3] There are cultural variations in what might be perceived as neglect. The practice of putting young children into separate bedrooms, while considered normal practice in much of Western Europe, would be considered frankly neglectful in some Eastern cultures.

Notwithstanding these definition problems, four main types can be identified: physical, supervisory, cognitive, and emotional neglect. Neglect can occur first during the prenatal period, for example through maternal substance abuse, and may be observed throughout childhood. **Physical** neglect includes inadequate nutrition, clothing, shelter, but also exclusion and abandonment. This is the most common form of neglect reported to welfare agencies in North America, Australia and Western Europe. **Supervisory** involves inadequate parental overview, relative to the child's needs, for developmental needs of the child, but also employing unsafe alternative carers, and failure to use available health care. **Emotional** neglect includes insufficient parental affection, and inattention to the child's cues, which has been termed 'psychological unavailability'.[24] **Cognitive** is insufficient parental responsiveness, attention and speech, but also denying access to education opportunities.

Diagnosis and presentation

Although most reported cases involve younger children, neglect occurs at all ages. Many cases are followed for years before being

finally identified by professionals. Non-organic failure to thrive can precipitate earlier recognition. Otherwise neighbours, relatives, or school teachers report the child's plight to protection agencies, by which time the effects are severe and neglectful caretaking entrenched. Recognition may also come about through the child's presentation with developmental delay, language problems, school non-attendance, inadequate medical or dental care, or with significant psychological difficulties. Conclusions about neglect need to be linked with the individual's developmental needs. Additionally neglect must be distinguished from the effects of poverty. Conclusions are assisted by using multiple sources of information; from the children themselves, caregivers, reviewing longitudinal case records, direct observation and standardized measures.[3]

Aetiological and background factors

(a) Characteristics of neglectful parents

Neglectful parents are likely to be poor, have multiple difficulties, and display what has been described as the apathy-futility syndrome. Parents show immature personality characteristics, with low self-esteem, impulsivity, and an inability to plan or demonstrate choice in such important areas as adult partners, having children, or employment. Neglectful parents frequently hold inaccurate or unrealistic expectations about their children's development or behaviour. Neglect may derive from parental psychiatric illness such as schizophrenia, depression, or drug or alcohol abuse.

(b) Characteristics of neglected children

Neglected infants have anxious, disorganized attachments with their caretakers. Later in childhood they are more aggressive than comparison children, though less so than physically abused children. Neglected toddlers show non-compliance and become easily frustrated, later developing low self-esteem and self-assertiveness and showing less flexibility or self-control. Both in preschool and school they lack persistence and enthusiasm, and become socially isolated.

(c) Family aspects

Child neglect is normally embedded within broader family insularity, lack of cognitive stimulation, affection or emotional nurturing between its members, and significant household disorganization. Neglectful parents are likely to be unresponsive to both their infants and older children, showing a paucity of prosocial positive behaviours, less interactions and stimulation, and more negative behaviours than controls. Even though there is a strong link with poverty, parents who neglect children stand out among their equally materially impoverished neighbours.

Course and prognosis

The seeds for the neglected child's long-term difficulties with social interaction, relationships, and educational progress can be observed in infancy. Neglected children tend towards passiveness and helplessness under stress. They show significant developmental delays, especially language problems, attention-seeking behaviour, and superficial displays of affection, as well as conduct problems, persistent defiance, and hostile behaviour. Studies of children who as infants were subjected to psychologically unavailable caretaking reveal persisting difficulties with anger, non-compliance, low frustration tolerance, little enthusiasm or persistence for tasks, poor impulse control, relative rigidity, and lack of creativity. Similar negative developmental outcomes have been reported to occur following non-organic failure to thrive in infancy, especially where combined with physical neglect, leading to long-term cognitive delay and poor educational attainment. Children who as infants experienced psychologically unavailable parenting do even worse than physically abused children, showing a greater number of emotional problems, inattention, social withdrawal, and unpopularity with other children.[24]

Psychological maltreatment

Definition and clinical features

Emotional abuse (better termed 'psychological maltreatment') refers to those interactions with children that have the potential to damage the child psychologically, given his or her particular developmental needs.[3] Four broad groupings of acts are described: the need for psychological safety and security; for acceptance and positive regard; for age appropriate autonomy, and sufficient opportunities to explore environment and extra familial relationships. Included within **psychological safety** are exposure to domestic violence, threats of injury, suicide or abandonment, and discipline through intimidation. Within **acceptance and self-esteem** are verbal and non-verbal negativity, active rejection, ridiculing, inappropriate expectations and undermining. **Age appropriate autonomy** includes both inappropriate responsibility giving and prohibiting age appropriate socialization and placing a child in a reversed parental role. **Restriction** includes restrictive confinement and isolation.

Psychological maltreatment may be direct toward the child, or operate indirectly, for example through the child witnessing domestic violence, or observing parental involvement in antisocial activities. It may occur in institutional settings as well as within families. The overlap with neglect is evident from the list of acts of omission and commission listed above.

Diagnosis and recognition

Recognition may occur when other kinds of maltreatment are discovered, or when domestic violence is revealed. It may also occur when a child is noted to be living with, and/or providing care for a parent with mental or physical illness, personality disorder or substance abuse. Sometimes recognition follows a child's referral to developmental or mental health clinic, or through the reported observations of neighbours or professionals (e.g. teachers, police). Diagnosis requires detailed history, with examples, direct observations of parent-child interactions, and interviews with older children. Standardized data gathering schemas may assist diagnosis.

Aetiological and background factors

(a) Characteristics of abused children

Reports of emotional abuse in children become more frequent throughout childhood into adolescence. Reported cases are more likely to be linked with lower socio-economic status. There is no particular link with racial or ethnic groups. Psychological maltreatment is frequently integral to other forms of maltreatment and so distinguishing different aetiological factors and consequences is complex.

(b) Characteristics of abusers

Although not systematically studied, this probably varies according to the mixture of subtypes present, and whether any other kinds of abuse or neglect coexist.

(c) Course and prognosis

Psychological maltreatment in infancy has a very poor outlook (see discussion of neglect). Much less is known about the outcome of different mixtures of psychological maltreatment identified during childhood and adolescence. There are indications that the degree and extent of psychological maltreatment is a better predictor of case outcome than the extent of any coexisting physical or sexual abuse, thus underlining its importance to the developing person's mental health.

Fabricated or induced illness (Munchausen syndrome by proxy)

Definition and key elements

Fabricated or induced illness (FII) is where a parent or carer feigns an impression or induces a state of ill health in a child whom they are looking after. The key elements are parental falsification or deceit, and a triangular interaction between parent, child, and health professional, in which the doctor is misled by the parent, some parental need is met, and the child harmed (directly or indirectly). The harm occurs through: verbal fabrication of symptoms/signs; falsification of reports or specimens; or through inducing ill health (either actively or by withholding essential substances).[25]

Diagnosis

The presentation can be in any bodily system, but common forms are fabricated epilepsy, non-accidental poisoning, apparent life-threatening events in infancy (either directly induced suffocation, or fabricated), or multi-system disorders (e.g. gastrointestinal and renal problems).

The diagnosis of fabrication is almost always undertaken by paediatricians, whose awareness of the possibility of fabricated signs or symptoms is now much greater and leads to an earlier diagnosis than when first described by Meadow.[26]

There are several elements to the phenomenon:

1 the harm caused to the child through fabrication;

2 the impact on the child's development, both physically and emotionally;

3 the psychological status of the fabricator.

Psychological services are especially involved in (2) and (3)—assessing the child's developmental status, and considering the mental state of the fabricator and assessing family dynamics. Differentially, factitious illness by proxy needs to be distinguished from parental overanxiety or exaggeration, or frank malingering, though sometimes FII contains elements of all these.[25]

Epidemiology

The annual incidence among children in the United Kingdom has been calculated to be 0.5 per 100 000, but for those under 1 year it is 2.8 per 100 000.

Aetiological and background factors

(a) Characteristics of abused children

The majority of children are under 5 years of age, with boys and girls equally represented. Affected infants are likely to have feeding problems; withdrawal and hyperactivity are seen in school-age children, whereas adolescents may develop somatization themselves. Up to three-quarters of the children show evidence of other fabrications, or of physical abuse and neglect.

(b) Abuser characteristics

Most fabricators are female—79 per cent of whom have a somatization disorder themselves and half have a personality disorder, particularly so among fabricators who induce illness. Most abusers deny responsibility, at least initially.

(c) Family characteristics

Unusually, families are often intact, though 40 per cent have serious marital problems. Child-parent attachment difficulties are common, and other siblings in the family may be affected. Typically, fathers are not involved in family life.

Course and prognosis

Affected children may be damaged by the abuse itself, while mortality is between 5 and 10 per cent. About 20 per cent are reabused, though not necessarily in the same way as the original FII. Emotional harm, conduct problems, and educational difficulties occur in half the children on follow-up.

Prevention of maltreatment

Reducing the incidence and recurrence of child maltreatment are crucial initiatives, because of the resultant ill effects[27] and difficulty in instituting effective treatments, quite apart from the humanitarian prerogative. Family support—and education concerning parenting, child development, and the management of problems—has a positive impact on parental attitudes and knowledge as well as on observed behaviour. Significant effects on children's behaviour, cognitive outcome, and child maltreatment rates are less clear. Preventive efforts have more impact upon physical abuse than on neglect.[28]

Brief interventions are beneficial for low-risk parents, whereas more intensive approaches are needed for higher risk groups. High-risk groups include deprived, impoverished parents, young mothers, and those with a personal history of childhood abuse themselves. However, for a prevention approach to be effective it must be personalized to the needs of the individual families and include outreach components for the most negative and hard to access parents.[28] Including males, whether resident or occasional visitors, is crucial to maintaining and sustaining improvements in parenting and child care. Equally, effective programmes are more likely to be valued by the parents themselves, underlining the importance of matching the skills of staff and the contents of programmes with the families' specific needs.[29,30] Similarly, programmes must be culturally sensitive if they are to be effective and accepted by parents. Interventions are most effective when they impact upon a broad network of influences and relationships, ranging from those of the immediate family to broader neighbourhood and social influences on children's welfare.[3] Furthermore, primary, secondary and tertiary prevention approaches need to be integrated and carefully planned within each area.[31]

Sexual abuse prevention is probably ineffective when aimed at enhancing children's capacity to protect themselves.[28] On the other hand, programmes which include parents (and increase their capacity to keep their children safe) and incorporate antibullying tactics are much more likely to be effective, although this probably stems from increasing disclosure of early sexually abusive actions rather than primary prevention.[28] Additionally, broader social

initiatives to reduce child sexual abuse within schools and institutions are essential to a comprehensive area strategy.[27,31]

Intervention and psychological treatment

Treatments

Psychological treatments have been developed for different types of maltreatment, notwithstanding contemporary appreciation that co-occurrence of types of maltreatment is common place. Effective treatments for physical and sexual abuse are reasonably well established.

Interventions with empirical support are principally behavioural, and cognitive behavioural ones.[3] These are normally structured and emphasize skill building to overcome emotional distress and behavioural disturbance in children and parents. Trauma-focussed, cognitive behavioural therapies are significantly superior to family and general therapeutic treatments for both physical and sexual abuse. These treatments combine psycho-education, exposure therapy, cognitive procedures and restructuring as well as behavioural management. Children are also assisted with emotional recognition and regulation and attention to maladaptive ideas. Other empirically supported interventions include child parent psychotherapy (blending psychodynamic and cognitive behavioural components).

Psychological treatments in neglect have concentrated on improving parenting skills and sensitivity through direct encouragement of positive interactions in feeding, play, general care, combined with individual therapy for parents themselves.[3] Parallel psychiatric treatment of parental mental health problems, such as depression or substance abuse is also important, although there is debate about whether such treatments should precede or be delivered subsequent to treatments focussed on parenting. There has been some support for therapeutic day treatment programmes and multisystemic therapy for neglect.

Psychiatric interventions for psychological maltreatment have not been subject to empirical evaluation thus far. In the meantime it seems reasonable to focus on improving sensitivity and responsiveness within parent child relationships (through direct work and feedback), together with family based work, and individual work with parents who have been subject to deprived or abusive backgrounds themselves.

Treatments for FII are sparse, but combinations of treatments for neglect and psychological maltreatment, combined with behavioural management of somatization appears promising.[3]

Management

(a) Guiding principles[3,32]

The first priority is to establish the **child's safety** and/or freedom from neglect. This may require separation of child and abuser. Those cases involving neglect or psychological maltreatment may be managed through verifiable agreement, and providing services designed to promote the child's welfare within the family. Interdisciplinary planning and coordination will be essential in order to achieve this.

Interventions must be focused primarily on the **child's welfare**, rather than other objectives such as adult treatment or family preservation. Interventions should also focus on the child's **physical and emotional safety**, and be **developmentally focused**, while simultaneously tackling **parenting problems**. A developmental focus involves adapting treatments to a child's age and developmental status, and taking into account any developmental impairment. Additionally, it is essential that treatment approaches comprehensively address those developmental processes affected by maltreatment, i.e. affect regulation, attachment, the evolving self system, and peer relationships.[5]

Approaches to parenting problems range from interventions with parents individually, those focused specifically on parent child dyads, through to more broadly based social support services, including day attendance at family centres designed to improve parenting care, the use of family support workers, harnessing neighbourhood supports and other parents prepared to assist new parents at risk, and parenting classes. Services to improve family conflict resolution and parental management of hostility and aggression may also be needed for a comprehensive approach to tackling parenting problems. Parents with mental health problems need to be psychiatrically evaluated to see whether treatment could assist overall case management. Sexually abusive behaviour or physical violence perpetrated by adults may be amenable to psychiatric intervention. Persistently dangerous abusers need to be identified and child safety assured.

Further, interventions should specifically address the **child's experience** of maltreatment and any moral or legal dimensions to this. Interventions should not focus solely on maltreatment itself, but equally address any **general mental health issues**. They should also aim to **prevent future difficulties** as well as addressing current problems. Supportive and **nonabusive parents and carers** should be part of the intervention and incorporated into treatment plans. Obtaining **parental acknowledgement and recognition** of their part in maltreatment is an important part of successful interventions. Parental acknowledgement may be achieved through education and discussion, and through treatment approaches designed to alter hostile or neglectful views and attributions held by parents. First choice interventions should be those with the highest level of **empirical support**. Generally, interventions should involve school and **local networks** of child professionals in a systemic fashion. A systemic perspective is helpful with respect to all families. Sensitive working across disciplinary boundaries requires knowledge of **local** family justice system and child protection **working practices**.

(b) Case planning

When deciding what to do in the individual case, treatment may have to address several aspects of abuse and neglect, because co-occurrence of maltreatment type is common. Safety is of paramount consideration. There is little value in starting treatment if the child remains unsafe. This will require multidisciplinary or multiagency co-operation in order to develop a comprehensive plan for child and family. Even where intensive treatments are unavailable, psycho-educational approaches are helpful for all abused children. In reality, most maltreated children receive no systematic therapeutic intervention.

The majority of treatment approaches combine intervention for the child with adult treatment and often family based work. Supportive work with non-abusive carers is of proven value. In planning this, however, it must be remembered that failure to protect a child is harmful and dangerous for the child's future too. In addition, in one family there may be more than one maltreating carer, e.g. abuse by one adult carer and neglect by another.

Effectiveness of intervention is gauged through several dimensions. For example, child safety, improved carer availability and sensitivity for the child, whether the child has overcome effects of trauma, psychological symptom reduction, improved peer relationships, speech and language catch up and educational progress. Improvements in family functioning which are sensitive to interventions include; reduced conflict and violence, improved communication and emotional expression, changes to disturbed attachment patterns and improved child to parent attachment, and greater warmth in parent/child relationships. In addition, effectiveness may be gauged through target mental health objectives for individual carers, e.g. reduction in parental depression or improvements in anger management.

Developmental considerations will also contribute to evaluation of effectiveness. For example, a shorter timescale is appropriate for younger children in view of their developmental needs; and greater than average improvement in parenting may be needed for children with impairments or disabilities.

Risk management is a central issue for mental health practitioners when providing interventions in maltreatment cases, and underlines the need for systemic awareness and well-developed interdisciplinary practice. Risk factors for occurrence are reviewed elsewhere.[3] Recurrence risk has been the focus of systematic review[33], and narrative overview, using an eco-developmental perspective and identifying risk elevating as well as lowering factors.[34] Nonetheless, multiplicity of relevant factors as well as complexity of transactions in individual cases render actuarial approaches to risk management an unrealistic aspiration at this point. In its place, an approach that has been characterized as structured professional judgment appears the most appropriate.[35] Key risk elevating factors for recurrence include: domestic violence, child neglect, cases where there had been previous maltreatment, and parents with mental health disorders. A structured approach to risk management entails an approach to data gathering, diagnostic formulation and subsequent decision making that rests explicitly on available evidence about risk factors for recurrence.[34]

(c) Ethical and legal considerations

Decisions as to whether to share sensitive information with other professionals will be guided by child welfare and safety considerations, which are paramount, and override adult consent withheld. Sometimes children request confidentiality, in situations where their safety is potentially compromized. Normally, explaining to children why information must be shared, and involving them enables their trust to be maintained even where there is initial disagreement.

Generally, patient information will need to be shared during multidisciplinary planning meetings, but agreement can usually be made to respect confidentiality, except to the degree necessary to assure child safety. Many cases will involve family justice systems, which normally require overview of progress. Practitioners will need to provide reports for planning meetings and family courts, in order to contribute to safe care.

(d) Long term

Mental health effects may present at different points in the life course, particularly at key times of developmental change, e.g. at a first romantic relationship, or when becoming a parent. If possible, such possible future difficulties should be noted in a person's medical record, to facilitate future intervention. As part of an area's comprehensive prevention approach, new parents with a history of childhood maltreatment should be offered extra support, as they are a vulnerable group for parenting problems (although it is important to stress that discontinuity of parenting problems is more common than continuity).

Further information

Web sites

California Evidence-based Clearinghouse for Child Welfare. http://www.cachildwelfareclearinghouse.org/

Every Child Matters. http://www.everychildmatters.gov.uk/

International Society for the Prevention of Child Abuse and Neglect. http://www.ispcan.org/

National Society for the Prevention of Cruelty to Children. http://www.nspcc.org.uk/Inform/informhub_Wda4993.html

References

1. World Health Organization. (2002). *World report on violence and health*. Geneva: World Health Organization.
2. Cicchetti, D. (1989). How research on child maltreatment has informed the study of child development: perspectives from developmental psychopathology. In *Child maltreatment; theory and research on the theory and causes of child abuse and neglect* (ed. D. Cicchetti and V. Carlson), pp. 377–431. Cambridge University Press, Cambridge.
3. Jones, D P H. (2008). Child maltreatment. In *Rutter's child and adolescent psychiatry* (5th edn) (ed. M. Rutter, D. Bishop, D. Pine, *et al.*). Blackwell, Oxford.
4. Cicchetti, D. and Toth, S. (1995). A developmental psychopathology perspective on child abuse and neglect. *Journal of the American Academy of Child and Adolescent Psychiatry*, **34**, 541–65.
5. Leventhal, J. (2007). Children's experiences of violence: some have much more than others. *Child Abuse and Neglect*, **31**, 3–6.
6. Finkelhor, D. (1994). Current information on the scope and nature of child sexual abuse. *Future of Children*, **4**, 31–53.
7. Jones, D.P.H. (1997). Assessment of suspected child sexual abuse. In *The battered child* (5th edn) (ed. R. Helfer, R. Kempe, and R. Krugman), pp. 296–312. University of Chicago Press, Chicago.
8. Jones, D.P.H. (2003). *Communicating with vulnerable children: a guide for practitioners*. Gaskell, London.
9. Berliner, L., and Elliott, D. (2002). Sexual abuse of children. In *The APSAC handbook on child maltreatment* (eds. J. Myers, L. Berliner, J. Briere, *et al.*), pp. 55–78. Sage, London.
10. Westcott, H.L., and Jones, D.P.H. (1999). Annotation: the abuse of disabled children. *Journal of Child Psychology and Psychiatry*, **40**, 497–506.
11. Vizard, E., Monck, E., and Misch, P. (1995). Child and adolescent sex abuse perpetrators: a review of the research literature. *Journal of Child Psychology and Psychiatry*, **36**, 731–56.
12. Marshall, W., and Barbaree, H. (1990). An integrated theory of the aetiology of sex offending. In *Handbook of sexual assault* (ed. W. Marshall, D. Laws, and H. Barbaree), pp. 257–75. Plenum Press, New York.
13. Becker, J. (1994). Offenders: characteristics and treatment. *Future of Children*, **4**, 176–97.
14. Madonna, P., Van Scoyk, S., and Jones, D.P.H. (1991). Family interactions within incest and non-incest families. *American Journal of Psychiatry*, **148**, 46–9.
15. Jones, D.P.H., and Ramchandani, P. (1998). *Child sexual abuse—informing practice from research*. Radcliffe Medical Press, Oxford.

16. Glaser, D. (2008). Child sexual abuse. In *Rutter's child and adolescent psychiatry* (5th edn) (ed M. Rutter, D. Bishop, D. Pine, *et al.*). Blackwell, Oxford.
17. Rind, B., Tromovitch, P., and Bauserman, R. (1998). A meta-analytic examination of assumed properties of child sexual abuse using college samples. *Psychological Bulletin*, **124**, 22–53.
18. National Research Council (1993). Scope of the problem. In *Understanding child abuse and neglect*, pp. 78–105. National Academy Press, Washington, DC.
19. Johnson, C. (2002). Physical abuse: accidental versus intentional trauma in children. In *The APSAC handbook on child maltreatment* (ed. J. Myers, L. Berliner, J. Briere, *et al.*), pp. 249–68. Sage, London.
20. Kempe, C.H., Silverman, F.N., Steele, B.F., *et al.* (1962). The battered child syndrome. *Journal of the American Medical Association*, **181**, 17–24.
21. Kolko, D. (2002). Child physical abuse. In *The APSAC handbook on child maltreatment* (ed. J. Myers, L. Berliner, J. Briere, *et al.*), pp. 21–54. Sage, London.
22. Kaufman, J., and Zigler, E. (1989). The intergenerational transmission of child abuse. In *Child maltreatment; theory and research on the theory and causes of child abuse and neglect* (ed. D. Cicchetti, and V. Carlson), pp. 129–50. Cambridge University Press, Cambridge.
23. Jones, D.P.H., and Lynch, M.A. (1998). Diagnosing and responding to serious child abuse. *British Medical Journal*, **317**, 484–5.
24. Erickson, M., and Egeland, B. (2002). Child neglect. In *The APSAC handbook on child maltreatment* (ed. J. Myers, L. Berliner, J. Briere, *et al.*), pp. 3–20. Sage, London.
25. Jones, D.P.H., and Bools, C.N. (1999). Factitious illness by proxy. In *Recent advances in paediatrics* (ed. T. David), pp. 57–71. Churchill Livingstone, Edinburgh.
26. Meadow, S.R. (1997). Munchausen's syndrome by proxy: a hinterland of child abuse. *Lancet*, **ii**, 232–5.
27. Daro, D., and Cohn Donnelly, A. (2002). Child abuse prevention: accomplishments and challenges. In *The APSAC handbook on child maltreatment* (ed. J. Myers, L. Berliner, J. Briere, *et al.*), pp. 431–48. Sage, London.
28. Wolfe, D., Reppucci, N., and Hart, S. (1995). Child abuse prevention: knowledge and priorities. *Journal of Clinical Child Psychology*, **24** (Supplement), 5–22.
29. Cox, A. (1998). Preventing child abuse: a review of community based projects. II: Issues arising from reviews and future directions. *Child Abuse Review*, **7**, 30–43.
30. Leventhal, J. (2005). Getting prevention right: maintaining the status quo is not an option. *Child Abuse and Neglect,* **29**(3), 209–13.
31. MacMillan, H.L., MacMillan, J.H., Offord, D.R., *et al.* (1994). Primary prevention of child physical abuse and neglect: a critical review—part I. *Journal of Child Psychology and Psychiatry,* **35**, 835–56.
32. Saunders, B.E., Berliner, L., Hanson, R.F. (eds.) (2004). *Child Physical and Sexual Abuse: guidelines for treatment (Final Report: January 15, 2003)*. Charleston, SC; National Crimes Victims Research and Treatment Center.
33. Hindley, N., Ramchandani, P., and Jones, D.P.H. (2006). Risk factors for recurrence of maltreatment: a systematic review. *Archives of Disease in Childhood*, **91**(9), 744–52.
34. Jones, D.P.H., Hindley, N., and Ramchandani, P. (2006). Making plans: assessment, intervention and evaluating outcomes. In *The Developing World of the Child* (eds, J. Aldgate, D. Jones, W. Rose, *et al.*) (Chap 15, pp 267–86). London: Jessica Kingsley.
35. Douglas, K.S., and Kropp, P.R. (2002). A prevention-based paradigm for violence risk assessment: Clinical and research applications. *Criminal Justice and Behavior*, **29**(5), 617–58.

9.3.4 The relationship between physical and mental health in children and adolescents

Julia Gledhill and M. Elena Garralda

Introduction

The link between physical and psychological disorder in children and adolescents is well established. Children with chronic illness are at increased risk of emotional and behavioural disorders. In addition, repeated presentations with physical symptoms may represent underlying psychological distress or psychiatric disorder.

Because of the inextricable links between young people and the family in which they live, it is inappropriate to consider symptoms in an index child in isolation. The effects of symptomatology on family functioning, parent, and sibling relationships should be considered. This may have important aetiological and prognostic significance.

Associations between physical and psychological symptoms

There are various ways in which physical and psychological disorders are related; these are summarized in Table 9.3.4.1.

In this chapter we shall consider the following:

- The psychiatric consequences of physical illness
- Helping the dying child and his or her family
- The effects of psychiatric disorder on the course and outcome of physical illness
- Aspects of assessment and treatment intervention

Table 9.3.4.1 Associations between physical and psychological symptoms

Nature of association	Examples
Psychiatric consequences of physical illness and treatment	*Organic*: acute confusional state, psychosis induced by brain disorder *Functional*: adjustment disorder after diagnosis of diabetes, specific needle phobia in young child with cancer receiving chemotherapy
Effects of psychiatric disorder on physical illness	Depression delaying the mobilization of a child following partial limb amputation after severe meningococcal disease, oppositional-defiant disorder affecting treatment adherence in diabetes
Physical complications of psychiatric problems e.g. deliberate self-harm, substance abuse	Liver failure following paracetamol overdose
Psychiatric disorders or psychological distress presenting with physical symptoms	Aches and pains in school age children, reduced physical well-being in adolescent depression, somatoform pain disorder, dissociative disorder

- Somatization and somatoform disorders, with a particular focus on recurrent abdominal pain, dissociative/conversion disorder, and chronic fatigue syndrome

Psychiatric aspects of chronic physical illness

Chronic physical illness and the risk of psychiatric disorder

Chronic physical illness in children, defined as disorders that last at least 1 year and are associated with persistent or recurrent handicap, affects about 4 per cent of children in Western countries.[1] This encompasses a broad spectrum of disorders including more common problems such as eczema, asthma, diabetes, epilepsy, and less prevalent conditions such as cystic fibrosis and cancer. Many children successfully adapt to living with a chronic illness, but it can be associated with a number of different types of stresses for children and their families.

The stress of chronic illness may operate at several levels. In addition to the presence of the illness itself, diagnostic and treatment procedures may be painful or have undesirable side-effects—changes in physical appearance such as alopecia, scars, and obesity may lead to difficulties in peer relationships. The demands of treatment such as dietary restrictions in diabetes may be difficult. The illness, together with hospital attendance for treatment, may lead to a considerable interruption to schooling as well as a reduced ability to participate in leisure activities and socialize with peers.

Although the majority of children and families successfully adapt to these stresses, children with chronic physical illness have a slightly increased risk for the development of associated psychiatric disorders. Specific factors related to the child and the illness have been shown to contribute to the likelihood of developing psychiatric disturbance and to influence the nature of the psychiatric disorder that develops (Table 9.3.4.2).[2]

(a) Nature of the physical disorder

Much of the increased prevalence of psychiatric disorder in children with chronic physical illness is accounted for by those with disorders affecting the brain, especially when epilepsy is involved.[3] They have a three-fold increased risk of psychiatric disorder over general population rates. The risk in young people with a chronic physical illness that does not involve the brain is considerably lower and only slightly increased over general population expectations.[3] The excess of psychopathology in children with brain anomalies may be attributable to the direct effects of organic pathology on behaviour, or may be mediated by the greater physical disability that frequently accompanies brain damage. Associated intellectual impairment may also be an important contributory factor.

Table 9.3.4.2 Factors related to the risk of psychiatric disorder and the form of its presentation

Nature of physical disorder (whether brain involvement)
Stage of illness (whether acute stresses involved)
Severity of illness
Degree of life threat
Psychosocial risk and protective factors in family
Age (developmental stage)
Effects of illness and treatment procedures

Whilst this dichotomy between disorders involving and not involving the brain is useful, there is little specificity in the behavioural pattern that may be attributable to intracerebral pathology. As a possible exception, children with brain dysfunction such as epilepsy or cerebral palsy may be more likely to exhibit externalizing disorders such as hyperactivity.[3] Psychiatric disorders in this group of children may be persistent, with 70 per cent still experiencing difficulties at 4-year follow-up. Overactivity, restlessness, and inattention are the best predictors of persistence.

For conditions not affecting the brain, the development of psychiatric disorder seems most likely to be linked with the accumulation of generic stress factors and family changes common to living with a chronic illness. These include life stresses such as hospitalization and daily difficulties such as specific dietary requirements and disruption of family routines.[4] A broad spectrum of psychiatric presentations are associated and these are not specific to the nature of the underlying disease processes. Children with non-neurological physical illnesses are more prone to developing emotional symptoms and eating anomalies as opposed to antisocial behaviour. Eating anomalies may arise from an emphasis on diet and a concern about poor appetite in the families of many children with chronic illnesses. Maternal anxiety may focus on feeding, especially in preschool children. The specificity of the relationship with emotional disorders is of interest. Physical illness in the child can generate family and social stresses and changes that are known risk factors for the development of emotional disorders in children. This includes mood disorders in parents and overinvolved and overprotective parenting.[2]

(b) Stage of the illness

Disorder at the time of initial diagnosis is not uncommon and is frequently short lived. In one study, 36 per cent of 8- to 13-year-olds with newly diagnosed insulin-dependent diabetes mellitus developed an adjustment disorder (most commonly dominated by depressive symptoms) within the first 3 months of diagnosis; 50 per cent had recovered within 2 months.[5] Similarly, in patients with chronic renal failure, psychological problems were reported in 60 per cent of children at the time of starting dialysis. One year later, after stabilization of their physical condition, the prevalence of disturbance was reduced to 21 per cent.[6] It is very likely therefore that in many children with chronic physical illness, psychiatric disorders are most frequently transitory adjustment disorders to stressful times in the illness.

(c) Severity of illness/degree of life threat

More severe physical disorders and those constituting a greater degree of life threat are associated with a higher risk of psychiatric disturbance. In children with end-stage chronic renal failure, those with more severe disorders (on hospital haemodialysis) have been found to have more psychiatric disorder than those not yet requiring dialysis.[7] More severely affected diabetic children and adolescents with a history of hospitalization for ketoacidosis in the previous year are more likely to exhibit psychiatric disorder than a control group of outpatients also with insulin-dependent diabetes mellitus.[8] Post-traumatic stress disorder (which by definition requires acknowledgement of perceived life threat), and high levels of post-traumatic stress symptoms have been found in children and parents up to a year after admission to Paediatric Intensive Care Units[9]; (a much higher proportion than following admission to general paediatric wards), and up to 10 years after treatment for childhood cancer.[10]

The link between illness severity and risk of psychosocial impairment may vary with the setting in which it is examined. Less severe physical impairment has been shown to be associated with a higher risk of behavioural problems in the school setting.[7] Teachers may be less aware of the presence of an underlying physical disorder in this group who have less visible physical signs, and may make less allowance for these children than for those with a more overt disorder.

(d) Psychosocial risk and intrafamilial protective factors

When a physically ill child develops psychological symptoms, these are frequently attributed by families and professionals to the presence of the illness and its stresses. It is important not to neglect consideration of other predisposing factors (i) within the child, for example genetic vulnerability, temperamental characteristics, (ii) in the family such as marital disharmony, lack of open communication, maternal mental illness affecting parenting, and (iii) within the broader social environment such as bullying at school and poor peer relationships. These factors contribute to child psychopathology in ill as well as in healthy children. Conversely, protective factors such as secure parent–child attachments, increased family social support in response to the physical diagnosis, as well as sensitive paediatric management of hospitalizations and stressful medical procedures may reduce the risk of developing psychiatric disorder.

(e) Age (developmental stage)

Manifestations of psychological distress in ill children vary with each developmental stage. Preschool children have fewer cognitive resources to cope with discomfort and stressful medical procedures and are likely to rely on maternal support and distraction to cope with illness. Between 4 and 7 years of age, children may believe that illness has been caused by something bad they have done and that they should be punished.[4] Clinginess to parents, fearfulness, sleep difficulty, and oppositional–defiant behaviour are seen in preschool children. The need for repeated painful procedures, for example with cancer chemotherapy, can lead to the development of specific needle phobia.

For school-age children, school life is a key aspect of their adjustment to illness. Return to school after cancer chemotherapy can be associated with the development of school phobia, loneliness, and social isolation. School absence and having to catch up with school work, teasing, or even bullying, especially of children who look different, may also occur and contribute to lowered self-esteem and the risk of affective disturbance. Cognitive development in adolescence allows a greater understanding about the implications of chronic illness and the realities of death; depression occurs more frequently in this age group. Adolescents may begin to challenge and experiment with their treatment; they may fail to come to outpatient appointments or attend erratically. There may also be a decline in compliance with medical advice and adherence to treatment regimens.[4] For example, diabetics may not follow dietary advice or pay reduced attention to their insulin regimen and monitoring of blood sugars leading to poorer diabetic control. Adherence may be influenced by family factors; poorer metabolic control is associated with less family cohesion and a parenting style that is perceived as critical and negative.[11] Adolescents aged 13–18 years with diabetes and co-morbid internalizing disorders, and discharged from hospital, have been found to be at greater risk of readmission up to 2 years later. This relationship was not found for younger children, suggesting that greater parental control of diabetes management (as is usual for younger children) may ameliorate the potential for psychiatric disorder to affect treatment adherence.[12]

The way in which psychiatric disorder presents may influence its perceived significance to health professionals and the likelihood of psychiatric referral. Presentations with behavioural disturbances such as screaming, struggling, panicking, or a failure to comply with treatment are more likely to precipitate referral than internalizing disorders such as depression.

Effects on parents and siblings

Whilst most families successfully adjust to the presence of a child with chronic illness in the family, this may act as a risk factor for psychological disorder. The incidence of marital break-up is not increased, but there are reports of increased marital distress. Interparental conflict may not be directly expressed but instead diverted to excessive worry and focus on the illness, which can be very stressful for the child.[13] In parallel with the heightened short-term psychological difficulties found in ill children immediately following diagnosis, a similar temporal pattern of disorder has been reported for parents and siblings. Most research has focused on mothers, who often undertake the practicalities of caring for a sick child. They may need to stop work themselves, leading to increased social isolation and a reduction in extra-familial support.[14] Fathers and mothers often cope differently with the diagnosis; mothers tend to react by emotional release, whereas fathers are more likely to withdraw and concentrate on practicalities.[14] Higher rates of maternal psychiatric treatment and negative affect have been found in families with a chronically ill child. The risk of maternal depression is greater for mothers of children with chronic as compared with newly diagnosed epilepsy; the burden of illness may impede parenting capacity and contribute to the development or maintenance of psychopathology in the children.[13] Siblings may resent both the extra attention an ill brother or sister is receiving, and repeated separations from parents during periods of hospitalization. Their psychological adjustment is related to the degree of functional impairment[15] and recent physical health of their ill sibling, the extent to which family life is disrupted by the illness, and the psychosocial support available. The need for improved communication with healthy siblings has been identified.

One disorder which highlights the complexities of interaction between living with a chronic illness and its effect on family members is AIDS. Vertical transmission from an infected mother to her unborn child has decreased in the last 10 years but there has been an increase in the number of adolescents with the virus due to survival of children with perinatally acquired HIV into adolescence in addition to adolescents acquiring the virus through other means. For many children with HIV, infection is also present in other family members, often the mother. Families have to cope with the disease itself and its treatment, the stresses of chronic illness which include an uncertain prognosis and the possibility of death as well as having to negotiate the stigma and social isolation that frequently accompany the diagnosis. Caregivers who are HIV positive themselves report poorer physical and emotional health compared with non-infected caregivers; this is associated with greater psychosocial impairment in the children—a higher risk of internalizing

problems such as anxiety and depression, more externalizing problems e.g. oppositional behaviour and poorer academic functioning. Disclosure of the diagnosis to affected children is often avoided; reasons include parental unease discussing their own HIV infection, fear of stigma, beliefs that the child is not emotionally ready to cope with the information and parents' own distress.[16] Children (aged 6–16) who are not told their diagnosis have been reported as having more internalizing problems than those informed.

Management

In the absence of rigorous treatment research in this area, the most important tenet of the psychological care of children with physical illness is based on good clinical practice, with clear and consistent communication between paediatricians, child psychiatrists and their multi-disciplinary teams. This allows early detection and intervention for psychological disorder.

Child psychiatrists frequently work closely with paediatric colleagues to assist in identifying young people at risk for psychiatric disorder, to provide assessment and treatment when indicated, and to give support and advice with regard to diagnosis and management. Many paediatric units have regular weekly psychosocial ward rounds where professionals both from within the hospital and from the community—representing paediatrics, child and adolescent psychiatry teams, social work, and education—can meet to discuss the progress of the child from each perspective.

A full psychiatric assessment involving the child and the family will be carried out in referred cases. This needs to be preceded by a careful explanation to families about the reasons why a psychiatric consultation has been sought.

Important information about premorbid concerns and the child's level of functioning may be obtained from schools, social workers, and other professionals involved with the family.

Specific psychiatric diagnoses should be treated appropriately. Children may develop acute confusional disorders associated with intracerebral infection or febrile illness. Manipulation of the ward environment to ensure: clear differentiation between night and day, that familiar toys are nearby, close family members are in attendance, and developmentally appropriate explanations are given to the child about where they are and what is happening, may help considerably. If behaviour is too difficult for staff to safely manage and is interfering with treatment, sedative medication may be needed and should be discussed with paediatricians.

Children with adjustment disorders may be helped by psychological interventions. Management may include ways of decreasing existing stresses or helping individuals to adjust to them. Possible interventions include supportive counselling, individual therapy using cognitive behavioural principles, and family therapy.

When there is a chronically ill child in the family, parents often find it difficult to maintain the usual boundaries. For example, disciplining an ill child may be associated with parental guilt; this can lead to increasing anxiety for children who exhibit increasingly oppositional behaviour in an effort to test the boundary limits. Discussion regarding parenting techniques in the context of these feelings may be helpful. Parents also tend to increase their protective responses to ill children and show more overinvolved parenting. If excessive it may impede the child's development, but to a modest degree it may be helpful and advantageous.

Systematic desensitization together with relaxation and distraction techniques may be used to treat a specific needle phobia. This needs to be carried out in collaboration with ward staff taking account of associated psychopathology, for example, oppositional behaviour, a generalized anxiety state, or an adjustment reaction. Treatment of the associated problems can often obviate the need for direct phobic treatment. When indicated, the latter's success is likely to be dependent on external changes that reduce anticipatory anxiety. These might include minimizing the time the child needs to wait for treatment and ensuring that more experienced members of the medical team are responsible for cannula insertion.

Generalized symptoms of anxiety are not uncommon in parents and children and may be manifested in different ways, for example, a young child may resume bed-wetting, a school-age child may become intensely distressed by being away from his parents, adolescents may experience difficulties sleeping, and anxious parents may become agitated with ward staff. Regular explanations from staff about the child's condition and treatment may help to alleviate this anxiety. Communication difficulties within the family may contribute to anxiety and be helped by family meetings where difficulties can be shared. Relaxation and distraction techniques together with cognitive behavioural interventions may also be of benefit. If symptoms are intense and interfering with physical treatment, anxiolytic medication may be indicated.

Antidepressant medication may be considered for children and adolescents with a depressive episode. This should be discussed with the medical team to minimize drug interactions and side-effects that may exacerbate the physical condition of the patient.

Treating children with severe illness who may be receiving distressing and painful treatment can arouse intense emotions in the most experienced of paediatric staff. Regular meetings with mental health professionals may help them to process some of these feelings and prevent them impeding patient care.

Prognosis of psychiatric disorder in children with chronic physical illness

Many of the psychological difficulties experienced by chronically ill children are short lived and do not continue into adult life. Overall, studies indicate that psychiatric outcome is not severely compromised in the majority of adult survivors. Persistence of disorder is related to the severity of childhood psychological symptoms (the more severe being more likely to last), persistence of physical symptoms into adulthood,[17] and to the presence of physical disorder affecting the brain (Table 9.3.4.3).

The form of psychiatric symptomatology in childhood and adulthood may also be different; for example, cystic fibrosis sufferers, aged 8–15, have been found to report more eating related symptoms whereas symptoms of anxiety and depression are more prevalent in the adult group. With regard to cystic fibrosis, a consistent association between disease severity in adulthood and psychiatric disorder has not been shown. However, increased disease severity in childhood is associated with lower educational attainment;

Table 9.3.4.3 Factors associated with persistence of psychosocial dysfunction into adulthood

Severity of childhood symptoms
Persistence of *physical* symptoms into adulthood
Physical disorder involving the brain

in adulthood, employment is associated with both higher academic achievement and less depressive symptoms.[18]

Many studies suggest that by adulthood, most survivors of childhood cancer are indistinguishable from the general population with regard to psychosocial outcome. However, more detailed analysis suggests that factors such as age at diagnosis, site of the tumour, and nature of treatment (e.g. cranial irradiation) may influence cognitive and psychological outcome. For children and adolescents up to age 18 diagnosed with brain tumours, cognitive deficits and psychosocial problems increased with age and time since diagnosis.[19] As survival has increased, adults are exposed to the chronic toxic effects of treatment such as endocrine abnormalities, cardiac or pulmonary problems, and infertility. Follow-up of childhood cancer survivors, to a mean age of 28, revealed that current physical functioning, including pain, was associated with suicidality even after accounting for treatment and depression variables. Younger age at diagnosis, longer time since diagnosis, and cranial irradiation were also important risk factors.[20] Survivors of acute lymphoblastic leukaemia and Wilms' tumour did not show increased psychopathology as adults but had more difficulties with interpersonal functioning and day-to-day coping.[21]

Individuals with intracerebral pathology maintain high levels of disorder in adulthood, especially with regard to behaviour and social isolation. By contrast, patients with congenital heart disease surgically corrected in childhood, are not at increased risk of psychiatric disorder as adults.

Although young adults with end-stage chronic renal failure report more episodes of psychiatric disturbance than healthy matched controls before 17 years, they do not necessarily have increased psychopathology in late adolescence and adulthood. In common with survivors of other chronic childhood disorders,[17] the majority of adult renal patients are reported as functioning well socially, but they are more likely than age-matched controls to be living with their parents, to have less school qualifications, higher rates of unemployment, and fewer intimate relationships outside the family.

Care of the dying child

Children at different developmental stages differ in their understanding of death. They gradually acquire components of the death concept; between 9 and 11 years of age, most children have reached a full understanding, acknowledging that it is permanent, inevitable, and universal. However, experience of serious illness and death interacts with the stage of understanding, so that children aged 5 or younger may have a more mature understanding and exhibit symptoms of anxiety about death. There is evidence that even young children with terminal illness are aware that they are dying, although they may not tell anyone that they know.[22]

Parents (and professionals) often find it difficult to talk about death with children. This is likely to interfere with coping for the whole family. Families with an open pattern of communication do better psychologically.[22] Mental health professionals may have a role in facilitating this discourse, promoting parents' confidence, and competence in communicating with their children. This will help the whole family to begin the process of mourning.

Children need information, reassurance, an opportunity to express their feelings, and adults with whom they can do so. As children lack the vocabulary of adults they may often exhibit their distress by behavioural changes, for example, bed-wetting, difficulty sleeping, and school refusal. Children and their siblings faced with death need clear, simple, and truthful explanations. They should not be pushed to talk, nor frightened with excessive medical detail.

Dying and grieving lead to a whole range of distressing feelings. This is part of a normal process, and mental health professionals can help their colleagues and families to acknowledge that this upset is acceptable.

Bereaved children frequently model their grief experience on what they perceive as being acceptable in the family, and an overt denial of upset by parents may lead to psychological difficulties in the child. The issue of whether to involve siblings after the death of the child in funerals or graveside visits often arises. If children are prepared for what to expect, involvement can be helpful in enabling them to acknowledge that a change has taken place and other people are feeling as sad as they are.[22]

Mothers are involved in nursing and caring for their dying children. They report an excess of depression, problems of helplessness, and a fear of being unable to cope with the child dying. Parents may feel that they can never fully recover from the loss of a child. Fathers tend to report more difficulties with feeling left out of the ill child's life and then with worry about their spouse being too preoccupied with the dead child. The effects of a child's death on family life can be traced even years after the death.[22] Formal follow-up after bereavement may help to identify those families and individuals experiencing psychological reactions that may benefit from more intensive support.

The effects of psychiatric disorder on the course and outcome of physical illness

Psychological disorder, as well as being a consequence of both acute and chronic physical ill-health, may also have an impact on the course of physical illness.

An increasingly recognized disorder in this respect is post-traumatic stress disorder. Sudden physical trauma, such as burns and road traffic accidents, are examples of antecedents. Victims of road traffic accidents between 5 and 18 years of age, particularly those who experience high levels of distress immediately after the accident, are at greatest risk of exhibiting post-traumatic stress symptoms 3 months later.[23] Such responses may be contributed to not only by the accident itself but also by the medical procedures that take place on arrival in hospital. Surgical collars, intravenous infusions, and monitoring equipment can be associated with intense fear.[23] Children with acute severe sepsis such as meningococcal disease admitted to paediatric intensive care units are also at risk of developing similar symptoms.[24] In turn, these reactions can have an effect on the child's ability to co-operate with future hospital attendance, medical, and surgical interventions. Stress reactions may be ameliorated, to some extent, by the provision of age-appropriate information about what has happened and what is going to happen.

The diagnosis and treatment of such disorders may be impeded by the fact that follow-up for young people may not be at the admitting hospital. Burns units and paediatric intensive care facilities are often at tertiary centres some distance from the patient's home. General practitioners and local paediatricians have a role to play in assessing how the family is coping, specifically regarding

symptoms of post-traumatic stress. Child psychiatry involvement may be appropriate if psychological treatment is required both during admission and at follow-up. Cognitive behavioural interventions with individual children and families may be used to alleviate symptoms.

Affective symptoms, particularly depression and anxiety, are not uncommon following an acute medical admission and may interfere with physical treatment. For example, an adolescent admitted to paediatric intensive care with meningococcal disease requiring a partial limb amputation could develop a depressive disorder. Symptoms of despair and hopelessness coupled with a lack of interest and energy may impede the physiotherapy programme, delay mobilization, and hospital discharge. Paediatric staff need to be alert to such potential sequelae and to have child psychiatric colleagues readily available for assessment and treatment.

Exacerbation of chronic illnesses such as asthma can be precipitated by emotional disturbance; adolescents aged 11–17 with anxiety or depressive disorders reported more asthma symptoms in the previous 2 weeks than young people without these affective diagnoses.[25] Adjunctive psychological treatments such as family therapy have been shown to lead to an objective improvement in airways disease,[26] compliance, and reduced hospital admissions.[27]

Somatization and somatoform disorders

Disorders presenting with functional physical symptoms and somatization

'Functional' somatic symptoms with no obvious organic explanation are frequent in childhood. Children have a limited vocabulary for expressing their emotions and often communicate their distress by means of physical symptoms. Somatization refers to this process. In some cases these symptoms become persistent with associated functional impairment; this may lead to consultation. The definitions of somatization disorder (one of the somatoform disorders) used in ICD-10 and DSM-IV are too stringent for children (in that diagnosis requires multiple physical symptoms over years). Other disorders (namely somatoform pain disorder, dissociative/conversion disorder, and neurasthenia) are seen in children and adolescents. The risk factors for somatization in this population are shown in Table 9.3.4.4.

Table 9.3.4.4 Risk factors for somatization in children and adolescents

Individual:	Personal experience of physical illness Enhanced focus on physical sensations Somatic attributions Conscientious, vulnerable, sensitive, anxious personalities with particular concerns about peer relationships High achievement orientation
Family:	Physical health problems Psychiatric problems Parental somatization Emotional overinvolvement Limitations in the ability to communicate about emotional issues
Environment:	Life stresses e.g. school, teasing or bullying, academic pressure

Aches and pains and somatoform pain disorder

Aches and pains (often abdominal pains and headaches) are a common manifestation in young children. Between 2 and 10 per cent of children in the general population have problems in this area. Mothers assess the child's symptomatology with specific regard to whether the child is 'pretending', 'upset', or 'ill' and generally respond appropriately. They recognize that children may experience symptoms as a result of stress or use them to avoid something they find difficult.

Abdominal pain commonly leads to a general practitioner consultation and may account for 10 per cent of new appointments with paediatricians. In only a few of these cases is serious organic pathology found. Lack of identifiable organic pathology does not imply a psychogenic aetiology. The latter is rather supported by evidence that psychological events influence the symptoms.

Children who somatize tend to have a family history of physical ill-health and parental illness. In some cases there are also psychosocial difficulties in the family. There is an association with stressful life events. Co-morbid internalizing disorders (depression and anxiety) are commonly present.[28]

In adolescence, headaches become a prominent symptom, peaking in prevalence at 12 years of age. As with abdominal pain in younger children, they are frequently preceded by physical or psychological precipitants, such as academic or social stresses in school or difficulties at home. Headaches lead to absence from school but are not associated with underachievement. A family history of migraine is common.

As defined in ICD-10, in persistent somatoform pain disorder, severe distressing pain occurs in association with emotional conflict or psychosocial problems that are sufficient to allow the conclusion to be drawn that they are the main causative influences. The result is usually a marked increase in support and attention, either personal or medical.

In trying to best manage severely affected children, close collaboration between paediatricians and child psychiatrists is helpful. The lack of demonstrable organic pathology should be communicated and professionals should help the family to make the link between physical symptoms and psychological precipitants with the help of a written diary if necessary. It is important to reduce the attention given to the physical symptoms in order to decrease the resulting functional handicap. Early return to school together with the resumption of normal activities should be encouraged. The short-term prognosis for presentations to medical services is good, with 75 per cent of children recovering within several months.

Dissociative/conversion disorder

Children and adolescents present as if having a physical disorder affecting voluntary motor or sensory functioning, although none can be found; the symptoms correspond to the patient's idea of physical disorder, which may not coincide with physiological or anatomical principles. Aetiologically, the disorder is believed to arise largely unconsciously and to represent an escape from an unbearable personal conflict. It usually manifests in adolescence and is more common in girls. The most common presentation is neurological with disturbance of motor function such as weakness of legs, paralysis of a limb, or bizarre gait. Multiple symptoms often occur.

Premorbid psychopathology in the child and family are often absent, although perfectionistic and conscientious traits with concerns about academic performance and a child and family focus on

high achievement have been noted. Overconcern with physical health and illness often characterize these families; frequently there is a family history of physical health problems. Families often present as being close, but communication, particularly regarding emotions, may be limited.

It is assumed children develop conversion disorders as an unconscious means of escaping a situation with which they cannot cope. This includes intolerably high academic expectations (often the child's own), unresolved family conflict, and, in a minority of cases, sexual abuse. The disorder is often precipitated by minor physical illness and may also occur in children with identified organic pathology, for instance the development of pseudoseizures in an individual with epilepsy.

The majority of these patients are managed by paediatricians. After investigations have excluded organic pathology and a psychogenic contribution is suspected, this needs to be communicated to the family. The shift from physical to psychological factors may be difficult for the family to accept; information may need to be conveyed slowly 'at a pace the family can cope with'. A collaborative approach between paediatricians and child psychiatrists is important. A persistent focus on physical aetiology may be unhelpful but in management it is more useful to focus on the handicap caused by specific symptoms rather than on their cause by introducing a programme of rehabilitation and physiotherapy directed at these features, including school attendance. Psychotherapeutic work, both individually and with the family, may help the family understand the factors maintaining the child's symptoms and explore any identified stressors or conflicts. Time can also be spent helping families to consider alternative strategies they may use to cope with future conflicts.

Chronic fatigue syndrome (neurasthenia in ICD-10)

This is operationally defined as disabling physical fatigue of over 6 months' duration, unexplained by primary physical or psychiatric causes. There are often other unexplained somatic symptoms and a strong belief by the patient and their family that the aetiology is physical.[29] It might be considered as one of the somatoform disorders, as similarities are shared with regard to aetiology and management.

There is no firm evidence that chronic fatigue syndrome results from a specific viral infection but a physical illness is often the precipitating factor. There is frequently a family history of physical illness and a preoccupation with physical symptoms. Parents invariably attribute the symptoms to an organic aetiology. Children tend to be described as high achieving and perfectionistic, as well as sensitive, vulnerable and anxiety-prone; onset can be temporally related to transitions at school, for example transfer to secondary school. Depressive mood changes are common and on assessment depressive disorder is found in one-third of cases.[29]

Chronic fatigue syndrome can be extremely disabling. A self-perpetuating cycle is set up whereby fatigue and the resultant inactivity lead to loss of muscle bulk and deterioration in physical fitness. Activity becomes increasingly difficult and is avoided, leading to a further deterioration in physical ability. An essential focus of treatment is to disrupt this cycle.

Management is often multi-disciplinary including paediatricians, physiotherapists, school teachers, and child psychiatrists. A clear explanation of the results of physical investigations and the fact that no serious organic pathology has been found is important. Focusing on improving symptoms, as opposed to debating aetiology, is most helpful. Helping the family to shift from a purely physical model to one that includes psychological factors in maintaining symptoms may be difficult and needs to be negotiated slowly. In particular, enabling them to see the disorder as an interaction between physical, social, and emotional factors can be useful.

Treatment ingredients usually include a graded exercise programme, a progressive return to school, and work with the family to facilitate engagement and address factors that may be impeding recovery. Antidepressants may be useful for co-morbid depressive disorder. There is an evidence base for both cognitive behaviour therapy (CBT) and graded exercise in adults and one study which demonstrates the effectiveness of CBT in adolescents.[30]

Treatment of functional symptoms and somatoform disorders in children and adolescents: research evidence

There are few satisfactory controlled studies on the effects of treatment for childhood somatization. Most work on children with somatoform disorders has been based on small groups with severe problems, and the management advice outlined above is derived from the conclusions of experienced clinicians and open-treatment case reports. However, there is some evidence from controlled studies indicating the efficacy of a cognitive behavioural family intervention for recurrent abdominal pains in children.[31] For hospitalized children with severe, recurrent abdominal pain, parental attribution of symptoms to psychological factors facilitated resolution.[32] The superiority of relaxation training over placebo in reducing migraine attacks has been shown.

Outcome of functional somatic symptoms and somatoform disorders

Of adults with a childhood history of abdominal pains, 50 per cent have recurrent symptoms in adult life despite a pain-free period in adolescence; they also have an increased prevalence of psychiatric disorders (particularly anxiety disorders).[33] Childhood conversion disorder is generally associated with a good outcome; recovery is usually complete by 3 months. When children (9–16 years) were followed up 4 years after a conversion disorder from which 85 per cent had recovered, 35 per cent had a mood or anxiety disorder; affective disorder was higher (100 per cent) in the minority who had failed to recover as compared with 23 per cent in those who were better.[34] In young people with chronic fatigue syndrome, recovery or marked improvement in symptoms can be expected in 50–75 per cent of cases in the short to medium term. School non-attendance may be over a year and time to full recovery 3 years or more. There are indications of an increased likelihood for the development of psychiatric disorder after recovery.[35]

Concluding remarks

Somatic and psychological symptoms are intimately linked. Changes in physical health can affect psychiatric outcome. Conversely, emotional distress may affect adherence to treatment, and it is sometimes expressed through physical symptoms. Awareness of this interplay is important and should be mirrored by a close working relationship and close communication between paediatricians and child psychiatrists.

Further information

Shaw, R.J. and DeMaso, D.R. (2006). *Clinical manual of pediatric psychosomatic medicine*. American Psychiatric Publishing Inc.

Rutter, M., Bishop, D., Pine, D., *et al.* (eds.) (2008). *Rutter's child and adolescent psychiatry* (5th edn). Blackwell Publishing, Oxford.

References

1. Cadman, D., Boyle, M., Szatmari, P., *et al.* (1987). Chronic illness, disability and mental and social well being: findings of the Ontario child health study. *Pediatrics*, **79**, 805–12.
2. Garralda, M.E. and Palanca, M.I. (1994). Psychiatric adjustment in children with physical illness. *British Journal of Hospital Medicine*, **52**, 230–4.
3. Rutter, M., Tizard, J., and Whitmore, K. (eds.) (1970). *Education, health and behaviour*. Longmans, London.
4. Eiser, C. (1993). *Growing up with a chronic disease. The impact on children and their families*. Jessica Kingsley, London.
5. Kovacs, M., Feinberg, T.L., Paulauskas, S., *et al.* (1985). Initial coping responses and psychosocial characteristics of children with insulin-dependent diabetes mellitus. *Journal of Pediatrics*, **106**, 827–34.
6. Wass, V.J., Barratt, T.M., Howarth, R.V., *et al.* (1977). Home dialysis in children. *Lancet*, **1**, 242–6.
7. Garralda, M.E., Jameson, R.A., Reynolds, J.M., *et al.* (1988). Psychiatric adjustment in children with chronic renal failure. *Journal of Child Psychology and Psychiatry*, **29**, 79–90.
8. Liss, D.S., Waller, D.A., Kennard, B.D., *et al.* (1998). Psychiatric illness and family support in children and adolescents with diabetic ketoacidosis: a controlled study. *Journal of the American Academy of Child and Adolescent Psychiatry*, **37**, 536–44.
9. Rees, G., Gledhill, J., Garralda, M.E., *et al.* (2004). Psychiatric outcome following paediatric intensive care unit (PICU) admission: a cohort study. *Intensive Care Medicine*, **30**, 1607–14.
10. Kazak, A.E., Alderfer, M., Rourke, M.T., *et al.* (2004). Posttraumatic stress disorder (PTSD) and posttraumatic stress symptoms (PTSS) in families of adolescent childhood cancer survivors. *Journal of Pediatric Psychology*, **29**, 211–19.
11. Fiese, B.H. and Everhart, R.S. (2006). Medical adherence and childhood chronic illness: family daily management skills and emotional climate as emerging contributors. *Current Opinion in Pediatrics*, **18**, 551–7.
12. Garrison, M.M., Katon, W.J., and Richardson, L.P. (2005). The impact of psychiatric comorbidities on readmissions for diabetes in youth, *Diabetes Care*, **28**, 2150–4.
13. Rodenburg, R., Meijer, A.M., Dekovic, M., *et al.* (2005). Family factors and psychopathology in children with epilepsy: a literature review. *Epilepsy and Behaviour*, **6**, 488–503.
14. Mastroyannopoulou, K., Stallard, P., Lewis, M., *et al.* (1997). The impact of childhood non-malignant life-threatening illness on parents: gender differences and predictors of parental adjustment. *Journal of Child Psychology and Psychiatry*, **38**, 823–9.
15. Sharpe, D. and Rossiter, L. (2002). Siblings of children with a chronic illness: a meta-analysis. *Journal of Pediatric Psychology*, **27**, 699–710.
16. Steele, R.G., Nelson, T.D., and Cole, B.P. (2007). Psychosocial functioning of children with AIDS and HIV infection: review of the literature from a socioecological framework. *Journal of Developmental and Behavioural Pediatrics*, **28**, 58–69.
17. Pless, I.B., Cripps, H.A., Davies, J.M.C., *et al.* (1989). Chronic physical illness in childhood: psychological and social effects in adolescence and adult life. *Developmental Medicine and Child Neurology*, **31**, 746–55.
18. Burker, E.J., Sedway, J., and Carone, S. (2004). Psychological and educational factors: better predictors of work status than FEV_1 in adults with cystic fibrosis. *Pediatric Pulmonology*, **38**, 413–18.
19. Poggi, G., Liscio, M., Galbiati, S., *et al.* (2005). Brain tumours in children and adolescents: cognitive and psychological disorders at different ages. *Psycho-oncology*, **14**, 386–95.
20. Recklitis, C.J., Lockwood, R.A., Rothwell, M.A., *et al.* (2006). Suicidal ideation and attempts in adult survivors of childhood cancer. *Journal of Clinical Oncology*, **24**, 3852–7.
21. Mackie, E., Hill, J., Kondryn, H., *et al.* (2000). Adult psychosocial outcomes in long-term survivors of acute lymphoblastic leukaemia and Wilms' tumour: a controlled study. *Lancet*, **355**, 1310–14.
22. Black, D. (1998). Coping with loss. The dying child. *British Medical Journal*, **316**, 1376–8.
23. Di Gallo, A., Barton, J., and Parry-Jones, W.L. (1997). Road traffic accidents: early psychological consequences in children and adolescents. *The British Journal of Psychiatry*, **170**, 358–62.
24. Shears, D., Nadel, S., Gledhill, J., *et al.* (2005). Short-term psychiatric adjustment of children and their parents following meningococcal disease. *Pediatric Critical Care Medicine*, **6**, 39–43.
25. Richardson, L.P., Lozano, P., Russo, J., *et al.* (2006). Asthma symptom burden: relationship to asthma severity and anxiety and depression symptoms. *Pediatrics*, **118**, 1042–51.
26. Gustaffson, P.A., Kjellman, N.I.M., and Cederblad, M. (1986). Family therapy in the treatment of severe childhood asthma. *Journal of Psychosomatic Research*, **30**, 369–74.
27. Godding, V., Kruth, M., and Jamart, J. (1997). Joint consultation for high-risk asthmatic children and their families, with paediatrician and child psychiatrist as co-therapists: model and evaluation. *Family Process*, **36**, 265–80.
28. Campo, J.V., Bridge, J., Ehmann, M., *et al.* (2004). Recurrent abdominal pain, anxiety and depression in primary care. *Pediatrics*, **113**, 817–24.
29. Garralda, M.E. (1996). Somatisation in children. *Journal of Child Psychology and Psychiatry*, **37**, 13–33.
30. Stulemeijer, M., de Jong, L.W.A.M., Fiselier, T.J.W., *et al.* (2005). Cognitive behaviour therapy for adolescents with chronic fatigue syndrome: randomized controlled trial. *British Medical Journal*, **330**, 14.
31. Robins, P.M., Smith, S.M., Glutting, J.J., *et al.* (2005). A randomized controlled trial of a cognitive-behavioural family intervention for pediatric recurrent abdominal pain. *Journal of Pediatric Psychology*, **30**, 397–408.
32. Crushell, E., Rowland, M., Doherty, M., *et al.* (2003). Importance of parental conceptual model of illness in severe recurrent abdominal pain. *Pediatrics*, **112**, 1368–72.
33. Plunkett, A. and Beattie, R.M. (2005). Recurrent abdominal pain in childhood. *Journal of the Royal Society of Medicine*, **98**, 101–6.
34. Pehlivantürk, B. and Unal, F. (2002). Conversion disorder in children and adolescents. A 4 year follow-up study. *Journal of Psychosomatic Research*, **52**, 187–91.
35. Garralda, M.E. and Chalder, T. (2005). Practitioner review: chronic fatigue syndrome in childhood. *Journal of Child Psychology and Psychiatry*, **46**, 1143–51.

9.3.5 The effects on child and adult mental health of adoption and foster care

June Thoburn

Introduction: mapping the terrain

Adoption and foster care are important 'solutions' to identified problems or risks, but potentially they are also contributors to problem behaviours or emotional difficulties. In their problem-solving role, they are seen as potential solutions, not only to actual

or future mental health problems of children, but also to the adverse effects of involuntary childlessness.

This chapter concentrates on the impact of adoption and foster care on the children placed, but their role in problem solution or problem generation for adults is also touched on. Adoption is more often than not a satisfactory way of meeting the need to become parents for those childless couples who succeed in having a child placed with them (a tiny minority of the involuntary childless). It is very rarely a solution to the problems of a parent who gives up a child for adoption whether voluntarily or involuntarily. Studies of adults who relinquished children indicate that the reaction to the loss of their child may be associated with moderate distress or may lead to a long-term grief reaction, which in turn will potentially harm children subsequently born to that parent. One must also note that some parents who lose a child to adoption or foster care are themselves children, sometimes not yet in their teens, whose needs are often overlooked in the interests of providing for the infant.

Fostering and adoption started as very similar processes, diverged in Europe and North America in the first half of this century, and are now much closer together again. The 'total severance' model of legal adoption—the type that most people in Europe, the United States, and Australasia immediately recognize—has a short history. In the United Kingdom it was not until the passing of the 1958 Adoption Act that secrecy became the norm. The 'sealing' of birth information started in the United States around 1948 but it was not until 1991 that Alabama 'sealed' its adoption records.[1] This experiment of totally closed adoption was short-lived, and many countries have introduced legislation to allow adult adoptees and/or birth relatives to access identifying information that allows them to seek each other out.[2] 'Open' adoptions, in which some degree of contact between the adopters, the birth parents, and the children is maintained after placement, are increasingly common.

As countries have become richer, the need to place children for adoption has diminished and the number of infants placed at the request of their parents has fallen well short of the 'demand' of those wishing to start a family through adoption. In consequence, it has been possible to encourage potential adopters to 'stretch' their notions of parenthood, and to place older children, those with disabilities, and those with behavioural or emotional problems with adoptive parents as well as with foster parents. The main remaining difference between adoption and foster care is that the majority of children placed in foster homes live there for comparatively short periods before returning to their families of origin. They are best seen as supplementary rather than substitute parents, although in all 'first world' countries long-term or 'permanent' foster care is an important option for a minority of those entering public care, especially in those countries (the majority) who rarely use adoption as a route out of care.

Adoption and foster care will impact on the mental health of the children in different ways, which may be considered along six main dimensions (see Box 9.3.5.1). The dimensions interact differently for different children. An infant placed from an Asian country might be adopted by childless relatives in Europe and have had a positive early experience of parenting, or might have experienced very adverse early nurturing and be adopted by strangers of a different ethnic origin. The child placed at six may have had good care from one or both parents until some traumatic event led to the need for an adoptive placement, or the child may have been seriously maltreated and had several placements before finally joining a substitute family.

The nature of the evidence on the impact of adoption and fostering on mental health

The actual or potential problems most obviously associated with child placement are those resulting from separation and loss. Brodzinsky *et al.*[3] have made significant contributions to our understanding of the psychology of adoption. Put simply:

> . . . for later-placed children, the loss of family or surrogate family connections is overt, often acute, and sometimes traumatic. In contrast, for children placed as infants, loss is, of necessity, more covert, emerging slowly as the youngster begins to understand the magnitude of what has happened. . . . In addition, there may be loss of a clear sense of genealogical connections and, in the case of transracial and inter-country adoption, loss of cultural, ethnic, and racial ties.

The impact of loss will also vary with the child's temperament, and the work of Rutter,[4] and of others who have written on 'resilience', are important sources. A 'born worrier' will go through life wondering what there was about him or her that was not worth keeping, and no amount of positive parenting will make this angst go away; a resilient child will shrug away the past and make the best of even not particularly good parenting by the substitute parents.

It is important, before considering the research findings, to take note of the limitations of our knowledge on the long-term outcomes of foster care and adoption. Turning first to the characteristics of the children, studies of family placement often include both infants and older children, those with emotional difficulties and those without. Some studies of foster care include children placed temporarily alongside others placed permanently, and in some US studies the term 'foster care' includes all children in out-of-home placements (for family placements the term 'foster family care' is used).

At the other end of the process a broad range of 'outcome' measures is used[5] and 'success' rates vary depending on the measures used and the length of time between placement and reported outcome. The well-being of the young adult (using a range of standardized instruments) is the most reliable outcome measure but more often 'output' measures are used. (Was the child placed? Was legal adoption completed? Was a satisfactory reunion with the birth parents achieved during childhood or as an adult?). Measures of satisfaction of the different members of the adoptive family are also used. Unsurprisingly, therefore, reported 'success' rates have varied between below 50 per cent and around 95 per cent.

Box 9.3.5.1 Dimensions of family placement

- The age of the child at placement.
- The degree of disturbance of the child prior to placement.
- The nature of attachments with birth family members and short-term foster carers.
- (For those in foster care) the duration of placements, the frequency with which they occur, and whether the child returns to the same foster carers on each occasion.
- (In the case of adoption) whether the child is adopted within the family (step-parent or relative adoptions), by foster carers to whom the child is already attached, or by parents not known previously (stranger adoption).
- Whether or not the child is adopted or fostered by parents of the same cultural and ethnic background or country of origin.

The placement process that researchers seek to evaluate is extremely complex. When, as with adoption or permanent fostering, the aim is for the child's life chances to be improved by their becoming fully a part of the new family, it becomes impossible to unpick the very many variables that will have had an impact on the mental health of the young person between placement at 6 weeks and maturity at around 26. (There is some evidence that adopted people move towards emotional maturity at a slower pace—not surprisingly with at least two extra hurdles to surmount: that of separation and loss, and that of making sense of their adoptive identity). In longitudinal studies, if numbers are large enough, it is possible to control for the major variables such as age at placement, disability, and emotional or behavioural problems at the time of placement. However, the many aspects of parenting, and the nature of any therapeutic input may all have had an impact on the placement. The researchers may seek the opinions of parents and children as to what they found helpful, but clear causal relationships between outcomes and variables such as parenting styles, models of social work practice, and therapy cannot be claimed.

In summary, whilst researchers have, for many years, sought to bring academic rigour to their studies, family placement remains an 'untidy' subject. The more complex the placement circumstances and the longer the timescale, the more difficult it is to attribute success to any one factor, type of placement, or model of intervention.

A review of the research evidence on outcomes

The above section explains why, although there are some random controlled trials of treatment approaches and of short-term foster care models, the literature contains more research syntheses of the different aspects of family placement[5–10] than 'classical' systematic reviews. The findings from the large volume of quantitative and qualitative research will be summarized under the broad headings of time-limited foster care placements and placements made with the intention that the child will become a full part of the adoptive or foster family. The emphasis will be on the second group, which will be further subdivided into placements of infants and placements of older children.

Time-limited placements

In general terms, short-term foster care is used along with other services in an attempt to improve family functioning so that the child may benefit from increased stability in the family home or as a short-term crisis intervention measure. The aims of short-term fostering can be summarized as: temporary care; emergency care; assessment; treatment and 'bridging'—to independence or between placements following placement break down.[7]

Generally short-term placements used as part of family support are successful in that few placements actually break down and most parents express satisfaction with the service. This is especially so if the placement follows careful preparation for the child, the birth parents, and the foster parents and if those who need a series of placements return to the same foster family. Several UK researchers have found that a 'keep them out of care at all costs' attitude tends to prevail in child welfare agencies, thus leading to too many ill-planned and ill-matched emergency placements, which in turn lead to placement break down and to unnecessary moves in care.

Testa and Rolock[11] conclude broadly positively from an overview of treatment foster care research in the United States, and Fisher and Chamberlain[12] report better outcomes for very troubled children in multi-systemic treatment foster care than for a 'service as usual' group. (These approaches involve placement with specially recruited, trained, and financially rewarded foster carers on a time-limited basis. Intensive multi-agency support is provided to the parents, foster carers, and children.) Though placement stability remains a problem, behavioural improvements are reported and these schemes are well rated by most of the young people and their foster carers. Some researchers report a problem of 'overstaying', but this should perhaps be reframed as a success, in that some young people settle in so well that, against the odds, the task-centred foster family becomes a 'secure base' and the foster parents continue to provide support to the young people as they move into adult life.

Associations have been found in some studies between positive child outcomes and practitioners who facilitate good contact between the birth parents, foster carers, and the child; provide support to the foster carers and the birth parents; and take a multi-agency approach to treatment of the child and parents before, during, and after placement.

Adoption and long-term foster-family placement

Whilst, in the United Kingdom and North America, adoption is considered to be the major placement option for most young children who cannot remain with their birth parents, opinion is divided (often along country lines) as to the importance of long-term foster family care as a placement of choice. Practice also varies in different countries in respect of placement with relatives. In most countries it is the exception rather than the rule for relatives to adopt (foster care, guardianship, or informal arrangements being preferred) whereas in the United States legal adoption by relatives tends to be encouraged.[13]

(a) Outcomes for children placed as infants

The largest volume of research on the long-term outcome of adoption concerns children placed 'voluntarily' as infants. However, inevitably, the practice referred to in these studies is already dated by the time the long-term outcomes can be measured some 20 years or so after the child was placed. Although some may have been born to mothers who had poor antenatal care, few of these early-placed children will have experienced neglect or maltreatment. However, with the growth of inter-country adoptions, studies of infants placed more recently are more likely to include substantial numbers of children who have experienced adverse conditions during their early months. It is likely that disruption rates will be higher than they have been in the past.

(b) The impact of placement in the short-term

An important source of detailed information on short-term outcomes of infant placements is the longitudinal study by Rutter *et al.*[14] which compares young Romanian children placed with British families with a cohort of English infants placed in 'stranger' adoptive families. Reactions to placement of the English infants who had generally good postnatal care are predictable in the light of knowledge about child development, attachment, separation, and loss. Those placed quickly settle with no obvious signs of stress; those with adverse early experiences including institutionalization (most of the Romanian infants) also appear to settle well if placed in their early months. Those placed when older than 6 months are more likely to show stress reactions at the loss of a carer to whom they are beginning to be attached, or to show

adverse reactions resulting from early maltreatment, neglect, or institutionalization.

(c) Signs of stress during childhood

The more robust studies of the mental health of adopted infants in their middle years and early adolescence are those that prospectively follow them as they grow-up. The conclusion drawn from these studies is that children placed with substitute families as infants tend to do better at each stage than non-adopted peers living in the generally adverse environments in which the children were likely to have lived had they remained with their birth parents.

All studies have found that, even for those with poor antenatal and birth history or who experience adverse circumstances in their early months, subsequent physical and cognitive development is generally good. However, children adopted as infants appear to be at a slightly higher risk of experiencing problems in their social, emotional, and behavioural development compared with other children raised in similar socio-economic circumstances. This is particularly the case with adopted boys. Information from longitudinal studies is supplemented by studies of clinical populations, such as those whose parents seek psychiatric help for them. Rates of maladjustment appear to be higher around the age of 11, and decrease as the children move into later adolescence. Some studies report that adopted children are more vulnerable on some measures of behavioural and emotional development than others, including an inability to settle, restlessness, a tendency to lie or fantasize, and difficulties in getting on with their peers and teachers. Low self-esteem and feelings of insecurity are also more likely to be present amongst children in their middle years and adolescence.

(d) Long-term outcomes

There is a lack of recently published quantitative studies of the well-being of adults adopted as infants. Summaries of the research[3,6,7] report that few of those placed as infants (around 5 per cent) will leave their adoptive families before the age of 18, in circumstances of conflict, which can be described as 'adoption break downs'. Qualitative studies have reported that around 80 per cent of both adopters and adoptees express broad satisfaction with the growing-up experience. Howe[15] uses in-depth interviews with the parents of adult adoptees to analyse the mental health problems that have persisted into adulthood and reports that, when more serious problems do emerge, the issue of adoptive identity often underlies a range of presenting symptoms. Also, amongst the over 80 per cent of adults who are generally satisfied with the experience of growing-up adopted are some who continuously or episodically have a sense of unease around questions of identity and the reasons why their birth parents 'gave them up' for adoption.[2] The most authoritative recent research on long-term outcomes is the Swedish cohort study of Lindblad *et al.*[16] These authors compared population data on nearly 6000 inter-country adopted adults (mostly placed when under the age of 5) with their non-adopted peers. They note that whilst there were more similarities than differences, the adopted children were more likely than peers brought up in similar circumstances to have psychiatric problems, including substance abuse, and there was a higher suicide rate.

In summary, when well-being and mental ill-health are the outcome measure used, adults who were adopted as infants tend to be healthier, have higher IQ scores, lower rates of criminal behaviour, and fewer psychiatric symptoms than non-adopted peers from similar backgrounds to those into which they were born and to be broadly similar to those brought up by birth parents living in similar circumstances to the adopters. However, the larger scale studies that allow for the control of the many intervening variables tend to lack detail on the children's experiences of family life and of any therapeutic interventions. It is therefore unclear whether any differences can be associated with adoption *per se* or with the more advantaged home circumstances of the adopted children.

(e) Children placed when past infancy

Researchers and clinicians tend to agree that beyond 6 months of age, the risks of moving children increase, and the older the child at placement, the more likely it is that there will be difficulties in the child's behaviour, which increase the risk of placement break down. Some delays in placement are caused by incompetence or poor practice. However, the main reason for delay in placement (sometimes referred to as 'drift') is contested legal proceedings. In most countries it is only possible in extreme circumstances to place a child for adoption without the consent of the birth parents, although adoption by long-term foster carers they have lived with for some years sometimes occurs. In the United Kingdom, United States, and Canada it is not uncommon for parental consent to be dispensed with by court order, but human rights legislation and the attempts at reunification mean that few children are placed from care before the age of 6 months. International adoptions tend to be delayed because of the search within the country of origin for an in country placement, or because of legal formalities.

(f) Medium-term outcomes

Many of the children placed when older bring problems with them into placement, to which may be added those discussed earlier, which are specifically associated with being adopted or fostered. For those placed from overseas, the difficulties are those commonly associated with institutionalization and privation of affection and consistent care. For a large proportion of those placed from care, the problems are those associated with maltreatment or neglect, including attachments with parent figures that may have been anxious, ambivalent, or avoidant, followed by the loss of those attachment figures. They may also have been separated from siblings and experienced multiple changes of carer.

Rushton and Dance[17] provide detailed accounts of the behaviour of 133 English children placed when over the age of 5. Eight years after placement, 19 per cent of the children had left their placements and only just over half of the 99 continuing placements were in the 'continuing/happy' group. Behaviours their parents had difficulty managing included over-activity, aggression, and destructiveness.

(g) Longer-term outcomes for late-placed children

Whilst some children placed in positive environments that provide committed and loving parenting and stability will recover from the adverse effects of early significant harm, developmental recovery cannot be anticipated in all cases. From a longitudinal study of over 1100 'hard-to-place' children placed in adoptive or permanent foster families not previously known to them, Thoburn[18] found that one in five of the placements had disrupted between 2 and 6 years after placement. There was a strong and statistically significant association between disruption and the age at placement (see Fig. 9.3.5.1). Of those aged between 7 and 8 years at placement, one in five experienced placement break down; this proportion rose to almost one in two for those placed between the ages of 11 and 12. The graph is less stable for teenagers, in part because numbers are smaller and statistics less reliable, and in part because families are

Fig 9.3.5.1 Age at placement and percentage of placements disrupting. (Reproduced from J. Thoburn, Evaluating placement: survey findings and conclusions. In *Permanent family placement: a decade of experience* (eds. J. Fratter, J. Rowe, and J. Thoburn), pp. 34–57, copyright 1991, Brirish Agencies for Adoption & Fostering (BAAF), London.)

more likely to 'hang on in' if they know the young person can be helped to leave home 'respectably' in a year or so. More recently, Rushton and Dance[17] found that the placements that disrupted did so between 6 months and 7 years after placement (a mean of 34 months). Rushton and Dance[17] followed up 90 children placed for adoption when aged at least three for an average of 7 years. Seventeen per cent of placements had disrupted (mostly before adoption was finalized), and for a third of those still in placement, there were many problems, which were often getting worse. Festinger followed up for 4 years 516 American children placed from care at a mean age of 2 years. The lower disruption rate (10 per cent) when compared to the UK studies may be explained by the younger age at placement and also by the fact that roughly half were adopted by relatives. She reported that many of the parents in the continuing placements reported difficulties and unmet needs.

Variables about the child and pre-placement history

Researchers and clinicians concur that, in addition to age at placement, variables relating to the child's behaviour and emotional well-being at the time of placement are most strongly associated with better or worse outcomes. These in turn are linked with biography, including experiences of early parenting and multiple caregivers. Having experienced early abuse or neglect has been found to be independently associated with less positive outcomes, whilst more positive outcomes are reported even for late-placed children who had formed a good-enough attachment with a parent or other main carer during the first few years of life.

Thoburn *et al.*[20] found that children of minority ethnic origin, whether placed with a family of the same or a different ethnic origin, were no more likely to experience break down than white children placed with white families. However, qualitative studies note that parents of a different ethnic and cultural background to the child have extra hurdles to overcome in the parenting process, and that some are unable to bring the child up to feel pride in his or her heritage, culture, and appearance, with consequent problems for self-esteem and identity.

Variables about the adoptive or foster families

Early studies reported an association between less positive outcomes and there being a 'home-grown' child younger or close in age to the placed child, but these have not been replicated more recently. The age of the adopters, whether they are single or in a partnership, experienced parents or childless, have not been consistently found to be significantly associated with placement break down. Two in-depth prospective studies of long-term foster care[21] and adoption[22] conclude that those new parents do best who can empathize with both the child and the family of origin, who enjoy a challenge, who have the skills to help the child with disabilities or emotional problems, and who, for older-placed children, can give out love even if the child gives little back and can find pleasure in tiny 'successes'.

Brodzinsky *et al.*[3] identify the importance of adopters treading a fine line between understanding and accepting the difference between parenting by adoption and parenting by birth, but not overemphasizing the difference.

Variables about placement practice and therapy

Few of the studies involving large enough numbers for statistical analysis look for statistically significant associations between outcome and placement practice or therapeutic interventions.

The child who remains for longer than a few weeks in temporary foster care is especially vulnerable to placement break down or moves made for bureaucratic reasons. Once settled in a planned long-term placement, when variables such as age at placement and behavioural difficulties are controlled for, break down rates for older-placed children are similar for children placed for adoption or in permanent foster families. Placements with relatives, and temporary foster placements confirmed as permanent (through adoption, guardianship, or administrative decision), have been found in most studies to have higher success rates that 'stranger' placements. Qualitative studies indicate that most children gain a sense of security from the legal status of adoption, but some can feel trapped and resentful, especially if they lose contact with birth relatives they want to see.[19,20,21,23]

Chapters by US and UK researchers in a book on birth family contact[24] indicate that post-placement contact, in itself, does not adversely impact on the attachment process, and that it can help new parents and children to be more comfortable in talking about adoption issues. For children placed when older, remaining in face-to-face contact with a birth parent, relative, or sibling, and being placed with a sibling, have been associated in several studies with more successful outcomes. However, this contact needs to be carefully managed and can sometimes be harmful to the child and to the stability of the placement.

The role of the psychiatrist

Child or adult psychiatrists will become involved in child placement work because they are asked to provide therapy for a child, young person, or adult who has been placed for adoption or in foster care, or for a teenager or adult who has lost a child to adoption. Whether working with task-centred or permanent carers, the special challenges of this different form of family life have to be acknowledged and incorporated into the therapeutic processes. It is for this reason that the need for specialist child mental health services for children in care or placed for adoption is now recognized, which adapt the full range of effective therapeutic approaches to the special needs of the adoptive or foster family. Barth and colleagues[25] have noted that adoptive parents tend to prefer attachment-based therapies to the parent-training approaches that have been demonstrated to be effective with non-adoptive families.

They hypothesize that in part this is because adopters believe that their special issues are better understood by the clinicians using these therapies, but they point to some potentially harmful effects of some of these methods, which can be experienced by the child as intrusive and coercive. Some clinicians who use attachment theories in their work also challenge the validity of some of these methods.[26,27]

Psychiatrists are often consulted about the advisability of adoption or a return to the birth parents or relatives. The message from research is that adoption and foster care will be better for most children than being left with parents who can not be helped to provide them with safe and loving care. However, they are not without risks, which have to be carefully weighed for each child, and the first step must always be to try to improve the quality of parenting in the birth family. At first sight, the break down rates for the placement of older children (now the majority of those needing placement) may appear discouraging. However, given their many difficulties, it should be welcomed that as many as 50 per cent of 11-year-old children, and more of those below that age, do find permanent substitute families. If permanent out-of-home placement does become necessary, the inherent risks demand the provision of the highest quality services provided for as long as needed.

Further information

For those who wish to pursue these issues in more depth look at references:

- 1, 2, 21, 22, 23 if you are providing therapy for a teenager or adult suffering the harmful effects of losing a child to adoption or foster care
- 7, 8, 10, 11, 12 on therapeutic and task-centred foster care
- 6, 9, 14, 15, 16, 17, 18, 19 to help with the decision-making process on substitute family placement options
- 2, 20, 21, 22, 23, 24 on birth family contact
- 3, 4, 14, 17, 21, 22, 25, 26, 27 on therapeutic approaches and methods when working with troubled children and their adoptive or foster families.

References

1. Carp, E.W. (2004). *Adoption politics: bastard nation and ballot initiative 58*. Kansas University Press, Lawrence.
2. Triseliots, J., Feast, J., and Kyle, F. (2005). *The adoption triangle revisited. A study of adoption, search, and reunion experiences*. BAAF, London.
3. Brodzinsky, D.M., Smith, D.W., and Brodzinsky, A.B. (1998). *Children's adjustment to adoption: developmental and clinical issues*. Sage, Thousand Oaks, CA.
4. Rutter, M. (1999). Resilience concepts and findings: implications for family therapy. *Journal of Family Therapy*, **21**, 119–44.
5. Bullock, R., Courtney, M., Parker, R., *et al.* (2006). Can the corporate state parent? *Children and Youth Services Review*, **28**, 1344–58.
6. Quinton, D. and Selwyn, J. (2006). Adoption: research, policy and practice. *Child and Family Law Quarterly*, **18**(4), 459–77.
7. Sellick, C., Thoburn, J., and Philpot, T. (2004). *Success and failure in permanent family placement*. Hants, Avebury, Aldershot.
8. Sinclair, I. (2005). *Fostering now: messages from research*. Jessica Kingsley, London.
9. Thoburn, J. (2003). The risks and rewards of adoption for children in the public care. *Child and Family Law Quarterly*, **15**(4), 391–402.
10. Wilson, K., Sinclair, I., Taylor, C., *et al.* (2004). *Fostering success: an exploration of the research literature in foster care*. SCIE, London.
11. Testa, M.F. and Rolock, N. (1999). Professional foster care: a future worth pursuing? *Child Welfare*, **LXXVIII**(1), 109–24.
12. Fisher, P. and Chamberlain, P. (2000). Multi-dimensional treatment foster care. *Journal of Emotional and Behavioural Disorders*, **8**(3), 155–64.
13. Festinger, T. (2002). After adoption: dissolution or permanence? *Child Welfare*, **LXXX1**(3), 515–33.
14. Rutter, M. and the English and Romanian Adoptees (ERA) Study Team. (1998). Developmental catch–up and deficit, following adoption after severe global early privation. *Journal of Child Psychology and Psychiatry*, **39**, 465–76.
15. Howe, D. (1997). Parent-reported problems in 211 adopted children. *Journal of Child Psychology and Psychiatry*, **38**(4), 401–11.
16. Lindblad, F., Hjern, A., and Vinnerljung, B. (2003). Inter-country adopted children as young adults-a Swedish cohort study. *The American Journal of Orthopsychiatry*, **73**, 190–202.
17. Rushton, A. and Dance, C. (2004). The outcomes of late permanent placements. *Adoption & Fostering*, **28**(1), 49–58.
18. Thoburn, J. (1991). Evaluating placements: survey findings and conclusions. In *Permanent family placement: a decade of experience* (eds. J. Fratter, J. Rowe, and J. Thoburn), pp. 34–57. BAAF, London.
19. Selwyn, J. and Quinton, D. (2005). Stability, permanence and outcomes: foster care and adoption compared. *Adoption and Fostering*, **28**(4), 6–16.
20. Thoburn, J., Norford, L., and Rashid, S.P. (2000). *Permanent family placement for children of minority ethnic origin*. Jessica Kingsley, London.
21. Beek, M. and Schofield, G. (2004). *Providing a secure base in long-term foster care*. BAAF, London.
22. Neil, E. (2003). Understanding other people's perspectives: tasks for adopters in open adoptions. *Adoption Quarterly*, **6**(3), 3–30.
23. Timms, J. and Thoburn, J. (2006). Your shout! Looked after children's perspectives on the Children Act 1989. *Journal of Social Welfare and Family Law*, **28**(2), 153–70.
24. Neil, E. and Howe, D. (2004). *Contact in adoption and permanent foster care: research, policy and practice*. BAAF, London.
25. Barth, R.P., Crea, T.M., John, K., *et al.* (2005). Beyond attachment theory and therapy: towards sensitive evidence-based interventions with foster and adoptive families in distress. *Child and Family Social Work*, **10**(4), 257–68.
26. Dozier, M. (2003). Attachment-based treatment for vulnerable children. *Attachment and Human Development*, **5**, 253–7.
27. Prior, V. and Glaser, D. (2006). *Understanding attachment and attachment disorders. Theory, evidence and practice*. Jessica Kingsley, London.

9.3.6 Effects of parental psychiatric and physical illness on child development

Paul Ramchandani, Alan Stein, and Lynne Murray

Introduction

A broad range of physical and psychiatric illnesses commonly affect adults of parenting age. For example, approximately 13 per cent of women are affected by depression in the postnatal period, and the prevalence of depression in parents of all ages remains high. Many parents will also experience severe physical illness; breast cancer affects approximately 1 in 12 women in the United Kingdom, about a third of whom have children of school age. Worldwide HIV has an enormous impact on adults of parenting age. In some parts of sub-Saharan Africa up to 40 per cent of women attending antenatal

clinics are HIV positive. Many of these parental disorders are associated with an increased risk of adverse emotional and social development in their children, and in some cases cognitive development and physical health are also compromized. It must be emphasized that a significant proportion of children at high risk do not develop problems and demonstrate resilience,[1] and, many parents manage to rear their children well despite their own illness. Nonetheless these risks represent a significant additional impact and burden of adult disease (both physical and psychiatric) that is often overlooked.

This chapter reviews the current state of evidence regarding selected examples of psychiatric and physical conditions, from which general themes can be extracted to guide clinical practice. Some of the key mechanisms whereby childhood disturbance does or does not develop in conjunction with parental illness are considered, and strategies for management and intervention reviewed.

Parental psychiatric illness

There is now reasonable evidence to suggest that most types of psychiatric disorder affecting parents are associated with an increased risk of difficulties for their children. There are some differences in risk by type of disorder; however, there are also some commonalities, suggesting that some of the mechanisms may be shared. Children's disorders may resemble those of their parents[2,3] but there is also evidence of a much broader range of problems, including adverse effects on children's social, emotional, cognitive and physical development. In the following section we will focus on parental depression, schizophrenia, eating disorders, alcoholism and substance abuse, and anxiety, but similar issues apply for other disorders not considered here.

Depression

Depression in either parent is associated with an increased risk of child psychopathology and other developmental difficulties, with the risks continuing into adulthood.[3] The longest running longitudinal study[4] found that, as well as a three-fold increase in major depressive disorder, the adult offspring of depressed parents had increased rates of anxiety disorders and substance dependence, as well as greater social impairment and physical health impairment. There has been a large body of research focusing on depression affecting mothers in the postnatal period, with studies demonstrating that infants and children have an increased risk of emotional and behavioural problems.[5] Some studies have suggested that children's cognitive development may also be affected, although the results from studies are not consistent.[6] Similarly, there is a suggestion that boys may be more affected than girls in early childhood. As the children enter adolescence an increased risk of mood and anxiety disorders emerges.[7] More recently research in developing countries has shown an association between postnatal depression and an increased risk of physical health problems in infants such as poor growth and diarrhoel illness.[8]

Much less work has been done on depressed fathers, although consistent evidence is now beginning to emerge of an independent effect of paternal depression on children's development.[9] The overall impact may be less than that of maternal depression, and there are also conjoint effects to consider, as depression in one parent can often co-occur with depression or another psychiatric disorder in the other parent. Similarly there may be protective effects if one parent remains well.[10]

While genetic factors clearly play an important role in the transmission of risk from parents to children, environmental factors, and the interactions that occur between genetic and environmental factors, also have substantial influence.[11,12] In the case of depression, the core symptoms of low mood, loss of interest and low energy can have a significant impact on parenting capacity and parent-child relations. These include a parent's capacity to be responsive, consistent, and warm when interacting with their children, particularly in the first few years of life. For example, depressed mothers may be less vocal, less positive, and less spontaneous than controls, more negative, unsupportive, and intrusive, and have more difficulty in communicating and listening to their young children.[13, 14]

Depression in either parent is strongly associated with marital discord.[15] This may play a key role in mediating the effects of parental depression and may be a more proximal predictor of child outcomes than depression.[13] The way in which conflicts are resolved may be very important and depressed parents are likely to use less effortful strategies, such as withdrawal. Children are generally more at risk as they are exposed to an increased number of risk factors, and children whose parents are depressed are particularly at risk if they are also socio-economically disadvantaged.

The direction of effects is not all from parent to child, and temperamental and behavioural factors in the child may also contribute to increasing family discord, parental psychiatric disturbance, and parenting impairments, and ultimately to disturbances in parent-child attachment. Infant irritability and poor motor control, measured before the onset of any maternal depression at 10 days postpartum, increase the risk that a mother will become depressed.[16] The influence of parental depression on child development thus represents a complex bidirectional interaction between individual vulnerability (which may be genetic), influences of depression on parenting characteristics, parent–child relationships, the wider context of the parental relationship, and other aspects of social disadvantage.

Schizophrenia

Parents with a diagnosis of schizophrenia have a greatly increased risk of having children who later develop schizophrenia themselves. Risks to child development are identified from birth, with an increased likelihood of obstetric complications, not fully accounted for by maternal behaviour during pregnancy, or by genetic risk.[17] During childhood, prior to the onset of any psychiatric symptoms, attentional problems similar to those found in adult schizophrenic patients have been identified and these problems not only persist into adulthood, but attentional problems have been identified as key neurobiological indicators of risk for subsequent schizophrenia or other psychopathology in adolescence and young adulthood.[18]

Social difficulties with peers and teachers are found in many longitudinal studies of children with schizophrenic parents,[19] although not necessarily to a greater extent than in children with parents suffering from affective disorder and a higher IQ can be protective. Social relationship problems and associated thought disorder may become more marked in adolescence and seem to be predicted by attentional problems.[18] As young adults, children of schizophrenic parents are at high risk for schizotypal behaviour, although this broader range of difficulties does not necessarily distinguish them from parents with affective illness.

A pattern of disturbed communication has been described in families with a schizophrenic parent,[20] but the importance of these interactions in explaining long-term outcomes has been questioned.

Overall, cognitive and attention difficulties appear to be largely associated with specific brain abnormalities linked with schizophrenia, but other kinds of childhood problems are probably influenced

more by the general family disruption associated with a parent who requires hospital admission and who may have difficulties with employment and other social relationships beyond the family.

Eating disorders

Eating disorders occur commonly among women of child-bearing age.[21] Studies have raised concern that mothers' attitudes and behaviours regarding food and body shape, may influence their children's feeding, and ultimately the children's own attitudes to body shape and eating.[22] Children are particularly vulnerable at two stages of development—infancy and adolescence. During infancy feeding and mealtimes take up a significant part of the day and provide important times for close communication between parents and children. A Scandinavian study has indicated that failure to thrive may be a risk in the first year amongst women with a history of anorexia nervosa.[23] One controlled observational study of 1-year-old children of mothers with eating disorders found that the mothers were intrusive with their infants during both mealtimes and play, and they expressed more negative emotion and conflict during mealtimes than controls, and allowed their children less autonomy.[24] Furthermore, infant weight was independently and inversely related to mealtime conflict.[25] Follow up studies of children of mothers who have experienced eating disorders in the postnatal period indicate that in middle childhood they are more likely than control children to value themselves by body shape and weight, and to use dietary restriction.[22, 26]

During adolescence children become more aware of societal pressures and develop increasing interest in body shape and attractiveness while preoccupied with their own concerns about food, body shape, and weight. Children may model themselves on their parents, and parents may influence their adolescent children directly by expressing attitudes towards their children's weight, shape, and eating habits. However, it should be emphasized that, in common with most parent psychopathology, the children of parents with eating disorders are not invariably adversely affected. Some parents manage well and their children develop without apparent problems.

Alcoholism and substance abuse

A substantial body of evidence has been amassed on the effects of parental alcoholism and substance abuse.[27] They are wide ranging, identifiable throughout development, and work has highlighted the importance of the social effects on the child in addition to physical and psychological outcomes.[28] Both genetic and environmental factors seem to be involved.

The impact of maternal alcoholism on the developing child can be found from the prenatal period.[29] Present in 0.01 to 0.03 per cent of normal births, foetal alcohol syndrome appears in 5.9 per cent of births of alcoholic women. There is also considerable evidence that infants exposed prenatally to heroin and cocaine are at increased risk of a number of developmental difficulties which may persist throughout childhood.[30]

Studies of outcome consistently describe impaired cognitive and social development in children of alcoholics and heroin users.[27,30] An increased risk of attention-deficit hyperactivity disorder, attention problems, and impulsivity in children of alcoholics is the most consistent finding. Children of drug abusers may also be more aggressive and have fewer friends and are at risk for criminality, depression, anxiety, and somatic problems. Children of alcoholics have an increased risk of becoming alcoholics themselves and, similarly, children of drug abusers also have an increased risk of drug abuse in adolescence, although many are resilient and do not develop similar problems themselves.

Social factors such as poverty and social isolation are known to influence child development adversely and it has been difficult to differentiate between the effects of the disorder and the associated adversity. Substance abuse in the parent may lead to impaired parenting and an impoverished social environment, leaving children vulnerable to neglect or abuse and contributing to impaired social and cognitive functioning, psychopathology, substance abuse, and delinquency but the relative impact of each factor has yet to be resolved. Studies have identified deficits in parenting behaviour and, in particular, neglect and harsh discipline.[31] Divorce and marital conflict are also more likely and there is evidence of assortative mating, all of which are likely to compound the risk for the children.

Stressful life events, and in particular those related to family conflict, have proved to be important in accounting for the link between paternal alcoholism and alcohol use in their offspring.[32] Overall, clinicians have emphasized a family perspective when conceptualizing the multiple levels of stress and vulnerability associated with parental alcoholism and opiate abuse and the need to enhance social supports for the family.[30]

Anxiety disorders

There is less information about children whose parents have anxiety disorders but there appears to be a considerable degree of specificity in the familial transmission of anxiety disorder.[33] One study showed a two-fold increased risk of anxiety disorders among offspring of parent probands compared with offspring of substance abusers or controls.[34] However, they are also at risk of other kinds of problems such as depression.[35]

In a search for the mechanism of transmission, it has been suggested that children at risk for developing anxiety disorders have a temperamental vulnerability characterized by behavioural inhibition and autonomic reactivity, identifiable in infancy by an increased startle response.[36,37] The question of the relative influence of genetic or environmental factors does suggest a lesser genetic component for anxiety disorders,[38] and a developing line of research has led to a number of aspects of parenting behaviour of parents with anxiety being considered important, including over-protection and the limiting of children's opportunities to develop new skills[39]. There appear to be some specific differences by type of anxiety disorder with, for example, mothers with social phobia demonstrating characteristic patterns of modelling of anxiety by the parent, and failure to provide encouragement/opportunities for child autonomy.[40]

Recent lines of research have also explored possible risks in utero to the developing foetus. Children exposed to maternal anxiety in utero are at an increased subsequent risk of behavioral problems. This increased risk appears to persist through childhood, leading to suggestions that the mechanism may be in part mediated through an antenatal effect on the HPA axis of the developing foetus.[41] More research is needed to clarify the mechanisms but it is clear that children whose parents have anxiety disorders are at risk of developing psychiatric disturbance themselves.

Parental physical illness

Many families experience chronic parental illness and paradoxically, as treatment techniques improve illnesses may extend over longer time periods, which may have greater impact on family members

including children. There have been limited reports about the impact of physical parental illness on children, however the importance of this area is beginning to be recognized, particularly in relation to parental cancer, and also HIV.[42] Similar associated difficulties can arise with many other parental illness, particularly chronic ones such as diabetes. However we will here confine our comments to parental cancer and HIV, and the general issues that these conditions illustrate for the developing child.

Cancer

Parental cancer is likely to be associated with depression and marital difficulties, both risk factors for the child. The balance of evidence indicates that their children are at increased risk of developing psychological disturbance.[43] The impact of parental cancer on family communication and child outcomes may vary according to the child's developmental level, their gender, the presence of disability in the child, and the parent's level of psychological distress and marital discord.[44]

A recent review[45] found that adolescents who had a parent with cancer had higher levels of emotional disturbance, than a normal population sample, but younger children were not consistently found to exhibit higher rates of problems, although some studies suggest this. Children's own responses to their predicament are likely to affect their eventual adjustment. Problem-focused or active coping affects the stressors (e.g. seeking information, positive reinterpretation of stressful events) and is expected to be more adaptive while emotion-focused coping (e.g. venting emotions, denial, apathy) draws attention away from the stressors but may place children at risk for anxiety and depression.[46] Health professionals may need to assist parents in recognizing and coping with their children's distress when it is present. Specifically, communication about the parental illness and how the children feel appears to be crucial to children's coping.[47] The levels of anxiety and distress amongst the children are related to whether they are told about the illness and the quality of the communication with the parents, with informed children having lower levels of anxiety than those who are uninformed.[44]

HIV status and AIDS

Women of child-bearing age account for an increasing number of sufferers of HIV/AIDS in the developing world and there is now increasing evidence that their children are at increased developmental risk, even if the children are not HIV positive themselves.[43] Young children may have fewer problems[48] but those of school age are at risk for externalizing and internalizing problems, lower social skills, and academic achievement difficulties.[49] Maternal depression is relatively common amongst women diagnosed with HIV during pregnancy,[50] and the impact of this on their caregiving capabilities may be one of the key mechanisms by which children are affected. Given the enormity of the HIV pandemic and possible parallels with other common infectious diseases such as Malaria and TB, there is serious need for further research in this specific field.

Summary of mechanisms

Some of the key mechanisms by which increased risk for child disturbance is transmitted from ill parents to their children have been described above. In cases of psychiatric disorder the link may in part reflect genetic transmission, but clearly a considerable amount of the variance is accounted for by environmental mechanisms and a number of such mechanisms have been proposed.[5,11] Most of these can apply equally to parental physical or psychiatric illness. First, parental illness may interfere with parental functioning and parent-child interactions, for example where a parent becomes withdrawn or preoccupied and relatively unavailable to the child. Second, a number of family and environmental factors such as marital/family discord and severe housing or associated economic deprivation are associated with mental illness and these constitute risks in their own right, as well as sometimes being a direct consequence of the parental disorder (for example, parental depression leading to increased marital conflict, although the reverse direction of causality can also occur). Third, in rarer instances parental symptoms may impinge directly on the child, for example where the parent incorporates a child into the core symptomatology such as a delusion or an obsession. Finally, it should not be assumed that influences are unidirectional. Child characteristics such as early temperamental difficulties or behavioural problems may influence the outcome of parental illness, particularly in the case of depression.[3] Child characteristics such as coping style, intellectual ability, or sociability may be particularly important in explaining resilience.

Implications—responding to a parent in clinic and preventive interventions

As the influences of parental illness on children's development become better understood there is increasing recognition of the potential importance of addressing parent-child links. This is both in making better enquiry about child welfare when a parent is seriously ill, but also enquiring about parental health when a child presents with a disorder such as depression (e.g. NICE guidance on depression in children and young people[51]) Two particular areas will be considered here; when a parent presents with a serious illness (psychiatric or psychological) and potential intervention strategies with high-risk groups who may or may not be presenting to health services.

When parents with serious physical or psychological illness are seen as outpatients, especially if they are subsequently admitted to hospital, it should be routine to enquire about children—their ages, developmental and scholastic progress, child care arrangements, and family support. In the limited time available, it makes most sense to enquire about the areas most likely to be affected by the parental disorder.[52] For example, where a parent with depression has young children, enquiry could focus on the level of care that the parent feels able to provide to the children, the feelings that the parent has for their children, and the presence of any marital or family discord, as well as the available support. These questions obviously require sensitive handling, as parents can easily feel blamed for any effect that they perceive their illness may have had on their children. In many cases provision of appropriate support and treatment to the parent will have a sufficiently beneficial effect for the whole family that no further intervention is required, (for example, findings from a recent large randomized controlled trial identify improvements in children's outcomes when their mother's depression is treated.[53]) However it is crucial to ask and make at least this preliminary assessment. Close collaboration with the primary care team, including the family's general practitioner, is of considerable importance. The general practitioner's involvement may be critical, both in terms of potential treatment and support for the family, and also as a

source of knowledge of the wider family system, and the resources available to the family.

In those cases where a child is more severely affected (either directly by a parent's illness or by other related factors), consideration should be given to including relevant children's services, either Child and Adolescent Mental Health Services (CAMHS) or children's care services. Close collaborative relationships between child and adult mental health services clearly ease potential joint working, and should be the norm. Communication with, and the close involvement of, the family's general practitioner, can also contribute to a plan to benefit the whole family system.

In those cases where a parent has a chronic illness and the clinician gets to know them over a period of time, it can be helpful to ask about children's understanding and knowledge of the parental illness and the extent to which it has been discussed within the family. These issues will need to be carefully and sensitively handled. Families have a range of ways in which they communicate and it is important not to compound the parent's problems by making them feel their illness is harming their children. As far as inpatients are concerned, it is important that facilities should be available for children visiting and, unless specifically contraindicated, regular contact should be encouraged, and appropriate play and other materials should be available. Discussions should be held with the patient and/or relatives about child-care arrangements, and the patient and family need to be helped to think about providing the child with an appropriate explanation about parental health and absence.

Communication either about the parental illness or the associated discord may help to alleviate some of the risks for childhood problems.[44] Studies of marital disruption and divorce have found that children cope better with marital conflict when they are given some explanation or told that the conflict has been resolved.[54] Children often feel left out and, without knowledge of the parental illness, they may be particularly likely to attribute any family conflict or disruption to themselves in their effort to understand the changes taking place. Intervention, including working with family communication, can have significant positive effects if well handled. Some guidelines are available.[55,56]

Less work has been conducted in families where the parent is physically ill, although reports of treatment once the family members are experiencing problems are providing some ways forward.[45] Parents may consciously avoid disclosure because of the questions they anticipate from their children, particularly about death. Communication in this context is not only a matter of disclosure of the illness but a starting point for ongoing discussion and questions, without which children may be at increased risk.[57] The most important role of support services may be to rehearse with parents the kinds of questions that might occur and how they could respond, a strategy which may also facilitate discussion of the ill parent's anxieties.

Prevention

Beyond the scope of the individual parent and family in clinic there have been a number of studies which have examined the possibility of preventative intervention in high-risk groups. This is most often conducted in the context of parental depression. One series of intervention studies with children of parents with major depression has shown marked improvement in family functioning and child outcome.[58] These interventions have taken a variety of formats, but all include a component of psycho-education directed to the whole family. However, in another trial, the involvement of parents in a group programme for adolescent offspring of parents with depression did not show any additional benefit.

Some preventive intervention can clearly be accomplished earlier. In infancy it is possible to promote the development of secure parent–child attachment, which should be protective even if the parental illness is chronic. Maternal sensitivity can be enhanced using videotaped mother-child interactions, which can increase the rate of secure attachment in at-risk families.[59] In the case of maternal eating disorders there is now some encouraging evidence for the effectiveness of video feedback treatment to enhance maternal responsivity to the infant, and to decrease mother-infant conflict.[60] A recent review of treatments for mothers and infants in the context of maternal depression found that treatment of maternal depression alone did not appear to mitigate the impact of the depression on the child. However, a number of different mother-infant psychotherapies did appear to confer benefit on mother child interaction and child outcome. However, much work remains to be done.[61]

Conclusions

In conclusion, children of parents with physical or psychiatric illness are at risk of a wide range of developmental and psychiatric difficulties, although not all will develop problems. Future work should be directed to developing and evaluating ways of providing support so that parents can best manage their illnesses and to prevent or mitigate any negative effects on their children.

Further information

The SCIE parental mental health network (www.scie.org.uk/mhnetwork/resources.asp)

The Children of Parents with a Mental Illness (COPMI) (www.aicafmha.net.au/copmi/index.html)

References

1. Rutter. M. (1990). Psychosocial resilience and protective mechanisms. In *Risk and protective factors in the development of psychopathology* (eds. J. Rolf, A.S. Masten, D. Cicchetti, K.H. Neuchterlein, and S. Weintraub), pp. 181–214. Cambridge University Press, New York.
2. Bernstein, G.A. and Borchardt, C.M. (1991). Anxiety disorders of childhood and adolescence: a critical review. *Journal of the American Academy of Child and Adolescent Psychiatry*, **30**, 519–32.
3. Downey, G. and Coyne, J.C. (1990). Children of depressed parents: an integrative review. *Psychological Bulletin*, **108**, 50–76.
4. Weissman MM, Wickramaratne P, Nomura Y, *et al.* (2006). Offspring of depressed parents: 20 years later. *American Journal of Psychiatry*, **163**(6), 1001–8.
5. Murray, L. and Cooper, P. (2003). Intergenerational transmission of affective and cognitive processes associated with depression: infancy and the preschool years. In *Unipolar Depression: a Lifespan Perspective* (ed. I. Goodyer), pp. chapter 2. Oxford University Press, Oxford.
6. Sohr-Preston, S.L. and Scaramella, L.V. (2006). Implications of timing of maternal depressive symptoms for early cognitive and language development. *Clinical Child and Family Psychology Review*, **9**(1), 65–83
7. Halligan, S.L., Murraym, L., Martins, C., *et al.* (2007). Maternal depression and psychiatric outcomes in adolescent offspring: a 13-year longitudinal study. *Journal of Affective Disorders*, **97**, 145–54.
8. Rahman, A., Iqbal, Z., Bunn, J., *et al.* (2004). Impact of maternal depression on infant nutritional status and illness: a cohort study. *Archives of General Psychiatry*, **61**, 946–52.
9. Ramchandani, P., Stein, A., Evans, J., *et al.* (2005). Paternal depression in the postnatal period and child development: a prospective population study. *Lancet,* **365**(9478), 2201–5.

10. Kahn, R.S., Brandt, D. and Whitaker, R.C. (2004). Combined effect of mothers' and fathers' mental health symptoms on children's behavioral and emotional well-being. *Archives of Pediatrics and Adolescent Medicine*, **158**(8), 721–9.
11. Rutter, M. (2007). Gene-environment interdependence. *Development Science*, **10**(1), 12–8.
12. Kim-Cohen, J., Moffitt, T.E., Taylor, A., *et al.* (2005). Maternal depression and children's antisocial behavior: nature and nurture effects. *Archive of General Psychiatry*, **62**(2), 173–81.
13. Stein, A., Gath, D.H., Bucher, J., *et al.* (1991). The relationship between post–natal depression and mother–child interaction. *British Journal of Psychiatry*, **158**, 46–52.
14. Murray, L., Fiori-Cowley, A., Hooper, R., *et al.* (1996). The impact of postnatal depression and associated adversity on early mother-infant interactions and later infant outcome. *Child Development,* **67**(5), 2512–26.
15. Cummings, E. M., Keller, P. S., and Davies, P. T. (2005). Towards a family process model of maternal and paternal depressive symptoms: exploring multiple relations with child and family functioning. *Journal of Child Psychology and Psychiatry and allied disciplines*, **46**(5), 479–89.
16. Murray, L., Hipwell, A., Hooper, R., *et al.* (1996). Cognitive development of 5 year old children of postnatally depressed mothers. *Journal of Child Psychology and Psychiatry*, **37**, 927–35.
17. Sacker, A., Done, D.J., and Crow, T.J. (1996). Obstetric complications in children born to parents with schizophrenia: a meta–analysis of case–control studies. *Psychological Medicine*, **26**, 279–87.
18. Dworkin, R.H., Lewis, J.A., Cornblatt, B.A., *et al.* (1994). Social competence deficits in adolescents at risk for schizophrenia. *Journal of Nervous and Mental Disease*, **182**, 103–8.
19. Weintraub, S. (1987). Risk factors in schizophrenia: the Stony Brook High–Risk Project. *Schizophrenia Bulletin*, **13**, 439–50.
20. Greenwald, D.F. and Harder, D.W. (1994). Outcome predictors in a longitudinal study of high-risk boys. *Journal of Consulting and Clinical Psychology*, **50**, 638–43.
21. Fairburn, C. G., and Harrison, P. J. (2003). Eating disorders. *Lancet,* **361**(9355), 407–16.
22. Stein, A., Woolley, H., Cooper, S., *et al.* (2006). Eating habits and attitudes among 10-year-old children of mothers with eating disorders. A longitudinal study. *British Journal of Psychiatry,* **189**, 324–9.
23. Brinch, M., Isager, T., and Tolstrup, K. (1988). Anorexia nervosa and motherhood: reproduction pattern and mothering behaviour of 50 women. *Acta Psychiatrica Scandinavica*, **77**, 611–17.
24. Stein, A., Woolley, H., Cooper, S.D., *et al.* (1994). An observational study of mothers with eating disorders and their infants. *Journal of Child Psychology and Psychiatry*, **35**, 733–48.
25. Stein, A., Murray, L., Cooper, P., *et al.* (1996). Infant growth in the context of maternal eating disorders and maternal depression: a comparative study. *Psychological Medicine*, **26**, 569–74.
26. Jacobi, C. Agras, W.S. and Hammer, L. (2001). Predicting children's reported eating disturbances at 8 years of age. *Journal of American Academy of Child and Adolescent Psychiatry*, **40**, 364–72.
27. Lieberman, D. (2000). Children of alcoholics: an update. *Current Opinion in Pediatrics,* **12**(4), 336–340.
28. Burns, E.C., O'Driscoll, M., and Wason, G. (1996). The health and development of children whose mothers are on methadone maintenance. *Child Abuse Review*, **5**, 113–22.
29. Nordberg, L., Rydelius, P.–A., and Zetterstrom, R. (1994). Parental alcoholism and early child development. *Acta Paediatrica*, **83** (Suppl. 404), 14–18.
30. Hogan, D.M. (1998). The psychological development and welfare of children of opiate and cocaine users: review and research needs. *Journal of Child Psychology and Psychiatry*, **39**, 609–20.
31. Famularo, R., Kinscherff, R., and Fenton, T. (1992). Parental substance abuse and the nature of child maltreatment. *Child Abuse and Neglect*, **16**, 475–83.
32. Hill, S.Y. and Hruska, D.R. (1992). Childhood psychopathology in families with multigenerational alcoholism. *Journal of the American Academy of Child and Adolescent Psychiatry*, **31**, 1024–30.
33. Biederman, J., Petty, C., Faraone, S. V., *et al.* (2006). Effects of parental anxiety disorders in children at high risk for panic disorder: A controlled study. *Journal of Affective Disorder*, **94**, 191–7.
34. Merikangas, K.R., Dierker, L.C., and Szatmari, P. (1998). Psychopathology amongst offspring of parents with substance abuse and/or anxiety disorders: a high risk study. *Journal of Child Psychology and Psychiatry*, **39**, 711–20.
35. Warner, V., Mufson, L., and Weissman, M. (1995). Offspring at high risk for depression and anxiety: mechanisms of psychiatric disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, **34**, 786–97.
36. Grillon, C., Dierker, L., and Merikangas, K.R. (1997). Startle modulation in children at risk for anxiety disorders and/or alcoholism. *Journal of the American Academy of Child and Adolescent Psychiatry*, **36**, 925–32.
37. Kagan, J., Reznick, J., and Snidman, N. (1987). The physiology and psychology of behavioral inhibition in children. *Child Development,* **58**, 1459–73.
38. Gordon, J. A., and Hen, R. (2004). Genetic approaches to the study of anxiety. *Annual review of neurosciences,* **27**, 193–222.
39. Wood, J.J., McLeod, B.D., Sigman, M., *et al.* (2003). Parenting and childhood anxiety: theory, empirical findings, and future directions. *Journal of Child Psychology and Psychiatry,* **44**, 134–51.
40. Murray, L. Cooper, P. Cresswell, C. *et al.* (2007). The effects of maternal social phobia on mother-infant interactions and infant social responsiveness. *Journal of Child Psychology and Psychiatry*, **48**, 45–52.
41. O'Connor TG, Ben-Shlomo Y, Heron J, *et al.* (2005). Prenatal anxiety predicts individual differences in cortisol in pre-adolescent children. *Biological Psychiatry*, **58**(3), 211–**7**.
42. Stein, A., Krebs, G., Richter, L., *et al.* (2005). Babies of a pandemic. *Archives of Disease in Childhood,* **90**, 116–8.
43. Welch, A.S., Wadsworth, M.E., and Compas, B.E. (1996). Adjustment of children and adolescents to parental cancer. *Cancer*, **77**, 1409–18.
44. Kroll, L., Barnes, J., Jones, A., *et al.* (1998). Cancer in parents: telling children. *British Medical Journal*, **316**, 880.
45. Visser, A., Huizinga, G. A., van der Graaf, W. T., *et al.* (2004). The impact of parental cancer on children and the family: a review of the literature. *Cancer Treat Rev,* **30**(8), 683–94.
46. Compas, B.E., Worsham, N.L., Ey, S., *et al.* (1996). When mom or dad has cancer. II. Coping, cognitive appraisals, and psychological distress in children of cancer patients. *Health Psychology*, **15**, 167–75.
47. Forrest, G., Plumb, C., Ziebland, S., *et al*, (2006). Breast cancer in the family—children's perceptions of their mother's cancer and its initial treatment: qualitative study. *British Medical Journal*, **332**, 998–1003.
48. Black, M.M., Nair, P., and Harrington, D. (1994). Maternal HIV infection: parenting and early child development. *Journal of Pediatric Psychology*, **19**, 595–616.
49. Forehand, R., Steele, R., Armistead, L., *et al.* (1998). The Family Health Project: psychosocial adjustment of children whose mothers are HIV infected. *Journal of Consulting and Clinical Psychology*, **66**, 513–20.
50. Rochat, T.J., Richter, L.M., Doll, H.A., *et al.* (2006). Depression among pregnant rural South African women undergoing HIV testing. *Journal of the American Medical Association*, **295**(12), 1376–8.
51. National Institute for Health and Clinical Excellence (NICE). (2005). *Depression in children and young people: identifi cation and management in primary, community and secondary care*. London: National Collaborating Centre for Mental Health
52. Rutter, M. (1989). Psychiatric disorder in parents as a risk factor for children. In *Prevention of mental disorder, alcohol and other drug use in children and adolescents*. (eds. D Schaffer, I. Phillips, and N.B. Enger), pp. 157–189. Office for Substance Abuse, Rockville, MD.
53. Weissman, M., Pilowsky, D. J., Wickramaratne, P. J., *et al.* (2006). Remissions in maternal depression and child psychopathology. A STAR*D-Child Report. *Journal of the American Medical Association*, **295**(12), 1389 – 98.
54. Cummings, E.M. and Davies, P.T. (1998). *Families, conflict, and conflict resolution: the children's perspective.* Guilford Press, New York.

55. Falkov, A. (ed.) (1999). *Crossing bridges: training resources for working with mentally ill parents and their children*. Pavilion Publications, Brighton.
56. Beardslee, W. (2004). *When a parent is depressed*. Little, Brown and company:USA.
57. Rotheram–Borus, M.J., Draimin, B.H., Reid, H.M., *et al.* (1997). The impact of illness disclosure and custody plans on adolescents whose parents live with AIDS. *AIDS*, **11**, 1159–64.
58. Beardslee, W.R., Gladstone, T.R., Wright, E.J., *et al.* (2003). A family-based approach to the prevention of depressive symptoms in children at risk: evidence of parental and child change. *Pediatrics,* **112**(2), e119–131.
59. Van Zeijl, J., Mesman, J., Van IJzendoorn, M.H., *et al.* (2006). Attachment-based intervention for enhancing sensitive discipline in mothers of 1- to 3-year-old children at risk for externalizing behavior problems: a randomized controlled trial. *Journal of Consulting and Clinical Psychology*, **74**(6), 994–1005.
60. Stein, A., Woolley H, Senior R, *et al.* (2006). Treating disturbances in the relationship between mothers with eating disorders and their infants: a randomized controlled trial. *American Journal of Psychiatry*, **163**(5), 899–906.
61. Nylen, K.J., Moran, T.E., Franklin, C.L., *et al.* (2006). Maternal depression: a review of relevant treatment approaches for mothers and infants. *Infant Mental Health Journal*, **27**(4), 327–43.

9.3.7 The effects of bereavement in childhood

Dora Black and David Trickey

Introduction

Bereavement is not an illness in itself, although it may cause illness or predispose to one. The reaction to the loss of a loved one may lead to temporary or long-term psychological distress and/or loss of function, and may occasion consultation with the general practitioner and referral to mental health professionals.

During the first two years of life, through instinctive behaviours which are modified by experience, infants and their main carers develop an attachment. This bond between a child and his caretaker(s) ensures the child's survival, enables his or her optimum physical, intellectual, and emotional development, and in due course ensures the survival of the species. The nature of the attachment between infant and carer(s) influences the way in which children come to view their social world; the pattern of attachment developed in the first two years of life often remains stable and is associated with the way in which children relate to other people later in their life. Attachment behaviour has been observed across different species and has obvious benefits for survival. However, part and parcel of attachment for the child is distress at separation. Infants who develop a secure attachment can gradually tolerate longer periods of separation from their carer and any distress is rapidly assuaged when they are re-united with their carer. When considered within the context of attachment theory, it is inevitable that permanent separation (e.g. through bereavement) will cause distress for the bereaved. Parkes reviews the body of attachment research and offers a comprehensive description of attachment with particular reference to its role in understanding the impact of loss.[1]

The DSM-IV-TR has a classification for 'Bereavement' (V62.82) differentiating it from 'Major depressive disorder' (296.2) which, unless the symptoms are severe, is generally not diagnosed until 2 months after the loss. ICD-10 has no separate classification for bereavement and suggests the use of 'Adjustment disorders' (F43.2) for temporary reactions to life-events, and 'Death of a family member' (Z63.4) for normal bereavement reactions not exceeding 6 months in duration.

In industrialized countries between 1.5 and 4 per cent of children are orphaned of at least one parent in childhood. Premature deaths in the parenting years may be due to illness, accident, war, civil conflict, natural and man-made disasters and the incidence of these are all higher in developing countries. It is estimated by UNICEF that, in some developing countries, 21 per cent of children are orphaned of at least one parent; with HIV AIDS responsible for up to three-quarters of the deaths.[2]

Reactions to the death of a parent

Research studies

It is generally accepted that loss of a parent in childhood is associated with harmful psychological consequences, however it is difficult to tease out the independent effects of adverse circumstances before the death, the loss itself and the subsequent disruption to the child's life, including the possibility of compromized parenting post-bereavement[3,4]. Most published research about bereaved children describes small-scale uncontrolled studies carried out on children and adolescents referred to mental health facilities. Dowdney comprehensively reviews the research examining the psychological impact of being bereaved of a parent in childhood. She concludes that despite methodological weaknesses, certain findings consistently emerge: 'Children do experience grief, sadness, and despair following parental death. Mild depression is frequent, and can persist for at least a year after parental death'. Bereaved children commonly exhibit a range of psychological symptoms that may not constitute a specific disorder, but the severity of which is likely to warrant referral to a specialist service for one in five bereaved children.[5]

Long-term effects of bereavement

There continues to be debate about a possible link between being bereaved of a parent as a child, and mental health as an adult. The debate is complicated by methodological weaknesses in studies, inconsistent results and difficulty in isolating the impact of experiences which may precede or follow the loss. Any long term consequences of parental bereavement can be mitigated by the subsequent provision of adequate parenting[6,7]. Furthermore, studies in behavioural genetics are increasing the understanding of how genetic endowment interacts with environmental hazards to lead to the presence or absence of mental health problems.[8]

Cultural and religious issues

Reactions to loss are biologically based and are therefore likely to transcend cultural differences, although culture may modify their expression.[3] Religious beliefs about what happens after death can be confusing to young children at the stage of concrete thinking and need to be presented taking account of their developmental stage. A helpful text[9] gives guidance on religious and cultural differences in the conceptualization of death.

Developmental issues

Young children react to the absence of a parent by developing an anxiety or depressive reaction, often expressed somatically (regression

in acquired control, anorexia, insomnia), but young children cannot distinguish temporary from permanent loss[3,10]. Research consistently demonstrates that children ordinarily do not develop a full understanding of the concepts of death before the age of 7 years, although younger children of 4 years and above can understand it with appropriate help.[11]

Pre-pubertal schoolchildren can be helped more easily to comprehend the reality of death, especially if they are given an opportunity to see for themselves the cessation of function. In cultures where viewing the body is the norm, there may be fewer misconceptions about death among children, but this should not be undertaken where the body is mutilated.[12] Although difficult to substantiate scientifically, clinical literature suggests that attending the funeral helps the grieving process.[5]

For adolescents the death of a parent may come at a time when they are freeing themselves from dependence and may have been in conflict with the parent who subsequently dies, leaving the young person with feelings of guilt and anger. Suicidal feelings are more likely to be acted upon if part of a depressive reaction. Adolescents are more able to sustain sad affects and express grief directly, but they may also react with behavioural and academic difficulties.

The reader is referred to Dyregrov for a more comprehensive description of common reactions to bereavement in childhood.[13]

Children and adolescents with learning difficulties may be at higher risk for developing psychological problems following bereavement, because of their cognitive difficulty in understanding the components of the concept of death and because of their greater dependency.[14] Everatt and Gale provide a helpful review of the available research and draw implications for bereaved children with learning disabilities.[15]

Traumatic bereavement

As with adults, children who witness horrific events involving the death or severe injury of people close to them, or upon whom they are dependent, are at risk of developing post-traumatic stress disorder (see Chapter 9.3.2). Traumatic symptomatology can impede the resolution of grief through mourning as for mourning to proceed, the child has to summon up an image of the dead person. However, if when she/he tries to imagine the deceased, a frightening picture appears or she/he experiences again the helplessness or terror he felt at the time of the death, she/he will tend to avoid recalling the person and thus will not be able to grieve for her/him. Similarly, children whose parents die through suicide or homicide are more likely to have difficulties. In such cases, not only is the nature of the death traumatic and more difficult to make sense of, but there is often an unhelpful media interest and, social support systems that would ordinarily be available may find the circumstances of the death unbearable. If there is a body at all it may be disfigured or its release to the relatives may be delayed the investigation or the authorities and mementoes or suicide notes may be retained by the authorities. Children who have been traumatized by experiencing the sudden, violent, or horrific death of someone close to them are unlikely to benefit from bereavement counselling or therapy until the post-traumatic stress symptoms have been treated[16,17].

Other losses

Much of our knowledge of the impact of bereavement on children and young people is drawn from research on children bereaved of a parent. Other losses have been less well studied; however reactions may be similar depending upon the relationship between the child and the deceased. The death of a grandparent, particularly if he or she lived with the child or carried out caretaking functions, can be devastating to child and parents. Sibling death carries a high morbidity for the survivors, but this can be mitigated by preparation for the death when possible and by participation in community rituals.[18] Adolescents losing a sibling often deny the finality and universality of death, even when these concepts are well established prior to the death.[19] The losses of friends, of pets, or of homes, whilst eliciting sadness, are less likely to provoke pathological grief reactions provided that the child is supported by parents and other adults who are not themselves withdrawn in grief. However, adolescents are affected by the suicide of a friend. A controlled study found that there was a higher incidence of depression than in a matched population sample, although the incidence of attempted suicide was no higher.[20]

Evaluation of treatments

Many of the adverse sequelae of childhood bereavement can be modified or prevented by an intervention before the death or shortly afterwards. In a controlled study, a brief family intervention 2 months after the death of a parent significantly reduced children's morbidity at 1 year post-bereavement. The differences between the treatment and the control group were no longer significant at 2 year follow up, but some of the more affected children had been lost to follow-up, making comparison difficult. But even if by 2 years the effect of the intervention is no longer significant, there is an argument for intervening to relieve symptoms and reduce suffering in the short and medium term.[21] Schut & Stroebe's most recent review concludes that, although in the general adult population a bereavement intervention is more effective for those with more complicated grief reactions, 'children are likely to be a special case, perhaps benefiting from primary intervention.' (A primary intervention is one which is open to all bereaved people rather than targetted at those who are at risk of difficulties such as following traumatic death, or those experiencing complicated reactions.).[22]

Management

Children whose symptoms reach the threshold for a diagnosable psychiatric disorder require a careful clinical assessment to determine the most appropriate treatment, and other sections detail the appropriate interventions for disorders such as depression (Chapter 9.2.7), anxiety (Chapter 9.2.6) or Post-traumatic Stress Disorder (PTSD: Chapter 9.3.2). Some studies have found that parents report fewer symptoms in their bereaved children than the children do themselves; this means that it is important in research and clinical practice to interview the children individually if possible.[5]

Children bereaved of a carer urgently need to be looked after and will transfer their attachment to a new caretaker who is available for them and responsive to them. Supporting a widowed parent in his or her grief, and enabling the process of mourning to occur by providing practical help (child care, financial advice, etc.), may be as important as counselling in helping the children. It may be also appropriate to offer support, supervision and guidance to other adults involved in the child's life such as teachers and religious leaders.

The therapeutic elements of appropriate interventions include the promotion of communication within the family about the dead

parent, the promotion of mourning through reminiscing, the appropriate expression of feelings, and making sense of the death. An overview of techniques used directly with children and young people is provided by Stokes.[23] Techniques for use with children, many of which can be done in groups, include:

- Using art and story-telling
- Writing letters to the deceased
- Creating memory boxes to store reminders of the deceased
- Rituals (such as lighting candles, releasing balloons)
- Making something that represents different aspects of the person (e.g. salt statues of different colours)
- Playing games which encourage children to open up
- Role playing

As with all direct interventions with children, the choice of which techniques to use depends upon the child factors such as age and intelligence, the therapist's or counsellor's training, skill, and experience, the nature of the therapeutic relationship and the organizational context. These interventions can be provided by carefully selected, well-trained and well-supervized volunteers. As part of an intervention, young children may require help to understand what has happened to the deceased by offering a careful and sensitive explanation of what death means in straight-forward biological terms ensuring that the child understands what they are being told. They may also need help to recognize, understand and cope with sad affects both in themselves and in the surviving family members.

Given the indications that problems may develop much later, a useful intervention strategy should include follow-up appointments after any time-limited intervention. Children who have been prepared for the death of a family member have been shown to fare better, in terms of anxiety levels, than those who have not.[24]

Conclusion

Bereavement in childhood, particularly the loss of a parent, represents a significant adversity, although the majority of bereaved children do not develop anything other than transient symptoms. Nevertheless, there is evidence that a brief preventive intervention can reduce subsequent morbidity. Children, who lose a parent through suicide, homicide, accident, or disaster, especially if they have witnessed the death, are at high risk of developing post-traumatic stress disorder and other psychiatric disorders and their treatment needs should be assessed by mental health professionals.

Further information

Dyregrov A (1991). *Grief in children: a handbook for adults*. London, Jessica Kingsley Publishers.

Cohen JA, Mannarino AP, Deblinger E (2006). *Treating Trauma and Traumatic Grief in Children and Adolescents*. New York, The Guilford Press.

Various useful resources for professionals, parents, carers and bereaved young people and children are provided by the UK child bereavement charities: The Child Bereavement charity (www.childbereavement.org.uk) and Winston's Wish (www.winstonswish.org.uk).

Help is at Hand: A resource for people bereaved by suicide and other sudden, traumatic death. This booklet is published by the National Health Service in the UK, and provides particularly good advice for parents of suddenly bereaved children. www.dh.gov.uk/assetroot/04/13/90/07/04139007.pdf

References

1. Parkes, C.M. (2006). *Love and Loss: The Roots of Grief and its Complications*. London and New York, Routledge.
2. UNICEF (2006). *Africa's Orphaned and Vulnerable Generations: Children Affected by AIDS*. Geneva: UNICEF, UNAIDS and PEPFAR.
3. Bowlby, J. (1980). *Attachment and Loss*. London, Hogarth Press.
4. Rutter, M. (1972). *Maternal Deprivation Reassessed*. Harmondsworth, Penguin.
5. Dowdney, L. (2000). Annotation: Childhood Bereavement Following Parental Death. *Journal of Child Psychology and Psychiatry,* **41**, 819–30.
6. Harrington, R. (1999). Unproven assumptions about the impact of bereavement on children. *Journal of the Royal Society of Medicine,* **92**, 230–3.
7. Bifulco, A., Harris, T., and Brown, GW. (1992). Mourning or early inadequate care? Re-examining the relationship of maternal loss in childhood with adult depression and anxiety. *Developmental Psychopathology,* **4**, 433–49.
8. Lau, J.Y.F., Rijksdijk, F.V., Gregory, A.M., *et al.* (2008). Pathways to childhood depressive symptoms: The role of social cognitive and genetic risk factors. *Developmental Psychology*, **43**, 1402–14.
9. Parkes, C.M., Laungani, P., and Young, B. (eds.) (1997). *Death and bereavement across cultures*. London, Routledge.
10. Furman, E. (1974). *A child's parent dies*. New Haven, CT, Yale University Press.
11. Slaughter, V. (2005). Young children's understanding of death. *Australian Psychologist,* **40**(3), 179–86.
12. Cathcart, F. (1988). Seeing the body after death (editorial). *British Medical Journal,* **297**, 997–8.
13. Dyregrov, A. (2008). *Grief in children: a handbook for adults (2nd edn)*. London, Jessica Kingsley Publishers.
14. Kloeppel, D., and Hollins, S. (1989). Double Handicap: mental retardation and death in the family. *Death Studies,* **13**, 31–8.
15. Everatt, A., and Gale, I. (2004). Children with learning disabilities and bereavement: A review of the literature and its implications. *Educational and Child Psychology,* **21**(3), 30–40.
16. Cohen, J.A., Mannarino, A.P., and Deblinger, E. (2006). *Treating Trauma and Traumatic Grief in Children and Adolescents*. New York, The Guilford Press.
17. Harris-Hendriks, J., Black, D., and Kaplan, T. (2000). *When Father Kills Mother. Second edition*. London, Routledge.
18. Pettle, M.S., and Lansdown, R. (1986). Adjustment to the death of a sibling. *Archives of Diseases of Childhood,* **61**(3), 278–83.
19. Hogan, N., and DeSantis, L. (1996). Adolescent sibling bereavement. In *Handbook of adolescent death and bereavement* (eds. C. Corr, D. Balk), pp. 173–95. New York, Springer.
20. Brent, D., Perper, J., Moritz, G., *et al.* (1992). Psychiatric effects of exposure to suicide among the friends and acquaintances of adolescent suicide victims. *Journal of the American Academy of Child and Adolescent Psychiatry,* **31**, 629–39.
21. Black, D., and Urbanowicz, M.A. (1987). Family intervention with bereaved children. *Journal of Child Psychology and Psychiatry,* **28**, 467–76.
22. Schut, H., and Stroebe, M. (accepted for publication). Interventions to Enhance Adaptation to Bereavement: A Review of Efficacy Studies. *The Lancet.*
23. Stokes, J.A. (2004). *Then, now and Always*. Cheltenham, Winston's Wish.
24. Rosenheim, E., Reicher, R. (1985). Informing children about a parent's terminal illness. *Journal of Child Psychology and Psychiatry,* **26**, 995–8.

9.4

The child as witness

Anne E. Thompson and John B. Pearce

Introduction

In the last 20 years, many societies have paid greater attention to children's rights and the importance of protecting children from abuse. As perpetrators of abuse have been tried in court, so more children have been called as witnesses. From being described as 'the most dangerous of all witnesses', children have become recognized to be able to provide valuable and credible testimony in the correct circumstances. Many jurisdictions are now making allowances for children so that their testimony can be delivered in court as fully and accurately as possible. It is no longer tenable to dismiss the capacity of a child to be a witness in court simply because of their age. **Children may be less reliable, as reliable, or more reliable than adult witnesses, depending on a variety of developmental and environmental factors**.

Developmental considerations for children as witnesses

Memory is immature below the age of 12

In the first 3 years of life recall is via preverbal memory (sometimes called eidetic memory), which allows early events to be recalled visually but not in words. Preverbal memory is very accurate but is largely lost around the age of 3 years as language develops. This form of memory remains longer in children who have delayed language development. The loss of this early form of memory explains why **adults do not usually have conscious memories from the first 3 years of life**. It is only around the age of 3 years that experiences begin to be memorized in 'explicit memory', which is accessed by verbal recall.[1] Memories laid down in explicit memory are organized according to hierarchical cognitive structures formed by the child's past experiences of the world. In young children, this organization is rudimentary. This means that new memories are stored with little selection or adaptation to reflect preconceptions, and **in younger children the retrieval of stored memories from poorly organized mental representations is difficult**. External prompts and cues help children recall more fully but in court witnesses are not generally allowed to be 'led' by questioning. Younger children are more likely than older children to forget information over time. **By the age of 12 years, children are generally considered to have the same capacities to lay down memories and recall information as adults**.

Surprisingly, children's **immature memories can sometimes actually improve the quality of information provided in witness statements**. For example, events may be memorized without being influenced by the prejudices that affect adult perceptions, and seemingly trivial details may be memorized by a child whose primitive cognitive schemas allow incoming sensory information to be memorized unselectively.[2] However, cognitive **immaturity more often acts as a barrier to children giving their testimonies fully and accurately**, especially when they are asked to talk about what they remember, rather than being allowed to communicate by behaviour or play.

Children's expectations of adults in conversation influence testimony

By the age of 6 years, many children are using the syntax and grammar of adults in their spoken language. However, their vocabulary is still limited and they are **easily confused by sophisticated or complex speech**. A particularly important aspect of language development for child witnesses is the role played by children and adults in conversational partnerships. **Children are used to being co-operative partners when talking with adults**. They attempt to please by providing answers to questions, even if they do not know the answer or have not understood the question.[3] Young children rarely answer 'I don't know' to a question they are uncertain about and prefer to give a false 'yes' or 'no'. This willingness to provide an answer is probably encouraged by the frequent experience of being 'tested' by adults who already know the answers to the question (for example, 'Look at this picture! Can you see the duck?' or 'How many sweets am I holding?'). As children expect adults to know the answers to their questions, and if a question is repeated, children may change the answer they gave in the assumption that the questioner feels it to be wrong. **The readiness of young children to please adults in conversation no doubt adds to their suggestibility.**

Young children are particularly suggestible

The suggestibility of children as witnesses has been a major concern in legal arenas. Modern research does not endorse the stereotype of the child witness as being highly suggestible. However, there is clear evidence that both children and adults are prone to suggestion at times, and that **pre-school children are particularly vulnerable to this**. The necessity to avoid suggesting information or answers to a child has led to the development of guidelines for interviewing child witnesses in several jurisdictions.

Immature moral development is not always a problem

According to Kohlberg's seminal description of children's moral development, children below 10 years of age operate with 'pre-conventional morality' and evaluate events according to whether the child themselves will gain reward or avoid punishment. Only after this age do children develop 'conventional morality' and begin to be motivated by the approval of other people and society. Therefore, **only older child witnesses have a full understanding of their moral obligations in court**. However, young children, who may hold concrete worldviews such as 'bad people must be punished' as moral imperatives, may be strongly motivated to tell the truth in court.

Children do not lie more than adults, but their lies are more easily detected

Contrary to popular belief, children above the age of 3 years have no more difficulty than adults in distinguishing fact from fantasy. Neither do children tell more lies than adults (although children's lies are more often unconvincing and therefore more easily discovered). **Children and adults are motivated to lie for similar reasons**.[4] The two motives for lying, which may particularly influence child witnesses asked to testify in cases of child abuse, are fear of personal recrimination and a wish to protect those to whom a child feels loyal.

Immature sense of time and short attention span influence testimony

Before 8 years of age, children do not generally have a clear sense of time. Young children therefore often have a muddled recollection of the timing of events and they may have difficulty saying how many times an event occurred. Similarly, **children below this age have short attention spans**, and readily become bored or overwhelmed by prolonged questioning. Both of these developmental limitations can cause difficulties in the preparation of witness statements when a child appears in court.

Traumatic memories may be recalled with more or less clarity

Many of the events child witnesses are called upon to remember were unpleasant and frightening at the time. These events will have been experienced in a state of high emotional arousal. Clinical experience suggests that **emotional arousal can either enhance or diminish recalled information**. For example, extremely traumatic events such as watching a parent being killed can be remembered by child witnesses in a series of highly accurate and detailed visual images that persist in memory over time. By contrast, some children process potentially overwhelming experiences using a variety of psychological defense mechanisms, which limit the amount and accuracy of material available in explicit memory. The psychological trauma associated with the witnessed event and the emotional state of the child during subsequent recall are both likely to have an influence on a child's capacity to give evidence in court.

Table 9.4.1 Guidelines for interviewing child witnesses

- Allowing a child to talk freely without being questioned maximizes the chances of accurate recall
- Questioning children may elicit additional information but lessen accuracy
- Accuracy is greatest when children give their statements as soon as an event occurred
- Children's accounts are most likely to be accurate when they tell the story for the first time
- Children are less likely to give a full account if they feel under pressure from the interviewer
- Younger children may attempt to please an interviewer by providing information even if it is untrue

Environmental considerations for children as witnesses

Skilled interviewing is essential

Because of their developmental immaturity, child witnesses face many disadvantages in the legal system, but these can be substantially offset by a skilled interviewer. Not only must an interviewer facilitate a child to say as much as he or she can remember, but the interview must be conducted so that its process and content will be considered to be acceptable evidence by the court.[5] Psychological research guides the practice of interviewing child witnesses. Some principles are outlined in Table 9.4.1.

Helpful adjustments to legal standards and courtroom process

Many jurisdictions have now made allowances for the developmental and emotional needs of children appearing in court as witnesses. Some legal systems have **relaxed their rules of evidence** where children are concerned so that hearsay evidence (in other words, reports from other people about what a child said or did) may be admissible and lawyers may be allowed to ask child witnesses leading questions. The detrimental effect of extreme stress on children's abilities to give accurate and credible evidence in court is well recognized. A variety of actions have therefore been taken by courts to make a child witness's appearance in court less stressful. The **courtroom may be rearranged to be less formal** and professionals may not wear gowns or wigs. The public may be excluded from courtrooms. Means of **preventing the child from having to face the accused** such as the use of screens or a live video-link to the child in a separate room may be used. Although many professionals agree that children generally find giving evidence by video-link less stressful than giving their entire evidence in open court, there is concern that the child's testimony may have less impact on a jury when viewed on video.[6]

Child witnesses need extra support in the courtroom

There is no doubt that **children find appearing in court a stressful event**. Typical concerns of child witnesses are shown in Table 9.4.2. Feelings of anxiety, confusion, humiliation, embarrassment, and the fear of retaliation or of not being believed

Table 9.4.2 Children giving evidence are often concerned about[7]

- People shouting at them in court
- Not understanding the questions
- Not being believed
- Giving 'wrong' answers
- Speaking in front of strangers
- Crying while giving evidence
- Needing to go to the toilet

(Reproduced from Hamilton, C. Working with young people: legal responsibility and liability, pp. 102–6, copyright 2005, The Children's Legal Centre.)

are common. The stress of appearing in court may be a further trauma for a child who was first traumatized by witnessing or experiencing the alleged crime in question. Some children report that although giving evidence was stressful, they also derived some satisfaction from playing their part in bringing a perpetrator to justice.

Anxiety created by the unfamiliarity of the surroundings and a lack of knowledge about what is happening in the courtroom can be addressed by **preparing child witnesses for their appearance in court**. Preparation often involves visiting the courthouse, receiving age-appropriate written information about the court process, and having the opportunity to ask questions. Ideally, the professional who prepares the child should be available to attend court with the child on the day of the trial. Some helpful information for child witnesses is shown in Table 9.4.3.

Receiving therapy prior to being a witness

Many child witnesses have been traumatized by witnessing or experiencing the alleged crime to which they will testify. Some of these children will develop emotional or behavioural problems as a result of the trauma and will be referred to child and adolescent mental health services. Treatment of trauma-related symptoms usually involves recounting the past experiences in a therapeutic setting. At this point **a conflict of interests arises between the needs of the child as a patient and as a witness**. The psychotherapeutic treatment of traumatized children centres on eliciting the child's subjective truth by the therapist using a variety of means to encourage communication. Within the legal system, the child's recollection of events must be examined in a neutral setting to determine the objective truth, while the rights of both the accused and the witness are protected. Therapy is seen as potentially detrimental to child witnesses because of the potential for their recollections to be altered by repetition or suggestion. The very fact that a child needs psychiatric help may be used to discredit the child as a witness in the eyes of the jury.

If a child witness clearly requires psychological or psychiatric treatment and cannot wait for many months until the trial is over for the treatment to begin, mental health professionals should discuss the child's needs with a representative of the legal service. **The therapy may be allowed to proceed with some restrictions about what can be discussed**. The therapist should be prepared to be called to court themselves in order to give evidence about the nature of their work. **The court will want to be satisfied that the child has not been coached by the therapist or told about information given by other witnesses**.

Table 9.4.3 A child appearing as a witness in court should know that[7]

- They have not done anything wrong
- They should always tell the truth
- They can take time to answer a question
- They should speak to the judge as clearly as possible
- Its OK to answer a question by saying 'I don't know' or 'I don't remember'
- Its OK to answer a question by saying 'I don't understand'
- They must not guess or make up an answer

(Reproduced from Hamilton, C. Working with young people: legal responsibility and liability, pp. 102–6, copyright 2005, The Children's Legal Centre.)

Children's testimony is worth hearing

The status of child witnesses has improved considerably in the last 20 years. Modern psychological research shows that although most children under 3 years of age lack the cognitive capacities to be competent witnesses, many older children are able to produce useful evidential information provided they are questioned competently. To make full use of what children can remember, they need to be allowed to talk in a comfortable setting, guided by professionals who are sensitive to developmental issues and aware of legal constraints. Child witnesses have often been traumatized by their experience. It is important that further distress caused by appearing in court is kept to a minimum and that children who need therapeutic help to deal with their trauma are not denied access to this in the pre-trial period.

Further information

www.childrenslegalcentre.com
www.victimsupport.org.uk
http://www.homeoffice.gov.uk/justice/what-happens-at-court/being-a-witness
http://www.homeoffice.gov.uk/documents/achieving-best-evidence/guidance-witnesses.pdf

References

1. Fundudis, T. (1997). Young children's memory: how good is it? How much do we know about it? *Child Psychology and Psychiatry Review*, **2**, 150–8.
2. Terr, L.C. (1986). The child psychiatrist and the child witness: traveling companions by necessity, if not by design. *Journal of the American Academy of Child Psychiatry*, **25**, 462–72.
3. Ceci, S.J. and Bruck, M. (1993). Suggestibility of the child witness: a historical review and synthesis. *Psychological Bulletin*, **113**, 403–39.
4. Flin, R. and Spencer, J.R. (1995). Annotation: children as witnesses—legal and psychological perspectives. *Journal of Child Psychology and Psychiatry*, **36**, 171–89.
5. Saywitz, K. and Camparo, L. (1998). Interviewing child witnesses: a developmental perspective. *Child Abuse and Neglect*, **22**, 825–43.
6. Williams, G.A. (1998). Video technology and children's evidence: international perspectives and recent research. *Medicine and Law*, **17**, 263–81.
7. Hamilton C. (2005). *Working with young people: legal responsibility and liability*, pp. 102–6. The Children's Legal Centre, Colchester.

9.5

Treatment methods for children and adolescents

Contents

9.5.1 Counselling and psychotherapy for children

John B. Pearce

Introduction

There is a remarkable lack of high quality research to support an evidence base for counselling and psychotherapy for children. And the words 'psychotherapy' and 'counselling' are so non-specific that they should always be clarified in more detail. Nevertheless, these approaches are used frequently in child mental health. While most psychotherapeutic approaches are based on work with adults it is important to note that there are marked differences between children and adults. In spite of these obvious differences, psychotherapy for children is usually based on techniques used for adults. However, psychotherapy that may work perfectly well for adults has to be modified to accord with the developmental level of each child.

Definitions

We each have a mental image of 'a child'. Often this is a stereotypical child aged about 5 to 10 years old. But the word 'childhood' covers the whole period from birth to adulthood, and of course every adult is also somebody's child. In this chapter **the term 'child' will be used to refer to anyone who is not an adult**, but who has matured sufficiently to develop a clear concept of themselves as individuals and of the nature of the real world around them. The ability to distinguish fact from fantasy is an important prerequisite for psychotherapy. This develops as a gradual process with an important stage at around 2.5 years of age when children normally start to refer to themselves as 'I' for the first time. Another stage occurs around 7 to 8 years of age when children develop a clear understanding of time and of the real world. If the therapist ignores these developmental issues it is likely that treatment will be harmful rather than helpful.

Psychotherapy is a very general term that implies treatment of mental dysfunction by psychological rather than physical methods. The aim is to improve function by changing cognition and emotions through the therapeutic relationship, by means of language, play, art, or drama. Dynamic child psychotherapy can be defined as a highly specialized technique where the primary aim is to explore a child's conscious and unconscious thoughts, feelings, and conflicts in such a way that inner resources become strengthened and enabled. It is child-led so that the child is able to follow and explore his or her own agenda, thus helping the child to make sense of the world and to find his or her own solutions to problems and dilemmas. Therapy is mediated by language, which can either be verbal or non-verbal and may use play or creative activities such as drawing, painting, and modelling. Counselling children is very similar, but the therapist usually takes a more passive role than in psychotherapy and would not be so concerned with the interpretation of

unconscious processes. Cognitive behaviour therapy on the other hand is a highly structured approach focused on challenging false cognitions in order to change behaviour and emotions. It is an approach that can be rather easily adapted to children and made sufficiently enjoyable to engage their interest and cooperation.

Differences from adult psychotherapy

A number of **interesting paradoxes and dilemmas** occur when treating children with psychotherapy (Table 9.5.1.1). For example, who should give consent for treatment. Should it be the child, a parent, both parents, or all three? Clearly, this depends on the age and understanding of the child, and each case should be approached in a way that puts the child's needs first. It is generally best to obtain consent from the child and both parents. Any other arrangement is likely to lead to problems at some stage. Psychotherapy and counselling are traditionally non-directive and patient-led, but children, unlike most adults, need to be given some direction otherwise they become easily lost and confused. They cannot be expected to find their own solutions without guidance and support. Most psychotherapeutic approaches for adults are based on coming to terms with and finding explanations for problems that are rooted in the past. However, children are still busy making their past, and their main focus of concern and interest is the present and the immediate future. A further dilemma in child psychotherapy concerns the management of the transference relationship between therapist and patient, which is a reflection of the parent–child relationship. A high degree of trust has to be established to use transference effectively. At the same time, it could be argued that it is not really appropriate for young children to develop high levels of trust and dependence on a therapist whom they only meet briefly in very artificial circumstances. Thus any interpretation of the transference relationship in child psychotherapy must be done carefully and with a good understanding of the subtle complexities of a child's dependency on the parent.

Adult psychotherapy is usually based on a single theoretical model that explains mental mechanisms. Children, however, benefit from the freedom to experiment with a number of different models of their inner world and to learn how to use these ideas in a flexible and constructive way. The use of a single-theory therapy in child psychotherapy is best avoided.

Counselling and psychotherapy

There are undoubtedly differences between child psychotherapy and counselling, but they are difficult to define precisely. This may be because there is a continuum of therapeutic interventions, from advice and guidance at one end of a therapeutic spectrum through more specific counselling and psychotherapeutic techniques to intensive child psychoanalysis at the other end. Counselling children requires a high level of skill, but less theory and technique than in psychotherapy, and it is focused primarily on normal reactions to abnormal events.

Psychotherapy is directed more at psychopathology than normal reactions to stress. It is therefore essential to know about the normal range of children's responses to life events. For example, a 5-year-old child whose mother has just died will grieve differently from a 10-year-old child, because at 5 years of age most children have not yet developed a clear concept of death. Grief in a 5-year-old is most strongly influenced by the way the adults around the child react to the death, whereas a grieving 10-year-old child, although responsive to guidance from the adults around, will also have his or her own unique way of coping with grief. As a general rule, the younger the child the more important it is to consider the attitude and mental state of the parents.

Table 9.5.1.1 Differences between children and adults in relation to therapy

Consent for treatment usually given by parent/carer rather than child
Child may nor understand or want therapy
Children need more help with problem solving than adults
Therapy should be relatively more directive and instructive
Problems tend to be rooted in present or recent past
Therapy should be mostly enjoyable for child to benefit
Transference is complicated because children live with and need parents/carers
Different approaches needed at various stages of development
Children are developing and changing quickly so therapy needs to do this too

Other psychotherapies

Dynamic psychotherapy based on the theories of Sigmund Freud and his daughter Anna Freud,[1] Melanie Klein,[2] and others has been the mainstay of individual child therapy. More recently, Virginia Axline[3] adapted the ideas that Carl Rogers[4] applied to counselling (trust, genuineness and understanding) and developed 'play therapy' as a specific technique for children. Subsequently, brief psychotherapy and interpersonal therapy have grown out of the need to update psychodynamic methods. Various forms of cognitive therapy are now increasingly used for children, although they were originally developed for the treatment of adults by Beck.[5] These therapies focus on a problem-solving approach to resolve current issues, rather than on resolving unconscious conflicts based in the past.

Natural emotional healing

Counselling and psychotherapy have been used with increasing frequency to help children cope with traumatic events such as death, divorce, abuse, illness, and so on. It is arguable whether this trend is at all helpful. Fortunately, the human psyche is remarkably resilient and there are powerful healing processes that take time, which in most cases achieve a satisfactory result. There are similarities between the way the body and mind respond to trauma and a strong correspondence between the natural healing processes that accompany both physical and emotional trauma. The initial healing process starts with a brief period where no pain or distress is felt whatever the cause of the trauma, and this is often accompanied by disbelief that such a thing could have happened. This first phase of shock and 'denial' is then replaced by the full impact of what has happened and is accompanied by high levels of physical or emotional pain. During the second phase the pain may be so severe that it interferes with everyday life, but this stage is usually over within 2 weeks. In the third stage, the healing process continues for a period of up to 6 weeks when the emotional or physical wound is normally healed sufficiently for the traumatized person to be able to return to everyday life, albeit with continuing pain and discomfort at times. The final phase of the healing process then continues over the next 6 to 12 months, leaving a scar that will always remain.

Routine counselling following traumatic events carries the risk of interfering with this normal healing process. Psychotherapy could also be misused to check that all is well, rather like opening up a wound unnecessarily, which will only serve to delay the healing process and might even introduce a secondary 'infection'. The parallel between physical and emotional healing provides some guidelines as to when and how counselling and psychotherapy should be used, as well as the dangers that can occur when they are misused.

The use of play in therapy

The fact that **children are still developing language and communication skills** means that play is often useful as a method of communicating in a therapeutic way with children. Play is essential for normal development. It helps children to develop a repertoire of responses and encourages behavioural flexibility. The main developmental stages of play need to be appreciated if it is to be used in an effective way in either counselling or psychotherapy (see Table 9.5.1.2).

Children will play with almost anything, and so the choice of play materials for use in psychotherapy requires careful thought. It is best to select toys that encourage imagination and creativity. It is important to remember that the type of play equipment provided will actually constrain the way the child plays. For example, there is a limit to what cars or toy animals can do. Similarly, a set of family dolls may have their usefulness restricted if there are not enough of them to represent all the key figures in the child's life. A thoughtful and appropriate choice of toys and play materials including drawing, modelling, and painting equipment can make the difference between the success and failure to engage a child in therapy.

Children's play can be a pointer to what is going on in the child's mind (either conscious or unconscious), but it can also mislead. A child's play will only give a very general indication of what the child thinks and feels and will not provide precise information. Perhaps the best way of viewing the use of play in therapy is that it is an aid to communication—and no more than that.

The treatment setting

There are a number of steps that can be taken to make the treatment setting more beneficial for the child and help to create a relaxed atmosphere. How the therapy and the therapist are introduced to the child is of some importance. The attitude of the therapist is undoubtedly more important than the way the therapy room is organized and the more anxious and tentative the introduction, the more uneasy the child will be. There are no hard and fast rules about the duration or frequency of the treatment session. It is best to arrange it to suit the child. Some children find more than 20 min with an adult extremely difficult to cope with whatever their age. Many children find the formal structure of therapy quite stressful and it may help to put the child at ease by talking about unimportant issues before the session itself begins. Anxiety can also be reduced by organizing the sessions to be as predictable as possible in place and time. Making it clear to children that the therapeutic time is specially for them and that they are the focus of all their therapist's attention can also motivate children to be more co-operative and to be less wary.

Table 9.5.1.2 The development of children's play

Age	Type of play
12–18 months	Symbolic play (e.g. block of wood is a car)
18 months–3 years	Imaginary play (e.g. blocks become a family)
3–7 years	Imaginary friends in 30 per cent of children
5–7 years	Rule-governed make-believe games
7 years onwards	Imagination founded much more strongly on the real world

The treatment process

Each child's experiences of distress must always be considered in the context of the child's family circumstances, the child's stage of development, and his or her temperamental characteristics. It follows that the therapeutic approach needs to be individually tailored and adjusted for each child. Nevertheless, it is possible to arrive at some general guidelines that will assist in treatment. The concern, interest, and supportive attitude of the therapist is central to the treatment process, since it is the interaction between the child's need and the adult's response that establishes the transference relationship.

The start of the first session is particularly important because it sets the tone for the future treatment. The child should know the role that the therapist has and what the aims of the treatment are. Something needs to be said about the limits to behaviour in the session and the nature and degree of confidentiality. It is important to remember that complete confidentiality cannot be assured. For example, the therapist may be presented with information that concerns the health or safety of the child as in the case of abuse, where information has to be disclosed for the benefit of the child.

The timing and number of sessions needs to be agreed at this stage. Therefore it is often a good idea to suggest a limited number of sessions in the first instance, together with an agreement to review whether or not further sessions are required. This type of introduction helps children to feel their needs are being taken seriously, especially if they are actively involved in the process. Sessions should start and finish on time and the links between the present and any previous sessions should be clarified.

Most children will remember their treatment experience for many years to come—if not for the rest of their lives. The child who feels uncomfortable, embarrassed, and misunderstood is likely to retain a memory that is painful and unhelpful. On the other hand, if the therapeutic experience was positive, where the therapist was seen as supportive, encouraging, and understanding, the memory is likely to be one that the child returns to again and again for emotional strength and support.

Different approaches to psychotherapy

There is a wide variety of jargon associated with various types of psychotherapy. Each approach has its own 'language' and associated special techniques. However, there is no evidence that any one method is better than another, and it would appear that the personal preference of the professional involved is more important in determining which approach is used rather than the characteristics

of the child. Whichever approach is used it is likely that there will be the same common themes in the focus of treatment. Common themes include dealing with feelings of anxiety and insecurity, difficulties in relationships, low self-esteem, and a feeling of failure. These emotions are often generated by difficulties with aggression, jealousy, sexuality, and death.

A focus on the past versus the present

The child comes to psychotherapy with a range of problems rooted in the past. A decision has to be made whether to focus the therapy on trying to understand and come to terms with the past, or to consider how a child might best cope with what is actually happening in the here and now. **The danger** in concentrating primarily on the past is that it may interfere with the child's ability to cope with the present and plan for the future. While it is important to learn from what has happened in the past, children tend to learn more from what is happening in the present in their daily lives. Understanding how their emotional stress was generated in the first place may not lead to a resolution or to a greater ability to deal with current problems. It is generally helpful to start therapy with an acknowledgement of what has happened in the past and a consideration of how that might affect what is happening in the present. In a few cases it may be helpful to focus more on the past, but only if the child has clearly become preoccupied with a particular issue from the past and is unable to move on. Normal development moves on so rapidly during childhood that any fixation with the past can have serious consequences, thus every effort needs to be made to promote and sustain developmental progress.

Theory vs. common humanity

While it is undoubtedly helpful to have a theoretical framework within which treatment can take place, the observation that different psychotherapeutic approaches for the same type of problem can be equally effective suggests that the precise theoretical framework underpinning treatment may not be that important. The basic human qualities of kindness, trust, caring and understanding are perhaps the most important qualities in psychotherapy. It is not an unusual experience for therapists to find that their early cases turn out to be the most successful, which is probably due to the enthusiasm and therapeutic optimism of the new therapist. It is clearly important to hold on to these therapeutic qualities as one becomes more experienced.

Supportive counselling vs. in-depth psychoanalysis

It is easy to assume that the more intensive the psychotherapy and the more it explores the deep unconscious world, the more effective it must be. Clearly there is no reason why this should be the case. For example, one would not expect a surgeon to cut deeper for greater effect or a physician to prescribe more medication than is necessary. This would only increase the adverse effects of the treatment. It is not difficult to see that regular psychoanalysis two or three times per week could be quite disruptive to family life merely as a result of the time commitment alone. There are also other potential problems for children who are treated with intensive psychoanalysis over a period of years, as this may delay or shape a child's development in an unhelpful way. On the other hand, it might be equally inappropriate to commence supportive counselling for a child who is deeply disturbed and whose need for loving care and protection is not being met. These primary and basic needs must always be given priority.

Cognitive therapy vs. psychoanalytically based psychotherapies

There has been an increased interest in cognitive therapy and cognitive behaviour therapy, where the emphasis is much more on the here and now and the behavioural consequences of abnormal thought patterns. The techniques used in cognitive therapy for anxiety and depressive disorder are described in Chapters 6.3.2.1 and 6.3.2.3 respectively. The only modification that is required for their use in children is to adapt them to the developmental stage and the level of cognitive ability that the child has reached. Cognitive therapy has a theoretical advantage for use in children in that its focus is more on the present and the future, in contrast to most psychoanalytically based psychotherapy. Its approach is strongly based on learning new ways of coping. Cognitive therapy is pragmatic and active rather than passive and reflective, making it generally more appropriate for the needs of younger children. Unfortunately, there is as yet limited evidence to support the theoretical underpinning of the various cognitive models of childhood disorders.

Limit setting vs. free expression

It is a common dilemma to know how much freedom children should be allowed to express themselves. Some children appear to enjoy pushing the limits to see how far they can go. Other children appear too inhibited and need encouragement to express themselves. Part of the art of child therapy is to strike a comfortable balance between control and freedom. Children gain nothing from disruptive and destructive behaviour, even if they normally tend to be quiet and inhibited. Indeed, they rapidly develop overwhelming feelings of anxiety and insecurity if they do not feel sufficiently contained. It is essential that the therapist retains a very clear notion of what behaviour is acceptable and what is not. It is therefore the therapist's responsibility to set the scene and establish the boundaries of acceptable behaviour within the therapeutic context. Should the child go beyond the limit then a warning should be given, and if the child persists it is quite acceptable to end the session early or at least until the disruptive behaviour has stopped. The therapist's reaction to bad behaviour should make it quite clear that it is unacceptable. However, the emotional response should be neutral or sad, in much the same way that one might behave in a shoe shop when a desirable new shoe does not fit as expected.

Closeness vs. distance

It is natural for an adult to be physically much closer to younger children and then to become more distant as they grow older. For example, it is quite natural for an adult to hold the hand of a 3- or 4-year old and to physically guide the child. On the other hand, any physical contact with a teenager can easily be misconstrued and is likely to be most unwelcome. In addition to the developmental perspective, every child has its own preferred degree of closeness or distance from other people. The task of the therapist is to judge what is right for each occasion. It is absolutely essential that the therapist must always avoid intruding into the child's space in any way that could be construed as abusive. This may prove difficult where children are unsure of their boundaries and seek out physical contact (in cases of sexual abuse, children may seek out

sexual contact). However, to maintain an artificial physical or emotional distance can be perceived as disinterest or even rejection by some children. To achieve a comfortable level of emotional warmth and physical closeness in therapy is obviously a very important matter, but a relaxed approach in the therapeutic relationship is generally best.

Individual vs. group therapy

There are no agreed guidelines to determine which child would benefit from an individual or from a group approach to psychotherapy. Some children, however, find the emotional intensity of undivided adult attention too much to cope with and learn better from others in a group situation. As the selection of cases for group psychotherapy tends to be determined by the therapist's skills, there is no clear evidence that one approach is better than any other. Nevertheless, there is a growing literature on group therapy for children and an increasing interest in this method of treatment if only because the cost per case is likely to be less.

Practical issues in child psychotherapy

Involving parents and the school

The younger the child, the more helpful it is to involve the parents in treatment. How this is managed will depend to some extent on the resources available. It is generally best for separate therapists to work with the child and with the parents. However, there will usually be occasions when it is helpful for the child's therapist to have some direct contact with the parents as a way of monitoring progress and keeping a link between the therapy sessions and the child's real world. As children grow older they tend to become increasingly inhibited by the involvement of their parents, so it is helpful to check with the young person how they would like the contact with their parents to be organized. All children should be considered in the context of the family and the school because so many risk factors are associated with these environments. Parental support for the treatment is a critical factor in a successful outcome. Equally, it is possible for parents to undermine treatment by their negative comments or overintrusiveness after each session.

The extent to which a child's school should be involved needs to be carefully considered on an individual basis. Although it is best to have the full co-operation of the school, this may not be possible.

Confidentiality and record-keeping

Confidentiality in relation to the individual therapy sessions is clearly important, but it should not dominate. The most vital issue is that the children feel safe in what they say and do during the session. They need to know that there is a reasonable level of confidentiality. However, there are obvious exceptions to the general rule of confidentiality: for example, in cases of child abuse or criminal activity, or in cases where the child may be at risk from harming itself. A suitable compromise is to indicate to the child that their session would be treated as confidential, but there may be occasions when it is best to inform their parents or somebody else about important information that was given during the session. This would only occur after discussion between the therapist and the child—the overarching rule is that it must be in the best interest of the child.

There is obviously a need for accurate record-keeping for each session, partly to keep track of the therapeutic process and also for medico-legal purposes. It is unnecessary to include the more detailed recording of the therapeutic process in the official notes if the record is purely for training purposes. These training notes then remain the property of the therapist and must be kept at the same level of security as the official notes.

Failure to attend

If there is a genuine and legitimate reason why a child is unable to attend for a therapy session then little or no importance needs be attached to this. On the other hand, if therapy sessions are cancelled repeatedly, cancelled with inadequate excuses, or simply not attended it is essential for the therapist to question whether therapy is actually achieving the original stated goal. Factors such as family disadvantage, parental stress, and the severity of the child's problems all increase the likelihood of failing to complete therapy successfully. It is all too easy to either blame the carers or to rationalize failure to attend as being some problem in the child, rather than an issue that could be due to the therapist. Of course, one good reason for failure to attend is that the child no longer needs to. It is usually unhelpful to send out repeated appointments that are not attended. A reasonable compromise is for one reappointment to be made, then if this is failed for a letter to be sent to the carers or to an older teenager asking them to contact you if further sessions are thought necessary. Occasionally, if treatment is at a very critical stage or if there is high level of concern about the child, it would then be appropriate for more effort to be put into arranging a further appointment.

Individual and cultural issues

A child's sociocultural background needs to be taken into account when planning therapy. Temperament, age, gender, and intellectual ability are also important factors to be considered. Ethnicity, on the other hand, is probably not that significant an issue. Some children enjoy, and benefit from, the freedom to express themselves openly and with little constraint; others need direction and structure if they are to resolve psychological problems. There is no point in trying to fit the child to the therapy rather than the other way round. One factor that supersedes all cultural and individual factors is that children generally wish to enjoy themselves and to have a good time. They are pleasure-seeking beings who, unlike most adults, see no benefit from the experience of pain and distress of therapy.

Ending therapy

Bringing therapy to a satisfactory conclusion is more likely to happen if the treatment was well set up in the first place and if achievable goals were agreed. Because it is so easy for therapy to lose its focus, it is helpful to consider how and when the treatment will be concluded at the same time as setting it up. Even if treatment goals have not been achieved it should still be possible to end the therapy on a positive note, identifying areas where self-knowledge has been increased and anything that has been positive in the therapeutic relationship.

Training and supervision

The objectives of training in psychotherapy are primarily to do with increasing knowledge and understanding of the issues that arise during psychotherapeutic treatment, many of which have been referred to above. It is essential for this to be underpinned by a sound knowledge of child development, but less important for

therapy to be founded in any particular psychodynamic theory. The key training method in psychotherapy is the conduct of therapy under supervision. Choosing a supervisor and the role of that supervisor are critically important issues for the trainee therapist, since the relationship between supervisor and trainee mirrors that of the therapist and client even if there is a clear understanding that the purpose of the supervision is not intended to be therapeutic. Most therapists find that supervision continues to be helpful even after they are 'trained'.

Measures of effectiveness and outcome

There is a lack of high quality research to support an evidence base for counselling and psychotherapy for children. One of the main difficulties in conducting research is in the selection of an adequate control group. Reviews of psychotherapeutic treatments indicate a general improvement in children that persists for many months after the intervention. However, it seems likely that the theoretical basis for the psychotherapeutic approach is less important than the caring and supportive relationship that develops between the child and the therapist.

Further information

Fitzgerald, M. (1998). Child psychoanalytic psychotherapy. *Advances in Psychiatric Treatment*, **4**, 18–24.

Jacobs, W.J. (2002). Individual and group therapy. In Child and Adolescent Psychiatry (4th edn) (ed M. Rutter, E. Taylor), pp. 983–97. Blackwell Science, Oxford.

Rutter, M., Bishop, D. Pine, D., *et al.* (eds.)(2008). *Rutters's Child and Adolescent Psychiatry* (5th edn.). Blackwell, Oxford.

Roth, A.D. and Fonagy, P. (2005). What works for whom? A critical review of psychotherapy research. Guilford Press. New York.

Useful web sites

www.youngminds.org.uk
www.psychnet-uk.com
www.babcp.com
www.rcpsych.ac.uk

References

1. Freud, A. (1974). *Introduction to psychoanalysis*. Hogarth Press, London.
2. Klein, M. (1932). *The psychotherapy of children*. Hogarth Press, London.
3. Axline, V. (1969). *Play therapy*. Ballentine Books, New York.
4. Rogers, C. (1951). *Client centred therapy in current practice: implications and theory*. Houghton Mifflin, New York.
5. Beck, A.T. (1976). *Cognitive therapy and emotional disorders*. International Universities Press. New York.

9.5.2 Psychodynamic child psychotherapy

Peter Fonagy and Mary Target

Introduction

Psychodynamic psychotherapy for children is based on a range of assumptions concerning mental functioning that have gradually evolved over the past 100 years out of the theories of Sigmund Freud. As these assumptions have been widely reviewed, we need to provide only a very brief introduction here.

Psychodynamic child clinicians assume that:

(a) The child's presenting difficulties may usefully be seen in terms of thoughts, feelings, wishes, beliefs, and conflicts. This entails the assumption that mental disorders can meaningfully be understood as specific organizations of a child's conscious or unconscious mental states.

(b) To understand conscious experiences, we need to consider non-conscious narrative-like experiences, analogous to conscious fantasies, which powerfully affect behaviour, affect regulation, and the capacity to handle the social environment. Modern neuroscience, with fMRI studies that show cortical response reflecting processing to meaning in the absence of awareness and non-conscious motivation,[1] has put the existence of an unconscious beyond debate.

(c) Intense relationship experiences are represented in the mind as structures of interpersonal interaction, and are aggregated across time before coming to form a schematic mental structure, which is often represented metaphorically as a neural network. Within many models, self-other relationship representations are also considered the organizers of emotion, as feeling states are seen as coming to characterize particular patterns of self-other and interpersonal relating (e.g. sadness and disappointment at the anticipated loss of a person).[2]

(d) Inevitably, wishes, affects and ideas will at times be in conflict with one another. The psychodynamic therapeutic approach sees such conflicts as key causes of distress and the lack of a sense of safety. Adverse environments either increase the intensity of conflict or fail to equip the child with the capacity to resolve such incompatibilities through mental work.[3] They may also set the child on a developmental trajectory in which the normal development of key psychological capacities is undermined, thereby reducing the child's competence to resolve mental conflict.[4] For this reason, while reviewers of psychodynamic psychotherapy often contrast conflict and development-focussed approaches, the reality of developmental trajectories means that the conflict and deficit often come together.[5]

(e) The child's mental mechanisms for dealing with intrapsychic conflict include defence mechanisms that distort mental representations in order to reduce conflict and unpleasure.[6] Such self-serving distortions of mental states relative to an external or internal reality are frequently demonstrated experimentally and have become accepted.[7–10] Classification of defences has frequently been attempted,[11,12] often as a method for categorizing individuals or mental disorders,[13,14] but few of these approaches have stood the test of time or achieved general acceptance.

(f) Behaviour may be understood in terms of 'complex meanings', that is, mental states that are not explicit in action or within the awareness of the person concerned. Thus, symptoms of disorders are classically considered as condensations of conflicting wishes together with the failed defence against conscious awareness of those wishes. Therapy is an effort to seek *personal* meaning,[15] and to elaborate and clarify implicit meaning structures—a process that may turn out to be

the essence of psychodynamic psychotherapy—rather than to give the patient insight in terms of any particular meaning structure.

(g) A relationship with a supportive and respectful empathic adult will benefit the young person, not least by enhancing their own understanding and emotional responsiveness. The nature of the relationship with psychodynamic therapists varies across therapies—from the highly transferential and fantasy oriented[17] to the quite practical and supportive,[18] although most therapies contain elements of both.[19] Establishing an attachment relationship with a clinician (i.e. with an interested, understanding, and respectful adult) may be a new experience for some young people[20] and is believed to trigger a basic set of human capacities for relatedness that appears therapeutic, apparently almost regardless of content.[21–27] The child's relationship with the therapist often appears to become the vehicle for disowned aspects of the child's thoughts and feelings, creating a process termed transference, which enables the psychoanalytic clinician to understand the child's representation of relationships and his or her feelings about them.[28]

Background

The roots of child psychoanalysis lie in Freud's observation of young children, most notably of the young Anna Freud's wishful dream for strawberries,[29] his grandson's separation game,[30] and his case study of Little Hans, a 5-year old with a phobic disorder who was treated by his physician father under Freud's supervision.[31]

Play therapy, incorporating both an insight-oriented interpretive approach and the developmental assistance perspective, was introduced by Hermine Hug-Helmuth.[32] Thereafter, two women, in strong opposition but frequently making reference to each other, established the field: Anna Freud[33] and Melanie Klein.[34]

Klein's approach was to regard children's play as essentially the same as free association with adults; that is, motivated by unconscious fantasy and activated by the relationship with the therapist (transference). The child's anxiety required verbalization (interpretation) if it was to be addressed. The focal point of therapy was the verbalization of anxieties concerning destructive and sadistic impulses, whilst the child's external relationships (with parents, teachers, etc.) were seen as peripheral and irrelevant.

A key construct was the notion of projective identification.[35] This term referred originally to the infantile tendency to project unwanted aspects of the self on to another person. The clinician, by understanding the child's perception of her as a person, could gain valuable insights about conflictual aspects of the child's experience of himself. Bion[36] described how the 'container's' capacity to understand and accept the projections could be critical both to successful therapy and to normal development. More recently, Kleinian child analysts have been less likely to offer early interpretations of deeply unconscious material; defences beyond projective identification are more commonly considered.[17,37]

Strongly influenced by Melanie Klein, Donald Winnicott firmly endorsed her emphasis on the impact of the first years of life on childhood psychopathology.[3] However, he also introduced new techniques (e.g. drawing) and various theoretical innovations, including the identification of a transitional space between self and other where the subjective object and the truly objective object could simultaneously be recognized.[38] The notion of transitional space, an intermediate area between the intrapsychic and the interpersonal, was critical to the development of an interpersonal[39,40] and intersubjectivist[41] approach within psychoanalysis.

Along this continuum, Anna Freud was perhaps most concerned with the child's developmental struggle with a social as well as an internal environment. Her background as a teacher may have led her to be as concerned with children's actual external circumstances as with their unconscious worlds.[42] Her focus was restricted to complications and conflicts arising from the child's libidinal impulses and, unlike Melanie Klein, she rarely focussed on innate aggression. The interpretation of defence was central to her technique.[43] Her approach paid careful attention to limitations on the child's cognitive capacities (ego functioning) and had as its explicit aim the restoration of the child to a normal developmental path.[44] Her concern with normal development led her to evolve a model of pathology as a disturbance of normal developmental processes, and she developed a systematic analysis of such anomalies using the concept of developmental lines.[45] Her propositions are in many respects consistent with modern developmental psychopathology.[46] Developmental help is aimed at facilitating the forward movement of the psychological processes that underpin social cognition and interpersonal function, and which include mentalization, impulse control and emotion regulation, symbolization and the use of metaphor, and the capacity for play.[47,48] Notwithstanding the curious historical fact that Anna Freudian, Kleinian, and Winnicottian approaches all originated in London, the Anna Freudian approach came to dominate child therapy in the United States,[49] whereas in the United Kingdom and in Latin America Melanie Klein's approach proved more popular.[50] Two comprehensive and detailed histories of the field have been provided.[51,52]

Techniques

Techniques of child therapy differ considerably depending on the degree of pathology manifested by the child. Two sets of technique may be distinguished: those with single diagnosis, usually involving anxiety, are offered what most would recognize as 'classical' forms of psychodynamic, insight-oriented therapy. Those with multiple diagnoses, severe behavioural problems and/or emergent personality disorders[53] require a different psychodynamic treatment approach. These will be discussed separately.

Principal features of 'classical' technique

Child psychotherapy involves the elaboration of distorted and, to a lesser or greater extent, non-conscious mental representations. The therapist, using the child's verbalizations, non-verbal play, and other behaviours, aims to provide a rational understanding of the child's non-conscious thoughts, feelings, and expectations. This understanding may encompass and integrate earlier modes of the child's thinking into a more mature, age-appropriate framework.[54] With young children, the treatment involves the use of toys, play, and any device that helps to engage the children in a process of self-exploration. The therapist works to elaborate the children's understanding of their emotional responses, their unconscious concerns about their body, and the way their symptoms might link together anxieties about relationships, including non-conscious aggressive or sexual thoughts and other conflictual feelings in relation to the parents, siblings, and peers.

The techniques used by the child therapist go beyond interpretive interventions and were usefully enumerated by Paulina Kernberg.[55] She delineated: (i) *supportive interventions*, which are aimed at addressing the child's anxiety and increasing the child's sense of competence and mastery through the provision of information, reassurance, empathy, and suggestions (ii) *facilitative statements*, which seek chiefly to maintain the therapeutic relationship with the child by reviewing, summarizing or paraphrasing the child's communications and (iii) *clarifications*, which review and summarize the child's communication, and which usually involve relabelling communication or behaviour. Clarifications also serve to focus the child's attention on certain patterns in his behaviour indicative of unconscious determination.

Interpretation may centre on: (i) the content of the child's communications (ii) the contents that the child systematically omits from verbalization (iii) the child's non-verbal behaviour (iv) the nature of the child's play, including the roles that he or she tends to assign to himself and to the therapist (v) the child's current emotional state, particularly sadness, anxiety, or guilt, and; (vi) dreams that the child recounts in his sessions. While the therapist may be able to link the child's therapeutic material to past experiences with attachment figures, such reconstructive interpretations are rare in child therapy. It is only gradually that the therapist hopes to be able to generate an emotionally meaningful understanding of the impact of past experiences on current anxiety and conflict.

Paulina Kernberg[55] distinguished between three types of child therapeutic interpretation. First and most common are interpretations of defences, which aim to show the child how it protects itself from thoughts, feelings, and actions that it considers unacceptable. For example, the therapist may draw the child's attention to repeated examples of self-denigration, and hint at his anxiety about being thought boastful; this serves a dual function in both bringing to the child's awareness what he is protecting himself from and also in prompting him to find alternative strategies to cope with warded off ideas. Second are interpretations that address the child's unconscious wishes, which are themselves thought to underpin behaviour. Frequently, these interpretations are made following interpretations of defences. Finally, child therapists might address the child's past experiences. The therapy may reveal traumatic experiences, and some therapists consider it helpful to bring these memories into consciousness. It should be noted that current psychodynamic theory in no way assumes that addressing such trauma directly is essential to cure. Far more important in terms of therapeutic progress is addressing the distorted relationship representations that are sequelae to early trauma.[56]

Whatever the interpretation, the child therapist aims to address the child's anxiety and how other emotions relate to it. Thus, destructive wishes would most likely be taken up in connection with the child's anxiety about his or her angry feelings. Child therapeutic technique also demands that the child's attempt actively to struggle with these wishes be clearly acknowledged. Interpretations are ideally tied to a highly specific context, such as the child's experience of anxiety associated with his anger that an ungenerous but otherwise valued therapist will not give him a special treat for his birthday.

An important part of child therapeutic work involves the child's parents. Some of this work is psychoeducational; in particular, parents often need guidance on appropriate, uncritical, warm, and playful methods of child rearing. Discussion of the child's symptoms may enable parents to gain greater awareness of the child's difficulties and how their own representation of the child may be distorted.[57]

Psychodynamic technique with complex childhood disturbances

The child psychotherapeutic approach has been extended to apply not only to so-called neurotic disorders, but also to the understanding and treatment of borderline, narcissistic, delinquent, and conduct disordered youngsters, as well as schizoid and even psychotic children.[53] The classical psychoanalytic approach as outlined above has clear limitations with these children: anxiety may not be accessible; there may be little evidence of conflict/of the child's struggle with wishes; defences may be hard to identify, and the child may be developmentally inaccessible to insight. Taking these issues into consideration, we have suggested that a dramatic modification of child psychotherapeutic technique may be in order[19] based on what Anna Freud called 'developmental help'. We have described this intervention in detail,[58] and colleagues in the Netherlands have elaborated and researched this form of therapy.[59]

Essentially, the therapist begins by performing mental functions of which the child is incapable, or by showing the child ways of performing these functions until he or she can take over and do it himself. These interventions are used with pathologies traditionally defined as ego defects, deficiencies in relationships, or developmental disturbances—pathologies understood here as mental process disturbances. These techniques have sometimes been labelled remedial education or ego-support, but, broadly, the therapist's aim is to free the mental processes from inhibition and to aid in the development of these processes. The therapist achieves this by: (i) providing a safe place and relationship within which the child can dare to change or wish to be different (ii) making up for some deficits in the parenting that the child has received by providing him with the missing elements (iii) stimulating delayed or stunted developmental processes by drawing the child's attention to what is missing, and encouraging his interest and desire to function better and (iv) using interpretations not to uncover the source of his difficulties but to help the child understand the extent and impact of his problems, his contribution to his developmental difficulties, and to confront the role played by his environment. The main foci of these modified forms of child therapeutic intervention are six-fold:

(a) *The enhancing of reflective processes*; that is, the understanding of how mental states (beliefs, desires, wishes, and emotions) determine human behaviour. This is achieved by encouraging the observation and labelling of both physical and psychological experiences in the immediate situation. It is assumed that, regardless of cause, the final common pathway of most personality disorders is a dramatic impairment of reflective function.[60]

(b) *The enhancing of impulse control* by identifying and helping the child to exercise ways in which impulses may be channelled into socially acceptable forms of behaviour. The therapist may initially have to control this herself by 'setting limits'. She tells the child what he may and may not do during the sessions, explaining that she will not let him hurt her or himself, or damage things in the room.

(c) *Affect regulation* is an important aim of developmental work. It can be assisted by the verbalization and labelling of affect, and by explaining to the child possible reasons for his feelings; for example, that his aggressive attacks are reactions to his sense of being threatened and endangered.

(d) The elaboration of strategies involving symbolization and metaphor for *enhancing cognitive self-regulation*. The therapist demonstrates her own capacities for reflection and the moderation of experience through mental representation rather than physical action or coercion.

(e) Focussing the child's awareness on *the mental states of others*. This is achieved initially by focussing interventions around the child's perception of the therapist's mental states, which can be a precursor to reflective processes in relation to the self.

(f) Developing the child's *capacity for play*, initially with physical objects, then with another person and, ultimately, with ideas. Play is not simply the creation of a pretend world but has the aim of creating a safe opportunity for alternative meanings of the child's experiences to emerge. This can show the child how his habitual ways of thinking and feeling represent but one of multiple ways of construing reality. Perhaps more importantly in this context, he can experience adults relating to him, as perhaps they have not often related to him before.

While the child therapist working with such severely disturbed children is still 'working in the transference', in the sense that the child's feelings about the therapist remain central, this is no longer thought to entail the displacement of feelings and ideas from one person to another, e.g. from the parent to the therapist. Rather, the clarification of the child's feelings about the therapist may be the most effective route towards assisting the child to acquire a reflective capacity. In this way, the therapist conveys that the child's affect can be understood and managed by another person.

Indications, contraindications, and the selection of procedures

There is general agreement on the indications for child psychotherapy.[61] These have traditionally included: (i) high IQ and verbal ability (ii) supportive environment (iii) conflict-related pathology (iv) adequate internal and external object relations and (v) the presence of anxiety. Contraindications include: (i) pervasive developmental disorder (ii) psychosis (iii) major deficiencies in psychological capacities and (iv) family constellations incompatible with adherence to treatment, for example, chaotic home environments or psychologically severely disturbed parents.

As described above, however, by substantially modifying traditional technique, child therapists have successfully worked with populations beyond this restricted group,[62] treating children with a variety of psychological deficiencies.

One way of conceptualizing the difference between the needs of the two groups is by using the classical distinction between mental representations and mental processes in cognitive science.[63] Classical techniques primarily impact on the organization and shape of the child's mental representations of self-other relationships.[64] By contrast, developmental help for the more severely disordered group aims at developing the function of mental processes, which may have been defensively distorted in early development,[65] by strengthening and supporting the patient's adaptive defences and helping them to label and verbalize their thoughts and feelings.

The distinction between classical technique and developmental help is a heuristic one. In reality, all child therapeutic treatments involve both, but it is nevertheless suggested that developmental help is essential for the effective treatment of severe disturbances, whilst it remains an 'optional extra' for children with neurotic disturbances.

Managing treatment

Starting treatment

At the beginning of psychodynamic child therapy, the therapist's aim is to communicate that: (i) sessions have the purpose of expressing thoughts and feelings through play and words and (ii) the therapist is trying to help the child make sense of his experience so that he can master his inner turmoil in a more effective manner. Children are generally able to develop a therapeutic alliance in the context of an empathic, respectful, non-exploitative relationship with an adult. A similar collaboration is established with the parents, and early meetings also allow the child therapist to acquire relevant information and assess family interactional patterns that may be relevant to the child's treatment. In early sessions, the therapist attempts to interest the child in meaning and to link mental states to activity by using child's play to address the child's understanding of momentary anxieties, and to explain that his struggles, thoughts, and feelings may be explored through his verbalizations and behaviour.[66] Children who are able to use therapy tend to respond to these interventions with an enhancement of the therapeutic alliance, showing greater freedom in their play and verbalizations. It is important to note that the same process of relationship building characterizes other individual therapies such as cognitive behaviour therapy (CBT).[23]

The middle phase

The middle phase of child therapy is expected to focus on the systematic use of transference interpretations and the initiation of the 'working through' process. The development of the transference is facilitated by a therapeutic structure that emphasizes regularity, consistency, and the specialness of the hour of therapy.[67] The child's play is 'interpreted', but normally these interpretations are kept within the context of the play situation. Clinical experience suggests that interpreting the play in reference to the child's 'real' feeling only serves to disrupt the child's communication.[68]

The major themes in children's lives are often expressed in the therapeutic relationship. Internal working models of attachment relationships, for example, will come to colour the child's relationship with the therapist,[69] be that a pattern of reluctance in forming trusting relationships (avoidant/dismissing) or excessive anxieties about separation and sometimes angry preoccupation with the therapist's thoughts and feelings (resistant/preoccupied). This, of course, opens up the possibility of correcting distortions and mastering conflicts and associated anxieties. The child gains information (insight) in a relational context that elaborates his understanding of his experience, provides him with a way of coping with thoughts and feelings that make him feel uncomfortable and interrupts maladaptive interactions.[70] Progress is often slow, and regression and progression tend to alternate as children struggle to face these anxieties without undue repression. Yet, gradually, given a capacity for the symbolic transformation of experience, they are able to bring their experience of internal and external

worlds into play and can therefore internalize the therapeutic attitude to solving their problems. Of course, development is helpful too in this regard.[71]

The ending of treatment

Indications for the termination of long-term child psychotherapy are both external and internal to the therapy. External indications are not restricted to symptomatic improvement, but include considerations of changes in family interaction and peer relationships. Relational changes indicate a capacity to use caretakers, teachers, and others as sources of protection, guidance, comfort, and models of identification. In general, the therapist seeks evidence that the child has returned to the path of normal development, can cope effectively with stress and conflict, and can respond with greater freedom to adaptive demands coming both from within and without.

There are also indications from the child that the therapy may be ended. Some were enumerated by Paulina Kernberg[72] as follows: (i) the therapist finds more opportunities for interpretation than for confrontation (ii) the child manifests more reflectiveness and the capacity to make interpretations (iii) there is greater freedom, expressiveness, and pleasure in play (iv) there is insight and humour and (v) the child assumes responsibility for his own actions.

The end of psychodynamic therapy often generates anxiety and can bring a temporary reactivation of symptoms and the re-emergence of dysfunctional patterns of interaction within the family. As the therapy creates an attachment relationship, its termination may be an essential part of resolving attachment-related concerns (e.g. fear of abandonment, rejection). Mourning the anticipated loss of the therapy and the therapist may well be an essential part of a successful ending to treatment. Other issues may involve children's disappointments—with themselves, with the unfulfilled promise of the therapy, and with adults who do not measure up to their expectations. Research and clinical experience suggests, however, that as this turmoil subsides, progress continues to be made after the end of treatment.[57,73,74]

Efficacy

The most comprehensive survey of outcome studies specifically concerned with psychodynamic treatment was undertaken by Kennedy[75] as part of a project sponsored by the British Association of Child Psychotherapists—although there are other comprehensive reviews of treatment of children[76–80] and of psychosocial interventions limited to 'evidence-based treatments'.[81,82] The general reviews highlight that research of psychodynamic treatment in the field of child therapy has lagged behind the evaluation of other approaches.

There are relatively few randomized controlled trials of psychodynamic psychotherapy.[83–88] All but one of these trials contrasted individual child psychotherapy with another (evidence-based) treatment. In addition, several studies employed quasi-randomized methods of assignment, such as postcode[89] or therapist vacancy.[90,91] Six studies reported on findings with matched comparison groups.[92–96] A further two studies reported non-matched control groups.[97,98] Two further studies used an untreated but poorly matched control sample.[99,100] In addition, there are a number of open trials of child psychotherapy employing no comparison groups.[101–107] Three studies used an experimental, single-case methodology.[108–110]

Considering studies of the common disorders of childhood separately, Muratori and colleagues[91] contrasted the efficacy of 11 weeks of psychodynamic therapy to treatment as usual in 58 children with depression and anxiety. At 2-year follow-up, 34 per cent of the treated group were in the clinical range on symptomatic measures, compared to 65 per cent of the controls. Treatment effects increased during the 2-year follow-up period (the so-called 'sleeper effect'), including a move into the non-clinical range for the average child with internalizing problems (in the psychodynamically treated group only).[90] It is encouraging that psychodynamic psychotherapy patients sought mental health services at a significantly lower rate than those in the treatment as usual comparison condition over the 2-year follow-up period.

In a multi-centred European trial,[88] moderate childhood depression was shown to be accessible to a brief individual psychodynamic psychotherapy. At 7-month follow-up, none of the moderately to severely depressed young people met criteria, which is comparable to children treated with a combination of fluoxetine and CBT.[111] The presence of anxiety/dysthymia signalled particular suitability for individual treatment, whilst comorbidity with oppositional defiant disorder (ODD) or conduct disorder (CD) contraindicated it. These children, however, appeared to do better with family based approaches that were also psychodynamic in orientation. A classic study, and for many years one of the only studies that considered the issue of intensity of psychological therapy, focussed on specific learning difficulties as a target of therapy.[93,112] Boys in middle childhood with serious reading problems benefited significantly from psychodynamic psychotherapy. There was a dose-response relationship over the 2-year treatment period. Children who received more intensive help (more sessions per week) benefited most from the therapy in terms of self-esteem, the capacity to form relationships, and the capacity to work, including frustration tolerance. Particularly interesting is the multi-centred randomized trial of the treatment of sexually abused girls treated in individual psychotherapy and psychoeducational group therapy.[87] Trowell *et al.*[87] randomized 71 sexually abused girls to either 30 sessions of individual psychoanalytic psychotherapy or 18 sessions of group psychotherapy with psychoeducational components. These young people presented with a range of psychiatric problems, most commonly post-traumatic stress disorder (PTSD) and depression. Psychodynamic treatment was somewhat superior to psychoeducation, but the difference was not as marked as might be expected. Superiority was particularly evident in relation to PTSD and generalized anxiety disorder (GAD). Depression, however, was relatively less likely to improve, as was separation anxiety. A subsequent report underscored the importance of the mother's support for the therapy as a predictor of improvement in the children and the benefit that the mothers gained in terms of their own mental health from the child's treatment.[113]

A chart review of the outcome of 763 cases in child psychotherapy has been carried out at the Anna Freud Centre in London.[103] While this retrospective methodology has severe limitations, the study reached a number of fairly robust conclusions, that need to be explored further in controlled, prospective investigations. The main findings were:

(a) Attrition was low compared to reports of other treatment approaches.

(b) Children with pervasive developmental disorders (e.g. autism) or intellectual disability did not do well, even with prolonged, intensive treatment. Children with serious disruptive disorders also had relatively poor outcomes.

(c) Younger children improved significantly during psychodynamic treatment, and gained additional benefit from 4–5 weekly sessions.

(d) Anxiety disorders, particularly specific rather than pervasive symptoms, were associated with a good prognosis, even if the primary diagnosis was of a different type, e.g. disruptive disorder.

(e) Children with emotional disorders and/or severe or pervasive symptomatology responded very well to intensive treatment (4–5 sessions per week), but did not show satisfactory rates of improvement in non-intensive psychotherapy.

(f) Predictors of improvement varied considerably between subgroups of the full sample, and, by subdividing the sample according to diagnostic group and developmental level, it was possible to predict a majority of the variance in outcome within the subgroups.

Limitations of the psychotherapeutic approach

The empirical status of all psychodynamic approaches remains controversial. The body of rigorous research supporting psychodynamic therapies for adults for most disorders remains limited, particularly relative to research supporting pharmaceutical treatments and even other psychosocial approaches such as CBT.[80] There are both practical and theoretical difficulties to mounting trials of dynamic therapies, and these go some way to explaining the lack of evidence (e.g. identifying suitable control groups for long-term intensive treatments, difficulties in operationalizing the treatment methods, the expense of mounting trials sufficiently powered to yield information on what treatments are appropriate for which disorder, the failure to tightly manualize psychodynamic treatments, etc.). Those who argue for continued investment in this approach (correctly in our view) point to the limitations of the evidence base supporting CBT[114] or pharmacological approaches.[115] Ultimately, however, such a negative case cannot persuade policy makers and funders, and, without intense research on the effectiveness of the method deeply rooted in and shaped by psychological models of pathology, the long-term survival of this orientation is not assured.[116] This is not to say that the techniques that have evolved as part of this approach will not survive (they are effective, and clinicians, being pragmatic people, will continue to discover and use them), but they will be increasingly absorbed into alternative models, and the unique approach pioneered by Freud and outlined in this chapter might not continue. Child therapists thus face formidable challenges to their clinical and theoretical convictions, to their professional status, and, as evidence-based medicine and managed care relentlessly expand their control over reimbursement and deny payment for child therapy, to their livelihood. If child psychotherapy is to have a future, its unique effectiveness for specific childhood disorders must be demonstrated in randomized controlled trials. We believe that children with severe disorders of personality and multiple psychiatric diagnoses are indeed well suited to and dramatically benefit from a psychotherapeutic approach. Such faith, however, now requires support from empirical investigations.

Further information

Fonagy, P., Target, M., and Gergely, G. (2004). Psychoanalytic perspectives on developmental psychopathology. In *Developmental psychopathology* (2nd edn) (eds. D. Cicchetti and D.J. Cohen), pp. 504–54. Guilford Press, New York.

Kennedy, E. (2004). *Child and adolescent psychotherapy: a systematic review of psychoanalytic approaches*. North Central London Strategic Health Authority, London.

Lanyado, M. and Horne, A. (2006). *A question of technique*. Routledge, London.

References

1. Pessiglione, M., Schmidt, L., Draganski, B., *et al.* (2007). How the brain translates money into force: a neuroimaging study of subliminal motivation. *Science*, **316**, 904–6.
2. Kernberg, O.F. (2005). Unconscious conflict in the light of contemporary psychoanalytic findings. *The Psychoanalytic Quarterly*, **74**, 65–81; discussion 327–63.
3. Winnicott, D.W. (1965). *The maturational process and the facilitating environment*. Hogarth Press, London.
4. Freud, A. (1965). *Normality and pathology in childhood: assessments of development*. International Universities Press, Madison, CT.
5. Fonagy, P. and Target, M. (2003). *Psychoanalytic theories: perspectives from developmental psychopathology*. Whurr, London.
6. Freud, A. (1936). *The ego and the mechanisms of defence*. International Universities Press, New York, 1946.
7. Blagov, P.S. and Singer, J.A. (2004). Four dimensions of self-defining memories (specificity, meaning, content, and affect) and their relationships to self-restraint, distress, and repressive defensiveness. *Journal of Personality*, **72**, 481–511.
8. Jorgensen, R.S., Frankowski, J.J., Lantinga, L.J., *et al.* (2001). Defensive hostility and coronary heart disease: a preliminary investigation of male veterans. *Psychosomatic Medicine*, **63**, 463–9.
9. Lyons-Ruth, K. (2003). Dissociation and the parent-infant dialogue: a longitudinal perspective from attachment research. *Journal of the American Psychoanalytical Association*, **51**, 883–911.
10. Shamir-Essakow, G., Ungerer, J.A., Rapee, R.M., *et al.* (2004). Caregiving representations of mothers of behaviorally inhibited and uninhibited preschool children. *Developmental Psychology*, **40**, 899–910.
11. Kaye, A.L. and Shea, M.T. (2000). Personality disorders, personality traits, and defense mechanisms. In Task force for the handbook of psychiatric measures (edn). *Handbook of psychiatric measures*, pp. 713–49. American Psychiatric Association, Washington, DC.
12. Vaillant, G.E. (1992). *Ego mechanisms of defense: a guide for clinicians and researchers*. American Psychiatric Association Press, Washington, DC.
13. Bond, M. (2004). Empirical studies of defense style: relationships with psychopathology and change. *Harvard Review of Psychiatry*, **12**, 263–78.
14. Lenzenweger, M.F., Clarkin, J.F., Kernberg, O.F., *et al.* (2001). The inventory of personality organization: psychometric properties, factorial composition, and criterion relations with affect, aggressive dyscontrol, psychosis proneness, and self-domains in a nonclinical sample. *Psychological Assessment*, **13**, 577–91.
15. Holmes, J. (1998). The changing aims of psychoanalytic psychotherapy: an integrative perspective. *International Journal of Psychoanalysis*, **79**, 227–40.
16. Allen, J.G. and Fonagy, P. (eds.) (2006). *Handbook of mentalization-based treatment*. Wiley, New York.
17. O'Shaughnessy, E. (1988). W. R. Bion's theory of thinking and new techniques in child analysis. In *Melanie Klein today: developments in theory and practice: mainly practice*, Vol. 2 (ed. E.B. Spillius), pp. 177–90. Routledge, London.

18. Kennedy, H. and Moran, G. (1991). Reflections on the aims of child psychoanalysis. *The Psychoanalytic Study of the Child*, **46**, 181–98.
19. Bleiberg, E., Fonagy, P., and Target, M. (1997). Child psychoanalysis: critical overview and a proposed reconsideration. *The Psychiatric Clinics of North America*, **6**, 1–38.
20. Hurry, A. (ed.) (1998). *Psychoanalysis and developmental theory*. Karnac, London.
21. Daniel, S.I. (2006). Adult attachment patterns and individual psychotherapy: a review. *Clinical Psychology Review*, **26**(8), 968–84.
22. Fonagy, P. and Bateman, A.W. (2006). Mechanisms of change in mentalization-based treatment of BPD. *Journal of Clinical Psychology*, **62**, 411–30.
23. Kazdin, A.E., Marciano, P.L., and Whitley, M.K. (2005). The therapeutic alliance in cognitive-behavioral treatment of children referred for oppositional, aggressive, and antisocial behavior. *Journal of Consulting and Clinical Psychology*, **73**, 726–30.
24. Krupnick, J.L., Sotsky, S.M., Simmens, S., *et al.* (1996). The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology*, **64**, 532–9.
25. PDM Task Force. (2006). *Psychodynamic diagnostic manual*. Alliance of Psychoanalytic Organizations, Silver Spring, MD.
26. Safran, J.D. and Muran, J.C. (2000). *Negotiating the therapeutic alliance*. Guilford Press, New York.
27. Wynn Parry, C. and Birkett, D. (1996). The working alliance: a re-appraisal. *British Journal of Psychotherapy*, **12**, 291–9.
28. Kernberg, P. and Chazan, S.E. (1991). *Children with conduct disorders: a psychotherapy manual*. Basic Books, New York.
29. Freud, S. (1900). The interpretation of dreams. In *The standard edition of the complete psychological works of Sigmund Freud*, Vols. 4 and 5 (ed. J. Strachey), pp. 1–715. Hogarth Press, London.
30. Freud, S. (1920). Beyond the pleasure principle. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 18 (ed. J. Strachey), pp. 1–64. Hogarth Press, London.
31. Freud, S. (1909). Analysis of a phobia in a five-year-old boy. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 10 (ed. J. Strachey), pp. 1–147. Hogarth Press, London.
32. Hug-Helmuth, H. (1921). On the technique of child analysis. *International Journal of Psychoanalysis*, **2**, 287–303.
33. Freud, A. (1946). *The psychoanalytic treatment of children*. Imago Publishing, London.
34. Klein, M. (1932). *The psycho-analysis of children*. Hogarth Press, London.
35. Klein, M. (1946). Notes on some schizoid mechanisms. In *Developments in psychoanalysis* (eds. M. Klein, P. Heimann, S. Isaacs, and J. Riviere), pp. 292–320. Hogarth Press, London.
36. Bion, W.R. (1959). Attacks on linking. *International Journal of Psychoanalysis*, **40**, 308–15.
37. Williams, G. (1997). *Internal landscapes and foreign bodies*. Duckworth, London.
38. Winnicott, D.W. (1971). *Playing and reality*. Tavistock, London.
39. Altman, N., Briggs, R., Frankel, J., *et al.* (2002). *Relational child psychotherapy*. Other Press, New York.
40. Mitchell, S.A. (1997). *Influence and autonomy in psychoanalysis*. Analytic Press, Hillsdale, NJ.
41. Trevarthen, C.A., Aitken, K.J., Vandekerckhove, M., *et al.* (2006). Collaborative regulations of vitality in early childhood: stress in intimate relationships and postnatal psychopathology. In *Developmental psychopathology: developmental neuroscience*, Vol. 2 (2nd edn) (eds. D. Cicchetti and D.J. Cohen), pp. 65–126. Wiley, New York.
42. Edgcumbe, R. (2000). *Anna Freud: a view of development, disturbance and therapeutic techniques*. Routledge, London.
43. Miller, J.M. (1996). Anna Freud. *The Psychoanalytic Study of the Child*, **51**, 142–71.
44. Freud, A. (1981). *A psychoanalytic view of developmental psychopathology*. International University Press, New York.
45. Freud, A. (1963). The concept of developmental lines. *The Psychoanalytic Study of the Child*, **18**, 245–65.
46. Cicchetti, D. and Cohen, D.J. (eds.) (2006). *Developmental psychopathology*, Vols. 1–3 (2nd edn). John Wiley & Sons, New York.
47. Fonagy, P., Edgcumbe, R., Moran, G.S., *et al.* (1993). The roles of mental representations and mental processes in therapeutic action. *The Psychoanalytic Study of the Child*, **48**, 9–48.
48. Fonagy, P. and Target, M. (1998). Mentalization and the changing aims of child psychoanalysis. *Psychoanalytic Dialogues*, **8**, 87–114.
49. Garber, B. (2001). Freud's impact on therapeutic work with children. *Annual of Psychoanalysis*, **29**, 133–43.
50. Rustin, M. (1999). The training of child psychotherapists at the Tavistock clinic: philosophy and practice. *Psychoanalytic Inquiry*, **19**, 125–41.
51. Geissmann, C. and Geissmann, P. (1998). *The history of child psychoanalysis*. Routledge, London.
52. Housman, A. (1997). *The history of child psychoanalysis*. Routledge, London.
53. Bleiberg, E. (2001). *Treating personality disorders in children and adolescents: a relational approach*. Guilford Press, New York.
54. Abrams, S. (1988). The psychoanalytic process in adults and children. *The Psychoanalytic Study of the Child*, **43**, 45–261.
55. Kernberg, P.F. (1995). Child psychiatry: individual psychotherapy. In *Comprehensive textbook of psychiatry* (6th edn) (eds. H.I. Kaplan and B.J. Sadock), pp. 2399–412. Williams & Wilkins, Baltimore, MD.
56. Fonagy, P. and Target, M. (1997). Perspectives on the recovered memories debate. In *Recovered memories of abuse: true or false?* (eds. J. Sandler and P. Fonagy), pp. 183–216. Karnac Books, London.
57. Sandler, J., Kennedy, H., and Tyson, R. (1980). *The technique of child psychoanalysis: discussions with Anna Freud*. Harvard University Press, Cambridge, MA.
58. Fonagy, P., Edgcumbe, R., Target, M., *et al.* (Unpublished manuscript). *Contemporary Psychodynamic Child Therapy: Theory and Technique*.
59. Verheugt-Pleiter, A.J.E., Zevalkink, J., and Schmeets, M.G.J. (eds.) (in press). *Mentalizing in child therapy: guidelines for clinical practitioners*. Karnac Books, London.
60. Bateman, A.W. and Fonagy, P. (2004). *Psychotherapy for borderline personality disorder: mentalization based treatment*. Oxford University Press, Oxford.
61. Dowling, A.S. and Naegele, J. (1995). Child and adolescent psychoanalysis. In *Psychoanalysis: the major concepts* (eds. B.E. Moore and B.D. Fine), pp. 26–44. Yale University Press, New Haven, CT.
62. Fonagy, P. and Target, M. (1996a). A contemporary psychoanalytical perspective: psychodynamic developmental therapy. In *Psychosocial treatments for child and adolescent disorders: empirically based approaches* (eds. E. Hibbs and P. Jensen), pp. 619–38. APA and NIH, Washington, DC.
63. Mandler, G. (1985). *Cognitive psychology. An essay in cognitive science*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
64. Sandler, J. and Rosenblatt, B. (1962). The concept of the representational world. *The Psychoanalytic Study of the Child*, **17**, 128–45.
65. Fonagy, P., Moran, G.S., Edgcumbe, R., *et al.* (1993). The roles of mental representations and mental processes in therapeutic action. *The Psychoanalytic Study of the Child*, **48**, 9–48.
66. Coppolillo, H.P. (2002). Use of play in psychodynamic psychotherapy. In *Child and adolescent psychiatry: a comprehensive textbook* (3rd edn) (ed. M. Lewis), pp. 992–8. Lippincott, Williams & Wilkins, Philadelphia.
67. Yanof, J.A. (1996). Language, communication, and transference in child analysis. I. Selective mutism: the medium is the message. II. Is

child analysis really analysis? *Journal of the American Psychoanalytic Association*, **44**, 100–16.
68. Fonagy, P. and Target, M. (1996b). Playing with reality. I. Theory of mind and the normal development of psychic reality. *International Journal of Psycho-Analysis*, **77**, 217–33.
69. Slade, A. (2000). The development and organization of attachment. *Journal of the American Psychoanalytic Association*, **48**, 1147–74.
70. Lewis, M. (2002). Intensive individual psychodynamic psychotherapy: the therapeutic relationship and the technique of interpretation: the use of play in psychodynamic therapy. In *Child and adolescent psychiatry: a comprehensive textbook* (3rd edn) (ed. M. Lewis), pp. 984–92. Lippincott, Williams & Wilkins, Philadelphia.
71. Ritvo, R.Z. and Ritvo, S. (2002). Psychodynamic psychotherapy. In *Child and adolescent psychiatry: a comprehensive textbook* (3rd edn) (ed. M. Lewis), pp. 974–84. Lippincott, Williams & Wilkins, Philadelphia.
72. Kernberg, P.F. (1991). Termination in child psychoanalysis: criteria from within the sessions. In *Saying goodbye: a casebook of termination in child and adolescent analysis and therapy* (eds. A.G. Schmulker), pp. 321–37. Analytic Press, Hillsdale, NJ.
73. Kolvin, I., MacMillan, A., and Wrate, R.M. (1988). Psychotherapy is effective. *Journal of the Royal Society of Medicine*, **81**, 261–6.
74. Muratori, F., Picchi, L., Casella, C., *et al.* (2002). Efficacy of brief dynamic psychotherapy for children with emotional disorders. *Psychotherapy and Psychosomatics*, **71**, 28–38.
75. Kennedy, E. (2004). Child and adolescent review of psychotherapy: A systematic review of psychoanalytic approaches. North Central London Strategic Health Authority, London.
76. Fonagy, P., Target, M., Cottrell, D., *et al.* (2002). *What works for whom? A critical review of treatments for children and adolescents*. Guilford, New York.
77. Kazdin, A.E. (2000). *Psychotherapy for children and adolescents: directions for research and practice*. Oxford University Press, Oxford.
78. Kazdin, A.E. (2003). Psychotherapy for children and adolescents. *Annual Review of Psychology*, **54**, 253–76.
79. Kazdin, A.E. (2004). Psychotherapy for children and adolescents. In *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th edn) (ed. M. Lambert), pp. 543–89. Wiley, New York.
80. Roth, A. and Fonagy, P. (2004). *What works for whom? A critical review of psychotherapy research* (2nd edn). Guilford Press, New York.
81. Hibbs, E.D. (ed.) (2004). *Psychosocial treatments for child and adolescent disorders: empirically based strategies for clinical practice* (2nd edn). American Psychological Association, Washington DC.
82. Weisz, J.R. (2004). *Psychotherapy for children and adolescents: evidence-based treatments and case examples*. Cambridge University Press, Cambridge.
83. Robin, A.L., Siegel, P.T., Moye, A.W., *et al.* (1999). A controlled comparison of family versus individual therapy for adolescents with anorexia nervosa. *Journal of the American Academy of Child and Adolescent Psychiatry*, **38**, 1482–9.
84. Sinha, U.K. and Kapur, M. (1999). Psychotherapy with emotionally disturbed adolescent boys: outcome and process study. *NIMHANS Journal*, **17**, 113–30.
85. Smyrnios, K.X. and Kirkby, R.J. (1993). Long-term comparison of brief versus unlimited psychodynamic treatments with children and their parents. *Journal of Consulting and Clinical Psychology*, **61**, 1020–7.
86. Szapocznik, J., Rio, A., Murray, E., *et al.* (1989). Structural family versus psychodynamic child therapy for problematic Hispanic boys. *Journal of Consulting and Clinical Psychology*, **57**, 571–8.
87. Trowell, J., Kolvin, I., Weeramanthri, T., *et al.* (2002). Psychotherapy for sexually abused girls: psychopathological outcome findings and patterns of change. *The British Journal of Psychiatry*, **180**, 234–47.
88. Trowell, J., Rhode, M., Miles, G., *et al.* (2003). Childhood depression: work in progress. *Journal of Child Psychotherapy*, **29**, 147–69.
89. Moran, G., Fonagy, P., Kurtz, A., *et al.* (1991). A controlled study of the psychoanalytic treatment of brittle diabetes. *Journal of the American Academy of Child and Adolescent Psychiatry*, **30**, 926–35.
90. Muratori, F., Picchi, L., Bruni, G., *et al.* (2003). A two-year follow-up of psychodynamic psychotherapy for internalizing disorders in children. *Journal of the American Academy of Child and Adolescent Psychiatry*, **42**, 331–9.
91. Muratori, F., Picchi, L., Casella, C., *et al.* (2001). Efficacy of brief dynamic psychotherapy for children with emotional disorders. *Psychotherapy and Psychosomatics*, **71**, 28–38.
92. Fonagy, P. and Target, M. (1994). The efficacy of psychoanalysis for children with disruptive disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, **33**, 45–55.
93. Heinicke, C.M. and Ramsey-Klee, D.M. (1986). Outcome of child psychotherapy as a function of frequency of sessions. *Journal of the American Academy of Child Psychiatry*, **25**, 247–53.
94. Reid, S., Alvarez, A., and Lee, A. (2001). The Tavistock autism workshop approach. In *Autism-the search for coherence* (eds. J. Richer and S. Coates), pp. 182–92. Jessica Kingsley, London.
95. Target, M. and Fonagy, P. (1994a). The efficacy of psychoanalysis for children with emotional disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, **33**, 361–71.
96. Target, M. and Fonagy, P. (1994b). The efficacy of psychoanalysis for children: developmental considerations. *Journal of the American Academy of Child and Adolescent Psychiatry*, **33**, 1134–44.
97. Apter, A., Bernhout, E., and Tyano, S. (1984). Severe obsessive compulsive disorder in adolescence: a report of eight cases. *Journal of Adolescence*, **7**, 349–58.
98. Boston, M. and Lush, D. (1994). Further considerations of methodology for evaluating psychoanalytic psychotherapy with children: reflections in the light of research experience. *Journal of Child Psychotherapy*, **20**, 225–9.
99. Lush, D., Boston, M., and Grainger, E. (1991). Evaluation of psychoanalytic psychotherapy with children: therapists' assessments and predictions. *Psychoanalytic Psychotherapy*, **5**, 191–234.
100. Target, M. and Fonagy, P. (2002). The long-term follow-up of child analytic treatments (AFC3). In *An open door review of outcome studies in psychoanalysis* (2nd edn) (ed. P. Fonagy), pp. 141–6. International Psychoanalytic Association, London.
101. Baruch, G. (1995). Evaluating the outcome of a community-based psychoanalytic psychotherapy service for young people between 12 and 25 years old: work in progress. *Psychoanalytic Psychotherapy*, **9**, 243–67.
102. Baruch, G., Fearon, P., and Gerber, A. (1998). Evaluating the outcome of a community-based psychoanalytic psychotherapy service for young people: one year repeated follow-up. In *Rethinking clinical audit* (eds. R. Davenhill and M. Patrick), pp. 157–82. Routledge, London.
103. Fonagy, P. and Target, M. (1996c). Predictors of outcome in child psychoanalysis: a retrospective study of 763 cases at the Anna Freud Centre. *Journal of the American Psychoanalytic Association*, **44**, 27–77.
104. Petri, H. and Thieme, E. (1978). Katamnese zur analytischen psychotherapie im kindes und jugendalter. *PSYCHE*, **1**, 21–54.
105. Vilsvik, S.O. and Va2glum, P. (1990). Teenage anorexia nervosa: a 1 to 9 year follow up after psychodynamic treatment. *Nord Psykiatr Tidsskr*, **44**, 249–55.
106. Winkelmann, K., Hartmann, M., Neumann, K., *et al.* (2000). Stability of therapeutic outcome after child and adolescent psychoanalytical therapy. *Praxis Kinderpsychol Kinderpsychiatr*, **49**, 315–28.
107. Zelman, A.B., Samuels, S., and Abrams, D. (1985). IQ changes in young children following intensive long-term psychotherapy. *American Journal of Psychotherapy*, **39**, 215–27.

108. Fonagy, P. and Moran, G.S. (1990). Studies on the efficacy of child psychoanalysis. *Journal of Consulting and Clinical Psychology*, **58**, 684–95.
109. Lush, D., Boston, M., Morgan, J., *et al.* (1998). Psychoanalytic psychotherapy with disturbed adopted and foster children: a single case follow-up study. *Clinical Child Psychology and Psychiatry*, **3**, 51–69.
110. Moran, G.S. and Fonagy, P. (1987). Psychoanalysis and diabetic control: a single case study. *British Journal of Medical Psychology*, **60**, 357–72.
111. Goodyer, I., Dubicka, B., Wilkinson, P., *et al.* (2007). Selective serotonin reuptake inhibitors (SSRIs) and routine specialist care with and without cognitive behaviour therapy: randomised controlled trial. *British Medical Journal*, **335**, 142–6.
112. Heinicke, C.M. (1965). Frequency of psychotherapeutic session as a factor affecting the child's developmental status. *The Psychoanalytic Study of the Child*, **20**, 42–98.
113. Rushton, A. and Miles, G. (2000). A study of a support service for the current carers of sexually abused girls. *Clinical Child Psychology and Psychiatry*, **5**, 411–26.
114. Westen, D., Novotny, C.M., and Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin*, **130**, 631–63.
115. Whittington, C.J., Kendall, T., Fonagy, P., *et al.* (2004). Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. *Lancet*, **363**, 1341–5.
116. Gabbard, G.O., Gunderson, J.G., and Fonagy, P. (2002). The place of psychoanalytic treatments within psychiatry. *Archives of General Psychiatry*, **59**, 505–10.

9.5.3 Cognitive behaviour therapies for children and families

Philip Graham

Introduction

Cognitive behaviour therapy (CBT) is derived from both behavioural and cognitive theories. Using concepts such as operant conditioning and reinforcement, behavioural theories treat behaviour as explicable without recourse to description of mental activity. In contrast, mental activity is central to all concepts derived from cognitive psychology. Both sets of theories have been of value in explaining psychological disorders and, in the design of interventions they have proved an effective combination.

Central to that part of cognitive theory that is relevant to CBT is the concept of 'schemas', first described in detail by Jean Piaget.[1] A schema is a mental 'structure for screening, coding, and evaluating impinging stimuli'.[2] The origin of mental schemas lies in the pre-verbal phase when material is encoded in non-verbal images that, as the child's language develops, gradually become verbally labelled. They form part of a dynamic system interacting with an individual child's physiology, emotional functioning, and behaviour with their operation depending on the social context in which the child is living. There are similarities but also differences between schemas and related concepts in psychoanalysis, such as Freudian 'complexes' and Kleinian 'positions'.

Schemas can be seen as organized around anything in the child's world, especially objects, beliefs, or emotions. They develop from past experience. The processing of new information in relation to such schemas can usefully be seen as involving the evaluation of discrepancies between information that is received and information that is expected. If there is a discrepancy, (the information not corresponding with that expected), then during the coding process information may be distorted so that it no longer creates discomfort, or, more adaptively, it may be incorporated into a modified schema.

Cognitive development

The theory of cognitive development that Piaget constructed on the basis of an immense amount of experimental work was characterized by stages of development. He described characteristic features of the sensori-motor (0–2 years), pre-operational (2–7 years), concrete operational (7–12 years), and formal operational (12 years onwards) stages. Before the end of a stage is reached the child is incapable of showing more advanced thinking. In particular, the child's thinking before the concrete operational stage is characterized by egocentricity and an inability to take the perspective of another person. Abstract reasoning is not possible for the child until the formal operational stage is reached.

Even though Piaget's views of the limitations of the cognitive abilities of young children have been strongly criticized especially on the grounds that he was judging egocentricity on the basis of findings obtained in highly artificial situations, Piaget remained a dominant influence in cognitive psychology and education throughout the twentieth century. It is now widely accepted that, although obviously young children are less competent than those in middle childhood and these are less competent than adolescents, cognitive competence advances much more rapidly than Piaget described and the social context in which a child's competence is investigated has a much more profound influence on performance than he allowed. Children do a great deal better in naturalistic circumstances than when they take part in experiments. Further, coaching can improve performance to a level not previously obtainable. For example, it has been shown that, with preliminary training, 3-year-old children understand that drawings of thought bubbles can represent what people think. They can distinguish between thoughts and actions, recognize that thoughts are subjective and that two people can have different thoughts about the same events.[3]

Investigation of the development of the 'theory of mind' held by children has revealed that between 3 and 4 years they begin to realize that other children can be deceived by appearances and hold false beliefs they themselves do not hold. This shows that, given the right circumstances, children of this young age are able to 'de-centre' and are not necessarily limited by egocentricity. By the age of 8 years children have such stable concepts of their own self-esteem that they are capable of reliably completing self-esteem questionnaires about their own feelings and performance in comparison to other children.[4] Some schemas in young children are however relatively unstable, gradually increasing in stability as they get older. For example, it has been shown that attributional style (the tendency to attribute adverse events either to the self or to external circumstances) does not become stable until early adolescence,

though it may be identified earlier if the events are particularly salient to the child in question.[5]

It has been hypothesized[6] that maladaptive schemas developed during childhood are responsible for the formation and maintenance of adult psychopathology. Building on this model, a therapeutic approach (schema-focused therapy) based on the identification of particular maladaptive schemas has been proposed for adults. Subsequently Stallard and Rayner[7] have developed a schema questionnaire that builds on adult work to identify such maladaptive schema in 11 to 16-year-old school children.

Technique and management in the paediatric age group

Although there are certain common principles, CBT does not involve, as will be seen, a single approach that can be applied across all disorders; it is better seen as a family of approaches with certain core elements in common. In adults the type of disorder and the individual circumstances of the patient will determine the choice of therapeutic methods. In children and adolescents the cognitive level of the patient will also need to be taken into account. Though the age of the child will give some indication of the cognitive level of the child, there is wide variation in competence amongst children of the same age. Further, the therapist may use the skills of an educationist to bring the child's competence up to a level at which the child can more actively participate in therapy. Kendall[8] suggests indeed that one of the therapeutic roles that the therapist should adopt is that of *educator*, who needs communication skills to assist children to learn to think for themselves.

Behaviour therapy or CBT?

In principle, the decision as to whether to include a cognitive component in therapy depends on whether the clinical formulation incorporates cognitive distortions or biases. In practice, because of their cognitive limitations CBT is rarely used in children under the age of 7 years. Treatment in children younger than 7 years is predominantly behavioural, with the cognitive component limited to coping self-talk. Conditioning approaches to the treatment of feeding and sleeping problems as well as enuresis and encopresis usually have a very small or no significant cognitive component.

In some conditions such as anxiety disorders, especially specific phobias, where desensitization and reinforcement approaches are widely used in adults, the use of a mainly behavioural approach does not reduce effectiveness. A cognitive component may nevertheless be incorporated because the CBT principles of collaboration, openness, and guided discovery, usually less marked when purely behavioural approaches are applied, are advantageous to the patient.

Aids to cognitive tasks

Where experience with adults suggests that cognitive tasks add significantly to the effectiveness of treatment, as in depressive disorders and problems of social relationships, even young people in early adolescence will usually be able to co-operate as well as adults. The cognitive treatment of younger children with these conditions may be helped by the use of age-adapted techniques.[9] For example, card-sorting *games* have been devised to help children distinguish between thoughts, feelings, and situations. *Puppets* can be used to facilitate discussion as part of the assessment process, to model alternative ways the child might cope with difficult situations and to engage the child in rehearsal and practice of new skills. *Story telling* can provide an insight into the child's inner world; they provide a way of externalizing and accessing the child's cognitions, allow an opportunity to challenge the child's assumptions, introduce the child to more positive ways of coping, and can be used to model success and help the child gain more functional assumptions and beliefs.

Working with parents

Parents play many roles in the delivery of CBT to children and adolescents. To begin with, even up to mid-adolescence, it is nearly always parents who identify the behaviour and emotional problems that lead to advice being sought. They are the people most likely to press for psychological help. It is they who have to persuade often reluctant children and adolescents to attend and participate in a service that their offspring may fear, not without reason, will result in stigmatization.

They are then likely to play a major part in the assessment process. From mid- to late adolescence, the patient or client will be the main source of information, but before that it is the parents and teachers who will often provide most relevant information. If treatment is proposed it is they who need to give consent, though their child will also need to assent if the therapy is to have any chance of succeeding.

Once treatment planning has begun, the part that parents play will depend very much on the age of the child or adolescent, the diagnosis, family circumstances (especially the quality of the relationships between parents and child), and the degree to which the assessment has revealed that the parents as well as being the main carers are also involved in the origin and maintenance of the problem. Most explanatory theories of anxiety disorders in children, for example, point to the ways in which parents can provide inappropriately anxious models for imitation by their children. In a small scale study it has been shown that changing parental attributions can, in itself, result in improvements in problem behaviour scores on a questionnaire.[10] Parents may also be seen as clients in their own rights in parallel sessions, as co-therapists or as facilitators of therapy for their children. Therapists dealing with adolescent offspring are often in a difficult position *vis-à-vis* parents in that they will wish to encourage autonomy and independent decision-making in the child or adolescent, while needing the parents to monitor homework, encourage further attendance, and provide information on progress.

The involvement of parents also brings ethical dilemmas. There are three main areas of ethical concern.[11] The therapist often has to balance the different viewpoints of parents and children, a particular problem in the management of oppositional and conduct disorders where children often fail to acknowledge the existence of problems that are causing distress to their parents. There is frequently need to address family issues such as marital conflict that are clearly relevant yet not the reasons why the child has been brought for treatment. Finally, there is the need to achieve genuine collaboration with parents, making explicit their role as co-therapists. This is made easier if children are also actively involved as fellow

collaborators, taking responsibility for progress and being encouraged to make suggestions for alternative approaches. A collaborative stance may however not be possible if it becomes clear that there are child protection issues with one or both parents involved in maltreatment of their children. Wolpert and her colleagues provide a useful checklist for clinicians to help assess how far they are attempting to balance different viewpoints in issues involving different family members and promoting collaboration.

Failure to engage and failure to respond

In adolescents, lack of motivation for change is often a major impediment to engagement in therapy. Not only is there often a failure to recognize the importance of a problem, to accept the need for change or to appear to understand why change is necessary, but there may also be an absence of the level of self-belief, self-confidence, or self-efficacy that is necessary before hopeful steps can be taken in the right direction. In these circumstances techniques of motivational interviewing will help the therapist to achieve engagement.[12]

The reasons for non-response to CBT in adults have been discussed by Kingdon *et al.*[13] Common problems include unsuitability for treatment possibly arising from misdiagnosis, resistance to treatment, an inadequate number of sessions, difficulties in the therapeutic relationship and the presence of concurrent social and/or physical pathology. Non-response in children and adolescents arises from similar issues, with, additionally, complicating problems arising from negative parental attitudes and behaviour.

Anxiety disorders

Cognitive distortions and deficits

A characteristic constellation of cognitive deficits and distortions underlies the presence of anxiety disorders in children and adolescents. A central feature is the exaggerated perception of threat arising from an inability to assess accurately the seriousness of danger. Thus a deficit in perceptual competence results in cognitive distortion. The characteristic nature of the threat involved will depend to a considerable degree both on the stage of cognitive development of the child and on the social demands that are encountered during that particular phase of life. Pre-school children are most likely to be threatened by separation from parents; children aged 5 to 12 years by feared situations at school and adolescents by social situations as well as wider concerns such as environmental pollution. Certain fears and phobias such as fear of spiders and snakes appear more biologically based and are present through childhood to adolescence.

These cognitive deficits and distortions both result in and are maintained and increased by abnormal levels of physiological arousal and by behavioural avoidance of the feared situations. Autonomic arousal produces symptoms such as dry mouth, palpitations, and abdominal pain and these may be misinterpreted as implying serious threatening illness. Panic attacks may be catastrophized and taken to mean that death is imminent. Avoidance of feared situations such as separation from parents in younger children, refusal to go to school in older children or to social events such as parties in adolescence prevent cognitive testing of the reality of the supposed threat and reinforce the cognitive distortion.

The fact that anxiety disorder is partly genetically determined means that children suffering from this condition have an increased risk of having anxious parents. Such parents are likely to model anxious behaviour, especially in the way they show over-protection to their children. Anxious children are therefore likely to be exposed to social learning situations at home that will increase the risk of avoidance of feared situations. Gene-environment interactions ensure that many parents who cannot bear to be separated from their children or who are anxious every time they leave the house will transmit their fears to their children both directly and indirectly. In adolescence, anxious young people may selectively choose shy, inhibited friends who reinforce their sense of unrealistic threat.

Techniques of assessment and intervention

The assessment of children with anxiety disorders by a cognitive behaviour therapist focuses on the identification of cognitive deficits and distortions and the manner in which they are currently being reinforced, especially by avoidant behaviour. Nevertheless it is important that before enquiry is made along these lines a full history is taken of the development of anxious symptoms, the presence of other symptomatology, the situations that increase and reduce anxiety, the presence of anxiety in parents, sibs, and friends, and the measures that have already been taken, especially by parents, to improve the condition. Skilled assessment involves listening to the anxious preoccupations of both children and parents sympathetically and without any hint of criticism.

There are a number of systematic cognitive approaches to the reduction of anxiety in children of which the most widely used is the four-step coping or FEAR plan, in which F = **F**eeling frightened (awareness of anxiety symptoms such as somatic aches and pains), E = **E**xpecting bad things to happen (awareness of negative self-talk), A = **A**ttitudes and actions that can help (problem-solving strategies), and R = **R**esults and rewards (rewarding for success, dealing with failure).[14] The 'Cool Kids' programme is generally similar but puts more emphasis on parent involvement.[15] When parents show significant levels of anxiety themselves, effectiveness of treatment is enhanced if parental anxiety management is included as part of treatment.[16] A self-help book for parents broadly based on the same principles provides a practical approach to the management of anxiety, using the so-called COPE programme.[17]

Treatment begins with one or two psycho-educational sessions in which the child and parent(s), together or separately, are given information about the way anxiety develops and is maintained, the manner in which the body shows anxiety (somatic symptoms), and the effects of avoidant behaviour and exposure to feared situations. It is important that these sessions are interactional with the child being encouraged to talk spontaneously about, for example, how he or she experiences somatic symptoms. The next few sessions involve children engaging in an exercise to identify their own negative thoughts, to test them against reality and to develop positive thinking in situations that have previously triggered anxiety. This will usually need to be done in imagination before it is tried out using 'graded exposure' in real situations. There are advantages in teaching relaxation techniques before the child embarks on exposure to feared situations. The use of imagery, such as the 'stepladder' approach to a hierarchy of feared situations may also be helpful. When the child makes progress, as is usually the case, rewards such as outings or other treats may be built in to the procedure.

Therapists vary in the degree to which they involve parents in management. The therapy can be delivered in a family context, parents can be seen separately from children, parents may not be seen at all, or the therapy may only be delivered to parents. Some centres use a group approach, with one or two therapists providing a group experience for parents and anxious children who go through the stages of treatment together and benefit from learning of each others' experiences. Some programmes have now been developed for use via the Internet with minimal personal contact with the child and family. Some therapists combine CBT with the use of medication, generally not anxiolytic agents because of the risk of dependency, but tricyclics or selective serotonin reactive inhibitors.

Evaluation of effectiveness and efficacy

A systematic review of the effectiveness of CBT for anxiety disorders in childhood and adolescence identified 10 randomized controlled trials that met inclusion criteria.[18] The outcome measure used was the remission of anxiety disorder. The remission rate was higher in the CBT groups (56.5 per cent) than in the control groups (34.8 per cent). The pooled odds ratio was 3.3 (CI = 1.9–5.6). The authors of this review conclude that CBT definitely provides benefit to children and adolescents with anxiety disorder, but that there is a lack of information concerning the value of CBT in younger children and that there are virtually no satisfactory studies comparing effectiveness with alternative treatments.

There is contradictory evidence concerning the importance of involving parents in therapy. Some[19,20] find little or no benefit, while others[21,22] find a trend towards benefit. A pilot study has found benefit from a programme that did not involve children directly but only involved parents seen in a group, who applied what they had learned in the group in managing the situations in which their children showed anxiety at home. Information on the use of therapy delivered via the Internet is limited, but those that exist suggest that Internet treatment is highly acceptable to families, creates minimal dropout and is effective when added to clinic treatment.[23] Dropout from more conventional treatment is likely to be high in single-parent families, ethnic minority families, and where anxiety levels are not conspicuously high.[24] There is evidence that the presence of co-morbid disorders does not reduce the efficacy of CBT.[25] The addition of antidepressants may increase the efficacy of CBT, especially in the treatment of school refusal.[26] Limited findings from long-term studies suggest that treatment benefits from the delivery of CBT to anxious children are maintained over at least 6 years.[27]

There is also evidence from controlled studies for the effectiveness of interventions, especially the FRIENDS programme[28] in the prevention of anxiety and depression in early adolescence. Stallard *et al.*[29] have shown how this programme can be delivered successfully by school nurses.

These evaluative studies have provided most encouraging findings for the effectiveness of CBT in this condition. However the findings also make clear that CBT, while producing worthwhile and persistent benefits in most children and adolescents with anxiety disorders, is not effective in a significant number of cases and in a significant number of others it is only partially effective. It is also less effective in socially disadvantaged groups. Finally, most evaluative studies have been carried out in highly specialist centres and there is a lack of evidence for their value in everyday practice.[18]

Depressive disorders

Cognitive distortions and deficits

The classical signs of depressive disorders, such as chronic misery and unhappiness, lack of interest in food, and motor retardation, may be seen as early as the first year of life. Infants and young children who show such symptomatology may well suffer depressive experiences similar to those of older people though in the pre-verbal phase there is no reliable method available to confirm this possibility. Awareness of feeling states develops towards the end of the second year of life.[30] By 2 to 3 years children realize that there can be a variety of personal reasons for an emotional reaction. By 4 years there is some consensus about the kind of situations that will provoke the common emotional reactions, including fear, sadness, and anger.[31] By 5 or 6 years a child is capable of understanding the concept of stability of mood, 'always being unhappy or just now and again', and by 7 or 8 years concepts of shame and guilt are understood at least in simplified form. Enduring and relatively stable negative attributions about the self become possible at around this age and the concept of death as a permanent state is established. By 13 to 14 years, emotional experiences of adult intensity occur and mature cognitions about different mood states will have been attained. Although the above account relates stage of development to chronological age, there is wide variation in the ages at which cognitive competence is gained. Further, the settings in which children are questioned or encouraged to express themselves freely and spontaneously, for example in play situations will greatly influence their capacity to show their abilities.

The cognitive model underlying CBT approaches to children and adolescents does not differ from that with adults. It is assumed that thoughts are the primary experience of depression and that depressed mood is secondary. Dysfunctional assumptions, including low feelings of self-worth, self-blame for events in the past, and hopelessness about the future are present either as stable features of a depressive personality or as a reaction to adverse experiences, real or imagined. Depressed children and adolescents systematically distort their experience to match their beliefs about themselves. At some point, these negative thoughts are automatically experienced without reflection. Increasingly situations are avoided because of a fear of negative outcomes. Therapy involves identifying and reality testing these negative thoughts. In addition the patient is encouraged to enter into activities that will be rewarding and disconfirm pessimistic assumptions.

Techniques of assessment and intervention

Initial assessment will involve taking a full history of the development of symptoms and the factors that reduce or exacerbate them, the child's functioning in different settings, and an account of family relationships. If the child is taken on for CBT, a typical approach[32] begins with the establishment of symptom status by the use of questionnaires such as the Children's Depression Inventory[33] in young patients and the Mood and Feelings Questionnaire[34] in adolescents. The goals of therapy are then discussed in a collaborative manner with emphasis on what the child or young person wishes to achieve. The proposed therapeutic approach is then explained together with the importance of homework outside the

therapy sessions. An indication of the number of sessions likely to be required, usually 12–16, is given. In early sessions an account of the child's current daily activities is obtained. Adolescents are helped to keep a diary of their activities and moods. In a form of 'affective education' a check is made on the vocabulary the child uses to describe feelings and links are then established between the child's mood and the activities he or she is undertaking.

During the next sessions, in collaboration with the child, homework is planned that aims to increase activity to the level previously undertaken. Emphasis is placed on the resumption of everyday activities rather than offering treats or special occasions. At this point a problem-solving approach may be indicated. This begins with problem definition, followed by brain-storming a number of different solutions. The outcomes for different solutions are discussed and a plan developed to achieve what seems to be a satisfactory outcome. Homework involves attempts to implement the plan while keeping a record of progress and how this has influenced mood.

At least from early adolescence it will usually be possible to introduce self-monitoring procedures, in which the child identifies and notes his level of mood in relation to the thoughts he is experiencing. The child is encouraged to imagine different situations and to record how each situation makes him feel. The child is encouraged to continue this process at home, recording what happens so that his experience can be discussed in the next session. This process is accompanied by self-evaluation training, a form of cognitive restructuring in which children learn to evaluate themselves in a more positive manner. They are encouraged to consider the evidence for having a poor opinion of themselves and then to examine carefully more positive alternative explanations. This process may be expected to reduce negative automatic thoughts.

Evaluation of effectiveness and efficacy

A comprehensive evidence-based review of controlled evaluation of cognitive behavioural psychotherapy for children and adolescents with depressive disorders[35] identified 12 studies that fulfilled methodological criteria. Most reported positive outcome for CBT post-treatment and at short-term follow-up. However, studies with longer follow-up periods from 9 months to 2 years found that a sizable percentage of subjects continued to report significant depressive symptoms or a recurrence of their depressive illness.

More recently the results of a major multi-centre trial, the Treatment of Adolescent Depression study (TADS) have been reported. In this study 479 adolescents, aged 12 to 17 years with depressive disorders were allocated randomly to a combination of fluoxetine and CBT, fluoxetine alone, CBT alone and an inert pill placebo. After 12 weeks of treatment the effects of combination therapy were clearly superior to either form of monotherapy and greatly superior to pill placebo. Fluoxetine alone was superior to CBT alone and to placebo, but CBT alone was not superior to placebo. On the other hand, fluoxetine alone was accompanied by higher rates of suicidal events and this did not occur in the combined group. It seemed therefore that CBT protected against the suicidality linked to fluoxetine use. The investigators concluded that the combination treatment produced the best outcomes.[36]

The published evidence mainly relates to children and adolescents with mild or moderate depressive disorders. There is some indication both from the TADS described and from other evidence that more severe depressive disorders do not respond as well or perhaps not at all to CBT alone.[37]

Attempts to use CBT to prevent depression in adolescents have met with varied success. One universal school-based approach found no difference at 2 to 4-year follow-up in children who received a teacher-administered cognitive behavioural intervention compared with a control group.[38] In contrast, positive effects were found for a CBT intervention targeting 13–18-year-old children of parents with depressive disorders.[39] Application of the Resourceful Adolescent Programme has also been shown to produce promising results in preventing depression in younger adolescents.[40]

Conduct disorder

Cognitive distortions and deficits

Both young children with oppositional disorders and adolescents with more severe conduct disorders show characteristic cognitive distortion in their thinking. They recall inaccurately high rates of hostile cues in social situations and when neutral remarks and movements are made by their peers, they see these as hostile.[41] In competitive situations with peers they exaggerate the aggressive behaviour of others and underestimate their own aggressiveness. These distorted attributions lead them into aggressive behaviour which then triggers angry behaviour from peers so that the originally neutral environment does indeed become more hostile.

Aggressive children and adolescents also have difficulties in problem-solving, both in experimental and naturalistic situations. They prefer rapid action-orientated solutions to those that require reflective thinking before any action is taken. Underlying this tendency to prefer rapid, aggressive solutions is the fact that their social goals relate more to the need for dominance and revenge than for affiliation.[42]

Parents of aggressive children also show cognitive distortions that are of relevance to the way they discipline their children.[43] This is of relevance both to the understanding and the management of childhood conduct disorders. For example, it has been shown that mothers of children with conduct disorder tend to attribute their children's difficult behaviour to deliberate wilfulness that is not within their children's capacity to control. They perceive themselves as helpless in the face of their children's behaviour. These cognitive distortions prevent them from acting effectively as parents, for example by drawing firm boundaries between acceptable and unacceptable behaviour.

Techniques of assessment and intervention

Until the early 1980s there were really no effective, evidence-based psychological interventions for children and adolescents with conduct disorder. Since that time a number of moderately effective psychological measures have been developed. All of these, including the cognitive behavioural techniques described below are only likely to be successful if they are combined with psychosocial measures directed towards the family as well as with appropriate education. All approaches require preliminary assessment of the child and family to identify the severity of the disorder and the possible presence of co-morbid disorders as well as to determine suitability for the approach envisaged.

Cognitive approaches to conduct disorder have been summarized by Lochman *et al.*[44] *Problem-solving skills training (PSST)*

has been developed for children aged 7 to 13 years. The programme is delivered over 25 sessions.[45] The group leaders teach problem-solving skills such as generating multiple solutions to a problem and reflecting on the different consequences of the alternatives. The skills are applied to interpersonal situations with teachers, parents, peers, and siblings. Parent participation is a major component of the training, with parents observing the sessions and acting as co-therapists in supervising the use of the new skills in the home.

The Anger Coping and Coping Power Programme is a school-based prevention programme delivered in group sessions to 13–14-year-old children.[46] The group sessions focus on enhancing emotional awareness, anger management training, attribution retraining and perspective-taking, social problem-solving and social skills training, behavioural and personal goal-setting, and handling peer pressure.

Multi-system therapy was designed as a multi-level intervention for 12 to 15-year-olds with multiple, severe antisocial problems.[47] Highly trained and closely supervised psychologists manage individualized programmes in the home setting. A variety of treatment approaches, including parent training, family therapy, school consultation, and individual therapy are employed in association with social measures such as helping lone mothers to find employment are used. The aim is to achieve change in one area before targeting another.

Functional Family Therapy combines family systems and cognitive behavioural approaches. The programme begins with an engagement and motivation phase in which the therapist addresses maladaptive beliefs in the family system thus aiming to increase expectations for change, reduce negativity and blaming, build respect for individual differences and develop a strong alliance between family members and the therapist. Practical behavioural interventions are designed to produce change and this is followed by a generalization phase in which the family is encouraged to interact effectively with the various systems in the community with which it is in contact.

Parent Management Training Programmes usually involve parents of young children with oppositional or conduct disorders. They derive from work originally carried out by Patterson and his colleagues at the Oregon Social Learning Center. His findings established the importance of coercive parental behaviour in the development of childhood aggression. Many treatment programmes have been based on their work,[48,49] which focus on ways of reducing parental coerciveness, often in group settings. Parents are taught to pinpoint problem behaviours, to apply positive reinforcement when their children's behaviour is more appropriate and to learn problem-solving and negotiating techniques. It has been suggested that the incorporation of a cognitive component into parent training, using a 'thoughts, feelings, behaviour cycle' can improve effectiveness of this approach.[50]

Evaluation of efficacy and effectiveness

All of the above programmes have been evaluated in controlled clinical trials and have been shown to be moderately effective.[51–53] However, virtually all the controlled studies have been carried out in highly resourced specialist centres. There is a conspicuous lack of studies of effectiveness carried out in routine clinical care.

Attention deficit hyperactivity disorder (ADHD)

Cognitive distortions and deficits

Children with ADHD show a range of cognitive deficits of attention and concentration with a strong predisposition to impulsivity, accompanied by explosive temperament and poor regulation of affect and impulses. Until recently these problems have been explained on the basis of deficits in one of two cognitive pathways. It has been proposed that there is a deficit in executive function, based on deficient inhibitory control arising from frontodorsal striatal brain networks.[54] Such failure of control results in deficits in self-monitoring, planning, attentional control, and executive skills. Stimulant medication remedies these deficits by increasing the activity of inhibitory pathways. Alternatively the condition has been attributed to disturbances in motivational processes, manifest as aversion to delay in gratification. More recently Sonuga-Barke,[55] has proposed that both these mechanisms are supported by the evidence and that there are two distinct but complementary neuro-developmental bases for ADHD which is thus, at least in the pre-school period, psychologically heterogeneous. As children get older and executive function matures, deficits in executive functions may become more prominent in affected children, especially in the areas of inhibition, set shifting, working memory, planning, and fluency.[56]

Techniques of assessment and intervention

Information about children with suspected ADHD needs to be obtained from both parents and school teachers as the child is likely to behave differently in the two settings. Both interviews and rating scales should be used. Observation of the child can confirm the presence of ADHD, but cannot rule it out as some children who are clearly showing symptomatology both at home and in school may appear normal in the clinic. Although it is helpful to reach a diagnosis, increasingly treatment approaches are focusing on the presence of specific impairments rather than on the presence of symptoms.[57]

Three types of psychological interventions have been found of value: ensuring the child's environment is structured and, when the child is engaged in a task that there is an absence of extraneous, distracting stimuli; counselling to parents and teachers, and behavioural and/or cognitive behavioural approaches directed to the child.

In the classroom the child will benefit if seated close to the teacher, task demands are kept short and there are interspersed periods of physical exercise. Teachers should be helped to reduce negative interactions by focusing on positive reinforcement for appropriate behaviour however brief this might be. Short periods of timeout before potentially problematic situations get out of hand may reduce the number of painful, angry confrontations. Similar principles can be applied in the home situation with parents being helped to understand and act on the principles of the identification of antecedents that result in problematic behaviour which will then have consequences that either increase or reduce the likelihood of recurrence. Positive reinforcement can be provided in the form of star charts, tokens, or other rewards. Training of parents of children with ADHD follows similar lines to training of parents with children with conduct disorder (see above) and, of course, many children show co-morbid ADHD and conduct disorder.

Cognitive behavioural approaches generally aim to achieve increased self-control. Most approaches involve encouraging appropriate self-instruction. The child is taught separate steps of self-instruction ('Stop: What is the problem?—Are there possible plans?—What is the best plan?—Do the plan—Did the plan work?) This approach can be applied when the child is faced with cognitive tasks which would usually be tackled impulsively or to social situations that often result in confrontations, such as arguments with parents or friends.[58]

Evaluation of effectiveness and efficacy

The delivery of behavioural treatment to children with ADHD presents particular problems. As a group they are slow to respond to conditioning procedures. Their distractibility and short attention span leads to problems in co-operation. Parents are likely to show similar behavioural and cognitive characteristics to their children, so collaboration of parents in treatment regimes may be problematic. The children's lack of reflectivity is a barrier to the use of cognitive approaches.

There is no good evidence that cognitive approaches alone are significantly effective in children with severe ADHD.[59] The most thorough evaluation of behavioural approaches to date is the Multi-modal Treatment of Attention-Deficit Hyperactivity Disorder (MTA) study carried out in the 1990s. The 579 children in the study were randomly allocated to one of three conditions: medication with stimulants, intensive behavioural treatment, and a combination of the two. The behavioural approach involved a parent training component, a two-part school intervention and an intensive summer treatment programme. It can therefore hardly be regarded as typical of psychological interventions applied in everyday clinical practice. There were slight advantages to combined treatment over medication alone. Behavioural treatments alone were much less effective for ADHD, though more useful for co-morbid anxiety disorders.[60] A 9-month follow-up revealed that the effectiveness of behavioural management approaches had been maintained over this period.[61] Interpretations of the findings of this study have been divergent. Re-analysis suggests that it may well be that medication alone or in combination with behavioural treatment is strongly indicated in severely affected children, while behavioural treatment and parent training are equally effective where impairment is mild or moderate.[58]

Parent training alone is effective in pre-school children with mild or moderately severe ADHD when delivered in a specialist setting, but is not when provided as part of routine primary care by non-specialist nurses.[62]

Obsessive–compulsive disorders

Cognitive distortions and deficits

The core cognitive distortion in children and adolescents with obsessive–compulsive disorder (OCD) is thought to lie, as it does with adult patients, in the appraisal of responsibility.[63] This is defined as 'the belief that one has power which is pivotal to bring about or prevent subjectively crucial negative outcomes'. Now we all do have responsibility for our actions; what makes patients with OCD different is that they take upon themselves quite unreasonable levels of responsibility. A 13-year-old might, for example, think 'I am responsible for making sure my mother does not die'. This sense of responsibility leads to attempts both to suppress and to neutralize the unwelcome thoughts of responsibility.

'Neutralizing' is defined as voluntary activity intended to have the effect of reducing the perceived responsibility. 'If I tap on my glass three times before I drink from it, my mother will not die'. But neutralizing activities increase discomfiting cognitions and this leads to further neutralizing activity. Attempts to suppress the intrusive thoughts also increase the likelihood of their recurrence. An additional complicating feature of the cognitive distortion is that, in the mind of the child with OCD, thoughts become imbued with unrealistic or magical powers. It is enough just to have a thought for it to be translated into action, so-called 'thought-action fusion'. 'If I allow myself to think about my mother dying this will mean that she will die'.

In general the cognitive distortions made by children and adolescents with OCD are similar to those seen in adults. However in a study comparing the various components of OCD cognitions in children, adolescents and adults it was found that children experienced fewer intrusive thoughts and these were less distressing and less uncontrollable than those experienced by adolescents and adults.[64] On the other hand, cognitive processes of thought-action fusion, perceived severity of harm, self-doubt and cognitive control were similar across the three age groups.

Techniques of assessment and intervention

The aim of cognitive therapy is to help the patient reach the view that obsessional thoughts, however distressing, are irrelevant to any activities that may be undertaken in the future. This is achieved by increasing the patient's sense of personal efficacy, predictability, controllability, and self-attributed likelihood of a positive outcome. The techniques used involve the conduct of tasks involving exposure to feared stimuli as well as response prevention, stopping the activities that reinforce the unwelcome thoughts.

The most widely applied treatment approach to OCD in children and adolescents is that developed by John March and his colleagues.[65] The treatment protocol involves 12 sessions of which the first two are spent on psycho-education and cognitive training and the next 10 sessions on exposure and response prevention with the first and last two sessions, as well as an intermediate session involving parents. The effectiveness of exposure depends on the fact that anxiety diminishes after repeated contact with a feared stimulus. Thus the anxiety of a child worried about germs will be reduced by prolonged contact with a surface the child thinks has germs on it. Encouraging parents not to provide reassurance to children who compulsively and repetitively demand it, removes reinforcement, and results in extinction of the behaviour. Some children become extremely distressed when their parents, on instruction, fail to provide such reassurance; more success is achieved by putting the child in control of reducing parents' inappropriate behaviour. Modelling and shaping behaviours are also helpful in giving the children or adolescents the skills to expose themselves to feared stimuli. Liberal use of rewards when the child behaves appropriately is also helpful in reinforcing desired behaviour.

Evaluation of effectiveness and efficacy

The most informative findings on efficacy come from the Pediatric OCD Treatment study (POTS) Team.[66] 112 patients aged from 12 to 17 years, suffering from OCD were divided randomly into four

groups: CBT alone, sertraline alone, combined sertraline, and CBT treatment and a pill placebo. Both sertraline alone and CBT alone were superior in outcome to pill placebo at 12 weeks after the beginning of treatment. But combined treatment was superior to both treatments administered separately with a clinical remission rate of 54 per cent, compared to 4 per cent for placebo.

Most studies report the results of individual treatment with children and adolescents, with limited input from parents. Initial findings suggest that for middle-school aged children, aged 8–14 years, CBT delivered with a stronger focus on parental involvement than is usually the case with adolescents is effective in reducing symptomatology[67] at least in the short-term. Success has also been reported for similar treatment provided in a group format. Group CBT is as effective as sertraline, and shows better results than sertraline at 9 months follow-up.[68] The presence of tics does not reduce the effectiveness of CBT in the treatment of OCD.[69] There are few studies investigating the longer-term effect of CBT on OCD. However it has been shown that improvement after both individual and group therapy is maintained for at least 18 months without attenuation.[70]

Application of CBT for miscellaneous purposes

There are a number of other conditions and adverse psychosocial situations occurring in childhood and adolescence in which the use of CBT is an important component of management. For a further discussion of these conditions and their management, see other sections of this book.

Chronic fatigue syndrome (CFS)

In this condition, characterized by severe fatigue and overwhelming exhaustion, with excessive sleepiness and a variety of other unexplained physical complaints, cognitive distortions involving an enhanced tendency to believe in the presence of disease in the absence of medical evidence (illness attribution), and deficits in the use of problem-solving techniques related to illness and disability have been identified.[71] The illness is not uncommon, occurring in around 2 per 1000, 11 to 15-year-olds. Rehabilitative methods, including the use of CBT have been found to be successful in adults and are also employed in the paediatric age group. A controlled clinical trial has found 10–17-year-olds with CFS to show greater improvement with CBT than a waiting list control group.[72]

Substance abuse

Cognitive distortions in young people presenting with substance abuse commonly relate to denial they have a serious, ultimately life-threatening problem, unwillingness to believe that effective help is available, and lack of belief in their own self-efficacy to change their behaviour. There is increasing evidence from controlled clinical trials that cognitive behaviour therapies can achieve positive results.[73] Motivational interviewing preceding the use of CBT is important with many children and adolescents and this is likely to be particularly the case with those suffering from substance abuse.

Eating disorders

Central features of both anorexia nervosa and bulimia nervosa include distorted cognitions about shape and weight. Cognitive behavioural approaches used with adults with these conditions require modification when used with adolescents, with greater emphasis on involvement of parents.[74] While CBT is the most effective treatment for bulimia in older adolescents and adults,[75] family counselling is now established as the most effective intervention for anorexia nervosa in younger patients.[76]

Post-traumatic stress disorder

This condition is characterized by disorders of thinking including repetitive, intrusive thoughts, phobic avoidance of the situation in which the individual was exposed to trauma, 'survivor guilt', and problems in concentration. CBT is the most effective, evidence-based technique in the management in both children and adolescents.[77,78]

Non-organic pain

Abdominal pain and headache for which no physical cause can be found are commonly seen in primary health care. Although when these conditions occur it is often difficult to establish a psychological mechanism for the pain, it is reasonably well established that management based on CBT is the most effective approach. CBT is also effective in reducing pain from organic disease as well as in reducing distress when painful paediatric procedures are carried out. For a review of the use of CBT in the management of pain in childhood, see McGrath and Goodman.[79]

Adverse psychosocial situations

Children and adolescents in adverse psychosocial situations are frequently troubled by distorted perceptions of their predicament. In particular, they may feel themselves responsible for the separation and divorce of their parents or that they have deserved the maltreatment, either physical or sexual, inflicted on them by adults who have abused them. CBT has a significant part to play in helping children adjust to parental separation and divorce.[80] It has also been shown to have demonstrable value when applied to victims of sexual abuse.[81]

Further information

Graham, P. (2005). *Cognitive-behaviour therapy for children and families* (2nd edn). Cambridge University Press, Cambridge.

Stallard, P. (2005). *A clinician's guide to think good-feel good*. John Wiley, Chichester.

Kazdin, A. and Weisz, J. (2003). *Evidence-based psychotherapies for children and adolescents*. Guilford Press, New York.

References

1. Piaget, J. (1970). Piaget's theory. In *Carmichael's manual of child psychology*, Vol. 1 (ed. P.H. Mussen), pp. 703–32. Wiley, New York.
2. Beck, A. (1964). Thinking and depression: 2. Theory and therapy. *Archives of General Psychiatry*, **10**, 561–71.
3. Wellman, H.M., Hollander, M., and Schult, C.A. (1996). Young children's understanding of thought bubbles and thoughts. *Child Development*, **67**, 768–88.
4. Hoare, P., Elton, R., Greer, A., *et al.* (1993). The modification and standardisation of the Harter self-esteem questionnaire with Scottish school children. *European Child & Adolescent Psychiatry*, **2**, 19–33.
5. Turner, J.E. and Cole, D.A. (1994). Developmental differences for cognitive diathesis in childhood depression. *Journal of Abnormal Child Psychology*, **22**, 15–32.

6. Young, E.J. (1990). *Cognitive therapy for personality disorders: a schema-focused approach*. Professional Resource Exchange, Sarasota, FL.
7. Stallard, P. and Rayner, H. (2005). The development and preliminary evaluation of a Schema Questionnaire for Children (SQC). *Behavioural and Cognitive Psychotherapy*, **33**, 217–24.
8. Kendall, P. (1991). *Child and adolescent therapy: cognitive-behavioral procedures*, p. 6. Guilford Press, New York.
9. Stallard, P. (2005). *A clinician's guide to think-good, feel-good*. John Wiley, Winchester.
10. Wilson, C. and White, C. (2006). A preliminary investigation of the effect of intervention on parental attributions and reported behaviour. *Behavioural and Cognitive Psychotherapy*, **34**, 503–7.
11. Wolpert, M., Doe, J., and Elsworth, J. (2005). Working with parents: ethical and practical issues. In *Cognitive behaviour therapy for children and families* (ed. P.J. Graham), pp. 103–20. Cambridge University Press, Cambridge.
12. Schmidt, U. (2005). Engagement and motivational interviewing. In *Cognitive behaviour therapy for children and families* (ed. P.J. Graham), pp. 67–83. Cambridge University Press, Cambridge.
13. Kingdon, D., Hansen, L., Finn, M., *et al.* (2007). When standard cognitive-behavioural therapy is not enough. *Psychiatric Bulletin*, **31**, 121–3.
14. Kendall, P., Chansky, T.E., Friedman, M., *et al.* (1991). Treating anxiety disorders in children and adolescents. In *Child and adolescent therapy* (ed. P.C. Kendell), pp. 131–64. Guilford Press, New York.
15. Allen, J.L. and Rapee, R.M. (2005). Anxiety disorders. In *Cognitive behaviour therapy for children and families* (ed. P.J. Graham), pp. 300–19. Cambridge University Press, Cambridge.
16. Cobham, V., Dadds, M., and Spence, S. (1998). The role of parental anxiety in the treatment of childhood anxiety. *Journal of Consulting and Clinical Psychology*, **66**, 893–905.
17. Dacey, J.S. and Fiore, L.B. (2000). *Your anxious child*. Jossey-Bass, San Francisco.
18. Cartwright-Hatton, S., Roberts, C., Chitsabesan, P., *et al.* (2004). Systematic review of the efficacy of cognitive behaviour therapies for childhood and adolescent anxiety disorders. *The British Journal of Clinical Psychology*, **43**, 421–36.
19. Heyne, D., King, N.J., and Tonge, B., *et al.* (2002). Evaluation of child therapy and caregiver training in the treatment of school refusal. *Journal of the American Academy of Child and Adolescent Psychiatry*, **41**, 687–95.
20. Nauta, M.H., Scholing, A., Emmelkamp, P.M., *et al.* (2003). Cognitive-behavioral therapy for children with anxiety disorders in a clinical setting: no additional effect of a cognitive parent training. *Journal of the American Academy of Child and Adolescent Psychiatry*, **42**, 1270–8.
21. Spence, S.H., Donovan, C., and Brechman-Toussaint, M. (2000). The treatment of childhood social phobia: the effectiveness of a social skills training-based, cognitive-behavioural intervention, with and without parental involvement. *Journal of Child Psychology and Psychiatry*, **41**, 713–26.
22. Wood, J.J., Piacentini, J.C., Southam-Gerow, M., *et al.* (2006). Family cognitive behavioral therapy for child anxiety disorders. *American Academy of Child and Adolescent Psychiatry*, **45**, 314–21.
23. Spence, S.H., Holmes, J.M., March, S., *et al.* (2006). The feasibility and outcome of clinic plus internet delivery of cognitive-behavior therapy for childhood anxiety. *Journal of Consulting and Clinical Psychology*, **74**, 614–21.
24. Kendall, P. and Sugarman, A. (1997). Attrition in the treatment of childhood anxiety disorders. *Journal of Consulting and Clinical Psychology*, **40**, 787–94.
25. Rapee, R.M. (2003). The influence of comorbidity on treatment outcome for children and adolescents with anxiety disorders. *Behaviour Research and Therapy*, **41**,105–12.
26. Bernstein, G.A., Borchardt, C.M., Perwien, A.R., *et al.* (2000). Imipramine plus cognitive-behavioral therapy in the treatment of school refusal. *Journal of the American Academy of Child and Adolescent Psychiatry*, **39**, 276–83.
27. Barrett, P.M., Duffy, A.L., Dadds, M.R., *et al.* (2001). Cognitive-behavioral treatment of anxiety disorders in children: long-term (6-year) follow-up. *Journal of Consulting and Clinical Psychology*, **69**, 135–41.
28. Barrett, P., Farrell, L., Ollendick, T., *et al.* (2006). Long term outcomes of an Australian universal prevention program of anxiety and depressive symptoms in children and youth. *Journal of Clinical Child and Adolescent Psychology*, **35**, 403–11.
29. Stallard, P., Simpson, N., Anderson, S., *et al.* (2007). The friends emotional health programme: initial findings from a school-based project. *Child and Adolescent Mental Health*, **12**, 32–7.
30. Kagan, J. (1982). The emergence of self. *Journal of Child Psychology and Psychiatry*, **23**, 363–81.
31. Terwogt, M.M. and Stegge, H. (1995). Emotional behaviour and emotional understanding: a developmental fugue. In *The depressed child and adolescent: developmental and clinical perspectives* (ed. I.M. Goodyer), pp. 27–52. Cambridge University Press, Cambridge.
32. Harrington, R. (2005). Depressive disorders. In *Cognitive behaviour therapy for children and families* (ed. P.J. Graham), pp. 263–80. Cambridge University Press, Cambridge.
33. Kovacs, M. (1981). Rating scales to assess depression in school-aged children. *Acta Paedopsychiatrica*, **46**, 305–15.
34. Angold, A., Costello, E.J., Messer, S.C., *et al.* (1995). The development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research*, **5**, 237–49.
35. Compton, S.N., March, J.S., Brent, D., *et al.* (2004). Cognitive-behavioral therapy for anxiety and depressive disorders in children and adolescents: an evidence-based review. *Journal of the American Academy of Child and Adolescent Psychiatry*, **43**, 930–59.
36. March, J., Silva, S., Vitiello, B., and TADS Team. (2006). The Treatment for Adolescents with Depression Study (TADS): methods and message at 12 weeks. *Journal of the American Academy of Child and Adolescent Psychiatry*, **45**, 1393–403.
37. Jayson, D., Wood, A.J., Kroll, L., *et al.* (1998). Which depressed patients respond to cognitive-behavioral treatment? *Journal of the American Academy of Child and Adolescent Psychiatry*, **37**, 35–9.
38. Spence, S.H., Sheffield, J.K., and Donovan, C.L. (2005). Long-term outcome of a school-based, universal approach to prevention of depression in adolescents. *Journal of Consulting and Clinical Psychology*, **73**, 160–7.
39. Clarke, G.N., Hornbrook, M., Lynch, F., *et al.* (2001). A randomized trial of a group cognitive intervention for preventing depression in the offspring of depressed parents. *Archives of General Psychiatry*, **58**, 1127–34.
40. Merry, S., McDowell, H., Wild, C., *et al.* (2004). A randomized placebo-controlled trial of a school-based depression prevention program. *Journal of the American Academy of Child and Adolescent Psychiatry*, **43**, 538–47.
41. Dodge, K.A. and Frame, C.L. (1982). Social cognitive biases and deficits in aggressive boys. *Child Development*, **53**, 620–35.
42. Lochman, J., Wayland, K., and White, K. (1993). Social goals: relationship to adolescent adjustment and to social problem solving. *Journal of Abnormal Child Psychology*, **21**, 135–51.
43. Baden, A.G. and Howe, G.W. (1992). Mother's attributions and expectancies regarding their conduct disordered children. *Journal of Abnormal Child Psychology*, **20**, 467–85.
44. Lochman, J., Phillips, N., McElroy, H., *et al.* (2005). Conduct disorders in adolescence. In *Cognitive behaviour therapy for children and families* (ed. P.J. Graham), pp. 443–58. Cambridge University Press, Cambridge.
45. Kazdin, A., Siegel, T., and Bass, D. (1992). Cognitive problem solving skills training and parent management in the treatment of antisocial

behavior in children. *Journal of Consulting and Clinical Psychology*, **60**, 733–47.
46. Lochman, J. and Wells, K. (2002). The coping power programme at the middle school transition. *Psychology of Addictive Behaviors*, **16**, S40–S54.
47. Henggeler, S. (1999). Multisystemic therapy: an overview of clinical procedures, outcomes and policy implications. *Child Psychology and Psychiatry Review*, **4**, 2–10.
48. Webster-Stratton, C., Hollinsworth, T., and Kolpacoff, M. (1989). The long-term effectiveness and clinical significance of three cost-effective training programs for families with conduct-problem children. *Journal of Consulting and Clinical Psychology*, **57**, 550–53.
49. Sanders, M., Markie-Dadds, C., Tully, L., *et al.* (2000). The triple-P Positive Parenting Program: a comparison of enhanced, standard and self-directed behavioral family intervention for parents of children with early onset conduct problems. *Journal of Consulting and Clinical Psychology*, **68**, 624–60.
50. White, C., McNally, D., and Cartwright-Hatton, S. (2003). Cognitively enhanced parent training. *Behavioural and Cognitive Psychotherapy*, **31**, 99–102.
51. Kazdin, A. and Wassell, G. (2000). Therapeutic changes in children, parents and families resulting from treatment of children with conduct problems. *Journal of the American Academy of Child and Adolescent Psychiatry*, **39**, 414–20.
52. Van de Wiel, N., Matthys, W., Cohen-Kettenis, P., *et al.* (2002). Effective treatments of school-aged conduct disordered children: recommendations for changing clinical and research practices. *European Child & Adolescent Psychiatry*, **11**, 79–84.
53. Scott, S., Spender, Q., Doolan, M., *et al.* (2001). Multicentre controlled trial of parenting groups for childhood antisocial behaviour. *British Medical Journal*, **323**, 194–8.
54. Sonuga-Barke, E. (2003). The dual pathway of AD/HD: an elaboration of neuro-developmental characteristics. *Neuroscience and Biobehavioral Reviews*, **27**, 593–604.
55. Sonuga-Barke, E. (2005). Causal models of attention-deficit/hyperactivity disorder: from common simple deficits to multiple developmental pathways. *Biological Psychiatry*, **57**, 1231–8.
56. Sergeant, J., Geurts, H., and Oosterlaan, J. (2002). How specific is a deficit of executive functioning in attention-deficit/hyperactivity disorder? *Behavior Brain Research*, **130**, 3–28.
57. Pelham, W.E. and Walker, K.S. (2005). Attention deficit hyperactivity disorder. In *Cognitive behaviour therapy for children and families* (ed. P.J. Graham), pp. 225–43. Cambridge University Press, Cambridge.
58. Taylor, E., Dopfner, M., Sergeant, J., *et al.* (2004). European clinical guidelines for hyperkinetic disorder—first upgrade. *European Child & Adolescent Psychiatry*, **13**(Suppl. 1), 7–30.
59. Abikoff, H. (1991). Cognitive training in ADHD children: less to it than meets the eye. *Journal of Learning Disabilities*, **24**, 205–9.
60. Jensen, P., Hinshaw, S., Swanson, J., *et al.* (2001). Findings from the NIMH Multimodal Treatment Study of ADHD (MTA): implications and applications for primary care providers. *Journal of Behavioral and Developmental Pediatrics*, **22**, 60–73.
61. Arnold, L., Chuang, S., Davies, M., *et al.* (2004). Nine months of multi-component behavioural treatment for ADHD and effectiveness of MTA fading procedures. *Journal of Abnormal Child Psychology*, **32**, 39–51.
62. Sonuga-Barke, E., Thompson, M., Daley, D., *et al.* (2004). Parent training for attention deficit/hyperactivity disorder: is it as effective when delivered as routine rather than as specialist care? *The British Journal of Clinical Psychology*, **43**, 449–57.
63. Salkovskis, P.M. and Kirk, J. (1997). Obsessive-compulsive disorder. In *Science and practice of cognitive behaviour therapy* (eds. D.M. Clark and C.G. Fairbairn), pp. 179–208. Oxford Medical Publications, Oxford.
64. Farrell, L. and Barrett, P. (2006). Obsessive-compulsive disorder across developmental trajectory: cognitive processing of threat in children, adolescents and adults. *The British Journal of Psychology*, **97**, 95–114.
65. March, J.S. and Mulle, K. (1998). *OCD in children and adolescents: a cognitive-behavioral treatment manual*. Guilford Press, New York.
66. Pediatric, O.C.D. Treatment Study (POTS) Team. (2004). Cognitive-behavior therapy, sertraline, and their combination for children and adolescents with obsessive-compulsive disorder: the Pediatric OCD Treatment Study (POTS) randomized controlled trial. *The Journal of the American Medical Association*, **292**, 1969–76.
67. Martin, J. and Thienemann, M. (2005). Group cognitive-behavior therapy with family involvement for middle-school-age children with obsessive-compulsive disorder. *Child Psychiatry and Human Development*, **36**, 113–27.
68. Asbahr, F., Castillo, A., Ito, L., *et al.* (2005). Group cognitive-behavioral therapy versus sertraline for the treatment of children with obsessive-compulsive disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, **44**, 1128–36.
69. Himle, J., Fischer, D., Van Etten, M., *et al.* (2003). Group behavioral therapy for adolescents with tic-related and non-tic-related obsessive compulsive disorder. *Depression and Anxiety*, **17**, 73–7.
70. Barrett, P., Farrell, L., Dadds, M., *et al.* (2005). Cognitive-behavioral family treatment of childhood obsessive-compulsive disorder: long-term follow-up and predictors of outcome. *Journal of the American Academy of Child and Adolescent Psychiatry*, **44**, 1005–14.
71. Garralda, E. and Chalder, T. (2005). Chronic fatigue syndrome in childhood. *Journal of Child Psychology and Psychiatry*, **46**, 1143–51.
72. Stulemeijer, M., De, J., Lieke, W., *et al.* (2005). Cognitive behaviour therapy for adolescents with chronic fatigue syndrome. *British Medical Journal*, **330**, 14.
73. Waldron, H. and Kaminer, Y. (2004). On the learning curve: the emerging evidence supporting cognitive-behavioral therapies for adolescent substance abuse. *Addiction*, **99**(Suppl. 2), 93–105.
74. Lock, J. (2002). Treating adolescents with eating disorders within the family context: empirical and theoretical considerations. *Child and Adolescent Psychiatric Clinics*, **11**, 331–42.
75. Thiels, C., Schmidt, U., Treasure, J., *et al.* (1998). Guided self-change for bulimia nervosa incorporating use of a self-care manual. *The American Journal of Psychiatry*, **155**, 947–53.
76. Eisler, I., Dare, C., Hodes, M., *et al.* (2000). Family therapy for adolescent anorexia nervosa: the results of a controlled comparison of two family interventions. *Journal of Child Psychology and Psychiatry*, **41**, 727–36.
77. Perrin, S., Smith, P., and Yule, W. (2000). The assessment and treatment of post-traumatic stress disorder in children and adolescents. *Journal of Child Psychology and Psychiatry*, **41**, 277–89.
78. Deblinger, E., Mannarino, A., Cohen, J., *et al.* (2006). A follow-up study of a multisite, randomized, controlled trial for children with sexual abuse-related PTSD symptoms. *Journal of the American Academy of Child and Adolescent Psychiatry*, **45**, 1474–84.
79. McGrath, P. and Goodman, P. (2005). Pain in childhood. In *Cognitive behaviour therapy for children and families* (ed. P.J. Graham), pp. 426–42. Cambridge University Press, Cambridge.
80. Herbert, M. (2005). Adjustment to parental separation and divorce. In *Cognitive behaviour therapy for children and families* (ed. P.J. Graham), pp. 170–83. Cambridge University Press, Cambridge.
81. Jones, D. and Ramchandani, P. (1999). *Child sexual abuse: informing practice from research*. Radcliffe Medical Press, Abingdon.

9.5.4 Caregiver-mediated interventions for children and families

Philip A. Fisher and Elizabeth A. Stormshak

This chapter summarizes interventions that have been developed to address child and adolescent behaviour problems and externalizing disorders within the therapeutic milieu of the family. Although it has long been recognized that caregiver-mediated treatments can be employed to address children's problems, research with families in the past two decades has resulted in numerous systematic, theory-driven approaches that have been subjected to rigorous scientific evaluation and have been found to be effective at improving outcomes. Although no intervention is certain to work for every child and it is not possible to engage every family in the intervention process, caregiver-mediated interventions are among the most promising approaches currently available to practitioners.

In recent years, progress in the field of caregiver-mediated interventions has included an expansion of the evidence base supporting specific intervention practices for use with the general population, with high-risk segments of the population (e.g. children in foster care and children in Head Start settings), and with underserved populations (e.g. girls and racial/ethnic minorities). In addition, an increasing emphasis has been placed on the dissemination of proven interventions on large-scale bases within community settings in North America, Europe, and Australia. Evidence is currently being gathered to evaluate the impact of many of these large-scale dissemination efforts. The chapter that follows contains background information on the theoretical underpinnings of caregiver-mediated interventions to address child behaviour problems. Specific interventions that have been developed for children in specific age groups—prenatal through early childhood, the school-age period, and adolescence—are then described. Finally, we discuss adaptations that have been made to address issues of gender and cultural diversity within this field.

Before providing a background on caregiver-mediated interventions, a disclaimer is necessary. The term 'caregiver-mediated' is employed throughout this chapter, rather than the term 'family-mediated', to convey the sense that these interventions need not occur specifically within the context of the child's biological family. Recognition of the diversity of family types in which children are raised requires a shift from a nuclear family conceptualization to include multigenerational families, lesbian/gay/bisexual families, and other nontraditional family configurations. In addition, many children are reared in contexts that include no direct biological relatives. For instance, increasing numbers of children are reared in foster care. To a certain extent, this is indicative of the need to address and prevent child maltreatment and to provide services that allow children to remain in their biological families. In addition, it represents a positive development to the extent that many children who have previously been cared for in institutional settings are now being placed in community families, which have the potential to provide more adequate rearing environments than institutions. However, it is often the case that caregivers in these foster/adoptive families will require additional support services to improve outcomes for children. Thus, we have adopted the term caregiver-mediated interventions to reflect the spectrum of existing rearing environments.

Background on caregiver-mediated interventions

There is no single predominant cause of the development of behaviour problems in an individual child. Rather, as is noted in Patterson *et al.*[1] research has implicated a broad array of factors that contribute to behaviours at virtually every level of analysis. From a *societal perspective*, factors such as poverty, discrimination, and unemployment have been implicated as contributing to higher rates of disruptive behaviour in children. In addition, children in various underprivileged contexts (e.g. children in foster care or with incarcerated parents) show higher rates of disruptive behaviour. *Neighbourhood factors* such as crowded living conditions and violence also appear to be associated with higher rates of behaviour problems. At the *individual level*, psychosocial and neurobiological factors have been associated with higher rates of behaviour problems. For example, poor social skills, and low cognitive functioning appear to be linked to the development of behaviour problems. Recent evidence also indicates that genetic factors play a role in whether or not a child develops externalizing behaviour.[2]

With this multitude of factors implicated in the development of disruptive behaviour, one might question why caregiver-mediated interventions have become so predominant. The answer is straightforward. In as much as children exist within the context of their families, caregivers exert the single largest influence on children's behavioural and developmental outcomes. Numerous longitudinal studies have shown that caregiver factors predict over and above individual-, neighbourhood-, and society-level variables in determining trajectories towards disruptive behaviour.[1] Moreover, randomized trials of caregiver-mediated interventions have provided information that manipulating caregiver behaviours (e.g. teaching them to use effective parenting strategies) is extremely powerful for improving child outcomes.[3] Thus, although child behaviour problems are multidetermined, the most direct method for improving child outcomes appears to be via the caregiver–child relationship.

The interventions that are described below are categorized by developmental epoch. It is also important to recognize that interventions exist across the spectrum of risk, including proven interventions to reduce problem behaviours in children with disruptive behaviour disorders and preventive efforts to deflect children from developing disruptive behaviour disorders. An overall synopsis of the interventions to be described is that they have shown great promise for addressing problems at a number of levels. Although a number of issues still confront the field, the progress that has been made in the last two decades has been remarkable in demonstrating that it is possible to address child behaviour problems through caregiver-mediated interventions.

Prenatal and early childhood caregiver-mediated interventions

Caregiver-mediated interventions for the prenatal period through infancy and early childhood are generally oriented towards

prevention. That is, rather than addressing concurrent problems with the child, they are based on the supposition that targeting known precursors of child problems is an effective way to prevent those problems. Among the most influential work in this area has been the home visitation programme developed and evaluated by Olds and colleagues to prevent antisocial behaviour in children.[4] This project involved a randomized control trial that included longitudinal data collection over more than 25 years.[5] A total of 400 pregnant women were enrolled in the original study according to one of the three following criteria: under 19 years of age, unmarried, or of low socioeconomic status. Those assigned to the intervention condition received an average of 9 home visits during pregnancy and an average of 23 home visits between birth and the child's second birthday. The home visits were conducted by nurses. The focus of visits was upon prenatal and neonatal maternal health behaviours, child care skills, and maternal life issues (e.g. education and employment).

Olds and colleagues reported 15-year outcomes for the children of the mothers in this study. Those in the intervention group reported fewer arrests than those in a comparison group.[6] In addition, among those in higher risk categories (as indicated by their mother being both low SES and unmarried at the time of the child's birth), youth in the intervention group reported lower rates of running away, arrests, criminal convictions and parole violations, and smoking and alcohol consumption than did youth in the comparison group. The authors concluded that this approach may be an effective means of preventing early-onset conduct disorder, which has been considered the more treatment-resistant and complex form of the disorder.[1] However, it is also noteworthy that, in subsequent randomized trial studies to examine how variations in the programme structure affect outcomes, the intervention was found to be less effective when paraprofessionals were used in place of nurses as home visitors.[7] More recently, and consistent with other evidence-based programmes, the emphasis of Olds and colleagues has shifted to a wide-scale dissemination of the intervention in community settings.[8] Within this context, emphasis has been placed on maintaining the programme's effectiveness by developing practices to ensure that the intervention is delivered with high fidelity to the original model.

It is important to recognize that not all caregiver-mediated preventive interventions for children in this age group have shown positive effects. For example, the Healthy Start programme was designed to prevent child abuse and neglect, using an initial screening process (usually in hospital settings) to identify families at risk for child maltreatment and employing a home-visitor model of service delivery for the intervention. Although modest positive effects were observed in individual sites implementing this programme, the overall results from a large-scale randomized trial of this intervention did not support the efficacy of the intervention.[9] A related problem in the evaluation of the Healthy Start programme was low rates of family participation. Of those recruited, many received very few home visitation sessions. With such low dosage rates, it can be difficult to determine whether the lack of positive intervention effects was due to a failure of the approach to impact targeted behaviours or was due to families receiving too little of the intervention for it to have been effective. Thus, as is true across the span of child development, it is important to consider what the critical elements of effective interventions are and to make sure that these are included in any intervention efforts.

Caregiver-mediated interventions in the preschool years

The preschool years are marked by a number of important developmental changes. Dramatic increases in the use of language and physical mobility in addition to an increase in autonomous behaviour provide challenges for many parents. Preschool children need substantial support from parents and caretakers in socialization in their family and school environments. As a result, as with interventions in infancy, the majority of caregiver-mediated interventions at this age are focused on parenting, and the parent–child relationship.

Within the parenting literature, there are many different types of parenting skills that have been found to be important in promoting healthy child development. Typically, interventions include a dual focus on increasing positive parenting and decreasing negative parenting. Positive parenting usually refers to supportive, warm, involved parenting that includes praise, positive support, approval, and responsiveness to children and their needs.[10] Negative parenting usually refers to parenting deficits, including poor limit setting, inconsistency, verbal and physical aggression, and harsh discipline.[11]

Caregiver-based interventions at this age typically focus on building a strong positive parent–child relationship in addition to teaching specific behaviour management techniques to promote healthy child adjustment and prevent later problems. Although some programmes target at-risk children, such as Head Start children or those with early behaviour problems, there is some evidence that more severe child behaviour problems are related to more limited effects of parent training programmes.[12]

One of the more successful caregiver-based programmes for preschoolers, *The Incredible Years*, was developed by Webster-Stratton and colleagues.[13] This videotaped programme includes a number of salient parenting skills for preschoolers, starting with building a positive parent–child relationship, praise, and rewards and moving to limit setting, problem solving, and discipline. Randomized trial evaluations have suggested that this programme is effective when administered in group and individual settings.[14] Results have suggested that the programme improves parent–child interactions immediately post-test and 1 year later. The programme was subsequently expanded from basic parenting skills to include more advanced parenting, such as anger management, communication, and self-control skills, which is also effective at improving the parent–child relationship and reducing child behaviour problems at short-term follow-up. The advanced programme is able to deal effectively with parent–child problems and the mediators that might influence the parent's ability to effectively manage the child, including depression and the marital relationship. Overall, this parent training programme has been effective with a number of different age groups and populations, including preschool- and school-aged children as well as clinic-referred and community Head Start families.[15]

Other caregiver-mediated interventions involving preschoolers have also shown success at enhancing the parent–child relationship and decreasing child behaviour problems. For example, *Parent–Child Interaction Training (PCIT)* was originally developed for at-risk children enrolled in Head Start.[16] Training involved teaching parents effective play skills and positive interaction. The programme has been effective at reducing teacher ratings of behaviour

problems 1 year after the intervention was completed. This approach has subsequently been used for other populations. For example, Chaffin *et al*. reported on the results of using *PCIT* to prevent child abuse among families with a history of child welfare system involvement.[17] The intervention significantly reduced future reports of maltreatment. Interestingly, an enhanced version of the intervention designed to provide individualized services to meet families' needs was no more effective than the original *PCIT* intervention.

Home visitation programmes in preschool have also shown promise as effective interventions. Head Start, for example, provides home-visiting services to all families in the programme. The goals of these visits are to work with families on meeting the Head Start performance standards in four areas: education, health–nutrition–mental health, social services, and parent involvement. Research has indicated that children who received home services have shown improvements in the parent–child relationship and early academic achievement compared to children and families who did not receive the home visits or parenting model.[18] Home-visiting programmes at this age also provide families with social support, self-efficacy, and a positive therapeutic relationship with the visitor. This relationship serves to enable parents to process and understand parenting and family histories that impede the development of successful parenting skills with their own child.

Caregiver-mediated interventions in the school-aged years

As children move into the school years, they face a new set of challenges at home and in the school environment, including academic achievement, negotiating peer relationships, and the demands of teachers and the school context. At this age, children who have begun developing problems in the context of their families may generalize these problems to the school environment, which is associated with increased risk of later difficulties in addition to problems across multiple domains of functioning.[19] As a result, the transition to school serves as an important target of preventive intervention programmes.[20] Furthermore, school problems and peer difficulties may exacerbate problems in the home as parents struggle with issues such as homework completion and their child's social skills and changing peer network. Thus, effective interventions aimed at school-aged children must support families and parents through an emphasis on parent training and improving or maintaining academic achievement and positive peer relationships at school.

Although the emphasis on increasing positive parenting skills and decreasing negative parenting is typically maintained in interventions at this age and across development, additional components targeting academic achievement, learning and early literacy, and parent–school involvement are important aspects of comprehensive, caregiver-mediated interventions for school-aged children. There are multiple examples of comprehensive, caregiver-mediated prevention programmes that have been associated with positive outcomes for school-aged youth and families, including the *Families and Schools Together (FAST) Track* programme,[21] the *Linking the Interests of Families and Teachers (LIFT)* programme,[22] and the *Schools and Families Educating (SAFE) Children* programme.[23]

The *FAST Track* programme began in 1991 and targeted children in early elementary school at risk of developing later conduct disorder and delinquent behaviour problems.[21] Children and families received a multifocused intervention package targeting development across multiple domains, including peers, the school environment, academic achievement, and the family context. The family intervention integrated successful approaches to parent training regarding the development of school-aged children, including parent–school involvement and early reading.[21,24,25] Parents met in groups weekly during the first-grade year and bi-weekly in second grade. One hour of parent–child learning activity that emphasized positive parent–child interactions in a controlled environment and early literacy was also provided. Home visits and individualized programming were implemented to meet the special needs of families, such as stress management, marital problems, and maternal depression. The intervention continued through adolescence, with new components adapted to the changing development of children and families. Results of this programme indicated that it was successful at the end of first grade and third grade in improving outcomes across a number of different domains, including parenting and child peer relations, emotional understanding, and reading skills. Specifically, parents showed less physical discipline, more consistent and appropriate discipline, and more warmth and positive involvement in their child's school.[24,26,27]

The *LIFT* programme was designed as a preventive intervention for at-risk children to decrease the development of conduct problems and delinquency.[28,29] *LIFT* was designed for first- and fifth-graders living in at-risk neighbourhoods. Intervention components consisted of a school-based intervention focused on social skills and problem solving, a parent training group, a playground behavioural programme, and communication between parents and teachers. Creative intervention techniques were employed to increase parent participation in schooling; for example, a phone answering machine was installed in each classroom, and newsletters were sent home to parents about school *LIFT* activities. Parents could call in to the answering machine or leave messages for teachers at any time. This programme was successful at decreasing aggression on the playground and increasing positive behaviours with peers during the year following the intervention. Additionally, the programme was successful at decreasing negative parenting during observed mother–child interactions.[28]

The *SAFE Children* programme was administered within different schools in high-risk neighbourhoods in the Chicago area with a diverse group of families.[23] Youth were randomly assigned within classrooms to receive the intervention, which included parenting and academic tutoring components. The parenting component focused on disseminating information about child development, parenting, skill practice, and home assignments to increase parenting skills and on group problem solving around skills when parents needed additional assistance with implementation. The intervention targeted children during the transition to elementary school, based on developmental research and theory suggesting that transition points are key points for intervention.[30] The intervention significantly impacted academic performance and parental involvement in school, with additional outcomes for high-risk families that included higher parental monitoring and reductions in child behaviour problems.[23]

In summary, effective school-aged, caregiver-mediated interventions for families are multifocused, typically including foci on parenting skills and school success for children. The majority of

successful interventions at this age combine a family component with additional interventions to support school adjustment (e.g. social skills training, academic tutoring). Clearly, interventions that target parenting, families, and other domains of children's functioning are the most efficacious interventions to administer during the transition to school.

Caregiver-mediated interventions involving adolescents

By adolescence, many youth have achieved a degree of independence from their parents and are embedded in the culture of peers, school, and/or their community. The extent of adolescents' autonomy in today's world is such that there may be a temptation to consider addressing adolescent mental health problems in a different manner than those of younger children. However, among the interventions that have demonstrated the biggest impact for adolescents are those focused on youth in the context of their families.[31] The only difference between these programmes and those discussed for younger children is that there is an increased emphasis on parenting skills that are more relevant to adolescent youth, such as problem solving, helping your adolescent gain autonomy, and communication. Adolescents may be more active participants in the interventions that target this age group, attending family sessions and practicing the skills outside of the treatment at home or at school.

There have been a variety of programmes associated with positive outcomes for adolescent youth. At this age, effective interventions tend to be brief, family-centred approaches that teach parents the skills needed to parent effectively during adolescence. Caregiver-mediated interventions at this age may be delivered in the school context or in a community centre or clinic. Within the school context, several programmes have shown success at decreasing adolescent problem behaviour, including substance use over time.

The *Adolescent Transitions Program* (ATP) is a caregiver-mediated treatment for adolescent problem behaviour and substance use. Over the past 15 years, the *ATP* has been shown to be effective in a number of different randomized control trials with parents recruited from the community and schools.[32,33] The *ATP* curriculum focuses on building a positive parent–adolescent relationship, decreasing known risk factors at this age for problem behaviour (e.g. lack of parental monitoring), and increasing communication and listening skills. In the first series of research studies using the *ATP* , the intervention was delivered in a group format and was successful at reducing problem behaviour.[34] More recently, the *ATP* curriculum has been developed into a caregiver-mediated, tailored approach to treatment called the *Family Check-Up* (*FCU*).[31] The *FCU* is a preventative intervention based on a health maintenance model appropriate for at-risk and high-risk youth. The *FCU* incorporates the content from the *ATP* curriculum into a model that targets parent engagement in treatment. In the *FCU* model, families receive a comprehensive, ecological assessment, videotaped observation, and feedback using motivational interviewing to engage families in treatment. Across numerous randomized controlled intervention trials, the *FCU* has been shown to be effective at reducing teacher-reported risk behaviour, reducing arrest rates, increasing attendance and achievement at school, reducing substance use, and reducing antisocial behaviour.[35–37] Interestingly, these outcomes have been mediated by an increase in family management skills, including parental monitoring.[36]

There have been several other comprehensive, caregiver-mediated approaches to the treatment of adolescents that have also shown positive outcomes. Many of these models are school-based and preventive, focusing on at-risk and high-risk youth and families via group format from schools. Spoth *et al.*[38] examined outcomes associated with both the *Preparing for the Drug Free Years* programme[39] and the *Iowa Strengthening Families Program*.[40] Both of these parenting programmes have been hypothesized to reduce the initiation of substance use in the high school years and to reduce antisocial behaviour and were both strength based and focused on teaching parents the skills necessary to prevent adolescent problem behaviour. Schools were randomly assigned to deliver the interventions at the school level as universal programmes. Results suggested that these programmes were associated with decreases in the growth of substance use from 6th through 12th grade at the school level. Interestingly, both programmes were brief (up to seven sessions), and effects were found all the way up to 12th grade. These results are promising because they suggest that caregiver-mediated interventions in adolescence can be delivered as universal school-based curricula to decrease problem behaviour in the school population.

The intervention programmes discussed previously are appropriate for typically developing youth, at-risk youth, and high-risk youth. Unfortunately, adolescent youth who have had serious problems with behaviour since early childhood may need caregiver-mediated support that is more intensive and focused around their specific problems. One intervention programme that is particularly promising for high-risk youth is *Multidimensional Treatment Foster Care* (*MTFC*).[3] *MTFC* is an alternative to institutional treatment for youth with severe emotional and behavioural problems, including those involved with the juvenile justice system. Within the *MTFC* intervention model, youths are placed with foster parents who have received specialized training and who receive a high level of support and supervision from professional staff during the placement. The *MTFC* approach includes the use of highly structured behaviour management programmes as a component of the treatment plan implemented in the foster home. This structure provides the youth with the opportunity to practice the skills that the foster parents are attempting to develop while adjusting the youth's access to risky situations at a rate that matches the youth's progress in treatment.

While the youth is in foster care, his/her biological parents receive intensive behavioural parent training. As the youth shows signs of progress in the foster home, he/she is allowed home visits. The duration of these visits is gradually increased, until the child is ready to return home. An extensive aftercare programme ensures that gains are maintained following reunification.

The *MTFC* approach was evaluated via a randomized control trial and was found to reduce recidivism, especially in comparison to group care.[41–43] Moreover, Chamberlain and Reid reported that youth in the *MTFC* programme had lower incidences of substance use and less contact with delinquent peers than youth in institutional settings.[3] Subsequently, variations of the *MTFC* programme have been developed for a variety of populations, including preschoolers[44–46] and adolescent girls.[47]

Mediators of intervention efficacy

There are many commonalities among the intervention programmes discussed in this chapter. First, each programme is

grounded in developmental research and theory suggesting that families are central to the process of intervening in the mental health issues of children and adolescents. Transitions are critical times for intervention, including transitions to school and puberty.[48] Each programme is strength-based, focusing on increasing protective factors for children and reducing risks in the environment. Lastly, each programme is multifocused, targeting parenting skills and other important factors in a child's life that impact development (e.g. academic achievement).

The literature in this area has suggested some common themes in the mediators that impact the success of interventions for families. First, recruiting parents into treatment can be difficult.[49] Heinrichs *et al.* examined recruitment issues by examining participation in parent training groups in their sample of about 600 families using the *Triple P* parenting programme.[50] They were only able to recruit 31 per cent of low-income families into their parenting groups but were able to retain 77 per cent of families once they began treatment, which included the retention of high-risk families. Similarly, Brody *et al.* found that having more children in the household and youth risk-taking behaviour was related negatively to attendance and engagement in their parenting intervention.[51] Recruitment and attendance issues have been a constant struggle in parenting interventions and have led to the development of brief parenting interventions[52,53] and more tailored, individualized approaches to family treatment.[31] For many families, issues such as childcare, work schedules, and the time commitment required to attend parenting groups prohibit participation. In addition, a parent's interpersonal problems, such as depression and parental resources, can impact participation and outcomes associated with family-based interventions.[54,55]

Second, *culture* plays a key role in the efficacy of family interventions. Parenting skills are inextricably connected to cultural values. For example, racial differences are evident in the literature linking parenting practices to child behaviour problems.[56–58] In previous decades, there was an emphasis on cultural sensitivity within interventions targeting diverse groups. Currently, interventions must be not only culturally sensitive but also adapted to meet the needs of the community in which they are implemented.

Several interventions for families have attempted to address the issue of culture by providing ecologically valid and culturally sensitive interventions to ethnic minority groups in the United States. For example, the *Effective Black Parenting Program* integrated a cognitive behavioural parent training programme with information of relevance to inner-city African-American families, including discussions of traditional discipline (physical punishment) versus modern discipline (internalizing standards of behaviour).[59,60] Components such as helping the child deal with racism at school and positive communication about ethnicity were also included. This programme has been effective at changing parenting behaviour and child behaviour 1 year later.

Another example may be found in the work of Szapocznik and colleagues, who have developed a programme to engage Latino families in treatment that includes less traditional forms of engagement and more of an emphasis on the ecosystems of families and the cultural context. These researchers have hypothesized that resistance to change occurs during the initial stages of therapy and that traditional forms of engagement may not be successful within the Latino population. Instead, engagement includes joining and encouraging the family to participate in home visits and meetings with significant family members. This type of engagement strategy has been effective in retaining families in treatment with the final goal of reducing drug use and other problem behaviours in adolescence.[61–63] Although these programmes serve as examples of specific approaches to working with populations of different cultures, intervention research is just beginning to emphasize these differences and integrate approaches to culturally diverse families into more traditional caregiver training curriculums. More work in this area is clearly necessary, both inside and outside of the United States.

Summary and conclusions

In this chapter, we have attempted to provide a framework for understanding caregiver-mediated, developmentally based interventions for parents and families and to provide examples of interventions that fall within that framework. As is noted previously, the field of caregiver-mediated interventions and the variety of interventions available is a much broader topic than has been addressed here. We hope that the information presented here serves both to inform about the specifics of certain interventions and to organize the study of the larger field.

Further information

http://www.oslc.org/index.html
http://www.mtfc.com/index.html
http://cfc.uoregon.edu/
Dishion, T.J. and Stormshak, E.A. (2007). *Intervening in children's lives: an ecological, family-centered approach to mental health care*. APA Books, Washington, DC.

References

1. Patterson, G.R., Reid, J.B., and Dishion, T.J. (eds.) (1992). *A social learning approach: 4. Antisocial boys*. Castalia, Eugene, OR.
2. Moffit, T.E. (2005). The new look of behavioral genetics in developmental psychopathology: gene–environment interplay in antisocial behaviors. *Psychological Bulletin*, **131**, 533–54.
3. Chamberlain, P. and Reid, J.B. (1998). Comparison of two community alternatives to incarceration for chronic juvenile offenders. *Journal of Consulting and Clinical Psychology*, **66**, 624–33.
4. Olds, D., Henderson, C.R., Tatelbaum, R., *et al.* (1988). Improving the life-course development of socially disadvantaged mothers: a randomized trial of nurse home visitation. *American Journal of Public Health*, **78**, 1436–45.
5. Olds, D. (2003). Reducing program attrition in home visiting: what do we need to know? *Child Abuse & Neglect*, **27**, 359–61.
6. Olds, D., Henderson, C.R., Cole, R., *et al.* (1998). Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15-year follow-up of a randomized controlled trial. *The Journal of the American Medical Association*, **280**, 1238–44.
7. Olds, D.L., Hill, P.L., O'Brien, R., *et al.* (2003). Taking preventive intervention to scale: the nurse-family partnership. *Cognitive and Behavioral Practice*, **10**, 278–90.
8. Olds, D.L. (2002). Prenatal and infancy home visiting by nurses: from randomized trials to community replication. *Prevention Science*, **3**, 153–72.
9. Duggan, A.K., McFarlane, E.C., Windham, A.M., *et al.* (1999). Evaluation of Hawaii's healthy start program. *The Future of Children*, **9**, 66–90.
10. Pettit, G.S., Bates, J.E., and Dodge, K.A. (1997). Supportive parenting, ecological context, and children's adjustment: a seven-year longitudinal study. *Child Development*, **68**, 908–23.

11. Campbell, S.B. and Ewing, L.J. (1990). Follow-up of hard to manage preschoolers: adjustment at age 9 and predictors of continuing symptoms. *Journal of Child Psychology and Psychiatry*, **31**, 871–89.
12. Ruma, P.R., Burke, R.V., and Thompson, R.W. (1996). Group parent training: is it effective for children of all ages? *Behavior Therapy*, **27**, 159–69.
13. Webster-Stratton, C. (1984). Randomized trial of two parent-training programs for families with conduct disordered children. *Journal of Consulting and Clinical Psychology*, **52**, 666–78.
14. Webster-Stratton, C. (1994). Advancing videotape parent training: a comparison study. *Journal of Consulting and Clinical Psychology*, **62**, 583–93.
15. Webster-Stratton, C. (1998). Preventing conduct problems in head start children: strengthening parenting competencies. *Journal of Consulting and Clinical Psychology*, **66**, 715–30.
16. Strayhorn, J.M. and Weidman, C.S. (1991). Follow-up one year after parent-child interaction training: effects on behavior of preschool children. *Journal of the American Academy of Child and Adolescent Psychiatry*, **30**, 138–43.
17. Chaffin, M., Silovsky, J.F., Funderburk, B., *et al.* (2004). Parent-child interaction therapy with physically abusive parents: efficacy for reducing future abuse reports. *Journal of Consulting and Clinical Psychology*, **72**, 500–10.
18. Peters, D.L., Bollin, G.G., and Murphy, R.E. (1991). Head start's influence on parental competence and child competence. In *Advances in reading and language research: a research annual: literacy through family, community, and school interaction*, Vol. 5 (ed. S.B. Silvern), pp. 91–123. Jai Press, Greenwich, CT.
19. Stormshak, E.A. and Bierman, K.L. (1998). Conduct Problems Prevention Research Group. The implications of four developmental patterns of disruptive behavior problems for school adjustment. *Development and Psychopathology*, **10**, 451–68.
20. Cicchetti, D. and Toth, S.L. (1992). The role of developmental theory in prevention and intervention. *Development and Psychopathology*, **4**, 489–93.
21. Conduct Problems Prevention Research Group. (1992). A developmental and clinical model for the prevention of conduct disorders: the FAST Track program. *Development and Psychopathology*, **4**, 509–27.
22. Reid, J.B., Eddy, J.M., Fetrow, R.A., *et al.* (1999). Description and immediate impacts of a preventive intervention for conduct problems. *American Journal of Community Psychology*, **27**, 483–517.
23. Tolan, P., Gorman-Smith, D., and Henry, D. (2004). Supporting families in a high-risk setting: proximal effects of the SAFEChildren preventive intervention. *Journal of Consulting and Clinical Psychology*, **72**, 855–69.
24. Conduct Problems Prevention Research Group. (1999). Initial impact of the FAST Track prevention trial for conduct problems. I. The high-risk sample. *Journal of Consulting and Clinical Psychology*, **67**, 631–47.
25. McMahon, R.J. and Slough, N. (1996). Conduct Problems Prevention Research Group. Family-based intervention in the FAST Track Program. In *Preventing childhood disorders, substance abuse, and delinquency* (eds. R.D. Peters and R.J. McMahon), pp. 90–110. Sage, Thousand Oaks, CA.
26. Conduct Problems Prevention Research Group. (2002). Evaluation of the first 3 years of the FAST Track prevention trial with children at high risk for adolescent conduct problems. *Journal of Abnormal Child Psychology*, **30**, 19–36.
27. Conduct Problems Prevention Research Group. (2004). The effects of the FAST Track program on serious problem outcomes at the end of elementary school. *Journal of Clinical Child and Adolescent Psychology*, **33**, 650–61.
28. Eddy, J.M., Reid, J.B., Stoolmiller, M., *et al.* (2003). Outcomes during middle school for an elementary school-based preventive intervention for conduct problems: follow-up results from a randomized trial. *Behavior Therapy*, **34**, 535–53.
29. Stoolmiller, M., Eddy, J.M., and Reid, J.B. (2000). Detecting and describing preventive intervention effects in a universal school-based randomized trial targeting delinquent and violent behavior. *Journal of Consulting and Clinical Psychology*, **68**, 296–306.
30. Kellam, S.G. and Rebok, G.W. (1992). Building developmental and etiological theory through thorough epidemiologically based preventive intervention trials. In *Preventing antisocial behavior: interventions from birth through adolescence* (eds. J. McCord and R. Tremblay), pp. 162–95. Guilford Press, New York.
31. Dishion, T.J. and Stormshak, E. (eds.) (2007). *Intervening in children's lives: an ecological, family-centered approach to mental health care.* APA Books, Washington, DC.
32. Dishion, T.J., Andrews, D.W., Kavanagh, K., *et al.* (1996). Preventive interventions for high-risk youth: the adolescent transitions program. In *Preventing childhood disorders, substance abuse, and delinquency* (eds. R.D. Peters and R.J. McMahon), pp. 184–214. Sage, Thousand Oaks, CA.
33. Dishion, T.J. and Kavanagh, K. (eds.) (2003). *Intervening in adolescent problem behavior: a family-centered approach*. Guilford Press, New York.
34. Dishion, T.J. and Andrews, D.W. (1995). Preventing escalation in problem behaviors with high-risk young adolescents: immediate and 1-year outcomes. *Journal of Consulting and Clinical Psychology*, **63**, 538–48.
35. Connell, A., Dishion, T.J., Jo, B., *et al.* (in press). An ecological approach to family intervention to reduce adolescent problem behavior: intervention engagement and longitudinal change. *Journal of Consulting and Clinical Psychology*.
36. Dishion, T.J., Nelson, S.E., and Kavanagh, K. (2003). The family check-up for high-risk adolescents: motivating parenting monitoring and reducing problem behavior. *Behavior Therapy*, **34**, 553–71.
37. Stormshak, E.A., Dishion, T.J., Light, J., *et al.* (2005). Implementing family-centered interventions within the public middle school: linking service delivery change to change in problem behavior. *Journal of Abnormal Child Psychology*, **33**, 723–33.
38. Spoth, R., Redmond, C., Shin, C., *et al.* (2004). Brief family intervention effects on adolescent substance initiation: school-level growth curve analyses 6 years following baseline. *Journal of Consulting and Clinical Psychology*, **72**, 535–42.
39. Catalano, R., Kosterman, R., Haggerty, K., *et al.* (1999). A universal intervention for the prevention of substance abuse: preparing for the drug free years. In *NIDA research monograph on drug abuse prevention through family interventions* (eds. R. Ashery, E. Robertson, and K. Kumpfer), pp. 130–59. National Institute on Drug Abuse, Rockville, MD.
40. Molgaard, V. and Spoth, R. (2001). Strengthening families program for young adolescents: overview and outcomes. In *Innovative mental health programs for children: programs that work* (eds. S.I. Pfeiffer and L.A. Reddy), pp. 15–29. Haworth Press, Binghamton, NY.
41. Chamberlain, P. (1990). Comparative evaluation of specialized foster care for seriously delinquent youths: a first step. *Community Alternatives: International Journal of Family Care*, **2**, 21–36.
42. Chamberlain, P. (ed.) (2003). The Oregon multidimensional treatment foster care model: features, outcomes, and progress in dissemination. *Cognitive and Behavioral Practice*, **10**, 303–12.
43. Chamberlain, P. and Friman, P.C. (1997). Residential programs for antisocial children and adolescents. In *Handbook of antisocial behavior* (eds. D.M. Stoff, J. Breiling, and J.D. Maser), pp. 416–24. John Wiley & Sons, New York.
44. Fisher, P.A., Burraston, B., and Pears, K.C. (2005). The early intervention foster care program: permanent placement outcomes from a randomized trial. *Child Maltreatment*, **10**, 61–71.
45. Fisher, P.A., Burraston, B., and Pears, K.C. (2006). Permanency in foster care: conceptual and methodological issues. *Child Maltreatment*, **11**, 92–4.

46. Fisher, P.A., Gunnar, M.R., Dozier, M., *et al.* (2006). Effects of a therapeutic intervention for foster children on behavior problems, caregiver attachment, and stress regulatory neural systems. *Annals of the New York Academy of Sciences*, **1094**, 215–25.
47. Leve, L.D., Chamberlain, P., and Reid, J.B. (2005). Intervention outcomes for girls referred from juvenile justice: effects on delinquency. *Journal of Consulting and Clinical Psychology*, **73**, 1181–5.
48. Coie, J.D., Watt, N.F., West, S.G., *et al.* (1993). The science of prevention: a conceptual framework and some directions for a national research program. *The American Psychologist*, **48**, 1013–22.
49. Gottfredson, D., Kumpfer, K., Polizzi-Fox, D., *et al.* (2006). The strengthening Washington, DC families project: a randomized effectiveness trial of family-based prevention. *Prevention Science*, **7**, 57–74.
50. Heinrichs, N., Bertram, H., Kuschel, A., *et al.* (2005). Parent recruitment and retention in a universal prevention program for child behavior and emotional problems: barriers to research and program participation. *Prevention Science*, **10**, 275–86.
51. Brody, G.H., Murry, V.M., Chen, Y., *et al.* (2006). Effects of family risk factors on dosage and efficacy of a family-centered preventive intervention for rural African-Americans. *Prevention Science*, **7**, 281–91.
52. Lim, M., Stormshak, E.A., and Dishion, T.J. (2005). A one-session intervention for parents of young adolescents: videotape modeling and motivational group discussion. *Journal of Emotional and Behavioral Disorders*, **13**, 194–9.
53. Stormshak, E.A., Kaminski, R., and Goodman, M.R. (2002). Enhancing the parenting skills of head start families during the transition to kindergarten. *Prevention Science*, **3**, 223–34.
54. Smith, K.E., Landry, S.A., and Swank, P.R. (2005). The influence of decreased parental resources on the efficacy of a responsive parenting intervention. *Journal of Consulting and Clinical Psychology*, **73**, 711–20.
55. Webster-Stratton, C. and Hammond, M. (1990). Predictors of treatment outcome in parent training for families with conduct problem children. *Behavior Therapy*, **21**, 319–37.
56. Deater-Deckard, K., Dodge, K.A., Bates, J.E., *et al.* (1996). Physical discipline among African-American and European-American mothers: links to children's externalizing behaviors. *Developmental Psychology*, **32**, 1065–72.
57. Florsheim, P., Tolan, P.H., and Gorman-Smith, D. (1996). Family processes and risk for externalizing behavior problems among African-American and Hispanic boys. *Journal of Consulting and Clinical Psychology*, **64**, 1222–30.
58. Lansford, J.E., Chang, L., Dodge, K.A., *et al.* (2005). Physical discipline and children's adjustment: cultural normativeness as a moderator. *Child Development*, **76**, 1234–46.
59. Alvy, K.T. and Marigna, M. (eds.) (1985). *Effective black parenting program: instructor's manual*. Center for Improvement of Child Caring, Studio City, CA.
60. Myers, H.F., Alvy, K.T., Arrington, A., *et al.* (1992). The impact of a parent training program on inner-city African-American families. *Journal of Community Psychology*, **20**, 132–47.
61. Coatsworth, J.D., Duncan, L., Pantin, H., *et al.* (2006). Patterns of retention in a preventive intervention with Hispanic and African American families. *The Journal of Primary Prevention*, **27**, 171–93.
62. Santisteban, D.A., Szapocznik, J., Perez-Vidal, A., *et al.* (1996). Efficacy of intervention for engaging youth and families into treatment and some variables that may contribute to differential effectiveness. *Journal of Family Psychology*, **10**, 35–44.
63. Szapocznik, J., Kurtines, W., Santisteban, D.A., *et al.* (1997). The evolution of a structural ecosystems theory for working with Hispanic families in culturally pluralistic contexts. In *Psychological interventions and research with Latino populations* (eds. J. Garcia and M.C. Zea). Allyn & Bacon, Boston, MA.

9.5.5 Medication for children and adolescents: current issues

Paramala J. Santosh

Introduction

Problems of mental health and behaviour in children are multidisciplinary in nature and optimal treatment is often multimodal. This article focuses on aspects of psychopharmacology that has special relevance in children and adolescents, especially the recent controversies. In general, this article provides information about classes of medication and not detailed information about specific medicines. Treatment recommendations of the specific disorders have been dealt within the appropriate chapters.

The use of psychotropic medication in children is higher in the United States than in many other countries, and polypharmacy is common. About 1 per cent of overall medical consultations visits by children and adolescents in 2003–2004 in the US resulted in a second-generation antipsychotic (SGA) prescription. The majority of the visits involving antipsychotics were by Caucasian boys aged over nine years, visiting specialists, without private insurance, with a diagnosis of bipolar disorder, psychosis, depression, disruptive disorder, or anxiety.[1]

Pre-school (2 to 4 year olds) psychotropic medication use, between 1995 and 2001 increased across the US for stimulants, antipsychotics, and antidepressants, while the use of anxiolytics, sedatives, hypnotics and anticonvulsants remained stable across these years, suggesting non-psychiatric medical usage.[2] Ethnicity may influence differential prescription rates; for example, as compared to Caucasian youths, African-American youths are less likely to be prescribed psychotropic medications especially methylphenidate.[3]

Information assisting psychopharmacological decision-making

Apart from a thorough diagnostic assessment, a full medical history including present, recent, and past prescribed and over-the-counter medication, response to treatment, and attitude towards interventions play a major role in deciding treatment. History of substance misuse needs to be elicited to ascertain potential medication-misuse liability and because certain medication significantly interact with illicit drugs. A family history of mental illness, suicide, substance abuse, neurological or medical conditions, especially early onset coronary artery disease, hyperlipidemias and diabetes, and specifically the response of family members to psychotropic medication are all important.

Medication as a part of multimodal treatment package

Disorders that have an extended course, where emergence of new problems is common, require continuous, dynamic treatment planning and monitoring to ensure effectiveness of the current treatment. The treatment should stress multi-modal intervention and address co-morbid psychiatric disorders. Treatment plans

should be individualized according to the pattern of target symptoms and strengths identified in the evaluation. A thorough functional analysis of problems or symptoms is central to pharmacological decision-making. Treatment should target situations in which symptoms cause the most impairment. Custom-designed target symptom scales or daily behavioural report cards are useful in monitoring treatment progress.

One way to conceptualize paediatric psychopharmacotherapy is the 'Symptom-based Approach' for core symptoms as follows:

- *Symptoms that require and are likely to respond to medication alone*: inattention, impulsivity, hyperactivity, tics, obsessions, psychotic symptoms, labile mood etc;
- *Symptoms less likely to respond to medication alone, requiring both medication and psychosocial interventions*: aggression, rituals, self-injury etc;
- *Symptoms that are unlikely to respond to medication that need specific remediation or rehabilitation:* skill deficits in academic, social, or sports domain.[4]

Psychosocial interventions may be required to address either primary or secondary relationship problems associated with the core deficits or to deal with co-morbidity.

The 'art' of prescribing medication

Pharmacological interventions do not necessarily work exclusively because of their 'neurochemical' effects. Response to medication also includes the inherent 'placebo response' or 'expectancy effect' that prescribing can induce, as well as the effect of the 'therapeutic concordance' achieved through getting the agreement and acceptance of why the medication is being prescribed and also what is the expected response.

Parents as well as patients respond better when they feel understood, accept why treatment is necessary and are in agreement with the prescriber regarding the need for treatment. Simple strategies help this process: collaborating with parents, patients, school and care providers; giving a clear explanation of diagnoses, of why medication (with or without psychosocial interventions) is necessary; setting realistic expectations (for example, aiming for a 40–60 per cent reduction in anxiety symptoms in a child with multiple co-morbidity); keeping track of the larger systems of care at school and home; providing clear, appropriate information sheets, websites etc to obtain further information if needed; prioritizing and tracking target symptoms; asking them to 'opt in' to treatment after having weighed all the pros and cons, as opposed to them perceiving that they are being 'told' that they 'have to' start medication; using short telephone-based medication monitoring during the stabilization phase in order to pick up emerging side-effects and monitoring dosage accordingly; initiating medication using small doses and titrating it up over a period of 4 to 6 weeks, to identify the **minimum effective dose (MED)**, which is the minimum dose with which 'acceptable' improvement with minimal side-effects is achieved; involving school in monitoring symptoms regularly even if it means maintaining a school-home dairy; and willingness to change treatment if the expected outcome has not been achieved.

Domains that each therapy focusses on should be clearly documented and periodically evaluated. The designation of a case manager is essential for chronically disabled individuals to coordinate the wide range of services necessary for their care and to ensure periodic diagnostic reassessments.

Developmental issues, pharmacokinetic and pharmacodynamic factors affecting pharmacotherapy in children

Generally, drug response may vary with age, weight, sex, disease state, absorption, distribution, metabolism, and excretion. Thus, developmental factors that influence these are important to consider. Although the extent of drug absorption for most medication is similar in children and adults, the rate of absorption may be faster in children and peak levels are reached earlier.[5]

Absorption is also dependent on the form in which it is administered, i.e. liquid versus tablet, and levels peak faster for liquid preparations. Generally speaking, hepatic metabolism is highest during infancy and childhood, 1 to 6 years, approximately twice the adult rate in pre-puberty at six to 10 years, and equivalent to adults by the age of 15.[5] This is clinically important as younger children may require higher mg/kg doses of hepatically metabolized medications than older children or adults.[6]

Adolescence is a period of particularly high ketosteroid levels, which have significant impact on brain neurotransmitter systems. A transient decrease in metabolism for some medication has been reported in a few months before puberty, which is believed to be due to the competition for hepatic enzymes with sex hormones.[7]

Fat distribution varies in children raising during year one then gradually falling until puberty and increasing with obesity. Substantial fat stores slow elimination of highly lipo-soluble drugs from the body (e.g. fluoxetine and pimozide).

Protein binding and volume of distribution affect the pharmacokinetics of medications. These parameters differ in children and have practical clinical implications such as the fraction of the drug that is active and unbound.[8] This is especially a factor when medicating children with eating disorders such as anorexia.

Overall, it is recognized due to the various factors covered in this article, non-pharmacological strategies are more effective in preschoolers. Pharmacotherapy in school-going children has a reasonable risk/benefit ratio and older adolescents behave more like adults. As patients mature, treatment plans often must be adapted to change according to the changing individual, family, and environmental conditions.

Antidepressant efficacy

The poor antidepressant response in childhood depression may have its basis on differences in the hormonal milieu of the brain, and incomplete maturation of the neurotransmitter systems involved in the control of affect, inclusion of adolescents who will over time become bipolar, adolescents with depressive phenocopies, and possible differences in pharmacokinetics and pharmacodynamics. The more rapid hepatic metabolism of imipramine and amitriptyline results in noradrenergically active metabolites, shifting the ratio of the noradrenergic to serotonergic activity of these compounds in children to a ratio higher than that seen in adults. This activity shift is significant because the noradrenergic system does not fully develop both anatomically and functionally until early adulthood.[9,10]

Cardiotoxicity

The maturation of vagal modulation of heart rate increases during the first decade of life, peaks sometime during the second decade,

and declines gradually with age through the sixth decade of life. Sympathetic modulation follows a similar pattern, but rate of maturation of the two branches differ. Furthermore, there is considerable variation between individuals of similar age in autonomic maturation. The relative loss of vagal modulation associated with tricyclic antidepressants may be accentuated in some younger subjects because of these maturational factors, leading to cardiotoxicity.[11]

Stimulants

Stimulants have been used for decades and good research evidence exists for their short-term use in ADHD. More recently, various stimulant delivery systems have been developed (the osmotic controlled-release system (OROS) – Concerta XL®; the wax-matrix-based beaded system – Metadate CD®; the patch release system – Daytrana®; etc) resulting in long-acting preparations which makes it possible to avoid medication needing to be administered in school. This once daily dosing schedule possibly reduces stigmatization and embarrassment. The release systems and preparation of stimulants (immediate release / slow release ratios) allow the tailoring of the long-acting preparations to suit the need of individual children.[12]

Precautions with stimulants

Stimulants are contraindicated in schizophrenia, hyperthyroidism, cardiac arrhythmias, angina pectoris, glaucoma, or a history of hypersensitivity to drug. They can be used with caution in those with hypertension, depression, tics (or family history of Tourette's syndrome), pervasive developmental disorders or severe intellectual disability. Occasionally tics can be made worse with stimulants.

(a) Rebound effects

It consists of increased excitability, activity, talkativeness, irritability and insomnia beginning 4–15 h after a dose. It may be seen as the last dose of the day wears off or for up to several days after sudden withdrawal of high daily doses of stimulants. This may resemble a worsening of the original symptoms and is encountered frequently by clinicians. Management strategies include increased structure after school, addition of a smaller dose of medication in the late afternoon, use of long-acting formulations or the addition of clonidine or guanfacine to the regime.

(b) Seizures

Methylphenidate can be used in the presence of well-controlled epilepsy. If seizures worsen or emerge during treatment, methylphenidate should to be changed to dexamfetamine, which is supposed to increase seizure threshold.

(c) Growth retardation

The MTA study indicates that there is significant growth reduction with stimulant use.[13] It would be advisable not to start stimulants in children who are short and are biologically predisposed to short stature (e.g. short parental stature, growth hormone deficiency etc.).

(d) Cardiac problems

Stimulants may increase pulse rate and rarely can lead to increased blood pressure. Importantly, African-American male adolescents may be at a higher risk from mild chronic elevation in blood pressure on methylphenidate, needing more rigorous blood pressure monitoring for hypertension than their Caucasian counterparts.[14] After several reports of death of patients on Adderall®, there is currently a warning to clinicians to be aware of possible cardiac side effects, especially in the presence of known cardiovascular illness. The reported rates of sudden death on Adderall® do not exceed reported rates of sudden death on other stimulants or the base rate of sudden death per age group in general off medication.

(e) Abuse potential of stimulants

There is little evidence that substance misuse or dependence results from the prescription of stimulants for ADHD. Self initiated increase in dose by emotionally unstable adults with substance use disorders is possible and needs to be suspected in those who repeatedly claim to have lost medication or in parents who repeatedly insist that higher doses are necessary to control symptoms, when the child is functioning well in other settings. In such situations it may be better to use long-acting preparations such as Concerta XL®, as the delivery system makes it difficult to abuse. Other drugs that are useful in this setting are atomoxetine and bupropion.

(f) Psychotic symptoms

Post-marketting surveillance led to enhanced labelling warnings regarding psychosis, mania and hallucinations as adverse events (*http://www.fda.gov/ohrms/dockets/ac/06/minutes/2006-4210m_Minutes%20PAC%20March%2022%202006.pdf*). Many such drug-related psychiatric adverse events may be self-limited and resolve with drug cessation. Stimulants are better avoided in those who have a first degree relative with a psychotic disorder or in children who have psychotic or quasi-psychotic experiences.

Non-stimulants

Atomoxetine is a non-stimulant noradrenaline reuptake inhibitor, which has a high affinity and selectivity for the noradrenaline reuptake site over serotonin and dopamine transporters. In extensive metabolizers (EMs), inhibitors of CYP2D6 e.g. *paroxetine, fluoxetine and guanidine* increase atomoxetine steady-state plasma concentrations. Atomoxetine is long acting and can be used once a day and does not result in rebound symptom worsening and has little potential for abuse. It does not worsen tics, anxiety or low mood and hence may be useful in some children with ADHD and co-morbid disorders.

(a) Precautions with atomoxetine

Atomoxetine is contraindicated in hepatic insufficiency/impairment, glaucoma, uncontrolled seizures or a history of hypersensitivity to drug. They can be used with caution in those with hypertension or with any condition that may predispose to it, tachycardia, cardiovascular problems in patients with congenital or acquired long QT or a family history of QT prolongation or with cerebrovascular disease.[15]

(b) Growth retardation

Acute treatment studies show that atomoxetine-treated patients lose some weight. Michelson *et al.* (2004)[16] reported a change of 2–3 percentiles in mean height, which appear to be similar to effects observed in stimulant-treated patients. Patients treated for extended periods should be monitored regularly.

(c) Seizure liability

A review of risks and benefits of atomoxetine in 2996 led to warnings on the risk of seizures when taking atomoxetine. It is not to be

used in uncontrolled seizures and should be discontinued in those who develop seizures or who experience an increase in frequency of their seizures.

(d) Cardiac problems

Atomoxetine increases noradrenergic tone and produces increased heart rate and small increases in blood pressure, which subsides on discontinuing atomoxetine. QT interval prolongation can occur but ECG monitoring is not necessary unless one suspects cardiac problems.

(e) Suicidal risk

In September 2005, a 'black box' warning was added to the product labelling of atomoxetine as a result of an analyses that showed that suicidal ideation was more frequently observed in clinical trials among children and adolescents treated with atomoxetine (5/1357 [0.37 per cent]) compared to those treated with placebo (0/851[0 per cent]) (*http://www.fda.gov/cder/foi/label/2007/021411s004s012s013s015s021lbl.pdf*). There was one suicide attempt in the atomoxetine treated group. No completed suicides occurred during these trials. There was, however, no evidence of increased suicidal thoughts in adults taking atomoxetine. Prescribers should monitor for signs of depression, suicidal thoughts or suicidal behaviour and refer for appropriate treatment if necessary.

(f) Hepatic dysfunction

Reports indicate that atomoxetine can cause severe liver injury in rare cases. The spontaneous adverse event database search identified two cases that were probably associated with atomoxetine use as a cause or contributor to the event. One spontaneously reported case of liver injury (fulminant hepatitis) appeared probably related to atomoxetine therapy by positive re-challenge in a population exposure of about 2.2 million patients within the 2-year period (2002–2004) after approval which is likely to be an underestimate. Less severe liver dysfunction indicated by abnormal liver enzymes is more common and such reactions may occur several months after therapy is started and atomoxetine should be discontinued in patients with jaundice abnormal liver enzyme levels, which should be done upon the first symptom or sign of liver dysfunction (e.g. pruritus, dark urine, jaundice, right upper quadrant tenderness or unexplained 'flu-like' symptoms).

Antidepressants

The pattern of antidepressant use in children and adolescents has changed significantly over the last couple of decades. Tricyclic antidepressants were predominantly being used for ADHD, enuresis, depression, and anxiety disorders during the eighties and early nineties. Over the last 10–15 years, this has changed to predominantly prescribing the newer antidepressants, especially for anxiety disorders and depression, despite little real evidence for its efficacy. Current data suggests that other than fluoxetine, no other antidepressant has evidence to clearly support its use in depression in children and adolescents.

Tricyclic antidepressants

Tricyclic antidepressants include amitriptyline, desipramine, nortriptyline, imipramine, and clomipramine. Historically, they have been used as a second-line pharmacologic treatment for ADHD (though only as an off-licence drug), following stimulant medications. Their use has declined in recent years due to concerns of cardiac arrhythmias and case reports of sudden death in the paediatric population. Drawbacks include potential cardiotoxicity, especially in pre-pubertal children, the danger of accidental or intentional overdose, troublesome sedation, anticholinergic side-effects, lowering seizure threshold and possibly declining efficacy over time.

New generation antidepressants

Fluoxetine, sertraline, fluvoxamine, citalopram, escitalopram, and paroxetine are all specific serotonin reuptake inhibitors (SSRIs) while venlafaxine is a specific noradrenaline and serotonin reuptake inhibitor (SNRI) and reboxetine a specific noradrenaline reuptake inhibitor (SNI) and mirtazapine. Few trials exist in pre-pubertal children with these drugs; however, the data suggest that the SSRIs have reasonable efficacy in severe anxiety disorders such as OCD (fluoxetine, sertraline) but only fluoxetine is effective in depression. Theoretically, the SNIs may help managing symptoms of ADHD but little research evidence exists.

Precautions with antidepressants

(a) Antidepressant-induced behavioural activation

Frequently children prescribed antidepressants (especially those with developmental disorders or intellectual disability) develop increased motor activity, restlessness, excitability and impulsivity. This usually occurs early in treatment and is often misconstrued as being a manic or hypomanic switch and wrongly treated as if the child had a bipolar disorder. This can be managed by reducing the dose and may need the cover of a benzodiazepine for a few days. This side-effect can be reduced by initiating antidepressants in vulnerable children in small doses (about a fourth of the final dose needed) and gradually increasing the dose over a few weeks.

(b) Antidepressant-related suicidal ideation and behaviour

It was realized two decades ago that imipramine was not effective in pre-pubertal major depression.[17] The Committee for Safety of Medicines (CSM) in December 2003 reviewed all the relevant trials with new generation antidepressants on remission, response to treatment, depression symptom scores, serious adverse events, suicide-related behaviour, and discontinuation of treatment because of adverse events and concluded that the evidence was adequate to establish effectiveness only for fluoxetine and contra-indicated all SSRIs (except fluoxetine) for depressed children (*http://www.mhra.gov.uk/home/groups/plp/documents/drugsafetymessage/con019472.pdf*).[18] This was then followed by the FDA asking for black box warnings on all SSRIs warning about the possibility of suicide-related behaviour as a side-effect in depressed children (*http://www.fda.gov/Cder/drug/antidepressants/SSRIPHA200410.htm*). It is currently advised that children or adolescents being started on or dose being increased of antidepressants should be monitored closely for emergence or worsening of suicidal ideation or behaviour.

Antipsychotics

Recent years have witnessed increased antipsychotic treatment of children despite limited long-term safety data in children. Second generation antipsychotics (SGAs) are the most frequently prescribed ones; for example, risperidone, quetiapine, aripiprazole, olanzapine, ziprazidone, and amisulpiride.

Risperidone is the commonest atypical antipsychotic used in children and adolescents to manage psychoses and disruptive behaviour in autism. It is a potent dopamine D2 receptor blocker (hence reduces positive symptoms but produces hyperprolactinaemia) and 5HT-2A receptor blocker (enhances dopamine release in certain brain regions, thus reducing motor side effects and possibly improving cognitive and affective symptoms). The Research Units for Paediatric Psychopharmacology (RUPP) studies have shown that risperidone is effective in managing disruptive behaviours in autism spectrum disorders.[19] Apart from the side effects of sedation and weight gain, hyperprolactinaemia is common and a few develop extrapyramidal side effects such as tardive dyskinesia.

Aripiprazole is dopamine partial agonist or dopamine stabilizer, used in managing schizophrenia and bipolar disorder. The partial antagonism of D2 receptors reduces dopamine output when dopamine concentrations are high, thus improving positive symptoms and mediating antipsychotic actions. Blockade of 5HT-2A receptors may cause enhancement of dopamine release in certain brain regions, thus reducing motor side effects and possibly improving cognitive and affective symptoms. Actions at D3 receptors and partial agonism of 5HT-1A receptors could also theoretically contribute to aripiprazole's efficacy.[20] Even though symptoms may improve in the first week, it is recommended to wait for four to six weeks to determine efficacy due to the pharmacokinetics of the drug. The mean elimination half-life is 75 h of aripiprazole and 94 h for the major metabolite dihydro-aripiprazole and is primarily metabolized by CYP450 2D6 and CYP450 3A4. ketoconazole, fluvoxamine, and fluoxetine all increase plasma levels of aripiprazole, while Carbamazepine decreases plasma levels of aripiprazole. Early experience of aripiprazole suggests that it produces less weight gain, diabetes, or hyperlipidaemia, compared to the other SGAs. It is however less sedating and rather activating. Little published evidence exists currently on its use in managing non-psychotic disruptive behaviour in developmental disorders but clinical experience suggests that very small doses (2 to 5 mg per day) are sufficient.

Quetiapine is an effective SGA with moderate effect on weight, but usually needs to be taken at least twice daily because of relative weak receptor binding. *Ziprasidone* is being increasingly used and is the only SGA that is weight neutral, but has greater impact on cardiac rhythm and QTc interval. *Clozapine* is used in those with resistant psychoses or those with tardive dyskinesia, but can lead to neutropaenia, sialohorrea, and significant weight gain. *Olanzapine* is used less in children and adolescents because of the propensity to weight gain and metabolic syndrome. Evidence from adults suggests that Clozapine, Olanzapine, and low-potency conventional antipsychotics such as chlorpromazine are associated with increased risk of insulin resistance, hyperglycaemia, and type 2 diabetes mellitus.[21,22]

Precautions with antipsychotics

(a) Movement disorders

Tardive dyskinesia and extrapyramidal side-effects are more common with conventional antipsychotics. Ethnicity may be a risk factor for dyskinesia in children as African-American children appear to be more prone to tardive dyskinesia when compared to European-American children.[23] Aripiprazole or clozapine are useful in those who require antipsychotics but have developed tardive dyskinesia.

(b) Weight gain and metabolic dysfunction

Weight gain is a serious side effect of SGAs and potential consequences of obesity include non-compliance with medication and significant morbidity and mortality.

(c) Risk for SGA-induced metabolic dysfucntion

High risk – clozapine, olanzapine

Moderate risk – risperidone, quetiapine, sertindole

Low risk – amisulpiride, aripiprazole, ziprazidone

(d) Baseline information before starting an SGA

Weigh the children and adolescents and track the BMI during treatment; get baseline personal and family history of obesity, dyslipidaemia, hypertension, and cardiovascular disease; children of parents with total cholesterol >24 mmol/l (parents with high BMI), and history of overt heart disease in parent or grandparent at age 55 or younger should be considered to be at high risk; get waist circumference at umbilicus, blood pressure, fasting plasma glucose, and fasting lipid profile.

(e) Monitoring after starting an SGA

BMI monthly for three months then quarterly with blood pressure, fasting plasma glucose, fasting lipids within three months and then annually but earlier and more frequently for patients with diabetes or who have gained greater than 5 per cent of initial weight; treat or refer for treatment and consider switching to another atypical antipsychotic for patients who become overweight, obese, prediabetic, diabetic, hypertensive, or dyslipidaemic while receiving an atypical antipsychotic. Elevated fasting plasma triglyceride or increased insulin levels may be important signals of potential insulin resistance. Increased low density lipoprotein (LDL) cholesterol and decreased high density lipoprotein (HDL) cholesterol is associated with increased adiposity, especially visceral adiposity. Even in patients without known diabetes, one should be vigilant for the rare but life threatening onset of diabetic ketoacidosis, which always requires immediate treatment by monitoring for the rapid onset of polyuria, polydipsia, weight loss, nausea, vomiting, dehydration, rapid respiration, weakness, clouding of sensorium, and even coma.

Treatment of SGA-induced metabolic dysfunction

Careful selection of treatments taking the metabolic-risk profile into account, preventative healthy lifestyle counselling, and regular monitoring of body composition and metabolic variables need to become clinical routine. Total caloric intake is more important than the content of the diet and motivational interviewing and cognitive behavioural techniques can be used to address unhealthy diet, physical inactivity and smoking. Clinically an effective method is to insist that the family (as opposed to only the child) go onto a healthy diet and activity schedule. In fact, informing the parents that they will also be weighed along with the child, helps address this issue effectively. Although preliminary, co-treatment with an SGA plus a mood stabilizer seems to be associated with a greater risk for age-inappropriate weight gain than treatment with one or even two mood stabilizers. A pilot study has shown Metformin therapy is safe and effective in decreasing weight gain, insulin sensitivity, and abnormal glucose metabolism in children aged 10 to 17 whose weight had increased by more than 10 per cent during less than one year of olanzapine, risperidone or quetiapine therapy.[24]

Developmentally appropriate monitoring guidelines[25] need to be implemented in routine clinical practice and the effectiveness of suggested behavioural and pharmacological interventions should be evaluated.

Mood-stabilizers

Mood stabilizers are usually antiepileptics (except SGAs and lithium) and are used to treat epilepsy or bipolar disorders and occasionally mood lability. Commonly used mood stabilizers include carbamazepine, sodium valproate, lamotrigine, and lithium carbonate. Lithium use warrants regular blood level monitoring, which is often a problem in children. *Sodium valproate or valproic acid* is the most used mood stabilizer and has a reasonable evidence base for its use in bipolar disorder. It however should be used very cautiously, if at all, in girls of child-bearing age due to its teratogenic effects, as well as possible side effect of polycystic ovarian disease. Lamotrigine is a more recent addition and hence being discussed.

Lamotrigine works by inhibiting effects on sodium channels and stabilizes neuronal membranes and thereby moderates the release of excitatory amino acids such as glutamate and aspartate. Lamotrigine is especially useful when significant depressive symptoms exist in bipolar disorder.[26] It has a moderately long half-life, especially as monotherapy and can be given once or twice a day. When used with hepatic enzyme inducing medications such as phenytoin and carbamazepine, the half-life is reduced to 12 h. Valproic acid markedly increases the half-life of lamotrigine to 48 to 72 hand the concomitant use of these medications increases the likelihood of developing severe drug rashes including Steven–Johnson syndrome. Lamotrigine is to be started at very low doses (as low as 5 mg/day) and increased slowly over a couple of months. In general, the initiation of this drug requires great patience. Evidence for anti-manic effects is limited.

Rapid tranquilization when managing severe maladaptive aggression in children and adolescents

Managing acute severe maladaptive aggression requires a quick functional analysis to identify the underlying cause – mood lability, mania, impulsivity, sensory trigger such as hypersensitivity to sound in autism, psychotic experience, autonomic arousal, anxiety or depression etc. The management involves:

- 'Talking down' as the first strategy.
- The patient should be offered the choice of having an oral medication before forcing parental medication.
- Oral medication – risperidone 0.5–1mg, or olanzapine 1–2.5 mg or lorazepam 0.5–1 mg.
- If the above fails, repeat giving either the antipsychotic or the benzodiazepine again in 30 min.
- If this fails, combine the antipsychotic and benzodiazepine after 30 min.
- If oral treatment is not accepted or has not helped, IV haloperidol 1–2 mg or IV diazepam 100–200 mcg/kg or IM haloperidol 1–2mg or IM lorazepam 50–100 mcg/kg can be used.
- The IV can be repeated after 10 min if not sufficient or the IM can be repeated after 30 min. Supportive staff and equipment are necessary as side-effects such as cardio-respiratory arrest can occur. *Maximum safe doses based on age and weight of the child or adolescent should not be exceeded over a 24-hour period.*
- Physical restraint and seclusion in safe environment may also be necessary.
- If **aggression is occurring in the context of a medical condition** (for example, cardiac delirium leading to pulling out catheters etc), midazolam 500 mcg/kg (max 15 mgs) can be used in the context of the medical setting.

Pharmacological preparation for medical or surgical interventions in children and adolescents with intellectual disability or severe behavioural disorders

Children with intellectual disability or mental health problems should receive the same level of medical care if they are physically unwell, as normal children. As they may not be co-operative, often, physical investigations (for example, MRI or dental X-rays) are avoided or postponed, often leading to late interventions and poorer medical outcomes. If the procedure is a minor one, lasting a few minutes, oral midazolam can be used; if it requires longer or for more serious procedures, an anaesthetist could administer oral ketamine or another general anaesthetic. Simple strategies such as anti-emesis prophylaxis, removal of cannulas before they recover consciousness etc., help minimize difficulties.

Conclusions

Paediatric pharmacovigilance for psychotropic agents are essential and more studies on efficacy in this population is necessary. Studies are urgently required in children and adolescents investigating the metabolic effects of individual medication, especially the SGAs and mood stabilizers, using sex and age adjusted measures of weight and body composition, including fasting blood work and blood pressure measurements, protective factors and long-term follow-up. Until such detailed data become available, it is safe to assume that paediatric populations are at least as or more vulnerable to adverse effects compared to adults. True long-term data is not available currently on most psychotropic agents. Evidence on treatment impact on co-morbid disorders, cost-effectiveness and impact on Quality of Life are sparse and urgently need to be addressed.

Further information

Stahl, S. M. (ed.) (2008). *Essential Psychopharmacology* (3rd edn). Cambridge University Press, New York.

Osterheld J.R. and Shader, R.I. (1998). Cytochromes: A primer for child and adolescent psychiatrists. *Journal of the American Academy of Child and Adolescent Psychiatry*, **37**(4), 447–50.

Martin, A., Scahill, L., Charney, D. S., *et al.* (2003). *Paediatric Psychopharmacology—Principles and Update.* Oxford University Press, New York.

Prescribing unlicensed drugs or using drugs for unlicensed indication (1992). *Drugs and Therapeutic Bulletin*, **30**(25), 97–100.

References

1. Aparasu, R.R. and Bhatara, V. (2007). Patterns and determinants of antipsychotic prescribing in children and adolescents, (2003–2004). *Current Medical Research and Opinion*, **23**(1), 49–56.
2. Zito, J.M., Safer, D.J., Valluri, S., *et al.* (2007). Psychotherapeutic medication prevalence in Medicaid-insured preschoolers. *Journal of Child and Adolescent Psychopharmacology*, **17**(2), 195–203.
3. Zito, M., Safer, D.J., DosReis, S., *et al.* (1998). Racial disparity in psychotropic medications prescribed for youths with Medicaid insurance in Maryland. *Journal of the American Academy of Child and Adolescent Psychiatry*, **37**, 179–84.
4. Santosh, P.J. and Baird, G. (1999). Psychopharmacotherapy in children and adults with intellectual disability. *Lancet*, **354**(9174), 233–42.
5. Bourin, M. and Couetoux du Tertre, A. (1992). Pharmacokinetics of psychotropic drugs in children. *Clinical Neuropharmacology*, **159**(1), 224A–5A.
6. Wilens, T.E., Biederman, J., Baldessarini, R.J., *et al.* (1992). Developmental changes in serum concentrations of desipramine and 2-hydroxydesipramine during treatment with desipramine. *Journal of the American Academy of Child and Adolescent Psychiatry*, **31**, 691–98.
7. Hughes, C. and Preskorn, S.H. (1989). Depressive syndromes in children and adolescents: diagnosis and treatment. *Annals of Clinical Psychiatry*, **1**, 109–18.
8. Paxton, J. and Dragunow, M. (1993). Pharmacology. In *Practitioner's Guide to Psychoactive Drugs for Children and Adolescents*, (eds. J. Werry and M. Aman), pp. 34–46. Plenum, New York.
9. Goldnan-Rakic, Patricia, S. and Brown, R.M. (1982). Postnatal development of monoamine content and synthesis in the cerebral cortex of rhesus monkeys. *Developmental Brain Research*, **4**, 339–49.
10. Thorn, G.W. and Lauler, D.P. (1972). Clinical therapeutics of adrenal disorders. *American Journal of Medicine*, **53**, 673–84.
11. Mezzacappa, E., Steingard, R., Kindlon, D., *et al.* (1998). Tricyclic antidepressants and cardiac autonomic control in children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, **37**, 52–9.
12. Banaschewski, T., Coghill, D., Santosh, P., *et al.* (2006). Long-acting medications for the hyperkinetic disorders. A systematic review and European treatment guideline. *European Child & Adolescent Psychiatry*, **15**(8), 476–95.
13. Swanson, J.M., Elliott, G.R., Green hill, L.L. *et al.* (2007). Effects of stimulant indication on growth rates across 3years in the MTA follow-up. *J Am Acand Child Adolesce Psychiatry*, **46**, 1015–27.
14. Brown, R.T. and Sexson, S.B. (1989). A controlled trial of methyjphenidate in black adolescents. *Clinical Pediatrics*, (Phila) **27**, 74–81.
15. Unni, J.C. (2006). Atomoxetine. *Indian Pediatrics*, **43**(7), 603–6.
16. Michelson, D., Buitelaar, J.K., Danckaerts, M., *et al.* (2004). Relapse prevention in paediatric patients with ADHD treated with atomoxetine: a randomized, double-blind, placebo-controlled study. *Journal of the American Academy of Child and Adolescent Psychiatry*, **43**(7), 896–904.
17. Puig-Antich, J., Perel, J.M., Lupatkin, W., *et al.* (1987). Imipramine in prepubertal major depressive disorders. *Archives of General Psychiatry*, **44**(1), 81–9.
18. Whittington, C.J., Kendall, T., Fonagy, P. *et al.* (2004). Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. *Lancet*, **363**(9418), 1341–5.
19. McCracken, J.T., McGough, J., Shah, B., *et al.* (2002). Risperidone in children with autism and serious behavioural problems. *The New England Journal of Medicine*, **347**(5), 314–21.
20. Davies, M.A., Sheffler, D.J. and Roth, B.L. (2004). Aripiprazole: a novel atypical antipsychotic drug with a uniquely robust pharmacology. *CNS Drug Reviews*, **10**(4), 317–36.
21. American Diabetes Association. (2004). *Consensus development conference on antipsychotic drugs and obesity and diabetes. Diabetes Care*, **27**, 596–601.
22. Casey, D.E., Haupt, D.W., Newcomer, J.W., *et al.* (2004). Antipsychotic-induced weight gain and metabolic abnormalities: implications for increased mortality in patients with schizophrenia. *Journal of Clinical Psychiatry*, **65**(7), 4–18.
23. Wonodi, I., Reeves, G., Carmichael, D., *et al.* (2007): Tardive dyskinesia in children treated with atypical antipsychotic medications. *Movement Disorders*, **22**(12), 1777–82.
24. Klein, D.J., Cottingham, E.M., Sorter, M., *et al.* (2006): A randomized, double-blind, placebo-controlled trial of metformin treatment of weight gain associated with initiation of atypical antipsychotic therapy in children and adolescents. *American Journal of Psychiatry*, **163**(12), 2072–9.
25. Correll, C.U. and Carlson, H.E. (2006). Endocrine and metabolic adverse effects of psychotropic medications in children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, **45**(7), 771–91.
26. Goldsmith, D.R., Wagstaff, A.J., Ibbotson, T., *et al.* (2004): Spotlight on lamotrigine in bipolar disorder. *CNS Drugs*, **18**(1), 63–7.

9.5.6 Residential care for social reasons

Leslie Hicks and Ian Sinclair

Introduction

Residential care for the young is an elusive object of study. Provided in the past by establishments as diverse as workhouses, orphanages, and reformatories, it has no clear definition marking its boundaries with foster care or boarding education; at the same time it variously aims to shelter, classify, control, and reform and it has no agreed theory or body of values. The need for residential care, and the difficulties of providing it, vary with time and place; the issues it raises are quite different in Romania than they are in California, or were in Victorian England.

Given this diversity, any discussion of residential care needs to outline the context within which it was written. In the case of this chapter the context is provided by current British social policy. Although the focus is on residential care provided to young people by Children's Services in England for social reasons, the conclusions drawn are applicable to the rest of the United Kingdom. The issues raised by this provision have similarities in other parts of the developed world, in virtually all of which the use of residential care is declining.[1,2] This chapter is written against the background of this decline. Its aims are as follows:

- to describe the current characteristics of residential child care in England, and by extension in Great Britain
- to outline the problems that have led to its numerical decline
- to identify practices that should overcome or reduce these problems
- to discuss the role that residential care might play in future.

The characteristics of residential care

In 1979, official statistics for England (the figures for Wales, Scotland, and Northern Ireland are published separately) showed

that there were approximately 95 000 children in substitute care (mainly foster care and residential care). By 2006 this figure had dropped to 60 300. While for much of this time numbers in foster care were fairly constant, the numbers in residential care fell from 35 000 to 6 600.[3] The great bulk of residential care is directly provided by local authority social services departments (more recently Children's Services), although provision by the private sector is increasing.

These figures give, in some respects, a misleading impression of the numerical importance of homes. The turnover in them is quite rapid—roughly 60 per cent leave a home within 2 months of arrival and just under half the placements result from movements within the care system rather than the breakdown of community care. On average homes still see around three times the number of residents in a year than they accommodate at any one time.[4] A recent study found that residential care accounted for about 28 per cent of the time that children aged 12 or over spent in the care system.[5]

The basic characteristics of the local authority homes are well known. They are 'open' in the sense that the residents are expected to go out to school or work and are not restricted at other times. They are typically small, 50 per cent of them have five or fewer places and the average capacity is around six.[6] The buildings are not markedly institutional, although identifiable to the practised eye, are usually located near to where the residents are likely to live, and aim to take local young people. Staff are non-resident. Many care staff do not possess a recognized appropriate qualification and although government targets in the form of National Minimum Standards[7] aim to rectify this situation, progress towards these remains relatively slow. Homes have more staff than residents and research has shown substantial variation in the number of care hours delivered for each resident young person each week (between 37 and 254 h).[6]

Children enter the care system in Britain to be 'looked after' because of a temporary emergency (for instance hospital admission of the carer) or because of abuse, extremely problematic behaviour on the child's part, or a breakdown of family relationships. The main characteristics that distinguish residents from other users of substitute care are as follows:

- Age: 81 per cent of those who start their period of being 'looked after' in local authority homes were aged 10–15 over and about 15 per cent are aged 16 or 17. By contrast, 62 per cent of those entering foster care are under 10 years of age and only 2 per cent are aged 16 or 17.[8]
- Sex: the proportions of males and females in foster care are roughly equal, but nearly 60 per cent of those in residential care are male.
- Geographical location: some local authorities use residential care much more than others. In 2006 the proportions of 'looked after' adolescents in residential care varied from just under 3 per cent in one local authority to slightly over 24 per cent in another.[3]
- Behaviour: on average children in residential care exhibit more 'challenging' behaviour than do foster children, as reflected for example in educational performance, measures of psychiatric ill health, delinquency, and the likelihood of being imprisoned as an adult.[9,10]

One study[4] carried out as part of a major tranche of research on residential child care in England showed that residents had relatively high levels of school exclusion, truancy, delinquency, violence towards adults and children, running away, risk-taking sexual behaviour, self-harm, and suicide attempts. Around two-thirds of residents entered 'care' for the first time as teenagers, generally because family relationships had broken down, and a further one-fifth entered because of abuse. They were unwilling to be fostered or seen as too disturbed for this. The great majority had previous placements in foster care, residential care, or with relatives.

The role of the homes was to return some as soon as possible to their families, attempt to improve the behaviour of others, or prepare them for independent living, and to keep a minority (around one-fifth) for the foreseeable future. Although there was some specialization, most homes attempted to fulfil all these roles and took all types of resident.

Problems of residential care

Residential care lacks the moral basis that it formerly had; the disciplines and virtues required for successful group life are no longer seen as imperatives. The Children Act 1989 ensured that, except in rare circumstances, young people could no longer be 'looked after' simply on the grounds of delinquency when their own welfare and those of others are not at risk.

Theoretical uncertainty accompanies these moral shifts. The best-known texts on residential care are now around 25 years old and are dubiously relevant to the current situation where staff no longer live on the premises, and where there have been major changes in staffing, turnover of residents, clientele, and the size and purpose of homes. In any event, there was never a consensus about which theory should underpin treatment or training. Clear evidence for the efficacy of any approach has been lacking, whereas the evidence for the harmful effects of bad residential care has been clear—and frequently repeated on social work courses. Young people reach residential care as the culmination of a process that marks, and possibly even exacerbates, their social exclusion. Typically, they are disturbed, poorly supported by their families, and lack educational qualifications, a combination which makes it difficult for them to compete subsequently in the job market or to lead happy lives. Descriptions of residential care do not suggest that it mounts the determined attack on these problems that might have some chance of success.

There are particular problems for those young people who leave residential care to live independently. Residential homes have difficulty in retaining young people until such time as they are 'properly launched'. To do so would create problems related to cost and to creating a regime suitable at the same time for younger and older teenagers. This means that young people leave care when they are still vulnerable, to cope with lives that are lonely and difficult, at an age much younger than their better qualified and supported contemporaries leave home. Their transitions to the adult roles of getting an income, maintaining a home, and living with a partner are made earlier than those of others and compressed into a shorter period.[11] Unsurprisingly, they have a much higher chance than their contemporaries of becoming lone parents, unemployed, imprisoned, or homeless.[12,13]

In addition to these problems, there are often pragmatic difficulties.

- Residential care is very expensive; estimated costs are around £78 000 per place per year in local authority homes and around £87 000 in the non-statutory sector.[6] The system costs

considerably more than foster care or preventive work.[14] Around 14 per cent of those who enter stay for prolonged periods and take up half the beds.

- Residential homes are prone to scandals involving sexual abuse, outbreaks of disorder, and suicidal behaviour.[4]
- Delinquency and running away are widespread within residential homes.[4]
- There is widespread bullying, sexual harassment, and personal unhappiness.[4]

Official reports[15] have emphasized the need to recruit a better trained workforce, and there are considerable efforts being made to raise standards in this respect.[16] However, the proportions of qualified staff remain low and studies that have looked for a relationship between the proportion of qualified staff and a measure of performance found no association.[4,6,13]

Contrary to common belief high staff ratios do not in themselves lead to better control in homes or better staff morale or better outcomes, although they do increase costs.[4,6,13]

Improving the quality of residential care

If residential care is to overcome these problems, it will need to reduce the incidence of difficulties in the home and increase its influence after residents have left.

In terms of immediate impact, residential homes and schools vary widely in the morale of the staff, the incidence of delinquent behaviour and running away, the proportion of residents who avoid going to school, the relationships between staff and residents, and the degree to which the residents report that they are bullied, offered drugs, or feel part of a friendly establishment to which they are committed. These characteristics are far from fully accounted for by intake, and establishments which do 'well' in terms of one of them tend to do 'well' in terms of the others.[4,6,13]

These studies suggest the following conditions are required for managing homes that are successful in the short-term.

- The residential units are either small in size or consist of a large unit divided into small subunits. Such small establishments seem better able to combat the influence of delinquent residents.
- The units have clear aims, with which the manager of the home is in agreement.
- The manager of the home has a clear philosophy on how the young people can be helped, and the staff are in agreement with this outlook.
- The residential unit is not based close to the residents' homes. An emphasis on local placement makes it harder to combat the influence of local negative cultures or to maintain a clear focus.

The most recent of these studies[6] was able to demonstrate the link between these immediate outcomes for young people and the practice of managers of homes. Young people experienced better outcomes where managers were accepted as embodying good practice from within a clear ethos, had positive strategies for working with the behaviour and education of young people, and importantly, could enable staff to reflect and deploy these. Without this bedrock, the systems, procedures, and management targets advocated in reports on the residential system will fail. The basic principle is to establish a set of shared expectations and approaches and this is more easily done in a small well-led establishment with a cohesive staff group.

Unfortunately, the ability of homes to affect the behaviour of residents in the short-term implies that the environments to which residents go next are equally powerful. This has been shown to be the case and creates problems for homes in achieving long-term change. (Evidence on the short- and long-term impact of residential care is discussed in more detail elsewhere.[4,6,13]) Overcoming these problems requires the following:

- Reduction of delinquent and problematic behaviour in the home—as noted above these are influenced by the residential environment, and there is evidence that they lead to future delinquency.
- Encouraging educational achievement and those skills required for 'success' in subsequent life and that are relevant to the resident's subsequent environment—and are seen by the residents as being so.
- Working to improve family relationships so that the resident either goes back to an improved family situation or can turn for support to his or her family even though he or she is not living with them.
- Providing continuing back-up from the residential home or an aftercare scheme and practical support (e.g. relating to accommodation).

Success is probably more likely where the home operates on a number of fronts simultaneously: seeking to improve skills, learning and educational achievement, and the way the residents see themselves and also the environment to which the residents return.

Future of residential child care

The case for residential care is essentially three-fold. First, it is able to tolerate behaviour that leads to foster-care placements, the main alternative at the moment, breaking down. Second, a number of young people choose it in preference to foster care. Third, it offers a major resource for adolescents, in preparation for adulthood, in providing supported accommodation, and for those who make a later entry to the care system, such as unaccompanied asylum seeking young people. Its future will depend, in part, on the degree to which other provisions can be developed that can match these advantages.

It seems likely that more alternative provision will be developed. Some intensive American fostering schemes do appear able to contain young people who are as challenging as those in British children's homes, and these schemes are being adapted to conditions in the United Kingdom. Some forms of fostering with outreach support from respite residential units are in their early stages of development. Remand fostering, 'crash pads', and occasional beds in family centres might provide for some young people currently accommodated briefly in children's homes. Supported lodgings could provide for some, particularly older adolescents who avoid foster care because they feel they cannot live in a family, or that it is time that they started to move out on their own. Boarding schools might provide for younger children who do not want to enter foster care because they feel it invites disloyalty to their family, or who enjoy the company of other teenagers.

The comparative advantages of these differing kinds of provision have not been evaluated and it is important that they should be.

The growth of alternative provision will allow the continuing reduction of the more traditional residential sector. It is also likely that the sector itself will change, with the current all-purpose children's home being replaced by more specialist provision that can focus upon clear goals. Research that could determine the shape of such provision is lacking and should be carried out. Possible models of provision include the following:

- Secure (namely, closed) provision—public opinion will not allow young people who have committed very serious crimes to remain in open conditions.
- Brief- to medium-term provision designed to allow time for young people to consider their situation and move on to new placements in a planned way—there is evidence that fostering placements made from residential care are less likely to disrupt, although it is not yet clear why this is.
- Medium-term provision designed to provide treatment—the case for such provision is that treatment involving a group can be more effective. Evidence for this is lacking, although the variations in the behaviour of residents in different establishments is evidence of the power of the group.

In addition to these kinds of residential care, there is a case for some long-stay accommodation in which groups of residents live together. Such provision would be similar to a large foster home or the former family group homes. There seems no reason why it should not work well. It would, however, need to be less generously staffed than current residential homes, otherwise there would be pressure to move residents on for purely economic reasons.

Conclusions

Residential care is very different in different parts of the world. In the United Kingdom it is provided by small community-based homes, who work with an extremely complex and challenging clientele, and face problems associated with high costs, scandal, and frequently a lack of a clear rationale. Despite these difficulties there are wide variations in the performance of such homes, particularly in terms of their immediate impact. Outcomes after the residents have left are harder to influence. However, an approach which combines attention to the residents' education, behavioural problems, and family environment seems most likely to be effective. In the longer-term, the residential sector is likely to continue to shrink and the part that remains may well need to become more specialized. Research that could guide such developments is presently lacking and should be undertaken.

Further information

Department of Health. (1998). *Children looked after away from home: messages from research*. Wiley, Chichester.

Sinclair, I. (2006). Residential care in the UK. In *Enhancing the well-being of children and families through effective interventions: international evidence for practice* (eds. C. McAuley, P. Pecora, and W. Rose). Jessica Kingsley, London.

References

1. Gottesman, M. (ed.) (1991). *Residential child care: an international reader*. Whiting and Birch in association with FICE, London.
2. Colton, M. and Hellinckx, W. (eds.) (1993). *Child care in the EC*. Arena, Aldershot.
3. Department for Education and Skills. (2007). *Children looked after by local authorities, year ending 31 March 2006* (Internet only) Department for Education and Skills. http://www.dfes.gov.uk/rsgateway/DB/VOL/v000721/index.shtml
4. Sinclair, I. and Gibbs, I. (1998). *Children's homes: a study in diversity*. Wiley, Chichester.
5. Sinclair, I., Baker, C., Lee, J., *et al.* (2007). *The pursuit of permanence: a study of the English care system*. Jessica Kingsley, London.
6. Hicks, L., Gibbs, I., Byford, S., *et al.* (2007). *Managing children's homes: developing effective leadership in small organisations*. Jessica Kingsley, London.
7. Department of Health. (2002). *Children's homes national minimum standards and children's homes regulations*. The Stationery Office, London.
8. Department for Education and Skills. (2006). *Statistics of education: children looked after by local authorities, year ending 31 March 2005, Vol. 1: National Tables*. HMSO, Norwich.
9. Department for Education and Skills. (2006). *Care matters: transforming the lives of children and young people in care green paper*. The Stationery Office, Norwich.
10. Koprowska, J. and Stein, M. (2000). The mental health of looked after young people. In *Young people and mental health* (eds. P. Aggleton, J. Hurry, and I. Warwick). Wiley, Chichester.
11. Biehal, J., Clayden, J., Stein, M., *et al.* (1995). *Moving on: young people and leaving care schemes*. National Children's Bureau, London.
12. Anderson, I., Kemp, P.A., and Quilgars, D. (1993). *Single homeless people*. HMSO, London.
13. Berridge, D. and Brodie, I. (1998). *Children's homes revisited*. Jessica Kingsley, London.
14. Carr-Hill, R.A., Dixon, P., Mannion, R., *et al.* (1997). *A model of the determinants of children's personal social services*. Centre for Health Economics, University of York, York.
15. Utting, W. (1991). *Children in the public care*. HMSO, London.
16. Training Organisation for the Personal Social Services. (2003). *National occupational standards for registered managers (child care)*. TOPSS, England.

9.5.7 Organization of services for children and adolescents with mental health problems

Miranda Wolpert

Introduction

This chapter aims to guide the thinking of practitioners who might be involved in developing services to meet the needs of children and young people with mental health difficulties. Anyone involved in this challenging but vital endeavour will need to address the following questions:

- Who should the service be for?
- What sort of interventions should be provided?
- How should the service be structured?

- Who should the staff be?
- How can the service be made most accessible?
- How can service quality be ensured?

This chapter will look at each of these issues in turn to explore how each might best be approached.

Who should the service be for?

Ever expanding conceptualization of mental health needs has meant that at least four groups of children are routinely referred to in the discussions about service development in this area:

1 Children and young people in difficult (and often terrible) circumstances
2 Children at risk of developing diagnosable mental health problems
3 Children with diagnosable mental health problems
4 Children with levels of impairment due to mental health issues that make it difficult for them to function within their community/culture

Lack of clarity when discussing the needs of the different groups can confuse service planning. In particular, it can result in the range of agencies that are all increasingly involved in collaborating in planning provision, talking at cross purposes. When resources are limited (as they generally are for these populations of children, even in the most economically developed countries of the world) it is likely that choices will need to be made about prioritizing between and within these groups. It is therefore vital to be clear at the outset which groups are seen as the priority for a given community and to achieve multi-agency agreement on this.

The needs of each of these groups in relation to service provision in this area is considered in turn below.

Children in difficult circumstances

Definitions of 'difficult circumstances' vary with national context. The estimated 14 million AIDS orphans (concentrated largely in Africa) would fall into this category as would the 12 million children in the United States living below the poverty line along with all those children who are in contexts of war, famine, or abuse.[1] This group is likely to include those children with high and complex mental health needs who are the hardest to reach and who are a policy priority in many areas. However it may also include children who do not have specific mental health needs. Services need to ensure they are accessible to these groups where they do have mental health issues but specialist mental health services are unlikely to be the main provider of care for the majority of children in such circumstances.

Children 'at risk'

Risk factors for developing diagnosable mental health problems include some aspect of difficult circumstances (such as violent environments, lack of warm family environments) but risk is also heightened by other individual and interpersonal factors such as: brain injury, low birth weight, poor parental mental health, low IQ, irritable temperament, family dysfunction, and the lack of a key supportive relationship with an adult. For children identified as 'at risk' the focus for mental health provision is likely to be on how to enhance resilience. There may be key opportunities for intervention, such as when the child is born or at key transition points (such as starting school, changing school, leaving school, etc). There is an argument for targeting particular groups such as children of parents with mental health problems. However, evidence for the effectiveness of health promotion and prevention initiatives is still limited and there is some evidence that the promotion of resilience may best be achieved by agencies other than child mental health specialists, such as welfare sectors creating greater neighbourhood cohesion, perinatal services reducing risk of low birth weight and educational services implementing appropriate programmes to promote emotional well-being and support children in times of stress. It is likely that services should only put major resources into targeting 'at risk' children if they have sufficient resources to meet the needs of those children with existing problems.

Children with diagnosable mental health problems

Epidemiological data from Europe and the United States suggests that around 10–20 per cent of children suffer from diagnosable mental health problems (using ICD-10 or DSM-IV criteria). There are indications of differences between countries, with slightly lower rates found in India and Norway, for example, and slightly higher in Brazil, Bangladesh, and Russia. However, not all of these children need direct service provision. Some diagnosable mental health problems may get better without intervention (such as depressions in mild form). For these reasons, amongst others (in particular the fact that not all current treatments are proven to do more good than harm—as will be discussed below), it is not advised to go down the route recently suggested by the American Psychiatric Association of screening all children in schools and treating all diagnosable difficulties.[2]

Children with impairment due to mental health difficulties

It is children in this category who are likely to be the main target for specialist child mental health provision. The impact of significant and impairing mental health difficulties if not effectively treated can be substantial. Worldwide, suicide is the third leading cause of death amongst adolescents; major depressive disorder, often starting in adolescence, is associated with substantial psychosocial impairment; conduct disorders amongst children tend to persist into adult life and are reflected in later drug abuse, antisocial behaviour, and poor physical health. This group covers a wide variety of children and young people with problems ranging from bed-wetting to psychosis. It includes those with chronic and multiple difficulties and those with discrete and defined difficulties. It encompasses those with difficulties where there are known to be effective interventions and those presenting with problems where the best course of action is much less clear (as will be discussed below). It is this very mixed range of problems that child mental health services must address.

Prioritizing needs

The way needs are prioritized in planning service provision will be heavily influenced by the context in which services are located. In some countries independent child mental health provision and unaffordable luxury when pitted against other basic needs. The World Health Organisation[3] has suggested that where resources

are particularly limited, priority for funds for child mental health provision should be given to those children with existing difficulties which are:

- occur frequently (and/or have highest cost implications)
- cause a high degree of impairment
- have the greatest long-term care/cost consequences
- have an evidence base for treatment and (particularly in those countries with the most limited resources)
- where the difficulties can be dealt with in primary care or universal services such as schools or GPs.

In countries with greater resources, child mental health professionals' input can be conceptualized as being provided at universal, targeted, and specialist levels. This involves supporting and working alongside universal provision to promote emotional well-being whether in schools or via primary health care workers. Targeted provision aims to promote emotional well-being in those deemed either at most risk (group 2 above) or in most need (group 1 above) with the groups being determined by policy imperatives. Here specialist mental health professionals will work alongside workers from other sectors who take a lead in relation to the needs of these groups as a whole, such as social welfare workers and primary care staff. Specialist provision aims to intervene primarily with those with existing impairment due to mental health difficulties (group 4 above) and can be provided at a local community level, though for more rare and specialized resources may be provided at a regional ore even at a national level.

Weighing mental health promotion initiatives against interventions for those with existing impairments requires particularly careful thought. Whilst there is evidence that promotion programmes may sometimes promote emotional well-being, research to date has not proved that this will reduce levels of significant disturbance and thus impact on specialist services, nor that such programmes are necessarily the most effective approach. It therefore does not seem warranted at this stage to assume that investment in prevention can be done at the expense of investment in services for those with existing impairing mental health difficulties.

In planning response to 'need', it is also important to consider the potential negative and even harmful impact of increased specialist mental health services. Mental health professionals are frequently in danger of assuming that more specialist mental health provision is unquestionably an unalloyed good. The need for more provision must be set in the context of other (sometimes competing) 'needs', such as: the primary need of children to be nourished, sheltered, and protected; their need not to be stigmatized or miss education; and their need not to receive inappropriate, ineffective, or harmful treatment. At times an inappropriate mental health focus can be an unhelpful drain on resources. One documented example is when well-meaning voluntary groups entered a country following a disaster to provide 'interventions for PTSD' that were not linked to other relief efforts and actually interfered with and undermined key initiatives.[1] Whilst the costs of not providing effective specialist mental health inputs can be high, it is important to remember there are also costs to providing unhelpful services.

What sort of interventions should be provided?

As yet, few services have been developed on the basis of considered evidence. Systems of care in CAMHS (as for many areas in health care) have historically been developed on the basis of beliefs, assertion, and innovation within the limits of given structures but with little reference to the slowly and tentatively emerging evidence base. The arguments for trying to promote evidence-based service development are compelling. Natural biases in reasoning mean that people tend to make decisions based more on things that fit their assumptive world view than those that challenge it and are more influenced by the charisma of those promoting a particular approach than by evidence for its effectiveness. When the evidence base is not used as the basis for service development, it makes it more likely that seemingly plausible but ineffective and/or harmful interventions may be introduced or continued and that new interventions that have been shown to do more good than harm may never be introduced.[4]

The evidence base in relation to child mental health interventions, whilst growing, is still limited both in extent and quality. Shortcomings include the sheer paucity of studies, the fact that most research is conducted in United States and there is lack of agreement over appropriate outcome measures. Even where interventions have been found to work in academic studies they are generally not as effective when applied in 'real life' settings. This difference may be due to differences in the populations of children seen, types of interventions made, and/or outcomes assessed. There is increasing evidence of the necessity of carefully implementing all aspects of a particular intervention if it is to be as helpful, and that lack of 'fidelity to model' may account for the lack of generalizability of some of the interventions. On the other hand, the role of non-specific factors such as therapeutic engagement, expectations of change and therapist warmth may need to be taken into account.[5]

Whilst bearing these limitations of the research evidence base in mind there is an emerging consensus in relation to some key interventions[6–8] and increasing attempts to develop guidance for practitioners based on the evidence in this area are being developed (e.g. National Institute of Clinical Excellence guidelines in England, American Psychiatric Association Practice parameters in the United States). Those interested in the detail of the sorts of interventions currently found by the evidence base to be most efficacious should go to these guidelines and may also be interested in attempts to summarize the evidence base[8, 9] (condenses current findings to inform the work of practitioners and others and is freely available via the internet).

It is hoped that as the evidence base develops, so should the sophistication with which it might be interrogated. For every intervention listed above as having 'good evidence' to support it, the following questions should be asked (modified from Kazdin[10], p. 113):

- What are the costs, risks, and benefits of this intervention relative to no intervention?
- What are the costs, risks, and benefits of this intervention relative to other interventions?
- What are the key components that appear to contribute to positive outcomes?
- What parameters can be varied to improve outcomes (e.g. including addition of other interventions, non-specific clinical skills, etc)?
- To what extent are effects of interventions generalizable across (a) problem areas, (b) settings, (c) populations of children, and (d) other relevant domains?

We are a long way from being able to answer these questions in relation to CAMHS currently but it is to be hoped that this will develop with time.

However, even as the evidence base grows from academic studies this must still be treated with caution. It is not suggested that findings from even the most rigorously undertaken randomized controlled trials can necessarily be applied wholesale to all individuals with similar problems. The full context of an individual's range of needs and circumstances must be taken into account and it is also hoped that understanding of the academic literature will be supplemented with a growing evidence base emerging from practitioners own routine evaluation of their own work (see discussion in 'ensuring service quality' section below).

How should the service be structured?

If the evidence base for types of intervention is limited, this is even more so for types of organization.

The 'Fort Bragg Study' (and the subsequent Stark County Study) conducted by Bickman and colleagues in the United States warrants particular attention as they generated such high levels of interest and controversy. This study evaluated a large-scale system change project designed to improve outcomes by providing an unrestricted set of coordinated inputs from a range of services. Results were compared with other sites using traditional services. Information was collected on service use, cost, satisfaction, clinical, and functional data over the course of the 3-year study and at follow-up since. The study found greater access and increased rates of satisfaction and less use of inpatient services, but no differences in behavioural–emotional functioning overall and the cost was much greater at Fort Bragg. The subsequent Stark County Study also found that a multi-agency system for care led to no significant difference in clinical outcomes when compared with routine services. Bickman concluded 'the current national policy of large investments in systems of care infrastructure is unlikely to affect children in the manner intended . . . we need to focus on the services or treatments themselves to improve outcomes'.[11]

This conclusion has been hotly contested. Flaws in the study design, implementation, and interpretation have been highlighted. It was argued that services were overwhelmed in the early months and provided less effective interventions, that the researchers failed to take into account some of the positive outcomes on some of the measures used and so on. It has been suggested that there is evidence that 'wrap around services' (whereby a range of services are provided in a coordinated fashion to children and families based on their needs and not those of the services) do produce better outcomes and that the poor results above may have been due to lack of fidelity to this model.[12]

There is some evidence that shared understandings at the highest level and pooled budgets seem to be crucial to positive multi-agency collaboration. 'MHSPY' in the United States is an initiative that involves five public and two private agencies who have come together with blended funding to provide a coordinated focus on a group of 'at risk' and ill children and where outcomes look positive though lacking a control group or clear cost-effectiveness data as yet.

There is little evidence available currently to guide service developers as to the best way to structure services in relation to a number of other key issues, such as: what age of children should be seen by different services? Is it best to structure services around particular problems or around different age groups? What should be the balance between inpatient and outpatient provision? Whilst it is hoped that with time, more evidence will emerge in relation to these issues, it may be hypothesized that, here too, non-specific factors, may be relevant. Thus the leadership skills of the clinicians who establish a given service, the strength of their commitment in a particular shape of service, and the ability of key individuals to collaborate across whatever boundaries are inevitably created, may impact as much, if not more, than the specifics of different forms of service structure.

Who should the staff be?

In terms of who the key staff should be who provide specialist mental health services for children, whilst there is much assertion and rhetoric as to what is the ideal workforce composition and an emphasis on multi-disciplinary and multi-agency team development wherever possible, there is, in fact, very little hard evidence as to what an ideal workforce should look like.

In many parts of the world specialist mental health professionals are in short supply, and there is an increasing focus on capitalizing on opportunities for mental health provision to piggyback on other sectors with more provision (such as HIV programmes in Africa or schools in many areas of the world). The potential contribution of paediatricians, primary care workers, and teachers is increasingly recognized, though care must be taken not to overload already stretched systems and research highlighting the importance of 'fidelity to model' (as discussed above) suggest that there may be a necessary threshold of amount of training combined with ongoing organizational support, for subsequent interventions to be effective.

How can the service be made most accessible?

Attempts to make provision more accessible as well as less stigmatizing can include use of education and primary care workers, community and religious leaders, and family networks. There is evidence that linking with relevant belief systems of the community, such as ayurvedic treatment and yoga in parts of India, may be crucial, and that providing training for traditional healers may help increase child mental health capacity. The work of voluntary organizations can be a major source of innovation and energy and communities need to find ways for statutory services to learn from them. For example the walk in community centre, Empilwent in South Africa pioneered a more accessible drop in approach.

It could be argued that child mental health services have been slow on the whole to 'get out of the clinic' and in particular to embrace the impact of new technologies that might increase access. However some interesting innovations are developing. These range from mobile services such as the peripatetic child mental health team which travels across Germany providing follow-up on previously hospitalized young people, providing new consultations, and supervision of institutions for children[13] to increased use of telephone and text contact generally, and telephone helplines and websites specifically.

Telephone helplines such as 'Kids Help'[14] in Australia, 'Childline'[15] in the United Kingdom, and 'Parents Information

Service' (YoungMinds Parent Information Service)[16] in the United Kingdom, are often run by charities independent of statutory provision but it may be in the future that greater links can be made by such resources and more traditional services. 'Telefono Azzuro' of Italy shows the way these can be combined with other provisions.[13]

A range of websites have been developed by and for young people some of which allow for discussions between young people and others offer access to specialist advice and support (e.g. 'ru-ok.com').[17] The use of emails and text messaging to communicate between professionals and young people is now increasing as are the use of IT aids to assessment and treatment and the pace of innovation in this area is likely to accelerate sharply in the future.

How can service quality be ensured?

The implementation of routine outcome evaluation is likely to be a crucial factor in ensuring service quality. There are increasing attempts to ensure routine outcome evaluation is in place. The CAMHS Outcome Research Consortium (CORC)[18] (chaired by the author) is collaboration between over half of all services across the United Kingdom who are implementing an agreed model of routine outcome and agreeing jointly ways to present the data in order to inform service providers and users and to help inform service developments (CORC).[8] The approach has parallels with that being employed in the United States and Australia. In all cases a small suite of measures is completed by service users and providers at initial contact and at some time(s) later. Increasingly models are being developed whereby individual practitioners evaluate outcomes over the course of treatment which can be immediately compared with outcomes from others with similar difficulties at the outset, allowing practitioners to get feedback about progress relative to others.

For child mental health service to institute routine outcome evaluation and to develop into self-reflective and learning organizations, requires functioning IT systems and an agreed core dataset.[19] There are now datasets for child mental health emerging in some countries.[20] But even in the most developed countries, the promise of coherent IT systems, remains more a hope than a reality. It is key that IT systems develop that are capable of supporting decision-making by providing easy access to information for practitioners so that they can get feedback on their work and learn accordingly.

There is an increasing emphasis on the importance of involvement of service users to inform service development priorities and to be part of any evaluation of services, to ensure quality. Some innovative models have been tried such as in Ohio which has created outcome systems from service user perspective and in parts of the United Kingdom where young people are trained as service evaluators though again the impact of these on quality of services in the future is yet to be researched.

Conclusion

There is no one ideal model for organizing services, however some key principles can be identified:

1 It is helpful to start with an analysis of the range of anticipated needs, and to seek clarity as to the current priorities for a given community.

2 Where resources are limited, specialist mental health provision should be targeted foremost on those with greatest levels of impairment for whom there are the most effective interventions.

3 Implementation of effective models of prevention and promotion should be supported so far as resources allow and not at the expense of provision for those with existing difficulties.

4 The existing evidence base in relation to effective interventions and models of delivery should be seen as a starting point for practice, though the limitations of the evidence base should be held in mind.

5 Successful service development is likely to rest on collaboration across health, social care, and education—using pooled budgets where possible.

6 Good information gathering systems are needed that allow individual practitioners and managers to audit and review work.

7 Staff members should be encouraged to adapt their practice in the light of the emerging evidence from literature and their own outcomes rather than retaining allegiance to any one theoretical model or framework.

8 Collaborations between service providers that focus on developing best practice may be of help.

9 The input and involvement of service users may be of relevance in helping develop accessible and acceptable services.

10 Remember not all service developments are benign—always weigh up possible harm that may be caused as well as potential for good.

Further information

Wolpert, M., Fuggle, P., Cottrell, D., *et al.* (2006). *Drawing on the evidence: advice for mental health professionals working with children and adolescents-revised version* (Child and Adolescent Mental Health Services Evidence Based Practice Unit-University College London) http://www.ucl.ac.uk/clinical-health-psychology/pdf Files/DotEBooklet2006.pdf

Wolpert, M. (2008). Organisation of services for children and adolescents with mental health problems. In *Rutter's child and adolescent psychiatry*, (eds. M. Rutter, D. Bishop, D. Pine, *et al.*) Blackwell, Oxford.

Goodman, R. and Scott, S. (2005). *Child psychiatry* (2nd edn). Blackwell, Oxford.

References

1. World Health Organization. (2005). *Atlas: child and adolescent mental health resources. Global concerns: implications for the future*. World Health Organization, Geneva.
2. American Psychiatric Association. (2002). *Quality indicators-defining and measuring quality in psychiatric care for adults and children.* American Psychiatric Association.
3. World Health Organization. (2003). *Caring for children and adolescents with mental disorders: setting WHO directions*. World Health Organization, Geneva.
4. Gray, M.J.A. (2001). *Evidence-based healthcare: how to make health policy and management decisions*. Churchill Livingstone, Edinburgh.
5. Fonagy, P., Target, M., Cottrell, D., *et al.* (2002). *What works for whom? A critical review of treatments for children and adolescents*. Guilford Press, London.

6. Goodman, R. and Scott, S. (2005). *Child psychiatry* (2nd edn). Blackwell, Oxford.
7. Rutter, M., Tizard, J., and Whitmore, K. (1970). *Education, health and behaviour*. Longman, London.
8. Wolpert, M., Fuggle, P., Cottrell, D., *et al.* (2006). *Drawing on the evidence: advice for mental health professionals working with children and adolescents-revised version* (Child and Adolescent Mental Health Services Evidence Based Practice Unit-University College London) http://www.annafreudcentre.org/dote_booklet_2006.pdf
9. CAMHS publications, London. (2007). Choosing what's best for you. What scientists have found helps children and young people who are sad, worried or troubled. http://www.annafreudcentre.org/ebpu/choosing.pdf
10. Kazdin, A.E. (2004). Evidence-based psychotherapies for children and adolescents: strategies, strengths and limitations. In *Facilitating pathways: care, treatment and prevention in child and adolescent mental health* (eds. H. Remschmidt, M. Belfer, and I. Goodyer), pp. 103–18. Springer-Verlag, Berlin.
11. Bickman, L., Lambert, E.W., Andrade, A.R., *et al.* (2000). The Fort Bragg continuum of care for children and adolescents: mental health outcomes over 5 years. *Journal of Consulting and Clinical Psychology*, **68**(4), 710–6.
12. Bruns, E.J., Burchard, J.D., Suter, J.C., *et al.* (2005). Measuring fidelity within community treatments for children and families. In *Outcomes for children and youth with emotional and behavioral disorders and their families*, Vol. 2 (eds. M. Epstein, A. Duchnowski, and K. Kutash). Pro-ED, Austin, TX.
13. Remschmidt, H., Belfer, M., and Goodyer, I. (eds.) (2004). *Facilitating pathways: care, treatment and prevention in child and adolescent mental health*. Springer-Verlag, Berlin.
14. Kids Help Line. http://www.kidshelp.com.au/home_KHL.aspx?s=6
15. Childline. http://www.childline.org.uk/
16. YoungMinds' Parents Information Service. http://www.youngminds.org.uk/parents
17. RU-OK.Com. http://www.ru-ok.com/
18. CAMHS Outcome Research Consortium. (CORC). http://www.corc.uk.net
19. Wolpert, M., Bartholomew, R., Domb, Y., *et al.* (2005). *Collaborating to improve Child Mental Health Services. CAMHS outcome research consortium handbook. Version1.* CORC, London.
20. http://www.ic.nhs.uk/our-services/standards-and-classifications/dataset-list/camhs

9.5.8 The management of child and adolescent psychiatric emergencies

Gillian C. Forrest

'Child' is used throughout this chapter to refer to anyone aged under 18, to avoid the repetition of 'child or young person'. 'His' and 'her' are used interchangeably.

Introduction

This chapter provides a practical approach to the management of psychiatric emergencies in children and adolescents. Such emergencies are challenging for a number of reasons. The professional resources available are usually very limited, and there is often confusion or even disagreement between professionals over what constitutes a psychiatric, as opposed to a social emergency. The parents or carers play a key role in the situation and need to be engaged and involved appropriately in the assessment and management; and issues of confidentiality and consent need to be taken into account. In addition, the psychiatrist may find himself or herself working in a variety of settings—the child's home, a hospital emergency department (A and E), a police station, a children's home, or residential school—where the facilities for assessing an angry, disturbed, or upset child may be far from ideal.

Most emergencies occurring in community settings involve externalizing behaviours: aggression, violence; deliberate self-harm, or threats of harm to self or others; or extreme emotional outbursts.[1,2] Some will involve bizarre behaviour which could be an indication of serious mental illness or intoxication by drugs or alcohol, or a combination of both. The emergency situation often arises in the context of acute family conflict or distress.

Frequently other agencies are involved before the psychiatrist is called in (for example, emergency room staff, social workers, or the police). The on-call psychiatrist needs to be familiar with or able to obtain immediate advice about his or her local child and adolescent psychiatric services, the local child protection and child care procedures, and with the relevant mental health and child care legislation.

Vignette 1:

A 12-year-old boy is in the police station, after attacking a neighbour and smashing a window. He punched a police officer when they tried to pacify him. He has refused to talk to the police, and is sweating and dishevelled, pacing up and down and muttering to himself. The police think he is psychotic. The neighbour is in the waiting room; his father has been called back from work.

After obtaining the history of the incident from the neighbour and the police, the psychiatrist interviewed the father. The boy had been diagnosed with an autistic spectrum disorder when he was 5. He was easily upset by any change of routine. His mother had gone away for a few days, leaving him in the care of a neighbour while his father was at work. This aggressive outburst was precipitated by the neighbour refusing to let him watch his favourite video. The boy calmed down with his father and a mental state assessment confirmed autistic features but no psychotic symptoms. The psychiatrist persuaded the neighbour and the police to drop charges and the boy was allowed to return home. The family were offered an out-patient appointment but declined this, and arrangements were made for his General Practitioner (GP) to review the situation in a few days' time.

Vignette 2:

A 15-year-old girl has locked herself in the bathroom at home with a carving knife and is threatening to cut her wrists. Her parents have been unable to persuade her to come out and are now distraught. The GP was called but she refuses to speak to him.†

When the psychiatrist arrived, he asked the family to withdraw so that he could talk to the girl through the bathroom door. She was very upset and described how she had been dumped by her boyfriend earlier that day, had had a row with her step-mother about a large phone bill, and now felt that she didn't want to live any more. The psychiatrist persuaded her to come out to talk things over with her parents. There was no evidence of a depressive disorder and after the psychiatrist helped the girl share her distress about her boyfriend with her

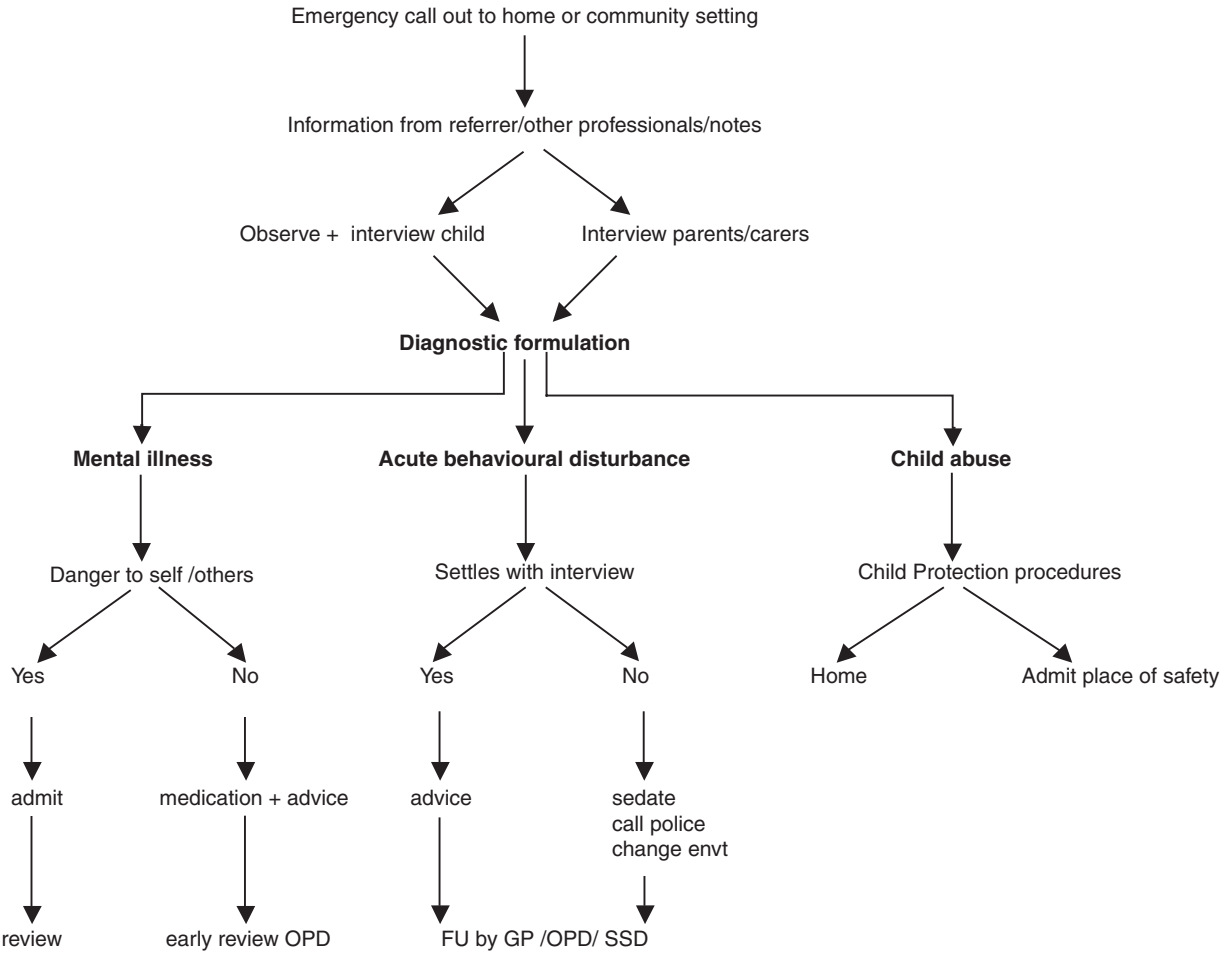


Fig. 9.5.8.1 Flow chart of management of child and adolescent psychiatric emergencies.

parents, she calmed down. There was a history of one previous attempt at self-harm, also connected with peer relationship difficulties, and the psychiatrist felt that there were some underpinning issues around family communication. He offered the family an out-patient appointment to explore these further.

In the emergency situation, the role of the psychiatrist is to:

- assess the disturbed behaviour
- make a diagnostic formulation which distinguishes behaviour associated with mental illness from that which is a reaction to an upsetting event or situation
- produce a care plan to manage the situation safely and effectively in the short-term.

A step-by-step approach to the management of child and adolescent emergencies is described here. The management of deliberate self-harm is dealt with in Chapter 9.2.10.

Step 1. Gathering information

Before attempting to assess the child and his or her family, it is very important to gather as much background information as possible about the incident or behaviour from any professionals who have been directly involved. They may be able to give an accurate description of the child's behaviour and any incident which preceded it; or useful background information about the family. Take time to read any notes that are available; and talk to those involved (GP, police, paediatric staff, A and E staff, etc.).

Step 2. Interview with the child/young person

Although this is described first, parents may expect to be seen before the child, or it may be preferable to talk to them first to gather background information (see Step 3).

A disturbed or upset child in this situation will be wary of any stranger, and the more out of control they are, emotionally and behaviourally, the more likely they are to be uncooperative initially. The psychiatrist needs to calmly establish that they have come to try and help the child, that they are non-judgemental and that they can be trusted. She should always ask to see the child alone, to establish a confidential relationship, and to be able to assess the child without the influence of parents.

Honest and open communication is essential, especially if there is any possibility of child protection issues. Often the child is terrified

that they are going to be forcibly taken away from their home or family (a desperate and angry parent will often threaten this), and this fear may need to be addressed before the assessment can begin.

A useful approach is to say to the child: 'I'm Dr ….. I'm a psychiatrist who sees children and young people with all sorts of worries and troubles. I've come to see you now, to try and understand what this is all about, and see if I can find a way to help'.

If the child refuses to talk at this stage, they may be angry, mistrustful, or psychotically withdrawn. Patience is required to see if this can be overcome by convincing the child that you have come to listen to them and not to 'take them away'. Reflecting back to the child how you perceive their emotional state may also help. For example, the psychiatrist might say: 'You are obviously very upset and angry. I wonder what it is that's made you feel this way'. Sometimes children will choose to communicate by writing notes or drawing, rather than by using speech, and if the child is mute, the psychiatrist should try offering paper and pencil.

If the child remains violent and out of control and the behaviour is being reinforced by attention and anxiety, it will be necessary to ask everyone to withdraw for a while to give the child the chance to calm down on their own. If this still does not help to calm the situation, the psychiatrist will need to set some limits, which could include calling the police to regain control of the behaviour. Medication should only be considered as a last resort when all other strategies to deescalate the situation and calm the child have failed (see Step 5: Management of emergencies in community settings section).

When the child starts to be able to communicate with the psychiatrist, they can then be encouraged to give their account of what happened to precipitate the behaviour. As this is explored, it will be possible to assess the child's mental state (see Chapter 9.1.3) and to form an impression about whether this is mental illness, or whether the disturbed behaviour is the result of emotional turmoil secondary to relationship problems, family conflict, or other psychosocial issues. Further information from the parents or carers will be needed to clarify this.

If child abuse is suspected, an interview with the child alone is crucial. The psychiatrist will need to enquire about the possibility of physical and sexual abuse without using leading questions[3] (see Chapter 9.3.2). A good starting point is 'Has anyone ever hurt you or made you do anything you didn't like?' The child needs to understand that you may need to breach confidentiality in order to protect them from further harm. (see Consent and confidentiality section).

Step 3. Interview with parents/carers

It is essential to see the parents or carers as part of the assessment, in order to fully understand the context of the crisis, and identify any relevant contributory factors. They may be seen first, or after the child's assessment, depending on the situation. It is often helpful to see the parents or carers on their own so that they can speak freely about their child.

These are the areas that will need covering:

- History of current episode/disturbance. Current medication?
- Previous emotional or behavioural problems.
- Brief developmental history: normal or unusual features? Learning difficulties? School and academic progress.
- Temperament; relationships with family/extended family members, other adults, peers.
- Family composition and history (parental separations/divorce; relationships mental and physical illness; recent life events).

During this interview, an assessment can be made of the family functioning; the parent's or carer's mental health, and their attitude to the child, and whether this seems to be playing a part in the current problem (e.g. high levels of family discord; rejecting, neglectful or hostile attitudes to the child; harsh parenting practices; parental depression, anxiety, drug or alcohol abuse).

Vignette 3:

> A 10-year-old has been raging around his house for several hours, following an argument with his mother about going out with his friends. He has been smashing toys and kicking in doors and wrecking his bedroom. His single mother and 6-year-old sister are terrified and cowering in the sitting room.

The psychiatrist spoke calmly to the boy through the door, stating that he wanted to talk to him. The boy continued being aggressive and abusive and the psychiatrist then said he would give him 5 minutes to calm down and come out, otherwise the police would have to be called in to help. The boy did then manage to calm down and talked about his resentment of his mother's rules, and his angry feelings about the loss of his father, who had had little contact with him since leaving the family.

His mother told the psychiatrist that she and the boy's father had divorced a year ago, and there had been on-going disputes with her exhusband about money and his erratic contact with the children. Recently she had become depressed, and her son had become increasingly irritable and oppositional at home. She felt unable to control her boy's temper outbursts.

The mother agreed to see her GP about her depression, try and talk to her ex-husband about seeing the children on a regular basis, and attend the clinic with the children for some family counselling sessions. The boy appeared relieved that his mother was going to seek some help.

Step 4. Making and sharing the diagnostic formulation

There should now be sufficient information available to make a diagnostic formulation. This will include the diagnostic category for the disturbed behaviour, the factors which led up to the acute disturbance; the factors which precipitated it, and the risk factors for continuing or recurring problems.

The Three Ps:

Predisposing Factors

Precipitating

Perpetuating

These factors may be present in the child herself (temperament, illness); in the family environment; at school or college; in the child's wider social environment (friends, neighbours; clubs, etc.); or any combination of these.

Apart from acute intoxication with drugs or alcohol, and deliberate self-harm, most child and adolescent psychiatric emergencies in the community fall into three categories:

- mental illness
- acute behavioural disturbance due to family/psychosocial factors
- child abuse

The psychiatrist should share his formulation with the child and parents/carers, using age-appropriate, and jargon-free language, to help their understanding of the events. Sometimes the parents or the child will need further discussions before a shared understanding between the psychiatrist and the family about the problems can be reached, and an acceptance of the proposed management/care plan by the family.

The formulation should also be shared with any other professionals present, as they may be needed to contribute to the immediate management plan, or even to take over the care of the child.

Step 5. Management of emergencies in community settings (see Fig. 9.5.8.1)

Mental illness

The emergency management of a child diagnosed with a mental illness will depend on the risk assessment. Is the child a danger to himself or others? If the psychiatrist considers this to be a high risk, then referral to inpatient care will be necessary to allow further assessment and treatment in a safe setting. Compulsory admission, using mental health legislation, may be required if the child or the parents refuse to cooperate with this plan (see Consent section). In some places, it may be possible to admit younger children to a paediatric ward for further observation and treatment. Where there are medical complications of a mental illness (for example, low output cardiac failure in severe anorexia nervosa), admission to a paediatric or medical ward will be needed.

If there is a low risk of any danger, and the child and family are cooperative, the psychiatrist can prescribe appropriate medication for the mental illness (see Chapter 9.5.5), give advice on the management of the symptoms in the short-term, and arrange for early follow-up and review in the outpatient clinic.

Acute behavioural disturbance (no mental illness)

Very often, the situation is defused by the psychiatrist's assessment, and the child's behaviour settles. In this case, the psychiatrist will need to decide whether there is a high risk of further episodes. If not, she may simply give advice to the family about how to avoid or deescalate future situations, and inform the child's GP about the incident. However, if she has identified significant ongoing issues underpinning or precipitating the child's disturbed behaviour, she will need to give the family advice about how to try and deal with these. For example, if the child has difficulties with family conflict, the psychiatrist could recommend referral for outpatient family therapy or anger management; if stress at school was linked to the outburst, she might advise the parents to arrange a meeting with the child's teachers; or she could advise a parent to seek help for their own mental health problems (such as depression or addiction) or social problems such as overcrowding or financial difficulties.

If the disturbed and violent behaviour continues in spite of the psychiatrist's intervention, a decision must be made about how best to regain control of the situation. The police may need to be called in, to provide sufficient manpower to safely control the violent behaviour. It may be possible to change the environment by calling in significant other people (for example, another relative) who can calm the child. The use of medication needs very careful consideration.[4] Children may react paradoxically to sedatives and tranquillizers, and the use of medications with the risk of respiratory suppression is not recommended unless life support equipment (and staff) is available. However, sometimes it may be possible to persuade a child to accept an appropriate dose of oral medication such as lorazepam or midazolam, and this may be beneficial.[5]

Child abuse

Although it is not common for child abuse to present as a psychiatric emergency, nevertheless it is vitally important for any psychiatrist on call to be able to recognize this, and manage the situation according to local child protection procedures. If the child is making allegations of abuse, these will need to be discussed with the parents/carers, and then it will be necessary to consult with other professional colleagues (including the police, social services, or paediatric staff) to decide on how best to care for the child. The welfare of the child must always be paramount, even if this involves the child being taken to a place of safety such as a paediatric ward or to alternative accommodation while further investigations are carried out. Clear and accurate notes of all aspects of the assessment are vital as they may be needed for legal proceedings later. (The assessment of possible child abuse is considered further in Chapter 9.3.3.)

Emergencies in paediatric/medical wards

Psychiatric assessment and advice is needed at times for children showing acutely disturbed behaviour on paediatric wards, as well as for deliberate self-harm.

The child may be acting bizarrely or aggressively. The psychiatrist will need to make a thorough assessment of the situation, including the background of the child and family and the medical aspects of the case. He will then need to assess whether the behaviour is part of the physical illness, (e.g. hypoglycaemia, hypoxia, delirium, pain), a primary mental illness (e.g. psychosis, somatoform reaction), the side effects of medication, or an acute behavioural reaction. Management will need to be discussed and agreed with the paediatric staff, and shared with the family if they are available. The psychiatric assessment and care plan should be carefully written up in the medical file to ensure continuity through shift changes of the ward staff.

The use of sedation needs to be given careful consideration, and used only if other strategies fail (such as finding out from the child the cause of any upset and talking it over; moving the child to a quieter setting; calling in the parent to help reassure or pacify the child). Oral lorazepam or midazolam are useful drugs of first choice.

If the primary problem is one of mental illness and the child cannot be cared for effectively and safely in a paediatric ward setting, bearing in mind the needs and safety of the other patients, then transfer to a psychiatric ward will need to be arranged as soon as possible.

Follow-up visits or phone calls to the ward will be needed, to make sure the child is improving or to reconsider the diagnosis or treatment.

Psychiatric problems in paediatric or medical wards are considered further in Chapter 9.3.4

Step 6. Follow-up and communication with other professionals

Effective management of the emergency situation can be undermined by inadequate follow-up arrangements or poor communication between professionals and agencies. It is therefore very important

that the on-call psychiatrist communicates with the child's general practitioner, the team providing routine care for the child, and any other relevant professional involved as soon as practicable, usually the following day. This will hopefully ensure that there is a seamless provision of care for the child and the family, and prevent or forestall future emergencies.

Consent and confidentiality issues

(a) Consent

Psychiatrists who assess and treat children and adolescents, like all mental health professionals, have a duty of care to ensure that the welfare of children is protected. They must also work within the limits of the laws of the land. This means that care and attention must be paid to the rights of children to consent to their treatment, while at the same time taking into account the circumstances, their mental capacity, their parents' views and the risks and necessity for treatment. In many cases, the sharing of information with the child and the family, and their involvement in the decision about the best care will ensure that an agreement can be reached. However, there are other occasions when either the child or their parents refuse to cooperate with treatment felt to be essential by the psychiatrist. If the child is refusing but the parents consent, it is usually possible to treat the child under parental consent, or for adolescents, to use other legal means (for example, the mental health legislation to allow the hospitalization and treatment of a mental disorder). If the child is consenting to treatment, but the parents refuse, care proceedings may need to be considered. If both child and parents are refusing treatment, the situation is very difficult and a second opinion on the need for treatment and legal advice may have to be obtained before anything further can be done.

(b) Confidentiality

The confidentiality of the doctor–patient relationship, as mentioned in the Hippocratic Oath, is a fundamental prerequisite for patients being able to trust their doctor with sensitive information. However when working with children and adolescents, confidentiality is complicated by the need to share information with parents, in order for them to fulfil their responsibilities to their child, and also with the other agencies which may be involved in providing services for the child (education, social care, etc.). Children and their families should be informed of the scope of any promises of confidentiality at the beginning, and it is always good to ask for the child's or adolescents' consent to share information where this is necessary and in the best interests of the child. Breaches of confidentiality may be clinically or legally necessary (such as where an adolescent with a mental illness is exposing them self to risk, or when child abuse is suspected).

Tan *et al.*[6] provide a full discussion of these issues, and decision-making algorithms. Effective management of an emergency should result in the child receiving appropriate care and treatment, the parents having a greater understanding of the issues associated with the disturbed behaviour, and any other professionals feeling supported by the psychiatrist's advice and intervention. This will all reduce the likelihood of further emergency situations arising for that child and family.

Further information

Jones, D.P.H. (2003). *Communicating with vulnerable children*. Gaskell, London.

References

1. Halamandaris, P.V. and Anderson, T.R. (1999). Children and adolescents in the psychiatric emergency setting. *The Psychiatric Clinics of North America*, **22**(4), 865–74.
2. Stewart, C., Spicer, M., and Babl, F. (2006). Caring for adolescents with mental health problems: challenges in the emergency department. *Journal of Paediatrics and Child Health*, **42**(11), 726–30. DOI: 10.1111/j.1440-1754.2006.00959.x
3. Jones, D.P.H. (2003). *Communicating with vulnerable children*. Gaskell, London.
4. Sorrentino, A. (2004). Chemical restraints for the agitated, violent or psychotic paediatric patient in the emergency room; controversies and recommendations. *Current Opinion in Pediatrics,* **16**, 201–4.
5. Taylor, D., Paton, C., and Kerwin, R. (eds.) (2007). Rapid tranquillisation (RT) in children and adolescents. In *The Maudsley 2005–2006 prescribing guidelines 8E,* p. 226. Taylor and Francis Group, London and New York.
6. Tan, J.A.O., Passerini, G., and Stewart, A. (2007). Consent and confidentiality in clinical work with young people. *Clinical Child Psychology and Psychiatry*, **12**(2), 191–210. DOI: 10.1177/1359104507075921.

9.5.9 The child psychiatrist as consultant to schools and colleges

Simon G. Gowers and Sian Thomas

Introduction

Those who provide public services for children and young people may have a role in the identification, prevention, and within reasonable parameters, the treatment of mental health problems. Social services and education in particular have a key responsibility to safeguard the physical and psychological health of children and identify potential areas of avoidable harm, including those which may develop within their institutions.

There is a well-recognized mismatch between the rates of child mental health problems identified in epidemiological studies and the number of children referred to child and adolescent mental health services (CAMHS). School staff will often be in the best position to identify unrecognized difficulties and also to help children and their families address prejudices associated with referral to CAMHS, though they may need training and help to do so.

The responsibilities of teachers have been confirmed by schools' inclusion within the broad concept of CAMHS in a number of countries. In the United Kingdom, the Health Advisory Service (now the Health and Social Care Advisory Service—HASCAS), proposed a model, subsequently adopted by the Department of Health of a tiered approach to service provision, in which schools, alongside primary medical care and social services formed the first Tier.[1] Within this model schools have been seen as offering unique opportunities to identify problems, provide simple assessments and refer up to more specialized tiers as judged appropriate and in negotiation with caregivers. Teachers though, often feel inadequately trained to fulfil this role and look to other professionals, including psychiatrists to advise and support them. Fortunately there are a number of professional roles, some employed within

education and some outside, forming a bridge between education and mental health services. Some of these roles vary in their detail between countries, but most developed countries will have professionals (possibly with different titles) filling roles comparable to those in the United Kingdom.

It is important to note that CAMHS generally work as multidisciplinary teams, hence any support and liaison may be offered by a range of professionals and not exclusively by psychiatry. One of the HAS recommendations was the creation of a new professional group—the primary child mental health worker—with the particular aim of liaising between Tier 1 and Tier 2 services. The following are some of the professionals involved in the interface between child mental health and education:

Primary child mental health worker

A practitioner, often with a mental health nursing background, employed either by education or CAMHS with the specific brief to liaise between the two in identifying children with mental health needs.

Special educational needs coordinator (SENCO)

Employed by the school as a teacher, the role of the SENCO is primarily to develop effective ways of identifying and removing barriers to learning, which may result from intellectual retardation, physical, or mental health problems. Alongside primary child mental health workers they have a role in the identification, management, and referral of children as well as a responsibility to contribute to in-service training for teachers.

Educational psychologist

Educational psychologists (EP) provide assessments of special educational needs. In the United Kingdom a consultation model has been adopted, whereby the Educational Psychologist meets with the person who has raised concerns as they are likely to be the person most motivated to bring about change. The new model recognizes that teachers are often skilled in assessing pupil attainment, learning styles, behaviour, strengths, and weaknesses. EPs have an important role in early identification and intervention and aim to promote child development and learning through the application of psychological theory using information gathered within a wider ranging context.

Education welfare officer (EWO)

In the United Kingdom, each school has an EWO assigned, to provide a support service to families and schools to help them meet legal obligations related to a child's education. They work with parents/carers to monitor attendance, with schools to consider courses of action of benefit to poor school attendees and with other agencies (e.g. health, social services, police, and youth offending teams) to provide a suitable programme that will help the child return to full-time education.

School nurse

An integral part of the school health team, the school nurse's responsibilities include supporting children with complex health needs, running immunization programmes, providing drop-in clinics, parenting programmes and bed-wetting clinics, assessing the health needs of every child on starting school, and providing health schemes for young people.

Learning mentor

Learning mentors have a broad remit including supporting the safe and effective transition from primary to secondary school, supporting provision for pupils with special educational needs and developing a relationship with identified pupils, based on a trusting individual relationship.

Connexions advisors

Primarily concerned with those in the 13–16 age range, connexions advisors offer one-to-one support and guidance similar to that previously carried out by careers officers. There is a strong emphasis on surveillance and monitoring. Young people who are seen to be 'at risk' of dropping out of education or who present behavioural problems are a priority for intervention. Personal advisers act as advocates, especially for those who are vulnerable or who have special needs.

Teachers' training in mental health

Despite their responsibility for identifying mental health problems, teachers in many countries are offered little specific training in this area. In the United Kingdom, most post graduate certificate of education courses offer only a very small amount of time, perhaps as little as a half day to the teaching of special educational needs. A survey of SENCO'S training and their wish for further teaching about mental health issues[2] revealed a significant lack of training. Many had no training in 3 years. In contrast they showed a great willingness to receive more and welcomed liaison from CAMHS professionals. There have been a number of useful initiatives to improve teachers' experience including the National Healthy Schools Programme[3] which aimed to improve learning by reducing emotional and health inequalities using a whole school approach; this involved improving the emotional literacy not only of pupils but of staff and parents too.

Developing a school liaison service

Establishing a liaison service between CAMHS and a school can have a number of benefits including:

- Early identification of child mental health problems
- Information sharing
- Monitoring and evaluation of treatment e.g. for attention-deficit hyperactivity disorder (ADHD)
- Establishing pathways of referral to higher tiers of service
- Offering school-based interventions for common problems
- Promoting the development of social skills and positive self-esteem.

The majority of CAMHS services do work with schools, the nature of the intervention ranging from consultation and support for school staff to direct work with children, including observation and assessment. However, joint working between CAMHS and schools has a record of patchiness across the United Kingdom, with a lack of key personnel often leading to a fragmented service.

Good examples of joint practice are characterized by secondments between organizations, shared working environments, a clear understanding of the different roles and expertise of team members, and a shared vision of joint working.

Where good practice is operating, schools are often faced with anxieties around short-term funding for specific projects, for example, the recent initiatives 'City Education Action Zones', 'Health Action Zones', and the 'Healthy Schools Standards'.

Schools in the United States tend to operate within multidisciplinary settings and research suggests that these are effective in breaking down professional barriers and also addressing the stigma associated with a young person being referred to external agencies such as CAMHS.

The provision of a key mental health worker within the school facilitates better communication between services and helps develop a greater understanding of how the culture of a school operates. Integrated links between CAMHS and the local authority, educational psychologists, behaviour and emotional support teams, and education welfare promotes a cohesive and collaborative service for children.

Practical issues

In order for a CAMHS service to establish an effective working link with a school, there are several issues to address:

(a) Gaining the cooperation of all the staff

Commitment of all staff (and indeed parents) rather than just one interested teacher is crucial. Effective prevention, treatment, and referral pathways require a 'whole school' approach.

(b) Negotiating realistic aims

Child mental health problems are common and often long-lasting; a realistic balance should be struck between prevention and management.

(c) Establishing a level of service

Both the school and CAMHS should be clear about who is providing the service, at what frequency and the expected level of commitment on both sides. There should be perceived benefits to the school and CAMHS. Does the service provide an urgent referral component or not? Who is the named contact?

(d) Confidentiality

Schools say that policies around sharing information act as barriers to effective joint working and so there is a need to determine a process whereby a joint strategy on confidentiality is agreed. At an individual level, young people can expect that private discussions on personal matters should be kept confidential unless they are told otherwise. However, teaching staff should not give unreserved assurances on confidentiality as these may have to be breached if the young person discloses information which leads an adult to believe that they or others are at risk. Sometimes the teacher will need to share information with others in the staff team in the young person's interests, or for supervision purposes. On occasions, (for example, where there is a serious risk of self-harm) it will be necessary to contact parents, but in these circumstances the young person should be told explicitly what has been shared with whom and why. Confidences should not be breached to other pupils.

(e) Pitfalls

There are several dangers of providing a mentoring/counselling service for the inexperienced teacher. Some of the commonest are: becoming over involved (emotionally and with time), giving unconditional guarantees of confidentiality, and dealing inappropriately with pupils concerns about another pupil. Obtaining advice or supervision from a more experienced member of the team or a CAMHS liaison worker is the most effective way of addressing these difficulties.

What do teachers need to know?

(a) Education about mental health problems

Surveys suggest teachers want to understand the common presentations of mental health problems in childhood, how they affect children's behaviour, and their impact on learning. They are often uncertain about aetiology and prognosis. Where disorders have a genetic component to their aetiology, teachers often mistakenly believe that this diminishes the potential impact of school-based interventions. They like to understand the distinction between generalized and specific learning difficulties.

Distinguishing disorder from bad behaviour is especially complex, particularly when attempting to differentiate between what a child can't or won't do. This is commonly an issue with hyperactive children whose attention in school is poor.

(b) Identification/detection

Teachers benefit from guidance on the detection of disorders, by learning about common symptoms and behavioural phenotypes. They can be helped in this by being aware of groups at risk and by the use of screening instruments designed to be used in school. Commonly used measures include the Conners Teacher Rating scale (CTRS)[4] and the Strengths and difficulties questionnaires.[5]

(c) Interventions/treatments teachers can deliver in school

These may include simple counselling interventions, e.g. to address anxiety at exam times or supporting a child during parental separation or after a bereavement. The teacher's role may involve supporting the administration of medication within school (e.g. for hyperactivity). This treatment (and the child's motivation to take it) can be severely undermined if teachers do not support its use.

(d) Knowledge/understanding of treatments given by CAMHS

Many myths about child mental health problems and their treatment may be shared by teaching staff. These include the aims, benefits, and likely adverse effects of medication. Some will mistakenly believe, for example that drugs for hyperactivity are sedative and will turn a child into a 'zombie'. Stigmatizing attitudes to inpatient child psychiatry units may impair a child's rehabilitation after admission.

(e) Pathways of referral

Schools should be clear about which problems should be referred to which agency. For example, acute self-harm should be referred to the accident and emergency department of a general hospital, whilst child protection issues should be referred to social services.

Specific issues posing a challenge for schools

A major concern for a school is how to manage disruptive disorders and their impact on other children. A disorder such as Tourette's syndrome will excite younger children and they will be easily distracted. Those whose attention is poor, who are more impulsive or who are low achievers are particularly vulnerable, whilst the subject may find their behaviour reinforced by their unexpected celebrity and status as the 'class clown'.

Appropriate behavioural management can be difficult to institute without it being punitive or unwittingly reinforcing. This applies to

conduct problems and for example where teaching staff may offer lunchtime supervision of a child with an eating disorder.

It is often difficult to address minor self-harm sympathetically without reinforcing the behaviour. Similarly the wish to be sympathetic towards those with eating disorders may be tempered by concerns to avoid 'epidemic' dieting.

The school's role in child protection

Schools in the United Kingdom have a responsibility to safeguard and promote the well-being of pupils under the Education Act of 2002[6] and, where appropriate, under the Children Act 1989.[7] Each school has a designated lead for child protection and if staff have concerns about the safety and well-being of a child they should report their concerns to them. The child protection lead will refer to the school Child Protection Policy and then directly to Children's Social Care services as necessary.

A guidance document 'safeguarding children and safer recruitment in education' was produced in the United Kingdom in 2006[8] and is a consolidated version of earlier guidance material. It focuses on the recruitment and selection processes, vetting checks, and duties for safeguarding and promoting the welfare of children in education. The document also forms a guide to inter-agency working. The guidance explains that a school should 'create and maintain a safe learning environment' and have the appropriate arrangements in place. Child protection arrangements, pupil health and safety, and bullying are all subject to statutory requirements. The guidance directs that if a child is the subject of an inter-agency child protection plan, the school should be involved with the preparation of that plan.

Through the delivery of personal, health, and social education (PHSE) the school may provide opportunities for children and young people to learn about keeping safe. Pupils should be taught to:

- Recognize and manage risk in different situations and then decide how to behave responsibly
- Judge what kind of physical contact is acceptable and unacceptable
- Recognize when pressure from others (including people they know) threatens their personal safety and well-being and develop effective ways of resisting pressure.

School-based intervention strategies

Primary prevention

A healthy school promotes physical and emotional health by providing accessible and relevant information and equipping pupils with the skills and attitudes to make informed decisions about their health. It understands the importance of investing in health to assist in raising levels of pupil achievement and improving standards. It also recognizes the need to provide both a physical and social environment that is conducive to learning.

The National Healthy School Standard was part of the Healthy Schools programme, led by the DFES and the Department of Health.[9] Launched in October 1999, it offered support for local programme coordinators and provided an accreditation process for education and health partnerships. It provided a model of partnership working between the health service and schools, with the aim of promoting a coherent and holistic message about the importance of a healthy lifestyle.

The standard covered four key themes:

- Personal, Health, and Social Education (PHSE).
- Healthy eating
- Physical activity
- Emotional health and well-being (including bullying)

PHSE is now included in many countries' teaching curricula. This provides education on social and emotional development and citizenship, including the individual's place in society, responsibility, and rights. In The United Kingdom, the PHSE syllabus now has sessions on mental health issues including the use of drugs and alcohol and the links between drug misuse and mental illness. It also covers self-harm and suicidal behaviour.

Social and Emotional Aspects of Learning—(SEAL) is a whole-curriculum framework for teaching social, emotional, and behavioural skills to primary school children in five areas: self-awareness, managing feelings, motivation, empathy, and social skills. In 2004, the scheme was piloted in 250 schools in 25 authorities in the United Kingdom with a subsequent planned extension to high schools.

Secondary prevention

Solution Oriented Schools (SOS) is an approach used in many United Kingdom local authorities, comprising training and resources to support the whole-school promotion of positive behaviour. The focus is on establishing small steps that can be taken to resolve conduct problems, attendance issues, poor peer group interactions and negative attitudes to learning.

The approach invites staff to consider: 'What works in school?'. It encourages them to take a pragmatic approach; learning from what is working, leaving behind practice that is failing to pay-off, recognizing 'the problem' as the problem (not the child, teacher, or professional), and building on strengths that each individual brings. It stems from the principles of solution-oriented brief therapy which focuses on finding and creating solutions to a problem whilst spending little time on the problem itself.

Circle time was developed in the 1930s particularly for primary schools, as a forum for children to share views and concerns about issues arising in school (such as bullying) or outside (within the family or neighbourhood). It is still widely practised.

Mobile phone, text, and Internet-based initiatives

A number of education authorities have developed pilot schemes using new technologies such as online counselling and support services. Various websites offer qualified counsellors and other support services. Services are confidential and young people book in for their session of online chat. In addition the young people can often access a Frequently Asked Questions area. Some sites have counsellors who can make referrals to CAMHS for the young people who access them.

Other initiatives include a pilot scheme in Wales in which pupils used mobile phones to text their school nurses for health advice. This short-term project ran in office hours, offered students instant help and provided those who might be wary of approaching adults with their problems face-to-face the chance to do so anonymously.

In Liverpool, *The Health and Education for Life Project (HELP)* was set-up in 2003, as an action research project that worked in

schools to change pupils' attitudes towards mental health issues and to provide coping strategies.

Meeting the needs of children with special educational needs

While most pupils with complex needs are educated in special schools, where the special needs of children can be met by a mainstream school, they are often taught in this setting. Attitudes to inclusion have changed over time. The latest draft guidance in the United Kingdom moves away from the inclusion drive of 2004 and advises councils that they should provide a 'range of provision' for children with special educational needs. The national curriculum requirements may often impact negatively on the experiences of a young person with specific learning needs; the challenge for teachers is to create a stimulating, engaging programme of study whilst still meeting the national requirements. The use of teaching assistants in a creative and well-planned way can facilitate the delivery of lessons in an inclusive setting. Ultimately, the needs of the individual child should be of primary concern in the inclusive–exclusive decision-making process.

Some special schools offer outreach to the mainstream, so that expertise can be shared and support given for inclusive practice. Some local authorities name specialist schools for each area to meet the needs for example, of autistic young people. Where a young person lives too far from the locality of the specialist school; the specialist school may adopt an advisory role.

Special schools attempt to offer a tailored and focussed response to the needs of specific groups of young people; it is arguable that a young person who has for example, high functioning Asperger's syndrome will feel more 'included' in a special school setting where his differences are less noticeable. The social exclusion that such a young person may experience in a mainstream setting can have detrimental effects on their progress. Successful inclusive education relies on a school approach that creates an inclusive culture; develops inclusive policies, and evolves inclusive practices.

Children and young people with emotional and behavioural difficulties (EBD) present a major challenge to schools attempting to become fully inclusive organizations. Through emotional literacy programmes such as SEAL, anger management groups, social skills, and self-esteem groups, schools can offer a variety of behavioural interventions that support an inclusive experience.

Hospital schools or Medical Pupil Referral Units offer education to young people who are unable to attend school because of medical needs. The term 'medical needs' includes those with mental illness; anxiety, depression, and school phobia. There is a strong emphasis on a strategic planning framework which ensures a continuum of provision; a focus on close liaison with all parties and the development of a robust reintegration plan.

The role of CAMHS within special education has a higher profile than in many mainstream educational provisions; children should all have statements of special educational need and therefore should be reviewed regularly. The reviews of children with a statement of emotional and social need or with autistic spectrum disorders (ASD) tend to require the involvement of a CAMHS worker; subsequently good relationships are developed through close and regular interactions with educationalists. CAMHS may offer consultation and advice or direct intervention with the child or group of children in a special educational setting.

The psychiatrist as advisor to higher/further education

Recent years have shown a growing awareness of the unmet needs of students in higher education.[10] Although higher education institutions often have quite sophisticated pastoral and counselling provision in place, they may need to consult with mental health services regarding issues with specific students. The psychiatrist should be aware of the following particular issues in relation to this group:

- The vulnerability of young people living away from home for the first time.
- Recent expansions in student numbers. In a number of countries, greater training opportunities for young people have resulted in access to higher education no longer being restricted to those from privileged backgrounds. One adverse consequence of this otherwise desirable state of affairs is that those with greater risks (or indeed histories) of mental health problems may take up places at colleges and universities.
- Vocational courses. Some courses (such as medicine, nursing, and social work) may restrict those suffering with particular mental health problems. The psychiatrist may have a duty to report such issues as, for example, drug dependence. Related to this a student may be reluctant to disclose a problem which would have implications for continuing on their course.
- Interface/communication issues. Young people living away from home at college may be vulnerable to falling into any gap that might exist between child and adult services and between services local to their home and those at their college. Effective communication between service providers is of paramount importance.
- Confidentiality. Students are often concerned about confidentiality from their parents, their peers, and their college's academic staff. As with the examples given earlier in this chapter, confidences should only be breached on a strictly 'need to know' basis.

Conclusions

As participation in education is almost universally compulsory for children, schools are in a unique position to offer prevention and identification of child mental health problems. Effective practice requires good liaison with CAMHS. There are a number of obstacles to effective working but recent times have seen a number of examples of good practice and policies to support these.

Further information

Royal College of Psychiatrists. (2003). *Mental health of students in higher education. Council report CR112*. Royal College of Psychiatrists, London.

The National Healthy Schools Programme. www.healthyschools.gov.uk

Safeguarding Children. www.everychildmatters.gov.uk

References

1. Health Advisory Service. (1995). *Together we stand: the commissioning, role and management of child and adolescent mental health services*. HMSO, London.

2. Gowers, S.G., Thomas, S., and Deeley, S.D. (2004). Can primary schools provide effective Tier 1 child mental health services? *Journal of Clinical Child Psychology and Psychiatry*, **9**(3), 419–25.
3. The National Healthy Schools Programme. www.healthyschools.gov.uk
4. Conners, C.K. (1973). Rating scales for use in drug studies with children [Special issue/Pharmacotherapy of Children]. *Psychopharmacology Bulletin*, **9**, 24–9.
5. Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire (SDQ). *Journal of the American Academy of Child and Adolescent Psychiatry*, **40**, 1337–45.
6. Department for Education. (2002). *Education Act*. HMSO, London.
7. Department of Health. (1991). *Children Act 1989; guidance and regulations*. Department of Health, London.
8. Safeguarding Children. www.everychildmatters.gov.uk
9. National Healthy Schools Standard. http://www.standards.dfes.gov.uk
10. Royal College of Psychiatrists. (2003). *The mental health of students in higher education. Council report CR112*. Royal College of Psychiatrists, London.

SECTION 10

Intellectual Disability (Mental Retardation)

This page intentionally left blank

10.1

Classification, diagnosis, psychiatric assessment, and needs assessment

A. J. Holland

Introduction

The general principles developed during the latter part of the twentieth century and continued into the twenty-first century guiding support for people with intellectual disabilities remain those of social inclusion and the provision of services to enable people to make, as far is possible, their own choices and to participate as full citizens in society. These are articulated in national policy documents, such as the White Paper for England, 'Valuing people[1] and also at an international level in the UN Declaration on the rights of people with disability.[2] However, given that people with intellectual disabilities represent a highly complex and heterogenous group with very varied needs, in order for such objectives to be achieved, a range of community based support and interagency and interdisciplinary collaboration is required. It is acknowledged that people with intellectual disabilities experience considerable health inequalities with the presence of additional disabilities due to the presence of physical and sensory impairments and co-morbid physical and mental ill-health, much of which goes unrecognized, and also the occurrence of behaviours that impact on their lives and the lives of those supporting them.[3,4,5] In the twenty-first century, few would now challenge the objectives of social inclusion and community support. The tasks for Government and society are to provide special educational support in childhood and also support to the families of children with intellectual disabilities, and the necessary range of services to meet the social and health needs of this diverse group of people in their adult life. This includes enabling adults with intellectual disabilities to gain meaningful support or full employment and to exercise their rights as citizens and to participate fully in society. To achieve such objectives there is a need to be able to characterize the nature and level of need, to establish the presence and significance of co-morbid illnesses and/or challenging behaviours, and to organize and provide support and services to meet such identified needs.

This complexity of need has meant that no single 'label', such as 'intellectual disability', can adequately describe this group of people. What individuals have in common is a difficulty in the acquisition of basic living, educational, and social skills that is apparent early in life, together with evidence of a significant intellectual impairment. However, for some this may be of such severity that, for example, meaningful language is never acquired and there are very substantial care needs. For others, there is the presence of subtle signs of early developmental delay, and evidence of learning difficulties that only becomes clearly apparent at school when there is an expectation that more sophisticated skills will be acquired. The nature and extent of disability and of any functional impairments in general, distinguishing those people with intellectual disabilities from those with specific learning difficulties, such as dyslexia.

In infancy and early childhood, the reason for any apparent developmental delay needs to be established. This is primarily the responsibility of paediatric and clinical genetic services. Such information helps parents understand the reasons for their child's difficulties and may guide, in a limited way, an understanding of future needs and potential risks. Later in childhood, the nature and extent of a child's learning difficulties and a statement of special educational needs is the main task and later still, the main focus may be the assessment of longer-term social care needs. Throughout life, there may also be questions about a child's or adult's behaviour or mental state or the nature and extent of physical or sensory impairments and disabilities. The role of assessment is essentially to determine need and to inform the types of intervention and treatments, whether educational, medical, psychological, or social, which are likely to be effective and of benefit to the person concerned. Systems of classification provide useful frameworks for such assessments.

Classification

The term 'classification' is unfortunate as it carries with it the stigma associated with previous legislation (e.g. Mental Deficiency Act, 1913) and the associated history of institutionalization consequent upon the eugenics movement at the beginning of the twentieth century.

However, systems of classification are an important way of organizing information and thereby enabling the reliable passing of that information to others and providing a framework to guide intervention. Whilst there are clear strengths to this process, any system of classification has serious limitations. It will tend to focus on a few particular characteristics to the potential exclusion of others, and none can impart a truly comprehensive picture. Methods of classification have inevitably changed over time in an attempt to better clarify the key issues and to minimize stigma that might be associated with any given label. However, the central principle of any system of classification is to bring order to knowledge in a manner that may then enable further advances or the instigation of interventions that previous research has shown to be effective. There is no single universal system—the system of classification used depends on the reasons for its use. These may be as diverse as being predominately administrative or for the purposes of guiding intervention and the use or not of specific treatments.

Classification systems also differ with respect to whether they are dimensional or categorical in nature. Intellectual disability illustrates this difference in that measures such as those obtained from IQ tests are clearly dimensional and continuous whereas labels such as 'intellectual disability' or the identification of particular syndromes are categorical. More recently such obvious categorical distinctions have begun to break down as the genetic basis for syndromes are more clearly elucidated. For example, in fragile X syndrome there is variation in the extent of the number of repeat sequences in the FMR-1 mutation, both within carrier and affected individuals that influence whether or not an intellectual disability is likely to be present.[6] Various different systems of classification are examined below and the relationship between assessment and classification is considered.

Mental retardation (DSM-IV)

From January 1st 2007, the previously named American Association on Mental Retardation changed its name, replacing 'mental retardation' with the term 'intellectual and developmental disabilities'. This followed similar changes in other organizations. However, in DSM-IV[7] the term 'mental retardation' remains for the moment. This standard diagnostic system provides a framework for multiaxial diagnosis with Axis II for personality disorders and mental retardation. Table 10.1.1 summarizes the DSM-IV criteria for mental retardation. The focus is not primarily one of aetiology but rather of quantifying the extent of 'mental retardation' through defining the level of intellectual impairment and listing the range of possible adaptive functions that might be impaired. The definition makes explicit that the onset is in the developmental period and that mental retardation is the final common pathway of a number of potential aetiologies. Significant sub-average intellectual function is defined as an IQ of 70 or below (using standard IQ tests). The IQ is also used to help determine the level of mental retardation (mild, moderate, severe, or profound).

The use of such a multi-axial system recognizes the fact that intellectual disability is a disorder of development, which is separate from other mental disorders, such as mental illness (Axis I), general medical conditions (Axis III), and which may be associated with particular psychosocial and environmental problems (Axis IV). Thus, the process of formulation requires that all these broad domains be considered in arriving at an understanding of an individual's particular difficulties.

International Classification of Functioning, Disability and Health (ICF)

In 2001, the World Health Organization published the International Classification of Functioning, Disabilities and Health (ICF).[8] This is a complete revision of the International Classification of Impairments, Disabilities, and Handicaps.[9] The latter classification was an advance at that time in that it had attempted to overcome the limitations of other methods of classification (particularly with respect to chronic disability) and, most importantly, aimed to guide intervention at several levels and in a more holistic manner than classification systems that were primarily focused on diagnosis, had been able to do. In this context, intellectual disabilities could be conceptualized at different levels. In the case of impairment, the organ system involved is that of the central nervous system. It is the impairment of this system for genetic, chromosomal, or environmental reasons that have primarily affected the acquisition of developmentally determined skills and the ability to learn. The associated disability is the effect of the impairment on a person's ability to learn and acquire new skills that come with development. The exact nature and extent of the disability may not only include the impact of an intellectual disability but also physical and sensory disabilities. The extent to which a given

Table 10.1.1 Summary of the diagnostic criteria for mental retardation (DSM-IV)

A. Significant subaverage general intelligence
B. Significant limitations in adaptive functioning in at least two of the following:
- Communication
- Self-care
- Home living
- Social/interpersonal skills
- Use of community resources
- Self-direction
- Functional academic skills
- Work
- Leisure
- Health
- Safety

C. Onset before age 18 years of age

Note

- Significant subaverage intellectual functioning is defined as an IQ of about 70 or below. The choice of testing instrument should take into account the individual's socio-economic background, native language, and other associated handicaps
- Adaptive functioning refers to how effectively individuals cope with common life demands and how well they meet the standards of personal independence expected of someone in their particular age group, sociocultural background, and community setting. Adaptive behaviour may be influenced by individual and/or environmental factors including the presence or not of additional mental or physical disorders. Information on adaptive behaviour should be gathered from one or more independent sources
- The degree of severity of mental retardation may be specified on the basis of intellectual impairment taking into account other aspects of functioning
  - Mild mental retardation: IQ level 50–55 to approximately 70
  - Moderate mental retardation: IQ level 35–40 to 50–55
  - Severe mental retardation: IQ level 20–25 to 35–40
  - Profound mental retardation: IQ level below 20 or 25

impairment results in a loss of function (disability) may well be influenced by the extent and nature of interventions such as special education, or the correction of hearing loss through the use of a hearing aid. The final level, that of 'handicap', is a consequence of an interaction between the disability and the extent to which support is available or environmental adjustments made. It is a measure of disadvantage that can be ameliorated through, for example, the presence of carers to enable individuals to go out, or environmental modifications (e.g. wheelchair ramps) that diminish the impact of physical disabilities.

The ICF attempts to take classification further and to conceptualize 'disability' within the context of society as a whole, recognizing that everyone can experience disability at one time or other—the stated aim of the ICF is to 'mainstream the experience of disability and recognize it as a universal human experience'. The intent is to encourage those using the ICF to take into account more fully the social aspects of dysfunction and not to see disability as only a medical or biologically determined dysfunction. This means of classification aims to enable the recording of the environmental effects on an individual's functioning. The ICF itself is divided into two Parts. Part 1 is concerned with 'Functioning and Disability' and Part 2 is concerned with 'Contextual Factors'. Each of the two components are expressed in both positive and negative terms in order to emphasize what a person is able to do as well as what he/she is not able to do. The ICF is more complex and more comprehensive than the 1980 WHO system of classification attempting to provide a conceptual framework that forces a much wider understanding by bringing together more comprehensively social and biological models of disability. In doing so, it does what a sound formulation should do, moving from the limitations of a diagnosis to an understanding of the individual within a biological, social and environmental context. This is illustrated in the diagram in Fig. 10.1.1 taken from the ICF.

As with the 1980 'Impairments, Disabilities and Handicaps' means of classification, the ICF is seen as complementing other WHO classification systems, such as the ICD-10.[10] The ICD-10 is focused on disease and the ICF on 'components of health'. The latter, in doing so, provides an appropriate means for characterizing need and for ensuring that people with chronic disabilities have such needs met in the context of their individual human rights and also based on rights established through national legislation. However, as knowledge has increased, for example, about the nature and extent of physical and psychiatric co-morbidity affecting people with intellectual disabilities, so then has the need to use different and other relevant systems of classification increased.

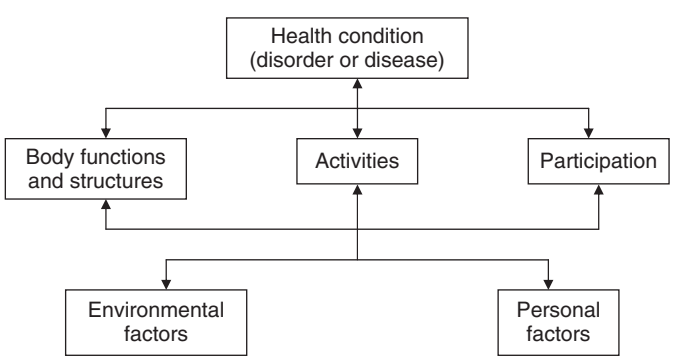


Fig. 10.1.1 Interactions between the components of ICF.

Such different approaches provide the necessary frameworks for a rigorous and comprehensive formulation of a person's needs and to guide treatment of any identified co-morbid illness.

Assessment

Assessment is a task that is undertaken to address specific questions and is informed by the relevant theoretical knowledge and conceptual framework. The type of assessment undertaken therefore, depends on the issue in question and in turn, on both the relevant theoretical background and the appropriate systems of classification, where they are required. The above systems of classification provide a broad framework for considering need. However, in the field of intellectual disabilities a truly holistic assessment will frequently require different theoretical perspectives because of the variability in and complexity of need. The precise form of any focussed assessment will depend on the reasons for undertaking the assessment. The nature of the assessment will vary considerably if it is primarily to determine a person's social care needs, as opposed to the reasons for a particular problematic behaviour. Even with the context of challenging behaviour, the assessments undertaken will vary. Increasingly, the skills expected of those working in community teams supporting people with intellectual disabilities is to be able to recognize what assessments are relevant and required and to be able to undertake such assessments in the community setting where the person with intellectual disability lives. For the sake of clarity, a distinction is made below between those assessments that are fundamentally directed at characterizing the persons, intellectual disability and those whose main focus is on psychiatric and behavioural aspects.

Intellectual disability, its characterization and causes

The term intellectual disability is not in itself a diagnosis as it does not inform in any reliable way about aetiology, prognosis, or specific treatments. Rather, it refers to a clinical state that is developmental in origin and affects intellectual and social functioning. The diagnosis is the identification of the underlying cause for the observed developmental delay. The extent of early developmental delay can be measured against standardized developmental scales (e.g. Bayley or Griffiths Developmental Scales), and during childhood and adult life. There are also well-established specific assessments of intellectual, language, and functional abilities. These assessments provide a profile of a person or group of people that can be compared against the norms for age and a given population. Adaptive functioning has to be measured against what would be expected for a person of that age, and the social and cultural experiences of the person have to be taken into account. The Wechsler Scales for IQ, and the Vineland Adaptive Behaviour Scales[11] or the revised Adaptive Behaviour Scales of the American Association for Intellectual and Developmental Disabilities[12] for characterizing functioning are established instruments for the measurements of these abilities and for which there are normative data for comparison.

Possible single major causes for an intellectual disability are covered in more detail in a separate chapter and are mainly the province of paediatrics or clinical genetics. However, where there is clear evidence for developmental delay in childhood and for the person having an intellectual disability and no obvious cause has been previously established, further investigation is indicated. This will be informed by the clinical history and by physical

examination (e.g. presence of a family history of similar disorder or not, evidence of the physical characteristics of a particular syndrome). As part of this process, the developmental history should also establish whether the developmental profile is not only characteristic of an intellectual disability but also whether the person meets criteria for having an autistic spectrum condition. This is important because of the high rates of autism within the population of people with intellectual disability and the implications such a diagnosis may have for the type of support.[13] Particularly where the person has evidence of dysmorphic physical characteristics, a moderate, severe or profound intellectual disability, and/or a family history of disability, chromosomal, biochemical, and/or molecular genetic studies may be indicated. With improving knowledge about the developmental differences and the different predispositions to specific co-morbidities that exist between genetically determined syndromes, establishing the cause of a person's intellectual disability is becoming increasingly important and may help inform psychiatric and psychological assessments.[14]

(a) Assessments of index problem and psychiatric diagnosis

There have been significant advances in the conceptual models used to help understand the occurrence and maintenance of problem behaviours or abnormal mental states in people with intellectual disabilities. The following four approaches are particularly relevant and each of these perspectives needs to be considered in any assessment. First, applied behavioural analytical studies have demonstrated how the occurrence of behaviours, such as self-injurious behaviour, can be shaped in the context of particular environmental or individual setting conditions.[15] This approach attempts to identify the 'functions' of behaviours. Specific behaviour (e.g. self-injurious behaviour, aggressive outbursts etc.) may, for example, be identified as being attention-maintained or demand avoidant. Secondly, some behaviours may be a consequence of arrested development. In such cases the behaviour itself (e.g. repetitive checking) may be similar to that which occurs as part of typical childhood development but it has continued into adult life. Thirdly, the behaviours may be a manifestation of co-morbid physical or psychiatric disorder. For example, as a consequence of increasing irritability and agitation associated with depression[16] or consequent upon pain or some other physical distress. Fourthly, the abnormal mental state or behaviour may be associated with the cause of the person's developmental disability. This is increasing referred to as 'the behavioural phenotype' of a particular syndrome.[14] These different models of understanding are not necessarily mutually exclusive. For example, the behaviours and psychiatric problems that commonly affect people with Prader Willi syndrome includes examples of each of the above. The increased propensity to temper outbursts and to repetitive and ritualistic behaviours are likely to be partly a consequence of arrested development and to be partly re-inforced and modified depending on environmental contingencies, the over-eating behaviour is a direct consequence of the syndrome and, if obesity is not prevented, it can lead to physical illness that may present with changes in behaviour (e.g. sleep disorders), and those with one genetic form of the syndrome have a high risk for developing co-morbid affective psychotic illness.[17,18] In case of each of these examples, rather different approaches will be required, varying from environmental change to the possible prescription of medication to treat co-morbid psychiatric illness. Assessments should therefore be informed by these different theoretical perspectives recognizing that similar 'behaviours' may have different aetiologies and that what predisposes to, precipitates, or maintains a particular behaviour and/or mental state may each be different. The challenge of assessment is to identify the developmental, biological, psychological, and social factors that relevant and the treatment/management implications.

In psychiatric practice, the referrals and the assessments are usually to determine the reasons for the occurrence of a particular maladaptive/challenging behaviour and/or apparent change in mental state or in cognitive and functional ability. The focus of the assessment is therefore to address these issues but, in doing so, it invariably requires an assessment of the developmental profile of the person concerned and consideration of the cause of any such disability. In principle, however, the psychiatric assessment of people with intellectual disability is not dissimilar to that undertaken in general child or adult psychiatry. The main differences are as follows:

1 Special care may have to be taken in assessing an individual's mental and cognitive state and whether this has changed over time and, where language development is impaired, a greater reliance may have to be placed on information from an informant;

2 a good developmental history is essential to map early development and the potential developmental origins of the individual's present state;

3 it is important to investigate the enabling and constraining aspects of a person's environment that facilitate or impinge on a person's life and thereby how environmental change might bring real benefit;

4 the possibility of multiple physical and mental health problems and the potential for complex interactions between the individual and his or her carers and the immediate and distant environment should be appreciated.

5 where co-morbidity is a possibility, evidence of a change in physical and mental state or behaviour or an exaggeration of previous states (e.g. obsessional behaviours) that might be indicative of the development of a physical or psychiatric illness must be enquired about. The development of physical illness may also manifest as behavioural change.

Psychiatric assessment is therefore invariably an iterative process, especially so with the uncertainties associated with the assessment of a person with an intellectual disability and the different conceptual models. In undertaking an assessment it can be helpful to draw a distinction between:

1 the characterization of the nature and extent of the person's intellectual disability and the identification of its aetiology;

2 the identification of the onset, nature and extent of additional problem behaviours or abnormal mental states;

3 the determination of the possible aetiological factors of the person's behaviour, such as whether there is a co-morbid psychiatric diagnosis or whether there are particular factors that might have predisposed to, precipitated and are now maintaining a particular behaviour or abnormal mental state.

A detailed history taking from both the person him- or herself and an informant covering both childhood development and

the presenting problem, a mental state examination and where indicated a physical examination, direct observation, and often detailed record keeping of particular behaviours (depending on the reason for referral) by care staff over time are the key components of a comprehensive assessment. Where referrals are about the occurrence of problem behaviours, the assessments should include a description of the behaviour as well as an attempt to identify those factors that might increase or decrease the likelihood of the behaviour occurring. These will include the identification of those factors affecting the person him- or herself and those that are particular to the environment. Specifically from a psychiatric perspective, this includes the identification of psychiatric or physical illnesses, the relationship of the index behaviour to any change in mental state or, for example, the occurrence of seizures, as well as the possible contribution of the developmental disability to the problem. For example, the presence of autism may account for observed ritualistic and obsessional behaviours and might also help to make sense of a person's aggressive outbursts that staff observation and record keeping over time has shown occurs at times of unexpected change in routine or activity. Knowledge about the chronicity of a particular behaviour might influence the understanding of the likely contributory factors that give rise to that behaviour. If the behaviour is of recent onset then the role of recent life events or the possibility that the person has developed a mental or physical illness should be investigated. However, if the behaviour has been present since early childhood, psychological models of understanding may be more relevant.

In terms of assessment instruments, a clear distinction needs to be drawn between those that are essentially descriptive in nature and those that are investigating the potential aetiology of the index behaviour. The former include, for example, the Aberrant Behaviour Checklist[19] and the latter, which are based on particular theoretical models, may include, for example, the Psychiatric Assessment Schedule for Adults with Developmental Disorders[20] (a structured assessment of mental state) and the Motivational Assessment Scale[15] (a structured assessment of the possible 'functions' of particular behaviours).

Mental state examination

Central to the practice of psychiatry in particular is the identification of specific mental phenomena that, when clustered together, are indicative of a specific psychiatric disorder. The gold standard for diagnoses are the diagnostic manuals ICD-10 and DSM-IV. Diagnostic criteria have also been modified to better fit what is observed when assessing people with intellectual disabilities. These have been based on the ICD-10, (DC-LD).[21] The process is two-staged and includes first a detailed history and mental state examination that establishes whether there has been a change in a person's mental state and if so, the characteristics of the change, and secondly, a comparison of these changes against diagnostic criteria to establish the presence or not of specific psychiatric disorders. Investigations may be necessary where there is uncertainty or other possible causes for what is observed need to be ruled out as part of the differential diagnosis.

For people with intellectual disabilities with spoken language, disorders such as schizophrenia present in similar ways to the general population, with the onset of the characteristic mental phenomena. [22] Similarly, disorders such as Alzheimer's disease can be diagnosed using very similar criteria, although the early features may be different.[23] In each of these examples, greater emphasis may have to be placed on informant observation, as some people with intellectual disabilities may have difficulty with concepts such as mood and with being able to describe the presence or not of abnormal mental phenomena.

The key to mental state examination and psychiatric diagnosis in people with intellectual disabilities is to be able to characterize the nature of any observed change. The assessment of mood, sleep, appetite, and concentration may be relatively easy as carers observe increasing distress, tearfulness, and agitation over time in the context of sleep and appetite changes, therefore giving rise to the suspicion of the presence of an affective disorder. Similarly, carers may observe personality and memory changes suggestive of dementia. The presence of hallucinations or delusions may have to be inferred by the development of odd behaviour that might reasonably be interpreted as responses to abnormal mental experiences, such as appearing to respond to auditory hallucinations. Family or other carers who have known the person for some time may well be able to describe more subtle mental state abnormalities such as a deterioration in a person's ability to express his or her thoughts indicating the possibility of thought disorder, or increasing perplexity or evidence of paranoid ideas. The interpretation of cognitive findings is more difficult because of the pre-existing intellectual impairments, but clear documentation of cognitive abilities is important as further deterioration or evidence of improvement over time can be very informative.

Needs assessment

A full needs assessment brings together the identified social, emotional, and health needs of an individual, including an understanding of the wishes of the person and views of other people who are concerned for and involved in the support and care of that individual. This whole process has, over the last few years in both mental health and intellectual disability services, become more formalized through 'needs-led assessments' and sometimes through the 'care programme approach'. Person-centred planning is the process by which this information is brought together in a manner that directs the allocation of support and any necessary treatment interventions.[24]

Social services are in general responsible for undertaking assessments at key transitions (e.g. prior to the end of statutory education) or if there is evidence that need has changed. If the person lacks the capacity to express his or her view then the question of what is in the person's best interests should be determined. The balance that this process is attempting to achieve is to respect the wishes of the person concerned and to balance an adult's right to autonomy, on the one hand, versus the need for care and support, on the other. More recently, the critical importance of assessing a person's ability to make decisions about his or her own life and the pivotal role of a person's 'decision-making capacity' in achieving the correct balance between the respect for autonomy and the need for care and support has received more attention. Legislation has been enacted first in Scotland and more recently in England and Wales that provides the framework for intervention where the person concerned lacks the capacity to make the relevant decision for him/herself. The assessment of a person's decision-making capacity requires an evaluation of a person's ability to understand and use information relevant to the decision, retain and balance the necessary information and then to communicate a choice.[25]

Although there is clearly an overlap, it can be helpful to distinguish between social care, educational, and health-care needs. In the case of children, the need for special education and for family support are likely to be central, depending on the extent of the disability and any other associated disabilities. For adults, accommodation with an appropriate level of support, meaningful daytime occupation, and companionship and practical help according to need will be key. Systematic assessments provide a structured way to determine whether a low, medium, or high service environment will best meet an individual's needs. The types and range of services that need to be provided are covered in other chapters. Health-care needs will include the same as those of the general population but high rates of physical and sensory impairments, mental health and behavioural problems, and impairment in communication are all likely to give rise to additional needs. Particularly for those people with more severe intellectual disabilities and limited and/or no language, there is also a clear responsibility for carers to ensure a healthy lifestyle, health screening, and to help identify health-related problems when they do occur. Special help may be required to ensure access to both primary and secondary health-care services and to ensure the person understands what is happening to them. This will include questions of consent to health treatment and how a person with intellectual disabilities can be best helped to give maximally informed consent or at least be able to assent when it is in his or her best interests to do so. A structured process of assessment, paying attention to the cause of the intellectual disability and its associated problems and the extent and nature of other impairments, disabilities, and handicaps will minimize the risk that health-related problems might go unnoticed or that social care needs might be ignored.

Further information

Emerson, E. (2001). *Challenging Behaviour: Analysis and intervention in people with severe intellectual disabilities*. Cambridge University Press, Cambridge.

Bouras, N. and Holt, G (eds.). (2007). *Psychiatric and Behavioural Disorders in Intellectual and Developmental Disabilities*. Cambridge University Press, Cambridge.

References

1. Department of Health. *Valuing People: a new strategy for the 21st Century*, Department of Health, London
2. United Nations. (2007). *Convention on the Rights of People with Disabilities*.
3. Beange, H., McElduff, A., and Baker, W. (1995). Medical disorders of adults with mental retardation: a population study. *American Journal of Mental Retardation*, **99**, 595–604.
4. Cooper S.A., Smiley E., Morrison J., *et al.* (2007). Mental ill-health in adults with intellectual disabilities: prevalence and associated factors. *British Journal of Psychiatry*, **190**, 27–35.
5. Lowe, K., Allen, D., Jones, E., *et al.* (2007). Challenging behaviours: prevalence and topographies. *Journal of Intellectual Disabilities Research*, **51**, 625–36.
6. Verkerk, A., Pieretti, M., Sutcliffe, J.S., *et al.* (1991) Identification of a gene (FMR–1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, **65**, 905–14.
7. American Psychiatric Association (1994). *Diagnostic and statistical classification of diseases and related health problems* (4th edn). American Psychiatric Association, Washington, DC.
8. World Health Organization (2001). *International Classification of Functioning, Disability and Health (ICF)*. World Health Organization, Geneva.
9. World Health Organization (1980). *International classification of impairments, disabilities and handicaps* (10th revision). World Health Organization, Geneva.
10. World Health Organization (1992). *International statistical classification of diseases and related health problems*, (10th revision). WHO, Geneva.
11. Sparrow, S., Ball, D., and Cicchetti, D. (1984). *Vineland Adaptive Behaviour Scale* (Interview Edition–Survey Version). American Guidance Association, Circle Pines, MN.
12. Nihara, K., Leland, H., and Lambert, N. (1993). *Adaptive Behavior Scales–Residential and Community* (2nd edn). Pro–Ed, Austin, TX.
13. Shah, A., Holmes, N., and Wing, L. (1982). Prevalence of autism and related conditions in adults in a mental handicap hospital. *Applied Research in Mental Retardation*, **3**, 303–17.
14. Editorial by Oliver, J..and Hagerman, R. (2007) Trends and challenges in behavioural phenotypes. *JIDR*, **51**, 649–52
15. Durrand, V.M. and Crimmins, D.B. (1988). Identifying the variables maintaining self-injurious behaviour. *Journal of Autism and Developmental Disorders*, **18**, 99–117.
16. Meins, W. (1995). 'Symptoms of major depression in mentally retarded adults' *Journal of Intellectual Disability Research*, **39**, 41–46.
17. Holland, A.J., Whittington, J.E. Butler, J., Webb, T., Boer, H and Clarke, D.J. (2003) Behavioural phenotypes associated with specific genetic disorders: evidence from a population-based study of people with Prader-Willi Syndrome. *Psychological Medicine,* **33**, 141–53
18. Boer, H., Holland, A.J., Whittington, J., Butler, J., Webb, T. and Clarke, D. (2002) Psychotic illness in people with Prader Willi Syndrome due to chromosome 15 maternal uniparental disomy *The Lancet*, **359**, 135–36.
19. Aman, M., *et al.* (1985). The Aberrant Behavior Checklist: a behavior rating scale for the assessment of treatment effects. *American Journal of Mental Deficiency*, **89**, 485–91.
20. Moss, S.C., *et al.* (1993). Psychiatric morbidity in older people with moderate and severe learning disability (mental retardation). Part 1: Developmental and reliability of the patient interview (the PAS-ADD). *British Journal of Psychiatry*, **163**, 471–80.
21. Cooper, S.-A., C. Melville, *et al.* (2003). 'Psychiatric diagnosis, intellectual disabilities and diagnostic criteria for psychiatric disorders for use with adults with learning disabilities/mental retardation (DC-LD)'. *Journal of Intellectual Disability Research*, **47**, 3–15.
22. Meadows, G., Turner, T., Campbell, L., *et al.* (1991). Assessing schizophrenia in adults with mental retardation. A comparative study. *British Journal of Psychiatry*, **158**, 103–5.
23. Ball, S.L., Holland, A.J., Huppert, F.A., *et al.* (2004).The modified CAMDEX informant interview is a valid and reliable tool for use in the diagnosis of dementia in adults with Down's syndrome. *Journal of Intellectual Disability Research*, **48**(6), 611–20.
24. Robertson, J., Emerson, E., Hatton, C. *et al.* (2007). Person-centred planning: factors associated with successful outcomes for people with intellectual disabilities. *Journal of Intellectual Disability Research*, **51**, 232–43.
25. Wong, J.G., Clare, I.C.H., Gunn, M.J., *et al.* (1999). Capacity to make health care decisions: Its importance in clinical practice. *Psychological Medicine*, **29** (2): 437–46.

10.2

Prevalence of intellectual disabilities and epidemiology of mental ill-health in adults with intellectual disabilities

Sally-Ann Cooper and Elita Smiley

Prevalence of intellectual disabilities

If intelligence quotient (IQ) was normally distributed in the population, with a mean of 100 and standard deviation of 15, then about 2 per cent of the population would have an IQ below 70. However, reported rates vary widely depending on the definition of intellectual disabilities used, the country and region of study, the time of the study, the age range and ethnicity of the population, and the method of population ascertainment.[1–3]

Definition

ICD-10 and DSM-IV-TR definitions of intellectual disabilities are similar:

- significantly sub-average intellectual functioning (IQ below approximately 70; mental age less than 12 years);
- concurrent impairments in present adaptive functioning; diminished ability to adapt to the daily demands of the social environment;
- onset before the age of 18 years.

IQ is a continuous measure, so when basing a definition of intellectual disabilities upon it the threshold is arbitrary, and changes in threshold can have a large impact on prevalence. For example, in the past the American Association on Intellectual and Developmental Disabilities' criteria had an IQ threshold of 84. This was subsequently changed to 70 in 1973, and then to 70–75 in 1992. ICD-10 and DSM-IV-TR definitions are not statistical constructions; the requirement of impaired adaptive functioning may half estimated prevalence rates compared with a statistical definition. Furthermore, different methods to assess intelligence and adaptive behaviour can lead to different prevalence rates.

Country and region

Prevalence of intellectual disabilities is reported to be much higher in developing than developed countries, due to socio-economic factors, although mild intellectual disabilities may possibly be less disadvantaging in non-literate societies. Iodine deficiency is the most common preventable cause of intellectual disabilities worldwide, and is indigenous in some regions of Asia and Africa. Exposure of populations to heavy metals and toxins can lower the average population IQ by a few points, hence some people who would otherwise have had low-average ability move into the intellectual disabilities range. Regions with a high level of consanguineous marriages can also have higher prevalence. For example, Tay-Sachs disease is prevalent amongst Ashkenazi communities, although premarital genetic counselling has markedly reduced it in the United States. The availability and sophistication of antenatal, perinatal, and neonatal care account for some differences between countries. Some studies show differences in ethnic groups, although the cultural suitability of the measures used may have contributed to these findings.

Time

Prevalence varies with time, due to preventative measures, and social developments. In developed countries, an increase in prevalence was seen in the early 1960s, with falling prevalence thereafter, due to developments in neonatal care with increasing survival of very low birth weight infants. Down syndrome is the most common chromosomal disorder causing intellectual disabilities, and survival rates of neonates and children with Down syndrome have increased substantially in recent decades, primarily due to access to surgery for congenital heart disease. The widespread introduction of antenatal screening for Down syndrome might have been expected to reduce the population prevalence of Down syndrome through lowering birth rate, but rising maternal age at birth, and increasing life expectation counter this, and there appears to be little change in the population prevalence. The widespread introduction of antenatal screening for phenylketonuria in the 1960s, and congenital hypothyroidism in the 1970s has virtually eliminated intellectual disabilities due to these conditions in

developed countries. Better living conditions, with individualized packages of support, and a political agenda for social justice and equality of access to health care and supports may all have contributed to the increasing life expectation for people with intellectual disabilities, although this is still lower than that of the general population. Increasing maternal age and increasing maternal alcohol consumption are expected to lead to higher birth rate of infants with foetal alcohol syndrome, and genetic causes of intellectual disabilities. Overall, there appears to have been little change in prevalence compared with 50 years ago.

Age

Prevalence is higher in child than adult cohorts, and lower in older than younger adult cohorts, with the highest prevalence at around age 10 years. This is due to intellectual disabilities having been identified by this age, combined with an earlier age of death for persons with intellectual disabilities compared with the general population. Children with the mildest intellectual disabilities are likely to benefit from additional support for learning at school, but will develop skills and experience over time, such that some no longer meet criteria for intellectual disabilities in adulthood.

Ascertainment

Reported prevalence varies with the methods of population ascertainment. For children, the ascertained prevalence doubled compared with case registers, when record linkage to education department data on educational attainments was included (giving an estimated prevalence of 1.4 per cent).[3] There was a disproportional increase in indigenous Australian children, who were possibly false positives. For adult populations, only a proportion with intellectual disabilities will be in contact with specialist health services for adults with intellectual disabilities. Ascertainment is higher when data is combined from primary health care, specialist health services, and social services, if the provision of day opportunities, supported work, respite care, funded support packages, and direct payments is considered. This is likely to identify almost all persons with moderate to profound intellectual disabilities, and adults with mild learning disabilities receiving support; it will not identify adults with IQ below 70 who no longer have impaired adaptive functioning (who therefore do not meet ICD-10 or DSM-IV-TR criteria for intellectual disabilities) and do not need support, or some people who receive all their support exclusively from unpaid carers. The assessment of IQ (culturally sensitive), adaptive functioning and support needs, plus medical assessment, of all individuals within a whole population or a representative sample would provide accurate prevalence data for that time point and area. However, this would be a substantial undertaking, in view of the tens or hundreds of thousands of participants required.

Prevalence

There have been many studies of prevalence of intellectual disabilities. For the reasons above, there is substantial variation in reported prevalence in developed countries, varying from 2 to 85/1000 general population, and there are few robust studies in developing countries. Less variability is found between studies of moderate to profound intellectual disabilities. Given the variation, it is inappropriate to provide average figures from across the studies. Interpreting the literature, we suggest that prevalence of intellectual disabilities in the United Kingdom may be in the order of 9–14/1000 childhood population, and 3–8/1000 adult population, varying with time and geography. However, it should be noted that the figure of 2 per cent is frequently assumed. Intellectual disabilities are more prevalent in males than females, particularly amongst children, young- and middle-aged adults: the reported ratio varies between 1:1 and 2:1. At older age, the gender ratio equalizes due to greater life expectancy of women compared with men (mirroring the general population), and at extreme old age, women may even outnumber men. The distribution of level of intellectual disabilities varies with age, due to the shorter life expectancy of people with more severe intellectual disabilities. Mild intellectual disabilities are associated with socio-economic status. These issues are explored in greater depth elsewhere.[1,2]

Prevalence and incidence of adult mental ill-health

Some genetic causes of intellectual disabilities have specific behavioural phenotypes. For example, Down syndrome confers protection from mania, and problem behaviours, whilst increasing risk for dementia, Prader Willi syndrome is associated with affective psychosis, and velo-cardio-facial syndrome increases risk for psychosis. Behavioural phenotypes are considered in greater depth in Chapter 10.4. In this section we consider mental ill-health of adults with intellectual disabilities of all causes.

Study methodologies

Mental ill-health is thought to be commonly experienced by adults with intellectual disabilities. Many of the existing prevalence studies have methodological limitations, accounting for the wide discrepancy in reported prevalence which ranges from 7 to 97 per cent. Limitations have included biased sampling; reliance upon existing case-note information, or instruments designed as screening tools only; lack of information on the extent of detail within assessments, the instruments, or diagnostic criteria used; and population-based studies limited by small cohort sizes. Other limitations include failure to indicate whether rates are lifetime, point, or period prevalence; reporting combined prevalence for children and adults; reporting mental ill-health in total, but not describing nor being comprehensive as to what is, and what is not, included (particularly with regards to problem behaviours, autistic spectrum disorders, attention-deficit hyperactivity disorder, and anxiety disorders); and studying selected subgroups such as only adults with verbal communication skills. All of these points must be carefully considered when interpreting and drawing conclusions from the existing literature.

Diagnostic criteria

Prevalence of mental ill-health varies, depending upon the diagnostic criteria employed. This is because many of the diagnostic categories within *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research* (DCR), and the DSM-IV-TR contain criteria that cannot be met due to the person's degree of intellectual disabilities and communication skills, and do not include other criteria that are important in this population. For these reasons, *Diagnostic Criteria for Psychiatric Disorders for Use with Adults with Learning Disabilities/Mental Retardation* (DC-LD) was developed for use specifically with this population. These, and

other important diagnostic issues are explored in further depth elsewhere.[4]

Prevalence

Population-based studies where participants received a psychiatric assessment are shown in the Table 10.2.1. [5–9] A high prevalence of mental ill-health was reported in all but Lund's study, which used assessment methods which would today be considered limited.[6] Point prevalence is higher than that observed in the UK general population. Specific types of mental ill-health with a higher prevalence compared with the general population including problem behaviours, autism, dementia, bipolar disorder, and psychoses. Dementia is part of the behavioural phenotype of Down syndrome, and also occurs three to four times more commonly amongst people with intellectual disabilities of other causes.[10,11] Bipolar disorder occurs at about double the prevalence of that reported for the general population. This is despite a high proportion of people (about 25 per cent) taking mood-stabilizing drugs (typically for epilepsy management). Depression is either more prevalent, or occurs at the same rate, depending upon the criteria used. Prevalence of non-affective psychotic disorders (including schizoaffective disorders) has consistently been reported to be higher than for the general population; a recent study found a point prevalence of 4.4 per cent including schizophrenia, in remission, or 4.0 per cent for psychosis, currently in episode. Problem behaviour is the most prevalent type of mental ill-health at 22.5 per cent (Consultant psychiatrist's opinion) or 18.7 per cent (DC-LD) in the most recent of the studies.[9]

Given the numerous biological, psychological, social, and developmental disadvantages experienced by adults with intellectual disabilities compared with the general population, a higher prevalence of mental ill-health is expected. The higher prevalence of bipolar disorder and of psychosis points to biological origins yet to be determined, and likely to be of relevance to the whole general population.

Incidence

There has been limited study of the incidence of mental ill-health in adults with intellectual disabilities. Longitudinal studies in the general population have not included persons with moderate to profound intellectual disabilities, but have demonstrated the higher prevalence of symptoms of depression and anxiety in adults with mild intellectual disabilities compared with the general population.[12,13] A recent study of a population-based cohort of 651 adults with intellectual disabilities found that incidence varied, depending upon the criteria used.[14] According to the Consultant psychiatrists opinion, the 2-year incidence of mental ill-health (of all types, except specific phobias) was 16.3 per cent (12.6 per cent excluding problem behaviours, and 4.6 per cent for problem behaviours), giving a standardized incident ratio for common mental disorders of 1.87 (95 per cent CI = 1.51–2.28). The comparison general population data was broadly similar, but not identical in its classification. The episodes with most frequent incidence were affective disorders at 8.3 per cent, followed by problem behaviours. The incidence of episodes of psychosis over the 2-year period was 1.4 per cent, of which 0.5 per cent were first episodes of psychotic disorders, giving a standardized incidence ratio of 10.0 (95 per cent CI = 2.1–29.3). The rate of full remission from an episode of psychosis within the 2-year period was low, at 14.3 per cent. The authors concluded that the high point prevalence of mental ill-health is explained by both a high incidence and high level of enduring mental ill-health, with slightly more enduring than incident cases.

Protective and vulnerability factors

Many of the factors that might afford protection from or increase vulnerability for mental ill-health are interrelated, for example level of ability, age, gender, and epilepsy. Few studies have attempted to tease these apart. Further work in this area is important, particularly as the pattern of related factors appears to differ from those found in the general population, suggesting that inferences cannot necessarily be drawn from general population data.

Ability

The relationship between ability level and mental ill-health has variously been reported to be absent, present with higher prevalence of mental ill-health at lower ability levels, or present with higher prevalence of mental ill-health at higher ability levels; these differences are explained by the limitations in the previous literature, as described above. Recent reports suggest lower ability is associated with mental ill-health in general, and specifically with

Table 10.2.1 Prevalence of mental ill-health among adults with intellectual disabilities

Study	Sample size	Ability	Diagnostic criteria	Prevalence (%)*
Corbett (1979)[5]	402	Borderline-profound intellectual disabilities	ICD-8	46.3†
Lund (1985)[6]	302	Borderline-profound intellectual disabilities	Modified DSM-III	28.1
Cooper and Bailey (2001)[7]	207	Mild-profound intellectual disabilities	Modified DCR	37.0
Deb *et al.* (2001)[8]	101	Mild-moderate intellectual disabilities	ICD-10	14.4‡
Cooper *et al.* (2007)[9]	1023	Mild-profound intellectual disabilities	Psychiatrists opinion DC-LD DCR DSM-IV-TR	40.9 35.2 16.6 15.7

*Excluding specific phobias.

†Excluding dementia.

‡Excluding problem behaviours, personality disorder, dementia, autism, alcohol problems, schizophrenia not in episode, and bipolar disorder not in episode.

problem behaviours, but not depression nor psychosis; and that lower ability also predicts incident problem behaviours.[6,7,9,14] However, diagnostic complexities may contribute to the lack of association between lower ability and some disorders.

Age

The prevalence and incidence of dementia is higher with increasing age. Whilst it has been suggested that problem behaviours are more prevalent at younger ages, there is no consistent evidence to support this. Of the population-based studies quoted above, no relationship was found between age and mental ill-health.

Gender

Autism and attention-deficit hyperactivity disorder are more common in males. Most studies have not otherwise found any association between gender and mental ill-health in this population, unlike the general population. A higher prevalence of problem behaviours has been reported in women,[8,9] but aggression has also been reported as more common in men[15]; these differences might possibly be due to whether the independent effects of gender and autism are investigated. Women have been reported to score higher than men on the 'affective/neurotic disorders' subdomain of a screening tool,[16] and to have a higher prevalence of mental ill-health,[9] and specifically depression; however, within the same cohort gender did not predict incident mental ill-health.[14]

Epilepsy and other physical ill-health

There are conflicting findings regarding whether there is a relationship between epilepsy and mental ill-health in this population, with the interaction between level of ability and epilepsy and use of antiepileptic drugs possibly contributing. It remains unclear whether physical and mental ill-health are related.

Life events

Associations have been demonstrated between preceding life events in adults with intellectual disabilities and scores for 'affective/neurotic disorders', and between life events and scores on the *Developmental Behaviour Checklist for Adults*.[17] A relationship has been reported with mental ill-health in general, and specifically with depression, but not psychosis, nor problem behaviours, nor incident mental ill-health (excluding problem behaviours),[9,14] but has been reported for incident problem behaviours.[14]

Accommodation/support

Both prevalence and incidence of mental ill-health, and of problem behaviours, is related to living in a setting other than with family carers, and this is independent of past psychiatric history. This highlights the need for engagement between professionals, service managers, and paid carers.

Deprived localities

Unlike for children, no relationship had been found between area-based measures of deprivation and mental ill-health in the adult population with intellectual disabilities, although there have been few investigations of this. It is possible that adults with intellectual disabilities may not have the same lifestyle characteristics as the general population living in the same area, due to being 'placed' in areas dissimilar from those they originated from and within which they acquired life-long habits and preferences, and through ongoing important relationships with family members whose own views and actions may be of greater influence than those of their paid carers or local community. It may be that the biological, social, and developmental causes and consequences of intellectual disabilities far outweigh some of the factors of relevance to the general population, in the aetiology of mental ill-health. This requires further study.

Smoking

As for the general population, smoking is associated with mental ill-health, and specifically with depression and psychosis.

Other factors

Other factors which may be independently related to mental ill-health in this population include urinary incontinence (as in the general population), not being immobile, visual impairment (for psychosis, as for the general population), abuse, neglect and exploitation, and parental divorce during childhood. However, few studies have investigated these factors. Few people with intellectual disabilities use alcohol or cannabis, so it is not known if there is any relationship between these behaviours and mental ill-health in this population.

Conclusions

- The prevalence of intellectual disabilities varies, depending upon definition, country, time, age range, and methods of population ascertainment. Reported rates vary substantially, and may be in the order of 9–14/1000 childhood populations, 3–8/1000 adult populations in developed countries, and higher in developing countries.
- Mental ill-health is more commonly experienced by adults with intellectual disabilities than the general population. Point prevalence is about 40 per cent, with problem behaviours being the most prevalent type.
- Dementia, problem behaviours, autism, bipolar disorder, and psychoses are more prevalent than for the general population.
- Incident mental ill-health is also greater than for the general population, at about 8 per cent per year. Common mental disorders and psychoses both have higher incidence than that for the general population.
- There is limited information on the protective and vulnerability factors for mental ill-health.
- Some factors related to prevalence and incidence of mental ill-health are similar to those found in the general population suggesting similar underlying causative mechanisms, but other factors differ, suggesting that inferences cannot necessarily be drawn from general population data and applied to the population with intellectual disabilities.
- Identifying high-risk groups within the population may allow for the provision of early interventions and supports, whilst some causative factors may be amenable to interventions to prevent or improve mental ill-health in this population. We need to gain a better understanding of these issues.

Further information

NHS Health Scotland. (2004). *Health needs assessment report. People with learning disabilities in Scotland*. NHS Health Scotland, Glasgow.

Royal College of Psychiatrists. (2001). *DC-LD [diagnostic criteria for psychiatric disorders for use with adults with learning disabilities/mental retardation]*. Gaskell Press, London.

Smiley, E. (2005). Epidemiology of mental health problems in adults with learning disability: an update. *Advances in Psychiatric Treatment*, **11**, 214–22.

http://www.intellectualdisability.info/home.htm

References

1. Leonard, H. and Wen, X. (2002). The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Mental Retardation and Developmental Disabilities Research Reviews*, **8**, 117–34.
2. Roeleveld, N., Zielhuis, G.A., and Gabreëls, F. (1997). The prevalence of mental retardation: a critical review of recent literature. *Developmental Medicine and Child Neurology*, **39**, 125–32.
3. Leonard, H., Petterson, B., Bower, C., *et al.* (2003). Prevalence of intellectual disability in Western Australia. *Paediatric and Perinatal Epidemiology*, **17**, 58–67.
4. Cooper, S.-A. and Simpson, N. (2006). Classification and assessment of psychiatric disorders in adults with learning disability. *Psychiatry*, **5**, 306–11.
5. Corbett, J.A. (1979). Psychiatric morbidity and mental retardation. In *Psychiatric illness and mental handicap* (eds. F.E. James and R.P. Snaith), pp.11–25. Gaskell, London.
6. Lund, J. (1985). The prevalence of psychiatric disorder in mentally retarded adults. *Acta Psychiatrica Scandinavica*, **72**, 563–70.
7. Cooper, S.-A. and Bailey, N.M. (2001). Psychiatric disorders amongst adults with learning disabilities—prevalence and relationship to ability level. *Irish Journal of Psychological Medicine*, **18**, 45–53.
8. Deb, S., Thomas, M., and Bright, C. (2001). Mental disorder in adults with intellectual disability I: prevalence of functional psychiatric disorder among a community-based population aged between 16 and 64 years. *Journal of Intellectual Disability Research*, **45**, 495–505.
9. Cooper, S.-A., Smiley, E., Morrison, J., *et al.* (2007). Prevalence of and associations with mental ill-health in adults with intellectual disabilities. *The British Journal of Psychiatry*, **190**, 27–35.
10. Cooper, S.-A. (1997). High prevalence of dementia amongst people with learning disabilities not attributed to Down's syndrome. *Psychological Medicine*, **27**, 609–16.
11. Strydom, A., Livingston, G., King, M., *et al.* (2007). Prevalence of dementia in intellectual disability using different diagnostic criteria. *The British Journal of Psychiatry*, **191**, 150–7.
12. Maughan, B., Collishaw, S., and Pickles, A. (1999). Mild mental retardation: psychosocial functioning in adulthood. *Psychological Medicine*, **29**, 351–66.
13. Richards, M., Maughan, B., Hardy, R., *et al.* (2001). Long-term affective disorder in people with learning disability. *The British Journal of Psychiatry*, **170**, 523–7.
14. Smiley, E., Cooper, S.-A., Finlayson, J., *et al.* (2007). The incidence, and predictors of mental ill-health in adults with intellectual disabilities. Prospective study. *The British Journal of Psychiatry*, **191**, 313–19.
15. McClintock, K., Hall, S., and Oliver, C. (2003). Risk markers associated with challenging behaviour in people with intellectual disabilities: a meta-analytic study. *Journal of Intellectual Disability Research*, **47**, 405–16.
16. Taylor, J.L., Hatton, C., Dixon, L., *et al.* (2004). Screening for psychiatric symptoms: PAS-ADD Checklist norms for adults with intellectual disabilities. *Journal of Intellectual Disability Research*, **48**, 37–41.
17. Hamilton, D., Sutherland, G., and Iacano, T. (2005). Further examination of relationships between life events and psychiatric symptoms in adults with intellectual disability. *Journal of Intellectual Disability Research*, **49**, 839–44.

10.3

Aetiology of intellectual disability: general issues and prevention

Markus Kaski

Causation

The complexity of causes

Intellectual disability can follow any of the biological, environmental, and psychological events that are capable of producing a decline of cognitive functions. Some factors do not directly or inevitably cause intellectual disability but add to the effects of a previous primary cause. Genetic causes may be hereditary or non-hereditary, and may or may not produce specific syndromes. Some lead to inborn errors of metabolism.[1]

Neurological symptoms during the neonatal period are strongly associated with prenatal developmental disturbances. For example in maternal pre-eclampsia, placental insufficiency may lead to malnutrition, foetal asphyxia, intrauterine growth retardation and prematurity, and subsequently to perinatal problems including asphyxia, intracranial haemorrhage, hyperbilirubinaemia, and hypoglycaemia. It is important to detect these coexisting conditions, because their effects may add to or interact with those of the primary cause.[2]

The biomedical cause of intellectual disability may lead to *additional disorders or disabilities*, or may itself be progressive.[1–3] In fact, intellectual disability exists commonly with many other symptoms in a patient, for example with sensory problems, dysphasia, cerebral palsy, epilepsy, or autism. These additional factors affect *opportunities for gaining experiences necessary for development*. Activity may be restricted by sickness, or the effects of medication. Motor disability may reduce mobility, or cause dysphasia. Sensory impairment may restrict vision or hearing. These restrictions add to the effects of the primary cause and interact with environmental and emotional factors to retard the development of the individual.[1–4]

How often do we know the cause(s)?

The various concepts of intellectual disability and its causes have led to *different epidemiological estimates*. The population at risk consists of survivors, but some figures include also all live born. Differences in the definition and detection of cases, the classification system used, the population studied, the timing of the study in relation to measures of the general population, the resources available for the study, and sources of the data make it difficult to compare both the frequency of persons with intellectual disability and the frequency of various causes obtained in different studies. However, a specific cause for intellectual disability can be identified for approximately 80 per cent for persons with severe intellectual disability (IQ <50, includes the groups of moderate, severe, and profound intellectual disability in ICD-10) and 50 per cent of persons with mild intellectual disability. The principal cause of intellectual disability is estimated to be prenatal in 50 to 70 per cent, perinatal in 10 to 20 per cent, and postnatal in 5 to 10 per cent of persons.[3–7] In general, knowledge of aetiologies is more accurate for people with severe intellectual disability than for those with mild intellectual disability.[8, 9]

Classifying causes

The causes of intellectual disability may be classified according to the particular clinical entity, the causative agent or presumed cause, or the *timing of the causative factor*. The newer and more successful classification systems are based on timing.[10] The principle is biomedical in nature and it is intended to elucidate the *earliest factor* that has affected the development of the central nervous system. The discovery of the primary cause (or sequential causes) aids family counselling and may lead to identification of a preventable general risk.

Viewed in this way, diagnoses can be divided into six main groups according to the probable cause and the timing of damage to the central nervous system (Table 10.3.1) [11]:

- genetic causes;
- central nervous system malformations of unknown origin;
- external prenatal factors;
- disorders acquired in the paranatal* period;
- disorders acquired postnatally; and
- untraceable or unclassified causes.

*The paranatal period is defined as the period from 1 week before birth to 4 weeks after birth.

Table 10.3.1 Aetiology based on the time and mechanism of the injury to the central nervous system and the history helping identification, timing, and diagnosis of intellectual disability

Aetiology	History[a]
Genetic causes Chromosomal disorders Malformations due to microdeletion Single-gene disorders Multifactorial intellectual disability Mitochondrial disorders	*Family history* Family tree, intellectual disability, learning disabilities, neurological diseases, congenital anomalies, psychoses, consanguinity of parents, recurrent abortions, previous stillbirths, low parity or infertility, and parental ages
CNS maformations of unknown origin Isolated malformation or malformation sequence of the CNS Multiple malformations	*Family and gestational history*
External prenatal factors Infections Physical, chemical, and toxic agents Maternal and gestational disorders Other	*Gestational history* Maternal infections of the TORCH group, HIV, radiation, trauma, chronic maternal diseases, drugs, pre-eclampsia, severe malnutrition, alcohol, bleeding, and abnormal intrauterine growth or fetal movements
Paranatally acquired disorders Infections Delivery problems Other newborn complications	*Birth and neonatal history* Gestational age, multiple pregnancy, birth order, placental abnormalities, labour or delivery complications/ mode and duration, asphyxia, intracranial haemorrhage, trauma, 5-min Apgar scores, newborn weight, length and head circumference, infections, hypoglycaemia, hyperbilirubinaemia, neurological problems, and weight gain
Postnatally acquired disorders Infections Other damage to CNS Psychosocial problems Psychoses without family history	*Childhood history* Feeding and sleeping patterns, nutrition, growth charts, developmental milestones, infections of CNS, head injuries, submersions, metabolic and endocrine disorders, vascular accidents or thromboses of cerebral veins, cerebral tumours. toxic agents, and psychosocial environment
Untraceable or unclassified causes Pure non-familial With CNS symptoms Not classified	*History of associating conditions* No history of adverse events or signs or history with CNS symptoms such as cerebral palsy, epilepsy, or autism in addition to mental retardation

CNS, central nervous system; TORCH, toxoplasmosis–other infection–rubella–cytomegalovirus–herpes.

[a]Gives the time of the risk event or cause, or appearing time of the sign whose cause(s) may also be earlier.

(a) Genetic causes

Chromosomal disorders include all intellectual disability caused by a proven chromosomal aberration or a clinically obvious chromosomal syndrome such as Down syndrome. However, chromosome analysis should be performed in Down syndrome because translocation, mosaicism, or other abnormalities are found in 5 per cent of cases. Chromosomal anomalies associated with intellectual disability account for up to 40 per cent of severe cases, and 10 to 20 per cent of mild cases.[3,6,12,13] However, the detection rates for chromosome abnormalities with the novel molecular karyotyping methods, such as micro**array**-based **c**omparative **g**enomic **h**ybridization (**array CGH**), range from 5 to 17 per cent in individuals with normal results from prior routine cytogenetic testing.[14] Array CGH has the ability to detect any genomic imbalance including deletions, duplications, aneuploidies, and amplifications.

Malformations due to microdeletion include many malformation syndromes whose causative agent is obscure. A new method of using DNA probes and fluorescence in situ hybridization has increased understanding of the causes of syndromes such as the Angelman, Cornelia de Lange, CATCH 22 (**c**ardiac defects, **a**bnormal face, **t**hymic hypoplasia, **c**left palate, and **h**ypocalcaemia) (velocardiofacial syndrome), Miller–Dieker, Prader–Willi, Rubinstein–Taybi, Smith–Magenis, Sotos, Williams, and Wolf–Hirschhorn syndromes. Parental imprinting modifies the expression of the genes involved in the Prader–Willi and Angelman syndromes.[12, 14–16]

Subtelomeric deletions or **chromosomal rearrangements** have been found in some persons with intellectual disability of hitherto unknown aetiology. Subtelomeric aberrations may explain up to 5 to 10 per cent of previously unknown causes.[13,15]

Single-gene disorders include states with intellectual disability in which the pedigree is highly suggestive of a single-gene origin. Some are caused by a mutant gene with simple Mendelian inheritance. Single-gene mutations may increase or diminish in frequency in areas with long-standing populations of the same origin, or in populations isolated by language or culture. For example, the so-called Finnish *disease heritage* includes 36 disorders, from which 10 manifest with central nervous symptoms and some others may have them.[17] Most of the specific disorders due to mutant gene have characteristic clinical phenotypic features, but there are a considerable number of non-syndromic individuals, especially in early infancy.[12,18,19]

Autosomal dominant inheritance causes tuberous sclerosis, myotonic dystrophy, Gorlin syndrome, neurofibromatosis I, Apert syndrome, Menkes syndrome, and Huntington's disease.

Autosomal recessive inheritance is the cause of most metabolic diseases with intellectual disability. These diseases include phenylketonuria, homocystinuria, maple syrup urine disease, aspartylglucosaminuria, mannosidosis, Salla disease, I-cell disease, mucopolysaccharidoses (except type II), neuronal ceroid lipofuscinoses, Tay–Sachs disease, metachromatic leucodystrophy, Smith-Lemli–Opitz syndrome, and Joubert syndrome.

X-linked inherited disorders include the fragile X, Aicardi, Lesch–Nyhan, Lowe, Norrie, and Coffin–Lowry syndromes, mucopolysaccharidosis II, Duchenne muscular dystrophy, α-thalassaemia intellectual disability syndrome, and Rett syndrome. The most

common intellectual disability syndrome caused by mutation of a single gene is fragile X syndrome. The pattern of its inheritance is X-linked dominant with decreased penetrance.[17,19,20] The prevalence of the 24 other genes identified to date in the X chromosome is low.[20] Dystrophic myotony, fragile X syndrome, and Huntington's disease are caused by so-called *dynamic mutation* in which the length of the repeated sequence of three DNA bases can vary from generation to generation increasing the variability in the phenotype.[21] In Rett syndrome female inactivation of X chromosome may be skewed. It explains the existing of the syndrome in a male or the very mild phenotype in a female.[22] Epigenetic regulatory factors are also involved in the aetiology of Rett syndrome.[23]

Mitochondrial disorders are inherited in most cases due to mutations in the *nuclear genes* encoding proteins targeted to this organelle. Autosomal dominant, recessive, or X-linked inheritances are possible. In addition, mitochondrial dysfunction is shown among others in patients with fragile-X, Rett, and Wolf–Hirschhorn syndromes or autism. [24–27] *Mitochondrial DNA* (mtDNA) is inherited maternally. Sporadic deletions and duplications are also found (Kearns–Sayre syndrome, sporadic deletion or partial duplication in mtDNA). Examples of the maternally inherited (mtDNA) syndromes with central nervous symptoms are the MELAS (mitochondrial myopathy, encephalomyopathy, lactic acidosis, and stroke-like episodes), MERRF (myoclonus epilepsy with ragged red fibres), and NARP (neurogenic muscle weakness, ataxia, retinitis pigmentosa), FSFD (facio-scapulo-femoral muscular dystrophy, familial cerebellar ataxia, recurrent Reye syndrome, cerebral palsy with intellectual disability), and cytochrome c oxidase (COX) deficiency (deafness, myoclonic epilepsy, ataxia, and intellectual disability) syndromes. Nuclear genes are often involved in mitochondrial DNA depletion and Leigh syndromes, which are severe progressive diseases in early childhood.[11, 18, 21]

Multifactorial intellectual disability may be a state of *pure familial* intellectual disability or associated to some *multifactorially inherited* conditions, for example neural-tube defects. One or more first-degree relatives are also affected. Similar pervasive developmental disorders or childhood or other psychoses in one or more of first-degree relatives or otherwise strong family background suggest a polygenic component of intellectual disability.[12, 21]

(b) Central nervous system malformations of unknown origin

Approximately 30 to 40 per cent of all malformations are *genetic* and 10 per cent have *exogenous* causes. The aetiology of the rest is unknown. The number included in this aetiological group decreases with more accurate diagnosis of genetic malformations, intrauterine infections, other teratogenic agents, or deficiencies of essential ingredients needed for the normal development.[28,29] The development of the central nervous system may be disturbed at the following stages[2,12,21,28,29]:

1 dorsal induction at the 3rd to 7th weeks of gestation, leading to anencephaly, encephalocele, meningomyelocele, or other neural-tube-closure defects;

2 ventral induction at the 5th to 6th weeks of gestation, causing prosencephalies and other faciotelencephalic malformations;

3 proliferation of the neurones at the 2nd to 4th month of gestation, leading to microcephaly or macrocephaly;

4 migration of the neurones at the 3rd to 5th month of gestation, causing gyrus anomalies and heterotopias;

5 organization of neurones from the 6th month of gestation to a year postpartum, leading to disturbances in the formation of dendrites and synapses;

6 myelination from 6th month of gestation to a year postpartum, disturbing the proliferation of oligodendrocytes and the formation of the myelin sheets.

The **malformation sequence** is a type of multiple malformation, which includes secondary anomalies caused by an earlier anomaly, for example equinovarus with meningomyelocele. **Multiple malformation syndromes** are caused by the disturbances in blastogenesis or organogenesis. Multiple malformation syndromes of unknown origin include some whose causes are unknown such as the Goldenhar and Kabuki syndromes, and research will show that some of these aetiologically unknown syndromes have a genetic cause.[30,31]

(c) External prenatal factors

The nature of the impairments or malformations, and the severity of resulting intellectual disability appear to relate, at least partially, to the timing of the causative factor as discussed earlier. Also dosage may be important. Effects are most serious when the cause acts *early in embryonic development*; during blastogenesis or organogenesis, when it may result in *multiple malformations*. Effects on the central nervous system of *causes acting later* may be severe even though outward signs may be lacking. These causes include congenital infections such as rubella, cytomegalovirus, herpes simplex type 2, parvovirus, and HIV infection, as well as toxoplasmosis and syphilis. Exposure to medication and other substances such as hydantoin, lipid solvents, alcohol, cocaine, and other drugs can affect the developing foetus.[2,3,28,32,33]

Maternal disorders that may contribute to the causes of intellectual disability include maternal diabetes, arterial hypertension, placental insufficiency, pre-eclampsia, pre- and postmaturity, multiple pregnancy, and foetal growth retardation. In other cases no specific causes can be identified with certainty but available data strongly suggest a prenatal external cause of central nervous system impairment such as exposure to ionizing radiation or trauma.[2, 34]

(d) Disorders acquired in the paranatal period

The effects of the last week of pregnancy extend to the neonatal period and are very important for the outcome of the newborn,[11,35] and combinations of the prenatal and postnatal factors are not rare. *Infections* are transmitted via placenta or the birth canal. They include neonatal septicaemia, pneumonia, meningitis, and encephalitis, which may lead by several mechanisms to neurological deficits, intellectual disability, and sometimes microcephaly, or in bacterial meningitides also to hydrocephalus. *Congenital infections* of herpes simplex and HIV as well as tertiary syphilis may manifest later. *Problems during delivery* may lead to asphyxia, intracranial haemorrhage, or other birth injuries and cause various symptoms of cerebral palsy and epilepsy. *Other newborn complications* include hypoglycaemia, hyperbilirubinaemia, and respiratory distress. Paranatal aetiologies may cause disorders of cognitive functions, as well as motor and sensory impairments.[2,35,36]

(e) Disorders acquired postnatally

Improved postnatal care has reduced the frequency of these causes, which include *infections* such as meningitides and encephalitides.[37] Other causes of postnatal damage to the central nervous system include *toxic agents*, *vascular accidents*, *brain tumours*, *hypoxia*, *and traumas*. Traffic accidents, other traumas, submersions, and cerebral tumours are common causes of disability in childhood. Lead poisoning has been a problem in the United States, iodine deficiency in some regions of the world, and malnutrition almost worldwide.[38] *Psychosocial problems* causing intellectual disability are not as common as was thought in the past, partly because of better identification of medical factors.[39] Severe maternal mental or chronic physical illness, parental alcohol or drug abuse, and some consequences of poverty may be contributory causes leading to inadequate care and stimulation. *Deprived environments* are linked to other risks such as malnutrition, poor medical care, child abuse, usage of alcohol and other substances, and teenage pregnancies.[40]

(f) Untraceable or unclassified causes

The aetiology of intellectual disability can be classified as unknown if the causative factor or timing of the brain damage cannot be established. ***Pure non-familial*** intellectual disability is the term used when there is no family history of intellectual disability and no signs and symptoms suggesting brain damage. It represents the extreme of normal variation. In the untraceable group, the second category is ***intellectual disability of unknown aetiology*** with other symptoms and signs of the central nervous system suggesting brain damage, but with no family history of intellectual disability and no identified malformations or dysmorphic features (see Table 10.3.1). Common examples are intellectual disability associated with cerebral palsy, epilepsy, or autism. Patients should not be assigned to the untraceable group if the diagnostic work-up is incomplete.[11] If so, the aetiology is still ***unclassified***.

How to assess causes

A comprehensive history and a *careful physical examination* are essential for identifying and timing the causative factor(s). In everyday practice it is appropriate to assess the family history, embryological, and postnatal development, possible pathogenetic mechanisms, and the time of the exposure to the supposed agent (Table 10.3.1). The finding of *more than three minor malformations* suggests genetic or early developmental disorder provided that the same dysmorphic features are not found in close relatives.[41] Infants at risk for external prenatal causes, perinatal causes, or postnatally acquired disorders should be examined carefully for dysmorphic features which may indicate alternative or additional cause. The diagnosis usually becomes evident by working out the history and clinical signs. *Databases for analyses of dysmorphology*, symptoms, or other findings are useful aids in the search of an aetiological diagnosis.[42,43]

Ophthalmological and audiological *examinations* can be arranged and other necessary investigations carried out when suggested by the history and physical examination (Table 10.3.2).[2,4,11,12,18,21,37,38,41–43] *If the findings are in accordance with history* and a person has no congenital anomalies, further examinations are seldom needed. However, the possibility of a metabolic disorder should be kept in mind, especially as congenital anomalies or dysmorphic features may occur in people with metabolic disorders. *Metabolic studies* should be performed for every patient with *progressive symptoms*. *If the history and physical findings do not match*, or if there are congenital anomalies or more than three minor dysmorphic features, additional studies are needed.[41] Because more accurate diagnostic methods are being developed, it is useful to keep for each person *a dated chart of examinations* performed.[11]

Prenatal diagnosis may be indicated when there is a known parental balanced translocation, chromosomal aberration of a sibling, a known hereditable disorder in family, a multifactorial disorder such as neural-tube defect in the family, or the mother is elderly.

The gravidity can be detected by *ultrasound examination* at 6th to 8th gestation weeks. Many structural changes can be found from 11th to 15th weeks and confirmed by repeated examinations to 22nd gestation week, for example neural-tube defects. The *nuchal translucency in relation to crown-rump length and gestation week* can be measured during the 9th and 14th for detecting Down syndrome. Other general screening methods for Down syndrome are based on the *applicable markers from the serum of the mother* to the 10th to 14th and 15th to 18th gestation weeks.[44,45] According to the positive screening result or abnormal morphology finding in the ultrasound examination foetal karyotype or some other further examination may be indicated. The age-specific screening of foetal chromosomes is based on the significantly increased probability of a Down syndrome child among mothers aged 35 years or older.[12]

The karyotype of a foetus can be identified after the 10th week of gestation from a *chorionic villus sample*, or after 15th week gestation from the *cells of amniotic fluid*. A known single-gene disorder can also be searched from the chorionic villus sample by *DNA*,

Table 10.3.2 Diagnostic examinations of intellectual disability

- Neurological, ophthalmological, audiological, cardiological, neuropsychological, etc., assessments
- Blood count, vacuolated lymphocytes, and thyroid function
- Antibodies, serology, and urine (TORCH, HIV)
- Radiographs of skull, vertebral column, chest, hands, feet, and long bones, and bone age
- Chromosomes: G-banding, high-resolution banding, and FISH array CGH
- FraX DNA, specific DNA tests, and other molecular genetic techniques
- Blood/urine: amino and organic acids, muchopolysaccharides, oligosaccharides, long- and very-long-chain fatty acids, Astrup, glucose, ammonia, lactate, pyruvate, uric acid, phytanic acid, carnitine, lead, copper, ceruloplasmin
- Fibroblast culture or white blood cell sample; specific enzymes
- Biopsies: muscle, skin, rectal
- Neurophysiological: EEG and evoked potentials
- Neuroimaging: cranial ultrasound, CT, MRI, MRS, functional MRI, SPECT, and PET
- Neuropathological examinations

TORCH, toxoplasmosis–other infection–rubella–cytomegalovirus–herpes; FISH, fluorescence in situ hybridization; MRI, magnetic resonance imaging; SPECT, single-photon emission CT; PET, positron emission tomography; MRS, magnetic resonance spectroscopy.

enzyme, or other specific methods. Cell cultures of few amniotic fluid cells can be used for karyotyping or diagnosing metabolic diseases, but it needs more time. Neural-tube defects, also the small ones, can be seen as elevated levels *of α-fetoprotein* in amniotic fluid. Sometimes a blood sample from the umbilical cord is needed after the 18th gestation week for confirmation of karyotype.

Prenatal diagnostic methods and identification of parental balanced translocations or aberrations in single gene are increasingly available. Microarray-based comparative genomic hybridization (array CGH), examination of foetal DNA or cells derived from maternal blood circulation, and preimplantation diagnoses becoming available for some diseases will change the prenatal diagnosing practice considerably.[12,21,46,47]

Why knowledge of causation is important

Intellectual disability is a confusing concept. The people with intellectual disabilities have more differences than common features. Developmental delay may appear in different ages and with different degrees of severity in different children. The development of a child can come to a stop or can even regress. There is a multitude of confirmed causes of intellectual disability. Single aetiologies are rare and the clinical picture within the same aetiology and between different aetiologies can vary greatly. It is now possible to detect causes that until recently were unknown. Associated disabilities and chronic diseases are common and modify further the complex interplay of individual and environmental factors.

The factors believed to be related to the incidence and prevalence of intellectual disability, such as personal history and gender, the age and the marital status of the parents, the number of siblings, and the living conditions and the social situation of the family, as well as the neighbouring community, vary in persons with intellectual disability. *Attitudes to disabilities* may differ in different families and societies. The permanence of the cognitive impairment is difficult to accept. Insufficient or inadequate information or a prolonged diagnostic process may lead the family or the child to become fearful about the cause of the condition or to try and identify some reason for it. The way is then open for misunderstandings, feelings of guilt, or projections.[4,11,48]

For the person with intellectual disability a confirmed aetiology is the basis of a correct awareness of his or her own disability; the limitations set by the disability and the possibilities for learning and development. The clinical manifestations of some developmental disorders, such as phenylketonuria, galactosaemia, or hypothyroidism can be prevented or arrested by dietary management or hormonal replacement therapy. Knowledge of the prognosis increases awareness of associated disease and disabilities such as sensory impairments, communication disorders, motor and joint problems, epilepsy, and behavioural or psychiatric problems. Thus, aetiology aids the planning of follow-up, rehabilitation, education, and living arrangements.[4,11] Knowledge of aetiology is particularly important at the time of transition from childhood to adult services, helping to ensure continuity of provision and to avoid drop-out.

For the family, knowledge of causes helps to dispel wrong beliefs, self-blame, and anxieties. The parents and siblings may change their preconceived ideas about the disability. It helps the parents to adopt appropriate standards for bringing up their child, and for life as an adult. It helps them to become aware of the child's special needs.[4,11] Aetiologic diagnosis is the necessary basis of reliable genetic counselling and helps the parents and siblings in family planning.[12,48]

In society the knowledge of the aetiologies of intellectual disability increases the likelihood that its people will adopt positive attitudes towards the disabled. Both the society and its service providers need understanding of the causes of intellectual disability, their prevalence, and their prognoses when planning primary prevention, organizing services and education, optimizing environmental factors, or preparing relevant legislation. Society needs experts continuously alert to advances in scientific research to keep this knowledge up to date.

When the causes are unknown, prognosis is uncertain and the planning and provision of the services is difficult. The risks of discontinuities in service provision and of drop-outs increase. Because families have limited information they are more likely to develop wrong beliefs, self-blame, and projections. They have unrealistic expectations about alternative therapies.[12, 48]

Prevention

Primary prevention

The identification of factors that contribute to intellectual disability, their removal or avoidance, and the protection of the population or individuals against them are the main principles of primary prevention. Immunization and other measures to prevent rhesus incompatibility, congenital rubella, measles encephalitis, tuberculosis and other bacterial meningitides, prionic diseases, and the provision of folic acid around the time of conception to prevent neural-tube defects have been successful preventive measures.[2,37]

Primary prevention includes *good medical follow-up, the identification and prompt removal of, or effective intervention in*, at-risk situations during pregnancy, delivery, the neonatal period, and childhood. Avoidable causes include intrauterine and perinatal infections due to many sexually transmitted diseases, and fetal alcohol spectrum disorders (FASD).[49] Lead intoxication, iodine deficiency, accidents in the home, and traffic accidents are preventable.[2,36,38] The primary prevention of genetic disorders has not been possible, although the disorder, the chromosomal aberration, or some associated abnormality can sometimes be identified.[21,29,44] However, all the means of the secondary prevention are available, and pregnancies may be terminated after counselling. Screening tests can identify some parents who are carriers.[48] The preimplantation diagnosis could give a new way for the early prevention of known genetic disorders in some cases.[47,50]

Secondary prevention

Early recognition and diagnosis, good medical care, and rehabilitation of injuries or diseases can avoid or reduce permanent damage which could lead to intellectual disability. Examples include the screening for and early treatment of congenital hypothyroidism and phenylketonuria.[2] Secondary prevention also includes planning or *genetic counselling* after the birth of a child with a genetic disorder.[48] The examination of asymptomatic parents and close relatives in order to detect carriers is part of genetic counselling. As yet there are no measures to prevent the underlying biological processes in genetic disorders.[12,18,21]

Once children with intellectual disability have been identified, accurate assessment of aetiology and associated conditions, therapy,

and rehabilitation lessen the risk of so-called **caused learning disability**, that is leaving the person with intellectual disability without the possibility of developing further because he or she is unable to share in the learning experiences of the peer group.[2,4,38]

Tertiary prevention

The aim of tertiary prevention is *to help an individual attain his or her full developmental potential.* Therefore tertiary prevention partly overlaps with secondary prevention. It encompasses all measures, which prevent or lessen persistent hindrances to the development of functional ability or social competence in people with identified damage or disease. It includes medical, psychological, social, and family support, environmental adjustments, aids, and education. In the ideal state, individual supports are so good that the mentally retarded can live a life similar to that of people without intellectual disability. However, even when this can be achieved, problems may reappear when tasks of life increase, or the supporting systems fail.[4,10]

Ethical problems of prevention

Usually there are no ethical problems. There are hardly any issues surrounding the prevention of causative diseases or injury, the attempt to reduce the prevalence of inheritable diseases by family planning or genetic counselling, the screening that leads to the treatment of the foetus or the newborn, preparation for the birth of a new infant, improving the care and rehabilitation, and planning a safer society.

Ethical questions arise if there are no measures to prevent the effects of an identified genetic or other cause of intellectual disability identified during pregnancy. Disordered genes, other genetic rearrangements, or accidental mutations cannot be removed from a population. Primary prevention of an intellectual disability becomes the prevention of childbirth or selective abortion. The latter is usually on the grounds that the burden of caring for the child would be too great for the parents, an indication that is more social than medical.[51,52] In general, the prevention of a disease does not exclude good therapy for it. However, the increasing trend to accept selective means of prevention may be a result of cultural change and increasingly negative attitudes towards disability.[53–56]

Prenatal diagnosis, detection of carriers and screening for disorders, which cannot be prevented or treated, usually lead to decisions about more or less harmful consequences. Ethical questions include *choices between knowing and not knowing*, between not knowing and worry or anxiety in risk situations, and between the need to take difficult decisions and to avoid them. It is very difficult to predict accurately all the long-term effects of these decisions. Parents have the right to be able to make voluntary decisions, but they need *accurate information* to be able to bear the burden of responsibility.[51,56–58]

Prevention requires a coordinated programme

On the basis of the aetiologies defined according to the timing principle, it is possible to search for epidemiological, psychological, sociological, and other explanations to answer the question of why some causes are more common in certain populations or subgroups than in others. *The starting point of prevention is knowledge of causation* and identification of factors that subject individuals to these causes. With better understanding of the causes of intellectual disability, attitudes towards it change. With knowledge, individuals needing support can be identified earlier, their development and life made easier, and the burden of care on the family lightened.

Public policy, education, public health, obstetric services, neonatal intensive care, general practice services, etc. have an important role in reducing and avoiding risk factors. Education, planning of a safe environment and alleviation of poverty has a general preventive effect on predisposing factors. *The preventive aspects should be taken into account in all general and specific legislation, in operating procedures, and professional practice*. Because multiple agencies are involved preventive measures need to be coordinated at both local and national levels.

The day-to-day prevention of infections, accidents in the home and traffic, exposure to toxic substances, drowning accidents, malnutrition, or child abuse is both *a general and a multiprofessional task*. Parents and educators should transfer their wisdom and experience to the next generation so that they will make better choices and avoid risks.

Further information

Accardo, P.J. (ed.) (2008). *Capute and Accardo's neurodevelopmental disabilities in infancy and childhood.* (Vol. 1–2, 3rd edn). Paul H. Brookes Publishing Co., Inc., Baltimore, MD.

Smith, M., Arfin, S.M., and Lott, I.T. (2006). *Mental retardation and developmental delay: genetic and epigenetic factors*. Oxford University Press, New York.

Cassidy, S.B. and Allanson, J.E. (2005). *Management of genetic syndromes* (2nd edn). Wiley & Liss, New York.

Fernandes, J., Saudubray, J.M., van den Berghe, G., *et al.* (eds.) (2006). *Inborn metabolic diseases: diagnosis and treatment* (4th edn). Springer, New York.

References

1. Filiano, J.J. (2006). Neurometabolic diseases in the newborn**.** *Clinics in Perinatology*, **33**, 411–79.
2. Volpe, J.J. (2001). *Neurology of the newborn* (4th edn). W.B. Saunders, Philadelphia, PA.
3. Heikura, U., Linna, S.-L., Olsen, P., *et al.* (2005). Etiological survey on intellectual disability in the northern Finland birth cohort 1986. *American Journal of Mental Retardation*, **110**, 171–80.
4. Rubin, I.L.M.D. and Crocker, A.C. (2006). *Medical care for children & adults with developmental disabilities* (2nd edn). Brookes Publishing Company, Baltimore, MD.
5. Ahuja, A.S., Thapar, A., and Owen, M.J. (2005). Genetics of mental retardation. *Indian Journal of Medical Sciences*, **59**, 407–17.
6. Hou, J.-W., Wang, T.-R., and Chuang, S.-M. (1998). An epidemiological study of children with intellectual disability in Taiwan. *Journal of Intellectual Disability Research*, **42**, 137–43.
7. Stromme, P. (2000). Aetiology in severe and mild mental retardation**:** a population-based study of Norwegian children. *Developmental Medicine and Child Neurology*, **42**, 76–86.
8. Heikura, U., Taanila, A., Olsen, P., *et al.* (2003). Temporal changes in incidence and prevalence of intellectual disability between two birth cohorts in Northern Finland. *American Journal of Mental Retardation*, **108**, 19–31.
9. Roeleveld, N., Zielhuis, G.A., and Gabreëls, F. (1997). The prevalence of mental retardation: a critical review of recent literature. *Developmental Medicine and Child Neurology*, **39**, 125–32.
10. American Association on Mental Retardation. (2002). *Mental retardation: definition, classification, and systems of supports* (10th edn).

American Association on Intellectual and Developmental Disabilities, Washington, D.C.
11. Wilska, M. and Kaski, M. (2001). Why and how to assess the aetiological diagnosis of children with intellectual disability/mental retardation and other neurodevelopmental disorders: description of the Finnish approach. *European Journal of Paediatric Neurology*, **5**, 7–13.
12. Smith, M., Arfin, S.M., and Lott, I.T. (2006). *Mental retardation and developmental delay: genetic and epigenetic factors*. Oxford University Press, New York.
13. Rooms, L., *et al.* (2006). Multiplex ligation-dependent probe amplification to detect subtelomeric rearrangements in routine diagnostics. *Clinical Genetics*, **69**, 58–64.
14. Shaffer, L.G. and Bejjani, B.A. (2006). Medical applications of array CGH and the transformation of clinical cytogenetics. *Cytogenetic and Genome Research*, **115**, 303–9.
15. Hodgson, S.V. (1998). The genetics of learning disabilities. *Developmental Medicine and Child Neurology*, **40**, 137–40.
16. Tatton-Brown, K., *et al.* (2005). Genotype-phenotype associations in Sotos syndrome: an analysis of 266 individuals with *NSD1* aberrations. *American Journal of Human Genetics*, **77**, 193–204.
17. Norio, R. (2003). The Finnish disease heritage III: the individual diseases. *Human Genetics*, **111**, 470–526.
18. Scriver, C.R., Beaudet, A.L., Childs, B., *et al.* (2000). *The metabolic and molecular base of inherited disease* (8th edn). (www.ommbid.com). McGraw-Hill, New York.
19. OMIM (Online Mendelian Inheritance in Man). http://www.ncbi.nlm.nih.gov/omim/
20. Raymond, F.L. (2006). X linked mental retardation: a clinical guide. *Journal of Medical Genetics*, **43**, 193–200.
21. Strachan, T. and Read, A.P. (2004). *Human molecular genetics* (3rd edn). Garland Science, London.
22. Huppke, P., Maier, E.M., Warnke, A., *et al.* (2006). Very mild cases of Rett syndrome with skewed X inactivation. *Journal of Medical Genetics*, **43**, 814–16.
23. Petel-Galil, Y., *et al.* (2006). Comprehensive diagnosis of Rett's syndrome relying on genetic, epigenetic and expression evidence of deficiency of the methyl-CpG-binding protein 2 gene: study of a cohort of Israeli patients. *Journal of Medical Genetics*, **43**, e56.
24. Rizzo, G., *et al.* (2006). A case of fragile X premutation tremor/ataxia syndrome with evidence of mitochondrial dysfunction. *Movement Disorders*, **21**, 1541–2.
25. Wenk, G.L. (1997). Rett syndrome: neurobiological changes underlying specific symptoms. *Progress in Neurobiology*, **51**, 383–91.
26. Schlickum, S., *et al.* (2004). LETM1, a gene deleted in Wolf-Hirschhorn syndrome, encodes an evolutionarily conserved mitochondrial protein. *Genomics*, **83**, 254–61.
27. Oliveira, G., *et al.* (2005). Mitochondrial dysfunction in autism spectrum disorders: a population-based study. *Developmental Medicine and Child Neurology*, **47**, 185–9.
28. Stevenson, R.E. and Hall, J.G. (eds.) (2005). *Human malformations and related anomalies* (2nd edn). Oxford University Press, New York.
29. Moore, K.L. and Persaud, T.V.N. (2007). *Before we are born: essentials of embryology and birth defects* (7th edn). W.B. Saunders, Philadelphia, PA.
30. Touliatou, V., Fryssira, H., Mavrou, A., *et al.*(2006). Clinical manifestations in 17 Greek patients with Goldenhar syndrome. *Genetic counseling*, **17**, 359–70.
31. Miyake, N., *et al.* (2006). No detectable genomic aberrations by BAC array CGH in Kabuki make-up syndrome patients. *American Journal of Medical Genetics*, **140**, 291–3.
32. Konstantinidou, A.E., Syridou, G., Spanakis, N., *et al.* (2007). Association of hypospadias and cardiac defect in a parvovirus B19-infected stillborn: a causality relation? *The Journal of Infection*, **54**, 41–5.
33. Sharp, N.J., Davis, B.J., Guy, J.S., *et al.* (1999). Hydranencephaly and cerebellar hypoplasia in two kittens attributed to intrauterine parvovirus infection. *Journal of Comparative Pathology*, **121**, 39–53.
34. Badawi, N., Kurinczuk, J.J., Keogh, J.M., *et al.* (1998). Antepartum risk factors for newborn encephalopathy: the Western Australian case–control study. *British Medical Journal*, **317**, 1549–53.
35. Badawi, N., Kurinczuk, J.J., Keogh, J.M., *et al.* (1998). Intrapartum risk factors for newborn encephalopathy: the Western Australian case–control study. *British Medical Journal*, **317**, 1554–8.
36. Stevenson, D.K., Benitz, W.E., and Sunsine, P. (2003). *Fetal and neonatal brain injury: mechanisms, management and the risks of practice* (3rd edn). Cambridge University Press, Cambridge.
37. Scheld, W.M., Whitley, R.J., and Marra C.M. (eds.) (2004). *Infections of the central nervous system* (3rd edn). Lippincott Williams and Wilkins, Philadelphia, PA.
38. David, R.B. (ed.) (2005). *Child and adolescent neurology* (2nd edn). Blackwell Publishing Limited, Massachusetts, MA.
39. Bosma, H., van de Mheen, H.D., and Mackenbach, J.P. (1999). Social class in childhood and general health in adulthood: questionnaire study of contribution of psychological attributes. *British Medical Journal*, **318**, 18–22.
40. Malamitsi-Puchner, A. and Boutsikou, T. (2006). Adolescent pregnancy and perinatal outcome. *Pediatric Endocrinology Reviews*, **3**(Suppl. 1), 170–1.
41. Jones, K.L. (ed.) (2006). *Smith's recognizable patterns of human malformations*. Elsevier Saunders, Philadelphia, PA.
42. POSSUM Database, version 5.7.3. (2006). *Pictures of standard syndromes and undiagnosed malformations*. Murdoch Children's Research Institute, Victoria, http://www.possum.net.au.
43. London Medical Databases. (2006). *The Baraitser-winter neurogenetics database, and the winter-Baraitser dysmorphology database* (www.lmdatabases.com). London Medical Databases Ltd., London.
44. Rosen, T. and D'Alton, M.E. (2005). Down syndrome screening in the first and second trimesters: what do the data show? *Seminars in Perinatology*, **29**, 367–75.
45. Canick, J.A., *et al.* (2006). First and Second Trimester Evaluation of Risk (FASTER) trial research consortium. Comparison of serum markers in first-trimester down syndrome screening. *Obstetrics and Gynecology*, **108**, 1192–9.
46. Bischoff, F.Z., *et al.* (2002). Cell-free fetal DNA and intact fetal cells in maternal blood circulation: implications for first and second trimester non-invasive prenatal diagnosis. *Human Reproduction Update*, **8**, 493–500.
47. Fasouliotis, S.J. and Schenker, J.Y. (1998). Preimplantation genetic diagnosis: principles and ethics. *Human Reproduction*, **13**, 2238–45.
48. Harper, P.S. (2004). *Practical genetic counselling* (6th edn). Arnold, Hodder Headline Group, London.
49. Harris, L.H. (2000). Rethinking maternal-fetal conflict: gender and equality in perinatal ethics. *Obstetrics and Gynecology*, **96**, 786–91.
50. Katz, M.G., Fitzgerald, L., Bankier, A., *et al.* (2002). Issues and concerns of couples presenting for preimplantation genetic diagnosis (PGD). *Prenatal Diagnosis*, **22**, 1117–22.
51. Louhiala, P. (2004). *Preventing intellectual disability: ethical and clinical issues*. Cambridge University Press, Cambridge, UK.
52. Alderson, P. (2001). Down's syndrome: cost, quality and value of life. *Social Science & Medicine*, **53**, 627–38.
53. Bromage, D.I. (2006). Prenatal diagnosis and selective abortion: a result of the cultural turn? *Medical Humanities*, **32**, 38–42.
54. Raz, A. (2004). "Important to test, important to support": attitudes toward disability rights and prenatal diagnosis among leaders

of support groups for genetic disorders in Israel. *Social Science & Medicine*, **59**, 1857–66.

55. Kaplan, D. (1993). Prenatal screening and its impact on persons with disabilities. *Fetal Diagnosis and Therapy*, **8**(Suppl. 1), 64–9.
56. Hall, S., Bobrow, M., and Marteau, T.M. (2000). Psychological consequences for parents of false negative results on prenatal screening for Down's syndrome: retrospective interview study. *British Medical Journal*, **320**, 407–12.
57. Searle, J. (1997). Routine antenatal screening: not a case of informed choice. *Australian and New Zealand Journal of Public Health*, **21**, 268–74.
58. van den Berg, M., Timmermans, D.R., ten Kate, L.P., *et al.* (2005). Informed decision making in the context of prenatal screening. *Patient Education and Counseling*, **63**, 110–17.

10.4

Syndromes causing intellectual disability

David M. Clarke and Shoumitro Deb

Introduction

Psychiatrists working with people who have intellectual disability (mental retardation) need expertise in the diagnosis and treatment of associated neuropsychiatric disorders. This entails knowledge of the causes of intellectual disability, and especially knowledge about those syndromal (often genetic) causes that are associated with neuropsychiatric manifestations. Such manifestations include vulnerability to behavioural and emotional disorders, epilepsy, and particular patterns of cognitive strength and weakness. This chapter provides an introduction to some such disorders and a discussion of the concept of behavioural phenotypes. For a detailed account of conditions causing intellectual disability texts such as Jones[1] should be consulted. The concept of behavioural phenotypes is discussed in detail in O'Brien.[2]

The genetic aetiologies of intellectual disability include chromosomal abnormalities (trisomy, deletion, translocation, etc), single-gene defects, and the effect of interactions between several genes. The last is thought to account for a substantial proportion of people with mild intellectual disability by setting a ceiling on possible cognitive attainment (life experiences, nutrition, education, and other factors then determining the extent to which potential is fulfilled or thwarted).

This chapter discusses the concepts of syndromes and behavioural phenotypes, then describes the clinical features of a number of syndromes that cause intellectual disability. Down syndrome, fragile-X syndrome, sex chromosome anomalies, and foetal alcohol syndrome are described in some detail. This is followed by a briefer alphabetical list of less common conditions.

Syndromes

A syndrome is a characteristic pattern of clinical features, including signs (that can be observed) and symptoms a patient may experience. They may be causes of intellectual disability (Down syndrome), associated with intellectual disability (syndromes of epilepsy, such as West syndrome), or coincidental (polycystic ovary syndrome). This chapter deals with some of the syndromes that increase vulnerability to intellectual disability. The vulnerability may be increased so much that all affected people have intellectual disability (Angelman syndrome) or increased to the extent that many, but not all, have intellectual disability (velo-cardio-facial syndrome).

There can be disadvantages to the labelling of people with disability, but the identification of a syndromal cause may have benefits for the affected person and for their families and carers. Benefits include an explanation of the cause of the person's disabilities or of unusual cognitive strengths and weaknesses, better understanding of risk of recurrence of the disorder among relatives, and the identification of complications or associated features.

Identification of a syndromal cause may give access to support organizations. A list of such organizations is given in the CaF (Contact a Family) Directory (www.cafamily.org.uk).

Behavioural phenotypes

Behavioural, social, linguistic, or cognitive aspects of a syndrome may be so striking and characteristic as to prompt diagnosis. Examples include the severe self-injury associated with Lesch–Nyhan syndrome and the combination of appetite abnormality, ritualistic behaviours, sleep abnormalities, skin-picking, repetitive speech, and vulnerability to psychiatric disorder associated with Prader–Willi syndrome. Such patterns of vulnerability to particular emotional or behavioural problems or peculiarities associated with biologically determined syndromes have been called behavioural phenotypes. Environmental factors may interact with a genetically determined vulnerability to a behaviour to determine whether or not it occurs in a given setting. Knowledge of the nature of this interaction may be important in order to determine effective treatment or management strategies. In Lesch–Nyhan disease, for example, all affected men self-injure, but whether a man with the syndrome injures himself at a particular time is influenced by environmental and internal psychological factors such as anxiety. A careful assessment of the causes and consequences of behavioural problems is essential before interventions are planned, particularly the use of psychotropic medication to influence behaviour.[3]

Specific conditions

Down syndrome

(a) Prevalence and genetics

J. Langdon Down originally described the syndrome in 1887. Trisomy 21 is associated with Down syndrome, and was first reported by Lejeune and colleagues in 1958. About 1 in 600 live

born children have Down syndrome. The rate increases with increasing maternal age, being about 1 in 2000 at maternal age 20 years and 1 in 100 at maternal age of 40 years.[4] There are three types of abnormalities affecting chromosome 21. In about 94 per cent of cases, Down syndrome is caused by primary non-disjunction leading to trisomy 21. The risk of recurrence of this abnormality is low if maternal age is also relatively low. In about 2 per cent of cases Down syndrome results from an unbalanced translocation (when material from one chromosome is separated and attached to another with some duplication). This often involves chromosomes 21 and 14. In some cases a parent also has a balanced translocation (with no overall disruption or duplication of genetic material), and this raises the risk of recurrence. Chromosome 21 to 21 translocations can occur. Mosaicism occurs when there are two or more cell lines within the body. In Down syndrome there may be one cell line with trisomy 21 and one without. In about 2 per cent of cases the syndrome results from mosaicism. Some cases may not be diagnosed. The proportion of affected and unaffected cell lines varies, as does the intellectual impairment.

(b) Physical characteristics

Muscular hypotonia at birth usually improves with development. Most adults are of short stature and have a characteristic facial appearance. The eyes seem to slope upwards and outwards, the nose has a wide bridge and the head has an unusual shape (brachycephaly). Limb abnormalities include a single transverse crease on the palm, a large cleft between the first and second toes, and relatively short upper arms. People with Down syndrome are prone to thyroid abnormalities. About 25 per cent develop hypothyroidism during childhood or adolescence. About half of affected people have a heart abnormality. Abnormalities of the gastro-intestinal tract occur in a significant minority. Life expectancy has improved markedly over the past 50 years. Survival into the eighth decade is unusual but not extraordinary. Changes in blood cells are relatively common. Older texts reported an association between Down syndrome and leukaemia, but recent research suggests that leukaemia is rare, affecting less than 1 per cent of people with Down syndrome.

(c) Behavioural and psychiatric aspects

Adults with Down syndrome are much more likely to develop dementia than the general population. On post-mortem examination, the brains of almost all adults with Down syndrome over the age of 35 show changes characteristic of dementia of Alzheimer type. Only about 38 per cent of those aged 50 to 59 have clinically apparent dementia, with a mean age at diagnosis around 51 years.[5,6]

The stereotype of people with Down syndrome as happy, placid individuals with a gift for mimicry is not borne out by recent behavioural research. Stubbornness and obsessional features seem to be over-represented, and many people with Down syndrome react adversely in situations involving changes to expected routines or conflict. Autism seems to occur more commonly than would be expected, but few methodologically sound studies have been carried out.[7]

Most adults with Down syndrome have moderate intellectual disability. Almost all children with Down syndrome have some degree of specific speech and language delay. About 25 per cent have features of attention-deficit disorder. Cognitive abilities tend to be greater among people whose Down syndrome is caused by mosaicism for trisomy 21.

Further information: www.downs-syndrome.org.uk

X-linked intellectual disability

The prevalence of X-linked intellectual disability is around 0.18 per cent.[8] The majority of affected men have non-syndromic X-linked intellectual disability (usually referred to as X-linked mental retardation or XLMR in international literature), with no associated dysmorphology. The most common syndrome resulting in XLMR is fragile-X syndrome (described below). Coffin–Lowry syndrome (CLS) is also described below. It is increasingly accepted that there is a spectrum of disorders associated with XLMR genes, ranging from defined syndromes such as CLS to XLMR with no dysmorphology. For example, the gene RSK2 is usually mutated in Coffin–Lowry syndrome but a missense mutation in exon 14 of RSK2 has been found in a family in which males have intellectual disability but no associated features of CLS.[9] An interesting article described a woman with mild intellectual disability, epilepsy, and some minor dysmorphology whose karyotype was reported as normal in 1993. Repeated testing was carried out after she was found to have a more severely affected brother with a duplication affecting his X chromosome showed 46,X dup (X)(p22.13p22.31). The authors concluded that genetic testing for individuals with intellectual disability should be considered even when there was a low index of suspicion for an X-linked disorder.[10] About 200 XLMR conditions and 45 cloned genes have now been described.[11] At least eight genes have so far been implicated in non-specific XLMR: *Rab-GDI*, *PAK3*, *AGTR2*, *TM4SF2*, FRAXE (*FMR2*), *ARHGEF6*/*αPIX*, and *FACL4*.[12] Readers are referred to specialized texts such as Jaquemont *et al.* (2005)[12] and web resources such as xlmr.interfree.it/home for further details.

Fragile-X syndrome

(a) Prevalence and genetics

The syndrome was first described in 1943. All ethnic groups are affected equally, with a frequency of about 0.3 per 1000 in men. More recent investigations with modern diagnostic techniques show lower figures than earlier studies.[13]

Fragile-X syndrome is an X-linked disorder with a very unusual pattern of inheritance. It is characterized by a bias to affected men but with some affected women and some unaffected men who have daughters who then have affected sons. When peripheral blood lymphocytes from affected individuals are grown in certain culture conditions, including a lack of folic acid, a fragile site becomes evident on the long (q) arm of the X chromosome at Xq27.3 (fragile site A). Fragile sites may not be seen in some unaffected men who transmit the abnormality to their carrier daughters. These men were historically termed 'normal transmitting males'. The probability that a child with a fragile-X chromosome will have intellectual disability depends on the sex of the parent from whom the chromosome was inherited (higher risk when the chromosome is passed from the mother). The 'fragility' of the X chromosome is now known to be associated with an unstable region of DNA within the fragile-X mental retardation (FMR-1) gene, which was first described in 1991.[14] This region of unstable DNA gradually increases in length and degree of instability in successive generations (a pre-mutation) until a critical point is reached and the gene no longer functions (a full mutation). The instability is caused by an increase in CGG (cytosine-guanine-guanine) repeats from the 50 or so repeats that are usual to 50–100 repeats (pre-mutation)

to over 230 repeats (full mutation). The chance of a child inheriting a lengthened gene is proportional to the length of the unstable region in the carrier mother. The severity of intellectual disability and other fragile-X related phenomena in women probably depends mostly on the proportion of cells in which the abnormal chromosome is inactivated, X inactivation being random. Most women who have children with fragile-X syndrome are premutation carriers of normal intelligence.[15] Carriers of the premutation are intellectually unimpaired but are more vulnerable than other women to anxiety and depression.[16] Variants of fragile-X syndrome have now been identified, with DNA expansions nearer to the end of the long arm of the X chromosome. These include FraX-E and FraX-F.

(b) Physical characteristics

Physical features are variable. The most characteristic feature is that about 95 per cent of affected men have large testes, although macro-orchidism is not usually apparent until after puberty. Other features include a long face with a large forehead, large ears, a large lower jaw, and high-arched palate. There is a connective tissue disorder that may lead to tissue laxity with hyper-extensible joints, flat feet, heart defects (especially valve abnormalities), and ear infections (the eustachian tube closes easily). Cataracts and other eye abnormalities may occur, and lead to impaired vision. About 30 per cent of affected men have epilepsy. Life expectancy depends on the severity of associated features such as epilepsy and cardiovascular anomalies.

(c) Behavioural and psychiatric aspects

There is usually some degree of social impairment, with social anxiety and avoidance of eye-to-eye contact, but with social responsiveness. Men with fragile-X are usually affectionate, and do not have the aloof quality typical of autism. Self-injury is relatively common, especially hand biting over the anatomical snuff-box (between the bases of the thumb and index finger) in response to frustration, anxiety, or excitement. Stereotyped behaviours such as hand flapping are common.

The associated intellectual disability is usually mild to moderate. Verbal intelligence scores exceed performance scores among populations of affected men and non-disabled women carriers. Speech and language development is delayed. Speech is often disorganized, with rambling and circuitous conversation, incomplete sentences, poor topic maintenance, tangential comments, echolalia, and perseveration. It may be rapid, or include peculiar changes in pitch.

There may be problems with attention and concentration that are disproportionate to the severity of the associated learning disability. Hyperactivity may be the presenting feature among boys with fragile-X who do not have intellectual disability.

Further information can be obtained from Hagerman and Hagerman (2002)[17] and www.fragilex.org.uk.

Sex chromosome abnormalities

The Y chromosome is small and has been completely mapped.[18] The X chromosome is much larger, containing over 1000 genes. Abnormalities of the X and Y chromosomes are more prevalent than those affecting autosomal chromosomes. Many affected children are not significantly dysmorphic and do not have major developmental disabilities.[19] Some remain undiagnosed.

Klinefelter syndrome

(a) Prevalence and genetics

This is a disorder characterized by additional X chromosomes in phenotypic males. Two-thirds have a 47 XXY chromosome complement. Prevalence at birth is about 1 in 1000 live males, with a frequency in prenatally karyotyped male foetuses of 1 in 470.[19]

(b) Physical characteristics

Height, weight, and head circumference are below average at birth. Increased growth, especially of legs occurs from 3 years of age onwards. Affected men are usually taller than their fathers, and mean heights are around the 75th centile. Head size remains small. Puberty normally occurs, but testosterone production falls in early adult life. Affected adults have a normal-sized penis but small testes. About 60 per cent have some breast enlargement. Life expectancy is thought to be normal.

(c) Behavioural and psychiatric aspects

Boys with XXY are typically introverted and less assertive and sociable than other children, with poorer school performance (especially with regard to reading and spelling). Adults may have increased rates of antisocial behaviour and impulsiveness. The IQ distribution is skewed downwards, although measured full scale IQs run from the 60s to the 130s. Performance scores usually exceed verbal scores. Most affected children receive speech and language therapy, and expressive language deficits are often more pronounced than problems with receptive language. One follow-up study has been reported.[20] Further information: www.klinefeltersyndrome.org.

Turner syndrome

(a) Prevalence and genetics

The genetic abnormality in Turner syndrome is the loss or abnormality of one X chromosome in women. The 45,X karyotype is found in about 1 in 10 000 live female births. The abnormality is much more common at conception. About 99 per cent of affected foetuses are miscarried, and 45,X is the most common karyotype found in chromosomally aborted foetuses. About 50 per cent have a 45,X chromosome complement (a very small proportion of normal cell lines may be present). Most of the other cases are the result of mosaicism, some are the result of structural abnormalities of an X chromosome.

(b) Physical characteristics

Affected children have a short stature in childhood. Ovarian failure occurs before birth, and puberty does not usually occur naturally, although childbirth has, rarely, been reported. Dysmorphic features include a webbed neck, low hairline at the rear of the head, widely spaced nipples and multiple pigmented naevi. About 12 per cent have coarctation of the aorta or a ventricular septal defect.

(c) Behavioural and psychiatric aspects

Hyperactivity and distractibility are common in childhood. Poor social skills, with immature social relationships and low self-esteem in adolescence were reported in one study.[21] Women with Turner syndrome are usually of normal intelligence and verbal abilities are usually unimpaired or enhanced. Specific cognitive abnormalities including deficits in spatial perception, visual motor integration, affect recognition, visual memory, and attention have been reported.[22] The relative strength in verbal tasks may lead to an

overestimation of abilities. There is considerable variation in cognitive profile between affected women. Further information: www.tss.org.uk

XXX syndrome

(a) Genetics and prevalence

The 47,XXX syndrome occurs about 1 in 1000 female births.[23] Many are not diagnosed. There is a primary non-disjunction of a maternal or paternal X chromosome. The 48,XXXX chromosome complement is much rarer (about 40 cases have been reported so far).

(b) Physical characteristics

In 47,XXX syndrome newborn babies have a low birth weight and small head circumference. Height in adult life is usually increased, with a low body mass index. Fertility is not usually impaired, although there are reports of premature ovarian failure and recurrent spontaneous abortion. There may be deficits in balance or fine motor coordination. Life expectancy is thought to be normal.

(c) Behavioural and psychiatric aspects

Underactivity and withdrawal have been reported. Emotional development may be slowed. About a quarter of affected women in one follow-up study had repeated episodes of abdominal pain as teenagers for which no organic cause could be found.[24] Most appear to adapt to adult life without difficulties. Women with the syndrome usually have IQs between 80 and 90. Women with XXXX syndrome have lower IQs (55 to 75). An expressive language delay is typical. Some have a relatively poor short-term auditory memory. Further information: www.triplo-x.org.

XYY syndrome

(a) Genetics, prevalence, and physical characteristics

This karyotype is associated with 1 in 1000 live male births.[23] There is a primary non-disjunction of the Y chromosome. About 10 per cent have mosaic 46,XY/47,XYY chromosome complement. Offsprings rarely have two Y chromosomes. Affected individuals show increase in body and leg length between years 4 and 9. Most are over 10 cm taller than their fathers as adults. Sexual development and fertility are unaffected. Balance and coordination may be minimally compromised. Life expectancy is normal.

(b) Behavioural and psychiatric aspects

Early research found an increased frequency of XYY men among inmates of special prisons.[25,26] More recent studies examining the relationship between 47,XYY karyotype and behaviour have concluded that affected men have lower mean intelligence scores (with a large overlap with the normal range) and poorer social adaptation. Distractibility, hyperactivity, temper tantrums, and speech and language problems appear relatively common in childhood. There is little evidence to suggest a significant link with seriously aggressive criminal conduct in adult life.[27,28]

Foetal alcohol syndrome

(a) Classification and prevalence

Exposure of the developing foetus to significant amounts of alcohol leads to cognitive impairment. The effect can occur during any stage of pregnancy, because brain development continues during all three trimesters. Dysmorphology, including a facial dysmorphology, can also occur. Foetal alcohol spectrum disorder (FASD) includes a number of subtypes including foetal alcohol syndrome (FAS), and more subtle abnormalities subsumed under the terms possible foetal alcohol effects (PFAE), prenatal exposure to alcohol (PEA), or alcohol-related neurodevelopmental disorder (ARND).

Foetal alcohol exposure is thought to be a common cause of intellectual disability in the United States and other developed countries. In the United States, an estimate of 0.33 per 1000 births has been given for the prevalence of foetal alcohol syndrome.[29] Alcohol inhibits *N*-methyl-D-aspartate receptors, which mediate postsynaptic excitatory effects of glutamate, and this is thought to have an effect on cell proliferation.[30]

(b) Clinical features

A number of abnormalities have been linked to FAS. Facial dysmorphology commonly includes a thin upper lip and smooth philtrum. The jaw may be small. Low-set abnormal ears and palate abnormalities can occur. Other abnormalities include growth retardation, skeletal abnormalities (deformed ribs and sternum, spinal curvature, dislocated hips, fused or webbed or missing fingers or toes, limited joint movement, small head), heart abnormalities, and urinary tract anomalies.

(c) Neurological and behavioural aspects

Central nervous system abnormalities include a small brain with abnormally arranged cells. Intellectual disability is usually mild or moderate but may be severe. Other problems commonly include reduction in attention span, overactivity, irritability in infancy; and coordination problems.

Angelman syndrome

(a) Prevalence and genetics

The prevalence of this syndrome is around 1 in 10 000 births.[31] Most cases are sporadic, and associated with deletions within 15q11q13 of maternal origin (Prader–Willi syndrome). Angelman syndrome is occasionally associated with paternal uniparental disomy (both chromosome 15s are of paternal origin) but this is less common than in Prader–Willi syndrome. Other genetic abnormalities leading to Angleman syndrome are an imprinting centre defect (this incorrectly 'marks' the chromosome, through methylation, as being from a parent of the opposite sex) and mutations in the gene responsible for the Angelman syndrome phenotype (UBE3A, coding for a ubiquitin ligase enzyme).[32] UBE3A is expressed only from the maternal chromosome, and in Angelman syndrome expression in relevant brain areas is only around 10 per cent of normal.

(b) Physical characteristics

Physical characteristics include a small head, characteristic face with wide mouth, 'hooked' nose, prominent lower jaw, widely spaced teeth, and tongue protrusion. Many affected children are hypopigmented compared to first degree relatives due to deletion of a gene related to pigmentation.[33] Voluntary movements are jerky and the gait ataxic with stiff legs. About 80 per cent develop epilepsy, and the EEG is highly characteristic.

(c) Behavioural and psychiatric aspects

Behavioural characteristics, including sudden bursts of laughter and the jerky ataxic gait, led to the term 'happy puppet' syndrome

being used in the literature of the 1960s and 1970s. It is no longer considered appropriate. Affected children enjoy social and physical contact, and mouthing objects. Many are fascinated by water. Intellectual disability is severe, with markedly delayed motor milestones. There is little speech development (no person reported in the literature has more than a six word vocabulary), but understanding of language may be better. Overactivity is often associated with a short attention span in childhood, but may improve with development. Behavioural studies have been reported[34] and genetic aspects reviewed.[35] Further information: www.assert.dial.pipex.com/

Coffin–Lowry syndrome

(a) Prevalence and genetics

Coffin–Lowry syndrome is one cause of X linked intellectual disability. The syndrome has been ascribed to a locus in the Xp22.1-p22.2 region. More than 100 cases have been reported.

(b) Physical characteristics

Physical features include short stature, facial dysmorphology including slanting eye fissures, prominent forehead, short broad nose, forward facing nostrils, large ears, large mouth and small, widely spaced teeth. Increased fatty tissue is deposited in the forearms. Hands are often large, with tapering fingers. Ligament laxity may lead to flat feet. Spinal and chest abnormalities occur. Behavioural and emotional characteristics are largely unknown, although depression and schizophrenia have been reported in association with the disorder and in female carriers.

(c) Behavioural and psychiatric aspects

Affected men usually have severe learning disabilities. Drop attacks and sleep apnoea syndrome have been reported. Further information: www.clsf.info.

Congenital hypothyroidism

(a) Prevalence

Following the introduction of a neonatal screening programme in the United Kingdom, the incidence of congenital hypothyroidism (identified through heel-prick screening and further investigation for at risk infants) is about 1 in 4000, and occurs more commonly in girls. In many cases the deficiency of thyroid hormones is mild, and there are few symptoms.

(b) Physical and cognitive characteristics

Severely affected children have a distinctive appearance with a puffy face and a large tongue that protrudes from a mouth that is kept open. Other features include dry, brittle hair, a low hair line, jaundice, sleep disorders, low muscle tone, constipation, and failure of cognitive development leading to intellectual disability. If untreated, even mild hypothyroidism may lead to intellectual disability.

Cri-du-Chat syndrome (CDCS, 5p-syndrome)

(a) Prevalence and genetics

Cri-du-Chat syndrome was originally described as a syndrome of multiple congenital anomalies, intellectual disability, microcephaly, abnormal face, and a mewing cry in infants with deletion of a 'B group chromosome', later identified as a 5p terminal deletion. The prevalence is about 1 in 35 000 births. Deletions vary in size, but the critical region for Cri-du-Chat syndrome is thought to be 5p15.2. About 85 per cent of the deletions arise spontaneously and the majority are of paternal origin. About 15 per cent of affected people have an unbalanced translocation, and the clinical features depend on the other chromosome involved. Fewer than 1 per cent of cases are due to inherited deletions, which are usually very small.

(b) Physical characteristics

In infancy there are feeding difficulties and the cry is abnormally high pitched (cat-like, hence 'cri-du-chat'), but this is not an invariable feature. The gene causing the abnormal (cat-like) cry has been located at 15p13. It is possible for infants with small deletions to have a cat-like cry but no other features of CDCS, or features of CDCS without the characteristic cry. A round face with widely spaced slanting eyes, a small head, a broad flat nose, and small lower jaw are characteristic. Ear abnormalities may occur. Larger deletions, and some translocations, are associated with more pronounced clinical features such as lower intelligence, smaller stature, lower weight, and smaller head. The face often lengthens with development and may be asymmetrical. Cleft lip or palate, curved fingers, hernias, and orthopaedic abnormalities may occur. Older individuals often have premature greying of the hair.

(c) Behavioural and psychiatric aspects

Hyperactivity is a problem for a substantial proportion of children, but may improve with age.[36] Language development is often markedly delayed. The IQ associated with the syndrome in one study varied from 6 to 85. Further information: www.criduchat.asn.au.

De Lange syndrome (Brachmann–de Lange syndrome)

(a) Prevalence and genetics

This syndrome is considered to occur about once in 60 000 live births, although some authors believe it to be more common. Mutations in a large gene on chromosome 5, the Nipped B like or *NIPBL* gene (named because its function resembles that of a fruit fly gene that produces a nipped wing), have been shown in about 40 per cent of people with the syndrome.[37]

(b) Physical characteristics

Affected individuals show growth retardation; distinctive facial features consisting of well-defined arched eyebrows which meet in the middle, long curled eyelashes, small nose with forward-facing nostrils, and down-turned mouth with thin lips and limb abnormalities such as small or shortened limbs, especially arms. Hearing impairments, gut malformations, and congenital heart defects also occur. Early mortality is high because of feeding problems with regurgitation and vomiting leading to aspiration pneumonia in some cases.

(c) Behavioural and psychiatric aspects

Self-injury, autistic features, and pleasurable responses to vestibular stimulation, e.g. spinning in a chair have been reported as part of behavioural repertoire. The degree of learning disability is usually severe, and speech is often very limited. However, some affected people have IQs within the normal range. Clinicians should be alert to the presence of pain and discomfort resulting from gastro-oesophageal reflux and other gastro-intestinal abnormalities. Further information: www.cdlsusa.org.

Duchenne muscular dystrophy

(a) Prevalence and genetics

This is an X linked recessive condition in which deletions, duplications, and mutations at Xp21 lead to failure to produce dystrophin, a protein component of muscle tissue. New mutations account for about 30 per cent of cases. The prevalence at birth is about 1 in 4000 male births.

(b) Physical characteristics

The syndrome is characterized by progressive muscle weakness, affecting the pelvis, upper leg, and upper arm muscles first. The onset is typically between 2 and 6 years of age. Respiratory muscles are involved later in the disease process. Heart muscle abnormalities may also occur. The disease is usually more severe in the lower limbs and trunk initially, with later involvement of the arms and respiratory muscles. Affected boys often need a wheelchair by around 11 years of age, with death in early adult life (typically in the mid twenties).

(c) Behavioural and psychiatric aspects

Low mood, anxiety, and social abnormalities are often problems, and may become more prominent as the disorder progresses. These features may be reactions to a chronic and progressive physical disease. Specific learning disabilities are common, especially specific reading disorder. About 25 per cent of those affected have a learning disability. Performance IQ is typically higher than verbal IQ. Further information: www.mda.org.au.

Lesch–Nyhan syndrome

(a) Prevalence and genetics

This X-linked recessive disorder results from a deficiency of a purine salvage enzyme, hypoxanthine-guanine phosphoribosyl transferase (HGPRT) leading to hyperuricaemia, and neurological disorder. Partial HGPRT deficiency results in gout. HGPRT is a 217 amino acid peptide coded for by one gene divided into nine exons, located on the X chromosome at Xq26q27. Many different genetic lesions can cause HGPRT deficiency. Complete and partial deletions, insertions, and duplication of exons have been reported. Most lesions appear to be point mutations. Affected males may have had spontaneous mutations or inherited mutations from asymptomatic female carriers. Carrier detection and prenatal diagnosis are possible. The incidence is around 1 in 380 000 births.[38]

(b) Physical characteristics

Neurological features include athetoid and other abnormal movements and spasticity. Growth retardation is usual. The presentation is usually with hypotonia and motor delay at about 4 months. Extrapyramidal signs (such as spasticity and choreo-athetoid movements) develop at about 9 months. Hyper-reflexia and clonus appear at about 1 year. Dystonic movements may also develop. Dysarthria is common. Affected individuals may survive to the second or third decade. Death is usually due to kidney failure secondary to uric acid deposition or infection. The syndrome is associated with abnormal neurotransmitter turnover in the basal ganglia.

(c) Behavioural and psychiatric aspects

Compulsive severe self-injury is very prevalent and usually consists of finger and lip biting, with self-splinting in an attempt to prevent the behaviour.[39] Other compulsive behaviours occur; men with the syndrome are reported to hit, spit, and swear at caregivers while apologizing for their behaviour at the same time.[40] The mean age at onset of self-injury is 3.5 years, with wide variation. The IQ is usually between 40 and 80, but dysarthria and neurological problems limit the validity of standard IQ tests. Further information: www.lndinfo.org.

Mucopolysaccharidoses

(a) Classification, genetics, and prevalence

The mucopolysaccharide group of disorders have both names (Hunter syndrome, Hurler syndrome, Sanfillipo syndrome, Morquio syndrome, Schie syndrome, Maroteaux–Lamy syndrome, Sly syndrome) and numerical designations (MPS IIA/B, MPS IH, MPS IIIA/B/C/D, MPS IVA/B, MPS IS, MPSVI, MPSVII, respectively). The disorders result from deficiencies in enzyme systems involved in the degradation of glycosaminoglycans leading to the accumulation of abnormal metabolic products. The prevalences among live born children are approximately 1 in100 000 for Hunter and Hurler syndromes, 1 in 200 000 for all types of Sanfillipo syndrome and for Morquio syndrome, and 1 in 500 000 for Schie syndrome. Hunter syndrome is much more prevalent in Israel. The transmission is autosomal recessive except in Hunter (IIA and IIB) which is X linked.

(b) Physical characteristics

Physical features vary. Coarse facial features ('gargoylism'), hepatosplenomegaly, joint stiffness, eye abnormalities, and short stature occur in many of the disorders. Life expectancy varies from death in the first decade in Hurler syndrome through survival into second or third decade in Sanfillipo syndrome, to survival to adult life in Hunter syndrome and Schie syndrome.

(c) Behavioural and psychiatric aspects

Sleep problems and abnormal nocturnal behaviours such as staying up all night, night-time laughing and singing, sudden crying out, and chewing of bedclothes have been reported in association with Sanfillipo syndrome, and have been shown to respond to behavioural management strategies. Other problem behaviours reported in association with MPS disorders include aggression, overactivity, restlessness, and anxiety. Cognitive abilities vary from normal intelligence in Schie syndrome to severe learning disability and progressive cognitive deterioration in Hurler syndrome. Sanfillipo syndrome is associated with slower progressive cognitive impairment than that seen in Hurler syndrome, but often with marked behavioural and psychiatric abnormalities consistent with the diagnosis of childhood disintegrative disorder. The susceptibility to tooth decay in Morquio syndrome can lead to pain and problem behaviours. Further information: www.mpssociety.org.

Neurofibromatosis type 1

(a) Prevalence and genetics

This autosomal dominant disorder was first described by von Recklinghausen in 1882 and occurs about once in 3000 births. The gene responsible is localized to 17q11.2. The gene product, neurofibromin, regulates cell division and is thought to suppress tumour formation. A high spontaneous mutation rate means that about a half of all cases of NF1 arise in unaffected families.

(b) Physical characteristics

Tumours arise from the connective tissue of nerve sheaths. Two or more of the following features are usually required for diagnosis:

six or more cafe au lait (light brown) skin lesions more than 5 mm in diameter before puberty or 15 mm after puberty, two or more neurofibromas or one plexiform neurofibroma (tumours of the nerve sheath); freckling of the inguinal or axillary region; two or more lisch nodules (benign iris hamartomas); an optic nerve glioma (tumour); a bony lesion characteristic of neurofibromatosis (usually shin bowing or scoliosis), a first degree relative with the disorder. About 45 per cent of affected people will have non-enhancing hyperintensities (or unidentified bright objects 'UBOs') on magnetic resonance imaging. These are commonly seen in the cerebellum, basal ganglia, brain stem, and thalamus.[41]

(c) Behavioural and psychiatric aspects

About 50 per cent of children have speech or language abnormalities. Distractibility and impulsiveness may be problems. Learning disability is present in about 10 per cent of affected people. Specific developmental disorders such as difficulties with reading, writing, or numeracy affect about half of the children. Visuo-spatial abnormalities and lack of coordination have also been described. Further information: www.geneclinics.org/profiles/nf1.

Phenylketonuria

(a) Prevalence and genetics

Classical phenylketonuria affects about 1 in 10 000 live born children in the United Kingdom. Other hyperphenylalaninaemias also occur. The disorder results from a deficiency of the enzyme phenylalanine hydroxylase. The extent of the deficiency varies, with a spectrum of resulting clinical conditions from classical phenylketonuria to benign hyperphenylalaninaemia. The gene regulating phenylalanine hydroxylase is located at 12q22-24.1. It is subject to various mutations. The classical form is inherited in an autosomal recessive manner. Prenatal diagnosis and the detection of heterozygotes with one defective copy of the gene are possible. About 2 per cent of cases are due to a deficiency of tetrahydrobiopterin rather than phenylalanine hydroxylase.

(b) Physical characteristics

Physical features include blond hair, blue eyes, eczema, and microcephaly (in half the suffers), epilepsy (in a quarter) and tremor and movement disorders or spasticity. Untreated infants have an unusual mouse-like body odour. In the United Kingdom all neonates are screened for the disorder. A low phenylalanine diet is usually continued through childhood. There is debate about the age at which it is appropriate to lift or relax dietary restrictions. Amino-acid supplements may be used to block phenylalanine uptake. Dietary control is essential when affected women become pregnant, because hyperphenylalaninaemia is toxic to the foetus leading to learning disability, microcephaly, and facial and heart abnormalities. Theories about the toxic effects of hyperphenylyalaninaemia include direct toxicity, competition for transport across the blood–brain barrier and dopamine depletion.[42]

(c) Behavioural and psychiatric aspects

Untreated phenylketonuria is associated with a number of maladaptive behaviours and behavioural syndromes including overactivity, self-injury, and autism. Autism and many of the other features do not occur in children managed with low phenylalanine diets. Those who have not been treated may have moderate to profound learning disabilities, irritability, and marked social impairments. Inadequate dietary control is associated with deficits in mathematical, visuo-spatial, and language skills. Further information: www.nspku.org.

Prader–Willi syndrome

(a) Prevalence and genetics

The prevalence is around 1 in 40 000 live born infants.[43] About 70 per cent of those affected have a deletion affecting the long arm of chromosome 15 (del 15q11q13), the deleted chromosome always being of paternal origin. About 29 per cent have maternal uniparental disomy (MUPD) in which both chromosome 15s are inherited from mother, with no paternal chromosome 15. About 1 per cent have an imprinting error, in which the paternal chromosome is incorrectly methylated so as to resemble a maternal one.

(b) Physical characteristics

Infants are hypotonic or floppy and have feeding problems associated with a failure to suck. Many are tube fed. In early childhood there is a switch to marked overeating. Affected adults are of short stature, have small hands and feet and a characteristic pattern of facial appearance, and a lack of sexual development. Affected people were often obese, as a result of the impaired satiety leading to overeating, but modern dietary management and treatment with growth hormone in childhood may lead to near normal body size and shape. There is an increased prevalence of curvature of the spine or scoliosis and other orthopaedic abnormalities, and diabetes or heart failure may result from obesity. Life expectancy depends on severity of obesity.

(c) Behavioural, cognitive, and psychiatric aspects

Affected individuals have an almost insatiable appetite. They may steal food and consume 'unpalatable' food such as rotting or frozen food or pet food. A variety of sleep abnormalities and a lowering of the threshold for loss of temper may be associated. About 80 per cent pick or scratch their skin. Insistence on routines, and compulsive behaviours are commonly reported.[44] Severe psychiatric disorders including affective and psychotic states are more prevalent, especially among people with MUPD.[45] Anecdotal reports suggest the pain threshold may be raised.

About 5 per cent of those with this syndrome have overall cognitive abilities with IQs in excess of 85, 27 per cent have borderline cognitive abilities with IQs between 70 and 85, 34 per cent have mild learning disabilities, 27 per cent moderate, 5 per cent severe, and less than 1 per cent have profound learning disability. There are deficits in auditory information processing, and relative strengths in visuo-spatial tasks. Further information: www.ipwso.org.

Rett syndrome

(a) Prevalence and genetics

Rett syndrome causes severe intellectual disability in women. The prevalence in the United Kingdom is around 1 in 10 000 women.[46] The syndrome results from a mutation in the MeCP2 gene located at Xq28. The mutation was considered to be lethal in males but there are a small number of males with the syndrome. The severity of the syndrome in women depends on the percentage of cells with the normal MeCP2 gene active after X inactivation. If more of the X chromosomes with normal MeCP2 gene have been inactivated, the syndrome is likely to be more severe. MeCP2 acts as a mechanistic bridge between DNA methylation and histone methylation.[47]

(b) Physical characteristics

The affected child appears normal at birth. For the first 12 months no major abnormalities are apparent though the child may be placid, lack muscle tone, or be relatively immobile. She acquires skills to about 1 year with regression and loss of skills from around 18 months onwards. Speech and use of hands are particularly affected. Physical problems increase with age, and include scoliosis, spasticity, and leg deformities. Epilepsy is common. Pathological changes include a reduction in brain size with reduced cortical thickness, reduced neuronal branching, and depigmentation of the basal ganglia. Many affected girls reach adulthood, but about 1 per cent of them die each year with early death more likely with increasing physical disability.

(c) Behavioural and psychiatric aspects

Sleep disturbance, withdrawal and episodes of crying occur during the phase of regression around 18 months of age. This is followed by a phase in which development stops. Extreme agitation and over-breathing interspersed with episodes of cessation of breathing then become apparent. The most prominent feature of the behavioural phenotype is the presence of stereotyped movements, especially midline 'hand-wringing' movements. Affected women and girls usually have profound learning disability. Further information: www.rettsyndrome.org.uk.

Rubinstein–Taybi syndrome

(a) Prevalence and genetics

This syndrome is one of the 25 most common multiple congenital anomaly syndromes seen in genetic clinics in the United States and has an estimated incidence at 1 in 125 000 live born infants. Microdeletions at 16p13.3 have been described in some cases, and mutations in the gene coding for CREB-binding protein (CBP) found at this locus have been reported to cause the syndrome.[48] A few apparently familial cases have been reported, and four sets of concordant monozygotic twins have been reported.

(b) Physical characteristics

The affected individuals are usually short, have a small head, a beaked or straight nose and downward slanting eyes. They have a stiff gait. The thumbs and first toes have broad terminal phalanges, often with an angulation deformity. Other congenital anomalies are not uncommon. Inadequate weight gain in infancy, congenital heart defects, urinary tract abnormalities, and severe constipation contribute to morbidity reflected in a hospitalization rate 10 times higher than the general population.

(c) Behavioural and psychiatric aspects

Findings from postal questionnaire surveys in the United States and United Kingdom indicate that people with the syndrome have a friendly disposition, a propensity to self-stimulatory activities such as rocking and an intolerance of loud noises. Reduced attention span, rocking, spinning, and hand flapping were common in the UK survey.[49] Intellectual disability is usually moderate. Further information: www.rubinstein-taybi.org.

Smith–Lemli–Opitz syndrome

(a) Prevalence and genetics

This disorder is thought to occur about once in 30 000 live births. Mildly affected people may be undiagnosed. The syndrome is said to be one of the commonest autosomal recessive conditions affecting people of White European origin in North America, but rare among people of African or Asian origin. The male to female ratio appears to be around 3 to 1 but this may be due to the relative ease with which sexual abnormalities are detectable in men. Abnormalities in the *DHCR7* gene, located at 11q12-13 results in deficiency of the enzyme 7-dehydrocholesterol reductase results in elevated levels of a cholesterol precursor. Treatment with cholesterol supplementation has been attempted.[50]

(b) Physical characteristics

During pregnancy, the foetus may show growth retardation. Affected individuals may have a small head, drooping eyelids, squint, forward-facing nostrils, small lower jaw, and finger abnormalities such as extra fingers and syndactyly. Males have abnormalities of their external genitalia such as small testes or penis, hypospadias, undescended testes, and female type genitalia. Cleft palate and abnormalities of almost all major organ systems may also occur.

(c) Behavioural and psychiatric aspects

There is little information available about behavioural and cognitive characteristics. Intelligence varies from normal to severe learning disability. Aggressive and self-injurious behaviours and autistic spectrum disorders have been reported. Further information: www.geneclinics.org/profiles/slo.

Smith–Magenis syndrome

(a) Prevalence and genetics

This syndrome, affecting around 1 in 50 000 births, is associated with deletions at 17p11.2.

(b) Physical characteristics

Affected individuals have flattened mid-face, abnormally shaped upper lip, short hands and feet, single transverse palmar crease, abnormally shaped or placed ears and sometimes a high arched palate or protruding tongue. The facial features may coarsen with development. Ear and eye disorders such as otitis media and squint are relatively common.

(c) Behavioural and psychiatric aspects

Newborn babies with the syndrome are usually placid, 'floppy' and feed with difficulty. This changes to hyperactivity, self-injury (e.g. head banging, pulling out finger and toe nails and the insertion of objects into body orifices), from about 18 months onwards. Self-hugging and mid-line hand clapping have been reported. Sleep disorders are common with some children waking repeatedly in a state of agitation. An absence of rapid eye movement (REM) sleep has been reported in some patients. Many affected children appear to be relatively insensitive to pain. Behavioural studies have been reported.[36,51] The severity of the cognitive impairment correlates with the size of the 17p11 deletion. Moderate intellectual disability is common. Speech delay is more pronounced than delay in motor achievements. Further information: www.prisms.org.

Tuberous sclerosis

(a) Prevalence and genetics

About 1 in 7000 people are affected. It is an autosomal dominant condition but up to 80 per cent of cases arise as a result of spontaneous mutations. The disorder is genetically heterogeneous, with gene linkage to 9q34 and 16p13.

(b) Physical characteristics

Physical features are very variable. The previously used diagnostic triad of epilepsy, intellectual disability and a characteristic facial skin lesion is seen in only about 30 per cent of people with the disease. The disorder is a multi-system one, with hamartomatous tumours (arising from primitive cells) affecting the brain (in about 90 per cent), skin, kidneys, heart, eyes, teeth, bones, lungs, and other organs. About 80 per cent of affected people have epilepsy. Brain tumours and kidney lesions are common causes of death.

(c) Behavioural and psychiatric aspects

Tuberous sclerosis is associated with autism and related disorders, hyperactivity and attention-deficit disorder, obsessive and ritualistic behaviour, sleep problems, and occasionally self-injurious or aggressive behaviours. Less than half of affected people have a learning disability. Attention-deficit is common. Of those with learning disability, many have an IQ less than 30. Further information: www.tuberous-sclerosis.org.

Velo-cardio-facial syndrome

(a) Prevalence and genetics

First described in 1978[52] this condition is relatively common, affecting about 1 in 2000 people. The disorder has also been called Shprintzen syndrome, Digeorge syndrome, Cayler syndrome, Takao syndrome, conotruncal anomalies face syndrome, 22q11 deletion syndrome, and CATCH22. It is associated with microdeletions at 22q11. About 90 per cent arise *de novo*, with 10 per cent having an affected parent.

(b) Physical characteristics

Physical features include cardiac abnormalities including ventriculoseptal defects, pulmonary stenosis, and cardiac outlet abnormalities; facial dysmorphology with a prominent nose with broad bridge and squared tip, small head or small lower jaw; ocular abnormalities; cleft palate; short stature and long, thin, hyperextensible fingers. The clinical features associated with the disorder are highly variable in both type and severity.

(c) Behavioural and psychiatric aspects

Many affected individuals have difficulties with reciprocal social interaction.[53] A high prevalence of severe psychiatric disorders is reported in later life, including high rates of bipolar affective disorder[54] and schizophrenia.[55] Anxiety, social withdrawal and other disorders have also been described. Over 90 per cent have a learning disability. Speech and language problems are common. Further information: www.vcfsef.org.

Williams syndrome

(a) Prevalence and genetics

Also known as idiopathic infantile hypercalcaemia, the syndrome affects about 1 in 15 000 infants. Most cases are sporadic though a few familial cases have been reported where the transmission seems to be autosomal dominant. The syndrome is a contiguous gene deletion disorder in which there is variable loss of genetic material involving the Elastin gene at 7q11.3 and sometimes as many as 17 nearby genes in the Williams syndrome critical region.[56]

(b) Physical characteristics

Infants have difficulties in feeding, are irritable, have constipation and fail to thrive. Over 60 per cent of children have high serum calcium concentrations. This can be treated with a low-calcium diet and vitamin D restriction. The face is distinctive, with prominent cheeks, a wide mouth and flat nasal bridge often described as 'elfin-like'. Kidney and heart lesions (especially supravalvular aortic stenosis and peripheral pulmonary artery stenosis) are common. Growth is usually retarded. Life expectancy is related to metabolic and heart abnormalities.

(c) Behavioural and psychiatric aspects

Social disinhibition with abnormal friendliness to strangers, overactivity, poor concentration, eating and sleeping abnormalities, abnormal anxiety, poor peer relationships, and abnormally sensitive hearing have been reported.[57,58] About 95 per cent of children with the disorder have a moderate or severe learning disability. Verbal abilities are better developed than visuo-spatial and motor skills. There is an unusual command of language in which expressive language is superficially fluent and articulate but comprehension is far more limited. Further information: www.williams-syndrome.org.

Further information

Information about individual syndromes is available from the source listed after the relevant syndrome in the text above. O'Brien (2002)[2] gives information about behavioural, cognitive, linguistic and psychiatric aspects of several genetic disorders. The Contact a Family Directory of Specific Conditions and Rare Disorders (CaF Directory) is widely used for basic information about characteristics and carer organizations. A new paper edition is published in January each year and it is also available in CD-ROM format on a quarterly subscription basis: www.cafamily.org.uk.

References

1. Jones, K. (2005). *Smith's recognizable patterns of human malformation* (6th edn). W.B. Saunders, Philadelphia, PA.
2. O' Brien, G. (ed.) (2002). *Behavioural phenotypes in clinical practice*. Mac Keith Press, Cambridge.
3. Deb, S., Clarke, D., and Unwin, G. (2006). *Using medication to manage behaviour problems among adults with a learning disability: quick reference guide*. University of Birmingham, Royal College of Psychiatrists and Mencap, Birmingham (www.ld-medication.bham.ac.uk).
4. Fryers, T. and Russell, O. (2003). Applied epidemiology. In *Seminars in the psychiatry of learning disabilities* (2nd edn) (eds. W. Fraser and M. Kerr), pp. 16–48. Gaskell, London.
5. Prasher, V.P. (1995). Age-specific prevalence, thyroid dysfunction and depressive symptomatology in adults with Down's syndrome and dementia. *International Journal of Geriatric Psychiatry*, **10**, 25–31.
6. Holland, A.J., Hon, J., Huppert, F.A., *et al.* (1998). Population-based study of the prevalence and presentation of dementia inadults with Down's syndrome. *The British Journal of Psychiatry*, **172**, 493–8.
7. Prasher, V. (2003). Health morbidity in adults with Downsyndrome. In *Seminars in the psychiatry of learning disabilities*(2nd edn) (eds. W. Fraser and M. Kerr), pp. 267–86. Gaskell,London.
8. Herbst, D.S. and Miller, J.R. (1980). Nonspecific X-linked mental retardation II: the frequency in British Columbia. *American Journal of Medical Genetics*, **7**, 461–9.
9. Merienne, K., Jaquot, S., Pannetier, S., *et al.* (1999). A missense mutation in RPS6KA3 (RSK2) responsible for non-specific mental retardation. *Nature Genetics*, **22**, 13–4.
10. Robertshaw, B.A. and MacPherson, J. (2006). Scope for more genetic testing in learning disability. *The British Journal of Psychiatry*, **189**, 99–101.

11. Chiurazzi, P., Tabolacci, E., and Neri, G. (2004). X-linked mental retardation (XLMR): from clinical conditions to cloned genes. *Critical Reviews in Clinical Laboratory Sciences*, **41**, 117–58.
12. Jaquemont, S., des Portes, V., and Hagerman, R. (2005). Fragile X and X-linked mental retardation. In *Genetics of developmental disabilities* (eds. M.G. Butler and F.J. Meaney), pp. 247–78. Taylor & Francis, Boca Raton.
13. Morton, J.E., Bundey, S., Webb, T.P., *et al.* (1997). Fragile-X syndrome is less common than previously estimated. *Journal of Medical Genetics*, **34**, 1–5.
14. Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., *et al.* (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, **54**, 905–14.
15. Hagerman, R.J. and Lampe, M.E. (1999). Fragile X syndrome. In *Handbook of neurodevelopmental and genetic disorders in children* (eds. S. Goldstein and C.R. Reynolds), pp. 298–316. Guilford Press, New York.
16. Hagerman, R.J. (1996). Physical and behavioural phenotype. In *Fragile X syndrome: diagnosis, treatment and research* (2nd edn) (eds. R.J. Hagerman and A.C. Cronister), pp. 3–87. Johns Hopkins University Press, Baltimore.
17. Hagerman, R.J. and Hagerman, P.J. (2002). *Fragile X syndrome: diagnosis, treatment and research*. Johns Hopkins University Press, Baltimore.
18. Tilford, C.A., Kuroda-Kawaguchi, T., Skaletsky, H., *et al.* (2001). A physical map of the human Y chromosome. *Nature*, **409**, 943–5.
19. Hammerton, J.L. and Evans, J.A. (2005). Sex chromosome anomalies. In *Genetics of developmental disabilities* (eds. M.G. Butler and F.J. Meaney), pp. 585–650. Taylor & Francis, Boca Raton.
20. Radcliffe, S. (1999). Long-term outcome in children of sex chromosome abnormalities. *Archives of Disease in Childhood*, **80**, 192–5.
21. McCauley, E., Ross, J.L., Kushner, H., *et al.* (1995). Self-esteem and behaviour in girls with Turner syndrome. *Journal of Developmental and Behavioral Pediatrics*, **16**, 82–8.
22. Romans, S.M., Stefanatos, G., Roeltgen, D.P., *et al.* (1998). Transition to young adulthood in Ullrich-Turner syndrome: neurodevelopmental changes. *American Journal of Medical Genetics*, **79**, 140–8.
23. Hook, E.B. and Hamerton, J.L. (1977). The frequency of chromosome abnormalities detected in consecutive newborn studies: results by sex and by severity of phenotypic involvement. In *Population cytogenetics* (eds. E. Hook and I.H. Porter), pp. 63–79. Academic Press, New York.
24. Robinson, A., Bender, B.G., and Linden, M.G. (1991). Summary of clinical findings in children and young adults with sex chromosome anomalies. In *Children and young adults with sex chromosome aneuploidy* (eds. J.A. Evans, J.L. Hamerton, and A. Robinson), pp. 225–8. Wiley-Liss, New York.
25. Jacobs, P.A., Brunton, M., Melville, M.M., *et al.* (1965). Aggressive behaviour, mental subnormality, and the XYY male. *Nature*, **208**, 1351–2.
26. Jacobs, P.A., Price, W.H., Court Brown, W.M., *et al.* (1969). Chromosome studies on men in a maximum security hospital. *Annals of Human Genetics*, **31**, 339–58.
27. Witkin, H.A., Mednick, S.A., Schulsinger, F., *et al.* (1976). Criminality in XYY and XXY men. *Science*, **193**, 547–55.
28. Fryns, J.P., Kleczkowska, A., Kubien, E., *et al.* (1995). XYY syndrome and other Y chromosome polysomies. Mental status and social functioning. *Genetic Counseling*, **6**, 197–206.
29. Abel, E.L. and Sokel, R.J. (1987). Incidence of fetal alcohol syndrome and economic impact of FAS-related anomalies: drug alcohol syndrome and economic impact of FAS-related anomalies. *Drug and Alcohol Dependence*, **19**, 51–70.
30. West, J.R., Perrotta, D.M., and Erickson, C.K. (1998). Fetal alcohol syndrome: a review for Texas physicians. *Medical Journal of Texas*, **94**, 61–7.
31. Petersen, M.B., Brondum-Nielsen, K., Hansen, L.K., *et al.*(1995). Clinical, cytogenetic, and molecular diagnosis of Angelman syndrome: estimated prevalence in a Danish county. *American Journal of Medical Genetics*, **60**, 261–2.
32. Kishino, T., Lalande, M., and Wagstaff, J. (1997). UBE3A/E6-AP mutations cause Angelman syndrome. *Nature Genetics*, **15**, 70–73 [erratum in *Nature Genetics*, **15**, 411].
33. King, R.A., Wiesner, G.L., Townsend, D., *et al.* (1993). Hypopigmentation in Angelman syndrome. *American Journal of Medical Genetics*, **46**, 40–4.
34. Clarke, D.J. and Marston, G. (2000). Problem behaviors associated with 15q-Angelman syndrome. *American Journal of Mental Retardation*, **105**, 25–31.
35. Williams, C.A. (2005). Angelman syndrome. In *Genetics of developmental disabilities* (eds. M.G. Butler and F.J. Meaney), pp. 319–35. Taylor & Francis, Boca Raton.
36. Clarke, D.J. and Boer, H. (1998). Problem behaviours associated with deletion Prader-Willi, Smith-Magenis and Cri du chat syndromes. *American Journal of Mental Retardation*, **103**, 264–71.
37. Krantz, I.D., McCallum, J., DeScipio, C., *et al.* (2004). Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of *Drosophila melanogaster* Nipped-B. *Nature Genetics*, **36**, 631–5.
38. Nyhan, W.L. and Wong, D.F. (1996). New approaches to understanding Lesch-Nyhan disease. *The New England Journal of Medicine*, **334**, 1602–4.
39. Deb, S. (1998). Self injurious behaviour as part of genetic syndromes. *The British Journal of Psychiatry*, **172**, 385–88.
40. Matthews, W.S., Solan, A., and Barabas, G. (1995). Cognitive functioning in Lesch-Nyhan syndrome. *Developmental Medicine and Child Neurology*, **37**, 715–22.
41. Gutmann, D.H., Aylsworth, A., Carey, J.C., *et al.* (1997). Diagnostic evaluation and multidisciplinary management of neurofibromatosis 1 and neurofibromatosis 2. *The Journal of the American Medical Association*, **278**, 51–7.
42. Waisbren, S.E. (1999). Phenylketonuria. In *Handbook of neurodevelopmental and genetic disorders in children* (eds. S. Goldstein and C.R. Reynolds), pp. 433–58. Guilford Press, New York.
43. Whittington, J.E., Holland, A.H., Webb, T., *et al.* (2001). Population prevalence and estimated birth incidence and mortality rate for people with Prader-Willi syndrome in one UK Health region. *Journal of Medical Genetics*, **38**, 792–8.
44. Clarke, D.J., Boer, H., Whittington, J., *et al.* (2002).Prader-Willi syndrome, compulsive and ritualistic behaviours. The first population-based survey. *The British Journal of Psychiatry*, **180**, 358–62.
45. Boer, H., Holland, A., Whittington, J., *et al.* (2002). Psychotic illness in people with Prader-Willi syndrome due to chromosome 15 maternal uniparental disomy. *Lancet*, **159**, 135–6.
46. Kerr, A.M. and Stephenson, J.B. (1985). Rett's syndrome in the West of Scotland. *British Medical Journal*, **291**, 579–82.
47. Fuks, F., Hurd, P.J., Wolf, N., *et al.* (2003). The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *The Journal of Biological Chemistry*, **278**, 4035–40.
48. Petrij, F., Giles, R.H., Dawerse, H.G., *et al.* (1995). Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature*, **376**, 248–51.
49. Boer, H., Langton, J., and Clarke, D. (1999). Development and behaviour in genetic syndromes: Rubinstein-Taybi syndrome. *Journal of Applied Research in Intellectual Disabilities*, **12**, 302–7.
50. Stark, L., Lovgren-Sandblom, A., and Bjorkhem, I. (2002). Cholesterol treatment forever? The first Scandinavian trial of cholesterol supplementation in the cholesterol-synthesis defect Smith-Lemli-Opitz syndrome. *Journal of Internal Medicine*, **252**, 314–21.

51. Dykens, E.M. and Smith, E.C. (1998). Distinctiveness and correlates of maladaptive behaviour in children and adolescents with Smith-Magenis syndrome. *Journal of Intellectual Disability Research*, **42**, 481–9.
52. Shprintzen, R.J., Goldberg, R.B., Lewin, M.L., *et al.* (1978). A new syndrome involving cleft palate, cardiac anomalies, typical facies, and learning disabilities: velo-cardio-facial syndrome. *Cleft Palate*, **15**, 56–62.
53. Swillen, A., Devriendt, K., Legius, *et al.* (1997). Intelligence and psychosocial adjustment in velo-cardio-facial syndrome: a study of 37 children and adolescents with VCFS. *Journal of Medical Genetics*, **34**, 453–8.
54. Popolos, D.F., Faedda, G.L., Veit, S., *et al.* (1996). Bipolar spectrum disorders in patients diagnosed with velo-cardio-facial syndrome: does a hemizygous deletion of chromosome 22q11 result in bipolar affective disorder? *The American Journal of Psychiatry*, **153**, 1541–7 .
55. Murphy, K.C., Jones, L.A., and Owen, M.J. (1999). High rates of schizophrenia in adults with velo-cardio-facial syndrome. *Archives of General Psychiatry*, **56**, 940–5.
56. Meng, X., Lu, X., Li, Z., *et al.* (1998). Complete physical map of the common deletion region in Williams syndrome and identification and characterization of three novel genes. *Human Genetics*, **103**, 590–9.
57. Udwin, O. (1990). A survey of adults with Williams syndrome and idiopathic infantile hypercalcaemia. *Developmental Medicine and Child Neurology*, **32**, 129–41.
58. Udwin, O. and Yule, W.A. (1991). A cognitive and behavioural phenotype in Williams syndrome. *Journal of Clinical and Experimental Neuropsychology*, **13**, 232–44.

10.5

Psychiatric and behaviour disorders among mentally retarded people

Contents

10.5.1 Psychiatric and behaviour disorders among children and adolescents with intellectual disability

Bruce J. Tonge

Introduction

Psychopathology is 2–3 times more common in intellectually disabled (ID) children than in the general population.[1,2] Psychiatric disorder is the most common source of additional handicap causing loss of educational, recreational, and social opportunity, burden for carers and cost to the community. Numerically, the size of this problem is approximately equal to schizophrenia, but is less well-recognized due to diagnostic over shadowing in which psychiatric disorder is not differentiated from ID as a separate condition open to diagnosis and treatment.[2] Although there is probably a significant reduction in overall prevalence of psychiatric disorders from approximately 43 per cent of children to 37 per cent of young adults with ID, psychopathology if present in childhood is likely to persist.[2] The profile of disorders varies from childhood into young adult life with the prevalence of attention-deficit hyperactivity symptoms decreasing, the frequency of symptoms of depression increasing and the prevalence of anxiety remaining stable with maturation.[2]

Diagnosis and classification

There are two approaches to the description and classification of psychopathology in young people with ID. First is the application of DSM and ICD diagnostic criteria. The reliability and validity of this approach is not well established when applied to children with ID.[3] Young people with more severe ID and language impairment are unable to report abnormalities of their emotions, thoughts, and perceptions, which are criteria for conditions such as obsessive-compulsive disorder (OCD) and schizophrenia. Some diagnoses, for example attention-deficit hyperactivity disorder (ADHD)[4] require a judgement that symptoms are inconsistent with developmental level, which in a child with ID is delayed relative to chronological age. The DSM-IV TR[4] specifies that either ADHD or separation anxiety disorder should not be made 'exclusively during the course of a pervasive developmental disorder'. These restrictions on comorbid diagnosis should not limit the necessity to describe the range of presenting symptoms and offer appropriate treatment, for example the use of stimulant medication in a child with autism and severe ADHD symptoms. Developmental level and degree of cognitive impairment influence the presentation of symptoms. For example, children with ID are more likely than children in general, to have externalizing symptoms such as disruptive, aggressive, impulsive, or avoidant behaviours; if psychotic to experience hallucinations without delusions; or if depressed to present with irritability and stereotypies. Self-absorbed, autistic, and withdrawn behaviours are more common in children with severe ID whereas anxiety, disruptive, and aggressive behaviours are more likely in children with milder levels of ID.[2] Some patterns of psychopathology recognized by DSM-IV TR[4] are specifically associated with more severe ID such as 'stereotypic movement disorder with or without self-injurious behaviour'. Other emotional and behavioural disturbances seen in people with ID receive non-specific, atypical, or not otherwise specified classifications and await better definition. Recent attempts to produce diagnostic criteria for psychiatric disorders in people with ID (the draft ICD-10 guidelines for the psychiatric assessment

of persons with mental retardation,[5] the Royal College of Psychiatrists diagnostic criteria for psychiatric disorders for use with adults with learning disabilities[6] and the DSM-IV TR for intellectual disability[7] are mainly designed for use with adults and require clinical validation).

The second approach to the definition of psychopathology in young people with ID is the use of informant questionnaires which rate disturbed emotions and behaviour. Factor analysis produces subscales which have clinical utility and refer to dimensions of disturbance such as disruptive/antisocial behaviours, social withdrawal, self-absorbed behaviours, communication disturbance, and anxiety/depression. Two reliable questionnaires validated for use in children and adolescents with ID are the Nisonger Child Behaviour Rating form[8] and the Developmental Behaviour Checklist.[9]

The multiaxial classification system of DSM or ICD, revised for use in people with ID, should form the basis of diagnosis of psychiatric disorder in young people with ID, but are usefully supplemented with standardized information gathered from informant questionnaires.

Contributing factors and context

Assessment of the psychopathology associated with ID requires consideration of the biopsychosocial context.

(a) Cognitive profile

A standardized cognitive assessment provides essential information to inform diagnosis and guide treatment. The level of intellectual and language ability gives an indication of the child's capacity to comprehend and communicate their perceptions, thoughts, and emotions. Subjective experiences such as grief, anxiety, hallucinations, and delusions cannot be assessed if the child is unable to communicate; therefore, psychopathology is more likely to be indirectly expressed by behaviour similar to that seen normally in younger children. For example, depression may be manifest as irritability, anxiety displayed by rocking or aggression and auditory hallucinations inferred from distressed covering of ears or self-injury.[7] Diagnosis is more speculative when the level of ID is more severe because the expression of emotions and behaviour is more atypical hence there is a greater use of unclassified or organic brain syndrome diagnoses. The cognitive subtest profile may also assist diagnosis. For example, children with autism usually perform better on visuo-motor tasks compared to verbal, imitation, and social comprehension tasks and therefore communicate and learn better if information is presented visually. The discovery of inattention and working memory deficits might help to confirm a diagnosis of ADHD.

(b) Temperament

As for the general population, difficult temperamental characteristics such as high levels of emotionality and activity and poor sociability, increase the risk of emotional and behavioural disorders, particularly in boys with mild ID. A difficult temperament might be enduring but improved parental understanding and management skills improve adaptation and reduce disturbed behaviour.

(c) Medical issues

A medical assessment is necessary, both to establish the cause of the ID, if known, and to determine if any medical conditions might be contributing to the emotional and behavioural problem. ID is associated with an increased risk of poor health in general, of brain disorders such as epilepsy (e.g. affecting 20 per cent of children with autism) and of medical complications associated with known causes of ID, such as cardiac and bowel abnormalities in Down syndrome, sensory impairments and deafness in Rubella embryopathy and the neuro-cutaneous brain lesions of tuberous sclerosis which are associated with tic disorder, autistic symptoms, and psychosis.[10] Disturbed behaviour might be the only manifestation of illnesses such as migraine, dental caries, and otitismedia in children with ID who are unable to talk about their pain. Psychoactive drugs are overprescribed in children with ID and their side effects are a well-recognized cause of behavioural and emotional disturbance and paradoxical effects. For example neuroleptic drugs may produce drowsiness, akathisia, and dystonic reactions. Irritability, anxiety, mood disturbance, and tics can be unacceptable side effects of stimulant medication. When prescribing drugs it is essential to systematically record behaviour and monitor side effects to confirm that the drug has a beneficial effect on target symptoms.[9]

(d) Behavioural phenotype

Specific genetic causes of ID often have characteristic patterns of psychopathology of relevance to diagnosis, treatment, and research (see Table 10.5.1.1).[11,12]

(e) Social and family influences

Children with ID are more likely than other children to experience adverse events such as poverty, socio-economic disadvantage, respite care and institutional care, rejection, social exclusion, teasing, school adjustment problems, abuse and neglect.[13] Their limited cognitive ability to comprehend adverse experiences may compromise adaptation. Parental stress, grief, guilt, and mental health problems and poor socio-economic circumstances are factors which are likely to adversely affect attachment and the quality of family care and aggravate child psychopathology.[14] In turn, behaviour problems, communication difficulties and lack of social responsiveness, for example in children with autism, predict maternal stress and mental health problems and placement of the child in out of home care.[15] Cultural responses, expectations, and attitudes may also influence parenting practices and the nature of care provided to children with ID. Observation and assessment of the quality of care, adverse events, parental mental health, family stress, resources, and community support is necessary to understand their contribution to psychopathology and implications for management. These factors are listed in AXIS V of the draft ICD-10 guide for mental retardation.[5]

Specific psychopathological disorders in children with ID

Behavioural disorders

(a) Attention-deficit hyperactivity disorder

Diagnosis of ADHD in children with ID is relatively straightforward because the DSM-IV TR[7] and the ICD[5] criteria are based on observable behaviour such as distractibility and fidgeting. This observed behaviour must also be 'inconsistent' with the child's developmental level. For example, the attention span of a 9-year-old child with a moderate ID would need to be less than that of a typical 3-year-old. Symptoms of inattention and hyperactivity

Table 10.5.1.1 Behaviour phenotypes

Syndrome	Genetics	Behavioural phenotype
Down	Trisomy 21 (1 in 800 live births)	Range of ID but usually moderate to severe Relatively lower rates of psychopathology (20–30%) Childhood: oppositional, attention-deficit problems Young adult: affective disorder, early-onset dementia
Fragile X	Expansion of CGG trinucleotide sequence at Xq27.3	Mild to moderate ID Verbal IQ > Performance IQ Shy, gaze avoidant, anxious inattentive, hyperactive, schizotypal disorder (females) 5–10% have autistic disorder but most responsive to social cues and form attachments Behaviour may settle with age
Prader–Willi	Paternal deletion long arm chromosome 15q 11–13 (70%) or maternal disomy chromosome 15 (25%) or a mutation	Mild/borderline ID Hyperphagia and food obsession Mild obesity, serious psychopathology (50%+) OCD (e.g. questioning, cleanliness) impulsivity, aggression, defiance, skin picking In adolescents anxiety, depression, psychosis (with maternal disomy)
Smith Magenis	Chromosome deletion at 17p 11.2	Moderate ID Severe psychopathology: hyperactive, impulsive, aggressive, insomnia, stereotypic movements (e.g. self-hugging) self-injury (e.g. nail pulling, head banging)
Williams	Micro-deletion on chromosome 7q 11.23 (elastin gene)	Moderate ID Visuo-spatial/motor deficits but recognize facial features, loquacious with stereotypic phrases Children: endearing, 'elfin-face', irrepressible, and affectionate, hyperacusis, phobias, anxiety, inattention, insomnia

must be present in at least two settings such as at home and at school. Symptoms of inattention and hyperactivity might also occur in reaction to stress such as bullying at school, but these adjustment disorders usually respond to psychosocial intervention and do not require medication. Anxiety and oppositional-defiant disorder are common comorbid conditions which need to be considered in a management plan.

Conduct disorder and oppositional-defiant disorder

Disruptive, aggressive, oppositional, and antisocial behaviours are problems in about 30 per cent of young people with mild/borderline ID, particularly males.[2] These children usually have language and learning difficulties and are likely to have experienced inconsistent care and sociocultural deprivation. The diagnosis requires a consideration of both the developmental age and the context. For example a 12-year-old young person with moderate ID might steal from a shop on the demand of a classmate without having a sufficient understanding of social rules and the rights of others. These diagnoses are not applicable in non-verbal children with more severe ID where, for example, aggressive behaviour might be the only means of communicating that an experience is stressful.

(a) Tic disorders (Tourette's disorder) and stereotypic movement disorder

Tics are sudden rapid non-rhythmic recurrent motor movements or vocalizations in response to an irresistible urge that can be delayed.[7] There is a comorbid association with autism or disruptive behaviour. Tics may emerge or deteriorate in a child with ADHD treated with stimulant medication. Relatively small doses of haloperidol or pimozide are often an effective treatment.[16] Stereotypic movements are persistent, driven, non-functional, complex motor behaviours that are differentiated from tics because they appear to be intentional. They occur in 2–3 per cent of children with more severe ID. These behaviours, such as self-biting, can cause serious injury and significantly interfere with daily activities, for example by rocking. They may be self-stimulating, occurring when the child is unoccupied, or might have communicative intent, for example to avoid an activity. The production of endogenous opioids might act to maintain the behaviours.

Emotional disorders

(a) Anxiety

Clinically significant symptoms of anxiety affect 10–12 per cent of both boys and girls with ID, compared to a prevalence in normal children of 2–5 per cent affecting twice as many females as males.[2] Fears are a common symptom and in children with ID are likely to be similar to the simple fears of young children in general, such as fear of the dark, loud noises, insects, and animals. Separation anxiety, typically seen in young children, can persist in older children with ID and is often complicated by fears, for example of school or other children. Children with ID are at high risk of suffering stressful experiences such as being placed in care or suffering physical and sexual maltreatment and neglect and have a limited capacity to understand stressful experiences.[17] Therefore, they are vulnerable to develop post-traumatic stress disorder (PTSD). Symptoms of PTSD are usually manifest as disturbed behaviours seen typically in traumatized young children such as

repetitive play, behaviours which re-enact the trauma, nightmares, withdrawal, increased startle response, and hyper-vigilance.[7] PTSD in children with ID is underdiagnosed and research on its phenomenology is required. The diagnosis of obsessive-compulsive disorder in young people with ID is problematic because they may not have sufficient language to describe persistent thoughts and their attempts to suppress these thoughts. They also do not recognize that their compulsive behaviour is unreasonable. Stereotypic self-injurious behaviours might be regarded as evidence for an OCD, but these behaviours do not usually respond to treatments for anxiety. Some drugs that are an effective treatment of OCD in young people in the general community, such as sertraline and clomipramine, are used for the treatment of anxious compulsive behaviours in young people with ID but their efficacy has not been investigated.

(b) Mood disorder

Depression in young people with ID is more prevalent than in other children.[2] Adolescents with moderate to severe levels of ID, provided they have some language, are able to reliably report sad feelings, but the diagnosis is confirmed by the presence of behavioural symptoms such as irritability, loss of interest in usual activities, loss of appetite and weight loss, sleep disturbance, crying, and withdrawn and regressed behaviours (e.g. rocking).[7] A daily carer completed record of mood and activity may reveal a pattern of cycling bipolar or unipolar mood disorder and is a useful record to document response to treatment.[9]

Pervasive developmental disorders

About 75 per cent of children with autism have ID. There is also an increased association with epilepsy, tuberous sclerosis, congenital rubella syndrome, and phenylketonuria and comorbidity with ADHD, OCD, anxiety, depression, and tic disorder. These comorbid symptoms might be epiphenomena of autism related to fronto-striatal dysfunction, but it is helpful to the child and family to identify any comorbid symptoms and to treat them appropriately.[18] Children with ID also have delayed language and may have stereotypic behaviours and a limited range of interests, but they can be differentiated from those who also have autism because they try to communicate, use gesture and imitation, have reciprocal play and respond with emotion in a manner appropriate to their developmental level.

Principles of management

Effective management begins with a multiaxial diagnostic formulation based on the DSM or ICD which describes the child's psychopathological disorder, cognitive profile, temperamental characteristics, genetic and associated medical conditions, level of adaptive functioning, and the family and sociocultural context. The delivery of effective management usually requires the involvement of a multidisciplinary team with a clear definition of roles and regular communication between professionals, parents, and teachers, for example at a case conference. The involvement of parents as partners in speech, physio, and behaviour therapy improves outcome and facilitates treatment compliance. Parent education and skills training reduces parental stress and improves parental mental health and child behaviour.[19] Family therapy may help reduce family conflict and improve communication and child management, but outcome research is required.

Psychological treatment

Effective behaviour modification techniques based on operant conditioning principles teach positive socially adaptive behaviours and reduce difficult behaviours.[20,21] Antecedent triggers, the functions of the behaviour and any rewarding consequences which reinforce the behaviour are identified. Intervention might focus on changing the antecedent events, for example by removal of an upsetting sound. Replacement of the disruptive behaviour might be achieved by rewarding alternative appropriate behaviour or by teaching a new behaviour which improves communication such as the use of pictures to communicate a need. A further approach which may help extinguish difficult behaviour is a non-aversive modification of the usual response to the behaviour such as moving away from a screaming child instead of paying them attention. Counselling and cognitive behavioural therapy including relaxation exercises, which is modified to take into account the developmental level and language ability of the child with ID might reduce anxiety and depression, but its effectiveness requires further research.

Pharmacotherapy

Drugs, if used to treat specific symptoms, should be part of a psychosocial and educational management plan, which includes informed parent/carer consent and participation. Evidence for the efficacy of psychotropic drugs in children with ID is limited,[16] but children with ID are prescribed or even overprescribed drugs on the basis of evidence for their use in either adults with ID or in normally developing children. These drugs include stimulant medication, atomoxetine, clonidine, neuroleptic drugs, and imipramine for ADHD; clomipramine and selective serotonin reuptake inhibitors (SSRIs) for OCD and stereotypic self-injurious behaviour; SSRIs for depression; neuroleptic drugs for tic disorders. Rigorous empirical studies have failed to demonstrate clear benefit for the use of 'typical' neuroleptic medication such as chlorpromazine in the treatment of disruptive behaviour. There is empirical evidence that low doses of haloperidol or risperidone (0.02–0.06 mg/kg/day) are an effective treatment for disruptive stereotypic behaviour.[22] Problematic side effects, particularly for haloperidol, are drowsiness, akathisia, dystonia, and tardive dyskinesia. Weight gain, prolactinemia, and metabolic disturbances such as diabetes are serious side effects of risperidone. Anticonvulsants used as mood stabilizers (sodium valproate, carbamazapine, and lamotrigine), lithium, β-blockers, and buspirone have been shown, mostly in open trials, to reduce self-injurious and episodic aggressive behaviours. Opiate antagonists (naloxone and naltrexone) may reduce self-injurious behaviour. Difficult sexual behaviour in adolescent males with ID can be treated with testosterone antagonists to reduce libido. In many countries this treatment requires approval and monitoring by an independent committee. The prescription of psychotropic drugs requires regular follow-up to monitor compliance, side effects, and response to treatment using behaviour observations and carer completed symptom checklist such as the DBC[9] or NCBRF.[8]

Early intervention

There is growing evidence that broad-based early intervention for young children with developmental delay, such as those with

autism, promotes adaptive behaviour and skill development. The components of effective early intervention include:

1 Parent education and skills training and the management of parent mental health problems such as depression.[19]

2 Regular medical review and treatment of associated conditions such as epilepsy and any inter-current illness or psychopathological disorder, which might compromise behaviour.[10]

3 A structured behaviour management and social and communication skills programme.[21]

4 Family support, respite care, home help, and holiday programmes.

5 Speech, occupational, and physiotherapy to develop communication, sensory, motor, and play skills.

6 Assisted education and socialization at preschool and school.

Conclusions

Children with ID often suffer the added handicap of emotional and behavioural disorder which seriously compromises their adjustment and causes significant extra burden and cost for their parents and the community. A comprehensive biopsychosocial assessment of the child and family provides the context for understanding psychopathological symptoms and the basis for a best practice management plan incorporating psychological, educational, family, and perhaps pharmacological interventions.

Further information

Bouras, N. and Holt, G. (eds.) (2007). *Psychiatric and behavioural disorders in intellectual and developmental disabilities* (2nd edn). Cambridge University Press, London.

Reiss, S. and Aman, M.G. (1998). *Psychotropic medication and developmental disabilities: the international consensus handbook*. Ohio State University, Columbus.

Therapeutic Guidelines. (2005). *Management guidelines: developmental disability* (Version 2). Therapeutic Guidelines Ltd., North Melbourne.

www.cddh.monash.org (provides an educational programme in the psychiatry of ID).

www.iassid.org (review of current research and information exchange on the psychiatry of ID).

References

1. Emerson, E. (2003). Prevalence of psychiatric disorders in children and adolescents with and without intellectual disability. *Journal of Intellectual Disability Research*, **47**, 51–8.
2. Tonge, B.J. and Einfeld, S.L. (2003). Psychopathology and intellectual disability: the Australian child to adult longitudinal study. *International Review of Research in Mental Retardation*, **26**, 61–91.
3. Einfeld, S.L. and Tonge, B.J. (1999). Observations on the use of the ICD-10 guide for mental retardation. *Journal of Intellectual Disability Research*, **43**, 408–12.
4. American Psychiatric Association. (2002). *Diagnostic and statistical manual of mental disorders* (4th edn, text revision). American Psychiatric Association, Washington, DC.
5. World Health Organization. (1996). *The ICD-10 guide for mental retardation*. World Health Organization, Geneva.
6. Royal College of Psychiatrists. (2001). *DC-LD (Diagnostic criteria for psychiatric disorders for use with adults with learning disorders/mental retardation)*. Gaskell, London.
7. Fletcher, R., Loschen, E., Stavrakaki, C., *et al.* (2007). *Diagnostic Manual – Intellectual Disability (DM-ID): a textbook of diagnosis of mental disorders in persons with intellectual disability*. NADD Press, Kingston, New York.
8. Aman, M.G., Tasse, M.J., Rojahn, J., *et al.* (1996). The Nisonger CBRF: a child behaviour rating form for children with developmental disabilities. *Research in Developmental Disabilities*, **17**, 41–57.
9. Einfeld, S.L. and Tonge, B.J. (2002). *Manual for the developmental behaviour checklist* (2nd edn). Centre for Developmental Psychiatry & Psychology, Monash University, Melbourne.
10. Therapeutic Guidelines. (2005). *Management guidelines: developmental disability* (Version 2). Therapeutic Guidelines Ltd., North Melbourne.
11. Dykens, E.M. (1999). Direct effects of genetic mental retardation syndromes: maladaptive behaviour and psychopathology. *International Review of Research on Mental Retardation*, **22**, 1–26.
12. Einfeld, S.L., Tonge, B.J., and Rees, V.W. (2001). Longitudinal course of behavioural and emotional problems in Williams syndrome. *American Journal of Mental Retardation*, **106**, 73–81.
13. Emerson, E., Graham, H., and Hatton, C. (2006). The measurement of poverty and socioeconomic position in research involving people with intellectual disability. *International Review of Research in Mental Retardation*, **32**, 77–108.
14. Baker, B.L., Blacker, J., Crnic, K.A., *et al.* (2002). Behaviour problems and parenting stress in families of three-year-old children with and without developmental delay. *American Journal of Mental Retardation*, **107**, 433–44.
15. Llewellyn, D., McConnell, D., Thompson, K., *et al.* (2005). Out-of-home placement of school age children with disabilities. *Journal of Applied Research in Intellectual Disability*, **18**, 1–16.
16. Reiss, S. and Aman, M.G. (1998). *Psychotropic medication and developmental disabilities: the international consensus handbook*. Ohio State University, Columbus, OH.
17. Hatton, C. and Emmerson, E. (2004). The relationship between life events and psychopathology amongst children with intellectual disabilities. *Journal of Applied Research in Intellectual Disability*, **17**, 109–17.
18. Tonge, B.J. and Rinehart, N.J. (2007). Autism and attention deficit/hyperactivity disorder. In *Neurology and clinical neuroscience* (ed. A.H.V. Shapira), pp. 129–30. Mosby, Elsevier, Philadelphia.
19. Tonge, B.J., Brereton, A.V., Kiomall, *et al.* (2006). Effects on parental mental health of an education and skills training program for parents with young children with autism: a randomized controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, **45**, 561–69.
20. Lucyshyn, J.M., Dunlap, G., and Albin, R.W. (eds.) (2002). *Families and positive behaviour support*. Brookes, Baltimore.
21. Emerson, E. (2001). *Challenging behaviour. Analysis and intervention in people with severe intellectual disabilities* (2nd edn). Cambridge University Press, Cambridge.
22. Aman, M.G., DeSmedt, G., Derivan, A., *et al.* and the Risperidone Disruptive Behaviour Study Group. (2002). Risperidone treatment of children with disruptive behaviour symptoms and sub-average IQ: a double blind, placebo-controlled study. *The American Journal of Psychiatry*, **159**, 1337–46.

10.5.2 Psychiatric and behaviour disorders among adult persons with intellectual disability

Anton Došen

Introduction

The behavioural and emotional difficulties that are experienced by adults with intellectual disability (ID) have been regarded for many years as manifestations of their intellectual deficits and maladaptive learning. The awareness that these persons may also suffer from mental illness was a notion that came into being in the mid-nineteenth century.

Nevertheless, the psychiatric problems of individuals with ID were continually ignored in the first half of the twentieth century. In the past three decades, the flourishing normalization philosophy has highlighted the psychiatric problems of this population once again and rekindled the interest of practitioners, scientists, and service providers. Systematic studies have been performed which indicate that the full spectrum of psychiatric disorders as we know them today can be identified among the persons with ID. Moreover, it is probable that they may be prone to a psychopathology that is determined by the specifics of their biological and psychosocial being.

Clinical features

Features affecting presentation

Studies indicate that the types of psychiatric symptoms and syndromes that are observed among persons with borderline and mild ID are similar to those encountered among the population in general. Amongst individuals with moderate and severe ID, however, the presentation of mental illness may be less typical, and the diagnosis more difficult to establish.

Sovner and Hurley[1] have categorized four factors which may influence the presentation of mental illness among the persons with ID: intellectual distortion (impaired ability to conceptualize feelings and to communicate them to others), psychological masking (lack of usual richness of the symptomatology found in general population), cognitive disintegration (inclination to become disorganized and to exhibit regressive behaviour), and baseline exaggeration (increase of pre-existing maladaptive behaviour by emotional stress or mental illness).

Hucker *et al.*[2] pointed to a generally banal symptomatology encountered among these individuals, often accompanied by regression to a child-like state of dependency and hysterical features. Behavioural disturbances were often more important than symptomatic complaints as indicators of psychiatric disorders.

Apparently, the lower the IQ, the more the symptoms of mental illness tend to lose their specificity, or take on a different meaning than is the case with the intellectually normal population. This, undoubtedly, makes it difficult to establish a confident diagnosis of mental illness in people with severe ID.[3]

The following is a concise survey of the striking clinical features of mental illness and behaviour problems as they occur among adult persons with ID.

Mental illness

(a) Psychosis

Among persons with mild ID, classical clinical features are present in psychotic states. The symptoms tend to be florid but banal. In schizophrenia, for example, there is a high incidence of delusions and hallucinations, which reflect the limited experiences, naive and wishful thinking, interests, and social horizon of the patient. Ideas, that have been influenced by radio, television, etc., are found frequently. Catatonic features with odd postures and slowness are common. Impulsive, aggressive, auto-aggressive, and bizarre behaviours may dominate the clinical picture. In the chronic-phase apathy, lack of motivation and social withdrawal are common.[4,5] Increase of 'negative' schizophrenic symptoms and decrease of functional abilities were observed in the group with ID when compared with the group from general population with schizophrenia.[6]

Establishing the diagnosis of schizophrenia in persons with more severe ID can be a difficult task because of verbal communicative difficulties. Reid[7] considers it to be impossible to establish such a diagnosis among persons who only communicate non-verbally. However, the problem of establishing the diagnosis does not mean that psychotic conditions do not occur in these individuals. To the contrary, it is likely that different sorts of psychosis occur in this population more frequently than in general population (see Chapter 4.5.2).

Short-term psychotic states (lasting several days or weeks), usually beginning rather suddenly after a stressful event, are found relatively often among adolescents and young adults with ID. Early Dutch psychiatric literature refers to these states as 'debility psychosis'. The symptoms may be heterogeneous, and remission is usually complete. These persons usually revert fully to the premorbid level of functioning. However, recurrence is frequent.

Establishing the diagnosis of atypical psychosis (contrary to diagnosing schizophrenia) is, for an experienced practitioner, possible even with patients who have no language development and are at a severe level of ID. The primary symptoms are changes in interactional patterns, changes of posture and movement, odd and bizarre behaviour, disturbances of the physiological functions, expression of emotional tension (e.g. anxiety, irritability), aggression, and self-injuring behaviour.

(b) Major depression and bipolar disorders

Depressed mood and vegetative symptoms are the most striking symptoms, even though complaints of depression are not always expressed. A depressive mood often is not verbalized, particularly among individuals at a lower intelligence level, but may well be observable. Similarly, the elevation of mood in mania is usually not expressed verbally either. Atypical features such as regression to child-like dependency, incontinence, loss of social skills, and hysterical symptoms such as pseudo-fits and paralysis may mask classical symptomatology. In persons with a more severe disability, depression should be suspected where there is a change or onset of behaviour problems like stereotypic behaviour, tantrums, aggression, and self-injuring behaviour.[4,8] Catatonic features and visual hallucinations, particularly among persons at lower intelligence levels, have also been reported.[2,9,10] The atypical symptomatology among persons on lower developmental levels may require modification of standard diagnostic criteria.

Aggressive behaviour was observed in 40 per cent of the depressed subjects.[11] Self-injuring behaviour has often been reported as well.

Suicidal behaviour has hardly ever been studied in this population, and suicidality is very rare among the more severely handicapped. This symptom is, however, not rare among depressive patients at a mild level of ID.

Some investigators report the relatively frequent occurrence of rapid-cycling affective disorder,[12] particularly in persons with more severe disability. Episodes of particular mood or of an undifferentiated mood-like dysphoria or irritability have a short duration, and may be expressed in terms of days or weeks.[9] Researchers assume that these disorders, in persons with ID are often related to organic brain disorders, that is, metabolic, neuroendocrine, and other neurological disorders.

In mixed bipolar disorder, there is either the simultaneous presence of manic and depressive features, or these features follow each other rapidly. Schizoaffective psychoses are also described among these individuals.

(c) Dysthymic disorder

Dysthymia is a relatively common disorder among persons with mild and moderate ID.[9,13] Nevertheless, publications on this disorder are rare. The symptomatology includes loss of energy and interest, negative self-image, feelings of helplessness, anxiety, and significant behavioural problems such as irritability, anger, destructibility, and aggression. The disorder is often related to a specific stress, for example, termination of an affective relationship, change in the surroundings, hospitalization, etc. Chronic states, dating back to the childhood or the teens, possibly caused by chronic overdemanding, social deprivation, or repeated abuse, may be interrupted by episodes of major depression, usually elicited by acute stress (so-called 'double depression', see Chapter 4.5.3). Došen and co-workers found this disorder relatively frequent in adolescents and young adults with ID and called it 'developmental depression'.[10, 14] Social interactional problems, poor social skills, and difficulties related to emotional development are considered to be predisposing factors for this disorder.[14]

(d) Anxiety disorders

The most commonly reported anxiety disorders are simple phobia, social phobia, and generalized anxiety disorder.[4] It seems that adults with ID have fears similar to those of children who are at the same mental age: fear of separation, fear of natural events, fear of injury, and fear of animals. The anxieties and fears are probably related to the traumatic events and cumulative failure experiences that these persons have. The presentation may be through behaviour problems, irritability, problems with sleeping, or somatic complaints.[5] In a panic disorder, a sudden onset, blackouts, aggression, sweating, and shaking may be observable. The obsessive–compulsive disorder may be difficult to diagnose in persons with ID because they do not resist against such feelings and the anxiety is often absent. According to some authors,[4] the diagnosis can be established with the emphasis being on the externally observable behavioural components, despite of absence of some internal states like anxiety and resistance. Post-traumatic stress disorder is likely to occur in this population, following relatively less severe stress than among general population. The diagnosis in those who are unable to communicate their experiences should be based on changes in a person's behaviour, mood, and level of functioning following a traumatic event.[4]

(e) Autism spectrum disorder

Autism spectrum disorders have been estimated to be present in 10 per cent of persons with mild ID and 40 per cent of those with severe ID, and account for a large proportion of behaviour disorders. It also appears that mental illness occurs frequently as a secondary disorder among these individuals.

A possible relationship between affective illness and pervasive developmental disorder has been suggested by various investigators[12]; however, this phenomenon has been examined insufficiently and is clearly an area that future research can be directed to. In clinical practice, we have encountered a number of cases of pervasive developmental disorder together with secondary atypical psychosis. Inexperienced practitioners are inclined to diagnose schizophrenia in such cases. However, thorough developmental history will reveal sufficient information to make diagnostic differentiation possible. Other problematic behaviours such as anxious, aggressive, auto-aggressive, or disruptive behaviour are frequently found among persons who have an autism spectrum disorder and ID.[15] In our opinion, these behaviours should be seen as being secondary disorders instead of as part of the autistic disorder.

(f) Dementia

In individuals with dementia, the typical features such as memory impairment, personality change, loss of social skills, and deterioration in habits are always present. Behavioural problems may be the most obvious manifestation. Nocturnal confusion, transient psychotic episodes, and late-onset epilepsy should always alert one to the possibility of a dementing illness in the ageing person with ID. Memory loss is generally difficult to identify in the early stages, but becomes more obvious as the illness progresses. Medical risk factors include a history of hypertension, ischaemic episodes, neurological symptoms, organic brain damage, and a family history of dementia. Dementia Alzheimer type in persons with Down syndrome presents a similar picture and is usually associated with generalized premature ageing.

Behaviour disorders—challenging behaviour

Behaviour disorders including aggression, self-injury, destructiveness, and disruptive, maladaptive, and antisocial behaviour occur commonly among persons with ID. Such behaviour has recently been called challenging behaviour, which emphasizes the need for appropriate care and supervision. These disorders are usually associated with severe ID, but can also occur in individuals who are at a moderate and mild level of ID.

Various attempts have been made to distinguish between behaviour disorders and psychiatric illness in these individuals. Gardner and co-workers[16] have proposed a bio-psycho-social diagnostic approach, which takes account of the multiple factors underlying and maintaining the behaviour disturbances of a particular individual. They point out that behaviour disorders with a neuropsychiatric and organic basis can still acquire a functional component if they are being reinforced by the environment or are of value to the individual. Another approach is from the developmental perspective,[3,17] viewing behaviour disorders as the result of a lack of real understanding of the person's developmental aspects and interactional problems.

(a) Aggressive behaviour

Aggressive behaviour is a common problem among persons with ID. The symptom of aggression is often a feature of the psychosis, depression, or antisocial personality disorder, and is often described in genetic disorders such as the fragile X, Prader–Willi, and

Klinefelter syndromes. Learned aggression through the imitation of aggressive models or as a function of communication is also found relatively frequently among people with ID.

(b) Self-injurious behaviour

Self-injurious behaviour occurs more often among persons with moderate and severe ID (IQ < 50), beginning sometimes in toddler age and most frequently between the ages of 10 and 20. The occurrence of self-injurious behaviour is related to genetic and organic disturbances and adverse environmental and development conditions. Certain psychiatric disorders such as depression and psychosis may also elicit self-injurious behaviour.

(c) Offending behaviour

Owing to their behaviour problems, these individuals may become involved in activities, that bring them into conflict with the law. Insufficient understanding of their problems and needs may result in their not receiving the appropriate support from the social services. The typical offender with ID is, according to Day,[18] a young male functioning in the mild to borderline intellectual range, from a poor urban environment, with a history of psychosocial deprivation, behaviour problems, and personality disorder. The most common offences are acquisitive and technical, but sex offences and arson are considerably overrepresented.

Personality disorders

Various investigators have reported personality disorders among persons with ID.[19,20] The relevance of the concept of personality disorder, in particular with regard to the persons with more severe ID, has been questioned by a number of investigators. Apparently, in these persons, besides the problem of personality disorders, there is a problem of personality development. Zigler and colleagues[21] have explored personality traits thought to be particularly salient in determining the behaviour of individuals with ID. Levitas and Gilson[22] have stressed the importance of a crisis period during the process of personality development and the related psychosocial aspects. Other developmentally oriented authors[14,23] make a link between, on the one hand, the problematic processing of particular phase–specific aspects of emotional development and ego structuring, such as the achievement of secure attachment, an intercompetitive separation-individuation process, and the establishing of ego functions, and, on the other hand, the increased vulnerability of these individuals to particular psychiatric disorders such as depression, social withdrawal, disruptive behaviours, etc. Classification of personality problems within the existing diagnostic categories for personality disorders in general population is questionable and requires modification of particular diagnostic criteria.[4, 5] It is unlikely that personality disorders could be diagnosed in persons with severe/profound ID.

It appears that the main problem at the root of the personality disorders of individuals with ID pertains to an underdeveloped personality structure in relation to a delay in psychosocial development. One is then inclined to speak of an immature rather than a disturbed personality.

Diagnosis and classification

Diagnosis calls for a full and detailed history, careful observation of the patient, knowledge of the natural history of the illness, and the elimination of irrelevant factors (for further information about assessment, see Chapter 10.1).

Assessment

In complex cases, a period of inpatient observation or explorative treatment may be necessary. As full a history as possible of the current illness should be obtained from the patient, together with corroborative histories from the relatives and carers. The standardized format of psychiatric history should be supplemented by a detailed developmental history, a description of current social functioning, environmental circumstances, associated somatic disorders and physical disabilities, and the aetiology of the ID. Enquiries should be focused on behavioural changes such as sleep disturbance, loss of appetite, weight loss, lack of interest, bizarre behaviour, restlessness, anxiety, withdrawal, and any other deviations from customary behaviour. Precipitating factors such as stressful events and possible predispositions to reacting in a particular way in a particular situation should be explored. Full details of previous psychiatric illnesses suffered by the patient and the family history of mental illness should be obtained. Because of the general paucity of subjective complaints by persons with ID, the examiner must rely more on objective data regarding the patient's appearance, manner of communication, facial expression, evidence of hallucinations, posture, etc. If called for, direct observations should be made in as wide a range of settings as is necessary. To these ends, a video recording of the patient in his or her natural surroundings may provide important information about the interactional pattern of the patient and his or her surroundings. The psychiatric examination is usually supplemented by somatic, neurological, neurophysiological, biochemical, and psychological examinations. Assessment of the level of ID is crucial for diagnostic consideration. Currently, in psychological assessment, besides examination of cognitive functioning, the emphasis is on determination of personality development and the level of emotional development.[3]

Diagnosis

IQ assessment, personality development, emotional level, and measuring adaptive behaviour can provide extra background information that may be useful to the diagnosis. Specific tests, for example, of thought disorder in schizophrenia, may be helpful in establishing the diagnosis, but are not yet standardized for this population. Non-invasive neuroimaging techniques promise to be potentionally valuable diagnostic tools, particularly for non-verbal persons with severe ID in the future. Structured interview schedules and rating scales are being used increasingly in an attempt to improve diagnostic accuracy. Instruments developed for use with persons without ID rely heavily on the ability of the patient to describe subjective feelings and are thus of limited value. Diagnostic rating scales that are to be used with the persons with ID should, as far as possible, reflect behavioural rather than subjective components. An early attempt was made by Hucker and colleagues,[2] who published diagnostic criteria for mania, depression, and schizophrenia for use with persons with ID; these have been further refined by Sovner and Hurley[1] and by Menolascino and Weiler.[24] Recently, a number of scales have been developed specifically for use with this population, and different other instruments are in development. These scales were primarily developed for use as research instruments, and whilst they play an invaluable role in epidemiological studies and population screening as well as being useful for monitoring the response to treatment, they are of limited value in clinical practice and rarely, if ever, solve a diagnostic problem.

(a) Integrative psychiatric diagnosis

For clinical purposes, it has been proposed that more elaborate and expanded diagnostic formulations be made.[25] Such diagnosis should incorporate diagnostic categories as well as the onset mechanisms of the psychopathology, biological aspects, psychological functioning, milieu characteristics, life problems, psychosocial needs, and individual strengths. Diagnostics of this sort have been called integrative diagnosis and are an attempt to adapt conventional diagnostic criteria to the complex problems of individuals with ID.[3,14] During the diagnostic process, particular attention is paid to describing the onset mechanism of the psychopathological phenomenon by which the dynamics of different nosological factors come into scope and become more understandable to the direct carers and professional helpers. A better understanding of the processes involved in the mental illness is important to the treatment approach.

Classification

Numerous attempts have been made to apply the traditional psychiatric diagnostic categories of ICD-10 and DSM-IV to the psychopathology of persons with ID. Their applicability to these individuals has, however, been questioned.[1,2,25] Whilst the ICD and DSM criteria may be applied to people functioning in the mild to borderline ID ranges without alteration or with little modification, they become increasingly unreliable as the severity of intellectual disability increases. The limited communication skills of these persons make it very difficult to ascertain the presence of certain symptoms such as delusions and hallucinations. As the role of underlying organic brain damage expands, the phenomenology becomes increasingly more characterized by a range of atypical symptoms. The non-specific nature of behavioural disturbances further confounds diagnostic endeavours. Szymanski,[25] among others, has pointed out that behaviour disturbance is not a psychiatric condition but a symptom, and Reid[26] has drawn attention to the fact that behaviours, that would be deemed abnormal in people functioning in the average intellectual range may be developmentally appropriate to the mental age of a person with severe ID.

Other authors[3,17,27] suggest using the developmental perspective when attempting to understand and diagnose the psychiatric and behavioural disorders of the persons with ID. They point out that there are findings which suggest that there may be a relationship between certain developmental levels and syndromes and specific neuropsychiatric disorders (see below). Some investigators argue that syndromes should be empirically derived, and a 'dimensional' approach would be a more effective way to classify psychopathology than the categorical system.[28,29] Van Praag[30] proposes a 'functional psychopathology model' emphasizing functional problems of the CNS on the background of psychiatric disorders.

Not surprisingly, there have been calls for some modifications of existing DSM and ICD diagnostic criteria with the introduction of behavioural equivalents for some symptom criteria as well as for development of a broader taxonomy, which takes account of the atypical presentation of mental illness in this population.[31]

Epidemiology

Overall prevalence

Psychiatric disorder appears to be more common among the persons with ID than in the general population. Overall prevalence rates range from 20 per cent to 74 per cent,[19,25,32] depending on the diagnostic criteria employed, the type of disorder screened (whether or not behaviour disorders are included, for example), the nature of the sample (community or institution), the type of data collected (case note studies or new data), and the level of ID, ages, and gender of the populations studied. Higher rates of psychiatric disorder have been reported in some studies among the individuals with severe ID in comparison with the person on mild ID level. For further information concerning epidemiology, see Chapter 10.2.

Aetiology

Investigators of aetiology agree that the high prevalence of psychiatric disorders among persons with ID is related to a wide range of neurological, psychological, social, and personality risk factors including impaired genetic factors, delayed cognitive, emotional and social development, organic brain damage, communication problems, environmental problems, and family psychopathology. Alone or in combination, these factors increase the vulnerability of the person with ID to psychiatric and behavioural problems.

Theories

Achenbach and Zigler[33] have pointed out the importance of social incompetence as a factor playing a role in interactional and intrapsychic problems. This theoretical perspective has received support from studies in which discrepancies between self-image and the expectations of others have been shown to be a fertile breeding ground for the onset of psychopathology. Menolascino[34] emphasized the importance of neurophysiological and sociological developmental processes which may have a different timing and take a different course in these individuals, causing deviations from normal development; this is known as the biodevelopmental theory. Matson[35] proposed the biosocial theory, which hypothesizes that due to specific biological factors (neurological, biochemical, genetic, etc.) together with specific social factors (family interactions, culture, and other environmental variables), and specific psychological processes (cognitive development, personality variables), the psychopathology of persons with ID differs in a number of ways from that of persons without ID.

Tanguay,[36] Došen,[3,14,17] and Gaedt and Gärtner[37] have applied the developmental approach to the understanding of the symptoms of psychopathology and to the psychiatric diagnostics in these individuals. These authors based their approach on Piaget's stages of cognitive development as well as on Mahler's and Bowlby's models of psychosocial development. Parallels could be drawn between the symptoms of psychopathology of children without ID at a particular chronological age and individuals with ID at the same developmental age, which indicates that the developmental level may specifically affect the exterior features of mental illness. According to these authors, although developmentally disabled adults who suffer from mental illness resemble other adults in many ways, it is the developmental level and consequent differentiation of psychosocial life (e.g. affect differentiation, personality structuring, moral development), that may be decisive for the symptoms linked to the disorder. For example, because of their underdeveloped psychosocial life, adults with severe ID display little differentiation in their symptomatology. They exhibit agitation, aggression, self-injurious behaviour, or other disruptive behaviours, which in other people may be indicative of various

underlying mental illnesses such as anxiety disorder, depression, or psychosis. Similarly, a toddler may have 'tantrums' when frustrated, anxious, or distressed by somatic pain or separation. Recently, problems of attachment development in these individuals and their vulnerability to stress have been related by some authors to the behaviour problems and psychopathological features.[38,39]

Behavioural phenotypes in relation to mental illness

Recently, the concept of a behavioural phenotype has been introduced as an attempt to assess the interrelationship between specific behaviours and genetic disorders (see Chapter 10.4). In addition to specific behaviours, an increased prevalence of psychiatric syndromes have been reported in association with particular genetic and other syndromes. The possible link between a behavioural phenotype and the particular psychiatric disorder of a person with ID is a highly challenging issue for investigators researching this field. For further information.

Down syndrome has always been associated with specific patterns of language, and cognitive and social development. In most studies, these persons were found to exhibit muted affect and have deficient language development, and yet they showed particular strengths in socialization. The tendency for such individuals to develop Alzheimer dementia has often been described. The occurrence of affective disorder among Down syndrome subjects had attracted the attention of scientists recently.[40,41] Diurnal mood variations, speech reduction, and an increase in aggression are the commonly reported symptoms, which are indicative of depression among persons with Down syndrome. The onset of mania in this syndrome has been disputed. However, cases of mania have been reported later in life among those having Down syndrome.

Persons with fragile X syndrome have often been described as having autistic-like social impairments. However, among these individuals, social anxiety is more characteristic than social indifference.[42] In addition, attention deficit, stereotyped behaviour, self-injury, and hyperactivity are common.

In Prader–Willi syndrome, hyperphagia and obsessive–compulsive and aggressive behaviour are common. The occurrence of brief psychotic episodes with heterogeneous symptoms have been described among these persons.

Different behavioural phenotypes and psychiatric disorders have also been described for other syndromes. For example, anxiety disorder, sleep disorder, and hyperactivity are common in Williams syndrome, depression is often found in Klinefelter's syndrome, and self-injurious behaviour occurs regularly in Cornelia de Lange syndrome.

Genetic and environmental factors and onset mechanisms

It should be stressed, however, that the very same behaviours may be found among individuals with genetic as well as with idiopathic ID, which probably means that these behaviours are not only the product of genetic factors, and that other factors may be the cause as well. Among other things, the process of psychosocial development and interactional patterns with the surroundings can affect the onset of a particular behaviour. Apparently, genetic characteristics can influence the psychosocial development, which then can be decisive for the onset of particular interactional patterns and particular behaviours (so called 'heightened probability', according to Dykens *et al.*[43]). The question is: Do these behaviours play a role in the onset of the aforementioned psychiatric disorders associated with these genetically based syndromes, and if so, how do they do this? Or, does a genetic disorder have a more direct role in the onset of a particular psychiatric disorder?

For a better understanding of how specific psychiatric disorders evolve among these individuals, it is necessary to make a scheme of the onset mechanisms involved in each psychiatric disorder. It is important to take account of the genetic disorder, the developmental processes, and the interaction patterns with the surroundings in which particular behaviours may play an important role. If one takes all these factors into consideration, it may be expected that in certain cases particular psychiatric disorders have a specific aetiology. An example is self-injurious behaviour, which is currently considered to be a 'challenging' behaviour, but, according to various professionals in this field, warrants being seen as a specific psychopathological phenomenon. However, there are also examples in which the connection between genetic and psychiatric disorder appears to be more direct. An example is the velocardiofacial or Shprintzen syndrome.[44,45]

Psychiatric problems in older persons with ID

In elderly persons with ID, besides dementia, other psychiatric disorders, like psychotic conditions, depression, and anxiety disorders, are common.[46] In a mixed community and institutionalized group of persons with ID older than 50 years, Patel and colleagues[47] found a prevalence of psychiatric disorders, mainly depression and anxiety in 11.4 per cent of the population. Other authors[48] ascertained that major psychiatric disorders could be found in 20 per cent of persons with ID aged 65 and above. Behavioural symptoms such as aggression and self-injury frequently accompany psychiatric problems among these individuals. Causes for frequent psychiatric problems in this population group are various; like changes in social milieu and living pattern, loss of beloved ones, social deprivation, and inactivation done to retirement from the work or previous activities and medical, in particular neurological, conditions at this age (see Chapter 9.2.1). Frequently, psychiatric disorders in these individuals can be found as an accompaniment of a dementia process. Differentiation between the beginning stage of dementia and other psychiatric disorders can be a difficult task.[49] In these cases, a total bio-psycho-social picture of the involved person, including a detailed patient history, as well as an insight into the surroundings circumstances and in interactions is necessary for appropriate understanding of the patient's symptoms. Use of some assessment instruments can be helpful (see Chapter 10.1).

Management

In principle, the diagnosis of psychiatric disorders in persons with mild ID and good verbal skills does not differ much from diagnosing these disorders in persons with average cognitive skills.[50] However, in persons on a lower level, the recognition of mental disorders can be difficult, partly because of linguistic limitations and partly because of difficulty in distinguishing between behaviour problems (challenging behaviour) and symptoms of psychiatric illness. Communication problems often ask for management through another individual ('management by proxy'), which means

that carers are relied upon to give description of a person's behaviour. The diagnostician should know that the reliability and validity of such information may be problematical and, besides this information, should search for more objective sources, like information through non-verbal communication with the patient and through observation of the patient in different situations and interactions. Distinguishing between behaviour problems and psychiatric illness is a very important issue for adequate aid. It asks for a multi-disciplinary assessment approach in which experienced (specialized) professionals from different disciplines, participate forming together a multi-disciplinary team. These teams have expertise in both intellectual disability and mental health, and provide direct services to patients and carers. Through the assessment of a specialist team, an insight is being obtained whether presenting behaviours of involved individual are the results of an organic condition, a psychiatric disorder, environmental influences, or a combination of these factors. The clinical symptoms must be viewed in a broad context of a patient's functioning, considering his/her deficits, strengths, and relevant bio-psycho-social factors. For example, a disruptive behaviour can be a symptom of a psychiatric disorder in one person, whilst in another person, and in other environmental circumstances, it may be a means of communication, serving as a vehicle for obtaining a caregiver's attention or avoiding an unwanted task. Delineation of a subtype of an established psychiatric disorder asks for the specialist knowledge of the diagnostician.

The specialist teams should be based locally, providing inpatient care as well as outpatient and community-based interventions (see Chapter 10.9). Involvement of parents and carers as partners in the management plan can have more advantages and is being recommended.

The diagnostic evaluation may require significantly more time than the evaluation of persons of normal cognition. An accurate diagnosis which integrates all assessment results of the multidisciplinary team members should serve as a starting point for the treatment.[51]

Conclusion

Persons with ID are more prone to psychiatric disorders and behavioural problems than the general population. The symptoms of their biological disorders, developmental processes, and interaction patterns may give atypical clinical pictures, particularly in individuals with moderate and severe ID. It is probable that certain disorders are specifically associated with or even unique to the ID. Traditional nosological classifications do not adequately accommodate the phenomenology of psychiatric disorders in this population. A broader taxonomy, that takes account of the atypical presentation of psychopathology in this population is necessary.

Further information

Bouras, N. and Holt, G. (2007). *Psychiatric and behavioural disorders in intellectual and developmental disabilities*. Cambridge University Press, Cambridge.

Deb, S., Matthews, T., Holt, G., *et al.* (2001). *Practice guidelines for the assessment and diagnosis of mental health problems in adults with intellectual disability*. Pavilion, Brighton.

Došen, A. and Day, K. (2001). *Treating mental illness and behaviour disorders in children and adults with mental retardation*. American Psychiatric Press, Washington.

Došen, A., Gardner, W.I., Griffiths, D.M., *et al.* (2007). *Practice guidelines and principles: assessment, diagnosis, treatment and related support services for persons with intellectual disabilities and problem behaviour*. Centre for Consultation and Expertise, Gouda.

Fletcher, R., Loschen, E., Stavrakaki, C., *et al.* (2007). *Diagnostic manua—Intellectual disability*. NADD Press, Kingston, New York.

Royal College of Psychiatrists. (2001). DC-LD: *diagnostic criteria for psychiatric disorders for use with adults with learning disabilities/mental retardation*. Gaskell Press, London.

References

1. Sovner, R. and Hurley, A. (1990). Assessment tools which facilitate psychiatric evaluation of treatment. *The Habilitative Mental Health Care Newsletter*, **9**, 11.
2. Hucker, S.J., Day, K.A., George, S., *et al.* (1979). Psychoses in mentally handicapped adults. In *Psychiatric illness and mental handicap* (eds. F.E. James and R.P. Snaith), pp. 27–35. Gaskell Press, London.
3. Došen, A. (2005b). Applying the developmental perspective in psychiatric assessment and diagnosis of persons with intellectual disability: part II—diagnosis. *Journal of Intellectual Disability Research*, **49**, 9–15.
4. Deb, S., Matthews, T., Holt, G., *et al.* (2001). *Practice guidelines for the assessment and diagnosis of mental health problems in adults with intellectual disability*. Pavilion, Brighton.
5. Royal College of Psychiatrists. (2001). *DC-LD: diagnostic criteria for psychiatric disorders for use with adults with learning disabilities/mental retardation*. Gaskell Press, London.
6. Bouras, N., Martin, G., Leese, M., *et al.* (2004). Schizophrenia-spectrum psychosis in people with and without intellectual disability. *Journal of Intellectual Disability Research*, **48**, 548–55.
7. Reid, A. (1993). Schizophrenic and paranoid syndromes in persons with mental retardation. In *Mental health aspects of mental retardation* (eds. R. Fletscher and A. Došen), pp. 98–110. Lexington Books, New York.
8. Evans, K.M., Cotton, M.M., Einfeld, S.L., *et al.* (1999). Assessment of depression in adults with severe or profound intellectual disability. *Journal of Intellectual Disability Research*, **24**, 147–60.
9. Day, K. (1990). Depression in mildly and moderately retarded adults. In *Depression in mentally retarded children and adults* (eds. A. Došen and F. Menolascino), pp. 129–54. Logon, Leiden.
10. Došen, A. and Gielen, J. (1993). Depression in persons with mental retardation: assessment and diagnosis. In *Mental health aspects of mental retardation* (eds. R. Fletcher and A. Došen), pp. 70–97. Lexington Books, New York.
11. Reiss, S. and Rojahn, J. (1994). Joint occurrence of depression and aggression in children and adults with mental retardation. *Journal of Intellectual Disability Research*, **37**, 287–94.
12. Sovner, R. and Pary, R. (1993). Affective disorders in developmentally disabled persons. In *Psychology in the mentally retarded* (eds. J. Matson and R. Barrett), pp. 87–148. Longwood, Boston, MA.
13. Day, K. (1985). Psychiatric disorder in the middle aged and elderly mentally handicapped. *The British Journal of Psychiatry*, **147**, 660–7.
14. Došen, A. (2005c). *Psychische stoornissen, gedragsproblemen en verstandelijke handicap*. Van Gorcum, Assen.
15. Holt, G., and Bouras, N. (2002). *Autism and related disorders*. Royal College of Psychiatrists, London.
16. Gardner, W.I., Došen, A., Griffiths, D.M., *et al.* (2006). *Practice guidelines for diagnostic, treatment and related support services for persons with developmental disabilities and serious behavioral problems*. NADD Press, Kingston.

17. Došen, A. (2005a). Applying the developmental perspective in psychiatric assessment and diagnosis of persons with intellectual disability: part I—assessment. *Journal of Intellectual Disability Research*, **49**, 1–8.
18. Day, K. (1993). Crime and mental retardation, a review. In *Clinical approaches to the mentally disordered offender* (eds. K. Howells and C. Hollin), p. 144. Wiley, Chichester.
19. Corbett, J.A. (1979). Psychiatric morbidity and mental retardation. In *Psychiatric illness and mental handicap* (eds. F. James and R. Snaith), pp. 11–25. Gaskell, London.
20. Reid, A.H. and Ballinger, B.R. (1987). Personality disorder in mental handicap. *Psychological Medicine*, **17**, 983–7.
21. Zigler, E., Bennett-Gates, D., Hodapp, R., *et al.* (2002). Assessing personality traits of individuals with mental retardation. *American Journal of Mental Retardation*, **107**, 181–93.
22. Levitas, A. and Gilson, S. (1994). Psychosocial development of children and adolescents with mild mental retardation. In *Mental health in mental retardation* (ed. N. Bouras), pp. 34–45. Cambridge University Press. Cambridge.
23. Gaedt, C. (1995). Psychotherapeutic approaches in the treatment of mental illness and behavioral disorders in mentally retarded people: the significance of a psychoanalytic perspective. *Journal of Intellectual Disability Research*, **39**, 233–9.
24. Menolascino, F.J. and Weiler, M.A. (1990). The challenge of depression and suicide in severely mentally retarded adults. In *Depression in mentally retarded children and adults* (eds. A. Došen and F. Menolascino), pp. 155–74. Logon, Leiden.
25. Szymanski, L. (1994). Mental retardation and mental health: concepts, aetiology and incidence. In *Mental health in mental retardation* (ed. N. Bouras), pp. 19–33. Cambridge University Press, Cambridge.
26. Reid, A. (1982). *The psychiatry of mental handicap*. Blackwell Science, Oxford.
27. Bregman, J.D. and Harris, J.C. (1995). Mental retardation. In *Comprehensive textbook of psychiatry VI* (eds. H. Kaplan and B. Sadock), pp. 2207–42. Williams and Wilkins, Baltimore, MD.
28. Sturmey, P. (1999). Classification; concepts, progress and future. In *Psychiatric and behavioural disorders in developmental disabilities and mental retardation* (ed. N. Bouras), pp. 3–17. Univesity Press, Cambridge.
29. Van Praag, H.M. (1997). The future of biological psychiatry. *CNS Spectrums*, **12**, 8–23.
30. Van Praag, H.M. (2000). Nosologomania: a disorder of psychiatry. *The World Journal of Biological Psychiatry*, **1**, 151–8.
31. Charlot, L. (2003). Mission impossible?: developing an accurate classification of psychiatric disorders for individuals with developmental disabilities. *Mental Health Aspects of Developmental Disabilities*, **6**, 26–35.
32. Einfeld, S.L. and Tonge, B.J. (1996). Population prevalence of psychopathology in children and adolescents with intellectual disability: epidemiological findings. *Journal of Intellectual Disability Research*, **40**, 99–109.
33. Achenbach, T.M. and Zigler, E.F. (1968). Cue-learning and problem-learning strategies in normal and retarded children. *Child Development*, **39**, 827–48
34. Menolascino, F.J. (1977). *Challenges in mental retardation: progressive ideologies and services*. Human Sciences Press, New York.
35. Matson, J.L. (1985). Bio-social theory of psychopathology: a three by three factor model. *Applied Research in Mental Retardation*, **6**, 199–227.
36. Tanguay, P.E. (1984). Towards a new classification of serious psychopathology in children. *Journal of the American Academy of Child Psychiatry*, **32**, 373–84.
37. Gaedt, C. and Gärtner, D. (1990). *Depressive Grundprozesse–Reinzeniering der Selbstentwertung*. Neuerkeroder Forum 4, Neuerkerode, Sickte.
38. Clegg, J. and Sheard C. (2002). Challenging behaviour and insecure attachment. *Journal of Intellectual Disability Research*, **44**, 503–6.
39. Janssen, C.G.C., Schuengel C., and Stolk, J. (2002). Understanding challenging behaviour in people with severe and profound intellectual disability; a stress-attachment model. *Journal of Intellectual Disability Research*, **46**, 445–53.
40. Collacott, R.A. (1999). People with Down syndrome and mental health needs. In *Psychiatric and behavioural disorders in developmental disabilities and mental retardation* (ed. N.Bouras), pp. 200–11, University Press, Cambridge.
41. Khan, S., Osinowo, T., and Pary, R.J. (2002). Down syndrome and major depressive disorder. *Mental Health Aspects of Developmental Disabilities*, **5**, 46–52.
42. Turk, J., Hagerman, R.J., Barnicoat, A., *et al.* (1994). The fragile X syndrome. In *Mental health in mental retardation* (ed. N. Bouras), pp. 135–53. Cambridge University Press, Cambridge.
43. Dykens, E.M., Hodapp, R.M., and Finucane, B.M. (2000). *Genetics and mental retardation syndromes*. Paul Brookes Publ., Baltimore.
44. Shprintzen, R.J., Goldberg, R., Goldning-Kushner, K.J., *et al.* (1992). Late-onset psychosis in the velo-cardio-facial syndrome. *American Journal of Medical Genetics*, **42**, 141–2.
45. Murphy, K.C. (2004). Review: the behavioural phenotypes in velo-cardio-facial syndrome. *Journal of Intellectual Disability Research*, **48**, 524–30.
46. Pary, R.J. (2002). *Psychiatric problems in older persons with developmental disabilities*. NADD Press, Kingston.
47. Patel, P., Goldberg, D., and Moss, S. (1993). Psychiatric morbidity in older people with moderate and severe learning disability. II. The prevalence study. *The British Journal of Psychiatry*, **163**, 481–91.
48. Day, K. and Jancar, J. (1994). Mental and physical health and aging in mental handicapped: a review. *Journal of Intellectual Disability Research*, **38**, 241–56.
49. Došen, A. (2002). Affective disorders in older people with developmental disabilities. In *Psychiatric problems in older persons with developmental disabilities* (ed. R.J. Pary), pp. 45–64, NADD Press, Kingston.
50. Szymanski, L.S., King, B.H., Bernet, W., *et al.* (1999). Practice parameters for assessment and treatment of children, adolescents and adults with mental retardation and comorbid mental disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, **38**, (Suppl. 12), 5S–31S.
51. Došen, A. (2007). Integrative treatment in persons with intellectual disability and mental health problems. *Journal of Intellectual Disability Research*, **51**, 66–74.

10.5.3 Epilepsy and epilepsy-related behaviour disorders among people with intellectual disability

Matti Iivanainen

Epilepsy is defined as at least one epileptic seizure; this in practice means two or more epileptic seizures unprovoked by any immediate identifiable cause during a relatively short period of time. Epileptic seizure is a clinical manifestation presumed to result from an abnormal and excessive discharge of a set of neurones in the

brain. An epileptic syndrome is a cluster of symptoms and signs including type of seizure, mode of seizure recurrence, neurological findings, and neuroradiological or other findings of special investigations, customarily occurring together. An epileptic syndrome can have more than one cause or the cause may remain unknown; consequently outcomes may be different. Pseudoseizure is used to denote epilepsy-like seizures without concomitant EEG changes.

Epilepsy and intellectual disability are symptoms of brain origin. The former is an unstable condition, where during the seizure or ictally the behaviour of a person with epilepsy is abnormal, but between the seizures or interictally there is no affect of epilepsy on his or her behaviour. Intellectual disability is a more or less stable condition. However, the categories of the degrees of intellectual disability are neither absolute nor static, as some children may move up or down between them.

This chapter deals with the diagnosis, manifestations, behavioural disorders, frequency, aetiology, treatment, effects of antiepileptic drugs on behaviour, and prognosis of epilepsy in people with intellectual disability.

Diagnosis and differential diagnosis of epilepsy

The diagnosis of epilepsy is clinical and requires the collection of historical data, physical and mental examination, EEG, and laboratory tests such as determinations of blood glucose and electrolytes.[1] In babies with epilepsy, attention should be paid to changes of skin colour or cardiac rhythm, sucking and smacking, which all may be epileptic phenomena. In people with intellectual disability it may be difficult or impossible to obtain an accurate clinical history from the patient. The clinician often has to depend on relatives or other professionals involved in the care of the patient. In addition to a description of the seizures, a history of the age at onset of epilepsy and the complete clinical picture of the epileptic syndrome are of value.

General factors that provoke seizures include fever, infection, hypoglycaemia, stress, excessive waking, alcohol withdrawal, hyperventilation, some medications, sudden discontinuation of sedative drugs, and specific activity. Fevers associated with infections such as those of the ears, sinuses, upper respiratory tract, or urinary tract are quite common. If seizures are exacerbated, ensure that any treatable infection is identified. The situation is further confused by the fact that seizures can produce the fever, which, however, resolves within an hour. Withdrawal seizures may be precipitated by a sudden discontinuation of drugs such as benzodiazepines which have an antiepileptic effect, although they may have been prescribed for another reason. Some seizures may be associated with a specific activity, especially if this activity induces excitement or anxiety. Exercise-induced seizures occur regularly in some patients.

Many people with intellectual disability have abnormal behaviour that resembles epileptic seizures but is not epileptic in origin. The diagnostic and other problems caused by non-epileptic seizures or pseudoseizures are well known.[2] Different dyskinesias, psychogenic attacks, and other non-epileptic episodes may be manifested as pseudoseizures at different ages (Table 10.5.3.1). Sudden aggression and other epilepsy-like conditions (Table 10.5.3.2) are in practice the most important reasons for the overdiagnosis of epilepsy, and consequently also for overmedication and subsequent intoxication in patients with intellectual disability. On the other hand, non-convulsive epileptic phenomena and even partial seizures (Table 10.5.3.3) may be difficult to diagnose in people with intellectual disability. The situation is more complicated when patients with intractable epilepsy have both real epileptic seizures and pseudoseizures, for example psychogenic seizures. In such cases the recognition of psychogenic seizures[3] (Table 10.5.3.4) helps to identify appropriate treatment.[4]

Table 10.5.3.1 Salient non-epileptic episodes at different ages

Age	Disorder
0–2 months	Tremor Dyskinesias associated with bronchopulmonar dysplasia Benign myoclonus during sleep Apnoea
2–18 months	Paroxysmal torticollins Ipsoclonus myoclonus syndrome Intestinal obstruction Breath-holding spells Jactatio capitis Masturbation Paroxysmal choreoathetosis Gastro-oesophageal reflux
1.5–5 years	Pavor nocturnus Benign paroxysmal vertigo Nodding puppet syndrome Enuresis nocturnus Confusion with fever Familial dystonic choreoathetosis
5–12 years	Tic Complicated migraine Attention disturbance Sleepwalking Paroxysmalis choreoathetosis
>12 years	Vertebrobasilar migraine Syncope Hyperventilation syndrome Obstructive sleep apnoea Psychogenic attacks Raving fits

Table 10.5.3.2 Conditions often misdiagnosed as epilepsy in subjects with intellectual disability

Sudden aggression
Self-abuse
Bizarre behaviour
Abnormal motor activity
Staring
Eye-blinking
Nystagmus
Exaggerated startle
Intermittent lethargy

Table 10.5.3.3 Underdiagnosis of epilepsy in subjects with intellectual disability

Absence seizures
Non-convulsive status epilepticus
Seizures with periodic headache
Seizures with vertigo
Seizures with paraesthesia
Seizures with visceral and vegetative disturbances
Loss of emotional control
Postictal effects
Simple partial seizures
Complex partial seizures

Using magnetic resonance imaging (**MRI**), it is possible to identify structural brain abnormalities, including neoplasms, dysplasia, heterotopia, or diseases in the brainstem and/or posterior fossa. If MRI is not available, CT is recommended.

Prolonged video-EEG monitoring of the patients is of use in selecting candidates for epilepsy surgery or in distinguishing between epileptic and non-epileptic seizures. Basically, this enables any behaviour to be analysed in relation to the EEG changes. If this investigation is not available, portable cassette recording of the EEG may also be of considerable value. The diagnosis of subclinical seizures, including minimal behavioural or cognitive changes in the absence of any obvious clinical seizures, can be demonstrated as lengthened reaction times during EEG discharges in the Romny test.

The brain function of people with epileptic seizures and syndromes can be examined by interictal and ictal single-photon emission CT, positron emission tomography, functional MRI, magnetic resonance spectroscopy, and magnetoencephalography together with simultaneous EEG. Such investigations can help to define the epileptogenic brain lesion and thereby guide management including decisions about epilepsy surgery.

Epilepsy and epileptic syndromes at different ages

The main categories in the classification of seizures and epilepsy are primary generalized seizures, focal seizures, and secondary generalized seizures[5] (Table 10.5.3.5). The semiologic seizure classification[6] seeks to provide common descriptive terms for typical ictal symptoms and for seizure evolution (Table 10.5.3.6). Epileptic syndromes[7] are quite frequent in people with intellectual disability ranging from early infancy through childhood to adolescence (Table 10.5.3.7). As understanding of the pathophysiologic and anatomic substrates of epileptic seizures, syndromes and disorders increases, these classifications may need to be reappraised. [8–10]

Infancy

Infants with **early infantile epileptic encephalopathy** or Ohtahara syndrome seem initially neurologically normal, but soon develop increasingly frequent seizures with tonic spasms that resemble infantile spasms and are usually resistant to treatment. Severe progressive intellectual disability becomes evident with age. Many die early and most survivals are handicapped. Some may evolve into the West syndrome and some later into the Lennox–Gastaut syndrome (see below). The EEG shows a 'burst suppression' pattern with an almost flat tracing for several seconds, alternating with diffuse, high-amplitude, slow wave-and-spike bursts, poorly modified by sleep–wake stages.[11] The aetiology of Ohtahara syndrome includes usually congenital or acquired malformations of cortical development and diffuse prenatal encephalopathies, the cause of which remains unknown, so far. A report on a case of Ohtahara syndrome included a metabolic defect with cytochrome oxidase deficiency.[12]

Early myoclonic epileptic encephalopathy is another epileptic syndrome occurring during infancy with a grim prognosis.[13] The predominant seizure pattern is erratic, paroxysmal, fragmentary myoclonus, often associated with other seizure types. Brain malformations are not so common as in Ohtahara syndrome.

Infantile spasms occur usually at the ages of 4 to 6 months and in 90 per cent of cases during the first year of life. The events resemble the Moro reflex with sudden, brief flexion of neck and trunk, raising both arms forwards or sideways, sometimes with flexion at the elbows, and flexion of legs at the hips. Less often, the legs extend at the hips. At the early stage flexion of the neck may be the only or main feature; this may be followed by more complex and dramatic attacks later on. A cry is often associated with the attack either as part of the attack or occurring afterwards as an expression of disquiet. The spasms are usually symmetric, but may be asymmetric or even unilateral. The EEG is chaotic with slow waves of high voltage intermixed with diffuse or asynchronous spikes in both hemispheres or in the contralateral hemisphere in unilateral cases. This

Table 10.5.3.4 Differential diagnosis of epileptic and psychogenic seizures

Typical features	Epileptic seizures		Psychogenic seizures
	Generalized tonic-clonic seizures	Complex partial seizures	
Comparison of seizures with known seizure types	Little variation in events	Wide range of events, but the most common are well described	Extremely wide range of events with bizarre or unusual behaviour
Ictal EEG	Abnormal and changed from preictal	Almost always abnormal and changed from preictal	Usually normal and unchanged from preictal
Postictal EEG	Almost always abnormal and changed from preictal	Frequently abnormal and changed from preictal	Usually normal and unchanged from preictal
Effect of antiepileptic medication on seizures	Prominent, especially in severely affected patients	Usually prominent	Usually no effect

Table 10.5.3.5 Classification of epileptic seizures (International League Against Epilepsy, www.ilae.org/Visitors/Centre/ctf/CTF table3.cfm, copyright ILAE)

Partial (focal, local seizures)
Simple partial seizures
With motor signs
With autonomic symptoms and signs
With somatosensory or special sensory symptoms: simple hallucinations (e.g. tingling, light flashes, buzzing), somatosensory, visual, auditory, olfactory, gustatory, vertiginous
With psychic symptoms (disturbances of higher cerebral functions): dysphasic, dysmnesic, cognitive, affective, illusions, structured hallucinations
Complex partial seizures (with impairment of consciousness, may sometimes begin with simple symptomatology)
Partial seizures evolving to secondarily generalized tonic–clonic seizures
Generalized seizures[a]
Absence seizures
Atypical absences
Myoclonic seizures
Clonic seizures
Tonic seizures
Tonic–clonic seizures
Atonic seizures

[a] Combinations of seizures listed here may occur.

Table 10.5.3.6 Semiologic seizure classification[7]

Aura
Somatosensory aura
Visual aura
Auditory aura
Gustatory aura
Olfactory aura
Autonomic aura
Abdominal aura
Psychic aura
Autonomic seizure
Dialeptic seizure
Motor seizure
Simple motor seizure
Myoclonic seizure
Epileptic spasm
Tonic seizure
Clonic seizure
Tonic–clonic seizure
Versive seizure
Complex motor seizure
Hypermotor seizure
Automotor seizure
Gelastic seizure
Special seizure
Atonic seizure
Astatic seizure
Hypomotor seizure
Akinetic seizure
Negative myoclonic seizure
Aphasic seizure

Reproduced from Commission of Classification and Terminology of the International League Against Epilepsy (1989). Proposal for revised clinical and electroencephalographic classification of epilepsies and epileptic syndromes. Epilepsia, **30**, 389–99, copyright 1989, International League Against Epilepsy.

Table 10.5.3.7 Salient epileptic syndromes which may be associated with intellectual disability and salient intellectual disability syndromes which may be associated with epilepsy

Infant
Early myoclonic encephalopathy
Early infantile epileptic encephalopathy with suppression bursts
Infantile spasms
Severe myoclonic epilepsy
Sturge–Weber syndrome
Down syndrome
Fragile X syndrome
Angelman syndrome
Children and adolescents
Epilepsia partialis continua Kojewnikow
Unverricht–Lundborg disease
Lafora disease
Progressive myoclonus epilepsy with intellectual disability
Other neuronal ceroid lipofuscinoses
Sialidosis
Myoclonic epilepsy with ragged red fibres (MERRF)
Rett syndrome
Landau–Kleffner syndrome
Continuous spike-wave discharge during slow-wave sleep

pattern is called hypsarrhythmia. Infants with unilateral spasms need to be examined using a positron emission tomography scan, as contralateral hypometabolism may be due to cortical dysplasia, a condition which may be treatable by resective epilepsy surgery. Aetiology is usually symptomatic including brain abnormalities due to intrauterine infections such as toxoplasmosis, cytomegalic inclusion disease, or rubella. Other aetiologies are brain malformations due to unknown cause. Infants with Down syndrome or tuberous sclerosis may develop infantile spasms. **West syndrome** comprises the triad of infantile spasms, hypsarrhythmia, and intellectual disability.

Progressive degenerative brain diseases and neoplasms are rare causes of infantile spasms. Also neurometabolic disorders such as phenylketonuria, maple syrup urine diseases, non-ketotic or ketotic hyperglycinaemia, and urea cycle defects may lead to infantile spasms.

Severe myoclonic epilepsy in infants includes generalized or unilateral febrile clonic seizures, secondary appearance of myoclonic jerks, and often partial seizures. All the children affected suffer from intellectual disability from the second year of life onwards. Ataxia, signs of upper motor neurone involvement, and interictal myoclonus may appear. [13]

Early childhood

Myoclonic epilepsy of early childhood shares many features with the **Lennox–Gastaut syndrome**.[13] The latter is a group of epileptic disorders of varied aetiology in childhood. West syndrome often evolves into Lennox–Gastaut syndrome characterized by atypical absences, axial, tonic and sudden myoclonic, atonic, partial, and generalized tonic–clonic seizures, diffuse slow interictal spike waves in the waking EEGs and fast rhythmic bursts (10 Hz) during sleep. A progressive decrease in IQ is often found in children with Lennox–Gastaut syndrome. **Myoclonic–astatic epilepsy** or **Doose**

syndrome resembles Lennox–Gastaut syndrome, but is not so severe.

Later childhood and adolescence

Progressive myoclonus epilepsies have the nosological picture of an evolving syndrome of symptoms including massive and segmental myoclonus, myoclonic or tonic–clonic seizures, partial seizures, cerebellar impairment, and higher neurological dysfunctions. [13] Unverricht–Lundborg disease is most common in the Finnish and North African population, but occur also elsewhere. The disease progresses only over a limited period and stabilizes thereafter. [14] The age of onset is around 7 years and the disease starts with myoclonus or nocturnal tonic–clonic seizures. The longest lifespans are more than 60 years. The intelligence level is slightly lowered or even normal. Patients with severe intellectual disability have often had drug intoxication. [15]

Progressive myoclonus epilepsy with intellectual disability (Northern epilepsy) and Lafora disease are more progressive disorders with different gene defects.[16] Sialidosis and mitochondrial encephalopathy with ragged red fibres may also show myoclonic seizures. Epilepsy is quite common in girls with Rett syndrome, affecting about 90 per cent of the patients. They may have several seizure types including partial, generalized tonic–clonic, and myoclonic seizures, atypical absences, short flexion or extensor spasms, and drop attacks or various combinations of such seizures.[17]

Of the progressive **partial epilepsies**, epilepsia partialis Kojewnikow or Rasmussen syndrome type 2 is especially important because the disease is fatal if untreated. The classical model of the association between frequent epileptiform discharges and permanent loss of function is provided by the Landau–Kleffner syndrome or acquired epileptic aphasia. There is increasing evidence that frequent epileptiform discharges, perhaps particularly overnight in the form of continuous spikes and waves during slow sleep, also called electrical status epilepticus, is associated with permanent intellectual impairment if allowed to continue for long periods.[18]

Adulthood and old age

The proportion of cerebrovascular disorders, brain tumours, chronic alcoholism, and sequelae of brain injuries is increasing with advancing age in the aetiology of epilepsy. From about 35 years of age onwards partial epilepsies become more common than generalized epilepsies. Patients with intellectual disability may also develop these disorders.

Behavioural disorders due to epilepsy

Psychic symptoms may be seen as epileptic manifestations of several epileptic seizure types (Table 10.5.3.5). Thus simple partial seizures may manifest themselves with somatosensory or special sensory symptoms including simple hallucinations such as tingling, light flashes, or buzzing or with psychic symptoms such as dysphasic, dysamnesic, cognitive, affective, illusional, or structured hallucinations. Complex partial seizures often include behavioural abnormalities reaching from confusional states to psychotic-like episodes. This is often the case in temporal-lobe and frontal-lobe epilepsies. Among generalized seizures absence status epilepticus resembles psychotic behaviour,[19] which the person in question does not remember afterwards.

In patients with intractable seizures, about one-fifth are non-epileptic in origin. Patients with psychogenic seizures do not generally have seizures when alone or when asleep. Their EEG shows normal activity preictally, ictally, and postictally. The courses of the psychogenic seizures do not show any uniform pattern as is the case with epileptic seizures (see Table 10.5.3.4). Instead their seizures include a variety of behavioural disturbances including conversion, depressive, anxiety, adjustment, somatoform, psychotic, or facitious disorders.[2] The antiepileptic drugs are ineffective against psychogenic seizures but appropriate psychotherapy is helpful in 70 to 80 per cent of cases. Also panic disorder may be difficult to distinguish from complex partial seizures especially in patients with mild intellectual disability. There exists a relationship between brain damage, epilepsy, ictal, and interictal aggressive behaviour, and socio-economic factors. Rarely, ictal aggression occurs in patients with epilepsy, but postictal confusional aggression, and aggression occurring in postictal psychotic states is more common.

Occurrence of epilepsy related to intellectual disability

The prevalences of epilepsy and intellectual disability in the general population are both close to 1 per cent. Epilepsy is more common and more difficult to diagnose and to treat in people with intellectual disability than in those with normal intellect. Population-based studies have revealed that intellectual disability occurs in at least 30 to 40 per cent of individuals with epilepsy,[20] while the prevalence of epilepsy in the population with intellectual disability is about 20 to 25 per cent.[21] It is higher in the more severely disabled (IQ < 50) than in less severely disabled (IQ 50–70)—30 to 50 per cent and 15 to 20 per cent, respectively. Brain damage tends to be more extensive when epilepsy, intellectual disability, or cerebral palsy are complicated by each other or by other conditions of brain origin. It is unlikely that specific causal factors of epilepsy, intellectual disability, or cerebral palsy will ever be positively identified, for these are non-specific clinical features of brain disorder. Epileptic fits themselves, especially if they are persistent, may produce brain damage and play a part in producing a progressive decline in the intellectual functioning of patients. Further apparent deterioration of intellectual functioning may be the result of excessively high doses of anticonvulsant drugs.

Epilepsy occurs in all the main aetiological categories of intellectual disability. In a series of 1000 mentally retarded patients, epilepsy was less frequent in the prenatal category than in the rest of the series (182/515 or 35.3 per cent versus 260/485 or 53.6 per cent). Of the main types of epilepsy, partial epilepsy is more frequent in the prenatal and postnatal aetiological categories and in the category of infections and intoxications.[22]

In people with Down syndrome the frequency of epilepsy is 5 to 10 per cent. There is an age-related bimodal distribution with about 40 per cent of seizures starting before the age of 1 year and another 40 per cent starting after the third decade.[23] Roughly 25 per cent of individuals with fragile X syndrome have epileptic seizures which are usually infrequent, mild, easily controlled, and typically disappear in adolescence, as in benign Rolandic epilepsy. In Angelman syndrome epilepsy is present in more than 90 per cent of the affected individuals.[23] In Rett syndrome epilepsy affects up to 90 per cent of patients[17] Seizures are usually benign

during the early years of life. In patients with aspartylglucosaminuria, epilepsy is found in 28 per cent of adults and in 2 per cent of children.[24] Epilepsy is common (up to 100 per cent) in patients with the various forms of neuronal ceroid lipofuscinoses, especially during the last years of life,[25] and also in other inborn errors of metabolism leading to intellectual disability such as sialidosis type 1, Tay–Sachs disease, type 3 Gaucher disease, mitochondrial encephalopathy with lactic acidosis and strokes, and myoclonic epilepsy with ragged red fibres.

Aetiology and pathogenesis of epilepsy

The presumed aetiology of intellectual disability is also the presumed aetiology of epilepsy in most patients.[21,22] In addition, patients with intellectual disability may develop an ischaemic or haemorrhagic lesion, a neoplasm, or another lesion in the brain which may lead to epilepsy.[21] The presumed aetiology of epilepsy can be found in about three-quarters of the patients. In the aetiological classification based on the time of the presumed cause of epilepsy and intellectual disability, prenatal aetiology is the most common (Table 10.5.3.8). In the aetiological classification based on presumed cause, the categories of unknown prenatal influence, infections and intoxications, trauma and physical agents, and other specified aetiological agents cover most of the cases (Table 10.5.3.9). In patients with intellectual disability and epilepsy it is important to try to find the cause of the intellectual impairment, epilepsy, or epilepsy syndrome. In some cases, the epilepsy syndrome or an underlying inborn error of metabolism may be relevant.

Basic mechanisms leading to epilepsy include disturbances in the balance between excitatory and inhibitory neurotransmitter function within brain cells and their connections to important channels such as voltage-gated sodium channels. For instance, hyperactivity of the excitatory neurotransmitter glutamic acid and/or hypoactivity of inhibitory neurotransmitter γ-aminobutyric acid (**GABA**) may lead to epileptic seizures. The existence of so many genetically determined disorders leading to intellectual disability and epilepsy[26] (Table 10.5.3.10) and the large variation in the prevalence of epilepsy in the specific intellectual disability syndromes and the use of new methods such as an array technology suggest that genetic factors play a more important role in producing epilepsy.

The three following examples, as well as those mentioned above, illustrate this variety of genetic explanations for epilepsy among intellectual disability syndromes.[23] Angelman syndrome is a contiguous gene defect most often caused by a maternally inherited deletion of chromosome 15q11–13. Several of the deleted genes code for GABA receptor subunits. Deficits of inhibitory GABAergic function could directly predispose affected individuals to seizures. This hypothesis is supported by knockouts of analogous chromosome region in mice, which produces an epileptic phenotype.

In Down syndrome or trisomy 21 the bimodal distribution of the frequency of epilepsy between young and older ages is interesting. The fact that more than 75 per cent of adults with Down syndrome develop late-onset epilepsy coincident with the onset of the neuropathological abnormalities in the brain compatible to Alzheimer's disease suggests an aetiological role of these abnormalities. However, as epilepsy is associated with Alzheimer's disease in only 10 per cent of patients without Down syndrome, it is unlikely that Alzheimer's neuropathological abnormalities are solely responsible for the late-onset epilepsy in patients with Down syndrome. As the EEG of most of these patients is characteristic of idiopathic generalized epilepsy and the gene for progressive myoclonus epilepsy is located in the Down syndrome region on chromosome 21, it is quite possible that this gene product predisposes for a senile myoclonus epilepsy in Down syndrome.[23] Abnormal neuronal circuits with fewer GABAergic neurones in certain cortical layers, cerebral dysgenesis particularly of dendritic spines, pathophysiological membrane ion channels, and altered neurotransmitter level are potential mechanisms of epilepsy in Down syndrome.[23]

In the tandem trinucleotide repeat disorder, fragile X syndrome, triplet expansion (CGG) results in shutdown of fragile X intellectual disability 1 gene transcription, which may alter overall neurologic development and lead to seizures.[23]

Thus, there exists a spectrum of epilepsy mechanisms among these three intellectual disability syndromes, ranging from deletion of a gene or genes that directly leads to hyperexcitability (Angelman syndrome), to a chromosomal triplication that alters several aspects of neuronal development and function (Down syndrome), to a

Table 10.5.3.8 Aetiological classification of epilepsy and intellectual disability according to time of presumed cause

	N = 129[a]	*N* = 442[b]
Prenatal (%)	35.1	41.2
Perinatal (%)	10.0	15.4
Postnatal (%)	8.7	18.8
Multiple (%)	14.7	4.3
Unknown (%)	31.4	20.4
Total (%)	100.0	100.0

[a] Data from Forsgren *et al.*[21]

[b] Data from Iivanainen.[22]

Table 10.5.3.9 Aetiological classification of epilepsy and intellectual disability according to presumed cause

	N = 442
Infections and intoxications (%)	18.1
Trauma and physical agents (%)	16.3
Disorders of metabolism (%)	3.2
Gross prenatal influence (%)	24.0
Prematurity (%)	0.5
Major psychiatric disorder (%)	0.5
Psychosocial deprivation (%)	0.0
Multiple causes (%)	7.2
Hereditary (simple) (%)	0.5
Other specified (%)	19.7
Unspecified (%)	0.0
Total (%)	100.0

(Reproduced from M. Iivanainen, Diagnosis of epileptic seizures and syndromes in mentally retarded patients. In *Paediatric epilepsy* (ed. M. Sillanpää *et al.*), pp. 233–41. Copyright 1990, Wrightson Biomedical, Petersfield)

Table 10.5.3.10 Genetic diseases with epilepsy and intellectual disability

Disease	Mode of inheritance	Gene location
Progressive epilepsy with intellectual disability (Northern epilepsy)	AR	8p
Unverricht–Lundborg disease (Baltic, Mediterranean)	AR	21q22
Infantile neuronal ceroid lipofuscinosis	AR	1p32
Late infantile neuronal ceroid lipofuscinosis	AR	11p15
Variant late infantile neuronal ceroid lipofuscinosis	AR	15q21-23
Juvenile neuronal ceroid lipofuscinosis	AR	16p12
Finnish variant neuronal ceroid lipofuscinosis	AR	13q22
Lafora disease	AR	6q23-25
Mitochondrial encephalomopathy, lactic acidosis (UUR) and stroke-like episodes (MELAS)	Maternal	tRNA-Leu (UUR)
Myoclonic epilepsy with ragged red fibres (MERRF)	Maternal	tRNA-Lys
Epilepsy and mental retardation limited to females	X-linked	Xq22
Tuberous sclerosis[a]	AD	9q34 16p13
Angelman syndrome[a]	AD	15q13
Neurofibromatosis type[a] 1	AD	17q11
Fragile X[a]	X-linked	Xq27
Rett syndrome[a]	X-linked	Xq28-McCP2

AR, autosomal recessive; AD, autosomal dominant.

[a] Epilepsy and/or intellectual disability may be manifested as part of the phenotype.

specific tandem repeat which alters neuronal function in a non-specific and probably benign manner (fragile X syndrome). It remains to be seen how much dissection of genetic mechanisms underlying other intellectual disability syndromes will provide additional insight into epilepsy mechanisms.

Treatment of epilepsy

The diagnosis of epilepsy and its underlying disorder needs to be made without delay. The identification and avoidance of provoking factors likely to precipitate seizures in each individual is an essential aspect of the overall management. If this is insufficient, antiepileptic drug treatment is needed. If this is still insufficient, epilepsy surgery should be considered. Important points to be considered include not only the nature and severity of an underlying disease, but also the degree and location of brain lesion, the age of the patient at onset of epilepsy, and possible pseudodisability (pseudoretardation) caused by epileptic seizures or by inappropriate medication. It is emphasized that treating frequent epileptiform discharges may not only reverse the intellectual disability which in such cases is pseudodisability or state-dependent intellectual disability,[18] but may also in some cases prevent permanent intellectual disability.

Antiepileptic drug therapy

Drug interactions

Antiepileptic drugs interact with each other by three principal mechanisms: enzyme induction, enzyme inhibition, and through altered protein binding. Phenytoin and phenobarbital induce a wide range of enzyme activity. Carbamazepine induces its own enzymatic metabolism, and may induce the metabolism of valproate and phenytoin, resulting in lower concentrations of these drugs. Valproate inhibits the metabolism of phenobarbital and the epoxide of carbamazepine, resulting in high concentrations of each of these. Lamotrigine is metabolized in the liver. Valproate inhibits the metabolism of lamotrigine, resulting in a longer half-life and higher blood levels of lamotrigine while the blood level of valproate may be decreased. Lamotrigine does not affect the blood level of carbamazepine. Gabapentin is not metabolized at all and is excreted in the urine unchanged. If levetiracetam is administered with enzyme-inducing drugs, its clearance may increase by 22 per cent, although the drug concentrations in serum do not change when using other antiepileptic drugs simultaneously. The free or unbound fraction of antiepileptic drugs is in equilibrium with the brain concentration and is considered to be more relevant than the total blood level. When two drugs with a high degree of protein binding are used together, for example phenytoin and valproate, there may be some displacement of each drug from protein binding, increasing the unbound fraction. This may result in clinical neurotoxicity even when the total (bound plus unbound) blood level is within the reference range. The antiepileptic drugs also interact with many other drugs and may affect their blood levels and action, and vice versa. For instance, chloramphenicol, cimetidine, anticoagulants, ibuprofen, imipramine, propranol, and some psychotropic drugs inhibit the metabolism of phenytoin and lead to an increase of phenytoin blood level and possibly to phenytoin intoxication unless the dose of phenytoin is reduced.

Choice of drug

Once the diagnosis of epilepsy has been made, the decision must be made as to whether antiepileptic medication is needed or not. If it is, the most appropriate antiepileptic drug is to be selected. The choice of antiepileptic medication depends primarily on an accurate classification of the seizure type and/or epilepsy syndrome.

Most treatment decisions have to be based on the results of studies of people who do not have intellectual disability. The exceptions are mainly studies of adults with the Lennox–Gastaut syndrome, where lamotrigine is recommended.[27] Valproate, carbamazepine, oxcarbazepine, and levetiracetam[28] are other antiepileptic drugs recommended in generalized and partial seizures based on uncontrolled studies or consensus. Valproate is the first choice for generalized epilepsies/seizures, while oxcarbazepine/carbamazepine is the choice for the focal epilepsies/seizures of people with intellectual disability.[29] Dosage and recommended drug levels in blood are presented in Table 10.5.3.11. If newer drugs are not available, phenobarbitone, primidone, and phenytoin may be used with caution, if special attention is paid not only to control of seizures, but also to behavioural, cognitive, and cerebellar functions which may

Table 10.5.3.11 Pharmacokinetic properties of antiepileptic drugs

Drug	Dose (mg/kg/day)	Doses per day	Therapeutic range	
			(μg/ml)	(μmol/ml)
Phenobarbitone	1–3	1	10–30	40–130
Primodone	10–15	2 or 3	6–12	25–50
Phenytoin	4–6	1 or 2	10–20	40–80
Carbamazepine	15–20	2 or 3	4–12	15–50
Oxcarbazepine	15–40	2 or 3	15–23	30–120[a]
Valproate	15–30	2 or 3	50–100	300–700
Ethosuximide	15–30	1 or 2	40–100	280–700
Levetiracetam	20–40	2	5–65	30–370
Vigabatrin	40–100	1 or 2	NA	NA
Lamotrigine	2–4(10)[b]	1 or 2	NA	NA
Gabapentin	20–40	3	NA	NA
Topiramate	400	3	NA	NA
Tiagabine	32–56	3	NA	NA
Clonazepam	0.01–0.2	2 or 3	20–75	60–240

NA = not available.

[a]For monohydroxy derivative.

[b]With enzyme-inducing drugs.

be affected adversely, and sometimes insidiously, by these drugs. However, it is stressed that because of the lack of well-designed properly conducted randomized controlled trials for patients with newly diagnosed generalized of focal untreated seizures/epilepsies and for children in general, it is impossible at present to develop evidence-based guidelines aimed at identifying the overall optimal recommended monotherapy antiepileptic drug.[30]

Most antiepileptic drugs may aggravate certain epilepsies. This is the case especially with phenytoin and carbamazepine in idiopathic generalized absence and myoclonic epilepsies. Although valproate has low risk of seizure aggravation,[31] it may cause weight gain, polycystic ovaries, and hepatitis. Oxcarbazepine may lead to hyponatraemia and vigabatrin to visual field defects.[32]

Withdrawal of treatment

Withdrawal of antiepileptic medication in patients with intellectual disability needs to be considered, provided their epilepsy is well-controlled and there are no specific contraindications for the withdrawal. Depending on the type of seizure or epilepsy syndrome and the individual history, it might be worth considering slow reducing the medication after a 2-year seizure-free period. However, if the individual is in a particularly poor prognostic category or if there is a history of severe, prolonged status epilepticus, it is worth waiting for longer before attempting medication reductions. Despite these reservations, attempts to discontinue antiepileptic medication in people with intellectual disability can be successful. It would appear that a better likelihood of a successful outcome may be suggested by later onset of epilepsy, i.e. after 2 to 2.5 years of age, a shorter duration, lower antiepileptic drug levels, and normal EEGs, together with complete control of the seizures.[33] The risk of recurrence of the seizures also depends on the type of epilepsy.[34] In complex partial seizures, where exogenic factors are more significant than genetic ones, the prognosis is good after 2 to 4 seizure-free years. The prognosis is worse in simple partial seizures, or in absence seizures with tonic–clonic seizures and grand mal tonic–clonic seizures, and at least four seizure-free years are recommended before ceasing medication. Patients with juvenile myoclonic epilepsy, or absence seizures with clonic–tonic–clonic seizures or grand mal clonic–tonic–clonic seizures, may need to take long-term, even lifetime medication. The relapse rate is likely to be high, even if these patients have been free of seizures for several years.

Status epilepticus

First-line treatment of status epilepticus is usually with intravenous benzodiazepines, either diazepam or lorazepam. If this is not effective, then intravenous phenytoin or intravenous or intramuscular fosphenytoin is recommended. Rectal paraldehyde may be of value in children. Rectal diazepam has been the pre-hospital treatment of first choice because it can be administered by non-medical personnel. A history of status epilepticus is liable to influence decisions about withdrawing regular antiepileptic medication. It would be wise to proceed with caution, ensuring that any recurrence of status epilepticus can be treated readily, if withdrawal of regular medication is to be undertaken.

Epilepsy surgery

Frequent severe epileptic seizures despite treatment with adequate antiepileptic medication for about 2 years means that epilepsy surgery needs to be considered. In addition there are certain disorders such as Sturge–Weber syndrome or unilateral infantile spasms where epilepsy surgery may be of benefit during infancy. The treatment of choice in children with Rasmussen syndrome type 2 currently is hemispherectomy, which needs to be done as early as the diagnosis is clear. Difficulties in cooperation and minimal psychosocial gains due to low IQ as well as progressive underlying disease may be contraindications for epilepsy surgery. The preoperative consideration includes extensive examinations such as video-EEG monitoring, high-resolution MRI, positron emission tomography, and neuropsychological and psychiatric evaluation according to the generally accepted principles.[35] The goal is to select those candidates who will benefit from epilepsy surgery. Surgical outcome varies according to the different pathologies of epileptogenic lesions. Thus, the results of surgery are better among patients with mesial temporal sclerosis, chronic encephalitides, infantile hemiplegia, focal cortical dysplasia, tuberous sclerosis, Sturge–Weber syndrome, or post-traumatic cicatrix than among patients with extratemporal focal sclerosis, polymicrogyria with or without heterotopia or hemimegalencephaly, anoxic brain damage, gliosis of obscure aetiology, or no structural pathology.[36] All these findings must be taken into account when selecting patients with epilepsy and intellectual disability for epilepsy surgery. As onset of intractable epilepsy within the first 24 months of life is a significant risk factor for intellectual disability, early intervention for epilepsy surgery is emphasized.[37]

Behavioural disorders caused by antiepileptic drugs

Most antiepileptic drugs may also cause behavioural disturbance and cognitive dysfunction.[38] For example, the diplopia caused by carbamazepine may result in considerable distress and consequent behavioural disturbance. Communication difficulties, which are

common in intellectually disabled patients, may add to the distress and make behavioural disturbance more likely.

Intellectually disabled people with epilepsy are especially vulnerable to harmful neurotoxic effects—sedation caused by phenobarbital or benzodiazepines or cognitive and cerebellar dysfunction caused by phenytoin alone or often together with other antiepileptic drugs. If these alarming effects are not taken into account, inappropriate medication may even jeopardize the rehabilitation of the patients.

Uncertainty about the long-term effects of antiepileptic drugs on brain function and development is largely due to conflicting results of often biased human observations. The problem can only be resolved in controlled experiments which by necessity must be done in animal models. Behavioural and structural consequences of epileptic activity and their modification by antiepileptic drugs are important points to be evaluated in experimental studies. It was reported[39] that enhancement of GABAergic inhibition by administration of vigabatrin prevented both pyramidal cell damage in CA1 and CA3 areas of hippocampus and the disappearance of somatostatin immunoreactive neurones from the dentate gyrus after perforant path stimulation in rats. Furthermore, the preservation of hippocampal structure was accompanied by prevention of the spatial memory deficits seen in control animals after such stimulations. Another study using a kainic acid status epilepticus model in adolescent rats[40] showed that animals that received kainic acid followed by valproic acid resembled control animals who had never received kainic acid with respect to their behavioural and memory performance and had fewer histological lesions. Animals that received kainic acid followed by saline or phenobarbital had impaired learning and behaviour, and more extensive lesions in the hippocampus. Thus, in this experiment valproic acid suppressed seizures and subsequent epilepsy while phenobarbital was only partly effective in suppressing seizures and did not prevent epilepsy. It is likely that seizures themselves, as opposed to the drugs, produced negative behavioural consequences in these rats. It is noteworthy that valproate at very high doses was protective against neuronal damage and prevented epileptogenesis in the kainic acid model.

Chronic phenytoin intoxication, especially in multiple drug therapy, may lead to ataxia, balance impairment and in the worst case finally to persistent loss of locomotion.[41] Another example of an insidious and dangerous effect of phenytoin was documented by the changed course of Unverricht–Lundborg disease. Its rather benign course worsened during phenytoin treatment, so that the patients became bedridden and pseudoretarded and their lifespan shortened from 50 to 60 years to under 30 years. When these patients were treated with valproate instead of phenytoin, their lifespan increased to the prephenytoin level.[15] In some of these patients, the loss of locomotion was reversible after valproate replaced phenytoin.

Prognosis

The long-term outcome of patients depends primarily on the underlying disorder; prognosis is better in idiopathic than in symptomatic cases. If it is not a question of a progressive brain disorder, epilepsy is quite easily treatable in patients with intellectual disability. Thus, about 70 per cent may obtain good seizure control with appropriate antiepileptic drug therapy or epilepsy surgery. The outcome of patients with specific disorders may vary considerably. For instance, the outcome of patients with Doose syndrome is variable but basically better than that of patients with Lennox–Gastaut syndrome. If hemispherectomy is not undertaken in time in Rasmussen syndrome type 2, the course of the disease, including neurological deficits, other types of seizures, and mental impairment is progressive.

Epileptic seizures themselves, and frequent epileptiform discharges, may produce brain damage and play a part in producing progressive decline in the intellectual functioning of patients[18] probably through at least two mechanisms: excitatory glutamate storm within cerebral neurones[42] and opening of the blood–brain barrier[43] during and after epileptic seizures.

Sleep disorders are often associated, in people with intellectual disability, with difficult-to-treat epilepsy and behavioral problems.[44,45] When sleep disorders are diagnosed and treated, antiepileptic and also psychotropic medication can be reduced successfully.

Conclusions

The quality of life in this population benefits from early diagnosis and differential diagnosis of epilepsy, including epilepsy-related behavioural disorder in patients with intellectual disability, identification of its aetiology, and appropriate antiepileptic drug treatment using firstly one drug therapy and, if needed later, rational multiple drug therapy. Currently, valproate is the first choice in generalized seizures while oxcarbazepine or carbamazepine are used for partial seizures with or without secondary generalization. Broad-spectrum drugs such as levetiracetam, lamotrigine, topiramate, or zonisamide are promising. The usefulness of epilepsy surgery should be considered in intractable cases no later than within 2 years, if adequate antiepileptic drug therapy does not help. To minimize sedative and other behavioural and other side-effects caused by antiepileptic drugs the fewest possible drugs should be administered at the lowest effective dose. This means that there should be careful clinical observation of the patients together with determination of drug concentrations in blood and other appropriate laboratory tests. Psychological aspects and sleep behaviour of the patients need to be taken into consideration in the treatment of epilepsy in patients with intellectual disability. Doctors and other personnel working in this field need special education.

Future prospects in the treatment of intractable epileptic seizures might involve the development of gene therapy, neuroprotective drugs, and drugs targeting epileptogenesis. Such treatment possibilities may bring new hope for people with epilepsies that are difficult to treat, including the population with a high proportion of refractory cases, namely that with intellectual disability.

Further information

Papayiopoulos, C.P. (2007). *A clinical guide to epileptic syndromes and their treatment* (2nd edn), pp. 480. Springer, New York.

Leppik, I., Loscher, W., McDonald, R.L., *et al.* (eds.) (2006). Epileptic syndromes in infancy and early childhood—proceedings of the international symposium of epileptic syndromes in infancy and early childhood—evidence-based taxonomy and its implications in the ILAE classification. *Epilepsy Research*, **70**(Suppl. 1), 1–280.

Birbeck, G. (2006). Interventions to reduce epilepsy-associated stigma. *Psychology, Health & Medicine*, **11**, 364–6.

Vallenga, D., Grypdonck, M., Tan, F., *et al.* (2006). Decision-making about risk in people with epilepsy and intellectual disability. *Journal of Advanced Nursing*, **54**, 602–11.

Ryvlin, P. (2006). When to start antiepileptic drug treatment: seize twice might not harm. *Current Opinion in Neurology*, **19**, 154–6.

References

1. Iivanainen, M. (1999). Diagnosing epilepsy in patients with mental retardation. In *Epilepsy and mental retardation* (eds. M. Sillanpää, L. Gram, S.I. Johannessen, and T. Tomson), pp. 61–72. Wrightson Biomedical, Petersfield.
2. Gram, L., Johannessen, S.I., Osterman, P.O., *et al.* (eds.) (1993). *Pseudoepileptic seizures*. Wrightson Biomedical, Petersfield.
3. Desai, B.T., Porter, R.J., and Penry, J.K. (1982). Psychogenic seizures. A study of 42 attacks in six patients, with intensive monitoring. *Archives of Neurology*, **39**, 202–9.
4. Ramani, V. and Gummit, R.J. (1982). Management of hysterical seizures in epileptic patients. *Archives of Neurology*, **39**, 78–81.
5. Commission of Classification and Terminology of the International League Against Epilepsy. (1981). Proposal for revised clinical and electroencephalographic classification of epileptic seizures. *Epilepsia*, **22**, 489–501.
6. Luders, H., Acharya, J., Baumgartner, C., *et al.* (1998). Semiological seizure classification. *Epilepsia*, **39**, 1006–13.
7. Commission of Classification and Terminology of the International League Against Epilepsy. (1989). Proposal for revised clinical and electroencephalographic classification of epilepsies and epileptic syndromes. *Epilepsia*, **30**, 389–99.
8. Engel, J. Jr. (1998). Classifications of the International League Against Epilepsy: time for reappraisal. *Epilepsia*, **39**, 1014–7.
9. Engel, J. Jr. (2001). Classification of epileptic disorders. *Epilepsia*, **42**, 316–17.
10. Engel, J. Jr. (2006). Report of the ILAE classification core group. *Epilepsia*, **47**, 1558–68.
11. Ohtahara, S., Ohtsuka, Y., Yamatogi, Y., *et al.* (1992). Early-infantile epileptic encephalopathy with suppression–bursts. In *Epileptic syndromes in infancy, childhood and adolescence* (eds. J. Roger, M. Bureau, Ch. Dravet, F.E. Dreifuss, A. Perret, and P. Wolf), pp. 25–34. John Libbey, London.
12. Williams, A.N., Gray, R.G., Poulton, K., *et al.* (1998). A case of Ohtahara syndrome with cytochrome oxidose deficiency. *Developmental Medicine and Child Neurology*, **40**, 568–70.
13. Roger, J., Bureau, M., Dravet, Ch., *et al.* (eds.) (1992). *Epileptic syndromes in infancy, childhood and adolescence*. John Libbey, London.
14. Magaudda, A., Ferlazzo, E., Nguyen, V.H., *et al.* (2006). Unverricht-Lundborg disease, a condition with self-limited progression: long-term follow-up of 20 patients. *Epilepsia*, **47**, 860–6.
15. Eldridge, R., Iivanainen, M., Stern, R., *et al.* (1983). 'Baltic' myoclonus epilepsy: hereditary disorder of childhood made worse by phenytoin. *Lancet*, **ii**, 838–42.
16. Ranta, S., Lehesjoki, A.E., de Fatima Bonaldo, M., *et al.* (1997). High-resolution mapping and transcript identification at the progressive epilepsy with mental retardation locus on chromosome 8p. *Genome Research*, **7**, 887–96.
17. Witt Engeström, I. (1992). Age-related occurrence of signs and symptoms in the Rett syndrome. *Brain and Development*, **14**(Suppl.), 11–20.
18. Besag, F.M.C. (1995). Epilepsy, learning and behaviour in childhood. *Epilepsia*, **36**(Suppl. 1), S58–63.
19. Bottaro, F.J., Martinez, O.A., Pardal, M.M., *et al.* (2007). Nonconvulsive status epilepticus in the elderly: a case-control study. *Epilepsia*, **48**, 966–72.
20. Jalava, M. and Sillanpää, M. (1996). Concurrent illnesses in adults with childhood-onset epilepsy: a population-based 35-year follow-up study. *Epilepsia*, **37**, 1155–63.
21. Forsgren, L., Edvinsson, S.-O., Blomqvist, H.K., *et al.* (1990). Epilepsy in a population of mentally retarded children and adults. *Epilepsy Research*, **6**, 234–48.
22. Iivanainen, M. (1990). Diagnosis of epileptic seizures and syndromes in mentally retarded patients. In *Paediatric epilepsy* (eds. M. Sillanpää, S.I. Johannessen, G. Blennow, and M. Dam), pp. 233–41. Wrightson Biomedical, Petersfield.
23. Sillanpää, M., Gram, L., Johannessen, S.I., *et al.* (eds.) (1999). *Epilepsy and mental retardation*. Wrightson Biomedical, Petersfield.
24. Arvio, M., Oksanen, V., Autio, S., *et al.* (1993). Epileptic seizures in aspartylglucosaminuria: a common disorder. *Acta Neurologica Scandinavica*, **87**, 342–4.
25. Santavuori, P., Vanhanen, S.L., and Autti, T. (2001). Clinical and neuroradiological diagnostic aspects of neuronal ceroid lipofuscinoses disorders. *European Journal of Paediatric Neurology*, **5**(Suppl. A), 157–61.
26. Elmslie, F. (2000). Gene table: epilepsy (update). *European Journal of Paediatric Neurology*, **4**, 87–90.
27. Working Group of the International Association of the Scientific Study of Intellectual Disability. (2001). Clinical guidelines for the management of epilepsy in adults with an intellectual disability. *Seizure*, **10**, 401–9.
28. Brodie, M.J., Perucca, E., Ryvlin, P., *et al.* (2007). Comparison of levetiracetam and controlled-release carbamazepine in newly diagnosed epilepsy. *Neurology*, **68**, 402–8.
29. Iivanainen, M. and Alvarez, N. (eds.) (1998). Drug treatment of epilepsy in people with intellectual disability. *Journal of Intellectual Disability Research*, **42**(Suppl. 1), 1–92.
30. Glauser, T., Ben-Menachem, E., Bourgeois, B., *et al.* (2006). ILAE treatment guidelines: evidence-based analysis of antiepileptic drug efficacy and effectiveness as initial monotherapy for epileptic seizures and syndromes. *Epilepsia*, **47**, 1094–120.
31. Gayatri, N.A. and Livingston, J.H. (2006). Aggravation of epilepsy by anti-epileptic drugs. *Developmental Medicine and Child Neurology*, **48**, 394–8.
32. Werth, R. and Schadler, G. (2006). Visual field loss in young children and mentally handicapped adolescents receiving vigabatrin. *Investigative Ophthalmology & Visual Science*, **47**, 3028–35.
33. Marcus, J.C. (1998). Stopping antiepileptic therapy in mentally retarded, epileptic children. *Neuropediatrics*, **29**, 26–8.
34. Delgato–Escueta, A.V., Treiman, D.M., and Walsh, G.O. (1983). The treatable epilepsies. *The New England Journal of Medicine*, **308**, 1508–14, 1576–84.
35. Engel, J. Jr. (ed.) (1993). *Surgical treatment of the epilepsies* (2nd edn). Raven Press, New York.
36. Goldring, S. (1987). Pediatric epilepsy surgery. *Epilepsia*, **28** (Suppl. 1), S82–102.
37. Vasconcellos, E., Wyllie, E., Sullivan, S., *et al.* (2001). Mental retardation in pediatric candidates for epilepsy surgery: the role of early seizure onset. *Epilepsia*, **42**, 268–74.
38. Aldenkamp, A.P. (2001). Effects of antiepileptic drugs on cognition. *Epilepsia*, **42**(Suppl. 1), 46–9.
39. Ylinen, A.M.A., Miettinen, R., Pitkänen, A., *et al.* (1991). Enhanced GABAergic inhibition preserves hippocampal structure and function in a model of epilepsy. *Proceedings of the National Academy of Sciences of the United States of America*, **88**, 7650–3.
40. Bolanos, A.R., Sarkisian, M., Yang, Y., *et al.* (1998). Comparison of long-term effects of valproate and phenobarbital in adolescent rats. *Neurology*, **51**, 41–8.

41. Iivanainen, M. (1998). Phenytoin: effective but insidious therapy for epilepsy in people with intellectual disability. *Journal of Intellectual Disability Research*, **42**(Suppl. 1), 24–31.
42. Rothman, S.M. and Olney, J.W. (1987). Excitotoxicity and the NMDA receptor. *Trends in Neurosciences*, **10**, 299–302.
43. Sokrab, T.-E.O., Johansson, B.B., Kalimo, H., *et al.* (1988). A transient hypertensive opening of the blood-brain barrier can lead to brain damage. *Acta Neuropathologica*, **75**, 557–65.
44. Iivanainen, M., Partinen, M., and Kaski, M. (2005). Why is it important to study sleep of people with intellectual disability? In *Proceedings of the first congress of the world association of sleep medicine, Berlin* (eds. T. Penzel, I. Fietze, and S. Chokroverty), pp. 131–4. Medimond International Proceedings, Bologna (Italy).
45. Manni, R., Terzaghi, M., and Zambrelli, E. (2006). REM sleep behavior disorder and epileptic phenomena: clinical aspects of the comorbidity. *Epilepsia*, **47**(Suppl. 5), 78–81.

10.6

Methods of treatment

T. P. Berney

Introduction

The presence of intellectual disability (mental retardation) affects the character of treatment in a number of ways.

1 Limited communication will hamper diagnosis so that much more has to be inferred from observable behaviour and greater weight given to the interpretation of carers.

2 Diagnosis is provisional, a therapeutic hypothesis that provides the basis for a programme of treatment that is, essentially, a therapeutic trial. There is a wide variation in individual characteristics, with intellectual disability (mental retardation) extending over an enormous range of ability, associated disabilities, aetiology, and psychopathology. Multiple pathology means that the response can be unexpected and ambiguous. For example, diminished aggression with carbamazepine may simply reflect its psychotropic effect but may also be the result of better control of unrecognized epilepsy. A well-designed behavioural milieu may produce a rapid improvement in disturbed behaviour. However, the improvement might simply reflect the effect on someone with autism of moving to a more settled, structured, and predictable environment. It may also reflect an improvement in organic disorder (such as epilepsy or gastritis) in response to a reduction in stress. The therapist has to tailor the treatment to the individual, to try to be specific as to what aspect of the disorder is being targeted and be very selective as to whom they treat.

3 Ill-defined treatment objectives often leave it unclear whether a treatment is aimed at a disorder (e.g. autism), a symptom (panic), or an associated disorder (depression or epilepsy).

4 Normal developmental change may be misattributed to a coincident treatment programme. For example, autism and epilepsy often show spontaneous improvement at about 4 years age and again in late adolescence, both times when the person is likely to be moving into new programmes. This propensity for maturational improvement, coupled with a tradition of care, has led developmental psychiatrists to have a greater therapeutic optimism for many problems, such as personality disorders, that are often considered intractable in those of normal ability.

5 Natural cycles of change can give a misleading sense of success. For example, a behavioural programme may be credited with the remission of a self-limiting episode of disorder such as depression. The true diagnosis may become clear only when it recurs and fails to respond to booster programmes.

6 There is limited evidence for the effectiveness of most treatments. Except for the behavioural therapies, most treatment is based on small series, open trials, the theoretical, and the ideal.

7 A large component of the therapeutic relationship is indirect, being with the family, carers, or professionals rather than directly with the patient. Many programmes utilize the power of the placebo effect, a dynamic that confounds scientific trials but one that should be used to its full in everyday clinical practice. Limited communication and greater dependency lead to work with the systems around the patient; many of the approaches, such as family therapy, deriving from child psychiatry.

8 The ability to consent to treatment is often underestimated. Circumstances often make it difficult for people with mental retardation to choose or refuse a particular therapy, particularly behavioural programmes and drug treatments. Their capacity to give or withhold consent should be assessed automatically and their care should fall within a legal framework that safeguards their rights and protects them from abuse.

Treatment services for people with intellectual disability (mental retardation) have two components. There is the routine support that should be available to all with intellectual disability (mental retardation). Its aim is to help people to grow up as normally as possible, offsetting the effects of their disability, and to establish the therapeutic environment. Second is the provision of treatment for individuals with disturbance; aimed at specific symptoms or disorders based on a multiaxial diagnosis[1] that includes the following:

- Axis I: The nature and degree of intellectual disability (mental retardation) for, in addition to the overall developmental delay other, specific, disabilities, and abilities are often present. For example, a discrepancy between receptive and expressive language may result in someone understanding little of what is said to him while sounding falsely fluent.
- Axis II: The aetiology of the retardation—there is increasing recognition of the contribution of a behavioural phenotype. Of particular note are autism and its imitators, drawing on the ubiquities of social impairment, obsessionality and communication problems, which are being teased apart.

- Axis III:
  - Level A: Other developmental disabilities that are associated with intellectual disability (mental retardation) such as autism, attention deficit disorder, and epilepsy,
  - Level B: Psychiatric disorder—the way this is defined will define the mode of treatment. A functional analysis with antecedents, triggers, and consequences leads into a behavioural programme. A more biological label (such as psychosis) opens the door to drug treatment. They are not mutually exclusive,
  - Level C: The patient's personality—this is often unusual and it may be difficult to distinguish from a pervasive developmental disorder (see Chapter 9.2.2) which runs through Intellectual disability (mental retardation) and which, once recognized, often explains the inexplicable,
  - Level D: Other disorders such as habit disorders and sexual preference disorders,
- Relevant comorbid, physical conditions such as hay fever, asthma, hypothyroidism, or gastro-oesophageal disorders. Particularly important are epilepsy and the antiepileptic agents that are dealt with in detail elsewhere (see Chapters 6.2.6 and 10.5.3, respectively).
- The patient's environment which includes not just their physical surroundings but also the people and their relationships with them.
- Contributory factors from the patient's past, notably the various forms of abuse.

The therapeutic environment

Support may be provided in different ways:

- Level A: General, the network of care provided for people with a intellectual disability (mental retardation) and their carers. This will include community teams, special schools, and the specialized residential placements that might be resorted to either as a short break or as a long-term home.
- Level B: Specific to a particular disorder, parental support groups exist for autism, epilepsy, and specific forms of intellectual disability (mental retardation) such as Prader–Willi, Fragile X, and Cornelia de Lange syndromes.

A primary aim is to integrate people into their community as far as possible. The concept of normalization implies that services should avoid the demarcation that leads to adverse discrimination. Conversely, those with disabilities too severe or too complex for their families or standard teaching or occupational placements may fare better in specialist settings. Examples of these are as follows:

- Some of those with autism, who are so distracted by the complexity of everyday life and the unpredictability of people that they need specialist environments which are well structured, predictable, and under their control.
- Those with severe or intractable seizures.
- Those with aggressive or disinhibited behaviour so disruptive as to block the progress of their peers.

The specialist setting encourages mutual support for people and families with similar problems; the staff gain experience; and it allows a concentration of expertise. However, it also encourages stigmatization and there is the risk that, in a group, disturbed people may copy or amplify each other's behaviour.

The family and other carers

Disturbance arises in the setting of a system of care which includes not only the family but also the staff of other placements, whether day or residential, educational or occupational. Its management will depend on the way these people perceive mental retardation and its care, their attitudes deriving from both past experience and present relationships (Fig. 10.6.1). A great deal depends on the extent to which they feel supported and assured in their roles, as much by each other as by the available system of care. For example, a mother or a teacher who is told frequently that, whatever it was that went wrong, it was her fault, is unlikely to cope confidently.

Disability and disturbance both hinder normal developmental experience; the child's atypical response spoiling the carers' efforts to learn parenting skills and leaving them demoralized or deskilled. They may need formal teaching in such skills as how to engage socially with the child, to play with them, to give clear and understandable instructions, and how to divert rather than confront. Disturbance may reflect boredom and be reduced simply by increasing the amount and variety of activities. Any approach must take a broad view, for carers have to work together comfortably enough to be consistent over time. A treatment programme may have to address the relationship between the carers as well as their needs. This may be sufficient in itself, improvement in the patient following an overall improvement in functioning in the family, school, or residential placement. There has been a growth in the conscious application of a systemic approach to work in this field.[2]

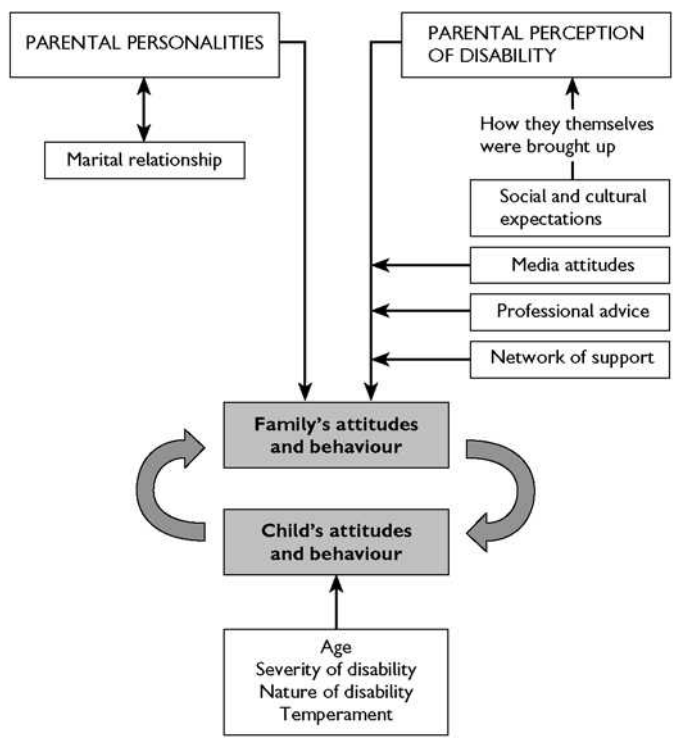


Fig. 10.6.1 The start of disturbance: factors influencing behaviour.

Education

This forms the core of any treatment programme, the individual getting formal instruction in the skills which others acquire in passing. Teaching and training are lifelong processes, invoking teacher–student relationships, and take place within a structured framework, essential for those students who have difficulty in understanding their environments. The difficulty is to gauge the degree of structure required: too much and the relationship risks becoming a battle about control, easily turning into trench warfare; too little and the student is distracted and learns little. It is a balance that shifts over time, promoting the student's self-control and autonomy.

(a) Independence and self-help skills

These include the basic skills of everyday life, such as feeding, dressing, or managing stairs or, at a higher level, the use of public transport, how to care for clothes, shop, and budget. Acquiring these, gives a sense of achievement and confidence as well as of increased independence.

(b) Communication skills

The frustration of living in an uncomprehending world frequently contributes to disturbance as the person falls back on various forms of attention seeking or violent behaviour to get their message across. Easier and more effective means of communication may range from simple gestures (e.g. pulling at the trousers to show the need for toileting), through a system of pointing to symbols or pictures, to complex signing which can convey abstract concepts such as emotional states. Language may be verbal or non-verbal and both modalities are taught simultaneously, reinforcing each other so that a course in signing can improve speech. Whatever the system used, it depends on, and will be limited by, the extent to which those around can understand it.

(c) Social and sexual relationships

People are vulnerable to exploitation until they learn the distinction between an acquaintance and a friend. Relationships become complicated by sexual behaviour, hedged around with cultural rules which vary between families making it an especially difficult area to teach and therefore tempting to ignore. Occasionally sexual arousal drives disturbance in someone for whom masturbation is inefficient, physically damaging, unlearned, or forbidden. Wider discussion can bring out strongly held beliefs and family conflicts which had been unsuspected or denied until then. This may lead on to other areas requiring resolution, such as whether a person with intellectual disability (mental retardation) should have a sexual relationship, marry, or have children.

Out-of-home placement

As described earlier, a move out of the home can be seen either as part of the wider programme of support and social d evelopment or else as part of the treatment of a specific disorder: the distinction is often blurred.

(a) Support

Short breaks allow individuals and their families some relief from the uninterrupted intimacy of care. They also widen social networks and pave the way towards the eventual departure from home, something that is ideally planned as part of an increasing, adult autonomy. Frequently however, this is left until there is a crisis, for example when the person's behaviour or dependency has outstripped their family's resources, too often the result of parental infirmity or death. At this point the unfortunate individual may find themselves in a series of short break (or even treatment) placements until something long-term can be arranged.

Placement for educational needs (e.g. in a residential school or college) becomes necessary where the person's disabilities (e.g. autism or intractable epilepsy) require specialist skills and settings. It can also be a compromise with a parent who, unable to care for their child themselves, will not accept more standard forms of social care.

(b) Treatment

The threshold for admission for assessment and treatment will depend on the extent of the supportive service available in the community and is often:

1 For the treatment of more complex disorders such as epilepsy or psychosis.

2 For the assessment of disturbance where it is difficult to disentangle the relative contributions of innate from environmental factors. For example, a patient's behaviour may be amplified by an exasperated or exhausted family, particularly where disturbed nights have left them short of sleep. This sets off a secondary, self-perpetuating cycle of disturbance involving the whole house which makes it impossible to discern the underlying, primary disturbance. The cycle may only be interrupted by changing the patient's circumstances, either by moving staff into the home or by moving the patient out.

3 For the management of behavioural disturbance where:

(a) The carers are unable to cope—the more frequent reasons include:

- marital disharmony
- the carer's inability to manage others simultaneously, for example single parents who have several children or a home with several disturbed children
- the loss of resilience in a demoralized carer
- an adverse or hostile neighbourhood where the carer has to give way in case the patient becomes so noisy that the neighbours complain, or where the patient is bullied or led astray.

(b) There has been a failure of earlier therapeutic trials with a family locked in their pattern of behaviour.

(c) The patient is at risk of harm.

(d) The patient presents a substantial risk to others, for example, of violence or sexual offending.

(e) There is the need for effective control of the patient with clear limits to disturbed behaviour.

Treatment methods

Discomfort is a potent and frequent cause of disturbance. Because the person may have difficulty in identifying and localizing pain, let alone communicating symptoms, it is frequently missed. It is essential that it is a matter of routine to seek and treat common ailments such as hay fever, toothache, earache, dyspepsia, and gastric

reflux. The last of these is particularly associated both with severe retardation and with certain disorders such as Cornelia de Lange syndrome as well as an increased possibility of oesophageal carcinoma.[3] The more severe the degree of disability, the greater the need for an active programme of health care.[4]

Behavioural treatments

Intellectual disability's communication barrier has led to an emphasis on observable behaviour rather than on reports of subjective emotions or perceptions. Although the function of a behaviour is often misinterpreted, the first, unthinking response is often a behavioural programme. At the same time, increasingly sophisticated and more solidly research-based than other forms of psychiatric treatment, such a programme can produce profound and rapid change.[5] Programmes are divided broadly into two areas:

1 Teaching appropriate habits and skills which can range from basic skills such as dressing, continence, communication, and sleep, through to the more sophisticated training in social skills, dating skills, and assertiveness.

2 The unlearning of other, maladaptive forms of behaviour.

These two areas are complementary—it is more effective to replace an undesirable behaviour than simply to remove it. Operant, incentive programmes dominate much institutional and offender work. While punishment techniques can be effective and in certain, very unusual situations may be justified, there must be concern about their effect on the trainer as much as on the patient and they need close, ethical control.[6]

Cognitive behavioural therapy

Behavioural principles can be used to target thought as well as behaviour. With disability comes a tendency to a polarized perception of the world—people seeing themselves either as acceptable, competent, and successful or else as worthless failures; similarly, others are seen as all good or all bad. Cognitive therapy is usually used with people in the mild or borderline range of retardation, and has a particular importance in the treatment of sex offenders, although there are some examples of work with non-verbal people or those with severe intellectual disability. Indeed, where problem-solving skills are formally taught, the improvement may be greater in people with moderate than with mild intellectual disability, suggesting some form of ceiling effect to their acquisition. Furthermore, performance may have more to do with the type of problem than with formal measures of ability.[7]

Rational emotive therapy, developed in an educational setting, seeks to change the perceptual set and thereby the impact and influence of events on the person. Its background in education resulted in a didactic format, giving the therapist a directive role. The orientation is behavioural, focusing on the development of skills, and it lends itself to being taught to groups. A number of open studies of people with moderate and mild intellectual disability have shown it to decrease irrationality and anxiety, and to increase internal control and self-esteem.[8]

Anger management was developed with people of average ability but now has a well-established place in intellectual disability.[9] Therapy has to cope with a number of obstacles including:

1 The inherent nature of anger which means that the problem is an excessive response rather than a deviant one. In this population organic factors are frequent, particularly brain damage, epilepsy, and medication.

2 Its usefulness where there is limited communication or more severe disability which, combined with personality factors, can make therapeutic engagement difficult.

3 The degree to which habitual use has made anger an entrenched response.

4 The emotion of anger can be difficult to distinguish and label, particularly where autism is a component. Here an aggressive response may be the result of excessive anxiety, often amounting to panic.

A programme of anger management may take place at several levels.

1 General clinical care—strategies to reduce anger which include ensuring that the person feels well, that their physical and social environment is suitable, and that there are suitable occupational and recreational programmes.[10]

2 Anger management—information is given to help the person recognize anger, its nature, the signs, and consequences, as well as ideas and information about changing their behaviour. This uses a more didactic group instruction which is general and involves less disclosure and engagement by the individual.

3 Anger treatment—an individually tailored programme which targets change in cognitive perception, autonomic arousal, and behaviour. Individual engagement is essential and transference and countertransference are important and likely to evoke distressing emotions.

Another target is anxiety reduction through treatments that range, depending on the degree of disability, from formal relaxation training through to physical activity. Some do not measure up to their promoter's promise and may even have a detrimental effect,[11] endorsing the principle of treatment as an individual therapeutic trial.

Psychodynamic therapies

Although we have moved away from the early belief that intellectual disability itself was the result of emotional abuse and psychological disturbance, we are beginning to recognize the extent to which these factors can reduce adaptive functioning and amplify a cognitive disability. Psychotherapy can complement educational programmes to produce someone who, although no more intellectually able, is more mature emotionally and better able to cope with the tasks of everyday life. However, the field is poorly researched so that its efficacy is uncertain.[12]

Both the family and the individual have to adjust to disability; a process akin to a series of grief reactions through which people come to terms with the loss of normality.[13] The process of adjustment occurs as a series of crises triggered by events such as the point of initial diagnosis, the failure of initial treatments, educational assessment and specialized placements, puberty, and leaving home. Each stage brings home afresh the degree and significance of the child's disability.

The therapist may require specific training and supervision in adapting standard approaches to cope with a number of potential elements peculiar to intellectual disability:

1 The therapy is likely to have to deal with the core themes of loss and disability.

2 Communication will be limited in various ways and there may be unexpected, conceptual barriers. For example, the patient may not understand or even notice gestures, facial expressions, and different tones of voice. They may be unable to identify many emotions, label them, or form abstract concepts. Communication has to be at the patient's level, being concrete and using simple words, short sentences, and their colloquial or slang terms as well as allowing sufficient time for the patient to process the thought. The therapist may circumvent some of these barriers by using alternatives modalities, such as music, art, play, and drama.

3 The patient may have a limited and distorted understanding of the roles and relationships of the people around. Unable to appreciate either that the therapist's knowledge is restricted or that they occupy a different world, the patient may assume that they know all about the patient's setting and routines. This means that, in order to make sense of what is said, the therapist must learn something of the patient's background.

4 A combination of memory problems, obsessionality, or poor executive function means that there may be repetition of information, ideas, and conclusions, often given as if they have never been mentioned before.

5 Many people have had lives marked by short and changing relationships and a nomadic change of accommodation. Engagement in therapy can be difficult with a disconcerting readiness to disengage.

6 Confidentiality may be difficult to establish for someone where dependency and total care has discouraged privacy. Carers frequently expect to be told what happens in therapy and the relationship may be complicated by a disclosure that can range from poor care to frank abuse.

Psychopharmacological treatments

Drugs are widely used in intellectual disability. Besides being used for the medical disorders which are more frequent in this population, such as epilepsy, Tourette syndrome, and attention deficit disorder, drugs play a part in the management of a wide range of symptomatology which includes symptoms as varied as aggression, self-injury, outbursts of distress, compulsive routines, and social withdrawal. A recurrent criticism is that a drug may be used for a purpose other than that suggested by its classificatory label (e.g. that antipsychotics and antiepileptics are used for conditions other than psychosis or epilepsy, respectively) or that they are prescribed outside their manufacturer's licence. While this often is simply semantic, there must be concern at the level of prescribing of psychotropic drugs.[14] The prevalence and number of drugs being prescribed is associated with, the degree of retardation, and the presence of autism; it is no less in the community than in inpatient institutions;[15] and is only reduced by a determined programme of rationalization.

The presence of intellectual disability colours prescribing:

1 Non-compliance usually results from an unpalatable formulation or a carer's prejudice rather than forgetfulness.

2 Coexistent disorders, as wide-ranging as epilepsy, constipation, and cerebral palsy, can make a patient more vulnerable to adverse effects.

3 Cerebral dysfunction may cause more frequent atypical responses, sensitivity being either increased or decreased to various aspects of the drug's effect. Most of these are dose-specific, paradoxical effects for therapeutic windows are frequent and, as frequently, forgotten. Prescribing details are subject to revision and should be checked with a current formulary but treatment should be started at a lower dosage and increased more gradually than is generally recommended. The unexpected should be expected so that, besides the routine warning of adverse effects, carers should be able to contact someone if they are in any doubt about the drug's effects.

4 Evidence of efficacy is largely anecdotal, with trials being mostly small, open, and uncontrolled, but providing some justification for almost any neuropharmacological adventure. The information that needs to be weighed up is more complex and is therefore less likely to be within the patient's capacity to decide whether to take it. This leaves the prescriber with a greater responsibility to put the patient's interests first.

(a) Neuroleptics

These are used frequently and for a variety of symptoms despite a shortage of consistent, demonstrable, and specific effects. For example, although a series of studies have shown haloperidol to be effective in autism, the response can be in any of a variety of areas including improved discrimination learning, a reduction in overactivity, anger, and in the frequency and intensity of outbursts.[16] A number of reports suggest violence, whether to others or self-directed, might be more responsive to fluphenazine or clozapine (although the use of the latter is severely limited by its potential for marrow toxicity).

The recent proliferation of neuroleptics has been driven by a search for greater effectiveness together with a reduced risk of adverse effects and has been steered by the theoretical clinical attributes of various neuroreceptor systems. The atypical neuroleptics are reputed to bring less risk of adverse effects such as the dyskinesias, but, as they become better known, are being linked with sedation, weight gain, and elevated prolactin level and there is growing concern about their potential to produce a metabolic syndrome.[17] The most established of these, risperidone, has shown itself effective in several random controlled trials in reducing behavioural disturbance in children and adults across a number of symptoms such as aggression, social withdrawal, inattentiveness, and overactivity.[18]

(b) Antidepressants

Depression, once noticed and identified, is as treatable as in the normal population. Obsessive–compulsive symptomatology is frequent and responsive to the drugs augmenting serotoninergic transmission although there is increasing anxiety about potential adverse effects, particularly suicide and dependence although there is, as yet, no evidence for this in the population with intellectual disability. More frequent is a paradoxical increase in anxiety that may be a partial serotonin syndrome. The use of the selective serotonin reuptake inhibitors and lithium is being extended beyond the management of apparently compulsive violence and self-injury to include bouts of non-specific distress. Clomipramine, fluvoxamine, and fluoxetine have both shown success in random controlled trials but the evidence for the effectiveness of other drugs largely consists of open label and retrospective case-series in which it can be unclear whether the underlying disorder is autism, intellectual disability, or both.[19, 20]

The serotoninergic system appears to be central to a wide number of vegetative functions suggesting a potential for these drugs in appetite disorders such as the compulsive search for food and lack of satiety of Prader–Willi syndrome.

(c) β-blockers

Propranolol and nadolol limit the autonomic response to anxiety and the propensity to panic. They have a particular place where acute anxiety underlies aggression as, for example, in someone with autism who lashes out or flees in panic when feeling crowded. They are non-sedative but can cause lethargy and even depression. Open-label series suggest propranolol may take some weeks to take effect with a dosage range of 50–960 mg/day; characteristics that may explain the lack of random controlled trials.

(d) Stimulants

There has been a widespread re-evaluation of the place of attention deficit-hyperactivity disorder (ADHD), its prevalence, and management. Well recognized in children in the normal range of ability, it is now being identified more frequently in adulthood, coexistent with autism spectrum disorder, and with intellectual disability. There is a general, unsubstantiated belief that ADHD is more intractable, the greater the degree of intellectual disability. Methylphenidate and amphetamine remain the standard treatments but amantadine might be more effective and less toxic.[21] Clonidine can be of use although sedative and short-lived and atomoxetine is finding its place. When effective, stimulants can produce a global improvement in behaviour that includes appetite, sleep, and mood, even though anorexia, insomnia, and depression head the list of potential adverse effects.

(e) Mood stabilizers

Antiepileptic drugs are being more widely used to reduce emotional lability, particularly outbursts of rage which have an organic, possibly epileptic, basis; the episodic dyscontrol syndrome. Supportive evidence is slight and occasionally the drug may make matters worse.[22] Aggression has a large variety of causes and, after the exclusion of physical discomfort, the primary approach is psychological. However, lithium has shown to be effective in reducing outbursts of aggression, particularly where there is irritability and explosiveness.[16]

(f) Opioid antagonists

The hypothesis is that opioid excess might underlie both autism and its frequent associate, an indifference to pain that encourages excessive self-stimulation. Naltrexone has been used to treat both autistic disturbance and repetitive self-injury. Where autism shows any response it is to a very low dosage (5–20 mg/day), a window in a therapeutic U-curve while, if self-injury should respond, it is likely to be to a higher dosage (100–200 mg/day). Unfortunately, many of the random controlled trials have fallen between the two dosages.

(g) Antilibidinal drugs

For a few, a strong sexual drive overrides the teaching, training, and psychological therapies which, as in other areas of psychiatry, are the main approach to sexual offending. An antilibidinal agent[23] can supplement the other approaches, reducing the drive to a level where it is under the patient's control. The sensitivity of this area, let alone potential adverse effects, makes it especially important that the patient, the family, and the carers are all part of any decision to use medication.

Conclusion

There are a large number and variety of treatments for any disorder in intellectual disability: a good indication of the complexity of disturbance and the inconsistency of the effectiveness of any one treatment. Therapists have an unusual responsibility not to exploit the limited ability of their patients to withhold consent as well as to try to see the world through their eyes. They have also to combine simultaneously the enthusiasm of the charismatic healer with the objectivity and scepticism of the scientist. As in any experiment, changes in treatment should be introduced singly to ensure that it is clear which manoeuvre produces which result. However, treatments complement each other and should not be used in isolation. This is the area above all in which the patient depends on teamwork and cooperation between therapists, disciplines, and agencies.

Further information

University of Birmingham. (2007). *Using medication to manage behaviour problems among adults with a learning disability.* Available from: http://www.ld-medication.bham.ac.uk/.

Gualtieri, C.T. (2002). *Brain injury and mental retardation: psychopharmacology and neuropsychiatry.* Lipincott, Williams & Wilkins, Philadelphia.

Handen, B.L. and Gilchrist, R. (2006). Practitioner review: psychopharmacology in children and adolescents with mental retardation. *Journal of Child Psychology and Psychiatry*, **47**, 871–82.

References

1. Royal College of Psychiatrists. (2001). *DC-LD: Diagnostic criteria for psychiatric disorders for use with adults with learning disabilities/mental retardation* (Occasional Paper OP48). Gaskell, London.
2. Baum, S. (2007). The use of family therapy for people with learning disabilities. *Advances in Mental Health and Learning Disabilities*, **1**, 8–13.
3. Bohmer, C.J., Niezen de Boer, M.C., Klinkenberg Knol, E.C., *et al.* (1997). Gastro-oesophageal reflux disease in institutionalised intellectually disabled individuals. *The Netherlands Journal of Medicine*, **51**, 134–9.
4. Cooper, S.A., Morrison, J., Melville, C., *et al.* (2006). Improving the health of people with intellectual disabilities: outcomes of a health screening programme after 1 year. *Journal of Intellectual Disability Research*, **50**, 667–77.
5. Emerson, E. (1995). *Challenging behaviour: analysis and intervention in people with learning disabilities.* Cambridge University Press, Cambridge.
6. Repp, A.C. and Singh, N.N. (eds). (1990). *Perspectives on the use of nonaversive and aversive interventions for persons with developmental disabilities.* Sycamore Publishing Company, Sycamore, IL.
7. Benson, B.A. and Valenti-Hein, D. (2001). Cognitive and social learning treatments. In *Treating mental illness and behavior disorders in children and adults with mentallly retardation* (eds. A. Dosen and K. Day). American Psychiatric Press, Washington, DC.
8. Schneider, N. (2001). A rational emotive group treatment approach with dually diagnosed adults. In *Treating mental illness and behavior disorders in children and adults with mental retardation* (eds. A. Dosen and K. Day). American Psychiatric Press, Washington, DC.
9. Taylor, J.L. and Novaco, R.W. (2005). Anger treatment for people with developmental disabilities: a theory, evidence and manual based approach. John Wiley & Sons, Chichester.
10. Harris, J., Allen, D., Cornick, M., *et al.* (1996). *Physical interventions: a policy framework*. BILD, Kidderminster.

11. Lindsay, W.R., Pitcaithly, D., Geelen, N., *et al.* (1997). A comparison of the effects of four therapy procedures on concentration and responsiveness in people with profound learning disabilities. *Journal of Intellectual Disability Research*, **41**, 201–7.
12. Royal College of Psychiatrists. (2004). *Psychotherapy and learning disability*. (CR116). Royal College of Psychiatrists, London.
13. Bicknell, J. (1983). The psychopathology of handicap. *British Journal of Medical Psychology*, **56**, 167–78.
14. Lott, I.T., McGregor, M., Engelman, L., *et al.* (2004). Longitudinal prescribing patterns for psychoactive medications in community-based individuals with developmental disabilities: utilization of pharmacy records. *Journal of Intellectual Disability Research*, **48**, 563–71.
15. Branford, D. (1997). A follow-up study of prescribing for people with learning disabilities previously in National Health Service care in Leicestershire, England. *Journal of Intellectual Disability Research*, **41**, 339–45.
16. Campbell, M. and Cueva, J.E. (1995). Psychopharmacology in child and adolescent-psychiatry—a review of the past 7 years, Part 1. *Journal of the American Academy of Child and Adolescent Psychiatry*, **34**, 1124–32.
17. McKee, J.R., Bodfish, J.W., Mahorney, S.L., *et al.* (2005). Metabolic effects associated with atypical antipsychotic treatment in the developmentally disabled. *The Journal of Clinical Psychiatry*, **66**, 1161–8.
18. Jesner, O.S., Aref-Adib, M., and Coren, E. (2007). Risperidone for autism spectrum disorder. *Cochrane Database of Systematic Reviews*.
19. Dinca, O., Paul, M., and Spencer, N.J. (2005). Systematic review of randomized controlled trials of atypical antipsychotics and selective serotonin reuptake inhibitors for behavioural problems associated with pervasive developmental disorders. *Journal of Psychopharmacology*, **19**, 521–32.
20. Posey, D.J., Erickson, C.A., Stigler, K.A., *et al.* (2006). The use of selective serotonin reuptake inhibitors in autism and related disorders. *Journal of Child and Adolescent Psychopharmacology*, **16**, 181–6.
21. Gualtieri, C.T. (2002). Psychostimulants, dopamine agonists and amantadine. In *Brain injury and mental retardation: psychopharmacology and neuropsychiatry*. Lipincott, Williams & Wilkins, Philadelphia.
22. Besag, F.M.C. (2001). Behavioural effects of the new anticonvulsants. *Drug Safety*, **24**, 513–36.
23. Cooper, A.J. (1995). Review of the role of two antilibidinal drugs in the treatment of sex offenders with mental retardation. *Mental Retardation*, **33**, 42–8.

10.7

Special needs of adolescents and elderly people with intellectual disability

Jane Hubert and Sheila Hollins

Introduction

Social health and mental health needs change throughout life, and this chapter highlights those particularly relevant for adolescents and elderly people. As a general rule, people with intellectual disabilities have the same needs as other members of the community, but they may also have additional needs for which they are entitled to extra support.[1]

Adolescents

Administrative prevalence of intellectual disability in adulthood vs. childhood

The UK Government White Paper 'Valuing people' estimates that there are about 2 10,000 people with severe and profound intellectual disabilities in the UK: around 65,000 children and young people, 1 20,000 adults of working age, and 25,000 older people. They estimate that there are some 1.2 million people in the UK with mild or moderate intellectual disabilities. Worldwide, it is estimated that there are some 20 million people with intellectual disabilities.[2]

Although there is now a trend in the UK towards mainstreaming children with special needs, and providing extra support, separate special schools for children with moderate and for severe learning difficulties are still provided in many places. At school leaving age, many young people with mild or moderate learning difficulties (roughly equating to IQ > 50) will not receive special services; only people who have severe intellectual disabilities, and those with additional disabilities, including epilepsy, autism, mental illness and/or behavioural problems will be referred on to adult specialist services. The administrative prevalence of adults with intellectual disabilities is thus much lower in adulthood as it is a measure of those in contact with services. The administrative prevalence rates should not be confused with true prevalence rates, which are far more difficult to assess.

Transition to adulthood

Among young people in the general population, there are certain important life events which are usually considered necessary for a successful transition to adulthood. These include getting a job or going to college, having economic and social independence from parents, and leaving home.[3] Although the transition to adulthood can be a difficult and painful process for anyone, for most people it is also a time at which choices and opportunities open up. For people with intellectual disabilities, the transition to adulthood does not usually follow the same pattern as it does for others. For many, the transition is marked simply by an abrupt move from the protective and relatively well-defined children's services to adult services, and by leaving school. These imposed transitions into adulthood are often abrupt and traumatic for the young people and for their parents.[4]

Overall, the criteria for 'successful' transition to adulthood are less likely to be fulfilled the greater the severity of the intellectual disability and other factors such as physical or mental health problems, communication difficulties and/or challenging behaviour. Those who have mental health problems, or aggressive challenging behaviour, are particularly unlikely to receive the necessary support and services to enable them to live independent 'adult' lives.[5] In 2001, a new service, Connexions, was established to improve the management of the transition to adulthood, by providing young people from 13–19 years of age with access to advice, guidance and support.[2]

In the UK, the majority of young people with intellectual disabilities attend Day Centres. Those who are more able may enter sheltered employment, workshops for disabled people or supported open employment. Relatively few people with intellectual disabilities are in paid employment, and although employment schemes are now being developed in many places, there are substantial barriers that are faced by people with intellectual disabilities in getting and maintaining employment in the open job market.[2]

In all parts of the world, the majority of young people with intellectual disabilities continue to live at home with one or both parents, and often have little or no social or economic independence, or participation in major, or even minor, life decisions. Although adolescence and leaving school imply a transition to adulthood, in many cases young people become more dependent

on their parents at this stage than they were before. For those who have severe or multiple disabilities, and/or challenging behaviour, there may be few practical alternatives and choices open to them.[6]

In different countries, various approaches have been developed to recognize the needs of young adults. One widely adopted approach, the development of small group homes in the community, has meant that more young people with intellectual disabilities are able to move away from home, even some who have severe intellectual disabilities and challenging behaviour. In many countries however, the responsibility remains firmly with carers, with institutional provision being the only backup when family care breaks down. An international carers' advocacy organization, Inclusion International, researched the views of carers in 80 or more countries in both developed and developing nations, and their report makes a number of recommendations about how communities and governments can provide better support to individuals and families.[7, 8]

Health needs

There is a high prevalence of epilepsy, psychiatric disorder, hearing and visual impairments and autism among people with intellectual disabilities. Children with intellectual disabilities are the responsibility of a paediatrician, and parents can discuss and monitor their children's needs and progress through one agency. When a child is transferred to adult services this situation changes, and there are many different agencies and individuals who become responsible for different aspects of the overall service to adolescents.

For the families concerned, the world of adult services can be bewildering. The situation is particularly problematic in relation to adolescents with severe intellectual disabilities, especially if there are also behaviour problems, during this transitional phase from child to adult services. Impairments in adaptive behaviour associated with intellectual disability lead to problems in developing normal social functioning, communication, and the ability to use community facilities. In addition, the relationship between parental and professional roles and responsibilities is often unclear. Multidisciplinary assessment is advisable, and parents should remain involved, but all too often are told that their opinion is no longer valid now their child is an adult.[2] The Royal College of Psychiatrists publishes leaflets for family carers to help them manage these changing professional relationships.[9]

It is often not apparent who, among the professionals, is directly responsible for someone in the context of services, and there may be inconsistencies between Health, Education, and Social Services in terms of policies and practice. Also, health professionals, including general practitioners, may be relatively inexperienced in dealing with people who have intellectual disabilities.

A coherent strategy for developing comprehensive health care services for young people with intellectual disabilities requires collaboration between service providers, to ensure that the health care needs of all people with intellectual disabilities, including those with autism, are properly identified, and that access to mainstream primary and secondary health care is supported. One initiative developed for the white paper 'Valuing people' was the introduction of Health Action Plans to try to address some of these information and knowledge gaps.[10]

Mental health needs

Diagnostic overshadowing of mental illness in people with intellectual disabilities was common in the past, but there is now increasing awareness and assessment of psychiatric disorders, and acceptance of dual diagnosis among people with intellectual disabilities. Although mental health needs can in some cases be met by general mental health services, some specialized mental health provision is still necessary to meet the needs of people with dual diagnosis, including those who also have challenging behaviour.[11]

Until recently, people with intellectual disabilities were seldom thought to suffer from depression, but recent research shows that adolescents with intellectual disabilities report more depression and other symptoms of psychopathology than others without intellectual disabilities.[12]

There is increasing awareness, and continuing evidence,[13,14] of the high prevalence of abuse of people with intellectual disabilities, of all ages, including emotional, physical, and sexual abuse, resulting in Post Traumatic Stress Disorder,[15] severe behavioural disorders[16] and damaging long term effects on the family as a whole.[17] Challenging behaviour in people with intellectual disabilities may be indicative of psychiatric disorders, such as psychosis, depression, and anxiety disorders.

A recent report concludes that people with intellectual disabilities who present behavioural challenges are often marginalized, stigmatized, disempowered, and excluded from mainstream society,[18] indicating the need for changes in policy and practice.

Sexual relationships, marriage, and parenthood

Long-term sexual relationships and parenting children are generally considered to be an integral part of being an adult. In adolescence, emotional and sexual interest and needs develop, and it is at this stage that most young people start to have sexual relationships. However, people with intellectual disabilities are seldom encouraged to develop sexual relationships. Parents tend to actively discourage it, and service managers and care staff, though they may not necessarily actively discourage it, often provide little opportunity, or privacy, to enable it to happen. Many people in the general population find it difficult to accept that men and women with intellectual disabilities have ordinary sexual feelings and desires, let alone that they should be allowed to act on them.[19] The argument against allowing people with intellectual disabilities to have sexual partners often involves judgments about whether someone is deemed fit to be a parent. People with intellectual disabilities are discouraged from parenthood, and the experiences of childbearing and child rearing are still usually denied to women with intellectual disabilities. In Norway, 40 per cent of a study cohort of 126 children born to parents with intellectual disabilities were found to have suffered from 'failures of care'.[20] In England, however, research has demonstrated that some people with intellectual disabilities can become successful parents, provided they are given appropriate and effective support.[21]

Cultural differences

People with intellectual disabilities from black and ethnic minorities are less likely to have their needs met by service organizations as compared with the rest of the population.[22] This is not only the result of difficulties in accessing services, and lack of appropriate

information, but also because too little attention is paid to the different social norms, beliefs and preferences of people from different cultural backgrounds.

It is vital that service planners and providers know in what ways and to what extent the belief systems of the people they provide services for coincide and/or conflict with their own. They must also be aware of the implications of these differences for the acceptability, expectations and outcomes of the services offered to people from different cultural groups.

Elderly people with intellectual disability

Life expectancy

People with intellectual disabilities have an increased risk of death compared with the general population. Whereas the majority of deaths (83 per cent) in the whole population in the UK occur in people aged 65 years and over, less than 50 per cent of deaths among people with intellectual disabilities are in this age group.[23] In a study of young people in one state in the US, the mortality rate was almost three times higher than average,[24] and in Denmark 'preventable' mortality was four times higher than average.[25] However, life span is increasing among people with intellectual disabilities,[26] especially among people with Down's Syndrome.

As a result of this increasing longevity, causes of death common in a normal ageing population are becoming more prevalent among people with intellectual disabilities, such as stroke, heart disease and cancer. The most common cause of death for people with intellectual disabilities is still respiratory disease, which occurs far more frequently than in the whole population, suggesting lack of effective care.[27] This cause of death is linked to pneumonia, swallowing and feeding problems, and gastro-oesophageal reflux disorder.

People with intellectual disabilities frequently suffer from epilepsy, and it is suggested that the mortality rate for people with epilepsy and intellectual disabilities 'may be as high as five times that of the general population'.[28] This high mortality rate is related to seizure type and frequency, rather than directly to seizures.

Health needs

Cooper[29] reviews the effects of age on the physical health of people with intellectual disabilities, and stresses the existence of significant health needs among this population. These needs arise not only from the normal ageing process but also from the specific social health and mental health needs of people with intellectual disabilities, including dementia.

There are serious problems relating to access to health care,[30] which is further complicated by their failure, and the failure of their carers, to recognize the signs and symptoms of illness. Overall, uptake of services by elderly people with intellectual disabilities is poor. There is an increased risk for a number of medical conditions in people who have Down's Syndrome,[31] including sensory impairments, thyroid disease, leukaemia and atlanto-axial instability. The later consequences of congenital heart disease include pulmonary hypertension and congestive heart failure.

Carers may not recognize that changes in behaviour are due to physical or mental illness, instead attributing changes to the learning disability itself. It is important to determine the aetiology of any learning disability, even late in life, because of the possible health implications.

People with intellectual disabilities currently access health screening less than others in the general population.[32] Pictorial health education materials are available to help health care professionals provide information about illness, medical procedures and treatment to people with limited verbal communication (for example, see www.rcpsych.ac.uk/bbw).

Signs of poor physical care among elderly people with intellectual disabilities, e.g. eye infections or tooth decay, may indicate a deterioration in functioning, but may also reflect the fact that carers are not coping effectively. This emphasizes the need to ensure access to primary care.

Recent reports from the Disability Rights Commission[33] and Mencap[34] identify the existing health inequalities faced by people with intellectual disabilities, and cite evidence suggesting the discrimination they face at every level of the health service.

Recent studies of the prevalence of mental ill-health problems among people with intellectual disabilities have shown that the prevalence rate of mental ill-health among adults with intellectual disabilities is higher than those recorded in the general population, with anxiety states and depression increasing with age.[35] People with intellectual disabilities are prone to the same risk factors as other people, but there are additional ones, such as living in more deprived areas, not having any daytime occupation, single marital status and epilepsy, all factors associated with mental ill-health. These factors need to be addressed if the existing inequality gap is not to be widened further.

People with Down's syndrome have an increased risk of developing Alzheimer's disease in middle age. Although the neuropathological changes of Alzheimer's dementia are widespread, development of clinical dementia is not inevitable.[36] Dementia occurs more frequently among elderly people with intellectual disabilities in general than among the rest of the population, and with the increasing longevity of people with intellectual disabilities the number with dementia is rising.[37]

Complexity of care needs of elderly carers and the people they care for at home

In households where an older person with intellectual disability is living with an elderly carer, there will be a complex set of individual and joint needs. Both are likely to be vulnerable at this stage in their lives, and their needs may not always be compatible.

Parents may reach a point where physical mobility and capabilities are declining, and in some cases dementia (in either), may complicate the situation. Input from services becomes essential, even if families have managed with few or no services until now. Many families will have relied on informal sources of support such as family members, friends and neighbours, but in later life these networks tend to break down, and households such as these become increasingly isolated. This isolation in the community tends to coincide with the increasing frailty of ageing carers. [38, 39] Parents of children with severe disabilities, and/or challenging behaviour, may well become isolated from kin and friends at a much earlier stage as a result of their dedicated caring role, increasing the likelihood of social isolation in later years.

Some elderly parents who come to light at this late stage will have made a decision not to accept help many years ago. This decision will have been made on the basis of information about services which were available up to 40 or 50 years ago, and elderly carers may not have received information about the range and nature

of current services. There are also some parents who originally concealed their child's disabilities in order to avoid institutionalization, and these parents may still be unaware of less institutionalized alternatives to care at home.

Systems of mutual caring often develop in households such as these, and such families may continue to be independent, in spite of serious long-term problems, and only come to the attention of the services when crisis intervention becomes necessary.

Families such as these present a double challenge to the services, requiring co-ordination between all relevant service providers. Co-operation between services for the elderly and those for people with intellectual disability is often inadequate, and some older people with intellectual disability fall into a limbo between them, with no one taking overall responsibility for assessing and meeting their needs. However, part of the brief of local Partnership Boards in the UK is to ensure that there is co-ordination between the services, so that individuals can obtain the range of services that they need.[2]

Planning for the future

Ageing parents often wish to continue to look after their adult child until they can no longer cope. If a son or daughter has mild intellectual disability this may not cause problems for the parents, and may represent a welcome continuation of family life, especially when other children leave home. Also, as parents grow older they may become increasingly dependent on their adult child with intellectual disability. This may outweigh their wish to see their son or daughter settled in a new home before they themselves die, or become too frail to care for them.

Although keeping an adult with severe disabilities at home may cause considerable hardship, many parents are unwilling to let their child go into residential care, because they believe that only they can provide the quality of care that he or she is used to.

Although separation and/or bereavement are likely to occur in the relatively near future, professionals working with families find that many elderly carers are reluctant to plan for their child's future care, and attempts to develop care plans are often fraught with anxiety.[40]

Bereavement

The experience of bereavement is often ignored by people involved with people with intellectual disabilities, although the effects of bereavement among this population are particularly severe and long-lasting, and there may be a significant increase in aberrant behaviours, and an increase in psychopathology. The experience is often made more difficult by their exclusion from the rituals and processes associated with dying and death. The onset of grief may also be delayed and thus there is a greater chance that it will not be recognized as grief, but attributed to something else or labelled as challenging behaviour.[41]

The experience of bereavement of a parent is particularly hard because, in addition to trying to come to terms with this loss, there are often other major life changes, such as moving into a strange home, living with unfamiliar people, and being cared for by a number of new carers.

Life changes in old age

Any major life changes in old age can have serious emotional and physical consequences. Life events which can trigger physical or mental deterioration include institutionalization, when caring parents die or become too frail to continue caring. This often sudden upheaval following unexpected separation or loss of a parent can be extremely traumatic.[42] Conversely, deinstitutionalization, i.e. moving into the community after a lifetime in a long-stay institution, requires complex and sympathetic planning and monitoring.[43]

Cultural differences

The implications of the lack of specific policies for older people with intellectual disability are particularly relevant to those from black and ethnic minority groups. Their needs may not be appropriately met by existing day services, but the alternative, i.e. living in residential homes in which staff and residents do not share the same background or language, may result in increased isolation in old age. Appropriate community services, which respond to different cultural preferences and expectations, have been developed in some areas of Britain. In general, however, service providers tend to focus on the issues of age and intellectual disability, rather than the cultural background of those individuals who have different expectations and preferences in the context of age and approaching death.[44]

Further information

St George's, University of London www.intellectualdisability.info Down's Syndrome Association UK http://www.downs-syndrome.org.uk

British Institute of Learning Disability www.bild.org.uk

The Department of Health – Learning Disabilities www.doh.gov.uk/learningdisabilities

Down's Syndrome Medical Interest Group www.dsmig.org.uk

The Foundation for People with Learning Disabilities www.learningdisabilities.org.uk

International Association for the Scientific Study of Intellectual Disabilities www.iassid.org

References

1. Royal College of Nursing (2006). *Meeting the health needs of people with learning disabilities*. Royal College of Nursing, London.
2. Department of Health (2001). *Valuing people: a new strategy for learning disability for the 21st Century*. The Stationery Office, London.
3. O'Brien, G. (2006). Young adults with learning disabilities: a study of psychosocial functioning at transition to adult services. *Developmental Medicine and Child Neurology,* **48**, 195–9.
4. Morris, J. (1999). *Hurtling into a void: Transition to adulthood for young disabled people with 'complex health and support needs'.* Pavilion Publishing Limited, Brighton.
5. Hollins, S. and Hollins, M. (2005). You and your child: making sense of learning disabilities. Karnac Books, London.
6. Mansell, J., Ashman, B., Macdonald, S., *et al.* (2002). Residential care in the community for adults with intellectual disability: needs, characteristics and services. *Journal of Intellectual Disability Research*, **46**(8), 625–33.
7. Inclusion International (2006). *Hear Our Voices: People with an Intellectual Disability and their Families Speak out on Poverty and Exclusion: a global report*. Inclusion International, London.
8. Dawson, E., Hollins, S., Mukongolwa, M., *et al.* (2003). Including disabled children in Africa. *Journal of Intellectual Disability Research,* **47**(3), 153–4.
9. Royal College of Psychiatrists (2004). *Factsheet 10: The child with general learning disability: for parents and teachers*. Mental Health and Growing

Up (3rd ed.). The child with general learning disability. Royal College of Psychiatrists, London.
10. McCoubrie, M., Hollins, S., Beckmann, R., *et al*. (2006). Health Action Plans: Some Guidelines for General Practitioners and Primary Care Teams. *Learning About Intellectual Disabilities and Health*. http://www.intellectualdisability.info/how_to/HAPs.htm
11. Bernal, J. and Hollins,S. (1995). Psychiatric illness and learning disability: a dual diagnosis. *Advances in Psychiatric Treatment,* **1**, 138–45.
12. Maag, J. W. and Reid, R. (2006). Depression Among Students with Learning Disabilities: Assessing the Risk *Journal of Learning Disabilities,* **39**(1), 3–10.
13. Commission for Healthcare Audit and Inspection and Commission for Social Care Inspection (2006). *Joint investigation into the provision of services for people with learning disabilities at Cornwall Partnership NHS Trust*. Commission for Healthcare Audit and Inspection, London.
14. Commission for Healthcare Audit and Inspection (2007). *Investigation into the service for people with learning disabilities provided by Sutton and Merton Primary Care Trust*. Commission for Healthcare Audit and Inspection, London.
15. McCarthy, J. (2001). Post-traumatic stress disorder in people with learning disability. *Advances in Psychiatric Treatment,* **7**, 163–9.
16. Sequeira, H., Howlin, P., Hollins, S., *et al*. (2003). Psychological disturbance associated with sexual abuse in people with learning disabilities: case-control study. *British Journal of Psychiatry,* **183**, 451–6.
17. O'Callaghan, A.C., Murphy, G., Clare, I.C.H., *et al*. (2003). The impact of abuse on men and women with severe learning disabilities and their families. *British Journal of Learning Disabilities,* **31**(4), 175–80.
18. Royal College of Psychiatrists. (2007). *Challenging Behaviour: A Unified Approach*. Report CR144. Royal College of Psychiatrists, London.
19. Heyman, B. and Huckle, S. (1995). Sexuality as a perceived hazard in the lives of adults with learning difficulties. *Disability & Society,* **10**, 139–56.
20. Morch, W. T., Skar, J., Andersgard, A. B., *et al*. (1997). Mentally retarded persons as parents: prevalence and the situation of their children. *Scandinavian Journal of Psychology,* **38,** 343–8.
21. Booth, T. www.intellectualdisability.info (accessed August 13 2007).
22. Mir, G., Ahmad, W., Jones, L., *et al*. (2001). *Learning difficulties and ethnicity*. Department of Health, London.
23. Hollins, S., Attard, M. T., von Fraunhofer, N., et al. (1998). Mortality in people with learning disability: risks, causes, and death certification findings in London. *Developmental Medicine and Child Neurology,* **40**, 50–6.
24. Decoufle, P. and Autry, A. (2002). Increased mortality in children and adolescents with developmental disabilities. *Paediatric and Perinatal Epidemiology,* **16**(4), 375–82.
25. Mölsa, P. K. (1994). Survival in mental retardation. *Mental Handicap Research,* **7**, 338–45.
26. Holland, A.J.(2000). Ageing and learning disability. *British Journal of Psychiatry,* **176**, 26–31.
27. S. D. R. Stoddart, E. C. Griffiths, and R. J. Lilford. (2005). *A Confidential Inquiry into Excess Mortality in Learning Disability*: Scoping Report to the National Patient Safety Agency.
28. Wilcox, J. and Kerr, M. (2006). Epilepsy in people with learning disabilities. *Psychiatry,* **5**(10), 372–7 (373).
29. Cooper, S-A. (1998). A clinical study of the effects of age on the physical health of people with intellectual disabilities. *American Journal on Mental Retardation,* **102**, 582–9.
30. NHS Service Delivery and Organization (SDO) (2004). *Access to health care for people with learning disabilities*. NHS National Institute for Health Research, London.
31. Holland, T. and Benton, M. (2004). *Ageing and its consequences for people with Down's Syndrome*. Down's Syndrome Association, London.
32. NHS Service Delivery Organization (SDO) (2004). *Access to health care for people with learning disabilities*. NHS National Institute for Health Research, London.
33. Disability Rights Commission (2006). *Equal Treatment: closing the gap*. A formal investigation into physical health inequalities experienced by people with learning disabilities and/or mental health problems. Disability Rights Commission, London.
34. Mencap (2007). *Death by Indifference.* London: Mencap
35. S-A. Cooper, E. Smiley, J. Morrison, A. Williamson and L. Allan. (2007). Mental ill-health in adults with intellectual disabilities: prevalence and associated factors. *British Journal of Psychiatry,* **190**, 27–35.
36. Stanton, L.R., and Coetzee,R.H. (2004). Down's syndrome and. dementia. *Advances in Psychiatric Treatment,* **10**: 50–8.
37. Cooper, S. (1997). High prevalence of dementia among people with learning disabilities not attributable to Down's syndrome. *Psychological Medicine,* **27**, 609–16.
38. McGrath, M. and Grant, G.(1993). The life-cycle and support networks of families with a person with a learning difficulty. *Disability, Handicap and Society,* **8**, 25–42.
39. Hubert, J. and Hollins, S. (2000). Working with elderly carers of people with learning disabilities and planning for the future. *Advances in Psychiatric Treatment,* **6**, 41–8.
40. Walker, C. and Walker, A. (1998). *Uncertain Futures: people with learning difficulties and their ageing family carers*. Pavilion Publishing/ Joseph Rowntree Foundation, Brighton.
41. Hollins, S., and Esterhuyzen, A. (1997). Bereavement and grief in adults with learning disabilities. *British Journal of Psychiatry,* **170**, 497–501.
42. Mencap (2002). *The Housing Timebomb: the housing crisis facing people with a learning disability and their older parents*. Mencap, London.
43. Owen, K. (2004). *Going Home? A study of women with severe learning disabilities moving out of a locked ward*. University of London/Judith Trust, London: St George's.
44. Mir, G., Ahmad, W., Jones, L., *et al*. (2001). *Learning difficulties and ethnicity*. Department of Health, London.

10.8

Families with a member with intellectual disability and their needs

Ann Gath and Jane McCarthy

Introduction

It is now more than 30 years since children with intellectual disability were among the first patients to emerge from long-stay hospitals, where the mothers had been persuaded to part with the disabled children. The argument given was that not only would that child have the best possible chance of a happy life, but so also would the other children. Fears that there would be adverse effects on parents and on brothers and sisters prompted much of the early research.[1]

Assessment of effects on the family

Recent sophisticated methodology has been used to explore a variety of factors, including the family as a whole as well as the parents, and what impinges on the family such as wider social, economic, and cultural influences.[2,3] Positive adaptation and coping strategies within families were identified and are highly relevant in providing a basis for intervention. Other life events and protective or compensating influences are not ignored. Families with a disabled member are exposed to the same risk of adverse factors, such as poverty, divorce, unemployment, or mental illness as any other and, in most cases, will have the same strengths, such as humour, good friends, or staunch relatives as their neighbours. Previously, all the complaints were added together as a measurement of 'stress', a concept too amorphous to be the basis of helpful intervention.

The early impact on a family of a disabled child

Diagnosis of an abnormality now frequently happens in pregnancy from screening tests or from ultrasound scans, all of which are routinely offered. Termination of pregnancy is offered when results are positive. Although negative tests by no means guarantee normality, they are often interpreted by the parents as meaning that major disability is ruled out. Hence their disappointment when a child is born with a defect is even more intense, and it often follows a prolonged highly anxious period in which the baby is in special care. Initial hope is followed by temporary relief and then by the reality of gross developmental delay. Others believe they have a normal child until they become aware of the slowness of development or the onset of seizures occurs in the second half of the first year. Parental reaction to these tragedies is often anger mixed with grief.

A major change in recent years has been the rapid expansion of intra-vitro fertilization. Although accurate figures are not available, it is estimated that at least 10 000 babies are born in the United Kingdom. Foetal abnormalities are more common than following normal conception as evidenced from clinical reports, but there is no data available about the effect on the parents who already have had much anxiety.

The mixed feelings at the time of the initial impact

The feelings that parents experience have been likened to those of grief occurring with a sudden loss. It is a useful comparison as there is a loss. Every expectant parent daydreams about the child and the arrival of a sick or damaged baby destroys many of those dreams. Commonly the first stage is shock or a numb disbelief. The next phase is often denial, 'This cannot be happening to me', followed by anger, which may be directed against the other parent, the doctor, or God. The last two phases are constructive active adaptation, which might involve learning about the condition or joining a parents association, leading on to resolution. Unfortunately, not everyone goes through all these stages and certainly not at the same rate. The mother might still be feeling as if she is shell-shocked while the father is making contact with a particular society or support group on the Internet.

The effect of diagnosis

Down syndrome is the most common disorder recognizable at birth and known to be likely associated with intellectual disability. Despite screening in early pregnancy, the condition is still common but is now often in the child of younger parents. Genetic diagnosis of unusual children is more rapid than hitherto. Parents are almost without exception relieved by a clear diagnosis, providing an explanation of why the condition has arisen as well as an estimate of future risk. Most families are also greatly helped by meeting others with similar problems. Parent support groups for each specific diagnosis are now worldwide and recruits are quickly introduced

to information via the Internet. Some diagnoses have implications for other members of the family, as with fragile X syndrome or tuberose sclerosis, as some relatives may have a minor form or be carriers, with a risk of further children in the family being affected. Genetic counselling is essential and may be requested very early on, in pregnancy or even before conception. To be preventative, decisions have to be made fast and at times when young couples are at their most vulnerable.[4]

The effect on the parents of a child with intellectual disability

Informal support from the family or neighbours is much more effective than more formal, professionally led support. Frequent outpatient appointments where little happens are not cost-effective for the family; often the father loses money as he must take time off work and the cost of travel with a difficult baby or toddler is high. Not all can benefit from discussion groups of parents, and other mothers find a succession of home visits from a variety of professionals very disruptive. Families with active participation in religion can strengthen family ties, especially among immigrant groups, such as Hispanic people in the United States, and Indian families, where Hinduism is central to family life and where children have specific roles to play, such as that of sons in funeral rites. A child with an abnormality makes many parents look again at their fundamental beliefs, but few make lasting changes. Other families find membership of other groups (social, cultural, even sporting) supportive, provided that the family feels that they and their child are unconditionally accepted.

Family functioning with a child with intellectual disability at school

The finding concerning the greater efficacy of informal support holds true for families of children across the whole age range, from school entry to adult life. The children are enrolled into school or special preschool groups earlier than normal brothers and sisters, thus widening the informal network of friends and confidants. Conversely, policies about choice of school, and the frequent necessity for children to be sent to schools at a greater distance away from home than other children, can lead to ostracism often felt more by the mother than by the child himself. An advantage of a special school is the relatively small size, allowing personal teacher–parent association and an open-door policy to parents, who are agreeably surprised to find themselves enjoying the school years of their disabled children.

Transition to adult life

For many years, families have learnt to work in partnership with schools, and enjoyed frequent contact with teachers, face to face or thorough the progress book that goes to and from school everyday. The last few years at school are much concerned with the choice of type of further education, sometimes residential, and with encouraging independence. The process of finding a suitable and acceptable place is often described as a lottery or a battle, and is very stressful for the parents. Many have struggled through the school years, hoping that a permanent placement will be found when they come to an end. Others fear the loss of their close contact with their child, particularly if given adult rights to make choices, with which the parents do not agree. The possible outcome is one of three: independence, semi-independence, and dependence.

For those who remain at home, the other children in the family leave home as expected, leaving their disabled brother or sister in what one mother described as 'a ghetto of the middle-aged'. However, in many families, the provision of good further education programmes or day centres plus club or leisure activities can lead to liberation and more happiness for all members of the family. It is those with severe behaviour difficulties who are not accepted by further education establishments and who become increasingly frustrated and difficult to manage at home.

Other members of the family

In many countries, the family has changed markedly in the last 30 years. There are many more divorces so that the 'parents' involved in the care of a child or young person are frequently one natural parent and one step-parent. There are also many divorced mothers living alone with the young adult after all the siblings have gone. Although some have adapted very positively, others feel very lonely, particularly if there are no members of the wider family to share the care and, often more pressing, to share the worries. Grandparents are as important in families with a member with intellectual disability as they are in ordinary families, although initially grandparents can become severely affected by the grief, take sides in attributing blame, or offer unsought advice. One mother described her mother as an enormous help because 'she was always behind me in every decision I took'. When no helpful grandparent is available, an older neighbour, another member of a parents group, or a teacher at the school or day centre could provide the sort of informal support that has proved to be so valuable.

Ageing parents and ageing 'children'

The physical work of looking after a still dependent and sometimes very heavy adult takes its toll on parents. With a severely physically handicapped adult 'child', some help can be provided with people coming in to help with bathing, but few houses can be adapted to minimize lifting which may be needed many times during the day and night. Sooner or later the work gets too much, particularly for a sole parent. Parental frailty makes aggressive behaviour much more frightening and potentially dangerous. Where there are few opportunities for outside contact, the adult, with intellectual disability complicated by severe behaviour disorder, can become possessive and may show jealousy, sometimes making visits even from grandchildren impossible. Despite the evident difficulties, a study in Wisconsin,[5] found that many parents, reaching the end of their lives after many years of looking after a disabled child, were fitter than others of the same age and had a much greater sense of having achieved something in life. Other studies of ageing people with intellectual disability have also shown the role reversal that occurs when the adult child for whom they have cared so long tenderly looks after a very old parent. Because of the increasing longevity of disabled adults as well as of older people in the general population, there are increasing numbers of very old frail parents left with a disabled offspring for whom they feel responsible and for whom they often feel anxiety about the future, commonly saying 'I always thought he would go before us'.

Brothers and sisters

The well-being of the family members continues to be an area of interest with an emphasis on siblings.[3, 6] The initial decision had

been made not just for the sake of the child with intellectual disability, but also in the belief that it was also in the interest of the other brothers and sisters, but it was possible that they who would pay the price. There is now a considerable body of literature confirming the early findings that siblings are by no means invariably damaged. Thirty years ago, there was evidence that the older girls in the families did suffer, or were difficult or distressed at school while having more than usual amounts of responsibility at home. As services improved, these findings were no longer replicated except in those countries with few facilities and many social problems. In general, the other children in the family have identified themselves with their parents' decisions and take some part in the caring. This 'assistant' parent role comes easily to older siblings, but younger siblings who grow up fast, first catch up with the disabled sibling and then overtake in development terms and in the privileges earned by greater maturity. Parents describe this period of catching up and gradual overtaking as one of the most difficult in bringing up their children because of rivalry or jealousy. However, subsequent interviews with parents show that they are as sensitive to the needs of their 'normal' children as to those of the disabled child, and the balance between the siblings is readjusted. There is little evidence of long-term damage, but on the contrary, a consistent finding that the brothers and sisters are drawn to the caring professions, particularly medicine, nursing, or special needs teaching. The majority of families with children with intellectual disability are ordinary families with 'one feature in common'.[7]

Mental illness

In the early months following the birth of a child recognized as having a major developmental disorder, such as Down syndrome, there was clearly much distress and disappointment, but little evidence that the mothers had a higher incidence of postpartum psychiatric disorder. Later in the childhood of the affected child, particularly in families with many other problems, depression was more common in the mothers of children with Down syndrome than in mothers of normal children. But when the mothers of children with a variety of disorders all producing intellectual deficit were compared with mothers of children with Down syndrome, there were less reported health problems in the Down group. Children with brain damage and severe hyperkinesis and those with autism were rated as the most stressful. Hyperkinesis and autism both occur in Down syndrome and their families report a similar degree of stress, as recorded by the questionnaire. For all families with a disabled child, many of the same factors appeared to be protective, for instance a good relationship with a partner and, in addition, support and affection from female relatives like the woman's mother or sisters. There were professionally led groups, assigned social workers, and parent–teacher associations at the school, but the informal sources of support were consistently more effective than formal, with studies in the United States showing very similar results to those in the United Kingdom.

The mental heath, composition, social background, and functioning of the family can increase the risk of psychopathology in the child with intellectual disability as can these risk factors for all young people.[8] There is no evidence that severe psychiatric illness is more common among the families of people with intellectual disability than in anyone else. However, the combination of a severe mental illness, such as bipolar disorder coexisting in the same family with intellectual disability is overwhelming for any family, particularly if one person has the dual diagnosis.

People with intellectual disability who become parents themselves

Although sexuality and pregnancy is a fear of many parents of severely intellectually disabled adolescents, their fertility appears to be very low and there are very few pregnancies in people who are totally dependent. The majority of people identified as having mild retardation during the period of education disappear from services when they leave school and so it is not possible to estimate how many women with mild or borderline retardation become mothers. However, a certain number do come before the family courts or are already known to services for other reasons. However, problems arise with planning ahead and the constant protection from danger that young babies require. There are now techniques to help teach these skills. The secret of success in such teaching is a positive attitude of enhancing skills and not one of undermining the mother. With a partner who is both stable and more able, many quite limited young women cope. As the children grow older, the problems increase as the balance between protection and encouraging new skills becomes more difficult. However, intellectual limitation in itself is not an absolute bar to parenthood. Sadly, many young women with difficulties in intellectual and emotional immaturity are likely to find partners with even more problems and have, for example, a high risk of being hurt by a violent man and of failing to protect children from similar abuse. It is problems such as these rather than the intellectual deficit that make the courts question the safety of the children.

International perspective

The very many worldwide studies that exist show a remarkable consistency in their findings. For all families of whatever ethnic origin, economic status, or religious persuasion, there is grief at the birth of a child, who is in any way defective, and anxiety and sadness about a child who later is seen to fail. In some cultures, an affected boy is harder to bear than an affected girl, as boys have special roles, for instance taking part in the funeral of the parents. Obviously, a high infant mortality will mean an even higher rate in children with any sort of disability. There are a few studies that have come from countries in an early state of development. The authors of these papers are anxious not to repeat what they understandably see as the mistakes of Europe and America. There are for example excellent community services in Asian countries based on the strengths and the beliefs of the local people,[9] whereas others model their services on those in the West, and thus have similar problems but cannot reach a significant proportion of the population. Other countries in Europe are struggling to establish new services for children at the same time as they deal with the very many older people who have been poorly treated for many years. The changes are difficult for the families of these older ones, yet even after many years, families have cooperated with rehabilitation and, in some cases, taken the adult 'child' home.

Needs and priorities

Today in the United Kingdom the social and economic needs of these families are often unmet with a significant number of families with a disabled child living in poverty.[10] Most families will agree that the needs of the disabled member should be given priority—provided that other members of the family are in good

agreement. The needs concerning the best possible communication and training or appropriate treatment for behaviour problems are very much in the interests of other members of the family. Hence the needs of the others would include:

- education that supports the development of the child and the needs of the family
- as accurate a diagnosis as possible
- genetic advice to other members of the family likely to produce children
- to be treated as informed partners by therapists and teachers
- available and interested primary health care
- a key worker to coordinate access to the different agencies and further financial support
- informed specialist care within reasonable distance
- advice and help to 'get through' to the child should communication be a problem, with the chance to learn sign language, symbols, or computer aids
- specialist and domiciliary help with behaviour problems
- respite care, arranged in partnership with the family
- when things go wrong, support and if necessary psychiatric care that treats the other members of the family as whole people with many other facets and not stereotyped family members of a disabled child.

Conclusion

Having a child with intellectual disability is a major and usually totally unexpected blow to any family. However, most families show great resourcefulness and adapt to give their normal child as well as themselves a happy, rewarding life. Parents strongly resent being treated as potential psychiatric patients and have vigorously thrown out the concept of 'the handicapped family'. They do suffer understandable grief. From the point of discharge, the encouragement of informal support is more useful than providing hospital-based services. Children with all sorts of disability go to school early and the provision of unobtrusive familiar services is helpful. Unfortunately, there is often a gap in services between children's services and those for older adolescents and adults. The gap occurs at the worst time for parents who of all times require a familiar knowledgeable person who can offer a service throughout the transition period. The services required by the parents are practical help, such as appropriate equipment, respite care, advice about behaviour, and the ability to find emergency or specialized help at short notice. Parents also require some notice to be taken of their increasing age and/or infirmity, the financial difficulties arising out of the disability, and their anxiety that a humane plan can be made for their son or daughter when they die.

Further information

www.downs-syndrome.org.uk
www.fragilex.org.uk

References

1. Tizard, J. and Grad, J.C. (1961). *The mentally handicapped and their families*. Maudsley Monograph 7. Maudsley Hospital, London.
2. Gath, A. (1997). A review of psychiatric and family research in mental retardation. *International Review of Research in Mental Retardation*, **20**, 137–55.
3. Hastings, R., Turnbull, A., and Tonge, B. (2006). Special issue on family research. *Journal of Intellectual Disability Research*, **50**, 861–988.
4. Richards, M. (1998). Annotation: genetic research, family life and clinical practice. *Journal of Child Psychology and Psychiatry*, **39**, 291–305.
5. Seltzer, M.M. and Krauss, M.W. (1989). Aging parents with adult mentally retarded children. Family risk factors and sources of support. *American Journal of Mental Retardation*, **94**, 303–12.
6. Blacher, J., Cameron, L.N., and Paczkowski, E. (2005). Families and intellectual disability. *Current Opinion in Psychiatry*, **18**, 507–13.
7. Byrne, E.A., Cunningham, C.C., and Sloper, P. (1988). *Families and their children with Down's syndrome: one feature in common*. Routledge, London.
8. Emerson, E. (2003). Prevalence of psychiatric disorders in children and adolescents with and without intellectual disability. *Journal of Intellectual Disability Research*, **47**, 51–8.
9. Miles, M. (1997). Afghan children and mental retardation: information, advocacy and prospects. *Disability and Rehabilitation*, **19**, 496–500.
10. Sloper, T. and Beresford, B. (2006). Families with disabled children. *British Medical Journal*, **333**, 928–9.

10.9

The planning and provision of psychiatric services for adults with intellectual disability

Nick Bouras and Geraldine Holt

Introduction

The functioning of people with intellectual disability (ID) is affected by many factors. As well as their ID, their ability to communicate with others, their social competency, personality, life experiences and circumstances, and their health (including mental health) also influence their behaviour and adjustment.

This chapter focuses on the development and provision of services for adults with ID who have additional psychiatric and behavioural disorders. Developments have taken place in various parts of the world in recent years and a wide range of services has emerged.

History and concepts

In the mid-nineteenth century the conceptualizations of the needs of people with ID and of those with mental illness, and of how to meet these needs were separated. Intellectual disability was not then included in psychiatric training curricula and generations of psychiatrists did not see people with ID, apart from those involved with administrative functions or the prescription of psychotropic medications in institutions. The mental health needs of those with ID at this time were largely unrecognized and so ignored.

Ideologies, sociological theories, civil rights issues, and the normalization philosophy[1,2] together with families' organizations inspired current care practices and directed the way ID services developed.

Policy initiatives originating in the United States during the 1960s and 1970s produced profound and far-reaching changes offering the integration of people with ID into mainstream community life. Similar policies were adopted gradually around the world, particularly in North America, Europe, and Australasia, and in several countries the number of people with ID remaining in institutions has been drastically decreased. Deinstitutionalization of people with ID has been probably the largest social policy experiment of our time. Vivid accounts have been published recently offering enlightening narratives from individuals who were resettled in community living.[3] Overall people with ID and their families have benefited, having a better quality of life. Nevertheless, there are significant variations in the quality of community-based services and of the experiences of people who use them.[4]

Psychiatric disorders and ID (dual diagnosis)

Many service planners and providers assumed that psychiatric disorders in this population would substantially diminish when community care programmes had been put in place. With the implementation, however, of the deinstitutionalization process the need for services for people with ID and psychiatric disorders emerged as a major issue.

This is because a significant number of people with ID, 5 to 12 per cent of children[5] and 15.7 to 40.9 per cent of adults with ID[6] have psychiatric disorders and despite progress in care delivery systems, require appropriate input to manage their mental health needs, sometimes over considerable time. Behavioural or psychiatric disorders can impair people's quality of life, cause regression of adaptive and intellectual functioning, and create unnecessary escalation of family stresses.

The presence of severe behaviour or psychiatric disorders in people with ID is one of the main reasons for the breakdown of community placements and of retention in residential environments that are more restrictive than otherwise required. Such people are at risk of being placed in out of area facilities[7] if local resources are not adequate to meet their assessment and treatment needs or ongoing support needs. These placements are often expensive and divert resources from developing local initiatives. The care provided may be inadequate and difficult to monitor. People may lose contact with families, friends, and those people and structures that previously supported them.

It has become clear that people with ID and mental health problems need services from both the ID network and the mental health system. The overall position of governmental policy has been that people with ID should have access to generic (i.e. for anyone with or without ID) health services, but with additional specialist (specifically for people with ID) support when needed.[8,9]

The argument for the provision of mental health care for people with ID from generic services appears sound and is supported widely.[10] Some argue that specialized services lead to stigmatization,

labelling, and negative professional attitudes. Others argue that special expertise is required for the diagnosis and treatment of psychiatric disorders in this population, because although it is theoretically possible to train staff in generic settings, the relatively small number of cases gives little opportunity for staff to gain or maintain the necessary skills.[11]

Problems arise particularly when admissions to adult acute inpatient units occur, as people with ID often require longer admissions, and may be vulnerable without additional support on the ward. Furthermore, people with ID represent a very heterogeneous group with a varied range of highly complex mental health needs which generic staff may feel ill equipped to meet.[10]

Menolascino[12] recommended that services be provided according to need and be delivered in the context of both ID and psychiatric disorders coexisting allowing for more appropriate treatment, support, service planning, and development. The result is to create a partnership between the mental health and ID service structures to ensure responsive supports and treatments to previously underserved individuals.

Models of services for people with psychiatric disorders and ID

There has been a growing interest internationally as to how to address this issue. Davidson and O'Hara[13] offer a comprehensive review of service developments for this population. Long-term resolution of behavioural or psychiatric disorders in persons with ID requires community-based activities. Hence since the year of publication of the first edition of the *New Oxford Textbook of Psychiatry* new developments in most countries of the world are community-based. The pace and form of change depends on each country's unique historical perspective and national philosophies about care for people with ID.[14] However, resolution of an acute crisis may require, in addition to community-based psychiatric or behavioural resources, inpatient acute psychiatric assessment and treatment services, specialized outreach, emergency respite, or emergency behaviour stabilization services.

The most common models of services for adults with ID and psychiatric disorders that have emerged in recent years in the United Kingdom can be described as: (a) generic ID community-based multidisciplinary (interdisciplinary) teams, (b) specialist community-based mental service for people with ID.

Generic ID community-based multidisciplinary (interdisciplinary) teams

A multidisciplinary (interdisciplinary) team offers assessment and specialist services to people with ID. Initially, most of these teams were involved with deinstitutionalization, carrying out tasks such as identifying appropriately adapted and staffed houses, matching clients to live together, assessing health and social needs, and so on. Most of them have input from clinical psychologists and usually some input from a psychiatrist specializing in people with ID. Some teams have developed innovative ways of working with people with challenging behaviour often with severe ID. Members specializing in functional analysis and/or behavioural treatments strengthen such teams.

One considerable problem with this model has been the lack of links with mainstream mental health services. Despite the psychiatric input, such services may experience difficulties in meeting the mental health needs of people particularly those with mild ID and mental illness. The problems are extended to people with ID who may have additional forensic mental health problems, autistic spectrum disorders including Asperger's syndrome and co-morbid conditions as well as those with borderline intellectual functioning.

Specialist mental health service for people with ID

Since 1982, the Community Mental Health in ID Service in South East London[10,15] has operated using this model. It has secondary and tertiary care functions. This Service includes outpatient clinics, outreach work, inpatient assessment and treatment, and consultation with community agencies. The clinical team comprises of psychiatrists, community psychiatric nurses, and administrative staff, and has a regular interface with clinical psychologists and behaviour support specialists. The clinical team also receives regular input from occupational therapists, speech therapists, and social workers. The composition and functions of the Service have evolved over a number of years. An integrated part of the Service is the provision of training to direct support care staff and others to promote and sustain the development of a competent workforce at every level, from direct care staff to managers and organizations.

There are three phases in providing clinical services: assessment, intervention, and follow-up.

The clinical team carries out a structured clinical assessment on all referrals with the additional application of standardized instruments, e.g. Aberrant Behaviour Checklist[16] and CANDID.[17]

Therapeutic interventions are based on multidisciplinary work and include medication and environmental manipulation, as well as psychological treatments such as anxiety management and cognitive behaviour therapy. Regular weekly clinical team meetings are held to review progress. Crisis prevention plans are developed to help families and service providers identify early signs of breakdown and to take appropriate action. Training is offered to improve the capacity of families and service providers, to better understand and respond to the mental health needs of people with ID. This includes seminars, books and videos as well as modelling and role-playing exercises. Ongoing support and consultation is also provided while other specific therapeutic interventions are implemented.

Follow-up is provided for as long as it is required. Once a client seems stable and the agreed upon strategy appears to be effective the team maintains quarterly or half yearly contacts.

If an inpatient stay is warranted for acute psychiatric crises, admission is into generic mental health facilities with consultative advice and support from the community-based team. Patients can also access a six-bed specialist unit at a tertiary level. The function of this unit is to provide comprehensive assessment of the mental health problems when this cannot be achieved in a community setting or within generic mental health services, to make recommendations and implement therapeutic interventions and to ensure the appropriate care plans are transferred to the community setting on discharge. Care is delivered and coordinated via a person centred, Care Programme Approach (CPA),[18,19] to help ensure effective links with the full range of psychiatric health and social care services.

This Service is compatible with the development of other specialist services in the United Kingdom over the last few years to address specific needs for example of children and adolescents, older adults, those with forensic problems, mothers and babies, those with

eating disorders, home treatment teams, assertive community treatment services, eating disorders teams, early intervention teams for psychosis, etc.[20]

Outcomes

Evaluation and measuring of outcomes in mental health care for people with ID is very complex. This is because most health care service developments and reforms are politically and socially driven rather than evidence led and researchers cannot embargo change until they have defined systems. Accumulating evidence from chronological studies will still require judgement and interpretation.[20]

Moss *et al.*[21] considered a variation of the Matrix Model, first described for non-disabled people with mental health problems by Thornicroft and Tansella,[22] for the evaluation of mental health services for people with ID. This consists of two dimensions, one determined by the level within the service system (i.e. national, local, or individual), and the other by the point in the temporal sequence of service provision (i.e. inputs to the service, the process of providing the service, and the resulting outcome). Bouras *et al.*[15] adopted the Matrix Model partially (inputs and processes) to evaluate their model of service and found that over 18 years statistically significant changes in referrals trends in ethnicity, type of residence, level of ID, the number of admissions to inpatient units and psychiatric diagnoses. In addition they also found that patients admitted to the specialist unit—in contrast to those admitted to a generic inpatient unit—showed a significant decrease in psychiatric symptoms, an increase in overall level of functioning, a reduction in severity of their mental health problems, and an improvement in behavioural function on discharge, at 6 and 12 months following discharge.[23]

In an attempt to compare the effectiveness of assertive and standard community treatment in people with psychotic spectrum disorders and ID, with a randomized controlled study, no significant differences were found between the two treatments.[24]

Clinical effectiveness studies in mental health care for people with ID still have to overcome important methodological limitations. At present the Matrix Model[21] seems to offer the most advantageous way of evaluation, providing a framework to conceptualize the factors that influence service developments in the field.

Residential programmes for people with psychiatric and ID

Successful community-living opportunities for people with ID require a comprehensive and collaborative service structure, including appropriate residential and vocational facilities. However, whilst these have been developed for many people, services to meet the needs of those with psychiatric disorders have lagged behind.

Housing for people with mental health problems must be compatible with all the main principles of 'ordinary housing'. It should be located in an acceptable community setting that offers opportunities for community integration, be designed to provide services and supports to meet the needs and desires of the person residing there, and be affordable, safe, and comfortable. This requires that staff have the necessary skills and service structures to meet client needs.

As institutions have been closed residential facilities have been developed in their stead. The trend across North America, Europe, and Australasia has been for larger residential homes (sometimes on the sites of the old institutions) to be replaced by smaller group homes for 3–8 people supported by staff. More recently, 'supported living schemes' have become more common, where people rent or own the property, and receive support from agencies that do not control the accommodation.[4] The pace of change varies between and within countries.

The aim is to empower individuals in smaller settings, organized to respond to a wide range of needs, creating environments that promote physical and mental health. However, no one model will necessarily meet the needs of all individuals with mental health problems and ID. Some people may become isolated and lonely in one or two-person settings, or have difficulties that cannot be managed in housing where additional staff or clinical support is not readily available. Some people may simply prefer to live in a supervised group living situation rather than supported living and should be given the opportunity to live in a place they prefer.

Residential services should include a full range of alternatives to enhance an individual's capacity for community living. The individual receiving residential services should be allowed to have as much comfort, ownership, and autonomy as possible. Housing can offer a wide range of options, and maximize opportunities for community integration and personal independence. Specialist mental health services for people with ID should work in collaboration with residential providers, to provide clinical support and a safety net when difficulties arise. Delivery of services in this manner represents one of the most important organizational challenges for services for people with mental health problems and ID.

Vocational programmes

Vocational services should also offer work in integrated settings in a person's community, opportunities, and supports that are manageable and productive for the worker and the workplace with adequate salary compensation.

There have been significant changes in employment and vocational services for people with disabilities and several have moved from traditional workshop settings to integrated supported employment. The majority of placements have been in the service sector consistent with shifts towards entry and low-skill jobs in the national employment market. Individual placement has had the greatest positive effect on wages. Supported employment enhances the quality of life of people with ID. Although there is an acceptance in society that people with mild levels of disabilities can be meaningfully employed, traditional views of the capabilities of people with severe disabilities continue to be major obstacles to their access to the most progressive contemporary, educational, and rehabilitation practices. People with mental health problems and ID may be under-represented in both the sheltered and supported employment workforces.

Staffing issues and training

The availability of specialist training varies markedly between countries, and not surprisingly bears a close relationship to the level of service development.

In the United Kingdom the need for specialist mental health services for people with ID and psychiatric disorders was recognized in the early 1970s. Specialist training programmes for psychiatrists, nurses, and other health care professionals including family doctors,

community nurses, and direct care staff have been developed. However, whilst such training is available its uptake is dependent on the interests of individuals or of their employees. Only for some is such training mandatory, e.g. psychiatrists.[25] Attention has focused in recent years on the training needs of first-level care workers in community day and residential facilities. They often receive little or no training in the psychiatric aspects of ID with the consequence that psychiatric illness amongst their clients frequently goes unrecognized and untreated.

Benefits of training

Staff finds working with people with ID and mental health problems stressful. Giving them skills in this area so that they can manage, with support, people with mental health problems enables them to find this work more rewarding. The most basic and vital role of support staff in this context is the awareness that a person with ID may suffer a mental illness, as we all may. They need to be aware of the range of therapeutic options that might be helpful, including environmental changes, behavioural strategies, psychotherapeutic techniques, medication, and so on. A fuller knowledge and consideration of this topic will help to dispel myths and prejudices, for example that medication is to be avoided at all costs, or that its use signifies that staff has in some way failed the client. Specific knowledge about some disorders will provide insights into why and how interventions must be tailored around someone's strengths and needs, for example someone with autistic spectrum disorder may hit himself when his routine is changed. The intervention chosen may be to provide a timetable, which the staff and client follow. This may need to be in pictorial form to meet the client's communication needs, and small and durable enough for him to carry at all times.

Training materials

Flexible training materials (e.g. The Training Package in the Mental Health of Learning Disabilities),[26] which can be used by staff groups in their own settings, are now available. It is often useful to design training around particular clients. Training should be a part of the culture of an organization. Including managers in training activities is helpful. It allows them to share a knowledge base with their staff, and to set-up processes, which facilitate the continued development of issues identified by the training. For instance, each client's mental health might be considered in his or her individual planning meeting. Actions agreed can then be regularly discussed in individual staff supervision, at staff groups, and at meetings with the mental health and multidisciplinary teams. Raising awareness on mental health issues for people with ID is also important for carers and families as they are a pivotal source of support.[27]

Commissioning services

The deinstitutionalization of people with ID in the United States, United Kingdom, and other parts of the world is well-advanced. A variety of service models are provided. Comprehensive local services systems for those with additional mental health needs are emerging. The old institutions represented a complete system of care, inasmuch as they provided accommodation, health care, social care, and occupation in a single setting. Current provision, by contrast, involves a range of agencies and settings. This requires that care is integrated and organized around an individual. This is not an easy task for those with complex needs.

Those commissioning services need to determine what services are needed locally and decide how they should be provided, monitored, and reviewed. This chapter provides an overview of the social and policy context and some models of services for adults with psychiatric disorders and ID. Local demographics and resources will of necessity shape services. To ensure that the commissioning of services is well-informed all planning partners should be involved. Using the Matrix Model[21] described earlier the commissioners might consider:

1 Joint commissioning. Various policies and legislation (**national inputs**) have proposed joint commissioning by health and social care services so that a joint strategy drives the joint commitment of resources (**local inputs**).
2 Client and carer participation (**individual inputs**). This is essential to ensure that appropriate priorities are set and that services are satisfactorily delivered (**individual and local processes and outcomes**).
3 Involvement of statutory and voluntary agencies (**local inputs**). This will enable the commissioners to make informed decisions.
4 A baseline needs assessment of the population to be served (**local inputs**) e.g. from census statistics, epidemiological research, local register data.
5 Local and national policies (**local and national input**) to enable them develop a vision, e.g. community-based services, mixed economy of provision.
6 Desired **outcomes at local level** (e.g. increased use of generic provision, reduction in number of people placed out of area) and translate these into
7 Service specifications
8 Purchase services, which have the necessary skills (**local and possibly national inputs,** e.g. for people with very particular needs) to deliver processes (**local and possibly national processes**) that will provide these outcomes.
9 Set in place monitoring systems, which may include individual and local outcomes e.g., complaints and incidents monitoring, scrutiny of statistics derived from CPA documentation (**individual and local processes and outcomes**).
10 Commissioning is a cyclical process and the monitoring and review of services (**local inputs**) will ideally enable more effective future commissioning.

Components of an effective psychiatric service for people with ID should include:

- Organizing services around clients' wishes and needs.
- Good interagency communication among health, social, and voluntary sectors.
- Good interagency communication between services for children and adults.
- High level of awareness of mental health issues by direct support staff in residential and day care services.
- High level of awareness of mental health issues by primary care staff.

- Multidisciplinary composition including psychiatrists, mental health nurses, clinical psychologists, behaviour support specialists, therapists, and social workers.
- Ability to provide consultation, assessment, and treatment.
- Provision of community-based interventions.
- Access to local specialist and generic community and inpatient assessment, treatment, forensic, and rehabilitation facilities.
- Resources to meet residential, recreational, and vocational needs of those with enduring needs.
- Clear coordination of inputs by a named person.
- Staff training.
- Measuring outcomes.

Conclusion

There has been considerable debate as to whether specialized mental health service for people with ID services should be established or generic mental health service providers should serve this population. Whatever strategy is undertaken it should be based on high professional standards. Standardized diagnostic and assessment tools should be used. Appropriate, individually tailored treatments should be given in the least restrictive environments. Staff must have the necessary expertise, training, and support.

Mixed results have been obtained for the evaluation of existing services.[28] No direct studies of comparative treatment effectiveness exist, and studies on single specialist services contain some methodological weaknesses. It is essential that the quality of services is monitored, to maintain and improve standards of care. Increasing fragmentation of provision makes this a complex but essential task.

Often mental health services for people with ID are provided in a crisis. This highlights and amplifies the existing deficiencies in care. In order to become more effective and accessible these services will have to address both individual needs and service systems complexities for children, adolescents, and adults with ID and mental health problems. Several years into the post-institutional period and the era of community care, meeting the mental health needs of people with ID remains a challenge. Though the initial philosophical concern about the coexistence of mental health problems in people with ID has been partially eroded with the emerged evidence-based research, nevertheless the clinically effective responses remain scanty. In times of hard fiscal constraints for mental health services in general, it is hoped that people with ID will not be further overlooked and marginalized but they will receive the required professional attention matched by adequate distribution of resources.

Further information

Bouras, N. and Holt, G. (eds.) (2007). *Psychiatric and behavioural disorders in intellectual and developmental disabilities*. Cambridge University Press, Cambridge.

Advances in Mental Health and Learning Disabilities published by Pavilion Publishing. (2007). http// www.pavpub.com

Estia Centre: http://www.estiacentre.org

National Association for the Dually Diagnosed: http://www.thenadd.org

References

1. Wolfensberger, W. (1969). The original nature of our institutional models. In *Changing patterns in residential services for the mentally retarded* (eds. R. Kugel and W. Wolfensberger), p. 235. President's Committee on Mental Retardation, Washington, DC.
2. Wolfensberger, W. (1991). Reflections on a lifetime in human services and mental retardation. *Mental Retardation*, **29**, 1–16.
3. Johnson, K. and Traustadottir, R. (2005). *Deinstitutionalization and people with intellectual disabilities: in and out of institutions*. Jessica Kingsley Publishers, London.
4. Mansell, J. (2006). Deinstitutionalisation and community living: progress, problems and priorities. *Journal of Intellectual & Developmental Disability*, **31**, 65–76.
5. Kerker, B.D., Owens, P.L., Zigler, E., *et al.* (2004). Mental health disorders among individuals with mental retardation: challenges to accurate prevalence estimates. *Public Health Reports*, **119**, 409–17.
6. Cooper, S.-A., Smiley, E., Morrison, J., *et al.* (2007). Mental ill health in adults with intellectual disabilities: prevalence and associated factors. *The British Journal of Psychiatry*, **190**, 27–35.
7. Pritchard, A. and Roy, A. (2006). Reversing the export of people with learning disabilities and complex health needs. *British Journal of Learning Disabilities*, **34**, 88–93.
8. Department of Health. (2001). *Valuing people: a new strategy for learning disabilities in the 21st century*. HMSO, London.
9. US Public Health Services. (2002). *Closing the gap: a national blueprint for improving the health of individuals with mental retardation*. Report of the surgeon general's conference on health disparities and mental retardation. US Department of Health and Human Services, Washington, DC.
10. Bouras, N. and Holt, G. (2004). Mental health services for adults with learning disabilities. *The British Journal of Psychiatry*, **184**, 291–2.
11. Day, K. (1999). Professional training in the psychiatry of mental retardation in the United Kingdom. In *Psychiatric and behavioural disorders in developmental disabilities and mental retardation* (ed. N. Bouras), pp. 439–57. Cambridge University Press, Cambridge.
12. Menolascino, F.J. (1977). *Challenges in mental retardation: progressive ideology and services*. Human Sciences Press, New York.
13. Davidson, P.W. and O'Hara, J. (2007). Clinical services for people with intellectual disabilities and psychiatric or severe behaviour disorders. In *Psychiatric and behavioural disorders in intellectual and developmental disabilities* (eds. N. Bouras and G. Holt), pp. 364–87. Cambridge University Press, Cambridge.
14. Holt, G., Costello, H., Bouras, N., *et al.* (2000). BIOMED-MEROPE project: service provision for adults with intellectual disability: a European comparison. *Journal of Intellectual Disability Research*, **44**, 685–96.
15. Bouras, N., Cowley, A., Holt, G., *et al.* (2003). Referral trends of people with intellectual disabilities and psychiatric disorders. *Journal of Intellectual Disability Research*, **47**, 439–46.
16. Aman, M.G. and Singh, N.N. (1986). *The aberrant behaviour checklist.* Slosson Educational Publications, New York.
17. Xenitidis, K., Slade, M., Thornicroft, G., *et al.* (2003). *Camberwell assessment of needs for adults with developmental and intellectual disabilities*. Gaskell, London.
18. Department of Health. (1999). *National service framework for mental health: modern standards and service models for mental health*. HSC1999/223HMSO, London.
19. Department of Health. (1999). *Effective care co-ordination in mental health services: modernizing the CPA*. HMSO, London.
20. Burns, T. (2001). Generic versus specialist mental health teams. In *Textbook of community psychiatry* (eds. G. Thornicroft and G. Szmukler), pp. 231–41. Oxford University Press, Oxford.

21. Moss, S., Bouras, N., and Holt, G. (2000). Mental health services for people with intellectual disability: a conceptual framework. *Journal of Intellectual Disability Research*, **44**, 97–107.
22. Thornicroft, G. and Tansella, M. (1999). *The mental health matrix: a manual to improve services*. Cambridge University Press, Cambridge.
23. Xenitidis, K., Gratsa, A., Bouras, N., *et al*. (2004). Psychiatric inpatient care for adults with intellectual disabilities: generic or specialist units. *Journal of Intellectual Disability Research*, **48**, 11–18.
24. Martin, G., Costello, H., Leese, M., *et al*. (2005). An exploratory study of assertive community treatment for people with intellectual disability and psychiatric disorders: conceptual, clinical, and service issues. *Journal of Intellectual Disability Research*, **49**, 516–24.
25. Costello, H., Holt, G., Cain, N., *et al*. (2007). Professional training for those working with people with intellectual disabilities and mental health problems. In *Psychiatric and behavioural disorders in intellectual and developmental disabilities* (eds. N. Bouras and G. Holt), pp. 364–87. Cambridge University Press, Cambridge.
26. Holt, G., Hardy, S., and Bouras, N. (2005). *Mental health in learning disabilities: a training resource*. Pavilion Publishing, Brighton.
27. Holt, G., Gratsa, A., Bouras, N., *et al.* (2004). *Guide to mental health for families and carers of people with intellectual disabilities*. Jessica Kingsley Publishers, London.
28. Chaplin, R. (2004). General psychiatric services for adults with intellectual disability and mental illness. *Journal of Intellectual Disability Research*, **48**, 1–10.

SECTION 11

Forensic Psychiatry

This page intentionally left blank

11.1

General principles of law relating to people with mental disorder

Michael Gunn and Kay Wheat

Introduction

This chapter provides a scheme for assisting in the analysis of two areas of law that provide some of the general principles that operate in relation to mentally disordered offenders. These two areas are (a) the law concerning decision-making and other action-taking to which the concept of competence is crucial, and (b) the law of responsibility in relation to liability for criminal offences and the tort of negligence. Whilst the focus of the chapter is on the law of England and Wales, it is clear that there are similarities in other common-law jurisdictions, and in other jurisdictions that have borrowed ideas from common-law jurisdiction, such as Japan, in relation to the concept of informed consent.

Decision-making and action-taking law and competence

Generally, the law in relation to decision-making and action-taking might take one of three approaches to mentally abnormal offenders.

- The law might adopt the same approach for mentally abnormal offenders as for anyone else.
- The law might adopt an approach dependent upon the competence of the individual that might be affected by the mental state of the mentally abnormal offender.
- The law might adopt an approach recognizing the impact of being a mentally abnormal offender that may be based upon the effects or mere status of the mental state.

  There is no reason to examine further the law that is not different for the mentally disordered.

Autonomy and Competence

The most appropriate approach that introduces different law is by reliance upon competence or capacity. Internationally, there is increasing acceptance that, where someone is incompetent to make their own decisions, there must be a route to making such decisions on their behalf. For example, in Japan, the approach of informed consent has been adopted and whilst there is not yet a fully developed concept of competence, it is accepted that that is the next necessary development. Where there is significant variation is in the approach to adopt if someone is not competent. Traditionally, the approach has been to adopt guardianship whereby someone is either under guardianship or not and if so that all decisions are taken by the guardian. More recently, a more varied approach has become preferred whereby the decisions taken by others are only those that the individual cannot take and the basis for taking those decisions is the decision that the incapacitous would have made if competent and otherwise the decision that is in their best interests. It must be accepted however that this may be viewed as an approach grounded in a particular approach to law and ethics, i.e. that grounded in Western societies. Even there, there is a tradition for making decisions on a paternalistic basis, that is largely now discredited, though there are also concerns about the focus on autonomy. However, in other societies much greater emphasis is placed on the importance of the family as decision-makers or the basis of a decision on the presumption that the individual is a part of a particular society that has a genuine and proper interests in decision to be made on their behalf. Having recognized that as an approach, this chapter will largely focus upon the Western legal systems' basis, using England and Wales as an illustrative jurisdiction.

Increasingly, there is recognition internationally that action should only be taken with regard to a person if either they are incapable of deciding or acting for themselves or if they present a harm to others (or self) and that harm is linked with a mental health problem. The extent to which different jurisdictions have a developed view of respect for the principle of autonomy and its legal application through a test for capacity or competency unsurprisingly varies in practice. But the major issue is how decisions are to be made if someone is incapable.

International statements of principle

Acceptance of respect for the principle of autonomy can be seen through at least two international instruments. Nascently, it can be identified in the provisions of the *United Nations* Declaration on the Rights of Mentally Retarded Persons (1971) which includes the following commitments.

1 The mentally retarded person has, to the maximum degree of feasibility, the same rights as other human beings.

2 The mentally retarded person has a right to proper medical care and physical therapy and to such education, training, rehabilitation, and guidance as will enable him to develop his ability and maximum potential. . . .

5 The mentally retarded person has a right to a qualified guardian when this is required to protect his personal well-being and interests.

6 The mentally retarded person has a right to protection from exploitation, abuse, and degrading treatment. . . .

7 Whenever mentally retarded persons are unable, because of the severity of their handicap, to exercise all their rights in a meaningful way or it should become necessary to restrict or deny some or all of these rights, the procedure used for that restriction or denial of rights must contain proper legal safeguards against every form of abuse. This procedure must be based on an evaluation of the social capability of the mentally retarded person by qualified experts and must be subject to periodic review and to the right of appeal to higher authorities.

More recently, the *Council of Europe* has agreed a set of recommendations that should be implemented across Europe and are attracting significant international attention, e.g. by the South African Law Commission. The *Principles Concerning the Legal Protection of Incapable Adults* (Council of Europe, 1999). As Jansen demonstrates, the Recommendation confirms the functional approach to capacity and seeks to provide the incapable adult, where necessary, with representation, assistance, measures of protection, and other arrangements. The Recommendation opens with a statement that underpins the general approach adopted in this chapter:

> 'I.1. The following principles apply to the protection of adults who, by reason of an impairment or insufficiency of their personal faculties, are incapable of making, in an autonomous way, decisions concerning any or all of their personal or economic affairs, or understanding, expressing or acting upon such decisions, and who consequently cannot protect their interests.'

This is followed up by an important statement that captures an underlying theme for most jurisdictions endeavouring to provide suitable approaches.

> 'II.1 In relation to the protection of incapable adults the fundamental principle, underlying all the other principles, is respect for the dignity of each person as a human being. The laws, procedures, and practices relating to the protection of incapable adults shall be based on respect for their human rights and fundamental freedoms, taking into account any qualifications on those rights contained in the relevant international legal instruments.'

The principles adopted are then:

- securing the maximum preservation of capacity that demands a functional approach to capacity and so not accepting that someone is either capable or not capable for all decisions, but may be able to make some decisions and not others (II.3.1) and that no step should be taken unless it is necessary (II.5.1).
- where steps are taken they must be proportional to the degree of capacity retained and they should be tailored to the needs and circumstances of the incapacitous person (II.6.1).
- there just be fair and efficient procedures for the taking of steps which must protect human rights and prevent possible abuses (II.7. 1 & 2). These requirements are expanded upon in Part III.
- the interests and welfare of the incapacitous person are the paramount consideration, thus ruling out a paternalistic basis for the taking of steps (II.8.1).
- the past and present wishes and feelings of the incapacitous person should be ascertained as far as possible, and should be taken into account and given due respect, and of most importance are the choices made by the incapacitous person themselves (II.9. 1 & 2).
- there is a preference for action taken without the intervention of a judicial or administrative authority but that such powers must be limited and their exercise controlled (IV.18.1)
- in the health field, no action should be taken if someone is capable of making the decision (V.22.1). The intervention may then be carried out if it is for the incapacitous person's direct benefit and authorization has been given by their representative or by an authority or person or body provided for by law (V.22.2). As not all jurisdictions are ready for this approach even in Europe, an alternative is provided so that where a person is under protective steps, the incapacitous person's consent should be sought even though there is someone with the power to make the decision (V.23.1). Where the incapacitous person cannot provide consent, the intervention is permissible where it is for their direct benefit and authorization has been given by their representative or by an authority or person or body provided for by law (V.23.2).

For Jansen, the key principles are, first, those in Principle 5, that is '*Necessity* and *Subsidiarity*' as they 'imply, first of all, that no measure of protection should be established unless it is necessary, taking into account the circumstances of the particular case. Secondly, in deciding whether a measure is necessary, account should be taken of any less formal arrangements which might be provided in particular by family members, or by public authorities or other means. The latter is the principle known as 'subsidiarity' . . ..' The second key principle is, that in Principle 3, that is 'that of maximum preservation of capacity . . .. In particular a measure of protection should therefore not result in an automatic complete removal of legal capacity.' The third key principle is that in Principle 6, that is 'Proportionality: where a measure of protection is necessary it should be proportional to the degree of capacity of the person concerned and tailored to the individual circumstances of the case. The measure should restrict the legal capacity, rights and freedoms of the adult by the minimum which is consistent with achieving the purpose of the intervention.'

The international picture is completed by the Convention on the Rights of Persons with Disabilities that was signed in 2006 but is not yet in force. This is a Convention of the United Nations and has 129 States as signatories to it. The Convention is a broad Convention and covers many areas not directly relevant to this Chapter. It takes a similar approach since, its General Principles (art. 3) are (a) respect for inherent dignity, individual autonomy including the freedom to make one's own choices, and independence of persons; (b) non-discrimination; (c) full and effective participation and inclusion in society; (d) respect for difference and acceptance of persons with disabilities as part of human diversity

and humanity; (e) equality of opportunity; (f) accessibility; (g) equality between men and women; and (h) respect for the evolving capacities of children with disabilities and respect for the right of children with disabilities to preserve their identities. In, for example, outlawing discrimination on the basis of disability (art. 4) and providing for freedom from exploitation, violence and abuse (art. 16), the Convention identifies the balance to be drawn between recognising the importance of decision-making with that of protecting those not capable of making their own decisions. It affirms the importance of the capacity to make decisions as a key requirement in the law. Article 12 states that State Parties reaffirm that persons with disabilities have the right to recognition everywhere as persons before the law and that State parties shall recognise that persons with disabilities enjoy legal capacity on an equal basis with others in all aspects of life.

Tests of capacity and competence

Thus, it can be seen that key to respect for the principle of autonomy is to have a workable concept of capacity or competence. The functional approach requires that the test of competence be related to the particular decision to be made, at the particular time that it must be made. There is a range of abilities that competence might involve. Much of the work on competence has been undertaken in the context of health-care law and in relation to consent to treatment. Much of this work has been undertaken in the United States. Two leading thinkers, Grisso and Appelbaum, have, with colleagues, identified four abilities that can be involved in competency:

- evidencing a choice
- understanding
- appreciation
- reasoning or rationality.

Any given jurisdiction will adopt one or more of these abilities[19] in what it looks for in relation to competency assessments. There is no consistency, currently, as to which one or more of the abilities must be satisfied, except to say that almost all jurisdictions require understanding to some degree. This lack of consistency reflects the developing international understanding of the concept of competence. If we take English health-care law as an example, it can be seen that, in the early stages, understanding was the prime ability that had to be established, though the patient also had to evidence a choice. But, more recently, it seems that the courts are being attracted to an approach that may ultimately see competence only being satisfied where all four abilities are satisfied. Requiring rather more of an individual to satisfy the requirement of competence may be regarded as a better means of satisfying the crucial bioethical principle of self-determination or respect for the principle of autonomy, since if someone is not truly able to exercise self-determination, there is no respect for autonomy if, nevertheless, that person's 'decisions' are legally binding. This means that rather more 'decisions' are open to the challenge on the basis that they are not made by someone competent to do so. A stringent approach to competence may be hard to accept. It must then be assessed (as a general matter) whether it would be better to reduce the standard and so enable more people to be assessed as competent or whether lowering the standard is illusory as being for the benefit of people whose competence may be open to question. Therefore it is hardly surprising that there continues to be debate as to the abilities that any individual must possess (and the level of functioning of that ability) in order to determine whether he or she is competent to make a particular decision. Wong *et al.*[19] make the point, drawing on the work of others, that the functional approach is not without problems. They point out that it is time consuming, legal standards vary between jurisdictions, and there is uncertainty about the threshold to be satisfied in determining competence.

The English Mental Capacity Act 2005, sections 2 and 3 creates a definition of capacity consistent with those abilities. The central elements of that definition are to be found in sections 2(1) and 3(1).

> 2(1) For the purposes of this Act, a person lacks capacity in relation to a matter if at the material time he is unable to make a decision for himself in relation to the matter because of an impairment of, or a disturbance in the functioning of, the mind or brain.
>
> 3(1) For the purposes of section 2, a person is unable to make a decision for himself if he is unable –
>
> (a) to understand the information relevant to the decision,
> (b) to retain that information,
> (c) to use or weight that information as part of the process of making the decision, or
> (d) to communicate his decision (whether by talking, using sign language or any other means).

Three approaches

What is key is that the approach is a functional one that is it is related to the abilities of the individual at the time a decision is required and is not dependent upon either status or outcome of the decision, though these clearly have formed part of either definitions or approach to capacity in the past and are relevant factors in identifying the possibility that someone may not be capable of decision-making and in exercising judgment about that capacity. To state that competence if to be interpreted functionally.[19] means that the status of the decision-maker is not determinative of the question of her or his competence. *A status approach* makes assumptions about an individual's decision-making competence on the basis of a particular characteristic, and there is no empirical evidence to support the validity of such an approach.[19,21] The mental state of the decision-maker may be the reason why competence is put into question, but mental state in itself is rarely, if ever, sufficient to determine the matter. Mental state may have relevance to decision-making in that certain states will impact on the ability to understand and process information. Furthermore, the outcome of a decision is also not in itself sufficient to determine the matter. The fact that any given decision is not reasonable does not mean that the decision-maker is not competent to have made that decision. For example, the simple fact that a patient disagrees with the doctor does not mean that the decision is that of an incompetent decision-maker, though lack of congruence with the proposals of a doctor may cause questions to be asked about decision-making competence. *The outcome approach* has been rejected in a number of jurisdictions.[19] It is internationally recognized that anyone can make what might be termed objectively silly decisions without necessarily giving rise to doubts about competence. However, the regularity with which silly decisions are made may raise doubts about competence as also will the inter-relation between mental state and quality of decisions. In the United Kingdom, these points are further reflected in the fact that there is a legal presumption

that a person is competent to make her or his own decision once adult state is reached.

The functional approach requires that the test of competence be related to the particular decision to be made, at the particular time that it must be made.[19] There is a range of abilities that competence might involve. Much of the work on competence has been undertaken in the context of health-care law and in relation to consent to treatment. Much of this work has been undertaken in the United States. Two leading thinkers, Grisso and Appelbaum,[20] have, with colleagues, identified four abilities that can

The functional approach to decision-making is not limited, in its application, to health-care decisions, even though that is where most of the debate has taken place. In principle, it may be applied to any type of decision. For example, the making of wills and the entering into of contracts are obvious examples where a functional approach applies, but it does not follow that the same abilities will be required for these decisions as for treatment decisions. Under the law prior to the Mental Capacity Act 2005, this is demonstrated by an old case which was, nevertheless, the leading case in relation to the making of wills. *Banks* v. *Goodfellow*, requires that a person:

> ought to be capable of making his will with an understanding of the nature of the business in which he is engaged, a recollection of the property he means to dispose of, of the persons who are the objects of his bounty, and the manner in which it is to be distributed between them.

This test demanded not just understanding, but also the appreciation and the reasoning ability noted above. Whilst the level at which the will writer must operate is not that of a lawyer, nevertheless he or she must be aware of the context in which the will is being made and must think through the competing potential demands on his or her estate. Of course, will writers can make silly dispositions, even going so far as to exclude financially dependent relatives. However, it must be recalled that an outcome that is questionable or unreasonable is not the same as the will or decision being made on the basis of an unacceptable reasoning process. An eccentric person might well, for example, not wish to leave anything to his or her relatives. Thus, in any jurisdiction, care must be taken to consider a particular test in deciding which of the four abilities are to be identified, and the answer to that question may demand very careful analysis.

Persons not competent to make decisions or take action

If a person is not competent to make a decision for themselves, there is more variability as to the approach to be taken. Some of this difference is related to the commitment to do the best for a vulnerable person and leads to a desire to act paternalistically, so making the decision that objectively is in the best interests of the individual A more frequent approach is to make the decision that that individual would have made for themselves. This is most recently reflected in the English Mental Capacity Act.

If a person is not competent to make a decision or to take certain action, the law increasingly provides mechanisms whereby these decisions can be made. In England and Wales, until the abolition of the sign manual by the Mental Health Act 1959, the Crown had the power and the process to make decisions in the best interests of a person not competent to decide or to take action. This power is the basis of many substitute decision-making procedures in common-law jurisdictions (including the United States). The specific adaptation is jurisdictionally specific. In some jurisdictions, it has been used to follow from generic decisions as to competence and to create overarching substitute decision-making procedures (e.g. those jurisdictions that adopt a full guardianship of the person approach). In England and Wales, the first approach, after the abolition of the sign manual procedure, was to allow for people to be received into guardianship (a process that was not, interestingly, dependent upon a finding of incompetence), but this proved not to be an acceptable procedure. Current English guardianship is very much only a community mental health power that does not enable anyone else to make decisions on behalf of the person received into guardianship.[33]

In England, the courts had to invent a procedure for making decisions on behalf of someone who was not competent. The House of Lords, the senior English court, provided a mechanism in relation to treatment decisions whereby, if a person was found to be incompetent, treatment could lawfully be provided if it was necessary, that is if it was for the life, health, or welfare of the patient and was in her or his best interests. Despite substantial improvements that made the approach identify what was the one, best approach that took into account the full range of interests (see below), this judicial approach has been replaced by a statutory format. The approach introduced by the Mental Capacity Act 2005, does not take a full guardianship or guardianship of the person approach and is not necessarily triggered by a judicial or administrative authority. Rather, a person may act on behalf of an incapacitous adult (as defined by sections 2 and 3, see above) provided she or he acts in that person's best interests (as defined by section 4, see below), but such actions are subject to significant procedural protections, since some areas must be referred to other procedures (e.g. in relation to research, where sections 30–34 allow intrusive, non-therapeutic treatment consistently with the European Convention on Biomedicine), some decisions cannot be made as they are too personal (so such matters as consenting to marry or to entering a civil partnership and consenting to sex do not fall within the Act, see section 27), some decisions necessitate, where carer's views are not available, the views of an independent mental capacity advocate to be taken into account (sections 35–41), some decisions may be made by a court or by a court appointed person where the court determines that is appropriate (through the new Court of Protection, with preference given to decisions made by the court rather than the appointment of a deputy), and all decisions are challengeable in court. This latter element is vital. Whilst this approach is clearly consistent with the provisions of Recommendation 99(4), it is challengeability which lies at the heart of its compliance with human rights obligations. What the Act does not do is require an initial judicial or administrative decision. It is not triggered by having to go to court or through some other governmental or quasi-governmental body. This will no doubt be challenging to many whose commitment to due process and procedural justice would rely upon judicial instigation of a procedure. However, this is not the only necessary approach, as is evidence in Recommendation 99(4). What is provided instead is a straightforward ability to challenge decisions by taking a matter about competence or about decisions made on behalf of someone who is or may be incompetent to the Court of Protection. Some people can launch a case as of right, some need the permission of the Court (section 50). This is clearly sufficient to meet

the demands of, for example, Article 6 of the European Convention on Human Rights, provided it operates in practice. If it fails to take cases that should get to the Court, then that might be a base of challenging the provision. The advantage of this approach, which are consistent with that of Recommendation 99(4), is that it more closely reflects the process applicable in relation to someone who is capable, it places emphasis on the fact that most decisions are taken on behalf of an incapable person by carers and those decisions are proper and appropriate and it provides a system that can work (that is the workload should be manageable and not be prohibitively expensive).

The focus for making decisions on behalf of another in England and Wales as in most jurisdictions, is individualized, function specific and based on the best interests of the individual. This *best interests approach* is, realistically, the only one available to the courts where there is no evidence of that person's preferences. Substituted judgement is, however, an appropriate approach where there is sufficient evidence of the decision that the person now incompetent would have made.[30] So, for example, decisions made in advance, advance health-care statements, are legally valid[22] (see also Kennedy and Grubb,[30] referring to such an approach in Florida, Ontario, Manitoba, and Victoria). Indeed, the old thinking that substituted judgement was an alternative to best interests should be re-thought so that what an individual wants is what is in her best interests, but if that is not known, nest interests is the only available approach. Where the person is not competent, the best interests approach, in England, was achieved by deciding whether what the doctor proposes in the given case is a treatment regime of which a responsible medical opinion would approve. This approach to 'best interests' was rightly severely criticized for it failed to address the issue by concentrating upon the interests of incompetent persons but professionalized it through one (medical) profession when it is possible to take a broader view of the issues in question when deciding upon what treatment to agree upon. Some of these criticisms were ameliorated by judicial developments (see, e.g. *Re S (Adult Patient: Sterilization)* and *Re A (Male Sterilization)*) that ensured two key changes. First, it was decided that, where there were options of a number of possible, acceptable approaches, only one of those options could be in the best interests of the individual. Secondly, in deciding what was in someone's best interests that should not be limited to scientific or medical matters, but should take in the whole range of social, welfare, and emotional factors. Despite such substantial judicial development, new law is in place through the Mental Capacity Act 2005. Section 4 provides a definition of best interests.

> 4(1) In determining for the purposes of this Act what is in a person's best interests, the person making the determination must not make it merely on the basis of –
>
> (a) the person's age or appearance, or
>
> (b) a condition of his, or an aspect of his behaviour, which might lead others to make unjustified assumptions about what might be in his best interests.
>
> (2) The person making the determination must consider all the relevant circumstances and, in particular, take the following steps.
>
> (3) He must consider –
>
> (a) whether it is likely that the person will at some time have capacity in relation to the matter in question, and
>
> (b) if it appears likely that he will, when that is likely to be.
>
> (4) He must, so far as reasonably practicable, permit, and encourage the person to participate, or to improve his ability to participate, as fully as possible in any act done for him and any decision affecting him.
>
> (5) Where the determination relates to life-sustaining treatment he must not, in considering whether the treatment is in the best interests of the person concerned, be motivated by a desire to bring about his death.
>
> (6) He must consider, so far as is reasonably ascertainable –
>
> (a) the person's past and present wishes and feelings (and, in particular, any relevant written statement made by him when he had capacity),
>
> (b) the beliefs and values that would be likely to influence his decision if he had capacity, and
>
> (c) the other factors that he would be likely to consider if he were able to do so.
>
> (7) he must take into account, if it is practicable and appropriate to consult them, the views of –
>
> (a) anyone named by the person as someone to be consulted on the matter in question or on matters of that kind,
>
> (b) anyone engaged in caring for the person or interested in his welfare,
>
> (c) any donee of a lasting power of attorney granted by the person, and
>
> (d) any deputy appointed for the person by the court,
>
> as to what would be in the person's best interests and, in particular, as to the matters mentioned in subsection (6).

The law relying on status

Whilst it has so far been asserted that approaches not facilitating decision-making by a competent person are the norm and that capacity should not be questioned on the basis of status, it is the case that status may play a role. Indeed, for example, in England and Wales status has, in the past, been the basis for effectively determining whether someone is capable, but increasingly these are being removed, as is the case across the globe. It is also a move demanded by international instruments, such as the Council of Europe Recommendation.

In England and Wales, it used to be the case that children under 16 were not able to consent, but the House of Lords changed this in 1985 when, in *Gillick* v *Wisbech AHA* it was recognized that, at least for some treatments, a person under the age of 16 could be capable of deciding upon treatment provided they had sufficient maturity to do so. The removal of the automatic barrier was an important step.

Further, the Sexual Offences Act was another example of law in a private area that was dependent upon status. For most people, it always has been a matter that is dependent upon their own consent. So, for example, it is rape for a man to have anal or vaginal sexual intercourse with a woman or a man who does not consent. The critical question is whether the victim is competent to make the decision. For example, does she or he understand what sexual intercourse is so that her or his apparent assent is indeed consent? However, under the 1956 Act this was not the case in all instances. First, there was an age of consent. Below the relevant age, the consent of the victim was irrelevant, however competent she or he may be. Second, a person who was a 'defective' in the terms of the Sexual Offences Act 1956 or had a 'severe mental handicap' could not consent. These terms were defined in the same way and referred to a person who had a state of arrested or incomplete development of mind that is associated with severe impairment of intelligence

and social functioning. If a person fell into this category, she or he could not, in law, consent to sexual intercourse or other sexual activity, however competent he or she might have been. This had real impact on some people with mental retardation (intellectual disability) and their carers. Interestingly, if a man married a woman with a severe intellectual disability (who was, therefore, a 'defective'), he did not commit the offence contrary to Section 7 of the Sexual Offences Act 1956 (whereby it was an offence for a man to have unlawful sexual intercourse with a woman who was a defective). Marriage was and is not dependent upon status, but is dependent upon an assessment of the competence of the particular individuals at the time of the marriage ceremony, and so falls into the second category of laws.[33] The law meant that a person could not prevent an indecent assault in any circumstances by consent. This could have unfortunate consequences for sex education for some people with severe mental retardation. It may be that no other form of sex education is possible than hands-on education to provide skills to enable appropriate behaviour by the person in question. Whatever the level of necessity, the individual could not consent in law. The only possible defence was to argue that the activity, contrary to appearances, was not indecent because of the purpose for which it was being undertaken as evidenced by the context, that is a carefully tailored and developed personal relationships programme. The difficulty here was that, when perceived from this stance, the law fails to allow sexual expression without good individual reason. Therefore it could be contrary to Article 8 (privacy) or Article 12 (founding a family) of the European Convention on Human Rights. The law presumably assumed that sexual experiences would, by definition, exploit or abuse such people. Law to prevent exploitation and abuse is very important, but it must not improperly limit a person's human rights. Subsequently, there has been an attempt to redress the balance somewhat and provide for a set of laws that is more acceptable. The Sexual Offences Act 2003 creates three types of offences that have an impact where the victim is a person with a mental disorder. The first consists of offences involving sexual activity with a person with a mental disorder and apply where that person cannot consent. The second consists of offences where the person's agreement is achieved through an inducement, threat or deception. The third comprises offences where the defendant is in some form of care relationship with the victim and these offences are committed regardless of consent and clearly deal with a significant form of exploitation and abuse. It is worth focusing, briefly, on an exemplar from the first type of offences. Section 30 of the 2003 Act makes it an offence for someone to engage in sexual activity with a person with a mental disorder impeding choice. As the offence relies upon touching, it involves sexual intercourse and other sexual activity short of intercourse but that involves touching. What it does not include is activity that would not involve touching, but would have been regarded as an assault under the old law. That touching must be sexual. It remains to be clarified whether sex education of a direct manner described above would be caught, but there must at least be an argument that it would not, provided that there is a non-sexual purpose, established by the education programme within which it falls and that a multi-disciplinary group identifies the need for the touching and the means of doing it. The offence applies potentially to anyone with a mental disorder, and is not so status driven as was the old law reliant upon being a defective or a person with a severe mental handicap. There is, though, still the status requirement of a mental disorder, which might be justified upon the basis of necessity and proportionality, as there must be an inability to refuse on the basis of that mental disorder, so the offence does not apply simply because of the presence of the disorder. This requirement is the equivalent of a capacity requirement tied to the specific requirements of the matter in hand. Finally, the defendant must know or could reasonably have known of the other's mental disorder and that he or she is thereby unable to refuse. These offences are likely to produce a much better balance of the need to protect the rights of vulnerable persons to understand and exercise their sexuality with their right to be protected by the law from sexual abuse and exploitation.

The law of responsibility

Criminal liability

Mental disorder is relevant to criminal liability in a variety of different ways, and this is true, at least, in all common-law jurisdictions. First, the presence of a mental disorder may be a reason to convince the decider of the fact that, contrary to external appearances, the defendant did not have the mental element for the crime with which he or she has been charged. Lack of the mental element is, of course, a complete defence. Whilst there is not a mental element requirement for all crimes (since there are some crimes for which conviction is based on strict liability), most serious crimes require a mental element that takes the form of intention, recklessness, knowledge, or belief and demand a consideration of what was the individual's purpose, awareness, foresight, or realization at the time that the crime was committed.

Second, the presence of a mental disorder may give rise to a defence. This will arise either by it being raised by the defence (even if the defence does not have the formal burden to prove it, but will have the burden of raising the matter for consideration) or by the prosecution challenging a point made by the defence (so, for example, if the defence raises mental disorder as an explanation for lack of criminal intent, the prosecution may respond by arguing that the defence has, in fact, raised one of the defences concerned with mental disorder). The most obvious defence is that of insanity. Whilst this takes various forms in many jurisdictions, there is usually a relationship in common-law jurisdictions with the English defence that was established by judicial answers to questions posed in McNaghten's case in 1843. This defence demands that the following matters be established by the defence (this is one of those rare instances in which the burden of proof lies, on a balance of probabilities, on the defence).

The defendant must have a disease of the mind. There are at least two possible ways of approaching this concept. First, it may be regarded as a simple concept in that it is present if the defendant has a condition (loosely termed) that has an internal cause, whereas there is no disease of the mind if there is an external cause. This simplistic distinction means that a person who has a brain tumour has a disease of the mind as does a person who has arteriosclerosis, schizophrenia (or other mental illness), epilepsy diabetes (provided the defendant had not taken her or his insulin and caused the offence in a hypoglycaemic state, or is a sleepwalker). The person with diabetes who causes the offence after taking insulin (an external agent) but falls into a hypoglycaemic state does not have a disease of the mind because the cause (insulin) is external and not internal. Immediately it can be seen that there is no congruence

between the legal construct of disease of the mind and any medical approach. Furthermore, this definition of disease of the mind produces outcomes that are clearly unacceptable (no one would argue that many of the conditions identified above are identifiable with mental disorders). An alternative definition, which may reduce some of the impact of the concept is to follow Lord Denning: 'it seems to me that any mental disorder which has manifested itself in violence and is prone to recur is a disease of the mind. At any rate it is the sort of disease for which a person should be detained in hospital rather than be given an unqualified acquittal'.

One difficult area of applying the external/internal causes distinction is in relation to 'whether a 'dissociative state' resulting from a 'psychological blow' amounts to insane automatism'. This was recognized by the Supreme Court of Canada so that the psychological blow can be recognized as an external cause giving rise not to insanity but to automatism as a defence.

The disease of the mind must cause a defect of reason. Defect of reason means that 'the powers of reasoning must be impaired and that a mere failure to use powers of reasoning which one has is not within the [McNaghten] Rules'.(47)

The consequences of actions A and B must be that the defendant either does not know what he is doing or does not know that what he is doing is legally wrong. The latter is contentious, because the Rules simply state that the defendant must know that what he was doing was wrong. In *R* v. *Windle* it was the Court of Appeal that established that the requirement was that the matter is concerned with legal wrong. The High Court of Australia has refused to follow this approach. In *Stapleton* v. *R*, 'their view was that if D believed his act to be right according to the ordinary standard of reasonable men he was entitled to be acquitted even if he knew it to be legally wrong'.(45)

The outcome of a finding of insanity is that the defendant is found not guilty by reason of insanity and, since the amendments introduced by the Criminal Procedure (Insanity and Fitness to Plead) Act 1991, the disposal of the defendant is not limited to being sent to a mental hospital under the equivalent of a restriction direction, but extends to less draconian forms of disposal, including discharge.

Third, there are limited or partial defences. In English law there is a defence of diminished responsibility that is a defence only to murder and produces, if successfully raised by the defendant, a conviction for manslaughter. There are similar defences, often of more general application in most, if not all, common-law countries. The defence of diminished responsibility was created by the Homicide Act 1957, Section 2, which provides:

> Where a person kills or is a party to the killing of another, he shall not be convicted of murder if he was suffering from such abnormality of mind (whether arising from a condition of arrested or retarded development of mind or any inherent causes or induced by disease or injury) as substantially impaired his mental responsibility for his acts and omissions in doing or being a party to the killing.

The question of impairment of responsibility is one for the decider of fact to make. Interestingly, expert evidence usually contains an assessment of the degree of impairment, though this would appear not to be a matter upon which the expert has the relevant training or expertise. Were the expert witness not to proffer a view on the matter, the practical reality is that the court would find it very difficult to know how to react to a claimed defence. The original rationale for this partial defence was to avoid the rigour of capital punishment. Its current rationale is wide ranging. Amongst other reasons why this defence is important is that it allows the defendant to argue that he or she was incapable of resisting an impulse produced by mental disorder, an argument that is not permissible in the insanity defence as is made clear by the McNaghten Rules themselves.

Fourth, there are other defences which are or may be related to mental disorder. One obvious defence is that of intoxication. If a person is intoxicated by drink or drugs such that he or she does not have the criminal intent for an offence, he or she is not guilty of that offence provided the offence is one of specific intent (such as murder), whereas he or she will be guilty if the crime is one of basic intent (such as manslaughter). Although a range of theories have been propounded to establish when a crime is one of specific or basic intent, the only approach that actually works is to take previous decisions as precedents for future approaches, and so develop a list of crimes of specific and of basic intent.

Tortious liability

Liability in tort encompasses a wide variety of non-contractual civil wrongs, including such diverse matters as negligence, nuisance, defamation, and trespass to land. It is worth noting that trespass to the person (assault and battery) is a civil wrong as well as a criminal offence and concurrent liability will lie. A battery occurs when there is a non-consensual touching, which includes the administration of medical treatment and care so that regardless of any benevolent motivation, if this is given without a valid consent an action will lie. However, this section concentrates on the law of negligence which is the main area where liability problems might occur. Again, the authors must emphasize that we concentrate mainly upon English law, but within the area of tort the law in other English-speaking jurisdictions is very similar e.g. in the USA, Australia, New Zealand, and Canada, and even in codified jurisdictions, such as France, Spain and Germany, there are similarities of approach.

In this section we look first at the general principles of the law of negligence. Second, any special considerations when examining the liability of the mentally ill defendant in an action in negligence are discussed. Third, the liability of third parties for the acts of the mentally ill are examined, and finally, the liability of third parties towards mentally ill patients, particularly in the context of statutory duties, and the difficult and controversial issue of the liability towards patients who harm or threaten harm to themselves are discussed.

(a) Negligence

The law of negligence is heavily circumscribed by a conceptual framework which is designed to restrict the ambit of claims. Much of this need not concern us here, suffice to say that the most relevant parts of it relate to the difference between certain types of loss and the way in which the courts regard these differences, and the role of public policy in some of the principal judgments.

The basic requirements of the tort of negligence have a certain simplicity; it is in the application of these requirements that complexity and confusion result. The claimant must show (a) that the defendant owed him or her a duty of care; (b) that the defendant breached that duty by failing to meet the requisite standard of care, and (c) that the breach of duty caused the resulting damage. It is important to note the latter. Negligence is not actionable *per se*;

there must be some tangible damage caused by the breach. There is insufficient space here to consider duty of care in all its contexts, but it is important to note the test put forward in *Caparo Industries PLC* v. *Dickman*. In that case it was said that in order for a duty of care to arise, the damage suffered must be foreseeable; there must be sufficient proximity between the claimant and the defendant; and it must be 'just and reasonable' to impose a duty of care. The first two aspects are well illustrated by examining the difference between physical injury and other forms of damage; the third is an important consideration in the context of statutory duties. As far as the standard of care is concerned, the standard is objective and based upon reasonableness. Causation must always be proved as there will be no liability for falling below the requisite standard of care if a causal link to the damage cannot be proved. The standard of proof throughout is that of the balance of probabilities.

(b) Liability for clinical negligence

Duty of care is not usually an issue in the doctor/patient relationship. It should be noted that the standard of care relating to professionals such as doctors is set by the 'accepted practices' test, well known as the 'Bolam test' as framed in the case of *Bolam* v. *Friern Hospital Management Committee*, where diagnosis and treatment will not normally be negligent if supported by a responsible body of medical opinion. The only exception to this is if the doctor's actions do not withstand logical analysis; in those circumstances the fact that the practice is 'accepted' will not be sufficient to avoid a finding of negligence. The test is slightly different if the allegation of negligence relates to the provision of inadequate information about risks, side effects and alternative treatments. In such cases it seems that the courts are moving towards a more stringent standard of informed consent. Proving that the damage was caused by the breach can raise many problems in medical negligence claims, largely for two reasons. First, in almost all cases the patient will be suffering from some medical condition at the outset, and it might be arguable that the resulting condition would have materialized anyway. Second, if the resulting condition is the realization of some form of risk (e.g. side-effects of medication), even if the patient was not warned of the risk it is open to the defendant to argue that regardless of warnings, the patient would have gone ahead and consented to the taking of the medication in any event.

(c) Liability for different types of damage—psychiatric injury

The law distinguishes between different types of damage. This appears to stem from judicial fear that certain losses might result in an unacceptably large number of potential claimants i.e. the fear of opening the 'floodgates'. Thus, what is known as 'pure economic loss', that is financial loss, which is not consequential upon physical damage, will only be recoverable in certain circumstances. For similar reasons the law makes another significant distinction: the difference between physical damage and psychiatric damage.

There are a number of reasons for the disparity between the treatment of the two types of injury. The first stems from the misunderstanding in earlier cases of the nature of psychiatric injury. In the authors' view, it is no coincidence that some of the early cases involved pregnant women who had miscarried or given birth to damaged children (e.g. *Dulieu* v. *White & Sons*). In these cases the courts could see a tangible manifestation of the 'shock' suffered by the claimants. Psychiatry was in its infancy, and the myriad of subtle manifestations of mental disorder was unknown. A further illustration of this is the phenomenon of 'railway spine'. In 1875, J.E. Erichsen, Professor of Surgery at University College Hospital, London, published a number of findings as a result of studying spinal injury cases following railway accidents, where there was no obvious physical cause of the symptoms manifested. His conclusion was that trauma caused 'concussion of the spine'. Later medical opinion such as that of surgeon Herbert Page denied that the spine as such was affected by the trauma, but that the resulting condition was 'nervous shock'. It is therefore understandable that courts may have taken a somewhat crude approach to shock-induced injury, which crudity is still reflected in the law, which is commonly referred to as 'the law of nervous shock'. However, the second reason is, perhaps, less acceptable: the fear of the floodgates opening and admitting unacceptably large numbers of claims. It is worth pointing out that the floodgates argument, being only a spectre (but a highly influential one at that) and not overtly referred to by judges, has never been supported by hard empirical evidence. Nevertheless, the Law Commission, in its report on psychiatric injury claims supported the floodgates argument, largely on the basis of the lack of clear demarcation lines between general, uncompensatable, mental disturbance, and specific psychiatric illnesses. A third reason for the distinction between physical and psychiatric injury is the argument that a too-liberal attitude might result in fraudulent claims. However, again there is no clear evidence to show that it is easy to fake psychiatric injury, and furthermore it is an odd system of justice which takes the approach that, because there is a possibility of fraud, genuine claims should fail.

It is important to note that in cases where the claimant has suffered physical injury, any claim for damages for associated psychiatric injury will not be controversial, subject to the claimant establishing causation. However, where the claim is for psychiatric injury alone special considerations apply and, it is worth summarizing the essential elements of such a claim. The distinction is often drawn between primary and secondary victims. A primary victim is either someone who is physically injured or is within the range of foreseeable physical injury (see *Page* v. *Smith*).

A secondary victim is someone who has some sort of proximity to one or more primary victims. (There are similarities in other European jurisdictions; for example in France this is *dommage par ricochet*, and the rules are more generous than in English-speaking jurisdictions.) In English law this has two aspects: first, there must be close physical proximity to the trauma which injures or threatens the primary victim; second, the secondary victim must either have a close tie of love and affection with one or more of the primary victims, or he or she must be a rescuer of a primary victim and be within the range of foreseeable physical injury (see *Alcock* v. *Chief Constable of South Yorkshire Police*) and *White* v. *Chief Constable of South Yorkshire Police*). There is one other condition which a secondary victim must meet: the trauma must have the necessary quality of 'shockingness' about it so as to make a sudden impact on the senses to such an extent that the person of normal fortitude must be shocked by it (the 'impact rule'). Finally, in the cases of both primary and secondary victims, the negligence must result in a recognized psychiatric injury; emotional conditions such as grief, fear, and distress will not suffice. Of course, if the claimant has suffered distress as a result of physical injury, not amounting to a recognized psychiatric condition, then this should be reflected in the damages for pain, suffering, and loss of amenity. Although ordinary 'shock' is not compensatable, both DSM-IV and ICD-10 refer to 'acute stress reaction' which, although only a short-term

reaction to stress, should be compensatable as a recognized psychiatric condition, albeit that any compensation would be modest. (See *Phelps* v. *London Borough of Hillingdon*, where it was held that the failure to mitigate the adverse consequences of a congenital defect such as dyslexia was capable of constituting an 'injury'.)

The impact rule states that the claimant must be present at the traumatic event or its immediate aftermath, and it must be an event such as to make an impact upon the unaided senses of the claimant, and to be shocking to a person of normal fortitude. The case of *Sion* v. *Hampstead Health Authority* is illustrative of this point. The claimant was the father of a young man injured in a road accident. The defendant was the health authority because the father alleged that his son was treated negligently. The claimant watched his son deteriorate and die over a period of 14 days. The resultant psychiatric illness suffered by the claimant was not compensatable as the process of death was slow, predictable, and not 'shocking'. This can be contrasted with *Tredget and Tredget* v. *Bexley Health Authority*, which concerned the negligent delivery of a child, which took place in an atmosphere of 'chaos' and 'pandemonium', and resulted in the child being born in a distressed state requiring immediate resuscitation. The father, who was present at the birth, recovered damages for his subsequent psychiatric condition, as the event was sudden and impacted in a shocking way upon his senses. In *Walters* v. *North Glamorgan NHS Trust*) damages were awarded to a woman who spent 36 hours with her baby son from seeing him choking on blood and vomit to the termination of life support. This period was found to be a shocking series of events. However, in other jurisdictions the need to find something sufficiently shocking upon which to hang liability has been rejected. In the US case of *Ochoa* v. *Superior Court (Santa Clara County)*, a woman recovered damages when, despite her son begging her to stay, was forced to leave his hospital bedside on the basis that he only had flu, and never saw him alive again. In Singapore, it was said that in the context of medical negligence as the secondary victim would rarely witness the act of negligence, it was wrong to impose the 'shocking event' requirement. Thus, in *Pang Koi Fa* v. *Lim Djoe Phing*) damages were awarded when a negligent medical procedure took place in June 1985 and the claimant watched her daughter die from then until the following September.

These sorts of distinctions are important in the context of so-called 'creeping trauma' in which there is no sudden impact to the senses. This refers to cases where victims have been exposed to some form of risk, for example contamination or the administration of a drug which has caused no harmful effects yet, but which might do so. The absence of 'impact' might be thought to be fatal to such claims. However, in an Australian case (*AQP* v. *Commonwealth Serum Laboratories Ltd*) the Victoria Supreme Court held that such cases should be decided on the basis of foreseeability and proximity alone. Subsequently this was confirmed in an English decision, where a group of claimants had received the human growth hormone at a time when the Department of Health and the Medical Research Council was aware of the possibility of Creutzfeldt–Jakob disease contamination. They subsequently developed psychiatric illnesses through fear of contracting Creutzfeldt–Jakob disease, and recovered damages. At first sight it may be thought that the impact rule would not apply because the claimants were primary victims. Although this is the common-sense view inasmuch as they had all been directly administered a drug, the judge was not willing to concede this. This was on the basis that it would widen unacceptably the number of potential primary victims in a much more widespread case of possible contamination and would inhibit manufacturers and providers of drugs and other goods from warning the public even when the risk was very small. Therefore the claimants were held to be secondary victims, but to whom a duty was owed simply on the basis of foreseeability. A duty of care was found to exist because of the nature of the relationship between the defendants and the claimants, the small and readily identifiable size of the group of claimants, the ways in which the claimants might become aware of the risk of Creutzfeldt–Jakob disease, and the nature of the suffering in terminal Creutzfeldt–Jakob disease.

(d) The liability of the mentally ill in negligence

It might be thought that if the law of negligence is about fault and blame (as it is in the English-speaking jurisdictions, and some European jurisdictions as well), then those who have less appreciation of risk and the consequences of their own actions might be regarded as being less culpable than those with a reasonable degree of appreciation. However, this is not the case. It is important to understand that, although the historical development of the law of tort suggests a number of different principles behind tort compensation and that in rare circumstances punitive damages can be recovered, generally tort is concerned with compensating victims rather than punishing wrongdoers. In the light of this it is not difficult to see why the same standard of care is applied. The second reason for this is that courts do not want to apply a different standard of care depending upon factors such as intelligence, emotional reactions, and experience. Therefore the standard is objective, being that of the reasonable person who is reasonably competent at whatever task is being undertaken. (Note that there is an exception in the case of children who are judged in accordance with the reasonable foreseeability of the child of the relevant age.) The general principle is illustrated by the classic example of the learner driver who is expected to meet the standard of the reasonably competent driver (see the case of *Nettleship* v. *Weston*). In *Wilsher* v. *Essex Area Health Authority*, in the Court of Appeal, Mustill LJ stated: 'this notion of a duty tailored to the actor rather than to the act which he elects to perform, has no place in the law of tort' (the case subsequently went to the House of Lords but this aspect of the judgment was undisturbed). However, there have been some exceptions to this (albeit not in England) which have interesting implications for the liability of the mentally ill. In the case of *Cook* v. *Cook*, the High Court of Australia held that, in the case of an inexperienced driver, liability may be different depending on the identity of the defendant. For example, a car driver who injures a pedestrian will be subject to the same standard of care as the reasonably competent driver, but when a driving instructor is injured, the fact that he knew of the claimant's lack of experience might mean that there is no liability for his injuries. By analogy with *Nettleship* v. *Weston*,[68] a mentally disordered person who negligently injures a third party will be subject to the objective standard of care. These latter considerations, however, might have some relevance to the position of those who care for the mentally ill in that they might be expected to be aware of the increased likelihood of unstable behaviour. In other words, can it be argued, that if a patient negligently injures a doctor, nurse, or social worker, in the course of, say receiving treatment, then he might not be liable? It should be noted, however, that we are concerned with negligence, not deliberate acts, so this would not be a common scenario. Furthermore, the patient would very likely have

few resources and no insurance so may not be worth suing. Nevertheless, if an action were financially feasible, it may be that, for policy reasons, the courts would be unlikely to render someone in the position of a carer, without compensation. However, in *Mansfield* v. *Weetabix Ltd*, the Court of Appeal stated that in a case where the defendant could not reasonably have known about his condition, or the effect of it, there would be no liability.

(e) Liability to third parties for the acts of the mentally ill

The next question that arises is the extent, if any, of the liability of third parties for acts of the mentally ill, whether these are criminal or civil acts. The third party might be an individual who is caring for the defendant, but is much more likely to be a local authority or a health authority. This area of negligence has given rise to some emotive and controversial litigation. In *Hill* v. *Chief Constable of West Yorkshire Police* the House of Lords held that no liability attached to a police force that failed to arrest the serial killer Peter Sutcliffe before he killed his last victim. The decision did not turn upon whether there had been negligent conduct, but upon public policy. The interests of the public as a whole are best served, so runs the argument, if those responsible for public safety and so on are able to carry out their duties unfettered by the threat of litigation. It must be noted, however, that in the case of *Osman* v. *United Kingdom*, this so-called immunity from suit was held by the European Court of Human Rights to be contrary to Article 6.1 of the European Convention on Human Rights (the protection of the right to a fair trial in both criminal and civil cases).

In *Palmer v Tees Health Authority*, the claimant, whose four year old daughter was murdered by a psychopath, brought an action against the health authority, both in her own name as she had suffered from a psychiatric illness as a result of the murder, and on behalf of the child's estate. Some months prior to committing the murder, the defendant had been an outpatient at the defendant's psychiatric hospital, where he had allegedly told staff that he would kill a child. The Court of Appeal rejected her claims on the basis that there was not sufficient proximity between the defendant and the victim. If, however, the defendant had known his victim and made threats specifically against her then there may well have been sufficient proximity to establish liability.

(f) Liability of the providers of services to the mentally ill

In the case of *Clunis* v. *Camden and Islington Health Authority*, Clunis was a mentally disordered man who had been discharged from a psychiatric hospital and subsequently killed a man in an unprovoked attack. The prosecution accepted his plea of manslaughter on the ground of diminished responsibility. Clunis subsequently sought damages from the health authority alleging that they had failed to treat him in accordance with their common-law duty of care, and that if he had been treated he would not have carried out the killing and would not have been subject to the lengthy period of detention he was now facing. The court invoked the legal maxim of *ex turpi causa non oritur actio*, which means that they found that he could not establish a duty when it stemmed from the claimant's own wrong-doing. The court relied on this as an aspect of public policy, and, somewhat curiously, referred to the importance of the deterrent effect of such a maxim. Given that the claimant was mentally ill, and the motiveless nature of the attack, he was surely incapable of being deterred? It is tempting to consider that *Clunis* might have been decided differently in terms of *ex turpi causa*, if it had not also had implications for fettering the medical and resource-allocation discretion of health authorities to make them liable in negligence in this way.

It is well established that English health authorities can be both primarily liable (when there is an organizational failure) and vicariously liable (where an individual employee is negligent) for acts of negligence. However, the provision of statutory duties by local authorities has raised some difficult issues. The leading authority is *X* v. *Bedfordshire County Council* In this case (a number of different cases listed together) the court had to consider the claims of claimants who alleged damage as a result of the negligence of social workers and psychiatrists involved in child protection cases, and teachers and others involved in the provision of education to children with special educational needs. Applying the threefold test from Caparo the court found that the first two elements were satisfied, that is foreseeability of damage and proximity of relationship between the parties. However, on the third element, the finding was that it would not be 'just and reasonable' to impose a duty of care. In the context of child protection three reasons were given. First, the statutory system set up to protect children cuts across many disciplines: police, education bodies, doctors, and others. To disentangle these relationships and to ascertain where the blame must fall would impose 'impossible problems'. Second, because there is a very fine balance to be drawn between the protection of children and the disruption of removing them from their homes, there is often a conflict to be resolved. The professionals involved in making such decisions should not be further hampered by the threat of litigation. Furthermore, a common-law liability may result in too cautious and defensive an approach on the part of local authorities. Finally, on a practical point, because of the obvious tensions between the local authority and the parents involved, litigation would be commonplace thereby placing a drain on vital resources necessary for the carrying out of the statutory functions.

However, if the social workers and psychiatrists involved *personally* owed duties towards the claimants then, it was said, local authorities might be vicariously liable for any negligence on their part. Furthermore, the House of Lords speculated that educational services might be treated differently in that they were more in the nature of a 'service to the public'.

In *Phelps* v. *London Borough of Hillingdon*[60] the House of Lords confirmed this approach. It concerned a claim for the failure of an educational psychologist to diagnose dyslexia. The court stated that vicarious liability arose on the basis of the personal duty owed by the psychologist, notwithstanding the fact that the educational psychology service was part of the general statutory services provided under the Educations Acts of 1944 and 1981, and that there would be no liability on the part of the local authority for breach of those statutory duties. A similar issue arose in the case of *Clunis* v. *Camden and Islington Health Authority*[73] (the facts of which are outlined above). The court had to consider the extent of the after-care that should be provided under Section 117 of the Mental Health Act. Under this provision, there is a duty upon the district health authority and the local social services authority to provide after-care services for detained patients who leave hospital. It was alleged in this case that the psychiatrist responsible for monitoring the after-care of the claimant after discharge was negligent in failing to admit him to hospital, thereby preventing the act of homicide which followed. However, the Court of Appeal found that the services being provided by the psychiatrist after discharge

were 'essentially in the sphere of administrative activities in pursuance of a scheme of social welfare in the community'. These, it was stated, were different from the duties owed in the context of the doctor–patient relationship, and there was no common-law duty of care. Furthermore, under Section 124 of the Act, an allegation that inadequate Section 117 services are being provided should be dealt with by way of complaint to the Secretary of State. The court found that the wording of the statute indicated a parliamentary intention that breaches of Section 117 should not give rise to a cause of action in private law. In consequence, it seems that the outlook is bleak for claims resulting from the negligence of psychiatrists, psychologists, and social workers in the context of certain statutory functions of local authorities and health authorities alike.

It must be stressed that, although these cases are referring to English statutory provisions, they raise general issues of principle as to how far statutory bodies are to be controlled by the courts, and, in particular, how far scarce public resources should be subject to depletion by litigation.

(g) Liability towards the mentally ill who harm themselves

As indicated above, the issues of consent to treatment, and competence to both consent and refusal of treatment are central to many legal systems. The case of *Re C* illustrates the importance of the right to self-determination in cases where, on the face of it, there might be a strong temptation to find a patient incompetent because of the nature of the mental illness from which he suffers.[22]

Under Section 63 of the Mental Health Act 1983, treatment can be given for mental disorder for which the patient is detained without the consent of the patient, and competency is irrelevant. In *B* v. *Croydon Health Authority* treatment for mental disorder was construed very widely. In that case the patient was clearly a 'self-harmer', and force-feeding due to her refusal of food was held to be treatment for the mental disorder. If she had been incompetent, then treatment (of any sort, not just for the mental disorder) could have been administered in her best interests. The question then arises as to whether an incompetent patient not detained under the Act who, however, refuses food in an attempt to self-harm must be fed, failing which there would be a breach of the duty of care. It must be noted that the courts are extremely reluctant to question medical decisions about treatment. It is left to clinical judgement to decide whether to give, or withdraw, treatment (see the case of *Airedale NHS Trust* v. *Bland*). (In the same case it was confirmed that treatment includes feeding; see also the Californian case of *Bouvia* v. *Superior Court*.) The decision as to whether treatment is necessary is judged by reference to the Bolam test. In other words if there are good clinical reasons for treating or not treating, and these can be justified by reference to a responsible body of medical opinion, then there will be no liability for failure to treat.

The question that arises is whether the health authority would have been liable nevertheless for failing to force-feed the subject of *B* v. *Croydon Health Authority*,[77] even though she was deemed capable of refusing treatment. It would follow from the general principle of self-determination, well established in English law (see, for example, *Re T* (*Adult: Refusal of Treatment*) and *Re MB* (*Adult: Medical Treatment*),[23]) that such a refusal should be respected. However, recent case law in the context of the duty owed by prisons, and by implication hospitals, has considerably muddied the waters.

In the case of *Kirkham* v. *Chief Constable of Greater Manchester Police* the court took the opportunity to look at the difference between negligent acts and negligent omissions (something that has exercised the minds of tort lawyers for some time) and said that, in this case, responsibility for the omission to prevent the suicide had been assumed by the defendant. The prisoner was said to be 'of unsound mind', which presumably meant that, in the context of the right to self-determination, the prisoner was incapable of exercising his autonomy. In those circumstances the decision is not surprising on its facts. However, the judgment suggested that when one person is in the lawful custody of another, whether that be voluntarily such as a hospital, or involuntarily such as a prison, there is a duty to take all reasonable steps to avoid acts or omissions to prevent reasonably foreseeable injury. Consequently in the case of *Reeves* v. *Commissioner of Police of the Metropolis*, another custodial suicide, the same duty was said to be owed to a 'sane' prisoner. Again, the wording is unfortunate by its reference to sanity, but it can be assumed that this prisoner was deemed to be competent. The police surgeon had stated, after examining the claimant, that there was no evidence of mental disturbance. It must be stressed that it will be necessary in all cases including those concerning hospitals, for there to be evidence that it was reasonable to regard the patient as a suicide risk. Of course, in practice, it may be the case that it is undesirable for hospitals to be venues for acts of suicide. However, it is by no means clear that, in the case of someone who has capacity, there should be a positive duty to prevent suicide any more than there should be a duty owed to prevent that person from smoking. These decisions are invoking a strong form of paternalism. Certainly in the case of hospital patients many competent patients discharge themselves against medical advice whereupon they are free to take any form of self-harming action. It is difficult to see why there should be a positive duty to prevent the competent patient from carrying out the acts on hospital premises.

Further information

Barrie, P. (2005). *Personal Injury Law*. Oxford University Press, Oxford.

Bartlett, P. and Sandland, R. (2007). *Mental health law: policy and practice*, (3rd ed.), Oxford University Press, Oxford.

Bartlett, P. (2007). *Blackstone's Guide to the Mental Capacity Act 2005*, (2nd ed.), Oxford University Press, Oxford.

Grisso, T. and Appelbaum, P.S. (1998). *Assessing competence to consent to treatment: A guide for physicians and other health care professionals*, Oxford University Press, New York.

Harpwood, V. (2005). *Modern Tort Law,* Routledge-Cavendish, London.

Mackay, R.D. (1995). *Mental condition defences in the criminal law.* Oxford University Press, Oxford.

Weir, T. (2006). *An Introduction to Tort Law*, Oxford University Press, Oxford.

Wheat, K. (2002). *Recovering Damages for Psychiatric Injury*.Oxford University Press, Oxford.

References

1. Ashibe, N. (1997). *Kempou (The Constitutional Law) (*2nd edn.), Iwanam, Tokyo.
2. Kitamura, F., Tomoda, A., Tsukada, K., *et al.* (1998). Method for Assessment of Competency to Consent in the Mentally Ill: Rationale, Development and Comparison with the Medically Ill. *International Journal of Law and Psychiatry*, **21**, 223–44.
3. Kitamura, T., Kitamura, F., Mitsuhashi, T., *et al.* (1999). Image of Psychiatric Patients. Competency to Give Informed Consent to

Treatment in Japan: I. A Factor Analytic Study' *International Journal of Law and Psychiatry,* **22**, 45–54.
4. Nakatani, Y. (2000). Psychiatry and the Law in Japan: History and Current Topics. *International Journal of Law and Psychiatry,* **23**, 589–604.
5. Satoh, K. (1995). *Kempou (The Constitutional Law)* (3rd edn.). Seirin Shobo, Tokyo.
6. Friedman, P.R. (1976). *The Rights of Mentally Retarded Persons*. Avon Books, New York.
7. Herr, S.S. (1983). *Rights and Advocacy for Retarded Persons*. Lexington Books, Lexington.
8. Carney, T. (2001). Globalisation and guardianship: harmonisation or (postmodern) diversity?. *International Journal of Law and Psychiatry,* **24**, 95–116.
9. South African law Commission, 2004. *Assisted Decision-Making: Adults with Impaired Decision-Making Capacity*. Discussion Paper 105 on Project 122.
10. Ritchie, J., Sklar, R., Steiner, W., *et al.* (1998). Advance Directives in Psychiatry: Resolving Issues of Autonomy and Competence. *International Journal of Law and Psychiatry*, **21**, 245–60.
11. Callahan, D. (1984). Autonomy: A moral good not a moral obsession. *Hastings Center Report,* **14**, 40–2.
12. Mappes, T.A. and Zembaty, J.S. (1994). Patient choices, family interests and physician obligations. *Kennedy Institute of Ethics Journal,* **4**, 27–46.
13. Carney, T. (2001). Globalisation and guardianship: harmonisation or (postmodern) diversity?. *International Journal of Law and Psychiatry,* **24**, 9–116.
14. Atkins, M.R. (1997). Adult Guardianship Reforms – Reflections on the New Zealand Model. *International Journal of Law and Psychiatry,* **20**, 77–96.
15. Frank, J.B. and Degan, D. (1997). Conservatorship for the Chronically Mentally Ill: Review and Case Series. *International Journal of Law and Psychiatry,* **20**, 97–111.
16. South African Law Commissio, 2004. *Assisted Decision-Making: Adults with Impaired Decision-Making Capacity* Discussion Paper 105 on Project 122.
17. Jansen, S. (2000). Recommendation No. 99(4) of the Committee of Ministers to Member States on Principles Concerning the Legal Protection of Incapable Adults: An Introduction in particular to Part V Interventions in the Health Field. *European Journal of Health Law,* **7**, 333–47.
18. Jansen, S., (2000). Recommendation No. 99(4) of the Committee of Ministers to Member States on Principles Concerning the Legal Protection of Incapable Adults: An Introduction in particular to Part V Interventions in the Health Field. *European Journal of Health Law,* **7**, 333–47.
19. Wong, J.G., Clare, I.C.H., Gunn, M.J., *et al.* (1999). Capacity to make health care decisions: its importance in clinical practice. *Psychological Medicine*, **29**, 437–46.
20. Grisso, T. and Appelbaum, P.S. (1998). *Assessing competence to consent to treatment: a guide for physicians and other health professionals*. Oxford University Press.
21. Grisso, T. and Appelbaum, P.S. (1995). The MacArthur Treatment Competence Study. III Abilities of patients to consent to psychiatric and medical treatments. *Law and Human Behavior*, **19**, 149–74.
22. *Re C* (*Adult: Refusal of Treatment*) [1994] 1 Weekly Law Reports 290.
23 *Re MB* [1997] 2 Family Law Reports 426.
24. Gunn, M.J., Wong, J.G., Clare, I.C.H., *et al.* (1999). Decision-making capacity. *Medical Law Review*, **7**, 261–98.
25. President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioural Research (1983). *Making health care decisions*. US Government Printing Office, Washington, DC.
26. Weisstub, D. (1990). *Enquiry on mental competency: fi nal report*. Queen's Printer, Toronto.
27. Law Commission (1995). *Mental incapacity*. Law Commission Report No. 231. HMSO, London.
28. Scottish Law Commission (1995). *Report on incapable adults*. Scottish Law Commission Report No. 151. HMSO, Edinburgh.
29. Morris, C.D., Niederbuhl, J.M., Mahr, J.M., *et al.* (1993). Determining the capability of individuals with mental retardation to give informed consent. *American Journal of Mental Retardation*, **98**, 263–72.
30. Kennedy, I. and Grubb, A. (1994). *Medical law: text with materials* (2nd edn). Butterworths, London.
31. Grisso, T. (1986). *Evaluating competencies: forensic assessments and instruments*. Plenum Press, New York.
32. *Banks* v. *Goodfellow* (1870) Law Reports 5 Queen's Bench 549.
33. Hoggett, B.M. (1996). *Mental health law*. Sweet and Maxwell, London.
34. Carney, T. and Singer, P. (1986). *Ethical and legal issues in guardianship options for intellectually disadvantaged people*. Australian Government Publishing Service, Canberra. 35. *Re F* (1990) 2 Appeal Cases 1.
36. Ouslander, J., Tymchuk, A., Rahbar, B., *et al.* (1989). Health care decisions among elderly long-term care residents and their potential proxies. *Archives of Internal Medicine*, **149**, 1367–72.
37. *Re S (Adult Patient: Sterilization)* [2001] Fam 15.
38. *Re A (Male Sterilization)* [2000] 1 F.L.R. 549.
39. Gunn, M.J. (1996). *Sex and the law: a guide for staff working with people with learning difficulties* (4th edn.). Family Planning Association, London.
40. Stevenson, K., Davies, A. and Gunn, M., Blackstone's Guide to the Sexual Offences Act 2003, chap. 6.
41. Stevenson, K., Davies, A. and Gunn, M., Blackstone's Guide to the Sexual Offences Act 2003, pp. 91–92.
42. Stevenson, K., Davies, A. and Gunn, M., Blackstone's Guide to the Sexual Offences Act 2003, chap. 2.
43. Stevenson, K., Davies, A. and Gunn, M., Blackstone's Guide to the Sexual Offences Act 2003, p. 100.
44. Mackay, R.D. (1995). *Mental condition defences in the criminal law*. Clarendon Press, Oxford.
45. *McNaghten's Case* (1843) 10 Clarkson and Findlay Law Reports 200.
46. Lord Denning (1963) *Bratty* v. *Attorney General* [1963] Appeal Cases 386.
47. Ormerod, D.C. (2005). *Smith and Hogan: criminal law* (11th edn). Butterworths, London.
48. *Rabey* (1977) 79 Dominion Law Reports (3rd series) 414.
49. *R* v. *Windle* [1952] 2 Queen's Bench 826.
50. *Stapleton* v. *R* (1952) 86 Commonwealth Law Reports 358.
51. *Caparo Industries PLC* v. *Dickman* [1990] 1 All ER 568.
52. *Bolam* v. *Friern Hospital Management Committee* [1957] 2 All ER 118.
53. *Bolitho* v. *City & Hackney Health Authority* [1997] 4 All ER 771.
54. *Chester* v. *Afshar* [2004] UKHL 41.
55. *Dulieu* v. *White & Sons* [1901] 2 KB 669
56. Law Commission (1998). *Liability for psychiatric illness*. Law Commission Report No. 249. HMSO, London.
57. *Page* v. *Smith* [1995] 2 WLR 644.
58. *Alcock* v. *Chief Constable of South Yorkshire Police* [1992] 4 All ER 907.
59. *White* v. *Chief Constable of South Yorkshire Police* [1999] 1 All ER 1.
60. *Phelps* v. *London Borough of Hillingdon* [2001]] 2AC 619.
61. *Sion* v. *Hampstead Health Authority* [1994] 5 Med LR.
62. *Tredget and Tredget* v. *Bexley Health Authority* [1994] 5 Med LR.
63. *Walters* v. *North Glamorgan NHS Trust* [2002] EWHC 321.
64. *Ochoa* v. *Superior Court (Santa Clara County)*(1985) 703 P 2d 1.
65. *Pang Koi Fa* v. *Lim Djoe Phing* [1993] 3 SLR 317.
66. *AQP* v. *Commonwealth Serum Laboratories Ltd* (unreported, Sup Ct (Vic) 5 February 1995).
67. Creutzfeldt–Jakob disease litigation (unreported, QBD 18 December 1997).
68. *Nettleship* v. *Weston* [1971] 2 QB 691.

69. *Wilsher* v. *Essex Area Health Authority* [1986] 3 All ER 801.
70. *Cook* v. *Cook* (1986) 68 ALR 353.
71. *Mansfi eld* v. *Weetabix Ltd* [1997] PIQR P526.
72. *Hill* v. *Chief Constable of West Yorkshire Police* [1988] 2 All ER 238.
73. *Osman* v. *United Kingdom* [1999] 1 FLR 193.
74. *Palmer v Tees Health Authority* [1999] Lloyd's Rep Med 351.
75. *Clunis* v. *Camden and Islington Health Authority* [1998] 2 WLR 902.
76. *X* v. *Bedfordshire County Council* (1995) 2 AC 633.
77. *B* v. *Croydon Health Authority* [1994] 2 WLR 294.
78. *Airedale NHS Trust* v. *Bland* [1993] AC 789.
79. *Bouvia* v. *Superior Court* (1986) 179 Cal App3d 1127.
80. *Re T* (*Adult: Refusal of Treatment*) [1993] 4 All ER 649.
81. *Kirkham* v. *Chief Constable of Greater Manchester Police* [1990] 2 QB 283.
82. *Reeves* v. *Commissioner of Police of the Metropolis* [1998] 1 All ER 381.

11.2

Psychosocial causes of offending

David P. Farrington

Introduction

Scope of this chapter

Offending is part of a larger syndrome of antisocial behaviour that arises in childhood and tends to persist into adulthood. There seems to be continuity over time, since the antisocial child tends to become the antisocial teenager and then the antisocial adult, just as the antisocial adult then tends to produce another antisocial child. The main focus of this chapter is on types of antisocial behaviour classified as criminal offences, rather than on types classified for example as conduct disorder or antisocial personality disorder.

In an attempt to identify causes, this chapter reviews risk factors that influence the development of criminal careers. Literally thousands of variables differentiate significantly between official offenders and non-offenders and correlate significantly with reports of offending behaviour by young people. In this chapter, it is only possible to review briefly some of the most important risk factors for offending: individual difference factors such as high impulsivity and low intelligence, family influences such as poor child rearing and criminal parents, and social influences: socio-economic deprivation, peer, school, community, and situational factors.

I will be very selective in focussing on some of the more important and replicable findings obtained in some of the more methodologically adequate studies: especially prospective longitudinal follow-up studies of large community samples, with information from several data sources (e.g. the child, the parent, the teacher, official records) to maximize validity. The emphasis is on offending by males; most research on offending has concentrated on males, because they commit most of the serious predatory and violent offences. The review is limited to research carried out in the United Kingdom, the United States, and similar Western industrialized democracies. More extensive book length reviews of antisocial behaviour and offending are available elsewhere.[1]

I will refer especially to knowledge gained in the Cambridge Study in Delinquent Development,[2] which is a prospective longitudinal survey of over 400 London males from age 8 to age 40. Fortunately, results obtained in British longitudinal surveys of delinquency are highly concordant with those obtained in comparable surveys in North America, the Scandinavian countries, and New Zealand and indeed with results obtained in British cross-sectional surveys. A systematic comparison of the Cambridge Study with the Pittsburgh Youth Study showed numerous replicable predictors of offending over time and place, including impulsivity, attention problems, low school attainment, poor parental supervision, parental conflict, an antisocial parent, a young mother, large family size, low family income, and coming from a broken family.

Measurement and epidemiology

Offending is defined as acts prohibited by the criminal law, such as theft, burglary, robbery, violence, vandalism, and drug use. It is commonly measured using either official records of arrests or convictions or self-reports of delinquency. The advantages and disadvantages of official records and self-reports are to some extent complementary. In general, official records identify the worst offenders and the worst offences, while self-reports include more of the normal range of delinquent activity. The worst offenders may be missing from samples interviewed in self-report studies. Self-reports have the advantage of including undetected offences, but the disadvantages of concealment and forgetting. By normally accepted psychometric criteria of validity, self-reports are valid. Self-reported delinquency predicted later convictions in the Cambridge Study. In the Pittsburgh Youth Study,[3] the seriousness of self-reported delinquency predicted later court referrals. However, predictive validity was enhanced by combining self-report, parent, and teacher information about offending.

The key issue is whether the same results are obtained with both methods. For example, if official records and self-reports both show a link between parental supervision and delinquency, it is likely that supervision is related to delinquent behaviour (rather than to any biases in measurement). Generally, the worst offenders according to self-reports (taking account of frequency and seriousness) tend also to be the worst offenders according to official records. In the Cambridge Study between ages 15 and 18, 11 per cent of the males admitted burglary, and 62 per cent of these males were convicted of burglary. The predictors and correlates of official and self-reported delinquency were very similar.

Much is known about the epidemiology and development of offending and criminal careers, but there is not space to review these topics here.[4] For example, the prevalence of offending tends

to peak in the teenage years, and an early onset of offending predicts a long criminal career. Offenders tend to be versatile rather than specialized, in committing not only different types of offences but also different types of other antisocial acts. While there is considerable continuity over time, in the sense that the most antisocial people at one age tend also to be the most antisocial at another, only about half of antisocial juveniles tend to become antisocial adults.

Risk factors

A risk factor is defined as a variable that predicts an increased risk of offending. For example, children who experience poor parental supervision have an increased risk of committing offences later on. Since risk factors are defined by their ability to predict later offending, it follows that longitudinal data are required to discover them. Risk factors tend to be similar for many different outcomes, including violent and non-violent offending, mental health problems, alcohol and drug problems, school failure, and unemployment. Protective factors are also important. They are defined as factors that predict a low risk of offending or that counteract risk factors.

An obvious problem is that it is not clear to what extent any risk factor is a cause of offending. It is important to investigate causal mechanisms linking risk factors and offending. The best way of establishing a cause is to carry out a prevention experiment tackling that risk factor; preferably a randomized experiment, because the random assignment of people to conditions in principle controls for all other influences on offending. If a prevention experiment was carried out in which parental supervision was improved, and if offending was reduced as a consequence, this would be powerful evidence that the risk factor of parental supervision truly had a causal effect on offending. However, most knowledge about causes comes from quasi-experimental analyses.

Because of the difficulty of establishing causal effects of factors that vary only between individuals (e.g. gender and ethnicity), and because such factors have no practical implications for prevention (e.g. it is not practicable to change males into females), unchanging variables will not be reviewed here. In any case, their effects on offending are usually explained by reference to other, modifiable, factors. For example, gender differences in offending have been explained on the basis of different socialization methods used by parents with boys and girls, or different opportunities for offending by men and women. Similarly, risk factors that are or might be measuring the same underlying construct as delinquency (e.g. physical aggression) will not be reviewed; the focus is on risk factors that might be causes. For simplicity, risk factors are reviewed one by one. Biological factors are not reviewed. I will not attempt to review additive, interactive, independent, or sequential effects of risk factors, although these are important issues. Nor will I review developmental theories of offending.

Individual factors

Hyperactivity and impulsivity

Hyperactivity and impulsivity are among the most important personality or individual difference factors that predict later offending.[5] Hyperactivity usually begins before age 5 and often before age 2, and it tends to persist into adolescence. It is associated with restlessness, impulsivity and a short attention span, and for that reason has been termed the 'hyperactivity-impulsivity-attention deficit' or HIA syndrome. Related concepts include a poor ability to defer gratification and a short future time perspective.

Many investigators have reported a link between hyperactivity or impulsivity and offending. For example, in the Orebro (Sweden) longitudinal survey,[6] hyperactivity at age 13 (rated by teachers) predicted violent offending up to age 26. The highest rate of violence was among males with both motor restlessness and concentration difficulties. The most extensive research on different measures of impulsivity was carried out by Jennifer White and her colleagues in the Pittsburgh Youth Study. This showed that cognitive or verbal impulsivity (e.g. acts without thinking, unable to defer gratification) was more strongly related to delinquency than was behavioural impulsivity (e.g. clumsiness in psychomotor tests).

In the Cambridge Study, a combined measure of hyperactivity-impulsivity-attention deficit was developed at age 8–10, and it significantly predicted juvenile convictions independently of conduct problems at age 8–10. Hence, HIA is not merely another measure of antisocial personality, but it is a possible cause, or an earlier stage in a developmental sequence leading to offending. Similar constructs to hyperactivity, such as sensation seeking, are also related to delinquency. In the Cambridge Study, the extent to which the boy was daring or took risks at age 8–10, as well as restlessness and poor concentration, significantly predicted convictions and high self-reported offending. Daring was consistently one of the strongest independent predictors of offending.

Low intelligence and attainment

Low intelligence is an important predictor of offending, and it can be measured very early in life. In a prospective longitudinal survey of about 120 Stockholm males,[7] low IQ measured at age 3, significantly predicted officially recorded offending up to age 30. Frequent offenders (with 4 or more offences) had an average IQ of 88 at age 3, whereas non-offenders had an average IQ of 101. All of these results held up after controlling for social class. Similarly, low IQ at age 4 predicted arrests up to age 27 in the Perry preschool project.[8]

In the Cambridge Study, twice as many of the boys scoring 90 or less on a non-verbal IQ test (Raven's Progressive Matrices) at age 8–10 were convicted as juveniles as of the remainder. However, it was difficult to disentangle low intelligence and low school attainment. Low non-verbal intelligence was highly correlated with low verbal intelligence (vocabulary, word comprehension, verbal reasoning) and with low school attainment, and all of these measures predicted juvenile convictions to much the same extent. In addition to their poor school performance, delinquents tended to leave school at the earliest possible age (which was then 15) and to take no school examinations.

Low non-verbal intelligence predicted juvenile self-reported offending to almost exactly the same degree as juvenile convictions, suggesting that the link between low intelligence and delinquency was not caused by the less intelligent boys having a greater probability of being caught. Also, measures of intelligence and attainment predicted measures of offending independently of other variables such as family income and family size. Delinquents often do better on non-verbal performance tests, such as object assembly

and block design, than on verbal tests, suggesting that they find it easier to deal with concrete objects than with abstract concepts.

Low IQ may lead to delinquency through the intervening factor of school failure; the association between school failure and delinquency has been demonstrated consistently in longitudinal surveys. In the Pittsburgh Youth Study, Donald Lynam and his colleagues concluded that low verbal IQ led to school failure and subsequently to self-reported delinquency, but only for African-American boys. Another plausible explanatory factor underlying the link between low IQ and delinquency is the ability to manipulate abstract concepts. Children who are poor at this tend to do badly in IQ tests and in school attainment and they also tend to commit offences, mainly because of their poor ability to foresee the consequences of their offending and to appreciate the feelings of victims. Low IQ may be one aspect of cognitive and neuropsychological deficits in the executive functions of the brain.

Family factors

Child rearing

Many different types of child-rearing methods predict offending. The most important dimensions of child rearing are supervision or monitoring of children, discipline or parental reinforcement, warmth or coldness of emotional relationships, and parental involvement with children. Parental supervision refers to the degree of monitoring by parents of the child's activities, and their degree of watchfulness or vigilance. Of all these child-rearing methods, poor parental supervision is usually the strongest and most replicable predictor of offending. Many studies show that parents who do not know where their children are when they are out, and parents who let their children roam the streets unsupervised from an early age, tend to have delinquent children. For example, in Joan McCord's classic Cambridge–Somerville Study in Boston,[9] poor parental supervision in childhood was the best predictor of both violent and property crimes up to age 45.

Parental discipline refers to how parents react to a child's behaviour. It is clear that harsh or punitive discipline (involving physical punishment) predicts offending. In their follow-up study of nearly 700 Nottingham children, John and Elizabeth Newson[10] found that physical punishment at ages 7 and 11 predicted later convictions; 40 per cent of offenders had been smacked or beaten at age 11, compared with 14 per cent of non-offenders. Erratic or inconsistent discipline also predicts delinquency. This can involve either erratic discipline by one parent, sometimes turning a blind eye to bad behaviour and sometimes punishing it severely, or inconsistency between two parents, with one parent being tolerant or indulgent and the other being harshly punitive.

Cold, rejecting parents tend to have delinquent children, as Joan McCord found in the Cambridge–Somerville Study. More recently, she concluded that parental warmth could act as a protective factor against the effects of physical punishment. Whereas 51 per cent of boys with cold physically punishing mothers were convicted in her study, only 21 per cent of boys with warm physically punishing mothers were convicted, similar to the 23 per cent of boys with warm non-punitive mothers who were convicted. The father's warmth was also a protective factor against the father's physical punishment.

Most explanations of the link between child-rearing methods and delinquency focus on attachment or social learning theories. Attachment theory was inspired by the work of John Bowlby, and suggests that children who are not emotionally attached to warm, loving, and law-abiding parents tend to become offenders. Social learning theories suggest that children's behaviour depends on parental rewards and punishments and on the models of behaviour that parents represent. Children will tend to become offenders if parents do not respond consistently and contingently to their antisocial behaviour and if parents themselves behave in an antisocial manner.

Teenage mothers and child abuse

At least in Western industrialized countries, early child-bearing, or teenage pregnancy, predicts many undesirable outcomes for the children, including low school attainment, antisocial school behaviour, substance use, and early sexual intercourse. The children of teenage mothers are also more likely to become offenders. For example, Morash and Rucker[11] analysed results from four surveys in the United States and the United Kingdom (including the Cambridge Study) and found that teenage mothers were associated with low income families, welfare support, and absent biological fathers, that they used poor child-rearing methods, and that their children were characterized by low school attainment and delinquency. However, the presence of the biological father mitigated many of these adverse factors and generally seemed to have a protective effect. In the Cambridge Study, teenage mothers who went on to have large numbers of children were especially likely to have convicted children. In the Newcastle Thousand-Family Study[12] mothers who married as teenagers (a factor strongly related to teenage childbearing) were twice as likely as others to have sons who became offenders by age 32.

There is considerable intergenerational transmission of aggressive and violent behaviour from parents to children, as Maxfield and Widom[13] found in a retrospective study of over 900 abused children in Indianapolis. Children who were physically abused up to age 11 were significantly likely to become violent offenders in the next 15 years. In the Cambridge–Somerville Study in Boston, Joan McCord found that about half of the abused or neglected boys were convicted for serious crimes, became alcoholics or mentally ill, or died before age 35. In the Rochester Youth Development Study,[14] child maltreatment under age 12 (physical, sexual or emotional abuse or neglect) predicted later self-reported and official offending. Furthermore, these results held up after controlling for gender, race, socio-economic status, and family structure.

Numerous theories have been put forward to explain the link between child abuse and later offending. Timothy Brezina described three of the main ones.[15] Social learning theory suggests that children learn to adopt the abusive behaviour patterns of their parents through imitation, modelling, and reinforcement. Attachment or social bonding theory proposes that child maltreatment results in low attachment to parents and hence to low self-control. Strain theory posits that negative treatment by others generates negative emotions such as anger and frustration, which in turn lead to a desire for revenge and increased aggression. Based on analyses of the Youth in Transition Study, Brezina found limited support for all three theories.

Parental conflict and disrupted families

Many studies show that broken homes or disrupted families predict offending. In the Newcastle Thousand-Family Study, marital disruption (divorce or separation) in a boy's first 5 years doubled his risk of later convictions up to age 32. Similarly, in the Dunedin Study in New Zealand,[16] children who were exposed

to parental discord and many changes of the primary caretaker tended to become antisocial and delinquent. The same study showed that single parent families disproportionally tended to have convicted sons; 28 per cent of violent offenders were from single parent families, compared with 17 per cent of non-violent offenders and 9 per cent of unconvicted boys.

The importance of the cause of the broken home is shown in the UK National Survey of Health and Development.[17] Boys from homes broken by divorce or separation had an increased likelihood of being convicted or officially cautioned up to age 21, in comparison with those from homes broken by death or from unbroken homes. Homes broken while the boy was under age 5 especially predicted offending, whereas homes broken while the boy was between ages 11 and 15 were not particularly criminogenic. Remarriage (which happened more often after divorce or separation than after death) was also associated with an increased risk of offending, suggesting a possible negative effect of step-parents. The meta-analysis by Wells and Rankin[18] also shows that broken homes are more strongly related to delinquency when they are caused by parental separation or divorce rather than by death.

Most studies of broken homes have focussed on the loss of the father rather than the mother, simply because the loss of a father is much more common. Joan McCord in Boston carried out an interesting study of the relationship between homes broken by loss of the natural father and later serious offending of the children. She found that the prevalence of offending was high for boys reared in broken homes without affectionate mothers (62 per cent) and for those reared in united homes characterized by parental conflict (52 per cent), irrespective of whether they had affectionate mothers. The prevalence of offending was low for those reared in united homes without conflict (26 per cent) and—importantly—equally low for boys from broken homes with affectionate mothers (22 per cent). These results suggest that it is not so much the broken home which is criminogenic as the parental conflict which often causes it, and that a loving mother might in some sense be able to compensate for the loss of a father.

In the Cambridge Study, both permanent and temporary separations from a biological parent before age 10 (usually from the father) predicted convictions and self-reported delinquency, providing that they were not caused by death or hospitalization. However, homes broken at an early age (under age 5) were not unusually criminogenic. Separation before age 10 predicted both juvenile and adult convictions, independently of all other factors such as low family income or poor school attainment, and was an important predictor of adult social dysfunction.

Explanations of the relationship between disrupted families and delinquency fall into three major classes. Trauma theories suggest that the loss of a parent has a damaging effect on a child, most commonly because of the effect on attachment to the parent. Life course theories focus on separation as a sequence of stressful experiences, and on the effects of multiple stressors such as parental conflict, parental loss, reduced economic circumstances, changes in parent figures, and poor child-rearing methods. Selection theories argue that disrupted families produce delinquent children because of pre-existing differences from other families in risk factors such as parental conflict, criminal or antisocial parents, low family income, or poor child-rearing methods.

Hypotheses derived from the three theories were tested in the Cambridge Study.[19] While boys from broken homes (permanently disrupted families) were more delinquent than boys from intact homes, they were not more delinquent than boys from intact high conflict families. Overall, the most important factor was the post-disruption trajectory. Boys who remained with their mother after the separation had the same delinquency rate as boys from intact low conflict families. Boys who stayed with their father, with relatives, or with others (e.g. foster parents) had high delinquency rates. These living arrangements were more unstable, and other research shows that frequent changes of parent figures predict offending. It was concluded that the results favoured life course theories rather than trauma or selection theories

Criminal parents

Lee Robins and her colleagues showed that criminal, antisocial and alcoholic parents tend to have delinquent sons.[20] She followed up over 200 males in St. Louis and found that arrested parents tended to have arrested children, and that the juvenile records of the parents and children had similar rates and types of offences. Joan McCord also reported that convicted fathers tended to have convicted sons. She found that 29 per cent of fathers convicted for violence had sons convicted for violence, in comparison with 12 per cent of other fathers, but this may reflect the general tendency for convicted fathers to have convicted sons rather than any specific tendency for violent fathers to have violent sons.

In the Cambridge Study, the concentration of offending in a small number of families was remarkable. Less than 6 per cent of the families were responsible for half of the criminal convictions of all members (fathers, mothers, sons, and daughters) of all 400 families. Having a convicted mother, father, brother, or sister significantly predicted a boy's own conviction. As many as 63 per cent of boys with a convicted parent were themselves convicted up to age 40. Furthermore, convicted parents and delinquent siblings predicted self-reported as well as to official offending. Same-sex relationships were stronger than opposite-sex relationships, and older siblings were stronger predictors than younger siblings. Therefore, there is intergenerational continuity in offending.

It is not entirely clear why criminal parents tend to have delinquent children. In the Cambridge Study, there was no evidence that criminal parents directly encouraged their children to commit crimes or taught them criminal techniques. On the contrary, criminal parents were highly critical of their children's offending; for example, 89 per cent of convicted men at age 32 disagreed with the statement that 'I would not mind if my son/daughter committed a criminal offence'. Also, it was extremely rare for a parent and a child to be convicted for an offence committed together. The main link in the chain between criminal parents and delinquent sons seemed to be poor parental supervision.

There are several possible explanations (which are not mutually exclusive) for why offending tends to be concentrated in certain families and transmitted from one generation to the next. First, there may be intergenerational continuities in exposure to multiple risk factors. For example, each successive generation may be entrapped in poverty, disrupted families, single and/or teenage parenting, and living in the most deprived neighbourhoods. Second, the effect of a criminal parent on a child's offending may be mediated by environmental mechanisms such as poor parental supervision. Third, the effect of a criminal parent on a child's offending may be mediated by genetic mechanisms. Fourth, criminal parents may tend to have delinquent children because of official (police and court) bias against criminal families, who also tend to be known to official agencies because of other social

problems. At all levels of self-reported delinquency in the Cambridge Study, boys with convicted fathers were more likely to be convicted themselves than were boys with unconvicted fathers. However, this was not the only explanation for the link between criminal fathers and delinquent sons, because boys with criminal fathers had higher self-reported delinquency scores and higher teacher and peer ratings of bad behaviour.

Large family size

Large family size (a large number of children in the family) is a relatively strong and highly replicable predictor of offending. It was similarly important in the Cambridge and Pittsburgh studies,[21] even though families were on average smaller in Pittsburgh in the 1990s than in London in the 1960s. In the Cambridge Study, if a boy had four or more siblings by his tenth birthday, this doubled his risk of being convicted as a juvenile, and large family size predicted self-reported offending as well as convictions. It was the most important independent predictor of convictions up to age 32 in a logistic regression analysis.

In the National Survey of Health and Development, Michael Wadsworth found that the percentage of boys who were convicted increased from 9 per cent for families containing one child to 24 per cent for families containing four or more children. The Newsons in their Nottingham study also concluded that large family size was one of the most important predictors of offending. A similar link between family size and antisocial behaviour was reported by Israel Kolvin and his colleagues in their follow-up of Newcastle children from birth to age 33.

There are many possible reasons why a large number of siblings might increase the risk of a child's offending. Generally, as the number of children in a family increases, the amount of parental attention that can be given to each child decreases. Also, as the number of children increases, the household tends to become more overcrowded, possibly leading to increases in frustration, irritation, and conflict. In the Cambridge Study, large family size did not predict delinquency for boys living in the least crowded conditions. This suggests that household overcrowding might be an important intervening factor between large family size and delinquency.

Brownfield and Sorenson[22] reviewed several possible explanations for the link between large families and delinquency, including those focussing on features of the parents (e.g. criminal parents, teenage parents), those focussing on parenting (e.g. poor supervision, disrupted families), and those focussing on economic deprivation or family stress. Another interesting theory suggested that the key factor was birth order: large families include more later-born children, who tend to be more delinquent. Based on an analysis of self-reported delinquency in a Seattle survey, they concluded that the most plausible intervening causal mechanism was exposure to delinquent siblings. In the Cambridge Study, co-offending by brothers was surprisingly common; about 20 per cent of boys who had brothers close to them in age, were convicted for a crime committed with their brother.

Social factors

Socio-economic deprivation

The voluminous literature on the relationship between socio-economic status (SES) and offending is characterized by inconsistencies and contradictions, and some reviewers[23] have concluded that there is no relationship between SES and either self-reported or official offending. British studies have reported more consistent links between low social class and offending. In the UK National Survey of Health and Development, the prevalence of official juvenile delinquency in males varied considerably according to the occupational prestige and educational background of their parents, from 3 per cent in the highest category to 19 per cent in the lowest. It has been suggested that low SES families tend to produce delinquent children because their child-rearing tends to be poor.

Numerous indicators of SES were measured in the Cambridge Study, both for the boy's family of origin and for the boy himself as an adult, including occupational prestige, family income, housing, and employment instability. Most of the measures of occupational prestige (based on the Registrar General's scale) were not significantly related to offending. Low SES of the family when the boy was aged 8–10 significantly predicted his later self-reported but not his official delinquency. More consistently, low family income and poor housing predicted official and self-reported, juvenile and adult, offending.

It was interesting that the peak age of offending, at 17–18, coincided with the peak age of affluence for many convicted males. In the Cambridge Study, convicted males tended to come from low income families at age 8 and later tended to have low incomes themselves at age 32. However, at age 18, they were relatively well paid in comparison with non-delinquents. Whereas convicted delinquents might be working as unskilled labourers on building sites and getting the full adult wage for this job, non-delinquents might be in poorly paid jobs with prospects, such as bank clerks, or might still be students. These results show that the link between income and offending is quite complex.

Socio-economic deprivation of parents is usually compared with offending by children. However, when the children grow up, their own socio-economic deprivation can be related to their own offending. In the Cambridge Study, an unstable job record of the boy at age 18 was one of the best independent predictors of his later convictions between ages 21 and 25. Also, having an unskilled manual job at age 18 was an important independent predictor of adult social dysfunction and antisocial personality at age 32.

Between ages 15 and 18, the study boys were convicted at a higher rate when they were unemployed than when they were employed, suggesting that unemployment in some way causes crime, and conversely that employment may lead to desistance from offending.[24] Since crimes involving material gain (e.g. theft, burglary, robbery) especially increased during periods of unemployment, it seems likely that financial need is an important link in the causal chain between unemployment and crime.

Peer influences

Having delinquent friends is an important predictor of later offending; peer delinquency and gang membership predicted self-reported violence in the Seattle Social Development Project.[25] Delinquent acts tend to be committed in small groups (of two or three people, usually) rather than alone. Large gangs are comparatively unusual. In the Cambridge Study, the probability of committing offences with others decreased steadily with age. Before age 17, boys tended to commit their crimes with other boys similar in age and living close by. After age 17, co-offending became less common.[26]

The major problem of interpretation is whether young people are more likely to commit offences while they are in groups than

while they are alone, or whether the high prevalence of co-offending merely reflects the fact that, whenever young people go out, they tend to go out in groups. Do peers tend to encourage and facilitate offending, or is it just that most kinds of activities out of the home (both delinquent and non-delinquent) tend to be committed in groups? Another possibility is that the commission of offences encourages association with other delinquents, perhaps because 'birds of a feather flock together' or because of the stigmatizing and isolating effects of court appearances and institutionalization. Terence Thornberry and his colleagues in the Rochester Youth Development Study concluded that there were reciprocal effects, with delinquent peers causing delinquency and delinquency causing association with delinquent peers.

In the Pittsburgh Youth Study, the relationship between peer delinquency and a boy's offending was studied both between individuals (e.g. comparing peer delinquency and offending of boy X with peer delinquency and offending of boy Y at a particular age and then aggregating these correlations over all ages) and within individuals (e.g. comparing peer delinquency and offending of boy X at different ages and then aggregating these correlations over all individuals). Peer delinquency was the strongest correlate of offending in between-individual correlations but did not predict offending within individuals.[27] In contrast, poor parental supervision, low parental reinforcement, and low involvement of the boy in family activities predicted offending both between and within individuals. It was concluded that these three family variables were the most likely to be causes, whereas having delinquent peers was most likely to be an indicator of the boy's offending.

Associating with delinquent friends at age 14 was an important independent predictor of convictions at the young adult ages in the Cambridge Study. Also, the recidivists at age 19 who ceased offending differed from those who persisted, in that the desisters were more likely to have stopped going round in a group of male friends. Furthermore, spontaneous comments by the youths indicated that withdrawal from the delinquent peer group was an important influence on ceasing to offend. Therefore, continuing to associate with delinquent friends may be a key factor in determining whether juvenile delinquents persist in offending as young adults or desist.

School influences

The prevalence of delinquency among students varies dramatically between different secondary schools, as Michael Power and his colleagues[28] showed many years ago in London. Characteristics of high delinquency-rate schools are well-known. For example, such schools have high levels of distrust between teachers and students, low commitment to school by students, and unclear and inconsistently enforced rules. However, what is far less clear is how much of the variation between schools should be attributed to differences in school organization, climate and practices, and how much to differences in the composition of the student body.

In the Cambridge Study, attending a high delinquency-rate school at age 11 significantly predicted a boy's later offending and antisocial personality scores. The effects of secondary schools on delinquency were investigated by following boys from their primary schools to their secondary schools. The best primary school predictor of juvenile delinquency was the rating of the boy's troublesomeness at age 8–10 by peers and teachers, showing the continuity in antisocial behaviour. The secondary schools differed dramatically in their official delinquency rates, from one school with 21 court appearances per 100 boys per year to another where the corresponding figure was only 0.3. Moreover, going to a high delinquency-rate secondary school was a significant predictor of later convictions. It was, however, very noticeable that the most troublesome boys tended to go to the high delinquency-rate schools, while the least troublesome boys tended to go to the low delinquency-rate schools. Most of the variation between schools in their delinquency rates could be explained by differences in their intakes of troublesome boys. The secondary schools themselves had only a very small effect on the boys' offending.

The most famous study of school effects on delinquency was also carried out in London, by Michael Rutter and his colleagues.[29] They studied 12 comprehensive schools, and again found big differences in official delinquency rates between them. High delinquency-rate schools tended to have high truancy rates, low ability pupils, and low social class parents. However, the differences between the schools in delinquency rates could not be entirely explained by differences in the social class and verbal reasoning scores of the pupils at intake (age 11). Therefore, they must have been caused by some aspect of the schools themselves or by other, unmeasured factors.

In trying to discover which aspects of schools might be encouraging or inhibiting offending, Rutter and his colleagues found that the main school factors that were associated with delinquency were a high amount of punishment and a low amount of praise given by teachers in class. Unfortunately, it is difficult to know whether much punishment and little praise are causes or consequences of antisocial school behaviour, which in turn may be linked to offending outside school. In regard to other outcome measures, they argued that an academic emphasis, good classroom management, the careful use of praise and punishment, and student participation were important features of successful schools.

Community influences

Offending rates vary systematically with area of residence. For example, the classic studies by Shaw and McKay[30] in Chicago and other American cities showed that juvenile delinquency rates (based on where offenders lived) were highest in inner city areas characterized by physical deterioration, neighbourhood disorganization, and high residential mobility. A large proportion of all offenders came from a small proportion of areas, which tended to be the most deprived. Furthermore, these relatively high delinquency rates persisted over time, despite the effect of successive waves of immigration and emigration of different national and ethnic groups in different areas.

Living in a deprived neighbourhood (whether based on parent ratings or on census measures of poverty, unemployment and female-headed households) significantly predicts convictions and self-reported offending. However, it is difficult to establish how much the areas themselves influence antisocial behaviour and how much it is merely the case that antisocial people tend to live in deprived areas (e.g. because of their poverty or public housing allocation policies). It has been suggested that neighbourhoods have only indirect effects on offending because of their effects on individuals and families. However, Robert Sampson and his colleagues[31] argued that a low degree of 'collective efficacy' of a neighbourhood (a low degree of informal social control) caused high crime rates.

One key question is why crime rates of communities change over time, and to what extent this is a function of changes in the

communities or in the individuals living in them. Answering this question requires longitudinal research in which both communities and individuals are followed up. The best way of establishing the impact of the environment is to follow people who move from one area to another. For example, in the Cambridge Study, moving out of London led to a significant decrease in convictions and self-reported offending. This decrease may have occurred because moving out led to a breaking up of co-offending groups, or because there were fewer opportunities for crime outside London.

Situational influences

It might be argued that all the risk factors reviewed so far in this section—individual, family, socio-economic, peer, school, and community—essentially influence the development of a long-term individual potential for offending. In other words, they contribute to between-individual differences: why some people are more likely than others, given the same situational opportunity, to commit a crime. Another set of influences—situational factors—explain how the potential for violence becomes the actuality in any given situation. Essentially, they explain short-term within-individual differences: why a person is more likely to commit crimes in some situations than in others. Situational factors may be specific to particular types of crimes: robberies as opposed to rapes, or even street robberies as opposed to bank robberies.

The most popular theory of offending events is a rational choice theory suggesting that they occur in response to specific opportunities, when their expected benefits (e.g. stolen property, peer approval) outweigh their expected costs (e.g. legal punishment, parental disapproval). For example, Clarke and Cornish[32] suggested that residential burglary depended on such influencing factors as whether a house was occupied, whether it looked affluent, whether there were bushes to hide behind, whether there were nosy neighbours, whether the house had a burglar alarm, and whether it contained a dog. A related theory is the 'routine activities' idea of Cohen and Felson.[33] They suggested that, for a predatory crime to occur, the minimum requirement was the convergence in time and place of a motivated offender and a suitable target, in the absence of a capable guardian. They argued that predatory crime rates were influenced by routine activities that satisfied basic needs such as food and shelter. Changes in routine activities led to changing opportunities for crime. For example, the increasing number of working women meant that more homes were left unattended during the day.

Much work on describing situations leading to violence has been carried out in the United Kingdom under the heading of crime analysis. This begins with a detailed analysis of patterns and circumstances of crimes and then proceeds to devise, implement, and evaluate crime reduction strategies. For example, it was found that most street robberies in London occurred in predominantly ethnic minority areas, and most offenders were 16–19 year old Afro-Caribbean males.[34] The victims were mostly Caucasian females, alone, and on foot. Most offences occurred at night, near the victim's home. The main motive for robbery was to get money, and the main factor in choosing victims was whether they had a wealthy appearance.

In their Montreal longitudinal study of delinquents, LeBlanc and Frechette[35] provided detailed information about motives and methods used in different offences at different ages. For example, for violence at age 17, the main motivation was utilitarian or rational. For all crimes, however, the primary motivation changed from hedonistic (searching for excitement, with co-offenders) in the teenage years to utilitarian (with planning, psychological intimidation, and use of instruments such as weapons) in the twenties. In the Cambridge Study, motives for physical fights depended on whether the boy fought alone or with others.[36] In individual fights, the boy was usually provoked, became angry, and hit out to hurt his opponent and to discharge his own internal feelings of tension. In group fights, the boy often said that he became involved to help a friend or because he was attacked, and rarely said that he was angry. The group fights were more serious, occurring in bars or streets, and they were more likely to involve weapons, produce injuries, and lead to police intervention. Fights often occurred when minor incidents escalated, because both sides wanted to demonstrate their toughness and masculinity and were unwilling to react in a conciliatory way.

Much is known about the situations in which violence occurs. For example, in Sweden, violence preceded by situational arguments typically occurs in streets or restaurants, while violence preceded by relationship arguments typically occurs in homes.[37] In England, stranger assaults typically occur in streets, bars, or discotheques, non-stranger assaults typically occur at home or work, and robberies typically occur in the street or on public transport. Most violence occurs on weekend nights around pubs and clubs, and involves young males who have been drinking.[38] More research on situational influences on offending needs to be incorporated in prospective longitudinal studies, in order to link up the developmental and situational perspectives.

Conclusions

Offending is one element of a larger syndrome of antisocial behaviour that arises in childhood and tends to persist into adulthood, with numerous different behavioural manifestations. However, while there is continuity over time in antisocial behaviour, changes are also occurring. It is commonly found that about half of a sample of antisocial children go on to become antisocial teenagers, and about half of antisocial teenagers go on to become antisocial adults. More research is needed on factors that predict these changes over time. Research is especially needed on changing behavioural manifestations and developmental sequences at different ages. More efforts should especially be made to identify factors that protect vulnerable children from developing into antisocial teenagers.

A great deal has been learned in the last 20 years, particularly from longitudinal surveys, about risk factors for offending and other types of antisocial behaviour. Offenders differ significantly from non-offenders in many respects, including impulsivity, intelligence, family background, and socio-economic deprivation. These differences are present before, during, and after criminal careers. While the precise causal chains that link these factors with antisocial behaviour, and the ways in which these factors have independent, interactive, or sequential effects, are not known, it is clear that individuals at risk can be identified with reasonable accuracy. In order to advance knowledge about human development and criminal careers, new multiple-cohort longitudinal studies are needed.[39]

The identified risk factors for offending should be targeted in prevention programmes. Risk-focussed prevention has great

potential for crime reduction.[40] The continuity of antisocial behaviour from childhood to adulthood suggests that prevention efforts should be implemented early in life. Because of the link between offending and numerous other social problems, any measure that succeeds in reducing offending will have benefits that go far beyond this. Any measure that reduces offending will probably also reduce alcohol abuse, drunk driving, drug abuse, sexual promiscuity, family violence, truancy, school failure, unemployment, marital disharmony, and divorce. It is clear that problem children tend to grow up into problem adults, and that problem adults tend to produce more problem children. Continued efforts are urgently needed to advance knowledge about offending and antisocial behaviour, and to tackle the roots of crime.

Further information

For extensive reviews of risk factors and interventions, see *Saving Children from a Life of Crime: Early Risk Factors and Effective Interventions* by D. P. Farrington and B. C. Welsh (Oxford University Press, 2007).

For extensive information about the Cambridge Study, see *Criminal Careers up to Age 50 and Life Success up to Age 48: New Findings from the Cambridge Study in Delinquent Development* by D. P. Farrington and colleagues (Home Office Research Study No. 299), available from www.homeoffice,gov.uk/rds.

For other reviews of psychosocial causes of offending, see *Antisocial Behaviour by Young People*, by M. Rutter, H. Giller, and A. Hagell (Cambridge University Press, 1998).

References

1. Rutter, M., Giller, H., and Hagell, A. (1998). *Antisocial behaviour by young people*. Cambridge University Press, Cambridge.
2. Farrington, D.P., Coid, J.W., Harnett, L.M., *et al.* (2006). *Criminal careers up to age 50 and life success up to age 48: new findings from the Cambridge Study in Delinquent Development*. Home Office, London (Research Study No. 299).
3. Loeber, R., Farrington, D.P., Stouthamer-Loeber, M., *et al.* (1998). The development of male offending: key findings from the first decade of the Pittsburgh Youth Study. *Studies on Crime and Crime Prevention*, **7**, 141–71.
4. Piquero, A.R., Farrington, D.P., and Blumstein, A. (2007). *Key issues in criminal career research: new analyses of the Cambridge Study in Delinquent Development*. Cambridge University Press, Cambridge.
5. Farrington, D.P. and Welsh, B.C. (2007). *Saving children from a life of crime: early risk factors and effective interventions*. Oxford University Press, Oxford.
6. Klinteberg, B.A., Andersson, T., Magnusson, D., *et al.* (1993). Hyperactive behaviour in childhood as related to subsequent alcohol problems and violent offending: a longitudinal study of male subjects *Individual Differences*, **15**, 381–8.
7. Stattin, H. and Klackenberg-Larsson, I. (1993). Early language and intelligence development and their relationship to future criminal behaviour. *Journal of Abnormal Psychology*, **102**, 369–78.
8. Schweinhart, L.J., Barnes, H.V., and Weikart, D.P. (1993). *Significant benefits: the High/Scope Perry preschool study through age 27*. High/Scope, Ypsilanti, Michigan.
9. McCord, J. (1979). Some child-rearing antecedents of criminal behaviour in adult men. *Journal of Personality and Social Psychology*, **37**, 1477–86.
10. Newson, J. and Newson, E. (1989). *The extent of parental physical punishment in the UK*. Approach, London.
11. Morash, M. and Rucker, L. (1989). An exploratory study of the connection of mother's age at childbearing to her children's delinquency in four data sets. *Crime and Delinquency*, **35**, 45–93.
12. Kolvin, I., Miller, F.J.W., Scott, D.M., *et al.* (1990). *Continuities of deprivation? The Newcastle 1000 family study*. Avebury, Aldershot.
13. Maxfield, M.G. and Widom, C.S. (1996). The cycle of violence revisited 6 years later. *Archives of Paediatrics and Adolescent Medicine*, **150**, 390–5.
14. Smith, C.A. and Thornberry, T.P. (1995). The relationship between childhood maltreatment and adolescent involvement in delinquency. *Criminology*, **33**, 351–481.
15. Brezina, T. (1998). Maltreatment and delinquency: the question of intervening processes. *Journal of Research in Crime and Delinquency*, **35**, 71–99.
16. Henry, B., Caspi, A., Moffitt, T.E., *et al.* (1996). Temperamental and familial predictors of violent and non-violent criminal convictions: age 3 to age 18. *Developmental Psychology*, **32**, 614–23.
17. Wadsworth, M.E.J. (1979). *Roots of delinquency: infancy, adolescence and crime*. Martin Robertson, London.
18. Wells, L.E. and Rankin, J.H. (1991). Families and delinquency: a meta-analysis of the impact of broken homes. *Social Problems*, **38**, 71–93.
19. Juby, H. and Farrington, D.P. (2001). Disentangling the link between disrupted families and delinquency. *British Journal of Criminology*, **41**, 22–40.
20. Robins, L.N., West, P.J., and Herjanic, B.L. (1975). Arrests and delinquency in two generations: a study of black urban families and their children. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, **16**, 125–40.
21. Farrington, D.P. and Loeber, R. (1999). Transatlantic replicability of risk factors in the development of delinquency. In *Historical and geographical influences on psychopathology* (eds. P. Cohen, C. Slomkowski, and L.N. Robins), pp. 299–329. Lawrence Erlbaum, Mahwah, NJ.
22. Brownfield, D. and Sorenson, A.M. (1994). Sibship size and sibling delinquency. *Deviant Behaviour*, **15**, 45–61.
23. Hindelang, M.J., Hirschi, T., and Weis, J.G. (1981). *Measuring delinquency*. Sage, Beverly Hills, California.
24. Farrington, D.P., Gallagher, B., Morley, L., *et al.* (1986). Unemployment, school leaving, and crime. *British Journal of Criminology*, **26**, 335–56.
25. Hawkins, J.D., Herrenkohl, T., Farrington, D.P., *et al.* (1998). A review of predictors of youth violence. In *Serious and violent juvenile offenders: risk factors and successful interventions* (eds. R. Loeber and D.P. Farrington), pp. 106–46. Sage, Thousand Oaks, California.
26. Reiss, A.J. and Farrington, D.P. (1991). Advancing knowledge about co-offending: results from a prospective longitudinal survey of London males. *Journal of Criminal Law and Criminology*, **82**, 360–95.
27. Farrington, D.P., Loeber, R., Yin, Y., *et al.* (2002). Are within-individual causes of delinquency the same as between-individual causes? *Criminal Behaviour and Mental Health*, **12**, 53–68.
28. Power, M.J., Alderson, M.R., Phillipson, C.M., *et al.* (1967). Delinquent schools? *New Society*, **10**, 542–3.
29. Rutter, M., Maughan, B., Mortimore, P., Ouston, J., and Smith, A. (1979). *Fifteen thousand hours: secondary schools and their effects on children*. Open Books, London.
30. Shaw, C.R. and McKay, H.D. (1969). *Juvenile delinquency and urban areas* (rev. edn). University of Chicago Press, Chicago.
31. Sampson, R.J., Raudenbush, S.W., and Earls, F. (1997). Neighbourhoods and violent crime: a multilevel study of collective efficacy. *Science*, **277**, 918–24.
32. Clarke, R.V. and Cornish, D.B. (1985). Modelling offenders' decisions: a framework for research and policy. In *Crime and justice*, Vol. 6, (eds. M. Tonry and N. Morris), pp. 147–85. University of Chicago Press, Chicago.

33. Cohen, L.E. and Felson, M. (1979). Social change and crime rate trends: a routine activity approach. *American Sociological Review*, **44**, 588–608.
34. Barker, M., Geraghty, J., Webb, B., *et al.* (1999). *The prevention of street robbery*. Home Office Police Department, London.
35. LeBlanc, M. and Frechette, M. (1989). *Male criminal activity from childhood through youth*. Springer-Verlag, New York.
36. Farrington, D.P. (1993). Motivations for conduct disorder and delinquency. *Development and Psychopathology*, **5**, 225–2241.
37. Wikström, P.-O.H. (1985). *Everyday violence in contemporary Sweden*. National Council for Crime Prevention, Stockholm.
38. Allen, J., Nicholas, S., Salisbury, H., *et al.* (2003). Nature of burglary, vehicle and violent crime. In *Crime in England and Wales 2001/2002: supplementary volume* (eds. C. Flood-Page and J. Taylor), pp. 41–68. Home Office, London (Statistical Bulletin 01/03).
39. Tonry, M., Ohlin, L.E., and Farrington, D.P. (1991). *Human development and criminal behaviour: new ways of advancing knowledge*. Springer-Verlag, New York.
40. Farrington, D.P. (2007). Childhood risk factors and risk-focussed prevention. In *The Oxford handbook of criminology* (4th edn) (eds. M. Maguire, R. Morgan, and R. Reiner), pp. 602–40. Oxford University Press, Oxford.

11.3

Associations between psychiatric disorder and offending

Contents

11.3.1 Associations between psychiatric disorder and offending

Lindsay Thomson and Rajan Darjee

Introduction

The associations between psychiatric disorder and offending are complex. There has been a great deal of research into certain disorders and violent offending particularly over the last two decades. In summary, this has found a clear and consistent association between schizophreniform psychoses and violence, the importance of premorbid antisocial behaviour in predicting future violence, and the adjunctive effect of co-morbid substance misuse and antisocial personality disorder in the prevalence of violence. In addition, it has allowed the development of neuropsychiatric models to begin to explain violence in the context of mental disorder. Substance use disorders and learning disability are discussed in Chapters 11.3.2 and 11.3.3.

Mental disorder and offending: a problematic relationship

Criminal behaviour is common in our society but there is evidence that violent crime rates have declined in Europe and North America over the last decade.[1] Mental disorders are also common. It is important to study the overlap between mental disorder and offending to consider those mental health and criminogenic factors that may be amenable to change. The social and economic factors relevant to offending are discussed in Chapter 11.2.

Before considering any associations between mental disorder and offending it is useful to consider the methodological problems in studying these:

- Offences are man-made concepts and not static. For example, many jurisdictions have created laws against stalking in the last twenty years which did not previously exist.
- Psychiatrists see a limited range of offenders but often base their research on these.
- Research is generally carried out on a captive population in prison or secure hospital. Offenders in these settings are likely to include those with characteristics that disadvantage them in the criminal justice system, for example ethnic status, low economic status, homelessness, unemployment, and mental illness.
- The generalizability of any findings must be queried given that offending is dependent on the wider social context such as rates of unemployment or crime, prevalence of substance misuse, and weapon carrying culture.
- It can be difficult to standardize populations studied, and severity of crimes.
- Criminal and mental health records may be unreliable.

Evidence for neurobiological determinants of offending or aggressive behaviour

There is resistance to any oversimplified idea of seeking a genetic or neuropsychological explanation to offending behaviour as a whole but research in this area is expanding slowly. Neuropsychological abnormalities are commonly found in offenders and there is evidence for specific brain deficits in aggressive or violent behaviour.[2] These findings include:

- An association between specific traumatic damage to the frontal lobes, particularly orbitofrontal injury, and poor impulse control and aggressive outbursts.
- Abnormalities on neuropsychological tests of frontal lobe function in aggressive and antisocial subjects, indicating prefrontal executive dysfunction.

- Abnormalities found in clinical neurological testing of offenders. Antisocial behaviour is associated with EEG abnormalities particularly frontal slowing and these are commonly found in more than half of prisoners with a history of repetitive violence. Clinical signs of frontal lobe dysfunction are also associated with recurrent aggression.
- Neuroimaging changes. Structural and functional studies examining patients in forensic services and patients with antisocial personality disorder have consistently found changes in the frontal lobes of aggressive patients, typically reduced prefrontal cortical size and activity. Predatory rather than impulsive, emotionally charged (affective) perpetrators of homicide show functional patterns of blood flow similar to controls suggesting that these neuroimaging findings are relevant to impulsive or affective aggression rather than premeditated, purposeful violence. Two groups of people with aggressive behaviour are postulated: the first with an acquired frontal lobe lesion due to injury or disease which impairs social judgement, risk avoidance, and empathy; the second, shows increased aggressive behaviour associated with deficits in executive functioning correlating with dorsolateral prefrontal dysfunction which may occur in foetal or birth brain injury, developmental learning disorders, attention deficit hyperactivity disorder, substance misuse, and antisocial personality disorder with episodic aggressive dyscontrol.
- In addition, there is evidence for biochemical abnormalities. Reduced serotonin function is largely related to impulsivity rather than directly to violence.[3] Serotonin has a role in emotional states such as impulsiveness, aggression, anxiety, and depression. This has provided potential avenues for treatment with selective serotonin reuptake inhibitor medication.[4] Cortisol abnormalities have been recognized for a long time and more recently dietary insufficiencies have been explored.[5]
- Lastly, the role of genetics in offending behaviour has been examined through family, twin, and adoption studies.[6] Adoption studies have found a consistent association between biological parents and adoptee for property offences but a more complex association for violent crime with a relationship discovered in the Danish birth cohort study between paternal violence and adoptee schizophrenia. A link between a genotype and disturbed behaviour was found in maltreated male children. Those with the gene encoding the neurotransmitter metabolizing enzyme monoamine oxidase A (MAOA) moderated the effect of maltreatment and had reduced antisocial behaviour in later life.[7] MAOA genes have a known association with aggression in animals and humans. The low expression variant leads to increased aggression, limbic volume reductions and hyper-responsive amygdala during emotional arousal with decreased reactivity of regulatory prefrontal regions.[8] Twin studies have shown that antisocial behaviour in early childhood particularly when associated with callous unemotional traits has a strong genetic influence which weakens in antisocial behaviour displayed initially in adolescence when environmental influences are important.[9]

Such changes are not a necessarily a cause for aggressive behaviour which is dependent on so many environmental factors, nor do they necessarily predict aggressive behaviour but they are important to study in that they indicate potential management strategies.

Clinical implications of the relationship between mental disorder and offending

The nature of the relationship between psychiatric disorder and offending is complex both at an individual and population level, and has important implications at both. Epidemiological data point to the following conclusions:

- Schizophrenia, personality disorder, and substance related disorders are significantly overrepresented in offenders.
- A number of factors associated with violence and offending in the non-mentally disordered are relevant to violence and offending in the mentally disordered.
- Amongst offenders there is the need to provide psychiatric assessment and treatment for significant numbers of people with mental disorder.
- Amongst mentally disordered offenders it is important to address criminogenic factors as well as providing traditional clinical treatment
- Appropriate treatment and preventative measures may prevent some offending and violence by individuals with mental disorders

It is important to know and understand the epidemiological research, but also to know and understand the patient that is being assessed. Not all factors of relevance to offending in people with mental disorders act in the same way in every case, and some factors not found to be of relevance in study samples may be crucial in individual cases. There is no simple or straightforward approach to such an assessment, and narrowing one's focus to a tick list of a few factors (perhaps from an actuarial tool) is at best lazy and at worst negligent. The psychiatrist should be able to articulate a formulation incorporating the factors which account for previous offending and which may be of relevance to future offending.

Every violent act involves a perpetrator, a victim, and a context, and this is no different where mental disorder is involved. A number of factors in the perpetrator, victim, and context in any violent incident are relevant to that violent act, and these three interact with each other. A simple example is that a drunk aggressive victim may frighten a suspicious impulsive perpetrator in the context of alcohol intoxication and an available knife. Where a person with mental disorder is violent one should not assume the former straightforwardly causes the latter. Sometimes mental disorder is the major determinant of an offence and at others mental disorder is entirely coincidental. In most cases mental disorder is one of a number of interacting factors (Fig. 11.3.1.1). Even where an offence seems explicable by psychotic hallucinations or delusions it is essential to consider wider factors (e.g. alcohol and substance misuse, social networks and personality).[10] A thorough assessment of these factors is crucial when evaluating criminal responsibility (see Chapter 11.1), assessing risk of future offending (see Chapter 11.15) and planning treatment for patients who have been violent. Treating mental disorder without addressing these factors will not adequately address the risk of further offending.

Scizophrenia and offending

For over a decade, studies have consistently shown a small but significant association between schizophrenia and violence with an

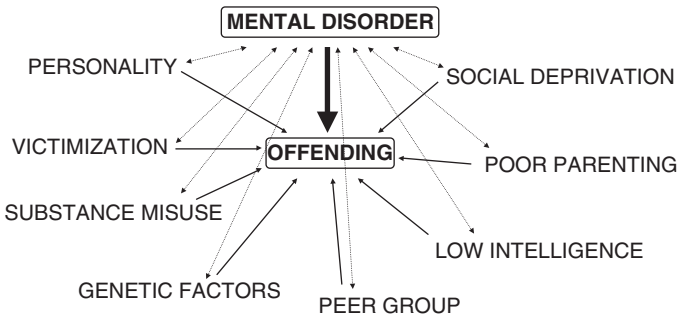


Fig. 11.3.1.1 A schematic representation of the complex interplay between factors of relevance to offending in people with mental disorder.

increased risk of violence of between 2–4 for men and 6–8 for women after controlling for marital status, socio-economic background and substance abuse.[11] The proportion of violent crimes in the population of Northern European countries attributable to individuals with severe mental illness is around 5 per cent.[12] A systematic review of mental disorder in 23,000 prisoners found psychoses in 3.7 per cent of men and 4 per cent of women which was 2–4 times more than in the general population. Aggression in first episode psychosis occurs consistently in one-third of cases.[13]

Two patterns of aggression associated with schizophrenia are seen: firstly, those in whom features of conduct disorder (20-40 per cent) and later antisocial personality disorder predate the onset of overt schizophrenia; and secondly, those in whom violence occurs at a later stage.[11] Psychotic symptoms and violence in childhood are strongly related to later violence, as is a history of conduct disorder.[14]

Studies of people convicted of homicide have consistently shown an excess of schizophrenia with rates of 5–15 per cent.[15,16] Family members are significantly more likely to be the victim than strangers. Those with comorbid antisocial personality disorder are less likely to be actively psychotic at the time of the offence, and more likely to be intoxicated and to kill non-relatives.

Neurobiological correlates of schizophrenia and violence

Naudts and Hodgins[17] examined neurobiological correlates of schizophrenia and violence. The literature is small, as are the sample sizes, and the measures used are diverse. At least one-fifth of men with schizophrenia show antisocial behaviour from childhood onwards. Overall, these men perform better on tests of specific executive functions and verbal skills despite greater impulsivity; and worse on tests of orbitofrontal functions than men with schizophrenia alone. Structural brain imaging studies of men with schizophrenia and a history of repetitive violence found reduced whole brain and hippocampal volumes; impaired connectivity between the orbitofrontal cortex and the amygdala; and structural abnormalities of the amygdala. Functional neuroimaging studies found reduced prefrontal cerebral blood flow during completion of a test of executive function in violent patients. This reduction in blood flow may result in loss of inhibition and therefore aggression, or it may reflect that these violent men found the test easier and did not need increased blood flow to complete it. Studies of acquired brain lesions show that ventromedial orbitofrontal cortex is necessary for inhibiting impulsive decision-making and behaviour and for physiological anticipation of secondary inducers such as punishment. Studies at different ages suggest that an intact amygdala in early life is necessary for the normal development of this orbitofrontal system to recognize and process emotions. This fits with the work of Silver *et al.*[18] who found that patients with schizophrenia with a history of severe violence differ from non-violent patients with schizophrenia in their perception of the intensity of emotions but not cognitive function. Failure to assess the intensity of emotions may contribute to conflict generation, failure to recognize resolution signals, conflict escalation, and violence.

Clinical features

People with schizophrenia may be violent because of their psychotic experiences, impaired judgment and impulse control, or situational factors. The evidence of the effect of symptoms is conflicting however, probably because of the way questions are asked; failure to take account of affective symptoms; variation in the timing between symptoms and violence; failure to control for medication, compulsion and previous violence; and variations in statistical procedures.[19] Swanson *et al.*[20] in a study of 1410 people with schizophrenia in a six month period found 19.1 per cent had been violent and this was serious in 3.6 per cent of cases. Positive symptoms increased the risk of minor and serious violence whereas negative symptoms decreased the risk of serious violence perhaps because these individuals lived alone. Serious violence was associated with psychotic and depressive symptoms, childhood conduct problems, and victimization. Severe psychotic symptoms and threat-control override (TCO) were antecedents of violent behaviour of patients in the community even after controlling for psychopathy and substance abuse.[21] TCO consists of persecutory delusions, passivity phenomena and thought insertion, resulting in perceived personal threat combined with loss of self control. Excessive perceptions of threat explained violence in people with schizophrenia-spectrum disorder alongside a history of conduct disorder.[22]

Hallucinations, acute suicidal ideation, acute conflict and stressors such as separations or housing problems, and lack of insight have all been associated with an increased risk of violence.[23]

Co-morbidity

The likelihood of violence in patients with schizophrenia increases in the presence of comorbid substance use (3-fold in men, 16-fold in women) or personality disorder (4-fold men, 18-fold women).[11] Swanson *et al.*[20] found however, that the effects of substance abuse were non significant after controlling for age, positive symptoms of schizophrenia, childhood conduct problems and recent victimization thereby suggesting that the effects of substance abuse on violence may be mediated by these factors. There is a strong association between comorbid antisocial personality disorder and substance abuse in patients with schizophrenia with common origins in conduct disorder.[24] Further evidence for this is found in neuropsychological tests. Patients with comorbid schizophrenia and substance abuse perform as well as those with schizophrenia alone on neurological tests although it is known that the brains of men with schizophrenia are particularly sensitive to effects of drugs and alcohol. It has been suggested that these men belong to the antisocial group who do better on these tests at the onset of their schizophrenia.[17]

Management

The care of patients with schizophrenia and a history of serious violence is centred around ongoing assessment and management of factors contributing to risk of violence and treatment of the schizophrenic illness. Predictors of violence include recent assault, a previous violent conviction, lower educational attainment and attending special education, a personal and family history of alcohol abuse, and lower but normal range IQ.[25,26] The management of schizophrenia is discussed in Chapter 4.3.8 but in patients with a history of violence methods of legal compulsion for detention and treatment, and programmes to address comorbid and criminogenic needs are particularly important. There is evidence that cognitive behavioural therapy can modify delusional beliefs that have lead to violence[27] and that clozapine has a particular role in the treatment of aggression in patients with schizophrenia separate from its antipsychotic or sedative components.[28]

Outcome

There is no indication that deinstitutionalization has lead to increased offending by patients with schizophrenia. Rates of violence have increased but in proportion to the increase in society as a whole and with increased comorbid alcohol abuse.[29] Outcome studies of patients transferred from high security found a recidivism rate of 34 per cent and 31 per cent in England and Scotland respectively, and a violent recidivism rate of 15 per cent and 19 per cent after ten years.[30, 31] Patients with a primary diagnosis of schizophrenia during a follow up period of 8 years had a recidivism rate of 15 per cent and a violent recidivism rate of 5 per cent.

Delusional disorders

Delusional disorders are described in Chapter 4.4. There are a number of subtypes including somatic, grandiose, mixed and unspecified but it is particularly the jealous, erotomanic and persecutory forms that are associated with offending and violence. Individuals with delusional disorders remain organized allowing them to target and plan any violence more effectively. Sixty percent of people with delusional jealousy are violent towards their partners. In erotomania the patient has the delusional belief that s/he is loved from afar by another and may attack individuals perceived as standing between them and their loved one, or the object of their affections if they feel slighted. Erotomania is associated with anger, harassment, stalking, and violence.[32]

Delusional disorders are found in excess in homicide (6-fold increase) and are associated with stalking. Stalking is persistent harassment in which a person repeatedly intrudes on another in an unwelcome manner that evokes fear or disquiet. One study of stalkers found that 30 per cent had a delusional disorder, 10 per cent schizophrenia, less than 5 per cent bipolar disorder or anxiety; 50 per cent had a personality disorder and 25 per cent abused substances. The classification of stalkers is based on their motivation. It is the rejected stalkers who pursue ex-intimates for reconciliation, revenge or both (includes delusional jealousy); intimacy seekers (includes erotomania); and the resentful stalker who pursue victims as revenge for an actual or perceived injury (includes persecutory) who are most likely to have a delusional disorder. This is not the case for incompetent suitors or predatory stalkers. See Chapter 11.10

Risk of harm to others can be reduced by treating the individual's delusional disorder with medication; controlling his or her environment by use of hospital or legal orders to restrict movement; and by advice to any potential victim on protective measures. There have been no randomized controlled trials of the treatment of delusional disorders but novel antipsychotics are now most commonly used due to their lower side-effect profile and better compliance, and it is recognized that pimozide has no particular role in the treatment of delusional disorder.[33] Cognitive behavioural therapy may assist compliance and insight but this has not been systematically demonstrated. Eye movement desensitization and reprocessing (EMDR) interwoven with cognitive therapy has been reported as a treatment of morbid jealousy. It is essential to warn any identified potential victims and such a breach of confidentiality is permitted by medical governing bodies.

Organic disorders

Organic disorders may be related to offending directly because of the brain injury with disinhibition and impaired judgement, or secondarily to socio-economic deprivation and exclusion.

Traumatic brain injury (TBI)

There is a 4-fold increase in the relative risk of developing any psychiatric illness in the six months post moderate to severe TBI and almost 3-fold in mild TBI excluding those with a history of mental illness. Common psychiatric sequelae of traumatic brain injury include mood (50 per cent), anxiety (25 per cent) and substance use disorders (28 per cent) along with post traumatic stress disorder and post traumatic brain injury attention deficit hyperactivity disorder in children, in addition to any cognitive problems.[34]

Posttraumatic irritability occurs in up to 70 per cent of people with TBI. Aggressive behaviour occurs in approximately one-quarter and this is significantly increased by the presence of major depression or substance abuse. Psychological tests of executive function and neuroimaging suggest that defective pre-frontal modulation of medial limbic structures may explain aggressive and impulsive behaviour seen in these patients. TBI patients who display aggression are characterized by significantly more impulsivity, disinhibition and social withdrawal; and poorer drive and motivation.[35] A study of sex offenders found that half had a history of head injury leading to loss of consciousness with significant neurological impairment in a quarter.[36] The head injury group offended more against adults than children, had more exhibiting and wide ranging sexual behaviours, and abused more substances than their non TBI sex offender controls.

Epilepsy

Contrary to previous views, the evidence now suggests that there is no association between epilepsy and criminal behaviour. A systematic review of epilepsy in prisoners found a prevalence rate of about 1 per cent which is comparable with a community sample of a similar age and gender.[37] Epilepsy is considered in detail in Chapter 11.3.3.

Dementia

A diagnosis of dementia and being charged with a sexual offence were factors most likely to distinguish those over 60 years of age referred by courts for a psychiatric assessment in Sweden compared to their younger counterparts although all forms of psychoses

(31 per cent), personality disorder (20 per cent), and substance use disorders (14.8 per cent) were more frequently diagnosed than dementia (7 per cent).[38] Offending is much less common in the elderly. Physical aggression in dementia however, is not uncommon and may be associated with exposure of intellectual deficits during testing (catastrophic reaction), severe cognitive impairment, impaired expression and comprehension, a history of premorbid aggression, physical illness particularly involving delirium or pain, post ictal confusion, depression and changes within the environment. Frontotemporal or orbitotemporal degeneration are more frequently associated with aggression in dementia, as are reduced serotonin and acetylcholine.[39]

Autism spectrum disorders

Autism spectrum disorders (ASD) (including autism and Asperger's syndrome) are described in Chapter 9.2.3. In the severe form (autism) incapacity is frequently major and affected individuals are consequently unlikely to come into contact with the judicial system. The frequency of the lesser forms is inadequately researched. Although the majority of individuals with ASD are law abiding—and many indeed may be less likely to offend as a result of 'concrete' rule-based thinking—they are found in excess in high security psychiatric settings. Those who offend may do so because of feelings of resentment caused by bullying or rejection at school or in the community, because of an over-sensitivity to sound, occasionally because of a powerful interest in guns, fire or killing[40] or because of a lack of appreciation of social norms. Theft may occur in order to pursue a particular interest or obsession.

Personality disorder

The link between personality disorder and offending and the role of psychiatry in managing personality disordered offenders is complicated by a number of issues: poor diagnostic systems; variable and subjective approaches to assessment; confusing legislative approaches seeking to define and impose measures on certain individuals; societal preoccupation with 'psychopaths' and 'sexual predators'; and uncertainty as to what treatment, if any, is effective. The concept of anti-social personality disorder, introduced in DSM-III and rehashed in DSM-IIIR and DSM-IV, has been unhelpful. Its reliance on socially deviant behaviour, rather than personality traits, reliably demarcates a broad heterogeneous group with varying underlying psychopathology and social problems. Personality disorder should not be diagnosed primarily on the basis of repeated anti-social behaviour. Assessment of personality in offenders should be based on pervasive and persistent emotional experience and expression, thoughts about self and others, interactions with others and behavioural control. Personality disordered offenders show a range of personality pathology rarely falling neatly into the diagnostic boxes of ICD 10 or DSM-IV. Focussing on personality disorder as a clinical entity misses relevant developments in assessment and management of such individuals from research and practice in criminal justice.

Psychopathy

The term psychopathy has a long and chequered history (see Chapter 4.12.1). Its clinical use should probably be reserved for individuals fulfilling Hare's criteria[41] (based on Cleckley's 'Mask of Sanity'[42] describing superficial, self-centred, callous, parasitic, impulsive, aggressive, predatory individuals. Hare operationalized psychopathy with the Psychopathy Check List-Revised (PCL-R). Interview and case file material is used to assess 20 items on a 3-point scale giving a maximum total score of 40. In North America a cut-off of 30 is used to diagnose psychopathy. In Europe the cut-off may be lower, but this is a contentious area. The total score may be useful for research purposes, but the profile of characteristics is important in clinical assessment and formulation. Underlying the construct are emotional, interpersonal and behavioural domains. Hare originally described two underlying factors: Factor 1 (emotional and interpersonal) correlated with narcissistic and histrionic disorders, and Factor 2 (socially deviant life style and behaviour) correlated with anti-social and borderline disorders. This has been refined into a four factor model subdividing factor 1 into interpersonal and affective domains, and factor 2 into lifestyle and antisocial behaviour domains. It has been suggested that the antisocial behaviour domain is not core to the condition.[43] Psychopathy is a narrower construct than dissocial (ICD 10) or antisocial (DSM-IV) personality disorder, so research findings cannot be extrapolated. It is an important concept in theory, research and practice concerning offender management and recidivism. But psychopathy only focuses on a limited, albeit important, set of personality traits of relevance in assessing offenders. Comprehensive consideration of personality pathology involves broader assessment. Patrick's book[44] gives an authoritative account of psychopathy.

Relationship between personality disorder and offending

Personality disorder rates in prisoners range from 10 per cent[45] to 78 per cent[46] depending on study method. Lower rates are found with comprehensive clinical assessment, higher rates with structured tools administered by non-clinicians. In male prisoners the most prevalent disorder category is, unsurprisingly, antisocial followed by paranoid, then borderline, then obsessive compulsive, avoidant and narcissistic at similar rates.[46] In Fazel and Danesh's systematic review[37] 65 per cent of male prisoners were personality disordered (47 per cent anti-social); amongst female prisoners the rate was 42 per cent (25 per cent borderline, 21 per cent anti-social). Psychopathy has been found in 10 to 30 per cent of prisoners.[47]

Blackburn and Coid highlighted the heterogeneity of personality pathology in violent offenders.[48] Using cluster analysis they found six diagnostic patterns: antisocial-narcissistic, antisocial-paranoid, borderline-antisocial-passive aggressive, borderline, compulsive-borderline, and schizoid. The three antisocial groups displayed more psychopathy. Half of homicide offenders in Sweden were personality disordered[49] but the rate in the UK was about 10 per cent[15] (an underestimate due to the method of the study). Rates of personality disorder are high in sexual offenders.[50] Adult rapists have higher rates of psychopathy than child molesters who have higher rates of avoidant and dependent disorders. Sexual homicide offenders have high rates of psychopathy and other personality disorders, particularly in serial offenders and sexual sadists. Sadistic personality disorder (defined in the DSM-IIIR appendix) is found in up to a quarter of violent and sexual offenders, but is present with other personality pathology and is associated with sexual sadism.[51] There is little support for its recognition as a discrete disorder. Most serial arsonists have personality disorders. Borderline, narcissistic and antisocial subtypes of domestically violent men have been

identified.[52] Personality disorder of various types is common in stalkers.[53] Psychopathy is particularly associated with instrumental violence, but reactive violence is also displayed.[44]

Population based cross sectional and birth cohort studies show offending and violence to be increased in individuals with personality disorders, particularly cluster B disorders.[54,55] Follow-up studies show higher rates of offending and violence in personality disordered subjects in general and forensic samples, particularly associated with psychopathy.[56,57,58] High rates of comorbid personality disorder are found in offenders with mental illness[59] and learning disability.[60] A range of personality pathology is found amongst patients detained under 'psychopathic disorder' and 'mental illness' legal categories in English secure hospitals.[59] Most legal 'psychopathic disorder' patients are not psychopaths.[61]

Various traits, individually and combined, are relevant to offending. Some lead to interpersonal conflict (suspiciousness, hostility, argumentativeness, rigidity, arrogance, clinginess), others to behavioural dyscontrol, others to angry emotional reactions, others to not considering consequences for self or others, others to taking pleasure in violating rules and others. Personality pathology may lead to inability to form intimate relationships, maintain work, establish a stable lifestyle or meet basic needs, which may predispose to offending. Impulsivity, need for stimulation, intolerance of dysphoric affect and inability to regulate affect predispose to drug and alcohol misuse leading to offending (see Chapter 11.7).

Assessment

Assessment, as in other cases, involves a comprehensive history, current mental state examination, and information from records and informants. Structured approaches to personality assessment (e.g. PCL-R and International Personality Disorders Examination (IPDE)), may be incorporated but 'psychometrics' should not be used in isolation. Comorbid mental illness should be considered and structured approaches to risk assessment should be used, especially with serious violent or sexual offenders. A comprehensive formulation should consider the relationship between dysfunctional personality traits and offending.

Management

Discussion of treatment is easily hijacked by the political agenda (largely focussed on protecting the public from dangerous psychopaths), the criminal justice context and legal issues. Although it would be naïve to consider treatment in isolation from these other issues, clinical management should be the starting point and primary concern of psychiatrists.

Most personality disordered offenders are not serious offenders, and their treatment needs are similar to individuals with personality disorder generally (see Chapter 4.12.7). General (e.g. long-term support and attention to relationships) and specific (e.g. psychological therapies and medication) approaches are relevant to many offenders. The application of such approaches within the criminal justice system or secure hospitals is complicated by coercion, the legal context and institutional factors. How should the success of treatment be measured? Amelioration of distress, improved social functioning or diminished reoffending? The three may be related, but the public, understandably, expects treatment to prevent offending.

Studies of consenting community patients cannot be extrapolated to serious offenders. The widespread view that psychological treatment programmes increase risk in psychopathic offenders is not backed by the literature, with mixed views and empirical findings.[62] The answer to 'Is psychopathy treatable?' is that the jury is still out and will be for some time. This does not mean treatment is ineffective or inappropriate for the range of personality pathology in serious offenders.

Cognitive behavioural programmes for offending behaviour have a moderate positive impact on recidivism.[63] Many of the areas targeted in such treatments relate to dysfunctional personality traits and many who go through these programmes have personality disorders. Adaptations for individuals with high levels of personality dysfunction and psychopathy have been made incorporating: greater focus on motivation, engagement, maintaining participation; using more appropriate learning styles; addressing underlying core beliefs; greater flexibility; emphasis on individual formulation, and positive psychology.[41] Whether such approaches lead to reduced recidivism awaits evaluation.

Although treatment in hospital may seem appropriate if there is to be a therapeutic approach, there are descriptions of prisons where a therapeutic environment has been achieved.[64] Detention in hospital primarily for indefinite incapacitation of dangerous offenders is controversial.

Personality disordered violent and sexual offenders have treatment needs, but they do not fit into traditional psychiatric systems, geared towards psychotic offenders. The approach with the latter is diversion from criminal justice to health care. With personality disordered offenders, the appropriate approach is probably to provide assessment and treatment within a criminal justice framework, in prison or community. Hospital treatment, as a scarce and expensive resource, should be reserved for those who have the potential to engage and benefit. Meux *et al* comprehensively describe hospital treatment approaches.[65] Staff working with personality disordered offenders, particularly with more severe pathology, need appropriate selection, training, support and supervision. Psychodynamic supervision of staff is important.

Management in the community involves monitoring, supervision and treatment by various agencies. In the UK Multi-Agency Public Protection Arrangements (MAPPA) are used to co-ordinate the management of violent and sexual offenders in the community. Within this mental health professionals can focus primarily on clinical assessment and treatment whilst criminal justice agencies focus on monitoring and supervision.

Outcome

Certain personality disorders are associated with risk of recidivism. Psychopathy is associated with high rates of general and violent recidivism.[41] Amongst sex offenders psychopathy predicts violent rather than sexual offending, although sexual deviation with psychopathy is a malignant combination.[66] Psychopathy is associated with violence towards victims and diverse offending. Successful social integration following discharge has been found to be associated with not being reconvicted.[67] Personality disorder is associated with suicide in offenders.[68]

Legal issues

An insanity defence, although historically relevant in some jurisdictions, is not usually available where the primary diagnosis is personality disorder. The partial defence of diminished responsibility is open to personality disordered homicide perpetrators in

England and Wales. Hospitalization does not often follow such a finding.

Detention in hospital is rather arbitrarily applied to some personality disordered offenders in England and Wales. Indefinite detention in hospital as a sentence can become lifelong incapacitation in a clinical setting.[69] To ensure that hospital detention is treatment focussed there should be prolonged assessment first. Some argue treatment in hospital should only occur via transfer to hospital during a prison sentence. In the Netherlands and Germany, personality disordered offenders are routinely detained in forensic hospitals.

Sexually violent predator laws allow civil commitment (and hospital detention) of mentally disordered high risk sex offenders at the end of a prison sentence in some US states.[70] Most offenders are personality disordered and these laws have survived constitutional challenge.

In many jurisdictions indeterminate sentences are applied to some offenders who are considered to pose an ongoing risk of serious offending.[71] Many such offenders are personality disordered. Psychiatric assessment of risk in such cases is considered unethical by some. Assessments of risk and personality disorder by psychiatrists have been used to under-write the application of the death penalty in some US States.[72]

Mood disorders

Mood disorders are less related to offending than schizophrenia. Disinhibition and grandiosity in individuals with mania leads to public order offences, driving offences, theft, fraud and minor violent or sexual offences. Serious violent or sexual offending is rare. Treatment non-compliance and comorbid substance misuse are associated with offending.

Depression is not uncommon in prisoners (10 per cent in Fazel and Danesh's systematic review[37]) but rarely leads to offending. The association between depression and shoplifting in the historical literature is probably spurious. When depression is encountered in an offender, it may not have been present prior to the offence. If it was it is unlikely to have played a direct role. High life-time rates of depression are reported in sex offenders, and low mood may be one factor of importance in the path to re-offending in some.

Suicide follows 5 per cent of homicides in the UK.[73] Depression sometimes plays a role but other psychopathology is also seen. Types of offences include:

- Spousal: with pathological jealousy.
- Spousal: elderly men with poor health and/or ailing spouses who feel either or both cannot cope with declining health, adversity, or loneliness.
- Filicide-suicide: depressed mother kills her child(ren) and herself to save them from a worse fate.
- Familicide-suicide: depressed, paranoid, or intoxicated male kills the family and himself in the context of financial, marital, or other social stresses.
- Extra-familial murder-suicide: disgruntled, paranoid, narcissistic individuals who feel slighted or humiliated, take revenge on single or multiple victims, either as specific targets or as bystanders. This pattern is seen in spree killings.

Altruistic homicide and extended suicide are terms used where a depressed person kills one or more family members to 'spare them suffering' and commits or attempts suicide. Most are filicide-suicide and familicide-suicide cases. Depressive psychosis with nihilistic delusions is seen in a minority. It is probably impossible to identify these cases in advance, but it is prudent to explore thoughts about children in parents presenting with suicidal ideation.

Publicity has been given to the association between SSRI anti-depressants and aggression and suicide. There is no epidemiological data to back these claims, and some evidence SSRIs reduce impulsivity and associated aggression.[74]

Severe depression or mania may make an accused unfit to plead; affective psychosis may be accepted as the basis of an insanity defence; and milder forms of mood disorder may found a defence of diminished responsibility in homicide cases. Very few patients in secure hospitals are detained on the basis of mood disorder. In England and Wales, mothers with postpartum disorders who kill their children are usually convicted of infanticide rather than murder.

Neurotic disorders

Neurotic symptoms are common in offenders, but it is unusual to find a close relationship between an offence and an ICD-10 neurotic, stress-related, or somatoform disorder. In many offenders it is possible to identify neurotic conflicts and find symbolic meaning in criminal behaviour.[75] Acute stress reactions and adjustment disorders may be associated with offending, particularly where the underlying personality is abnormal or substance misuse occurs. Social phobia may be relevant through its association with alcohol misuse. Although violent and sexual themes often feature in obsessions, obsessive compulsive disorder is negatively correlated with violence and offending. Obsessional features are sometimes seen in the paraphilic behaviour of some sexual offenders.

Post-traumatic stress disorder (PTSD) may be a contributory factor to offending. Offending rarely results directly from flashbacks or other re-experiencing phenomena, although there are cases where rape victims have attacked sexual partners. PTSD may contribute to the violence of women towards abusive partners. 'Battered Woman Syndrome' has been suggested as a sub-category of PTSD in such cases.[76] PTSD rarely leads to a defence of insanity, but may lead to diminished responsibility.

UK psychiatrist read with astonishment and incredulity accounts from the US of multiple personality disorder,[77] where the diagnosis has justified separate legal representation for alters, separate testimony from alters, unfitness to plead, the insanity defence, and malpractice suits when the diagnosis is 'missed'. James and Schramm[78] give guidance to avoid repetition of the US experience in the UK. Psychiatrists are advised not to collude by looking for further alters, investigating those that 'appear', and treating the patient as if he were more than one person. Frankel and Dalenberg[79] give a comprehensive account of dissociation and dissociative identity disorder in the US context.

Factitious illness by proxy

Also called Munchausen syndrome by proxy, this is the fabrication of symptoms in, or the injury of, a child by its carer who presents

the child for medical attention.[80] There is a wide range of injurious behaviours from fabricating symptoms and tampering with specimens and charts to poisoning, smothering, and withholding of nutrients. It is a behaviour rather than a psychiatric disorder. It may feature in systematic and serial child abuse, and in serial killings by health-care workers. Underlying disorders include personality disorders (narcissitic, borderline, and antisocial), somatization, mood, eating and substance misuse disorders. Psychodynamic literature offers various perspectives. In a number of highly publicized cases in the UK, expert evidence regarding families with multiple unexplained child deaths has led to mothers being convicted of murder. Several cases have been overturned on appeal with subsequent disciplinary action against the experts involved.

Impulse control disorders

Impulse control disorders are described in Chapter 4.13.1. Two specific 'disorders' have relevance in forensic psychiatry: kleptomania and pyromania. Psychiatric aspects of arson are described in Chapter 11.9

In clinical practice, it is rare to find a shoplifter who meets criteria for kleptomania without underlying psychiatric disorder (mood, anxiety or eating disorder).[81] Apparently inexplicable acts of shoplifting may be perpetrated by people who have no features of kleptomania. Goldman[82] in an extensive review concluded that the disorder is more common than previously suggested.

The diagnosis of pyromania is of equally dubious validity. Most repeated arson offenders with the clinical features are excluded as they have other underlying disorders (psychosis, intellectual disability, or personality disorder) or are intoxicated when they set fires.[83]

Further information

Patrick, C. J. (2006). *Handbook of psychopathy*. Guilford Press, New York.

Hodgins, S. and Muller-Isberner, R. (2000) *Violence, Crime and Mentally Disordered Offenders: Concepts and Methods for Effective Treatment and Prevention*. John Wiley and Sons Ltd, Chichester.

Gunn, J. and Taylor, P. *Forensic Psychiatry*. Hodder Arnold Health Sciences (in press)

Soothill, K., Rogers, P., and Dolan, M. (2008). *Handbook of Forensic Mental Health*. Willan Publishing, Cullompton, UK.

References

1. Home Office. Crime Statistics for England and Wales. National Statistics, London. 2006 www.crimestatistics.org.uk. US Department of Justice (2006) Violent crime rates have declined, reaching the lowest level ever in 2005. Bureau of Justice Statistics, Washington, DC, USA. www.ojp.usdoj.gov/bjs/
2. Brower, M.C. and Price, B.H. (2001). Neuropsychiatry of frontal lobe dysfunction in violent and criminal behaviour: a critical review. *Advances in Neuropsychiatry*, **71**, 720–6.
3. Dolan, M., Deakin, W.J.F., Roberts, N., *et al.* (2002). Serotonergic and cognitive impairment in impulsive aggressive personality disordered offenders: are there implications for treatment? *Psychological Medicine*, **32**(1), 105–117.
4. Hollander, E., Phillips, A., Chaplin, W., *et al.* (2005). A Placebo Controlled Crossover Trial of Liquid Fluoxetine on Repetitive Behaviors in Childhood and Adolescent Autism. *Neuropsychopharmacology*, **30**(3), 582–9.
5. Gesch, C., Bernard, H., Sean, M., et al. (2002). Influence of supplementary vitamins, minerals and essential fatty acids on the antisocial behaviour of young adult prisoners: Randomised, placebo-controlled trial. *British Journal of Psychiatry*, **181**(1), 22–8.
6. Brennan, P.A., Mednick, S.A., Jacobsen, B., *et al.* (1996). Assessing the Role of Genetics and Crime Using Adoption Cohorts. *Ciba Foundation Symposium*, **194**, 115–28.
7. Caspi, A., McClay, J., Moffitt, T., *et al.* (2002). Role of Genotype in the Cycle of Violence in Maltreated Children. *Science*, **297**, 851–4.
8. Meyer-Lindenberg, A., Buckholtz, J.W., Kolachana, B.R., *et al.* (2006). Neuromechanisms of Genetic Risk for Impulsivity and Violence in Humans. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(16), 6269–74.
9. Viding, E., Blair, R., James, R., *et al.* (2005). Evidence for Substantial Genetic Risk for Psychopathy in 7 Year Olds. *Journal of Child Psychology and Psychiatry*, **46**(6), 592–7.
10. Hiday, V.A. (1997). Understanding the connection between mental illness and violence. *International Journal of Law and Psychiatry*, **20**, 399–417.
11. Brennan, P.A., Mednick, S.A., Hodgins, S., (2000). Major mental disorders and criminal violence in a Danish birth cohort. *Archives of General Psychiatry*, **575**, 494–500.
12. Fazel, S. and Grann, M. (2006). The Population Impact of Severe Mental Illness on Violent Crime. *The American Journal of Psychiatry*, **163**(8), 1397–403.
13. Foley, S.R., Kelly, B.D., Clarke, M., *et al.* (2005). Incidence and clinical correlates of aggression and violence at presentation in patients with first episode psychosis. *Schizophrenia Research*, **72**, 161–8.
14. Arseneault, L., Moffitt, T., Caspi, A., *et al.* (2000). Mental Disorders and Violence in a Total Birth Cohort: Results from the Dunedin Study. *Archives of General Psychiatry*, **57**, 979–86.
15. Shaw, J., Hunt, I., Flynn, S., *et al.* (2006). Rates of mental disorder in people convicted of homicide: National clinical survey. *British Journal of Psychiatry*, **188**(2), 143–7.
16. Nordstrom, A., Dahlgren, L., Kullgren, G., *et al.* (2006). Victim relations and factors triggering homicides committed by offenders with schizophrenia. *The Journal of Forensic Psychiatry and Psychology*, **17**(2), 192–203.
17. Naudts, K., and Hodgins, S. (2006). Schizophrenia and violence: a search for neurobiological correlates. *Current Opinion in Psychiatry*, **19**, 533–8.
18. Silver, H., Goodman, C., Knoll, G., *et al.* (2005). Schizophrenia Patients with a History of Severe Violence Differ from Nonviolent Schizophrenia Patients in Perception of Emotions but Not Cognitive Function. *Journal of Clinical Psychiatry*, **66**(3), 300–8.
19. Applebaum, P.S., Robbins, P.C., Monahan, J., *et al.* (2000). Violence and Delusions: Data From the MacArthur Violence Risk Assessment Study. *The American Journal of Psychiatry*, **157**(4), 566–72.
20. Swanson, J.W., Swartz, M.S., Van Dorn, R.A., *et al.* (2006). A National Study of Violent Behavior in Persons with Schizophrenia. *Archives of General Psychiatry*, **63**, 490–9.
21. Hodgins, S., Hiscoke, U., Freese, R., *et al.* (2003). The Antecedents of Aggressive Behavior Among Men with Schizophrenia: A Prospective Investigation of Patients in Community Treatment. *Behavioral Sciences & the Law*, **21**(4), 523–46.
22. Arsenault, L., Moffitt, T., Caspi, A., *et al.* (2000). Mental Disoders and Violence in a total Birth Cohort: Results from the Denedin Study. *Archives of General Psychiatry*, **57**, 979–86.
23. Haggard-Grann, U., Hallqvist, H., Langstrom, N., *et al.* (2006). Short-term effects of psychiatric symptoms and interpersonal stressors on criminal violence, a case-crossover study. *Social psychiatry and psychiatric epidemiology*, **41**, 532–40.
24. Moran, P. and Hodgins, S. (2004). The Correlates of Comorbid Antisocial Personality Disorder in Schizophrenia. *Schizophrenia Bulletin*, **30**(4), 791–802.

25. Miller, PMcC., Johnstone, E.C., Lang, F.H., *et al.* (2000). Differences between patients with schizophrenia within and without a high security psychiatric hospital. *Acta Psychiatrica Scandanavica*, **102**, 12–8.
26. Walsh, E., Gilvarry, C., Samele, C., *et al.* (2004). Predicting violence in schizophrenia: a prospective study. *Schizophrenia Research*, 67, 247–52.
27. Taylor, P.J. (2006). Delusional Disorder and Delusions: Is There a Risk of Violence in Social Interactions About the Core Symptom? *Behavioural Sciences and the Law*, **24**, 313–31.
28. Krakowski, M.I., Czobor, P., Citrome, L., *et al.* (2006).Atypical antipsychotic agents in the treatment of violent patients with schizophrenia and schizomood disorder. *Archives of General Psychiatry*, **63**(6), 622–9.
29. Wallace, C., Mullen, P.E., Burgess, P., *et al.* (2004). Criminal Offending in Schizophrenia Over a 25-Year Period Marked by Deinstitutionalization and Increasing Prevalence of Comorbid Substance Use Disorders. *American Journal of Psychiatry*, **161**(4), 716–27.
30. Buchanan, A. (1998).Criminal conviction after discharge from special (highsecurity) hospital. *British Journal of Psychiatry*, **172**(6), 472–6.
31. Thomson, L.D.G. (2005). *Mental Disorder and Psychiatric Services, Mental Health and Scots Law in Practice.* W. Green & Sons, Edinburgh, Chapter 1.
32. Kelly, B.D. (2005). Erotomania: Epidemiology and Management. Therapy in Practice, *CNS Drugs*, (8), 657–69.
33. Sultanta, A. and McMonagle, T. (2007). Pimozide for schizophrenia or related psychoses. *Cochrane Database of Systematic Reviews*, 1.
34. Jorge, R.E. (2005). Neuropsychiatric consequences of traumatic brain injury: a review of recent findings. *Current Opinion in Psychiatry*, **18**(3), 289–99.
35. Wood, R.L. and Liossi, C. (2006). Neuropsychological and Neurobehavioral Correlates of Aggression Following Traumatic Brain Injury. *Journal of Neuropsychiatry and Clinical Neurosciences*, **18**(3), 333–41.
36. Langevin, R. (2006). Sexual offences and traumatic brain injury. *Brain and Cognition*, **60**(2), 206–7.
37. Fazel, S., and Danesh, J. (2002). Serious mental disorder in 23000 prisoners: a systematic review of 62 surveys. *Lancet*, **359**, 545–50.
38. Fazel, S., and Grann, M. (2002). Older Criminals: a descriptive study of psychiatrically examined offenders in Sweden. *International Journal of Geriatric Psychiatry*, **17**, 907–13.
39. Hall, K.A. and O'Connor, D.W. (2004). Correlates of aggressive behavior in dementia. *International Psychogeriatrics*, **16**(2), 141–58.
40. Wing, L. (1997). Asperger's syndrome: Management requires diagnosis. *Journal of Forensic Psychiatry*, **8**, 253–57.
41. Hare, R.D. (2006). Psychopathy: a clinical and forensic overview. *Psychiatric Clinics of North America*, **29**, 709–24.
42. Cleckley, H. (1998). (5th edn.). *The Mask of Sanity: An Attempt to Clarify Some Issues About the So Called Psychopathic Personality.* Emily S.Cleckley, Augusta, GA.
43. Cooke, D.J., Michie, C., Hart, S.D., *et al.* (2004). Reconstructing psychopathy: clarifying the significance of antisocial and socially deviant behavior in the diagnosis of psychopathic personality disorder. *Journal of Personality Disorders*, **18**, 337–57.
44. Patrick, C.J. (2006). (ed.). *Handbook of Psychopathy*. Guilford Press, New York.
45. Gunn, J., Maden, A., Swinton, M. (1991). Treatment needs of prisoners with psychiatric disorders. *British Medical Journal*, **303**, 338–41.
46. Singleton, N., Meltzer, H., Gatwood, R., *et al.* (1998). *Psychiatric morbidity among prisoners in England and Wales.* Office for National Statistics, London.
47. Andersen, H.S. (2004).Mental health in prison populations. A review–with special emphasis on a study of Danish prisoners on remand. *Acta psychiatrica Scandinavica. Supplementum*, **424**, 5–59.
48. Blackburn, R., and Coid, J.W.(1999). Empirical clusters of DSM-III personality disorders in violent offenders. *Journal of Personality Disorders*, **13**, 18–34.
49. Fazel, S., and Grann, M. (2004).Psychiatric morbidity among homicide offenders: a Swedish population study. *American Journal of Psychiatry*, **161**, 2129–31.
50. Harsch, S., Bergk, J.E., Steinert, T., *et al.* (2006). Prevalence of mental disorders among sexual offenders in forensic psychiatry and prison. *International Journal of Law and Psychiatry*, **29**, 443–9.
51. Berger, P., Berner, W., Bolterauer, J., *et al.* (1999). Sadistic personality disorder in sex offenders: relationship to antisocial personality disorder and sexual sadism. *Journal of Personality Disorders*, **13**, 175–86.
52. Johnson, R., Gilchrist, E., Beech, AR., *et al.* (2006). A psychometric typology of U.K. domestic violence offenders. *Journal of Interpersonal Violence*, **21**, 1270–85.
53. Mullen, P.E., Pathe, M., and Purcell, R. (2000). *Stalkers and Their Victims*. Cambridge University Press.
54. Coid, J., Yang, M., Tyrer, P., *et al.* (2006). Prevalence and correlates of personality disorder in Great Britain. *British Journal of Psychiatry*, **188**, 423–31.
55. Hodgins, S., Mednick, S.A., Brennan, P.A., *et al.* (1996).Mental disorder and crime: evidence from a Danish birth cohort. *Archives of General Psychiatry*, **53**, 489–96.
56. Steadman, H., Mulvey, E., Monahan, J., *et al.* (1998). Violence by people discharged from acute psychiatric inpatient facilities and by others in the same neighborhoods. *Archives of General Psychiatry*, **3**, 401.
57. Grann, M., Langstrom, N., Tengstrom, A., *et al.*(1999). Psychopathy (PCL-R) predicts violent recidivism among criminal offenders with personality disorders in Sweden. *Law and Human Behavior*, **23**, 205–17.
58. Quinsey, V.L., Harris, G.T., Rice, M.E., *et al.* (2005). Violent Offenders: Appraising and Managing Risk. American Psychological Association.
59. Blackburn, R., Logan, Donnelly, *et al.* (2003). Personality disorders, psychopathy and other mental disorders: co-morbidity among patients at English and Scottish high-security hospitals. *Journal of Forensic Psychiatry*, **14**, 111–37.
60. Hogue, T., Steptoe, L., Taylor, J.L., *et al.* (2006). A comparison of offenders with intellectual disability across three levels of security. *Criminal Behaviour and Mental Health*, **16**, 13–28.
61. Coid, J.W. (1992). DSM–III diagnosis in criminal psychopaths: a way forward. *Criminal Behaviour and Mental Health*, **2**, 78–94.
62. D'Silva, K., Duggan, C., McCarthy, L., *et al.* (2004).Does Treatment Really Make Psychopaths Worse? A Review of the Evidence. *Journal of Personality Disorders*, **18**(2), 163–77.
63. McGuire, J. (2002).Offender Rehabilitation and Treatment: Effective Programmes and Policies to Reduce Reoffending. John Wiley and Sons Ltd.
64. Morris, M. (2002). Managing the unmanageable: psychotherapy in Grendon Prison. *Criminal Behaviour and Mental Health*, **12**(2 Suppl), S54–67.
65. Meux, C., Newrith, C., Taylor, P., (2006). Personality Disorder and Serious Offending: Hospital Treatment Models. Hodder Arnold.
66. Olver, M.E. and Wong, S.C. (2006). Psychopathy, sexual deviance, and recidivism among sex offenders. *Sex Abuse*, **18**, 65–82.
67. Reiss, D., Grubin, D., Meux, C., *et al.* (1996).Young 'psychopaths' in special hospital: treatment and outcome. *British Journal of Psychiatry*, **168**, 99–104.
68. Kullgren, G., Tengstrom, A., Grann, M., *et al.* (1998). Suicide among personalitydisordered offenders: a follow-up study of 1943 male criminal offenders. *Social Psychiatry and Psychiatric Epidemiology*, **33**(Suppl 1), S102–6.
69. Chiswick, D. (1992). Compulsory treatment of patients with psychopathic disorder: an abnormally aggressive or seriously

irresponsible exercise? *Criminal Behaviour and Mental Health*, **2**, 106–13.
70. Miller, H.A., Amenta, A.E., Conroy, M.A., *et al.* (2005). Sexually violent predator evaluations: empirical evidence, strategies for professionals, and research directions. *Law and Human Behavior*, **29**, 29–54.
71. Heilbrun, K., Ogloff, J.R.P., Picarello, K., *et al.* (1999). Dangerous offender statutes in the United States and Canada: Implications for risk assessment. *International Journal of Law and Psychiatry*, **22**, 393–415.
72 Edens, J.F., Buffington-Vollum, J.K., Keilen, A., *et al.* (2005). Predictions of future dangerousness in capital murder trials: is it time to 'disinvent the wheel'? *Law and Human Beahvior*, **29**, 55–86.
73. Shaw, J. and Flynn, J. (2003). Homicide followed by suicide. *Psychiatry*, **2**, 32–5.
74. Walsh, M.T. and Dinan, T. G. (2001).Selective serotonin reuptake inhibitors and violence: a review of the available evidence. *Acta psychiatrica Scandinavica*, **104**, 84–91.
75. Cordess, C. and Cox, M. (eds.). (1996). Forensic psychotherapy. Jessica Kingsley, London.
76. Walker, L.E. (2006). Battered woman syndrome: empirical findings. *Annals of the New York Academy of Science*, **1087**, 142–57.
77. Kluft, R.P. (1991). Clinical presentations of multiple personality disorder. *Psychiatric Clinics of North America*, **14**, 605–29.
78. James, D. and Schramm, M. (1998). 'Multiple personality disorder' presenting to the English courts: a case study. *Journal of Forensic Psychiatry*, **9**, 615–8.
79. Frankel, A.S. and Dalenberg, C. (2006). The forensic evaluation of dissociation and persons diagnosed with dissociative identity disorder: searching for convergence. *Psychiatric Clinics of North America*, **29**, 169–84.
80. Adshead, G. and Brooke, D. (2001). Munchasen Syndrom by Proxy. *Current issues in assessment, treatment and research*. Imperial College Press, London.
81. McElroy, S.L., Pope, H.G., Hudson, J.I., *et al.* (1991). Kleptomania: A report of 20 cases. *American Journal of Psychiatry*, **148**, 652–7.
82. Goldman, M.J. (1991). Kleptomania: making sense of the nonsensical. *American Journal of Psychiatry*, **148**, 986–96.
83. Lindberg, N., Holi, M., Tani, P., *et al.* (2005). Looking for pyromania: Characteristics of a consecutive sample of Finnish male criminals with histories of recidivist fire-setting between 1973 and 1993. *BMC Psychiatry*, **5**, 47.

11.3.2 Offending, substance misuse, and mental disorder

Andrew Johns

This chapter deals with the relationship between offending, substance misuse, and mental disorder, and also describes approaches to clinical and medico-legal assessment.

Relationship between offending, substance misuse, and mental disorder

The nature of this relationship is complex, yet has to be understood in order to manage the health and risk of offending of individual patients. Offences related to substance misuse can be categorized as (i) violent offences, often involving an altered mental state, (ii) acquisitive offences, and (iii) miscellaneous offences such as breaking laws to control the misuse of drugs, driving under the influence of alcohol, and impact of substance misuse on parenting.

Violent offences

Aggression is not an inevitable pharmacological consequence of misusing alcohol or any particular drug, but arises from many possible factors including expectancy effects, pattern of consumption, individual responses to intoxication or withdrawal, peer influences, and interpersonal issues.

Alcohol consumption is the single factor most associated with violence. It is a repeated finding from the British Crime Surveys that alcohol is a key factor in at least half of interpersonal assaults, with a greater contribution to assaults on strangers and domestic violence. By comparison, drug misusers are overwhelmingly more likely to commit acquisitive offences.

The co-morbidity of major mental illness and substance misuse further increases the risk of violence. For example, in the Epidemiological Catchment Area (ECA) survey of 10 000 individuals,[1] the prevalence of violent behaviour in the previous year was 2 per cent for those with no mental disorder, 7 per cent in 'major mental illness', 20 per cent for 'substance misuse disorder', and 22 per cent for co-morbid respondents. Among patients with first-episode psychoses,[2] just under 10 per cent demonstrated serious aggression when psychotic and 23 per cent showed lesser degrees of aggression. Those co-morbid for drug misuse were nine times more likely to show aggression after service contact—primary drug-related psychoses or alcohol misuse were not so associated.

There is particular concern in the UK about the risk to the public from serious violence by the mentally disordered. Nationally,[3] alcohol or drug misuse contributes to two-fifths of homicides and 17 per cent were committed by patients with severe mental illness and substance misuse. Alcohol- and drug-related homicides were generally associated with male perpetrators who had a history of violence, personality disorders, mental health service contact, and with stranger victims.

However, these epidemiological studies cannot define a causal link between substance misuse and mental disorder. There are many ways in which violence may arise in substance misusers and in co-morbid individuals. Simple intoxication on alcohol or other depressants such as benzodiazepines or barbiturates, leads initially to apparently excited behaviour. Stimulants such as cocaine or amphetamines, may produce arousal and irritability. Most forms of intoxication are also associated with impaired judgement, perception, and impulse control. Severe intoxication on alcohol, Cannabis, sedatives, or stimulants can lead to a toxic psychosis and highly disturbed behaviour. Even at levels of consumption insufficient to intoxicate, disinhibition, and autonomic arousal may facilitate recklessness and aggression. Pathological intoxication, in which aggression is supposed to occur within minutes of consuming moderate amounts of alcohol, is of doubtful validity and in most cases, better explained by alcohol-induced hypoglycaemia, head injury, or other organic disorder.

The association between withdrawal effects and potential for violence is often overlooked. Withdrawal from alcohol and most drugs of dependence, is a highly aversive state in which irritability and aggression may occur. Cessation from alcohol or sedatives, may lead to more severe withdrawal syndromes such as delirium tremens which are commonly associated with impaired perception, affect, judgement, and impulse control.

Acquisitive offending

The relationship between acquisitive crime and drug misuse problems was studied among 753 clients recruited to the National Treatment Outcome Research Study (NTORS).[4] More than 17 000 offences were reported during the 90-day period prior to treatment. Half of the clients committed no acquisitive crimes during this period, whereas 10 per cent committed 76 per cent of the crimes.

Such work does not demonstrate a causal relationship between illicit drug use and acquisitive crime. From a large survey of British youth,[5] the average age of onset for truancy and crime are 13.8 and 14.5 years respectively, compared with 16.2 for drugs generally and 19.9 years for 'hard' drugs. Thus, crime tends to precede drug use rather than vice versa. It is clear that heavy drug use is strongly associated with impulsive acquisitive offending, including street robbery, and burglary, which involve violence.

Other offences

In Britain, the non-medical use of drugs is subject to the Misuse of Drugs Act 1971, as subsequently amended, and which contains a classification based on perceived harm. Class A drugs include Ecstasy, LSD, heroin, cocaine, crack, magic mushrooms (if prepared for use), amphetamines (if prepared for injection); Class B drugs include amphetamines and methylphenidate (ritalin); Class C drugs include Cannabis, tranquilizers, some painkillers, Gamma hydroxybutyrate (GHB), ketamine. In January 2004, Cannabis was reclassified from a Class B to a Class C drug, it is still illegal. This legislation defines the penalties for supply, dealing, production, trafficking, and also possession.

Other offences include driving cars, or public conveyances such as trains whilst under the influence of alcohol or other drugs.

Responding to the drug or alcohol using offender

Clinical- and risk-assessment

The following is a practical guide to assessing the drug or alcohol-using offender.

1 Obtain a detailed life history, with corroboration from other informants and agencies. This should include relationships, work record, and current social situation.

2 Take a detailed history of all of the substances of misuse, including onset of regular use, dosage, route of administration, and pattern of use in a typical week. Ask about the desired effects of substance misuse, and also the actual effects. If there is also a serious mental illness, ascertain the effect of substance misuse on symptoms and behaviour. Has previous or recent substance misuse been associated with self-harm or aggression? Note any history of substance misuse treatment and the effects of this.

3 Take a detailed history of previous and recent offending with reference to the effects of substance misuse and mental illness, on mental state and behaviour before, during, and after each offence. Corroborate where possible from witness statements and independent sources.

4 Take a detailed mental state including some assessment of intelligence and personality, and also degree of insight into their offending, illness, and substance misuse.

5 Consider with the patient, the practical implications of this assessment for immediate clinical management and any need for medico-legal reporting that may arise.

There is increasing recognition of the role of actuarial instruments such as the HCR-20 (Historical/Clinical/Risk-management 20-item scale)[6] which allow for previous and current substance misuse to evaluated in the context of other significant risk factors.

Informed by the above, carry out a risk-assessment, firstly by defining the nature of any risk such as self-harm or relapse in substance misuse or of a primary mental illness. If a risk of violence is identified, this may involve particular individuals such as family, partners, or carers. Assess the probability and severity of each risk, and whether there are any early warning signs, such as particular behaviours or symptoms, or non-compliance with treatment.

Risk- and clinical management

Maden[7] (2007) argues 'the first step in improving risk-management is to recognize that that the prevention of violence is a central task of mental health services'. As a general approach to risk-management of the drug or alcohol misusing offender, ascertain what risk factors may be changed, and how the provision of support, care, or security may reduce the risk.

It is clearly important to achieve cessation or control of drug or alcohol misuse. These are not easy aims, but it is important to dispel therapeutic nihilism and to appreciate that a range of interventions have been shown to be effective. Details of specific interventions are given in Chapter 4.2.2.4 and Chapters 4.2.3.1–4.2.3.7.

Treatment can reduce re-offending. The NTORS[4] found that 5 years after treatment, convictions for acquisitive, drug selling, and violent crimes had reduced.

The National Confidential Inquiry[8] has concluded that provision for dual diagnosis should be central to modern mental health care and should include: staff training in substance misuse management, joint working with drug and alcohol teams, local clinical leadership and use of enhanced Care Programme Approach for all those with severe mental illness, and a destabilizing substance misuse problem.

Medico-legal issues

Possible defences related to substance misuse

Generally speaking, the acute effects of having voluntarily taken drugs or alcohol are not a mitigating factor and it is argued that a drunken intent is still an intent. There are however narrowly defined circumstances in which an altered mental state due to substance misuse can raise the question of a possible defence.[9]

Amnesia

Amnesia is common after violent offending, may relate to acute intoxication especially on alcohol or sedatives. In the absence of organic disease, amnesia does not affect fitness to plead, though it clearly complicates assessment of the perpetrators mental state at the time of an offence.

Simple intoxication

Self-induced intoxication is generally no defence to a criminal charge. However, in England and Wales, case-law has determined

that crimes such as murder, wounding with intent, theft, and burglary, require a *specific intent*, for which self-induced intoxication on alcohol or drugs may be a defence, but only if it can be shown that the accused was so intoxicated as to be unable to form the necessary intent. The psychiatrist can only comment as to whether the accused had the capacity to form the specific intent. It is a matter for the jury to determine whether the specific intent was present or not. If the specific intent is not demonstrated, then the accused may still be convicted of a lesser offence, so that acquittal on a charge of murder may lead to a conviction for manslaughter. It is a matter of clinical judgement as to whether an individual was so intoxicated as to be unable to form a specific intent, and the degree of purposiveness before, during, and after the offence, may be a useful indication.

Other crimes such as manslaughter, rape, and unlawful wounding, require only a *basic intent*, which cannot be negatived by intoxication. For these offences, the recklessness of voluntary intoxication may provide the necessary mental guilt.

Insanity

Alcohol or drug misuse may give rise to a psychotic illness, such as delirium tremens, which may meet the requirements of the McNaughton rules, but the inanity defence is rarely used. In theory, consumption of drugs or alcohol could lead to a state of insane automatism, but the defence of insanity is not available if the consumption has been voluntary.

Diminished responsibility

In England and Wales, Section 2 of the Homicide Act 1957 provides a defence of diminished responsibility in a charge of murder. The defence has to demonstrate that an *abnormality of mind* arises from one of the causes specified in the Act and those of possible relevance to substance misuse are *disease*, *injury* or *inherent causes*. An abnormality of mind due to intoxication is no defence. Alcohol dependence could meet criteria for *disease*, provided that the first drink of the day was shown to be involuntary. Diminished responsibility may become an issue when the effects of substance misuse interact with other factors such as organic brain damage, depression, or personality disorder. For legal purposes, the effect of intoxication has to be set aside and the defence must show that the associated condition was in itself severe enough to lead to an abnormality of mind.

Psychiatric recommendations to the courts

In reporting to the court, the task of the psychiatrist is to explain the possible contribution of substance misuse to a particular offence, in the context of the life history, psychiatric, and offending history of the individual. The aim of such a report is (i) to consider relevant psychiatric issues and their bearing on the offence, (ii) to indicate whether treatment could usefully prevent re-offending, and (iii) to help the court to protect society. There are a range of legal interventions that can facilitate engagement in treatment in community settings.

References

1. Swanson, J.W. (1994). Mental disorder, substance abuse, and community violence: an epidemiological approach. In *Violence and mental disorder—developments in risk assessment* (eds. J. Monahan and H.J. Steadman), pp. 101–36. University of Chicago Press, London.
2. Milton, J., Amin, S., Singh, S.P., *et al.* (2001). Aggressive incidents in first-episode psychosis. *The British Journal of Psychiatry*, **178**, 433–40.
3. Shaw, J., Hunt, I.M., Flynn, S., *et al.* (2006). The role of alcohol and drugs in homicides in England and Wales. *Addiction*, **101**, 1071–2.
4. Gossop, M., Trakada, K., Stewart, D., *et al.* (2005). Reductions in criminal convictions after addiction treatment: 5-year follow-up. *Drug and Alcohol Dependence*, **79**, 295–302.
5. Pudney, S. (2002). Home office research study 253, the road to ruin? Sequences of initiation into drug use and offending by young people in Britain. Home Office Research, Development and Statistics Directorate, UK.
6. Webster, C.D., Douglas, K.S., Eaves, D. *et al.* (1997). *HCR-20: assessing risk for violence* (version 2). Mental Health Law and Policy Institute, Simon Fraser University, Vancouver.
7. Maden, T. (2007). *Treating violence; a guide to risk management in mental health*. Oxford University Press, Oxford.
8. Avoidable deaths. Five year report of the National Confidential Inquiry into suicide and homicide by people with mental illness. (2006). University of Manchester.
9. Haque, Q. and Cumming, I. (2003). Intoxication and legal defences. *Advances in Psychiatric Treatment*, **9**, 144–51.

11.3.3 Cognitive disorders, epilepsy, ADHD, and offending

Norbert Nedopil

'Cognitive disorders' is a broad and heterogeneous diagnostic category, which includes different disorders, each with a distinct aetiology. They affect individuals in different ways depending on the age in which they occur. The term may be applied to a child, who has experienced perinatal trauma as well as to an older person with a beginning dementia of the Alzheimer type. The scientific literature on offenders with cognitive disorders is sparse. Most authors in forensic psychiatry do not systematically differentiate between the diagnostic subcategories and tend to use broad terms, such as organic disorder, organic psychosis, organic brain syndrome, neuropsychological deficit, dementia, mental handicap, mental retardation to include a number of different disorders in their studies. The number of patients with any kind of brain disorder in forensic hospitals and institutions is comparatively small and ranges from 1 to 10 per cent of all forensic inpatients. The same numbers apply for individuals assessed for criminal responsibility or risk of reoffending.[1–3] Compared to major mental disorders like schizophrenia or affective disorders or to personality disorders, patients with cognitive disorders account for only a small proportion of individuals seen by forensic psychiatrists. Subdividing this group any further would be statistically irrelevant. The way forensic psychiatry and the law deals with offenders suffering from organic brain disorders is rather derived from case reports and convention than from empirical knowledge.

DSM-IV-TR cites several disorders where aggression is either a diagnostic or associated feature and among them are four with an organic aetiology.

- Dementia of the Alzheimer type (DAT)
- Dementia caused by head trauma

- Personality change due to general medical condition (aggressive type)
- Postconcussional disorder

The psychiatric and general medical literature lists several other organic brain disorders that are either believed to be or in fact are associated with violence and offending, although their link is not as well proven.[4]

- Epilepsy
- Huntington's chorea
- Korsakow psychosis
- Brain tumours
- Mental retardation

From the experience of the author two other disorders should be added to this list:

- Traumatic brain injury
- Frontotemporal dementia

Systematic analyses of epidemiological data and of other research findings show that patients with clinically relevant brain damage do not commit violent crimes more often than would be expected according to their proportion in the general population.[5–7] These findings do not contradict the knowledge we have about aggressive and disruptive behaviour of certain patients with brain damage. The estimates of the frequency of such behaviours range from 18 per cent in demented patients to 60 per cent in patients with frontal lobe injuries.[8] Most of these patients are not seen by forensic psychiatrists, but are treated in special institutions or in outpatient settings. Apparently, the violent behaviour of patients with brain damage does not lead to interventions by the criminal justice system as often as could be expected from the above mentioned numbers. Similar findings are reported from demented patients: although not appearing in criminal court files, aggression and agitation of demented patients is a major problem in nursing homes and for caregivers of the elderly in outpatient settings. Again, exact definitions and robust data on how much violence really occurs are lacking, but estimates range from 18 to 48 per cent.[8] Rabins *et al.*[9] reported that 75 per cent of caregivers considered aggression as the most serious problem in agitated demented patients.

Offending and contact with the criminal justice system can be expected to be more frequent in patients who suffer less from cognitive impairments—which would prevent skilled or planned criminal activity—but rather from personality changes, like irritability, impulsivity, lack of concern for others, and for the consequences of one's action which is the case in frontotemporal dementia. Offending, but rarely violent offending, occurs sometimes as a first sign of this disorder.[10] Violent crimes are sometimes associated with cognitive disorders when delusions are among the first symptoms of a beginning dementia. Especially delusions of jealousy, envy, or revenge are prone to result in violent acts, which may leave partners or neighbours as victims. These crimes contradict the previous occupational and social life of the perpetrators and are paradigmatic examples of offending as a result of a mental disorder, leading to inculpability of the patient.

Offending can also be expected to be more frequent in patients between 18 and 35 years old and therefore in an age, where offending is statistically more frequent than in other age groups. Males of the same age group have the highest rate of traumatic brain injury. They also belong to the age group with the highest rate of criminality and especially of violent criminality. This same age and sex group also has the highest rate of substance abuse. Given the high prevalence of brain injury among young men and their propensity to use alcohol and drugs it is surprising how few are seen by forensic psychiatrists or sentenced to prison. The actual numbers of such patients found in forensic hospitals and in prisons do not reflect the high risk of violent crime by persons with brain injury. Hodgins[11] found that only 0.4 per cent of male penitentiary inmates warranted a diagnosis of organic brain syndrome (which is a much broader term than traumatic brain injury). Similarly the proportions of patients with organic brain syndrome in forensic hospitals is below 10 per cent and not greater than that in general psychiatric hospitals.[2,5]

Several studies suggest that the criminality of individuals with brain injury may, to a large extent, be attributed to premorbid personality traits, to the social disintegration which follows the injury, and hence not only to the injury itself. Kreutzer *et al.*[12] studied a sample of 327 patients with varying severities of traumatic brain injury. Those arrested after the brain injury were more likely to have had a history of police contacts before the brain injury, than those who were not arrested.

Two disorders have to be presented in greater detail:

Epilepsy, because it was historically one of the disorders of great concern for forensic psychiatrists and served as a model of the mentally ill offender not responsible for his crimes, and ADHD, because it is one of the disorders for which a relationship to antisocial behaviour and offending is most intensively researched.

Epilepsy and offending

Throughout history epilepsy has been associated with violence. Devinsky and Bear[13] observed 'it would be difficult to cite, either from case reports or a literature review, another medical or neurologic illness in which aggressive behaviour is described so regularly'. Not only seizures were frightening for lay people and caused them to consider epileptics as being cursed by gods or being possessed by witches (Malleus Maleficarum, 1487) and dangerous to others, these patients were seen as threat because of their personality changes. At the turn of the twentieth century most lay persons and professionals believed that people with epilepsy had pathological personality traits and displayed aggression, sociopathy, and psychosis.[14] Kraepelin too reported aggression in epileptic patients and mentioned that almost always an intensification of mental irritability occurs. Jackson took it as given that epilepsy was a cause of insanity '…often of a kind that brings epileptics in conflict with the law'.[15] Even in 1973 Sjöbring[16] noted, that patients suffering from epileptic seizures become torpid and circumstantial, sticky and adhesive, effectively tense, and 'suffer from explosive outburst of rage, anxiety and so on'.

Epidemiologic research,[17] literature reviews,[18] and experimental studies[19] have not supported these beliefs. Although epilepsy was found to be three to four times more frequent among prisoners in the United Kingdom than in the general population,[17] their offences did not differ from those of the rest of the prison population. Similar findings were reported from the United States (King and Young, 1978). In a extensive survey of mentally ill

offenders in Germany, who had committed acts of violence, Häfner and Böker[5] found only 29 patients with epilepsy out of 533 hospitalized violent offenders (5.4 per cent of the total sample). They compared their sample to an unselected population of 3392 non-violent mentally ill hospital patients and reported that 5.2 per cent of them had also received the diagnosis of epilepsy. They concluded that epilepsy was statistically not a risk factor for violence. A thorough analysis of the crimes of the epileptic patients showed that marital status (single), educational level, socio-economic state, and alcohol consumption were more important risk factors than epilepsy. This is in accordance with studies in other countries. Eight of the 29 patients in the Häfner and Böker study had committed their crimes in an epileptic confusional state (which corresponds to the medico-legal term of organic automatism), but 11 had a quarrel with their victim before their offence.

Although specific personality changes were not confirmed for epileptics in general, and some authors attributed them rather to institutionalization, brain damage, comorbid disorders, medication, or social changes than to epilepsy itself, a specific association has been made between temporal lobe epilepsy (TLE) and special personality traits. 18 characteristic personality traits were summarized from the literature to constitute the Gastaut–Geschwind syndrome,[20] among them aggression, emotional lability, and 'hypomoralism', traits that must be considered as risk factors for offending. The empirical and neurobiological database to support an increased risk of criminality or violence even in these patients is, however, small. MRI studies found that severely aggressive epileptic patients were characterized by severe amygdalar atrophy or by left temporal lesions affecting the amygdala and the brain regions around them,[21] confirming the assumption that violence has to be attributed to specific brain damage rather than to epilepsy itself.

Epilepsy is associated with a number of psychopathological and behavioural symptoms, among them mood changes, anxiety, rigidity, and aggression. Epidemiologic studies show a high comorbidity with cognitive impairment, ADHD, personality disorders and psychotic disorders. Behavioural abnormalities and comorbid disorders, as well as epilepsy itself can contribute to the risk of offending in these patients.

Summarizing the results from newer empirical studies, epilepsy does not increase the risk for offending or violence, and the number of cases, in which epilepsy was successfully used to claim inculpability after offending violently is small.[22] Forensic psychiatrists confronted with the assessment of epileptic patients who have committed crimes have to consider the following key questions:[23]

- Is the association between offending and epilepsy due to the occurrence of the epileptic seizure itself?
- Is it due to the associated brain damage that may be the cause of the seizure?
- Is it the result of socio-economic factors or of medication?
- Is the offending independent of the epilepsy and due to other criminogenic factors?

There are theoretically several possibilities, why epileptic patients could offend or act violently as a consequence of the disorder:

1 Aggression or impulsive acts could be a manifestation of a seizure or an equivalent to a seizure (violent automatism); it could also be the reaction of the patient to negative aura experiences. Delgado-Escueta *et al.*[24] collected 5400 videotaped seizures and concluded that violence appears to be extremely rare event in epileptic seizures. Only 13 individuals in their sample acted violently and only three attacked other people. This study was criticized because it did not take the patient environment interaction into account, which also plays a major role in outbreaks of aggression and which can be modified by epileptic seizures. Nevertheless, the number of cases of ictal aggression reported in the literature is small. Ictal violence erupts out of a normal non-aggressive situation within seconds, appears to be inappropriate to the circumstances, lasts for 1 to 3 min and subsides as suddenly as it has erupted. The patient returns to consciousness immediately or returns to normality after a few minutes of confusion, appears to be puzzled over what has happened and has no memory of it. Complex criminal acts cannot be attributed to seizures, seizure equivalents, or epileptic automatism.

2 Violence can be the result of tension and irritability in the prodromal phase of an epileptic seizure. In this case the irritability usually proceeds in waves, often triggered by the reactions of bystanders. Violence appears to be goal-directed and limited to avoid personal harm. The aggressive tension may last for several minutes up to an hour. Sometimes the patients seem out of control and do not remember their behaviour at al.

3 Postictal confusion can be associated with poriomania, somnambulism, and offending. Again, complex acts and adequate reactions to new situations cannot be attributed to postictal confusion or twilight states.

4 Interictal offending is quite often not associated with the disorder itself but with a criminal or aggressive family background, lower socio-economic status, and brain trauma.[23,25] Violence can be due to epileptic psychosis, to cognitive impairment, to impulse control disorder or to personality disorder, all of which could be consequences of epilepsy, although the data on these sequels of epilepsy are not undisputed.

A recent review of the literature by Schachter[26] can be summarized to the fact that postictal aggression was rarely due to confusion but significantly more frequent in postictal psychosis, and interictal aggression was associated rather with male sex, brain damage, social disadvantages, and chronic behavioural difficulties than with specific EEG findings or characteristic brain scans. In conclusion several risk factors have to be taken into account when the offending or violence of epileptic patients is considered:

Disorder-related risk factors

- Brain damage
- Early onset of epileptic seizures
- Postictal psychosis

Disorder-unrelated risk factors

- Male sex
- Growing up in a criminal or aggressive environment
- Low socio-economic class
- Antisocial personality disorder
- Alcohol abuse

In most legislations forensic psychiatrists and courts would agree, that the criteria proposed by Hindler,[27] would be minimal requirements to relate offending or violence to epilepsy:

- An unequivocal past history of epileptic attacks
- The crime is out of character with the person's previous personality
- The crime is motiveless and unpremeditated
- EEG studies are compatible
- An altered state of consciousness during the event
- Total or partial amnesia for the crime

More stringent criteria were proposed by a panel of epileptologists in the United States in 1981.[24] They included videotaped documentations of epileptic automatism and of the presence of aggression during such seizures; also the aggressive or violent acts should be characteristic of the patient's habitual seizures. In most Central European countries videotaped proof would not be necessary to attribute the offending of an epileptic to his disorder.

Attention-deficit/hyperactivity disorder (ADHD) and offending

Attention-deficit/hyperactivity disorder (ADHD) is a relatively new diagnostic term. It has been introduced in the current form in DSM-III-R in 1987, although the condition has been known to psychiatrists for more than 100 years.[28] Only with DSM-IV-TR some allowance was made to extend the diagnosis to adults[29] but DSM-IV-TR continued to assert that the majority of the patients loose their symptoms during late adolescence. The diagnostic criteria were empirically evaluated in school-children and it can be questioned whether they are appropriate for adolescents and adults. It can be questioned even more, how the association between offending and ADHD can be adequately established. The core symptoms of ADHD, inattention, hyperactivity, and impulsivity, are likely to cause social conflicts with peers and caregivers and to lead to resentment and aggressive interactions. This dissocial functioning may appear as the precursor of offending in later life. A number of studies tried to investigate the relationship between ADHD in childhood and the diagnosis of antisocial personality disorder (APD) in adulthood. Although there is converging evidence that about 25 per cent of children with ADHD will later receive the diagnosis of APD,[30,31] and that in individuals who met the criteria of psychopathy according to Hare,[32] ADHD was diagnosed retrospectively four times more often than in the general population.[33] Most of the studies ignore the possible impact of comorbid disorders, which are especially frequent with ADHD. Conduct disorder (CD), oppositional defiant disorder (ODD), alcohol dependence or abuse and other drug dependence or abuse are significantly more frequent in ADHD individuals than in control groups,[34] and all contribute substantially to offending and to the diagnosis of APD. The longitudinal study of Satterfield & Schell[35] found that half of the children with a combination of ADHD and CD were later diagnosed as APD, while APD was only found in 12 per cent of those who had ADHD without CD in their childhood.[36]

A number of studies found ADHD largely overrepresented in prison populations, where it is calculated to be up to 10 times more frequent than in the general population.[37,38] Except for drug-related crimes,[30] no special criminal profile could be attributed to the disorder.

In psychiatric assessments for courts and tribunals ADHD is rarely considered as a cause for diminished responsibility, and most offenders with this disorder will not be sent to forensic psychiatric institutions. ADHD has, however, to be regarded as risk factor for reoffending, especially if it is or was combined with a conduct disorder. Prisoners with ADHD had been reconvicted four times more often than other prisoners.[39]

Further information

Barkley, R.A. (2006). *Attention-deficit/hyperactivity disorder* (3rd edn). The Guilford Press, New York.

Etinger, A.B. and Kanner, A.M. (2007). *Psychiatric issues in epilepsy* (2nd edn). Wolters Kluwer, Philadelphia.

Nedopil, N. (2000). Offenders with brain damage. In *Violence, crime and mentally disordered offenders: concepts and methods for effective treatment and prevention* (eds. S. Hodgins and R. Müller-Isberner), pp. 38–42. Wiley & Sons, Chichester.

References

1. Gunn, J. (1991). Human violence, a biological perspective. *Criminal Behaviour and Mental Health*, **1**, 34–54.
2. Grant, I. (1997). Canada's new mental disorder disposition provisions: a case study of the British Columbia criminal code review board. *International Journal of Law and Psychiatry*, **20**, 419–43.
3. Leygraf, N. (1988). *Psychisch kranke Rechtsbrecher*. Springer, Berlin.
4. Fava, M. (1997). Psychopharmacologic treatment of pathological aggression. *The Psychiatric Clinics of North America*, **20**, 427–51.
5. Häfner, H. and Böker, W. (1982). *Crimes of violence by mentally abnormal offenders. The psychiatric epidemiological study in the Federal German Republic*. Cambridge University Press, Cambridge.
6. D'Orban, P., Gunn, J., Holland, A., *et al.* (1993). Organic disorder, mental handicap and offending. In *Forensic psychiatry* (eds. J. Gunn and P.J. Taylor), pp. 286–328. Butterworth-Heinemann, Oxford.
7. Toone, B. (1990). Organically determined mental illness. In *Principles and practice of forensic psychiatrie* (eds. R. Bluglass and P. Bowden), pp. 385–92. Churchill Livingstone, Melbourne.
8. Nedopil, N. (2000). Offenders with brain damage. In *Violence, crime and mentally disordered offenders: concepts and methods for effective treatment and prevention* (eds. S. Hodgins and R. Müller-Isberner), pp. 38–42. Wiley & Sons, Chichester.
9. Rabins, P.V., Mace, N.L., and Lucas, M.J. (1982). The impact of dementia on the family. *The Journal of the American Medical Association*, **248**, 333–5.
10. Diehl, J., Ernst, J., Krapp, S., *et al.* (2006). Frontotemporale Demenz und delinquentes Verhalten. *Fortschritte der Neurologie-Psychiatrie und ihrer Grenzebiete*, 203–10.
11. Hodgins, S. (1995). Assessing mental disorder in the criminal justice system. *International Journal of Law and Psychiatry*, **18**, 15–28.
12. Kreutzer, J.S., Marwitz, J.H., and Witol, A.D. (1995). Interrelations between crime, substance abuse, and aggressive behaviours among persons with traumatic brain injury. *Brain Injury*, **9**, 757–68.
13. Devinsky, O. and Bear, D.A. (1984). Varieties of aggressive behavior in temporal lobe epilepsy. *The American Journal of Psychiatry*, **141**, 651–6.
14. Devinsky, O., Vorkas, C.K., and Berr, W. (2007). Personality disorders in epilepsy. In *Psychiatric issues in epilepsy* (2nd edn) (eds. A.B. Etinger and A.M. Kanner), pp. 286–305. Wolters Kluwer, Philadelphia, Baltimore, New York, London.
15. Jackson, J.H. (1931). *Selected writings of John Hughlings Jackson*. Hodder and Stoughton, London.

16. Sjöbring, H. (1973). Personality structure and development, a model and its application. *Acta Psychiatrica Scandinavica Supplementum*, **244**, 1–204.
17. Gunn, J. (1977). *Epileptics in prison*. Academic Press, London.
18. Treiman, D. (1986). Epilepsy and violence: medical and legal issues. *Epilepsia*, **27**, 77–102.
19. Rodin, E.A. (1973). Psychomotor epilepsy and aggressive behavior. *Archives of General Psychiatry*, **28**, 210–3.
20. Bear, D.M. and Fedio, P. (1977). Quantitative analysis of interictal behavior in temporal lobe epilepsy. *Archives of Neurology*, **34**, 454–67.
21. van Elst, L.T., Woerman, F.G., Lemieux, L., *et al.* (2000). Affective aggression in patients with temporal lobe epilepsy: a quantitative MRI study of the amygdala. *Brain*, **123**, 234–43.
22. Treiman, D.M. (1999). Violence and the epilepsy defense. *Neurologic Clinics*, **17**, 245–55.
23. Herzberg, J.L. and Fenwick, P.B.C. (1988). The aetiology of aggression in temporal-lobe epilepsy. *The British Journal of Psychiatry*, **153**, 50–5.
24. Delgado-Escueta, A., Mattson, R., and King, L.M. (1981). The nature of aggression during epileptic seizures. *The New England Journal of Medicine*, **305**, 711–6.
25. Stevens, J.R. and Hermann, B. (1981). Temporal lobe epilepsy, psychopathology and violence: the state of evidence. *Neurology*, **31**, 1127–32.
26. Schachter, S.C. (2007). Aggression in epilepsy. In *Psychiatric issues in epilepsy* (2nd edn) (eds. A.B. Etinger and A.M. Kanner), pp. 306–20. Wolters Kluwer, Philadelphia.
27. Hindler, C.G. (1989). Epilepsy and violence. *The British Journal of Psychiatry*, **155**, 246–9.
28. Hechtman, L. (2004). Attention-deficit/hyperactivty disorder. In *Kaplan & Sadock comprehensive textbook of psychiatry* (8th edn) (eds. B.J. Sadock and V.A. Sadock), pp. 3183–98. Lippincott, Williams & Wilkins, Baltimore.
29. McGough, J.J. (2004). Adult manifestations of attention-deficit/hyperactivity disorder. In *Kaplan & Sadock comprehensive textbook of psychiatry* (8th edn) (eds. B.J. Sadock and V.A. Sadock), pp. 3198–204. Lippincott, Williams & Wilkins, Baltimore.
30. Barkley, R.A., Fischer, M., Smallish, L., *et al.* (2004). Young adult follow-up of hyperactive children: antisocial activities and drug use. *Journal of Child Psychology and Psychiatry*, **45**, 195–211.
31. Mannuzza, S., Klein, R.G., Bessler, A., *et al.* (1993). Adult outcome of hyperactive boys: educational achievement, occupational rank, and psychiatric status. *Archives of General Psychiatry*, **50**, 565–76.
32. Hare, R.D. (1991). *Manual for the Hare psychopathy checklist-revised*. Multi-Health-Systems Inc., Toronto.
33. Johannsson, P., Kerr, M., and Andershed, H. (2005). Linking adult psychopathy with childhood hyperactivity-impulsivity. Attention problems and conduct problems though retrospective self-reports. *Journal of Personality Disorders*, **19**, 94–101.
34. Barkley, R.A. (2006). *Attention-deficit/hyperactivity disorder* (3rd edn). The Guilford Press, New York.
35. Satterfield, J.H. and Schell, A. (1997). A prospective study of hyperactive boys with conduct problems and normal boys: adolescent and adult criminality. *Journal of Child and Adolescent Psychiatry*, **36**, 1726–35.
36. Mannuzza, S., Klein, R.G., Bessler, A., *et al.* (1998). Adult psychiatric status of hyperactive boys grown up. *The American Journal of Psychiatry*, **155**, 493–8.
37. Rösler, M., Retz, W., Retz-Junginger, P., *et al.* (2004). Prevalence of attention-deficit/hyperactivity disorder in male young prison inmates. *European Archives of Psychiatry and Clinical Neuroscience*, **254**, 365–71.
38. Vermeiren, R. (2003). Psychopathology and delinquency in adolescents: a descriptive and developmental perspective. *Clinical Psychology Review*, **23**, 277–318.
39. Ziegler, E., Blocher, D., Groß, J.,*et al.* (2003). Assessment of attention-deficit/hyperactivity disorder in prison inmates. *Recht and Psychiatrie*, **21**, 17–21.

11.4

Mental disorders among offenders in correctional settings

James R. P. Ogloff

Incontrovertible evidence now exists to show that the prevalence of mental disorders among prisoners far exceeds that found in the general community. A surprising concordance is emerging from several large international studies to show that, in western developed societies at least, the rates of major mental disorders in prisons are quite consistent. This chapter will provide an overview of relevant research examining rates of mental illness in prisons with those found in the community. Some observations regarding trends and implications for prisons also will be provided.

At the outset it is useful to reflect on the scope of illnesses which have been subsumed under the 'mental disorder' umbrella as it has been applied to the prison research. Most of the research that exists has focussed serious mental illnesses within the Axis I disorders—namely psychotic illnesses, mood disorders, and anxiety disorders. Considerable attention has been paid regarding the prevalence of personality disorders within prisons. Over the past 20 years much of that work has investigated antisocial or dissocial personality disorder and psychopathy.[1] By comparison, relatively little attention has been paid to other personality disorders. A growing area of importance concerns substance abuse and dependence disorders and, of course, co-occurring substance use and mental illness disorders. Considerable research also exists exploring the prevalence of mental retardation or intellectual disabilities in prison. Thereafter, fragments of research exist exploring any number of mental syndromes and conditions. The focus of this chapter will be on the major mental disorders which fall into Axis I. Some mention will be made of substance use disorders and personality disorders. In addition, with the growing number of women in prisons, information will be provided regarding this important group.

The prevalence of mental illness among male and female prisoners

Recent research exists from Britain and Wales that shows that the prevalence of mental illness in prisons is many times greater than that found in the community. Brugha and colleagues[2] compared rates of mental illness among some 3,000 remanded and sentenced male and female prisoners in Britain and Wales and more than 10,000 community residents in Great Britain. The rate of psychotic illnesses in the community was 4.5 per 1,000 (0.045 per cent) compared with 52 per 1,000 in prisons (0.52 per cent). While the ten-fold increase in prevalence from the community to the prisons was remarkable, the results revealed further that the prevalence rate of psychotic illness for female prisoners was an astonishing 110 per 1,000 (0.11 per cent), compared to 50 per 1,000 for males (0.05 per cent).

One of the key studies that has helped provide information regarding the rate of mental illness in gaols and prisons is a meta-analysis conducted by Fazel and Danesh[3] that was published in *The Lancet*. Their analyses included 62 studies that included 22,790 prisoners. The majority of prisoners (81 per cent) were male. Nonetheless, enough studies that included women prisoners were available to provide information regarding the rate of mental disorder among them. Data from the Fazel and Danesh meta-analysis are presented in Table 11.4.1. The results show that approximately one in seven prisoners have a psychotic illness or major depression. As the authors report, this is between two and four times greater than would be expected in the general population.

Table 11.4.1 Representative prevalence of mental illness and personality disorder among male and female prisoners (international samples)

Disorder	Males per cent (95 per cent C.I.)	Females per cent (95 per cent C.I.)
Psychotic Illness (k = 49, N = 19,011)	3.7 (3.3–4.1)	4 (3.2–5.1)
Major Depression (k = 31, N = 10,529)	10 (9–11)	12 (11–14)
Personality Disorder (k = 28, N = 13,844)	65 (61–68)	42 (38–45)
Antisocial Personality Disorder	47 (46–48)	21 (19–23)

k = number of studies; N = number of subjects; C.I. = Confidence Intervals.

Source: Fazel & Danesh (2002).

With respect to personality disorders, half of males and approximately 20 per cent of females are found to have a personality disorder—which is ten times greater than would be seen in the community.

There was some variability across studies, some (but not all) of which was explained by differences between research that used validated diagnostic procedures (3.5 per cent) and those that did not (4.3 per cent). Studies from the USA also showed higher prevalence rates than elsewhere. Psychosis among female prisoners was found to be slightly higher than that in males (4.0 per cent *c.f.* 3.7 per cent).[3]

A limitation of the Fazel and Danesh[3] meta-analysis is that relatively limited information was provided regarding the type and nature of mental illness. Brinded, Simpson, Laidlaw, Fairley, and Malcolm[4] reported the results of one of the most well conducted studies on the prevalence of mental illnesses among inmates ever published. *All* female sentenced and remanded inmates and a random sample of 18 per cent of sentenced male inmates in New Zealand were interviewed. Interviewers used standardized measures to identify inmates with mental illnesses and personality disorders. The final sample consisted of approximately 1 200 inmates. The results of prevalence rates for mental disorder in the last month are presented in Table 11.4.2.

As the results in Table 11.4.2 show, the prevalence rates obtained by Brinded and colleagues[4] in New Zealand essentially parallel those obtained by the Fazel and Danesh[3] meta-analysis. The New Zealand results, however, include data for post-traumatic stress disorder, substance abuse, and dependence disorders. As with psychosis and major depression, the prevalence rates of the other disorders is significantly greater than what would be seen in the general population.

Of late, increased attention is being paid to the prevalence of mental disorders among female inmates.[5] In recent years, the rate of growth among women in custody has far surpassed the growth rate for male prisoners. For example, data show that in the 10 years ending 2005, the percentage of women in prisons in the United States has increased by 57 per cent, compared to a growth rate of 34 per cent for men during the same period.[6] Similar findings exist in Australia, where research shows that the number of women in prison increased by 66 per cent from 1991 to 1999, while it increased by 24 per cent for men during the same period.[7]

Even more than for male prisoners, 'the prevalence of childhood and adulthood sexual and violent victimization, poverty, and poor educational and employment attainment reported by female inmates is nothing short of alarming'.[8] While a comprehensive review of the studies of mental illness among women offenders is beyond the scope of this chapter, Ogloff and Tye[5] have shown that the prevalence rates of mental disorders for women is now surpassing those identified for men in most published studies. This is particularly the case for mood disorders and anxiety disorders.

Table 11.4.2 Prevalence rates for mental disorder in last month (New Zealand samples)

Diagnosis	Women N = 167 n (per cent)	Remanded Men N = 441 n (per cent)	Sentenced Men N = 636 n (per cent)
Mental Illness			
Schizophrenia and related disorders	7 (4.2)	15 (3.4)	14 (2.2)
Bipolar affetive disorder	2 (1.2)	4 (1.0)	7 (1.1)
Major depression	18 (11.1)	47 (10.7)	38 (5.9)
Obsessive-compulsive disorder	7 (4.3)	22 (5.0)	21 (4.8)
Posttraumatic stress disorder	27 (16.6)	42 (9.5)	55 (8.5)
Substance-Related Disorders			
Alcohol abuse	7 (4.3)	25 (5.7)	8 (1.2)
Alcohol dependence	4 (2.5)	19 (4.3)	3 (0.5)
Cannabis abuse	6 (3.7)	38 (8.6)	27 (4.2)
Cannabis dependence	0 (0)	0 (0)	0 (0)
Other abuse/dependence .	6 (3.7)	27 (6.1)	12 (1.9)

Reproduced from Brinded *et al.* Prevalaence of pyschiatric disorders in New Zealand Prisons: A national study. *Australian and New Zealand Journal of Psychiatry*, **35**, 166–73, copyright 2001, John Wiley & Sons, Inc.

Implications, service needs, opportunities

The cause of the relatively high prevalence of mental illness among people in the prison system has been sometimes attributed to the deinstitutionalization movement that has occurred in mental health over the past 20 years. The contention that the mentally ill are entering gaols in increasing numbers has not been accepted by all, however.[9] It has been proposed that it is simply heightened awareness among professionals and the public of the problem of mentally ill in the gaols that has resulted in the perception that they are entering in increasing numbers.[10] In a recent study investigating the criminal offence history of every person in Victoria with schizophrenia in the public mental health registry in five year cohorts from 1975 to 1995, Wallace, Mullen, and Burgess[11] found that there was no subsequent increase in offence rate by year for those with schizophrenia, while the offence rate for the matched comparison group of people in the community without a mental illness increased significantly over the period. This is particularly interesting since during that time the process of deinstitutionalization was completed in Victoria. Indeed, there are no more psychiatric hospitals in Victoria (except for a 100 bed secure forensic psychiatric hospital).

A number of contributing factors have been identified that help explain the high numbers of people with mental illnesses in the criminal justice system. Considerable concern has been raised about the capacity of community-based mental health services to address the needs of mentally ill offenders. Community-based mental health services work best for those who have reasonable connections and support within the community. Unfortunately, offenders (especially imprisoned offenders) tend to be poorly integrated into the community[12] and have poor access to a range of support services including accommodation, income support, health and mental health.[13, 14]

While the presence of mentally ill people in the criminal justice system presents challenges and raises concerns, the fact is that the justice system provides an opportunity to identify and deliver treatment to people who are otherwise likely to remain outside the reach of services. As such, it has been suggested that justice mental health services present an opportunity for identifying those with mental illnesses and making services available to them that would otherwise be non-existent.[10] Accordingly, taking a population

health perspective, efforts to identify and treat those with mental illnesses who are entering the criminal justice system can help provide much needed services to this otherwise under-serviced population.

Service development and provision to mentally ill offenders in prisons

Although it is beyond the scope of the chapter to discuss the provision of mental health services to prisoners, it may be helpful here to outline some of the requirements for services. The service model outlined here has been detailed elsewhere in the literature.[10] The service model recommended for use in prisons consists of six components outlined below, the nature and extent of which will vary depending on the needs arising in each institution.

Intake screening

A two-tier evaluation process is recommended. The first step involves a brief mental health screening for every inmate upon admission. Second, those prisoners identified as being mentally ill are referred to mental health professionals for a more complete assessment. All prisoners should be screened for mental illness soon after admission to a correctional facility (within the first 24 h). The *Jail Screening Assessment Tool (JSAT;* Nicholls *et al*., 2005) was developed to screen people being admitted to gaols for mental illness, as well as self-harm risk and risk of harm to others. The JSAT is administered by psychiatric nurses or other mental health professionals and takes approximately 20 minutes. It is validated for both male and female prisoners.

Ongoing monitoring/screening of prisoners

A process must be in place for ensuring that prisoners are monitored, both formally and informally. This should include self-referrals and referrals by all prison staff.

Comprehensive psychodiagnostic assessment

Comprehensive assessments by psychiatrists or clinical psychologists are required for all prisoners exhibiting symptoms of mental illness. The examination should include consideration of whether the prisoners who are acutely mentally ill should receive treatment in the institution or be transferred to hospital.

Mental health treatment

Once assessed as having a mental illness, the prisoner should be referred to an appropriate treatment program within the correctional facility or correctional system if possible and practical. The size of the gaol and its mandate affects the type of service available. Services should be at the standard available in the community.

Gradual post-release monitoring/supervision and continuity of services

Treatment should continue post-release in the community. The transition back to the community is often difficult, as evidenced by high recidivism rates and ongoing illness. Service needs for mentally ill people leaving prison include initiating psychiatric treatment and psychosocial services with a community mental health agency, locating housing, and finding employment.

Programme evaluation

Given the complexity of mental health services in prisons, it is critical that programmes are evaluated on an ongoing basis (e.g. Elliot, 1997). Such evaluations are as important as the other components in achieving assessment and delivery of mental health services. Also, wherever possible, assessment must be linked with treatment. Ongoing evaluations of the effectiveness of the assessment/treatment decisions should be built into the system. Evaluation informs decision-makers about the outcome of their decisions. Over time, this feedback can lead to improvements in the assessment, referral, and treatment phases of the model. Data on the base rates of mental disorder among women in custody-and the number of female prisoners who fall into the MDO categories—also can prove valuable in planning for future treatment needs.

Further information

Fazel, S., and Danesh, J. (2002). Serious mental disorder in 23 000 prisoners: a systematic review of 62 surveys. *Lancet*, **359**, 545–50.

Ogloff, J.R.P (2002). Identifying and accommodating the needs of mentally ill people in gaols and prisons. *Psychiatry, Psychology, and Law*, **9**, 1–33.

Sheehan, R., McIvor, G., Trotter, C. (eds) (2007). *What works with women offenders*. Willan Publishing, Devon, UK.

References

1. Ogloff, J.R.P. (2006). The Psychopathy/Antisocial Personality Disorder conundrum. *Australian and New Zealand Journal of Psychiatry*, **40**, 519–28.
2. Brugha, T., Singleton, N., Meltzer, H., *et al.* (2005). Psychosis in the community and in prisons: A report from the British National Survey of Psychiatric Comorbidity. *American Journal of Psychiatry,* **162**, 774–80.
3. Fazel, S., Danesh, J. (2002). Serious mental disorder in 23 000 prisoners: a systematic review of 62 surveys. *Lancet*, **359**, 545–50.
4. Brinded, P.M.J., Simpson, A.I.F., Laidlaw, T.M., (2001). Prevalence of psychiatric disorders in New Zealand prisons: A national study. *Australian and New Zealand Journal of Psychiatry*, **35**, 166–73.
5. Ogloff, J., Tye, C. (2007). Responding to mental health needs of women offenders. In *What works with women offenders.* (eds. R. Sheehan, G. McIvor, C. Trotter), pp. 142–81 Willan Publishing, Devon, UK.
6. Harrison, P. and Beck, A. (2006). *Prisoners in 2005*. Washington, DC: US Bureau of Justice Statistics Bulletin NCJ 215092.
7. Cameron, M. (2001). Women Prisoners and Correctional Programs, *Trends and Issues in Crime and Criminal Justice*, 194, Canberra, ACT: Australian Institute of Criminology.
8. Nicholls, T., Lee, Z., Corrado, R., *et al.* (2004). Women inmates' mental health needs: Evidence of the validity of the *Jail Screening Assessment Tool (JSAT), International Journal of Forensic Mental Health*, 3(2),167–84.
9. Monahan, J., Caldeira, C., Friedlander, H. (1979). The police and the mentally ill: A comparison of arrested and committed persons. *International Journal of Law and Psychiatry*, **2**, 509–18.
10. Ogloff, J.R.P. (2002). Identifying and accommodating the needs of mentally ill people in gaols and prisons. *Psychiatry, Psychology, and Law*, **9**, 1–33.
11. Wallace, C., Mullen, P.E., Burgess, P. (2004). Criminal offending in schizophrenia over a twenty-five year period marked by

deinstitutionalization and escalating comorbid substance abuse. *American Journal of Psychiatry,* **161**, 716–27.

12. Makkai, T., McGregor, K. (2003). Drug Use Monitoring in Australia: 2002 Report on Drug Use Among Police Detainees. *Research and Public Policy Series* No. 47. Australian Institute of Criminology, Canberra.
13. Baldry, E. (2003). Ex-prisoners, housing and social integration. *Parity*, **16**, 13–15.
14. Travis, J., Waul, M. (2003). From prison to home: *The effect of incarceration and re-entry on children, families, and communities*. The Urban Institute. Washington, DC.

11.5

Homicide offenders including mass murder and infanticide

Nicola Swinson and Jennifer Shaw

There is a widespread public perception of the mentally ill as violent.[1,2] Until the early 1980s there was a consensus view that patients with severe mental illness were no more likely to be violent than the general population. Emerging evidence from various countries over the past two decades, however, has established a small, yet significant, association between mental illness and violence.

Large-scale birth cohort studies, such as a 30 year follow-up of an unselected Swedish birth cohort, show a significantly increased risk of violent offences in men and women in the presence of major mental disorder.[3] Community epidemiological studies in New York[4] and in Israel[5] again show an increased risk of violence in psychiatric patients. An important contribution to this field is data from the Epidemiological Catchment Area study, showing that major mental illness increases the rates of violence over a 12-month period from a 2 per cent base rate to 8 per cent, but co-morbid substance abuse increases this rate further to 30 per cent.[6] Co-morbid substance abuse and personality disorder substantially increase the risk of violence, as demonstrated in the MacArthur Risk Assessment Study which showed rates of violence in discharged psychiatric patients of 18 per cent in those with major mental disorder, 31 per cent with major mental disorder and co-morbid substance abuse, and 43 per cent in those with personality disorder and co-morbid substance abuse.[7]

Public fears are often fuelled by media reporting of high-profile cases of homicide by people with mental illness.[8] Despite indications that rates of homicide among the mentally ill are relatively constant across countries,[9] studies of mental disorder in people convicted of homicide show that 8.7 per cent of homicides in New Zealand are 'abnormal',[10] yet evidence from Canada indicates that 35 per cent of perpetrators are mentally unwell.[11] Indeed rates ranging from 8 to 70 per cent have been found, varying with different definitions of mental disorder.[12]

The National Confidential Inquiry into Suicide and Homicide by People with Mental Illness was established at the University of Manchester in 1996. The core work of the Inquiry is to establish rates of mental disorder in homicide and the clinical care received by those in contact with services.

General population homicides

There are 500–600 homicides annually in England and Wales. Perpetrators and victims are predominantly young males, especially when the victim is unknown to the perpetrator. In such 'stranger homicides' perpetrators are less likely to have a lifetime history of mental illness, symptoms of mental illness at the time of the offence, or contact with mental health services.

In the UK the total number of both total homicides and stranger homicides increased between1973 and 2003 but neither category increased in people with mental illness.[13] Similar trends have been noted in work from both the UK,[14] and in other countries.[10]

The commonest method of homicide is with a sharp instrument; shooting is relatively rare, accounting for less than 1 in 10 homicides in the UK.

Around half of all convictions are for murder and just under half for manslaughter. One in 25 receives a verdict of Section 2 manslaughter, diminished responsibility.

Infant homicide

Despite an increasing rate of homicides in the general population, convictions for infanticide and the rate of infant homicide has remained relatively constant, at around 4.5 per 100 000 live births.[15] Infanticide has become a generic term for killing of infants, even though the criminal charge in England applies to a crime for which only a woman can be indicted.

Although the risk of homicide is higher in the first year of their life than at any other time, the rarity of infant homicide in absolute numbers means that there is a lack of high quality, systematic data at a population level which incorporates clinical characteristics.[16]

Data from the National Confidential Inquiry from 1996 to 2001 shows that 1 in 25 of the 2665 homicide perpetrators identified were convicted of infant homicide. Half of these infants were killed by their father and around a third by their mother. A quarter of perpetrators had symptoms of mental illness at the time of the offence and a third had a lifetime history of mental illness. Perpetrators of neonaticide were predominantly young, unmarried mothers experiencing symptoms of dissociation at the time of the homicide.

There were significant differences between male and female perpetrators, with males being more likely to have previous convictions for violent offending. Females were more likely to kill within a month of the birth and they were more likely to have affective disorder and symptoms of mental illness at the time of the offence but few of these women were under the care of mental health services.

Most males received a custodial sentence, whereas three quarters of women received a community sentence or hospital disposal.[16]

Multiple homicides

Multiple homicides, in particular serial homicides, have generated a great deal of public and media interest over recent decades yet this phenomenon is rare in the UK. The rarity of these events means that there is a lack of empirical evidence about the characteristics of perpetrators and victims in the UK, with most evidence emanating from the United States. Even then, however, there is an absence of systematic, robust evidence, with many studies being limited by small sample size.

Most definitions of multiple homicides include three criteria; number of victims, which can vary from 2 to 10 in different definitions,[17] time, and motivation. The temporal relationship distinguishes subcategories: mass murder consisting of a single episode and location, with serial, and spree murders occurring over time in separate locations. The latter two are differentiated by an emotional 'cooling-off' period, which is present in serial homicide. Other authors have discussed motivation, such as sexual gratification and internal psychological gratification, but the lack of robust evidence means that it seems premature to include motivation as part of any such definition.[17]

Mass murder has been classified by victim type such as family annihilators and classroom avengers. Mullen[18] proposed a category of 'autogenic (self-generated) massacre', which encompassed perpetrators indiscriminately killing people in pursuit of a highly personal agenda, arising from their own specific social situation and psychopathology. They were characterized by social isolation, being bullied in childhood and personality traits such as suspiciousness, obsessional behaviour, grandiosity, and persecutory beliefs. He concluded that these murders are essentially murder-suicides, where the intention is to kill as many people as possible before killing themselves. It would now appear, particularly with recent events in Virginia, that this form of multiple homicide is an established form and concerningly appears to becoming more common. Cantor *et al.*[19] propose that media-related modelling is a potential factor in the emergence of this crime, with perpetrators often seeing themselves as lone warriors, themselves modelled on media images, and well informed about previous, similar, massacres.

An exploratory study, incorporating a nested case control study, showed that serial homicide offenders were more likely to be male dominated, compared with single homicide offenders, and were more likely to use strangulation. Moreover, victims of serial murders were significantly more likely to be females who were unknown to the perpetrator and the motivation being sexual.[17] Unfortunately most classification systems of serial murder, including the FBI classification, have been criticized as being inherently flawed due to weak operational definitions and unsubstantiated assumptions regarding behaviour and characteristics.[20]

There is, unfortunately, a lack of robust evidence regarding multiple homicides. There seem to be clear similarities between serial and mass murderers, but also fundamental differences. Mass murderers appear more likely to use firearms as a method, whereas serial murderers tend to kill in a more personal manner, using methods that afford greater physical proximity, such as strangulation, in addition to a greater propensity for female victims. Clinically, some evidence indicates that a substantial proportion of mass murderers have a severe mental illness, often a psychotic illness. On the other hand, it is proposed that serial murderers can be distinguished by lower levels of severe mental illness and the presence of higher degrees of psychopathy.[21]

Female perpetrators of homicide

Around 1 in 10 perpetrators of homicide in England and Wales are female,[13] which is consistent with data from other countries, such as Finland.[22] Stranger homicide by females is rare. In one-quarter of cases the victims are the perpetrators' own children and a current or former partner in over a third.

As with men the commonest method is stabbing, although females are proportionally more likely to use suffocation or poisoning when compared with men.

There are no clear gender differences in the proportion of those with severe mental illness but females are proportionally more likely to have a diagnosis of alcohol or drug dependence than men. Females are less likely to receive a prison sentence and are more commonly placed on a hospital or community rehabilitation order.[13]

Homicide by older people

Homicides perpetrated by the elderly are exceptionally rare. In England and Wales they account for less than 1 in 50 homicides. The male to female ratio in perpetrators over 65 years is the highest of all age groups, at 19:1.[13]

There is a distinct lack of robust evidence regarding homicide in this population. Elderly spouse homicides have been described by Knight[23] as involving a couple perceived to have a close, caring relationship with the homicide of the wife, by the husband, occurring in an abrupt and unexpected manner. Depression is well recognized in elderly homicides, not infrequently with associated delusions of impoverishment and ruin. Perpetrators are often in care giving roles with physical or psychiatric disability in the victim. The homicide is often followed by suicide of the offender.[24]

Those over 65 years are more likely to receive a hospital order or community disposal, than a custodial sentence.[13]

Perpetrators of homicide with mental health service contact

The aim of the National Confidential Inquiry is to collect detailed clinical information on people convicted of homicide, focusing on those with a history of contact with mental health services.

The inquiry collects a national consecutive case series of patient homicides occurring since April 1996. Data collection involves collecting information on all homicides from the Home Office Homicide Index, which includes details of the perpetrator, victim, and method used. Where available, psychiatric reports prepared for the trial are obtained. Antecedent data (of previous offences) is collected from the National Crime Operations Faculty. Details on each case are submitted to mental health services in each individual's district of residence and adjacent districts to identify those with a history of mental health service contact. These individuals become Inquiry cases. Information on trust Inquiry cases is obtained from clinical teams via a comprehensive questionnaire sent to the consultant psychiatrist.

In the UK around 1 in 10 people convicted of homicide have been in recent contact with mental health services. In most of these cases the responsible service is a general adult psychiatry service, rather than a specialist service. The remaining cases are under alcohol and drug services, child and adolescent services, and forensic psychiatry services. Around one in five cases have had lifetime contact with services. This compares with data from other countries, such as Australia, where one in three perpetrators has had contact with psychiatric services.[25]

Of those in contact with mental health services in England and Wales, the most common diagnosis is schizophrenia, although less than half have severe mental illness (schizophrenia or affective disorder). There are high rates of co-morbid alcohol and drug dependence and personality disorder. Only one-third have previous admissions under the Mental Health Act (1983).

A high proportion of these patients have a history of violence, including convictions for violence, which, worryingly, are not documented in the case notes in a number of cases. Similar findings regarding the prevalence of violence were found in an examination of findings from public inquiries into homicides in the UK.[26] A small number of homicides are committed by patients who have previously been on a restriction order because of a violent offence.

Around half of those prescribed medication are non-compliant or disengaged from services at the time of the offence and relatively few are receiving any psychological intervention.[13]

Perpetrators with schizophrenia

There is a well documented increased risk of violence in those with schizophrenia.[27] This has been shown in studies from the UK[28] and in other countries such as New Zealand[29] and Denmark.[30]

Around 1 in 20 perpetrators of homicide have a diagnosis of schizophrenia, a half have been in recent contact with services and one-third have never had any service contact.[13] These findings are broadly consistent with other UK data from remand prisoners,[31] and with data from other countries with rates of schizophrenia ranging from 7 per cent in Finland,[32] 7.5 per cent in Australia[25] to 12.6 per cent in Canada.[11]

Of those with recent contact one-fifth have a secondary diagnosis, commonly personality disorder or substance dependence, and a history of violence is common. Nearly a half have a history of violence when psychotic, around one-quarter are psychotic at the time of the homicide. Victims are most commonly family members; in less than one in six cases the victim is a stranger. Similar rates of stranger homicide by the mentally ill are found in Australia and New Zealand.[10] The majority of these patients have symptoms of mental illness at the time of the homicide and one in four receives a verdict of diminished responsibility. Of those not in contact with services, the vast proportion are psychotic at the time of the offence.[13]

It is of concern that nearly one-third of all perpetrators with schizophrenia receive a prison disposal.

Despite clear evidence of an increased risk associated with schizophrenia it is important to present a balanced view to prevent unnecessary stigmatization. The proportion of violent crime in society which is attributable to schizophrenia is consistently less than 10 per cent.[27] Wallace *et al.*[25] showed an increased risk of serious violent offending in males with schizophrenia of five times that of the general population. However, he also highlights data which indicate that, in any given year, 99.97 per cent of all those with schizophrenia will not be convicted of a serious violent offence, and that the probability of patients with schizophrenia committing homicide is extremely low.

Risk assessment

Nearly one in three Inquiry cases were seen during the week before the homicide, a similar proportion within 1–4 weeks and the remainder between 1–12 months. A substantial proportion had mental state abnormalities at final contact, often distress, depressive symptoms, hostility, or increased use of alcohol or drugs. Despite this immediate risk was judged to be low or absent in 88 per cent cases at the last contact.

There are clear difficulties in predicting risk of serious violence, given the rarity of its' occurrence alongside the high prevalence of risk factors such as substance abuse and a history of violence within the patient population. In an examination of findings from public inquiries into homicides it was shown that only 28 per cent of homicides were judged 'predictable', yet 65 per cent were seen as 'preventable'. 'Preventability' was conferred by 'improved mental health care'.[26]

Use of enhanced CPA to manage risk

In the National Confidential Inquiry sample from 1999 to 2003, nearly three quarters of those with recent contact were not receiving care under the provisions of enhanced Care Programme Approach (CPA), including a substantial proportion of patients at high risk such as those with schizophrenia, personality disorder, a history of detention under Mental Health Act legislation, or a previous history of violence. Furthermore, one-third of those with severe mental illness, a history of violence, and detention under the Mental Health Act were not under enhanced CPA.

Among those who were being cared for under the provisions of enhanced CPA, a significant number were non-compliant with medication or disengaged from services at the time of the offence. It seems, therefore, that even if risk is recognized high-risk patients are not receiving the intensive care, commensurate with their level of risk, in the community.[13]

Preventability

Clinicians identified one case in five in recent contact where the homicide could potentially have been prevented. Factors viewed as increasing the chance of preventing the homicide included a diagnosis of schizophrenia, multiple previous admissions, and detention under the Mental Health Act. Factors which were seen to have made the homicide less likely were better patient compliance; closer contact with patient's family; closer patient supervision; improved staff communication; and better staff training.[13]

Longitudinal trends

When longitudinal data from the National Confidential Inquiry from 1997 to 2003 was examined it was apparent that, despite a rise in the homicide conviction rate in the general population, there has been no consistent change in rates of mental illness symptomatology at the time of the offence, contact with mental health services, lifetime history of mental illness, or specifically schizophrenia. Significant upward trends can be seen in the number of perpetrators with a history of drug and alcohol misuse, in particular the use

of cocaine and crack cocaine. There has been a significant decrease in those receiving a verdict of diminished responsibility but, surprisingly, no change in rates of those receiving a hospital order.[13]

Further information

http://www.medicine.manchester.ac.uk/suicideprevention/nci/ (the website for the National Confidential Inquiry which is regularly updated).

Maden, A. (2007). *Treating violence: a guide to risk management in mental health*. Oxford University Press, Oxford.

Taylor, P. and Gunn, J. (2009). *Forensic psychiatry: clinical, legal and ethical issues* (2nd edn). Hodder Arnold, London.

References

1. Phelan, J.C. and Link, B.G. (1998). The growing belief that people with mental illnesses are violent: the role of the dangerousness criterion for civil commitment. *Social Psychiatry and Psychiatric Epidemiology*, **33**, S7–12.
2. Crisp, A.H., Gelder, M.G., Rix, S., *et al.* (2000). Stigmatisation of people with mental illness. *The British Journal of Psychiatry*, **177**, 4–7.
3. Hodgins, S. (1992). Mental disorder, intellectual deficiency, and crime: evidence from a Danish birth cohort. *Archives of General Psychiatry*, **49**, 476–83.
4. Link, B.G., Andrews, H., and Cullen, F.T. (1992). The violent and illegal behaviour of mental patients reconsidered. *American Sociological Review*, **57**, 275–92.
5. Link, B.G., Stueve, A., and Phelan, J. (1998). Psychotic symptoms and violent behaviours: probing the components of 'threat/control-override' symptoms. *Social Psychiatry and Psychiatric Epidemiology*, **33**, S55–60.
6. Swanson, J.W., Holzer, C.E. III, Ganju, V.K., *et al.* (1990). Violence and psychiatric disorder in the community: evidence from the Epidemiologic Catchment Area surveys. *Hospital & Community Psychiatry*, **41**, 761–70.
7. Steadman, H.J., Mulvey, E.P., Monahan, J., *et al.* (1998). Violence by people discharged from acute psychiatric inpatient facilities and by others in the same neighbourhoods. *Archives of General Psychiatry*, **55**, 1–9.
8. Ritchie, J., Dick, D., and Lingham, R. (1994). *Report of the inquiry into the care and treatment of Christopher Clunis*. Stationery Office, London.
9. Coid, J. (1983). The epidemiology of abnormal homicide and murder followed by suicide. *Psychological Medicine*, **13**, 855–60.
10. Simpson, A.I.F., McKenna, B., Moskowitz, A., *et al.* (2004). Homicide and mental illness in New Zealand, 1970–2000. *The British Journal of Psychiatry*, **185**, 394–8.
11. Cote, G. and Hodgins, S. (1992). The prevalence of major mental disorders among homicide offenders. *International Journal of Law and Psychiatry*, **15**, 89–99.
12. Woodward, M., Nursten, J., Williams, P., *et al.* (2000). Mental disorder and homicide: a review of epidemiological research. *Epidemiologiae Psichiatria Sociale*, **9**, 171–89.
13. Appleby, L., Shaw, J., Kapur, N., et al. (2006). Avoidable deaths: five year report by the national confidential inquiry into suicide and homicide by people with mental illness. Published online 2006.
14. Taylor, P.J. and Gunn, J. (1999). Homicides by people with mental illness: myth and reality. *The British Journal of Psychiatry*, **174**, 9–14.
15. Home Office. (2005). Crime in England and Wales 2003/2004. In *Homicide and gun crime, Supplementary* Volume 1 (ed. D. Povey). National Statistics, London.
16. Flynn, S., Shaw, J., and Abel, K. *Homicide of infants: a crossectional study*. (submitted).
17. Kraemer, G.W., Lord, W.D., and Heilbrun, K. (2004). Comparing single and serial homicide offenses. *Behavioral Sciences & the Law*, **22**, 325–43.
18. Mullen, P. (2004). The autogenic (self-generated) massacre. *Behavioral Sciences & the Law*, **22**, 311–23.
19. Cantor, C.H., Mullen, P.E., and Alpers, P.A. (2000). Mass homicide: the civil massacre. *The* Journal of the American Academy of Psychiatry and the Law, ***28****, 55–63.*
20. Godwin, G.W. (2000). Hunting serial predators: a multivariate classific*ation approach to profiling violent behaviour. CRC Press, Boca R*aton, FL.
21. Meloy, J.R. and Felthous, A. (2004). Introduction to this issue: serial and mass homicide. *Behavioral Sciences & the Law,* **22**, 289–90.
22. Putkonen, H., Komulainen, E.J., Virkkunen, M., *et al.* (2003). Risk of repeat offending among violent female offenders with psychotic and personality disorders. *The American Journal of Psychiatry*, **160**, 947–51.
23. Knight, B. (1983). Geriatric homicides—or the Darby and Joan syndrome. *Geriatric Medicine*, **13**, 297–300.
24. Malphurs, J.E., Eisdorfer, C., and Cohen, D. (2001). A comparison of antecedents of homicide-suicide and suicide in older married men. *The American Journal of Geriatric Psychiatry*, **9**, 49–57.
25. Wallace, C., Mullen, P., Burgess, P., *et al.* (1998). Serious criminal offending and mental disorder. Case linkage study. *The British Journal of Psychiatry*, **172**, 477–84.
26. Munro, E. and Rumgay, J. (2000). Role of risk assessment in reducing homicides by people with mental illness. *The British Journal of Psychiatry*, **176**, 116–20.
27. Walsh, E., Buchanan, A., and Fahy, T. (2002). Violence and schizophrenia: examining the evidence. *The British Journal of Psychiatry*, **180**, 490–5.
28. Wessley, S.C., Castle, D., Douglas, A.J., *et al.* (1994). The criminal careers of incident cases of schizophrenia. *Psychological Medicine*, **24**, 483–502.
29. Arsenault, L., Moffitt, T.E., Caspi, A., *et al.* (2000). Mental disorders and violence in a total birth cohort: results from the Dunedin study. *Archives of General Psychiatry*, **57**, 979–86.
30. Brennan, P.A., Grekin, E.R., and Vanman, E.J. (2000). Major mental disorders and crime in the community. In *Violence among the mentally Ill* (ed. S. Hodgins). Kluwer Academic Publishers, Dordrecht.
31. Taylor, P.J. and Gunn, J. (1984). Violence and psychosis 1. Risk of violence among psychotic men. *British Medical Journal*, **288**, 1945–9.
32. Eronen, M., Hakola, P., and Tiihonen, J. (1996). Mental disorders and homicidal behaviour in Finland. *Archives of General Psychiatry*, **53**, 497–501.

11.6

Fraud, deception, and thieves

David V. James

Introduction

Dishonesty and deception are mundane and ubiquitous elements of human behaviour. Various forms are also categorized as criminal offences in the codes or statutes of all organized societies. In criminological terms, fraud, deception, and theft are forms of stealing, in other words dishonestly depriving others of goods or services. However, deception and fraudulent misrepresentation play a much wider role in human behaviour and interactions. This chapter will first consider briefly the broader picture, before considering in detail psychiatric aspects of stealing.

Fraud and deception

It is a common tendency to distort memory or reality through the lens of wishful thinking. To deny, to lie to others, and to engage in self-deception is part of the human condition. At one level, this can amount to innocuous forms of self-distraction, such as daydreaming or the childhood world of make-belief. Both blend the borders between fantasy and reality. Deception, which more bluntly put means lying, is on the other hand instrumental in purpose. The essential elements of lying are a conscious awareness of falsity, the intent to deceive and a pre-conceived goal or purpose. In some walks of life, such as advertising or politics, it may be part of the job, at least when the individual thinks that they can get away with it. Deception of others may slide into self-deception, the editing of memory to suit current desires. This may concern a wide range of life's extravagancies, from the fraudulent endeavours of the narcissistic fantasist to the imperatives of those who convince themselves, or wish to convince others, of their own supposed illnesses or infirmities. Particular forms of fraud, deception, and self-deception relevant to the medical context are factitious disorder, malingering, and pathological lying: these are dealt with in Chapter 5.2.9.

Theft

Types

Categorization of offences differs between jurisdictions. However, that for England and Wales forms a basic, illustrative framework. Offences are categorized under the Theft Act 1968 into theft, burglary, vehicle offences, and deception. The Theft Act 1978 further clarifies the area of deception (dishonestly obtaining services from another), dividing it into evading liabilities and debts, and making off without payment. The Fraud Act 2006 provides for a general offence of fraud with three ways in which it can be committed—by false representation, by failing to disclose information, and by abuse of position.

Rates/prevalence

Rates of criminal theft, fraud, and deception can be examined in terms of offences committed per unit of population. In England and Wales, with a population of 53.3 million, the number of reported fraud, theft, and deception offences in 2006 was 2 789 600, which equates to approximately 5.2 per 1 000 total population. These comprised 51 per cent of all crimes reported to the police.[1] However, crime is under-reported to the police and it is notable that the rate of offending as recorded by population survey in England and Wales is generally more than double. International comparisons of crime figures are hampered by differences in classification, measures, and time period.

A motivational classification of theft

Theft and dishonesty may arise from a range of motives. There is no necessity to invoke mental disorder in order to explain thieving. However, a minority of cases are related to serious mental disorder, such as mania, schizophrenia, depression, and organic brain disorder; and other forms of abnormal psychological processes can act as drivers for criminally dishonest behaviour. These go beyond the formal classification of such behaviours in, for instance, the DSM-IV-TR,[2] where these appear only in the specific syndrome of kleptomania and as a component of the definitions of conduct disorder and anti-social personality disorder.

The following classification is not exhaustive and the categories not mutually exclusive. The great majority of thefts will fit into the first two categories, and it is the remainder that are more likely to form part of a presentation to a psychiatrist.

(a) Ordinary theft

Ordinary theft may be planned or impulsive, but is deliberate and motivated by the usefulness of the object or its monetary value.

1 Professional—crime as a career choice.

2 Delinquent—theft as one component of a delinquent or anti-social lifestyle.

3 Survival offences—driven by poverty, desperation, or necessity.

(b) Emotionally driven theft

Emotionally driven theft is the consequence of emotion, rather than as a means of financial gain.

4 Anger or revenge—based upon depriving someone else, rather than personally acquiring.

5 Fear—coerced into committing an offence by threats from a third party.

6 Excitement—this can be as part of a dare or a rite of passage, particularly in adolescents.

(c) Secondary theft

Secondary theft is attributable to the presence of an underlying disorder.

(i) Pecuniary

7 To fund addictive behaviour—e.g. alcohol dependence, drug addiction, pathological gambling.

(ii) Non-pecuniary

8 Depressed stealing:

(a) cry for help (attention-seeking behaviour): stealing in a way that is sure to be detected in order to obtain support and other help.

(b) suicidal gesture: a depressed persons may steal articles of which they have no need, the action serving as a justification for feelings of guilt and a form of suicidal equivalent.

(c) substitution: the stolen object compensates for something else. For example a rejected wife might steal from her husband in order symbolically to establish a degree of control and compensate for the loss of affection.

(d) distraction: e.g. 'absent-mindedness' in a shop as a consequence of distraction by depressive ruminations, or distress in the context of bereavement or divorce.

9 Manic stealing: stealing in the context of manic disinhibition or delusion.

10 Psychotic: the result of delusional drive or command hallucinations.

11 Confusion: related to cognitive deficit, as a consequence of organic brain disorder, or to dissociative states.

12 'Kleptomania': compulsive stealing as an impulse-control disorder.

Management

A psychiatrist will only have a role in the management of theft cases where a particular psychiatric or psychological issue is central. The management will depend upon the nature of the underlying disorder. Treatment for the primary disorder will, in most cases, be supplemented by a psychological approach to helping the patient understand the reasons for their offending behaviour, recognize triggers or danger points, and develop strategies for dealing with these in the future.

Kleptomania constitutes a particular diagnostic entity, and shoplifting is a common behaviour which not uncommonly leads to requests for psychiatric reports from the courts. Both will be considered in more detail below.

Kleptomania

Definition

Kleptomania is an old term, first used by Marc and Esquirol in 1838 to indicate a 'stealing madness'. Kleptomania was designated a psychiatric disorder in DSM-III in 1980, and, in DSM-IIIR in 1987, it was grouped under the category 'impulse control disorder, not elsewhere classified'. The disorder is described further in Chapter 4.13.1 (Impulse Control Disorders) and only aspects most relevant to forensic practice are considered here.

The core characteristic is that objects are stolen despite the fact that they are typically of little value to the individual, who could have afforded to pay for them. Sometimes, they are hoarded and sometimes dispensed with or returned. DSM-IV-TR describes the essential features of the disorder in five diagnostic criteria (312.33):

(a) 'Recurrent failure to resist impulses to steal objects that are not needed for personal use or for their monetary value.

(b) Increasing sense of tension immediately before committing the theft.

(c) Pleasure, gratification, or relief at the time of committing the theft.

(d) The stealing is not committed to express anger or revenge and is not in response to a delusion or a hallucination.

(e) The stealing is not better accounted for by conduct disorder, a manic episode, or anti-social personality disorder'.

In addition, individuals with kleptomania 'experience the impulse to steal as ego-dystonic and are aware that the act is wrong and senseless'.

Diagnostic problems

Criterion A, which concerns the senselessness of the theft, is often considered to be the characteristic which separates kleptomania from ordinary theft. However, some kleptomaniacs may desire the items and be able to use them, even if they are not strictly needed: this may be particularly the case with those who hoard.[3] Concerning criteria B and C, some patients report amnesia surrounding the time of the theft and therefore deny feelings or tension immediately beforehand or relief at the time of committing the theft.[3,4] There is also some suggestion that kleptomaniacs who repeatedly steal over long periods eventually lose the feelings of tension and pleasure in a behaviour which has simply become a habit.

There may be overlap with other disorders.[5] Kleptomania and drug addiction share similar core qualities.[6] The presence of repetitive thoughts and behaviours suggests to some a link with the obsessive-compulsive spectrum.[7] With the high comorbidity with depressive disorders, kleptomania might be categorized within the affective spectrum.[8]

Epidemiology

The prevalence of kleptomania is unknown though it is thought to account for less than 5 per cent of shoplifting.[2] In the US the lifetime prevalence may be 0.6 per cent[3] though some consider that it may be higher because the associated embarrassment and illegality may deter people from reporting it.[9]

A US study of psychiatric inpatients found a lifetime prevalence of 9.3 per cent.[10] While of 107 inpatients with depression,

3.7 per cent had kleptomania[11] and, of 79 inpatients with alcohol dependence, 3.8 per cent reported symptoms consistent with kleptomania.[12]

Kleptomania was generally believed to be much more common in women, but in a summation of four studies of kleptomania 63 per cent were women,[5] suggesting that the preponderance of women is not as great as was once assumed.

Comorbidity

Kleptomania is highly comorbid with depression and anxiety.[13–15] Estimates of lifetime comorbid rates of mood disorders range from 59[9] to 100 per cent,[16] and of anxiety disorders from 60 to 80 per cent.[17] Some of the comorbidity with depression and anxiety may be secondary to the consequences of the behaviour. Compared with people with alcohol dependence and general psychiatric disorders, those with kleptomania scored significantly higher on measures of impulsivity, sensation-seeking, and disinhibition.[14] In one study,[18] 43 per cent of people with kleptomania met the criteria for at least one personality disorder, the most common being paranoid, schizoid, and borderline. However in a meta-analysis none satisfied the diagnostic criteria for anti-social personality disorder.[19]

Aetiology

The aetiology of kleptomania is uncertain. Suggested causes include attempts to relieve feelings of depression through stimulation,[20,21] or to make up for early deprivation.[5,22] The various theories are considered further in Chapter 4.13.1

Treatment

Treatment involves a combined psychological and pharmacological approach. There is little evidence to suggest one particular psychological approach, but a combination of cognitive behaviour therapy and psychosocial interventions is generally adopted. As regards drug treatments, there have been case reports and small series suggesting the efficacy of a range of drugs,[23] in particular selective serotonin re-uptake inhibitors, mood stabilizers, and opioid antagonists. However, there is a need for controlled trials before definitive conclusions can be drawn.

Shoplifting

Shops specialize in making items seem desirable and tempting to the shopper. Most people shoplift at least once at some point in their lives, usually opportunistically in adolescence. But the incidence of shoplifting remains unknown. An early observational study in the UK[24] suggested that 1 to 2 per cent of customers in an English department stores took items without paying, which compared with 1 in 12 in New York City and 1 in 18 in Dublin. The number of people convicted of shoplifting is evidently far smaller. Whereas earlier studies had suggested that shoplifting was more common in women,[25] this is now thought to have been an artefact of case selection, as samples in most studies have been limited to court samples or sub-samples referred to psychiatrists for reports.[26]

There is no unitary phenomenon of shoplifting. It has been suggested that those who shoplift fall into two groups—those who do so out of rational choice and those that suffer from depression.[27] A survey of 1649 shoplifting convictions in Montréal found that only 3.2 per cent were suffering from serious mental disorder, but that affective symptoms were relatively common.[28] A further study of 106 shoplifters[26] reported that depression was the most common psychiatric disorder associated with shoplifting and that the majority of shoplifters were poor and unemployed. The range of possible psychiatric disorders associated with shoplifting is indicated in the classification of theft above, but in the large majority of cases of shoplifting, there is no psychiatric reason for avoiding payment.

Management

A thorough assessment is needed to order to establish how and why the shoplifting occurred and to indicate the presence or absence of underlying mental illness or disorder. Where such illness or disorder is present, it should be treated. Where no such factors are in evidence, psychological therapies may still have a role in offender rehabilitation, usually in a group setting.

Further information

Grant, J.E. (2005). Kleptomania. In: *Clinical manual of impulse-control disorders,* (eds. E.R. Hollander and D.J. Stein). Arlington, American Psychiatric Publishing, USA, pp 175–201.

Ormerod, D. and Williams, D. (2007). *Smith's Law of Theft* (9th ed). Oxford: Oxford University Press.

References

1. Lovbakke, J., Taylor, P., and Budd, S. (2007). *Crime in England and Wales: quarterly update to December 2006*. Home Office Statistical Bulletin. Home Office, London.
2. American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th edn, text revision). American Psychiatric Association, Washington, DC.
3. Goldman, M.J. (1991). Kleptomania: making sense of the non-sensical. *The American Journal of Psychiatry*, **148**, 986–96.
4. Grant, J.E. (2004). Dissociative symptoms in kleptomania. *Psychological Reports*, **94**, 77–82.
5. Grant, J.E. (2005). Kleptomania. In *Clinical manual of impulse-control disorders* (eds. E.R. Hollander and D.J. Stein). American Psychiatric Publishing, Arlington, USA.
6. Grant, J.E. and Potenza, M.N. (2004). Impulse control disorders: clinical characteristics and pharmacological management. *Annals of Clinical Psychiatry*, **16**, 27–34.
7. McElroy, S.L., Phillips, K.A., and Keck, P.E. (1994). Obsessive-compulsive spectrum disorder. *The Journal of Clinical Psychiatry*, **55**, 35–51.
8. McElroy, S.L., Pope, H.G., Keck, P.E., *et al.* (1996). Are impulse control disorders related to bipolar disorder? *Comprehensive Psychiatry*, **37**, 229–40.
9. Grant, J.E. and Kim, S.W. (2002). Clinical characteristics and associated psychopathology of 22 patients with kleptomania. *Comprehensive Psychiatry*, **43**, 378–84.
10. Grant, J.E., Levine, L., Kim, D., *et al.* (2005). Impulse control disorders in adult psychiatric inpatients. *The American Journal of Psychiatry*, **162**, 2184–8.
11. Lejoyeux, M., Arbaretraz, M., McCloughlin, M., *et al.* (2002). Impulse control disorders and depression. *The Journal of Nervous and Mental Disease*, **190**, 310–14.
12. Lejoyeux, M., Feuché, N., Loi, S., *et al.* (1999). Study of impulse-control disorders among alcohol-dependent patients. *The Journal of Clinical Psychiatry*, **60**, 302–5.
13. Sarasolo, E., Bergman, B., and Toth, J. (1996). Personality traits and psychiatric and somatic morbidity among kleptomaniacs. *Acta Psychiatrica Scandinavica*, **94**, 358–64.

14. Baylé, F.J., Caci, H., Millet, B., *et al.* (2003). Psychopathology and comorbidity of psychiatric disorders in patients with kleptomania. *The American Journal of Psychiatry*, **160**, 1509–13.
15. Dannon, P.N., Aizer, A., and Lowengrub, K. (2006). Kleptomania: differential diagnosis and treatment modalities. *Current Psychiatry Reviews*, **2**, 281–3.
16. McElroy, S.L., Pope, H.G., Hudson, J.I., *et al.* (1991). Kleptomania: a report of 20 cases. *The American Journal of Psychiatry*, **148**, 652–7.
17. McElroy, S.L., Hudson, J.I., Pope, H., *et al.* (1992). The DSM-III-R impulse control disorders not elsewhere classified: clinical characteristics and relationship to other psychiatric disorders. *The American Journal of Psychiatry*, **149**, 318–27.
18. Grant, J.E. (2004). Co-occurrence of personality disorders in persons with kleptomania: a preliminary investigation. *The Journal of the American Academy of Psychiatry and the Law,* **32**, 395–8.
19. McElroy, S.L., Hudson, J.I., and Pope, H.G. (1991). Kleptomania: clinical characteristics and associated psychopathology. *Psychological Medicine*, **21**, 93–108.
20. Gudjonnson, G.H. (1987). The significance of depression in the mechanism of 'compulsive' shoplifting. *Medicine, Science, and the Law*, **27**, 171–6.
21. Fishbain, D.A. (1987). Kleptomania as risk-taking behaviour in response to depression. *American Journal of Psychotherapy*, **41**, 598–603.
22. Grant, J.E. and Kim, S.W. (2002). Temperament and early environmental influences in kleptomania. *Comprehensive Psychiatry*, **43**, 223–9.
23. Durst, R., Katz, G., Teitelbaum, A., *et al.* (2001). Kleptomania: diagnosis and treatment options. *CNS Drugs*, **15**, 185–95.
24. Buckle, A. and Farrington, D. (1984). An observational study of shoplifting. *British Journal of Criminology*, **24**, 63–73.
25. Bradford, J. and Balmaceda, R. (1983). Shoplifting: is there a specific psychiatric syndrome? *Canadian Journal of Psychiatry*, **28**, 248–54.
26. Lamontagne, Y., Boyer, R., Hétu, C., *et al.* (2000). Anxiety, significant losses, depression and irrational beliefs in first-offence shoplifters. *Canadian Journal of Psychiatry*, **45**, 63–6.
27. Gudjonsson, G.H. (1990). Psychological and psychiatric aspects of shoplifting. *Medicine, Science, and the Law*, **30**, 45–51.
28. Lamontagne, Y., Carpentier, N., Hétu, C., *et al.* (1994). Shoplifting and mental illness. *Canadian Journal of Psychiatry*, **39**, 300–2.

11.7

Juvenile delinquency and serious antisocial behaviour

Susan Bailey

Introduction

Juvenile crime and delinquency represent a significant social and public health concern. Both rates of mental disorders and offending are high during adolescence. This chapter reviews prevalence rates of mental disorders in young offenders, screening, and assessment of juveniles, principles of interventions with young offenders before describing principles of forensic mental health, policy and practice, how mental disorders in adolescence can impact on offending and antisocial behaviour, how policy is shaping practice in this field and how mental health practitioners may be involved in meeting mental health needs and undertaking medico-legal assessments

Delinquency, conduct problems, and aggression all refer to antisocial behaviours that reflect a failure of the individual to conform his or her behaviour to the expectations of some authority figure, to societal norms, or to respect the rights of other people. The 'behaviours' can range from mild conflicts with authority figures, to major violation of societal norms, to serious violations of the rights of others.[1] The term 'delinquency' implies that the acts could result in conviction, although most do not do so. The term 'juvenile' usually applies to the age range, extending from a lower age set by age of criminal responsibility to an upper age when a young person can be dealt with in courts for adult crimes. These ages vary between, and indeed within, countries and are not the same for all offences.[2,3]

Adolescence as a context

The adolescent population in the UK constitutes half of the child population with around 7.5 million young people in the transitional stage between childhood and adulthood, (age 10–19).[4] Adolescence is a transitional stage of development between childhood and adulthood—a stage of possibility and of promises and worries that attend this possibility. The developmental tasks of adolescence centre on autonomy and connection with others, rebellion and the development of independence, development of identity and distinction from and continuity with others. The physical changes of puberty are generally seen as the starting point of adolescence whilst the end is less clearly delineated. Adolescence ends with attainment of 'full maturity'. A range of social and cultural influences including the legal age of majority, may influence the definition of maturity.[5]

Mortality among adolescents, in contrast to almost all other age groups, did not fall during the second half of the twentieth century, the main causes being accidents and self-harm.[4]. Health needs are greater in this age band than in children in middle childhood (5 to 12 years) or of young adults, and arises out of mainly chronic illness and mental health problems. The main concerns of young people, in relation to health, focus on issues of immediacy that impact on their relations with peers and include problems with skin, weight, appearance, emotions, and sexual health including contraception.

The principal aim of the Youth Justice System (YJS) is to prevent offending by children and young people under 18 years of age. There are 157 Youth Offending Teams in England and Wales. The YJB commissions some 3000 custodial places at any one time for young people under the age of 18 years in 18 Prison Service Young Offenders Institutions, 15 Local Authority Secure Children's Home and 4 private sector Secure Training Centres. In 2005–2006 there were 301,860 recorded offences the 4 highest recorded offences being theft and handling 18.5 per cent, violence against the person 18.1 per cent, motoring offences 16.6 per cent, criminal damage 12.9 per cent. 16 and 17 year old were responsible for 49.6 per cent of offences with males responsible for 80.6 per cent and females 19.4 per cent of all offences resulting in a disposal. Offences by ethnicity were white 85.2 per cent and Black and Ethnic Minority 14.8 per cent. Of the 212,242 disposals 80 per cent received pre-court of first tier disposals with 17 per cent receiving a community sentence and 3 per cent a custodial sentence.[6]

Risk factors for, and pathways to antisocial behaviour are summarized in Tables 11.7.1 and 11.7.2 respectively. There is a significant overlap between the risk factors for offending, poor mental health and substance misuse and the number of assessed risk factors increases as a young person moves further into the Youth Justice System.[7] Many young offenders are not engaged in mainstream education and health services. It is critical that these young people are supported to access the mainstream and specialist services they require while under the supervision of the YOT or in custody. Otherwise, once their sentence ends they can become detached from services and their circumstances are likely to deteriorate, leading to more offending and greater demands on specialist services as they get older.

In the UK, the Children's National Service Framework for Children, Young People and Maternity Services (NSF) set out a vision of a comprehensive child and adolescent mental health service.[8]

Table 11.7.1 Major risk areas in children and adolescents with persistent antisocial behaviour[35]

Broad child-centred factors
Genetic vulnerability
Perinatal risk
Male sex
Cognitive impairment
School underachievement
Hyperactivity/inattention temperament
Family factors
Criminality in parents and siblings
Family discord
Lack of supervision
Lack of effective feeling
Abuse
Scapegoating
Rejection
Neglect
Influential contextual factors
Drug and alcohol abuse
Unemployment
Crime opportunity
Peer group interaction

A young person in contact with the criminal justice system, whether in custody, or in the community, should have the same access to this comprehensive service as any other child or young person within the general population. Treatment options should not be affected by a young offender's status within the criminal justice system. The Change for Children Programme has the aim of improving outcomes for all children in the following 5 areas: being healthy, staying safe; enjoying and achieving; making a positive contribution; and achieving economic well-being, and to narrow the gap in outcomes between those who do well and those who do not. If we do not address the mental health needs of young offenders then they are excluded from the opportunity to participate in improvements in these 5 outcomes and ultimately from the ability to achieve their full potential.

Table 11.7.2 Critical pathway to serious antisocial behaviour[64]

Family features
Parental antisocial personality disorder
Violence witnessed
Abuse, neglect, rejection
Personality features
Callous unemotional interpersonal style
Evolution of violent and sadistic fantasy
People as objects
Morbid identity
Paranoid ideation
Hostile attribution
Situational features
Repeated loss and rejection in relationships
Threats to self-esteem
Crescendo of hopelessness and helplessness
Social disinhibition
Group processes
Changes in mental state over time

There is a high prevalence of mental health problems among young people in custody.[9] YJB research published in 2005 reported the following findings.[10]

- 31 per cent had mental health problems
- 18 per cent had problems with depression
- 10 per cent suffered from anxiety
- 9 per cent reported a history of self-harm in the preceding month
- 9 per cent suffered from post-traumatic stress disorder
- 7 per cent had problems with hyperactivity
- 5 per cent reported psychotic-like symptoms

One in five young offenders were identified as having intellectual disability IQ<70. Additionally, needs were identified across education 48 per cent and social relationships 36 per cent. Needs were unmet because they were not recognized.

Research consistently reveals high levels of psychiatric disorders among detained juveniles, although rates vary widely by study, ranging from more than 50 per cent to 100 per cent.[11–21] The variations between the studies may reflect methodological differences or true variations between countries and samples. Advances in developmental psychopathology and increased understanding of the continuities between child and adult life[27] demonstrating that many childhood disorders once thought to resolve with age cast long shadows over later development.

There are several reasons why high rates of mental disorders may be expected in youth in contact with juvenile justice. First, prevalence rates of psychiatric disorders in community samples were shown to be around 15 per cent.[22] Also, severe delinquency is common in the adolescent population, with about 5 per cent showing an early-onset and persistent pattern of antisocial behaviour.[23] A substantial number of adolescents will show offending behaviour and will have a mental health disorder simply because of coincidental overlap between both conditions. Second, because delinquent and antisocial behaviour reaches high levels among juvenile justice populations, a diagnosis of conduct disorder (CD) will often be made. Because CD shows high comorbidity rates with several other psychiatric disorders,[24] increased levels of many types of disorder may be expected. Third, risk factors for youthful offending overlap substantially with those for several types of non-disruptive child psychiatric disorders, therefore identical risk factors may underlie both antisocial behaviour and emotional or developmental problems. Disorders for which mental health interventions are provided, such as substance use disorders (SUD's), may also lead to judicial involvement. Also, because of the prevalence of complex comorbidity, treatment in a regular mental health care programme may be intricate and often is not possible, thus increasing the likelihood of judicial involvement. In addition, severely disordered persons may be less likely to have the personal capability and have adequate resources to defend themselves and to avoid more drastic legal interventions.

Grisso and Zimring listed three principal reasons for concern regarding mental disorders in youthful offenders: a) the obligation to respond to mental health needs in those in custody, b) assurance of due process in adjudicative proceedings, and c) public safety.[25]

Mental health treatment within the juvenile justice system is often inadequate. It has been reported that only about 20 per cent of incarcerated youth with depressive disorders, 10 per cent with other mental disorders, and less than half with SUD receive intervention.[26,27]. Much more research is needed into the treatment needs of this population.

Risk and protective factors

Understanding is growing of how risk factors combine to both precipitate and maintain antisocial behaviour. Several environmental and individual risk factors including psychiatric pathology in childhood have been identified.[28] Not all risk factors need to be present in a single individual but multiple risk factors greatly increases the risk of a serious and long-term negative development.[29] Positive characteristics or experiences may act protectively. These protective factors may be specific interventions or experienced within the natural contact of development. When protective factors are present, young people may show positive social development despite high risk of antisocial behaviour, or they may abandon their problem behaviour after a difficult phase. Such trajectories are less well investigated than the risks.[30–33] It is also more difficult to implement adequate research designs in this field.[34] It could be assumed that the opposite to the risk value of the variables listed in Table 11.7.1 may promote positive development. However, truly protective effects need to compensate for a given high-risk constellation (moderator approach). The available research suggests a number of factors that may protect from the risks of antisocial behaviour. Table 11.7.3 reports a selection of such personal and social resources that have already been proven or may be promising (for a detailed review see[33]).

Table 11.7.3 Multilevel examples for protective factors against serious antisocial behaviour[35]

Biological/bisocial	- Non-deviant close relatives; no genetic vulnerabilities; high arousal; normal neurological and hormonal functioning.
Pre-and perinatal	- Non-alcoholic mother; no maternal smoking during pregnancy; no birth complications
Child personality	- Easy temperament; inhibition; ego-resiliency; intelligence; verbal skills; planning for the future; self-control; social problem solving skills; victim awareness; secure attachment; feelings of guilt; school and work motivation; special interests or hobbies; resistance to drugs
Cognitions/attitudes	- Non-hostile attributions; non-aggressive response schemes; negative evaluation of aggression; self-efficacy in prosocial behaviour; non-deviant beliefs; realistic self-esteem; sense of coherence.
Family	- No poverty; income stability; harmony; acceptance; good supervision; consistency; positive role models; continuity of caretaking; no disadvantage; availability of social support.
School	- Achievement and bonding; low rate of aggressive students; climate of acceptance; structure, and supervision.
Peer group	- Non-delinquent peers; support from close, prosocial friends.
Community	- Non-deprived, integrated and non-violent neighbourhood; availability of professional help.
Situational	- Target hardening; victim assertivesness; social control.
Legal	- Effective firearm and drug control; effective criminal justice interventions.
Cultural	- Low violence; tradition of moral values; shame and guilt-orientation; low exposure to violence in the media.

Pathways of care and the juvenile justice system

Juvenile justice is a high-volume system, which makes clear logistics and a clear pathway of care necessary. Early identification of mental health needs may result in diversion from custody by using community services rather than adjudication and derive economic benefit by affording non-custodial disposal. Nonetheless a significant number of young persons progress to pre-trial assessment, albeit from the home or a residential care setting.

Preadjudication dispositions should be informed therefore by best available screening and assessment processes. In this context specific tools may be used to derive markers of psychopathology and of ongoing risk to self and others as well as to address medico-legal questions posed by the criminal justice system including assessment on disposition, matters of public protection, treatment for mental disorders, and need for security and likelihood of recidivism.

For those detained in prison, screening must determine if urgent problems (such as suicidal intent or consequences of substance use) require immediate attention; a detailed diagnostic assessment of the young person may take a longer period of time and continue as the youngster moves from one institution to another. Later critical transitions, for which an additional screening may be useful, include re-entry into the community, assessment of readiness for re-entry, mental health planning for integrated continuing care post detention as part of a multiagency re-entry strategy, and, where necessary, community residential programmes monitoring emotion or reactions, especially where the young person is returning to stressful conditions such as a troublesome family.

General principles of assessment

Standard clinical assessment tools used in child and adolescent psychiatry cover many of the areas considered in forensic child and adolescent risk assessments.[36] This is especially important as juvenile justice systems in particular are not always equitable. In choosing between the many scales available it is important to question not just their proven scientific properties but also their feasibility for practitioners to use.

It is important to consider the purpose for which the scale is to be used (see table below). Scales that measure psychopathology may not be good ways of assessing the risk that the psychopathology poses. Measures used to map out types of symptom must have good content validity. An instrument required to pick out one group of symptomatic people from the rest of the community (e.g. mental health screening of young people in custody) needs to have

good criterion validity. A related issue is the extent to which the scale is intended to measure change.

Grid for specifying requirements of a structured scale in a juvenile forensic population

Assessment required (yes / no)

Purpose of assessment	Psychopathology	Need	Risk
Screening of all juveniles coming into contact with an agency			
Detailed assessment e.g. for sentencing, planning treatment			
Measuring change e.g. during treatment or sentence			

Child psychiatry uses multi-axial and developmental concepts of child psychopathology. Specific and general intellectual delays are very common among young people in the juvenile justice system[10] as is co-morbidity of disorders. Broad-band interviews, however, offer only poor coverage of rare conditions such as pervasive developmental disorders.

Needs assessment

Needs assessment may have advantages over more traditional ways of diagnosing disorders, mainly because this method also indicates whether specific conditions need attention and intervention. Especially in delinquent youth characterized by multiple problems, such an approach may carry substantial advantage. A health care need should be distinguished from a general need. One commonly used definition of a health care need is '*the ability to benefit in some way from (health) care*'.[37]

Needs and risk assessment are two separate but intertwined processes essential for clinical management (see Fig. 11.7.1). Assessment of danger to others and the need to address this problem is at the centre of legislative and policy decision-making. The attention of the public and media are focussed on this area. Needs assessment may both inform and be a response to the risk-assessment process.[38] The reciprocal process can be termed 'risk management' when accurate information about the risk assessment, combined with recurrent needs assessment, leads to risk-management procedures. A recurrent needs-assessment and risk-assessment process should identify changes in problem areas, thus leading to monitoring or intervention as part of risk management. Core to this assessment are appropriate mental health screening tools and processes that are available to the young person at any point in the system.[39]

Risk assessment

Risk assessment combines statistical data with clinical information in a way that integrates historical variables, current crucial variables, and the contextual or environmental factors. Structured risk assessment instruments have been developed that aim to increase the validity of clinical prediction. These scales typically contain a number of risk items selected from reviews of research, crime theories, and clinical considerations.[41,42] Items are summed to form a total risk score and may also reveal specific risk patterns (e.g. mainly family or child factors). Such instruments are used for screening, in-depth assessment and related risk management (e.g. for decisions on the child's placement or specific interventions). They can also be applied in differentiated evaluations of intervention programmes. Instruments vary with respect to the age and gender for their clients, problem intensity in the target groups, theoretical and empirical foundations, the number and domains of risk included, scoring procedures, time required for assessment, information sources, institutional contexts of administration and other issues.[43,44] Many instruments have been designed for application in the juvenile justice system.[43,45,46] Most instruments contain factors from various areas of risk (e.g. individual, family, neighbourhood).

Mental disorders and offending

Current concepts focus on a developmental approach to psychopathology in child and adolescent psychiatry and psychology. Physical aggression peaks at around the second year of life and subsequently shows distinct developmental trajectories.[47,48] Attachment enables the mastery of aggression, self-control being developed through the efficient exercise of attritional mechanisms and symbolization. Fonagy has suggested a primary developmental role for early attachment in the development of mentalization (the capacity to understand others' subjective experience). He suggests that impaired mentalization leads to later violence.[49] Threats to self-esteem trigger violence in individuals whose self-appraisal is 'on shaky ground' and are unable to see behind the threats to what is in the mind of the person threatening them. These processes are played out in the complex and toxic co-morbidities seen so much more frequently in child and adolescent than in adult mental health practice.

Oppositional disorders, conduct disorder, and ADHD

Substantially higher rates of physically aggressive behaviour are found in children and adolescents with attention deficit hyperactivity disorder, with those who meet the criteria for ADHD and conduct disorder having substantially greater risks of delinquent acts in adolescence, harmful acts in later adolescence and continued violence and offending into adulthood.[50] Children with hyperactivity, impulsivity, attention deficits and serious conduct problems may also be at risk for developing psychopathy.[51]

Distorted or biased thought processes have over time been implicated in the development of violence. Psychological treatments aimed at reducing violent behaviour in adolescents and young adults traditionally centre on violence as learned behaviour. Patterns of violence and criminal behaviour are seen as embedded in habits of thinking.[52] In juvenile delinquents significant cognitive attributional bias has been shown in aggressive children and youths. They are more likely to perceive neutral acts by others as hostile, and more likely to believe conflicts can be satisfactorily resolved by aggression. In the social context, as the young individual becomes more disliked and rejected by peers, the opportunity for viewing the world this way increases.[53] By their late teens they can hold highly suspicious attitudes and be quick to perceive disrespect from others. In the social context of juvenile incarceration,

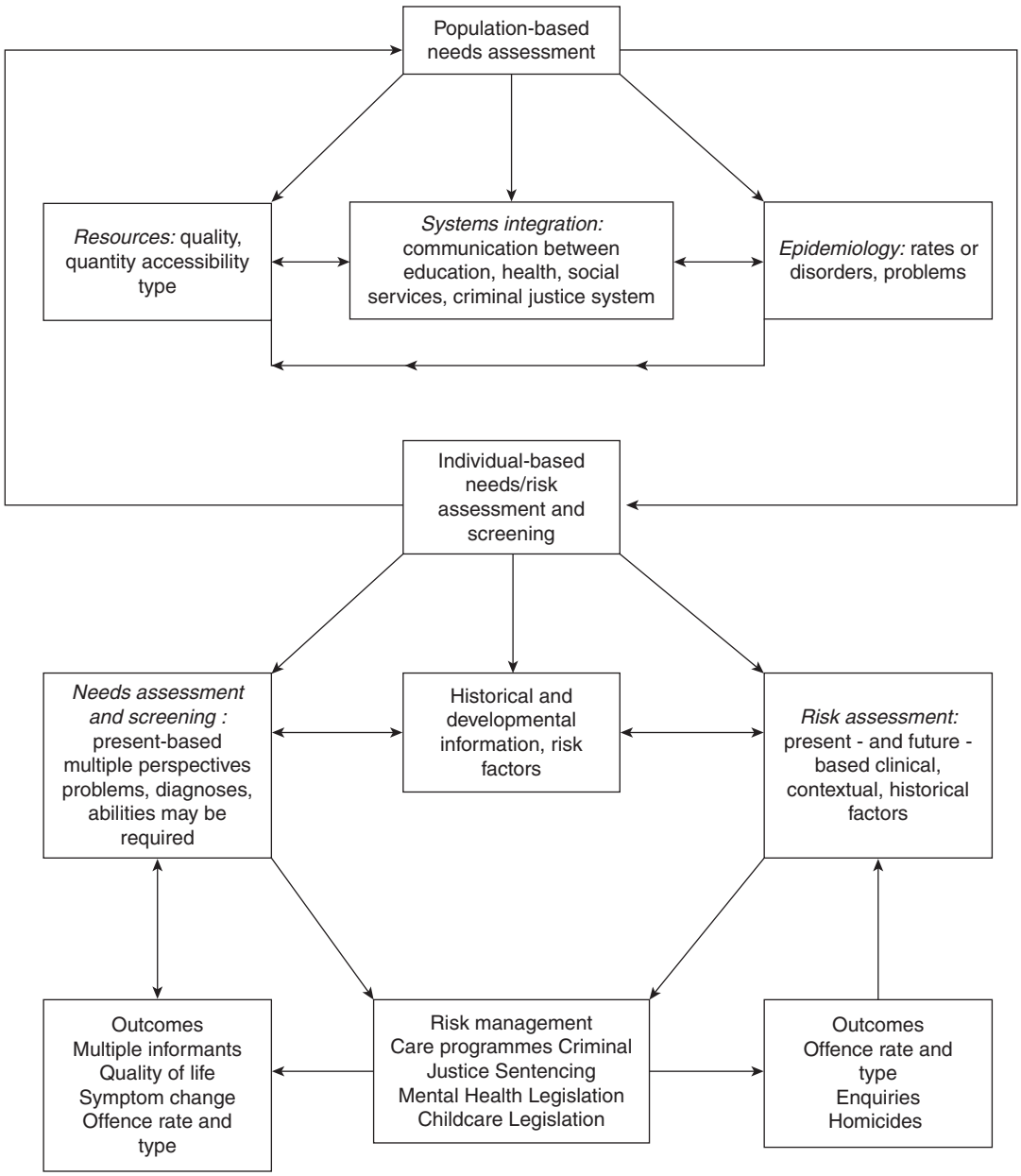


Fig. 11.7.1 Relationship between various screening, need assessment, risk assessment, and management approaches in juvenile justice systems.[40] (Reproduced from Kroll, L. Needs assessment in adolescent offenders. In Adolescent forensic psychiatry (eds. Bailey, S. Dolan, M.), copyright 2004, Hodder Education.)

being 'para'[54] can become in peer group interactions the shared norm.[55]

Depression anxiety and post traumatic stress disorder in childhood and adolescence

As well as the recognized feelings of low mood in depression there is also some evidence of irritability, hostility and anger when depression occurs in adolescence. Irritability in adolescence leads to interpreting annoyances by others as direct threats, increasing the risk of defensive aggression.[56] Nowhere is this more apparent than in juvenile justice populations.[57,58] A self-serving bias with a tendency to attribute good outcomes to the self and bad outcomes to external causes observed in ordinary people, is usually regarded as a mechanism for maintaining self-esteem in the face of threats to the self.

PTSD is related to the conditioning of neurobiological fear responses underlying tendencies to react aggressively to protect the self when exposed to reminders of earlier trauma.[59] In the recent escalating context of both children who have experienced violence in war torn countries and those who live in a context of 'urban war zones' Garbarino[60] sets out an ecological framework to explain the process and conditions that transform the 'developmental challenge' of violence into developmental harm in some children. He set out an accumulation of risk models for understanding how and when children suffer the most adverse consequences of

exposure to community violence and go beyond their limits of resilience, the concept of 'social maps' as products of childhood experiences and of trauma as a psychological wound.

The combination of depression, anxiety and severe PTSD is being increasingly recognized in as being linked to trajectory into adult anti-social personality disorder.[61]

Autism spectrum disorders and learning disability

Autism spectrum disorders are being increasingly recognized in adolescent forensic populations. Their identification is critical to the understanding of violent offending. This is particularly so if an offence or assault is bizarre in nature, the degree or nature of aggression is unaccountable and/or there is a stereotypic pattern of offending. Four reasons have been proposed for offending and aggression in autistic persons:[62,63]

1 Their social naivety may allow them to be led into criminal acts by others;

2 Aggression may arise from a disruption of routines;

3 Antisocial behaviour may stem from a lack of understanding or misinterpretation of social cues;

4 Crimes may reflect obsessions, especially when these involved morbid fascination with violence—there are similarities with the intense and obsessional nature of fantasies described in some adult sadists.[64]

It has been proposed that the paranoia observed in Asperger's syndrome has a different quality from that seen in people with a diagnosis of schizophrenia stemming from a confusion of not understanding the subtleties of social interaction and social rules.[54]

Early onset psychosis

Non-psychotic behavioural disturbance occurs in about half of cases of early-onset schizophrenia and can last between 1 and 7 years. It includes externalizing behaviours, attention-deficit disorder and conduct disorder. This emphasizes the need for mental health assessments repeated over time to include a focus on changes in social functioning (often from an already chaotic baseline level) to a state including perceptual distortion, ideas of reference, and delusional mood.[65]

As in adult life most young people with schizophrenia are non-delinquent and non-violent.[66] Nevertheless, there may be an increased risk of violence to others when they have active symptoms, especially when there is misuse of drugs or alcohol. The risk of violent acts is related to subjective feelings of tension, ideas of violence, delusional symptoms that incorporate named persons known to the individual, persecutory delusions, fear of imminent attack, feelings of sustained anger and fear, passivity experiences reducing the sense of self-control, and command hallucinations. Protective factors include responding to and compliance with physical and psychosocial treatments, good social networks, a valued home environment, no interest in or knowledge of weapons as a means of violence, good insight into the psychiatric illness and any previous violent aggressive behaviour and a fear of their own potential for violence. These features require particular attention but the best predictors of future violent offending in young people with mental disorder are the same as those in the general adolescent population.[67]

Psychopathic personality in young people

A three-factor structure has been proposed,[68] which includes:

- An arrogant, deceitful interpersonal style, involving dishonesty, manipulation, grandiosity and glibness;
- Defective emotional experience, involving lack of remorse, poor empathy, shallow emotions, and a lack of responsibility for one's own actions;
- Behavioural manifestations of impulsiveness, irresponsibility, and sensation-seeking.

Conduct disorder, antisocial personality disorder, and psychopathy are often seen as developmental disorders that span the life course and the terms are sometimes used interchangeably. Conduct disorder and antisocial personality disorder primarily focus on behavioural problems, psychopathy, as described by Hare,[69] emphasizes deficits in affective and interpersonal functioning.

A consensus is likely to be reached only when we have longitudinal studies demonstrating the stability of psychopathic traits over the lifespan and evidence that the same aetiological factors contribute to this disorder at all ages. As there is significant overlap between the behavioural aspects of juvenile psychopathy and ADHD and between the callous-unemotional dimension of psychopathy and autistic-spectrum disorders, future work needs to disentangle these constructs from a phenomenological and aetiological perspective. As yet, there are few treatment outcome studies in juveniles with psychopathic traits, although the limited data suggest that these traits might be a moderator of outcome. Most clinicians view youth psychopathy as a potentially treatable disorder, and there is some evidence that identification of psychopathic traits in young people has a number of benefits, which include:

- Identifying high-risk offenders;
- Reducing misclassifications that have negative ramifications for children and adolescents;
- Improving and optimizing treatment planning for young people with psychopathic traits, who may require more intensive and risk focused therapeutic approaches.

Interventions with juvenile delinquents

A large number of different treatments have been used to reduce antisocial behaviour. These include psychotherapy, pharmacotherapy, school interventions, residential programmes, and social treatments. Kazdin reported over 230 available psychotherapies, the great majority of which had not been systematically studied.[70] This chapter will focus on treatments with a testable scientific basis which have been evaluated in randomized trials and applied to populations of young offenders.[71]

Meta-analyses of treatment approaches to juvenile delinquency have produced reasonably consistent findings.[72–75] Lipsey[73] considered nearly 400 group-comparison studies published since 1950. The main finding was that there was an overall reduction of 10 per cent in re-offending rates in treatment groups as compared to untreated groups. As might be expected, there were of course considerable variations in the results of individual studies. The best results were obtained from cognitive behavioural, skills-orientated, and multi-modal methods. The results from deterrent trials were

particularly poor, though the numbers in these studies were relatively small. Specifically, treatment approaches that were participatory, collaborative and problem-solving were particularly likely to be beneficial. Family and parenting interventions also seem to reduce the risk of subsequent delinquency among older children and adolescents.[76]

McGuire and Priestley[52] identified six principles for effective programmes:

1 Intensity should match the extent of the risk posed by the offender.
2 A focus on active collaboration, which is not too didactic or unstructured.
3 Close integration with the community
4 Emphasis on behavioural or cognitive approaches.
5 Delivered with high quality with training and monitoring of staff.
6 Focus on the proximal causes of offending behaviour (peer groups, promoting current family communication, and enhancing self-management and problem-solving skills) rather than distal causes (early childhood).

The reviews suggest that there are a number of promising targets for treatment programmes, which include antisocial thoughts, antisocial peer associations, promotion of family communication and affection, promotion of family supervision, identification of positive role models, improving problem-solving skills, reducing chemical dependencies, provision of adequate living conditions, and helping the young offender to identify high risk situations for antisocial behaviours. Conversely, the systematic reviews have also suggested a number of approaches that are unlikely to be promising. For instance, improving self-esteem without reducing antisocial cognitions is unlikely to be of value. Similarly, it is unlikely that a focus on emotional symptoms that is not clearly linked to criminal conduct will be of great benefit.

Life experiences associated with treatment ressistance are summarized in table 11.7.4.

Promising interventions for adolescent antisocial behaviour

A rational starting point when considering interventions would be to consider the main causal factors and processes (see above), and design interventions around them. However, in practice many other considerations have shaped interventions, from the desire to punish offending youths, to making use of what is currently available at relatively low cost. For generic interventions for conduct disorder also highly relevant to juvenile delinquency and serious antisocial behaviour (see Chapter 9.2.5).

Table 11.7.4 Life experiences associated with treatment resistance

Early modelling experiences
Early exposure to related phenomena
Enduring antisocial behaviours and aggressive response patterns
Limited judgement skills
Low academic achievement
Clusters of confrontative acts
Personality traits of callousness
Jealousy and revenge
Limited parental/carer supervision
Erratic punishment schedules
Absent, neglecting, or abusive parenting
Parental psychopathology

Reproduced from Losel F, Bender D, Protective factors and resilience. In (eds. Farrington DP, Cold JW) Prevention of Adult antisocial behaviour, p. 130–204. copyright 2003, with permission of Cambridge University Press.

Working with young offenders with mental health problems—some practice points for interventions

Interventions with juvenile offenders, regardless of whether they are addressing offending behaviour or mental health problems, should take into account developmental and cognitive factors significant to this age group. Interventions designed for use with adults are usually highly structured and target driven. This style of intervention is often perceived by juvenile offenders as alienating and not relevant to their lives with the result that they are likely to disengage (either actively or passively) from the programme.

The skills required to engage a juvenile offender in a 'therapeutic alliance' are often different to those necessary with adult offenders. Adult offenders are more likely to see the value of participating in an enhanced thinking skills course, possibly as a means to an early release or to improve the quality of relationships within their family unit. Such goals may be perceived as too long range to have any meaning to a juvenile offender, or simply seem irrelevant. Juvenile offender's general experience of relationships with adults, particularly professionals, is of authority figures that give instructions, set limits on behaviour, and at best are givers of information. The typical responses to this are to adopt an aggressive posture or one of passive indifference. To actively engage with juvenile offenders professionals need to listen attentively and show interest in the young person's perspective. This does not mean agreeing with the young person's point of view; however, it is an opportunity to establish the 'middle ground' within which a therapeutic alliance may be fostered.

Adolescent offenders have often experienced unstable lives with disrupted attachments. Thus they often have difficulty in understanding the significance of life events such as trauma and bereavement that an adult will readily understand is likely to impact on emotional well being. A thorough assessment prior to commencing an intervention and drawing on material from multiple sources, particularly parents or professionals who have a detailed knowledge of the young person, is very helpful. Events such as the loss of a parent and the onset of conduct problems are often closely linked temporally, yet young offenders frequently do not see any connection between such events.

A formulation is a brief statement that summarizes the possible links between different aspects of the young person's life, for instance between a bereavement and the onset of behavioural difficulties. Juvenile offenders are often unable to make formulations because of their failure to understand how different elements of their lives are connected. The ability to make formulations can be seen as a developmental milestone and adolescent offenders, for the reasons identified above, often lag behind their own peer group as well as the adult population. Regardless of whether addressing offending behaviours or mental health problems, professionals working with juvenile offenders need to generate

such formulations collaboratively so that they make sense both within the therapeutic framework and also within the young person's life experience.

Establishing therapeutic goals also needs to be collaborative and developmentally appropriate. Adolescence in general is characterized as a time of heightened emotional responses (partly as a result of hormonal changes), a growing but still limited capacity for problem solving and the tendency to seek immediate advantage rather than long-term gain. All of these factors are likely to be enhanced in juvenile offenders in contrast with the rest of their peer group. Goal setting therefore needs to concentrate heavily on the short-term, i.e. within the young person's perceived time frame. Targets and rewards should be low key in order to reduce the likelihood of extreme emotional responses to success or failure. Therapists working with juvenile offenders need to be active in encouraging the generation of alternative solutions in order to extend the young persons range of problem solving skills.

Juvenile offenders' capacity to generate alternative strategies is often limited by their own, limited emotional range. They will typically respond to any adverse event with hostility and anger; events that would typically evoke a response of sadness or fear within adults. Work on emotional recognition with juvenile offenders will assist them in recognizing a wider range of emotions, both in themselves and others. This enhances their range of options when faced with future adverse events.

There is often a tendency to concentrate on behavioural objectives as the most easily recognized or measured outcomes. However, working on goals such as recognizing and managing arousal levels, or shifting cognitions or attributions in specific situations may prove more beneficial in the longer term even if immediate behavioural changes are not achieved.

Goal setting and intervention strategies should be individually tailored and take into account differences in cognitive ability, maturity and insight within this population. To ensure the young person remains actively engaged in the intervention process it is important to frequently check out their perception of how effective the therapy is, and whether the goals and strategies are relevant. Therapists should frequently check the young person's level of understanding to ensure that the communication is two way and repeats elements or themes as necessary to ensure good compliance and comprehension, rather than adhering to a timetable.

Treatment and special crimes

Juvenile homicide

Violent behaviour often involves a loss of sense of personal identity and of personal value. A young person may engage in actions without concern for future consequences or past commitments.

Violence denotes the 'forceful infliction of physical injury'.[77] Aggression involves harmful, threatening or antagonistic behaviour.[78] Longitudinal studies are invaluable in mapping out the range of factors and processes that contribute to the development of aggressive behaviour and in showing how they are causally related.[79] However, in attempting to work with any individual who has committed a violent act, the question to be answered is 'why this individual has behaved in this unique fashion on this occasion'.[73]

Studies show that children and adolescents who murder share a constellation of psychological, cognitive, neuropsychiatric, educational and family system disturbance.[80–82] In the UK, young people who commit grave sadistic crimes including juvenile homicide are liable to periods of lengthy incarceration. Detention itself can provide time for further neuro-developmental, cognitive, and emotional growth. Irrespective of treatment models, the provisions of education, vocational training, consistent role models and continued family contact are of critical importance.

The approach to juvenile delinquency, including juvenile homicide, in the Netherlands is determined by a policy of minimal intervention, with a strongly pedagogical point of view. The emphasis lies on education and treatment rather than on punishment. Cases recorded by the police as murder or manslaughter in the first instance may not ultimately be presented to court as such. Moreover, the pedagogical nature of the punishments is recognizable in the fact that treatment is ordered in most cases, in the form of placement in a juvenile institution, sometimes in combination with imprisonment, rather than a straightforward youth detention or imprisonment. There are no indications of special policies for prosecuting or handling the cases of legal minors suspected of murder or manslaughter. The limited prevalence of this phenomenon is undoubtedly partly due to the fact that every case is judged on its own merits. Incidentally, this is also noted in the USA where a considerably higher number of youths are sentenced for murder and manslaughter.[83]

Youths who have been prosecuted for murder or manslaughter vary only slightly or not al all from other juvenile delinquents on points such as age, gender and ethnic background, and only to a limited extent on risk factors. Murder and manslaughter are committed alone comparatively more often, and on average the perpetrators start their criminal activities at a later age and are much less likely to have previous convictions than other minors taken into judicial youth institutions. At the same time, it is clear that while the group of youths involved in murder and manslaughter may be small, it is anything but homogeneous. There is great variety in terms of motives, victims, modus operandi etc. In simple terms each case stands on its own.

The majority of young persons who have killed initially dissociate themselves from the reality of their act, but gradually experience a progression of reactions and feelings akin to a grief reaction. The young person whilst facing a still adversarial, and public pretrial and trial process has to move safely through the process of disbelief, denial, loss, grief and anger/blame. Post traumatic stress disorder arising from the participation in the sadistic act (either directly or observing the actions of co-defendants) has to be treated, as does trauma arising from their own past personal emotional, physical and/or sexual abuse.

A combination of verbal and non-verbal therapies are effective but qualities such as previous frequent and severe aggression, low intelligence and a poor capacity for insight weigh against a safe outcome.[84] In understanding the role of violence and sadism in a young person's life one has to understand the depth of their sensitivity and reaction to perceived threat and their past maladaptive behaviours aimed at allowing them to feel in control of their lives. In coming to terms with their internal rage, addressing victim empathy saying sorry and reattribution of blame, expression of anger and distress within sessions is expected and is often sexualized in both form and content. This can spill outside sessions when the young person and carers can become collusively dismissive and rejecting of therapists,[85] emphazing the importance of intensive

work to prepare the young person for transition from long term incarceration and re-entry into the community together with extended aftercare.

Sexually abusive behaviour

Sexually inappropriate behaviour in children and adolescents constitutes a substantial health and social problem.[86] Most, but not all, abusers are male, often come from disadvantaged backgrounds with a history of victimization, and sexual and physical abuse[87] and show high rates of psychopathology.[88] Of particular concern are a significant subgroup with mild intellectual disability whose treatment programmes have to be tailored to their level of development and cognitive ability. Young abusers come within the Criminal Justice System but also should be considered in their own right within the child protection framework. Most adult sexual abusers of children started their abuse when adolescents and yet neither ICD10 nor DSMIV has a diagnostic category for paedophilia in those under 16.

A structured carefully planned multi-agency approach is required when working with sexually aggressive younger children and sexually abusive adolescents. The three stages to assessment of juvenile sexual offenders are:

1 Clarification and rapport building

2 Mapping the abuse: the fantasies, strategies and behaviours; and

3 The future, placement treatment and personal change.

The treatment process occurs in the context of:

- The crisis of disclosure;
- Family assessment;
- Therapeutic work in a protective context for the victim; and
- Reconstruction and reunification of the family.

The 'family' in this context may include foster carers, or long term residential carers.

Treatment outcome

The earliest possible interventions with young over-sexualized children, before their patterns of sexually aggressive behaviours become entrenched, are likely to be most effective. However, there is a dearth of longitudinal follow-up studies looking at treatment outcomes with this younger group of children.

Outcomes may be measured by looking at recidivism, treatment outcome or other measures. At present, there are no longitudinal outcome studies of children and adolescents with sexually abusive behaviour which have measured other outcomes such as adult adjustment, attitudinal change or parenting.

Recidivism as a sole outcome measures for treatment is unlikely to be reliable since persistent sexual behaviour problems in children under the age of criminal responsibility will not appear in crime statistics and conviction statistics for sex offenders of all ages are notoriously unreliable for a variety of reasons including failure to report victimization experiences, failure to proceed with charges and a high rate of trial failures.

Other factors appear to be highly relevant to treatment outcomes with juvenile sexual abusers such as good interprofessional communication and a systemic context for treatment to occur.

New approaches to CBT with sexually abusing youths have recently been described within the context of relapse prevention and a more complex CBT intervention, Mode Deactivation Therapy (MDT), has been suggested for disturbed, sexually abusive young people with reactive conduct disorders or personality disorders[89] CBT group work with sexually abusing children and young people is widely practised in the UK and the principles of this work are described by Print and O'Callaghan.[90]

Other treatment approaches will take into account the living context of the young person and the need for his or her carers to be provided with support and explanation of the treatment process in order to maximize positive results. For instance, when children and young people who sexually abuse are still living at home or in contact with their parents, family work is usually needed. An approach to group work with parents of children with sexually abusive behaviour has been described.[91] In the case of children and young people who are living in the care system, concurrent work for the professionals and carers looking after the sexually abusing child or young person has been strongly advocated.

There are a significant number of mid-adolescent, recidivist, delinquent, sexually abusive youths who are too dangerous to other children and young people to be treated (with any treatment modality) alongside other young people. Many of these young people have been through the court system or are currently facing charges. For these reasons, treatment of the sexually abusive young person needs to be undertaken within a close supervized, intensive, community-based foster placement with specially trained foster carers who are experienced in dealing with young offenders, risk and dangerousness. This type of approach is known by various names such as Multidimensional Treatment Foster Care[92] or forensic foster care.[93] Early results from small-scale studies with this type of intervention are reasonably encouraging.

Outcomes claimed for these approaches include significantly fewer subsequent criminal referrals and more incarcerated boys returning to live with relatives, compared with those who received group home care alone.[92] In a seven-year study of forensic foster care at the Treatment for Appropriate Social Control (TASC), Yokely and Boettner[93] describe a social responsibility model to teach recidivist youths 'pro-social skills and values that compete with antisocial behaviour'.

Dynamic psychotherapy aims to work at an unconscious level with the sexually abusive young person to explore and understand the reasons for his persistent behaviour. However, evidence-based treatment outcome studies have not yet been undertaken for dynamic therapy with juvenile sexual abusers. A clinical description of long-term dynamic therapy[94] with these children emphasizes the need to establish a systemic child protection context for the safe delivery of such treatment.

In summary, the components of effective treatment interventions with children and young people who sexually abuse will include the following:

- A well planned, systemic, child protection orientated, treatment context.
- Treatment should be one of a number of positive interventions into the life of the young person and his or her family.
- All interventions should be part of an agreed inter-agency care plan for the young person.

- Offence-specific interventions, such as CBT, aimed at straightening out the distorted cognitions and self-justifications of sexually abusing young people should be the core of any intervention programme for this client group.
- Treatment programmes not focussed solely on the victimization of the young person.
- Interventions should occur at all possible levels including individual work with young person, family work (where relevant), support for foster carers or for professional care staff and consultation to the professional network.

Firesetting/arson

Arson can have a devastating impact on the victim and the wider society. Juvenile arsonists are not a homogenous group, with a wide range of familial,[95] social,[96] developmental interpersonal,[97] clinical and 'legal' needs. Kolko and Kazdin[98] highlighted the importance of attraction to fire, heightened arousal, impulsivity and limited social competence. As with other forms of serious antisocial behaviours, no single standard treatment approach will be appropriate for all individuals.[99] In addition to the general assessment of antisocial behaviour the specific domains to be considered include:

- history of fireplay;
- history of hoax telephone calls;
- social context of firesetting (whether alone or with peers);
- where the fires were set;
- previous threats/targets;
- type of fire, single/multiple seats of fire setting;
- motivation (anger resolution, boredom, rejection, cry for help, thrill seeking, fire fighting, crime concealment, no motivation, curiosity, and peer pressure).

For recidivistic firesetters therapy may include:

- Psychotherapy to increase the understanding of the behaviour, including antecedents defining the problem behaviour, and establishing the behavioural reinforcers;
- Skills training—to promote adaptive coping mechanisms;
- Understanding environmental factors to manage or self trigger solutions;
- Counselling to reduce psychological distress;
- Behavioural techniques to extinguish the behaviour;
- Education to promote understanding of cause and effect; and
- Supervision for the staff caring for the adolescent.

Early modelling experiences and early exposure to related phenomena militate against a good outcome.

The role of specialist child and adolescent mental health services in medico–legal assessment

In the case of T&V v United Kingdom[100] it was stated that a child's age, level of maturity and intellectual and emotional capacities must be taken into account when they are charged with a criminal offence and that appropriate steps should be taken in order to promote their ability to understand and participate in the court proceedings. A responsibility therefore falls on the defence lawyer to be aware of the possibility that a young person may not be able to participate effectively in the trial process, particularly if they are under 14 years old or have learning problems, or a history of absence from school.[101] In 1985, the Office of the High Commissioner for Human Rights, in reference to the age of criminal responsibility stated that there is a close relationship between the notion of responsibility for delinquent or criminal behaviour and other social rights and responsibilities.

All young defendants, regardless of the offences they charged with, should be tried in youth courts with permission for adult sanctions for older youths if certain conditions are met. This should enable a mode of trial for young defendants to be subject to safeguards that can enhance understanding and participation. Assessment of cognitive and emotional capacities should occur before any decisions on venue and mode of trial take place.

Capacity

One fundamental distinction in the criminal law is between conditions that negate criminal liability and those that might mitigate the punishment deserved under particular circumstances. Very young children and the profoundly mentally ill may lack the minimum capacity necessary to justify punishment. Those exhibiting less profound impairments of the same kind may qualify for a lesser level of deserved punishment even though they may meet the minimum conditions for some punishment. Immaturity, like mental disorder, can serve both as an excuse and as mitigation in the determination of just punishment. Capacity is sometimes thought of as a generic skill that a person either has or lacks. However, that is not so. To begin with, it is multifaceted, with four key elements. These are as follows:

1 The capacity to understand information relevant to the specific decision at issue (understanding).
2 The capacity to appreciate one's situation as the defendant is confronted with a specific legal decision (appreciation).
3 The capacity to think rationally about alternative course of action (reasoning).
4 The capacity to express a choice among alternatives.[102]

The second key point is that capacity is a feature that is both situation-specific and open to influence.

Any evaluation of competence should include assessment of possibly relevant psychopathology, emotional understanding as well cognitive level, the child's experiences and appreciation of situations comparable to the one relevant to the crime and to the trial, and any particular features that may be pertinent in this individual and this set of circumstances.[103] The general principles to be used in the assessment are broadly comparable to those employed in any clinical evaluation. However, particular attention needs to be paid to developmental background, emotional and cognitive maturity, trauma, exposure and substance misuse. The likely appropriate sources for obtaining clinical data relevant to assessment of a juvenile's competence to stand trial will include a variety of historical records, a range of interviews and other observations and in some cases, specialized tests. Records of the child's school functioning,

past clinical assessment, treatment history and previous legal involvements need to be obtained. In coming to an overall formulation, there should be a particular focus on how both developmental and psychopathological features may be relevant to the forensic issues that have to be addressed.

The main focus is on the youth's ability to understand and cope with the legal process. This comes from three sources, direct questioning of the defendant, inferences from the functioning in other areas and direct observation of the defendant's behaviour and interaction with others. It is useful to enquire about the youth's expectations about what the consequences of the court involvement might prove to be. Because of the course of juvenile proceedings can vary so widely, with consequences ranging from the extremely aversive to extremely beneficial, rational understanding will necessarily involve a high degree of uncertainty. Potentially relevant problems include: inattention, depression, disorganization of thought processes that interfered with the ability to consider alternatives; hopelessness, such that the decision is felt not to matter, delusion or other fixed beliefs that distort understanding of options (or their likely outcomes), maturity of judgement and the developmental challenges of adolescence.

In providing information to the court, written reports have the advantage of a standard format that helps the consultant to be sure that s/he has considered all the relevant questions; it also provides a familiar structure for readers. In essence, for the sake of consistency and clarity, competence reports need to cover the following areas:

1 Identifying information and referral questions.

2 The description of the structure of the evaluation including sources and a notation of the confidentiality expectations.

3 The provision of clinical and forensic data.

Grisso[104] suggested that psychiatric assessment of competence in young people should include assessment of:

- understanding of the charges and the potential consequences
- understanding of the trial process
- capacity to communicate with their defence lawyer
- general ability to participate in the courtroom proceedings.

In court, a child's ability to give an account of events can be impaired by a number of factors, including poor physical health on the day of the trial, overwhelming anxiety or anger about giving evidence, or intimidation by the physical surroundings of the court. From a psychological perspective, however, the basic evidential capacity of the child defendant will depend on two main components:

- The child's mental state—this needs to be stable, therefore any disturbance that might interfere with the child's perception of the world and the ability to understand it will impair evidential capacity;
- The child's cognitive ability—a concept that includes a large number of facets, such as memory, understanding and the ability to communicate. The last includes both verbal (speech) and non-verbal means, as well as the ability both to comprehend and to express thought. Any psychological assessment therefore has to be across a range of domains.

Discrepancies are particularly likely in the areas of educational achievement, adaptive skills and social and emotional development. A child's ability often is gauged on educational achievement and given as being equivalent to that of a certain age—e.g. a 15-year-old child might have the everyday living skills of a 7-year-old. However, a child who might be unable to cope with monetary change or public transport might well have the emotional and social experiences of an older child and the drives of an adolescent.

When discussing developmental psychology and child development, it is important to bear in mind that none of these processes operates in a vacuum. The child's experience of parenting (important in relation to physical and emotional development), the provision of appropriate role models (moral development and self-control depend heavily on appropriate modelling and social learning) and the learning environment (whether it fostered or hindered intellectual development) all have a vital role. For instance, during adolescence, as young people take on a wider and more social perspective and become integrated within a peer group, they will nevertheless tend to adopt social values and norms (i.e. ideas about 'right and wrong') that are very similar to those of their parents. Hence, despite any demonstrations of teenage rebellion (often short-lived), the majority of adolescents will tend to adopt parental mores, either law-abiding or delinquent.

It should be emphasized that clear-cut ages do not apply to the completion of physical, intellectual, emotional, and social development. For most young people, given appropriate parenting, normal biological development and a structured, emotionally supportive and stimulating environment, the bulk of the aforementioned processes should be achieved by late teenage years and a considerable degree of intellectual maturation may have occurred by the age of 14 years.

When delivering forensic mental health services for children and adolescents it is important that the services are developed in such a way that their needs are met and that the services build on established concepts of service design in line with a strategic framework. Doing so will require long term planning that actively addresses the requirements of an adequate size and composition of an appropriately trained, supervized and managed workforce. Such services should be developed with an awareness of the scope of existing services and recognition of current demands, analysing gaps current services.

Adolescent girls

Longitudinal data demonstrates that girlhood aggression contributes to a cascading set of negative outcomes as young women move into adolescence and adulthood.

Young girls who engage in disruptive behaviour and fight are at risk for:

- Being rejected by peers
- Feeling alienated
- Feeling unsupported in their relationships with peers and adults
- Struggling academically
- Affiliating with other peers prone to deviant behaviour
- Becoming involved in more serious antisocial behaviours
- Choosing antisocial romantic partners
- Initiating and receiving partner violence

- Becoming adolescent mothers
- Having children with more health problems
- Being less sensitive and responsive as parents

Some are sufficiently antisocial and even violent, they are incarcerated, if they are also mothers they may lose custody of their children and opportunities for stable employment and relationships are much diminished.

Given low base rates of girls engaging in physical aggression and violence, identifying girls at risk is a critically important step for prevention and intervention programmes.

High-risk groups of girls to target include those girls who are temperamentally overactive as toddlers and pre-schoolers (fewer than boys) but even more important to target for early identification and intervention. Those who have early pubertal development (girls report engaging high levels of bullying as they enter puberty). They may be likely targets as well as perpetrators and sexually abused girls—especially those abused by their biological fathers over a long period of time.

From the available literature interventions to reduce rates of aggression, relational aggression and violence in female children and adolescents should address the following:

1 Pre-natally delivery of programmes for high-risk expectant mothers, (especially young mothers and those themselves aggressive or disruptive as children).

2 Augment the parenting skills of at risk young mothers—the evidence show children of young mothers with histories of girlhood aggression may themselves be more prone to infection and injuries.
   (a) Provide additional parenting skills around key issues of hygiene, child proofing of homes, good nutrition, meal planning and household management.
   (b) Help young mothers to respond optimally to the perceived challenging behaviour of their infants and toddlers.

3 Middle childhood, girls episodes of physical aggression are often preceded by relational aggression. Interventions to help these girls may include how to:
   - Relate to others
   - Manage strong emotions
   - Understand own aggressive feelings and recognize when own aggression is adaptive in the immediate situation.
   - Understand how 'girl talk' ignites hurtful indirect social relational aggression.
   - Understand how relational aggression lead to physical violence for some girls towards peers, adults and partners

Conclusions

The major challenge of altering the trajectories of persistent young offenders has to be met in the context of satisfying public demands for retribution, together with welfare and civil liberties considerations. In England and Wales for example we lock up more than 3 000 juveniles at anyone time (age of criminal responsibility set at 10).

Treatment of delinquents in institutional settings has to meet the sometimes contradictory need to control young people, to remove their liberty and to maintain good order in the institution, at the same time as offering education and training to foster future prosocial participation in society and meeting their welfare needs. At least in England and Wales, the legislative overhaul of Youth Justice[105] has mandated practitioners to bridge the gap between residential and community treatments and to involve families using Youth Offending Teams (YOT's) to meet this complex mix of needs, but the public demand to remove antisocial youths from the street has led to the implementation of Antisocial Behaviour Orders including children with learning disabilities.

Over the last 30 years there has been a gradual shift in opinion regarding effectiveness of intervention with delinquents, from the 'nothing works' approach to a 'what works' approach. The pressure from politicians and public will remain, for a quick fix solution to problems that span cultures, countries and generations.

Provision of appropriately designed programmes can significantly reduce recidivism amongst persistent offenders. The mode and style of delivery is important; high quality staff and staff training are required. Community based programmes seem better than institution based programmes. In prison settings, the strongest effects are obtained when programmes are integrated into the institutional regimes.

What are the key levers for change?

- Developing an integrated comprehensive screening and assessment tool for mental health, substance misuse and physical health
- Better identification of mental health needs by courts with effective court diversion.
- Improving access to child and adolescent mental health services for 16 to 17 year olds across the board.
- Enhancing the role of health workers in YOT's.
- Integrating work of substance misuse workers in custody and YOT's
- Enhancing intensive resettlement and aftercare provision—RAP Schemes.
- Reviewing the demand and need of nationally commissioned adolescent psychiatric secure inpatient beds (88 by 2008).
- And ensuring all working with young offenders understand and are trained in normal and abnormal child and adolescent development together with awareness of the nature of mental health problems in this stage of the life course and how this impacts on all aspects of a young person's life.

Our knowledge of true prevalence rates of mental disorders in a young offending population is developing further. Child and adolescent mental health practitioners have the skills to set the understanding of delinquency in a developmental context and treat those young offenders with mental disorders. Knowledge is advancing rapidly by an established international research network in Europe and the USA.

Further information

Moffitt, T.E. and Scott, S. Conduct Disorders of Childhood and Adolescent. In *Rutter's Child and Adolescent Psychiatry* (5th edn) (eds. M. Rutter, D. Bishop, D. Pine, S. Scott, J. Stevenson, e. Taylor and A. thapar), Chapter 35, pp. 543–54.

Gives a comprehensive review of Conduct disorder classification, prevalence, subtypes, associated co-occurring dsorders and complicating conditions, tackling risk indicators, diagnostic challenges

and evidence based interventions as such it offers a good context in which to set the current chapter.

References

1. Frick, P.J. (2006). Developmental pathways to conduct disorder. *Child and Adolescent Psychiatric Clinics of North America*, **15**(2), 311–32.
2. Cavadino, P., Allen, R. (2000). Children who kill: trends, reasons and procedures. In (ed. G. Boswell) *Violent Children and Adolescents: sking the question why*. London: Whurr Publishers, p. 16–7.
3. Royal College of Psychiatrists. (2006). Occasional Paper: Child Defendants. Royal College of Psychiatrists.
4. Coleman, J., Schofield, J. (2003). *Key Data on Adolescence. Trust for Study of Adolescence*. Brighton: TSA Publishing Ltd.
5. Bailey, S. (2006).Adolescence and beyond: twelve years onwards. In *The Developing World of the Child* (eds. J. Aldgate, D. Jones, W. Rose, C. Jeffrey), p. 208–25. Jessica Kingsley, London.
6. Youth Justice Board. (2007). *Youth Justice Annual Statistics* 2005/06. youth Justice Board for Engalnd and Wales.
7. Youth Justice Board. (2005).*Risk and Protective Factors*. Research Undertaken by Communities that Care on Behalf of the Youth Justice Board for England and Wales.
8. Department of Health. (2004). *National Service Framework for Children, Young People and Maternity Services. (NSF)*. Department of Health.
9. Kazdin, A.E. (2000). Adolescent Development, Mental Disorders, and Decision Making of Delinquent Youths. In: T. Grisso, R.G. Schwartz, editors. *Youth on Trial, A Developmental perspective on juvenile justice 2*. Chicago: University of Chicago Press. p. 33–65.
10. Chitsabesan, P., Kroll,L., Bailey, S., *et al.* (2006). Mental health needs of young offenders in custody and in the community. *British Journal of Psychiatry*, **188**, 534–40.
11. Dixon, A., Howie, P., Starling, J. (2004). Psychopathology in female juvenile offenders. *Journal of Child Psychology and Psychiatry*, **45**(6), 1150–8.
12. Gosden, N.P., Kramp, P., Gabrielsen, G., *et al.* (2003). Prevalence of mental disorders among 15-17-year-old male adolescent remand prisoners in Denmark. *Acta Psychiatrica Scandinavica*, **107**(2), 102–10.
13. Lederman, C.S., Dakof, G.A., Larrea, M.A., *et al.* (2004). Characteristics of adolescent females in juvenile detention. *International Journal of Law and Psychiatry*, **27**(4), 321–37.
14. McCabe, K.M., Lansing, A.E., Garland, A., *et al.* (2002). Gender differences in psychopathology, functional impairment, and familial risk factors among adjudicated delinquents. *Journal of the American Academy of Child and Adolescent Psychiatry*, **41**(7), 860–7.
15. Ruchkin, V., Koposov, R., Vermerien, R., *et al.* (2003). Psychopathology and age at onset of conduct problems in juvenile delinquents. *Journal of Clinical Psychology*, **64**, 913–20.
16. Teplin, L.A., Abram, K.M., McClelland, G.M., *et al.* (2002). Psychiatric disorders in youth in juvenile detention. *Arch Gen Psychiatry*, **59**(12), 1133–43.
17. Vreugdenhil, C., Doreleijers, T.A.H., Vermeiren, R., *et al.* (2004). Psychiatric disorders in a representative sample of incarcerated boys in the Netherlands. *Journal of the American Academy of Child and Adolescent Psychiatry*, **43**(1), 97–104.
18. Wasserman, G.A., McReynolds, L.S., Lucas, C.P., *et al.* (2002). The voice DISC-IV with incarcerated male youths: Prevalence of disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41(3), 314–21.
19. Atkins, D., Pumariega, A.J., Rogers, K., *et al.* (1999). Mental health and incarcerated youth.I: Prevalence and nature of psychopathology. *Journal Child Family Studies*, **8,** 193–204.
20. Shelton, D. (2001). Emotional disorders in young offenders. *Journal of Nurse Scholarship*, **33**(259), 263.
21. Vermerien, R., Schwab-Stone, M., Ruchkin, V., *et al.* (2002). Predicting recidivism in delinquent adolescents from psychological and psychiatric assessment. *Comprehensive Psychiatry*, **43**(2), 142–9.
22. Roberts, R.E., Roseblatt, A. (1998). Prevalence of psychopathology among children and adolescents. *American Journal of Psychiatry*, **155**(715), 725.
23. Moffitt, T.E. (1993). Adolescence limited and life course persistent antisocial behaviour. A developmental taxonomy. *Psychological Review*, **100,** 674–701.
24. Angold, A., Costello, E.J., Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, **40**(1), 57–87.
25. Grisso, T., Zimring, F.E. (2004). *Double jeopardy: Adolescent Offenders with mental disorders*. Chicago: University of Chicago Press.
26. Johnson, T.P., Cho, Y.I., Fendrich, M., *et al.* (2004). Treatment need and utilization among youth entering the juvenile corrections system. *Journal of Substance Abuse Treatment*, **26**(2), 117–28.
27. Domalanta, D.D., Risser, W.L., Roberts, R.E., *et al.* (2003). Prevalence of depression and other psychiatric disorders among incarcerated youths. *Journal of the American Academy of Child and Adolescent Psychiatry*, **42**(4), 477–84.
28. Rutter, M., Giller, H., Hagell, A. (1998). Antisocial Bahviour by Young People: The main messeges from a major new review of the research. Social Information Systems Ltd. Knutsford. Ref Type: Pamphlet
29. Hawkins, J.D., Herrenkohl, T., Farrington, D.P., *et al.* (1998). A review of predictors of youth violence. In Serious *and violent juvenile offenders. Risk factors and successful interventions* (eds. D.P. Farrington), p. 106–45.Thousand Oaks: SAGE Publications,
30. Losel, F., Bliesener, T. (1994). 'Some high-risk adolescents do not develop conduct problems: A study of protective factors'. *International Journal of Behavioural Development*, **17**, 753–77.
31. Sampson, R.J., Laub, J.H. (2003). 'Life-course desisters? Trajectories of crime among delinquent boys followed to age 70'. *Criminology*, **41**, 319–39.
32. Werner, E.E., Smith, R.S. (1992). *Overcoming the Odds*: Ithaca. Cornell University Press.
33. Losel, F., Bender, D. (2003). Protective factors and resilience. In *Early Prevention of Adult Antisocial Behaviour* (eds. D.P Farrington and J.W. Coid), p. 130–204. Cambridge University Press, Cambridge.
34. Luthar, S.S., Cicchetti, D., Becker, B. (2000). 'The construct of resilience: A critical evaluation and guidelines for future work'. *Child Development*, **71,** 543–62.
35. Losel, F., Bender, D. (2003). Protective factors and resilience. In: Farrington DP, Coid JW, editors. *Early Prevention of Adult Antisocial Behaviour*. Cambridge: Cambridge University Press; 2003. p. 130–204.
36. Gowers. S. (2001). Assessing adolescent mental health. In *Adolescent Psychiatry in Clinical Practice* (eds. Gowers S), p. 258–77. Arnold, London.
37. Stevens, A., Rafteryy, J. (1994). Introduction: concepts of need. In *Healthcare needs assessment* (eds.A. Stevens and J. Rafteryy), p. 13–4. Radchliffe Medical Press, Oxford.
38. Bailey, S., Dolan, M. (2004). Violence. In *Adolescent Forensic Psychiatry* (Bailey S and Dolan M), p. 213–27. Arnold publishing, London.
39. Bailey, S., Tarbuck, P. (2006). Recent advances in the development of screening tools for mental health in young offenders. *Current Opinion in Psychiatry*, **19**(4), 373–7.
40. Kroll, L. (2004). Needs assessment in adolescent offenders. In *Adolescent forensic psychiatry* (eds.S. Bailey and M. Dolan), p. 14–26. Arnold Publishing, London.
41. Farrington, D.P., Tarling, R. (1985). *Prediction in Criminology*. Albany State: University of New York Press.
42. LeBlanc, M. (1998). Screening of serious and violent juvenile offenders: identification, classification, and prediction. In *Serious and Violent Juvenile Offenders. Risk Factors and Successful Interventions.* (eds. R. Loeber and D.P. Farrington), p. 167–93. Thousand Oaks, Sage Publications, CA.
43. Hoge, R.D., Andrews, D.A., Leschiedm, A,W. (1996). An investigation of risk and protective factors in a sample of youthful offenders.

Journal of Child Psychology and Psychiatry and Allied Disciplines, **37**(4), 419–24.
44. LeBlanc, M. (2002). The offending cycle, escalation and de-escalation in delinquent behavior, A challenge for criminology'. *International Journal of comparative and applied criminal justice*, **26**(1), 53–84
45. Barnoski, R. (2002). Monitoring vital signs: Integrating a standardized assessment into Washington State's juvenile justice system. In *Multi-problem violent youth: A foundation for comparative research on needs, interventions and outcomes* (eds. R. Corrado, R. Roesch, S.D. Hart, J.K. Gierowski), p. 219–31. IOS Press, Amsterdam, Netherlands
46. Borum, R., Bartel, P., Forth, A. (2002). *Manual for the Structured Assessment of Violence Risk in Youth (SAVRY)*. Tampa, FL: University of South Florida.
47. Shaw, D.S., Gillion, M., Ingoldsby, E.M., *et al.* (2003). Trajectories leading to school-age conduct problems. *Developmental Psychology*, **39**(2), 189–200.
48. Nagin, D., Tremblay, R.E. (2001). Parental and early childhood predictors of persistent physical aggression in boys from kindergarten to high school. *Archives of General Psychiatry*, **58**, 389–94.
49. Fonagy, P. (2003). Towards a developmental understanding of violence. *British Journal of Psychiatry*, **183**, 190–2.
50. Fischer, M., Barkley, R.A., Fletcher, K.E., *et al.* (1993). The Adolescent Outcome of Hyperactive-Children - Predictors of Psychiatric, Academic, Social, and Emotional Adjustment. *Journal of the American Academy of Child and Adolescent Psychiatry*, **32**(2), 324–32.
51. Seagrave, D., Grisso, T. (2002). Adolescent development and the measurement of juvenile psychopathy. *Law and Human Behavior*, **26,** 219–39.
52. McGuire, J., Priestley, P. (1995). Reviewing «what works»: past present and future. In *What Works: Reducing Reoffending* (ed. J. McGuire), p. 3–34. *Guidelines from Research and Practice*. Wiley, Chichester.
53. Dodge, K.A., Schwartz, D. (1997). Social information processing mechanisms in aggressive behavior. In *Handbook of antisocial behavior* (eds. D.M. Stoff, J. Breiling J.D. Maser), p. 171–80. Wiley, New York.
54. Craig, J.S., Hatton, C., Craig, F.B., *et al.* (2004). Persecutory beliefs, attributions and theory of mind: comparison of patients with paranoid delusions, Asperger's syndrome and healthy controls. *Schizophrenia Research*, **69**(1), 29–33.
55. Farrant, F. (2001). *Troubled Inside: responding to the mental health needs of children and young people in Prison*. Prison Reform Trust, London.
56. Dubicka, B., Harrington, R. (2004). Affective Conduct Disorder. In *Adolescent Forensic Psychiatry* (eds. S. Bailey, M. Dolan), p. 124–44. Arnold, London.
57. Harrington, R.C., Kroll, L., Rothwell, J., *et al.* (2005). Psychosocial needs of boys in secure care for serious or persistent offending. *Journal of Child Psychology and Psychiatry*, **46**(8), 859–66.
58. Kroll, L., Rothwell, J., Bradley, D., *et al.* (2002). Mental health needs of boys in secure care for serious or persistent offending: a prospective, longitudinal study. L*ancet*, **359**(9322), 1975–9.
59. Fletcher, K.E. (2003). Childhood posttraumatic stress disorder. In *Child psychopathology* (eds. E.J. Mash, R.A. Barkley), p. 330–71. 2nd ed. Guildford Press, New York.
60. Garbarino, J. (2001). An ecological perspective on the effects of violence on children. *Journal of Community Psychology*, **29**(3), 361–78.
61. Harrington, R., Bailey, S. (2004). NHS National Programme on Forensic Mental Health Research and Development. Expert Paper: The Scope for Preventing Antisocial Personality Disorder by Intervening in Adolescence.
62. O'Brien, G. (1996). The psychiatric management of adult autism. *Adv Psychiatr Treat*, **2**(173), 177.
63. Howlin, P. (1997).*'Autism' preparing for adulthood*. Routledge, London.
64. Bailey, S. (2002). Violent children: A framework for assessment. *Adv Psychiatr Treat*, **8**(2), 97–106.
65. Clark, A. (2001). Proposed treatment for adolescent psychosis. 2: Bipolar illness. Adv Psychiatr Treat, **7**(2), 143–9.
66. Taylor, P.J., Gunn, J. (1999). Homicides by people with mental illness myth and reality. *British Journal of Psychiatry*, **174,** 9–14.
67. Clare, P., Bailey, S., Clark, A. (2000). Relationship between psychotic disorders in adolescence and criminally violent behaviour - A retrospective examination. *British Journal of Psychiatry*, **177,** 275–9.
68. Cooke, D.J., Michie, C. (2001). Refining the construct of psychopathy: towards a hierarchical model. *Psychological Assessment*, **13,** 171–88.
69. Hare, R.D. (1991). *The Hare Psychopathy Checklist - Revised*. Multi Health Systems, Toronto.
70. Kazdin, A.E. (1993). Treatment of conduct disorder: progress and directions in psychotherapy research. *Developmental Psychopathology*, **5,** 277–310.
71. Sukhodolsky, D.G., Ruchkin, V. (2006). Evidence based psychosocial treatments in the juvenile justice system. *Child & Adolescent Clinics of North America*, **15**(2), 501–16.
72. Andrews, D., Zinger, I., Hoge, R., *et al.* (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology*, **28,** 369–404.
73. Lipsey, M.W. (1995). What do we learn from 400 research studies on the effectiveness of treatment with juvenile delinquents? In *What works* (ed. J.McGuire), Reducing offending. Wiley, Chichester.
74. Lipsey, M.W., Wilson, D.B. (1993). The efficacy of psychosocial, educational, and behavioural treatment: Confirmation from meta-analysis. *American Psychologist*, **48,** 1181–209.
75. Losel, F. (1995). The efficacy of correctional treatment: a review and synthesis of metaevaluations. In *What Works* (ed. J. McGuire), p. 57–82. Reducing Reoffending: Guidelines from Research and Practice. Wiley, Chichester.
76. Woolfenden, S.R., Williams, K., Peat, J. (2003). *Family and parenting interventions in children and adolescents with conduct disorder and delinquency aged 10-17 (Cochrane Review)*. The Cochrane Library, Issue 3, Oxford.
77. Blackburn, R. (1993). *The psychology of criminal conduct*. Wiley, Chichester UK.
78. Berkowitz, L. (1993). *Aggression: Its causes, consequences and control*. McGrow-Hill, New York.
79. Farrington, D.P. (1995). The Twelfth Jack Tizard Memorial Lecture. The development of offending and antisocial behaviour from childhood: key fi ndings from the Cambridge Study in Delinquent Development. *Journal of Child Psychology and Psychiatry and allied disciplines*, **36**(6), 929–64.
80. Cornell, D.G., Benedek, E.P., Benedek, B.A. (1987). Juvenile homicide. Prior adjustment and a proposed typology. American *Journal of Orthopsychiatry*, **57**(3), 383–93.
81. Myers, W.C., Burket, R.C., Harris, H.E. (1995). Adolescent psychopathy in relation to delinquent behaviours, conduct disorder, and personality disorder. *Journal of Forensic Sciences*, **40**, 436–40.
82. Myers, W., Scott, K. (1998). Psychotic and conduct disorder symptoms in juvenile murderers. *Journal of Homicide Studies*, **2**(2), 160–75.
83. Zimring, F. (1998). *The Challenge of Youth Violence*. Cambridge University Press, Cambridge.
84. Bailey, S. (1996). Current perspectives on young offenders: aliens or alienated? *Journal of Clinical Forensic Medicine*, Mar, **3**(1), 1–7.
85. Heide, K.M. (1999). *Young Killers: The Challenge of Juvenile Homicide*. Thousand Oaks, Sage publications: California.
86. James, A.C, Neil, P. (1996). Juvenile sexual offending: One-year period prevalence study within Oxfordshire. *Child Abuse and Neglect*, **20**(6), 477– 85.
87. Skuse, D., Bentovim, A., Hodges, J., *et al.* (1998). Risk Factors for Development of Sexually Abusive Behaviour In Sexually Victimised Adolescent Boys. Cross Sectional Study. *British Medical Journal*, **317,** 175–9.
88. Dolan, M., Holloway, J., Bailey, S., *et al.* (1996). The psychosocial characteristics of juvenile sexual offenders referred to an adolescent forensic service in the UK. *Medicine Science and the Law*, **36**(4), 343–52.

89. Apsche, J.A., Ward Bailey, S.R. (2005).Mode Deactivation Therapy: Cognitive Behavioral Therapy for Adolescents with Reactive Conduct Disorders and/or Personality Disorders/ Traits. In *Children and young people who sexually abuse: New theory, research, and practice developments* (ed. M.C. Calder), Russell House, United Kingdom.
90. Print, B., O'Callaghan, D. (1999). 'Working with young men who have sexually abused others'. In *Children and young people who sexually abuse others* (eds. M. Erooga and H. Masson), p. 124–45. Routledge, London.
91. Hackett, S., Telford, P., Slack, K. (2002). Groupwork with parents of children who sexually harm. In *Young People who Sexually Abuse* (ed. M.C. Calder), Building the Evidence Base for your Practice. Lyme Regis, Russell House Publishing.
92. Chamberlain, P., Reid, J.B. (1998). Comparison of two community alternatives to incarceration for chronic juvenile offenders. *Journal of Consulting & Clinical Psychology*, **66**(4), 624–33.
93. Yokely, J., Boettner, S. (2002). Forensic foster care for young people who sexually abuse: Lessons from treatment. In *Young people who sexually abuse* (ed. M.C. Calder), p. 309–32. Building the evidence base for your practice. Lyme Regis: Russell House.
94. Vizard, E., Usiskin, J. (1999). Providing individual psychotherapy for young sexual abusers of children. In *Children and young people who sexually abuse others: Challenges and responses* (eds. M. Erooga and H. Masson), p. 104–23. Taylor & Frances/Routledge, Florence, KY, US.
95. Fineman, K.R. (1980). Firesetting in childhood and adolescence. *Psychiatric Clinics of North America*, **3**(3), 483–500.
96. Patterson, G.R. (1982). Coercive family process. Eugene OR, Castalia.
97. Vreeland, R.G., Lowin, B.M. (1980). Psychological aspects of firesetting. In *In Fires and human behaviour* (ed. D. Canter). Wiley, New York .
98. Kolko, D., Kazdin, A.E. (1992). The emergence and re-occurrence of child firesetting: A one year prospective study. *Journal of Abnormal Child Psychology*, **201**, 17–37.
99. Repo, E., Virkunnen, M. (1997). Young arsonists, history of conduct disorder, psychiatric diagnosis, and criminal recidivism. *Journal of Forensic Psychiatry*, **8**, 311–20.
100. T & V v United Kingdom. (1999). European Centre of Human Rights Judgements.
101. Ashford. M., Chard, A., Redhouse, N. (2006). *Defending Young People in the Criminal Justice System*. 3rd ed. Legal Action Groups, Glasgow.
102. British Medical Association. (2001). *Health Care for Children and Young People: Consent Rights and Choices*. British Medical Association, London.
103. Grisso, T. (1997). The Competence of Adolescents As Trial Defendants. *Psychology, Public Policy and Law*, **3**, 3–32.
104. Grisso, T. What we know about youths' capacities as trial defendants. In *Youth on Trial* (eds. T. Grisso and R.G. Schwartz), p. 139–71. University of Chicago Press, Chicago.
105. Crime and Disorder Act: Section 37 (1998). The Stationary Office, London.

11.8

Child molesters and other sex offenders

Stephen Hucker

Introduction

In most Western societies sexual offenders are more reviled than almost any other type of offender. On both sides of the Atlantic this is reflected in the sanctions that specifically address this group such as Sexually Violent Predator laws in the United States, Dangerous and Long-Term Offender legislation in Canada, and Sex Offender Orders in the UK. Related approaches include the introduction of sex offender registries and the widespread requirement that children at risk from sexual predators be reported by professionals and others.

The general psychiatrist and the sex offender

Though constituting a relatively small proportion of all reported offences, sex offending affects large numbers of people in the general population. Though there are methodological difficulties associated with much of the research in this area, the World Health Organization has reviewed estimates of childhood sexual abuse from 39 countries and found that the prevalence of non-contact, contact, and intercourse in female children was about 6, 11, and 4 per cent, respectively; the corresponding figure for males was about 2 per cent in all categories.[1] However, these must represent minimal figures as it is known that many victims do not report their experience and individual sex offenders, when guaranteed confidentiality, will admit to many more offences than they were charged with or convicted of.[2] Various forms of sexual offending against adults are also underreported but it has been estimated that about 13 per cent of women and 3 per cent of men have been raped at some time during their lifetime.[3]

The long-term effects of such victimization has been extensively studied and it is clear that people with a history of child sexual abuse, for example, experience a wide range of long-term psychological consequences.[4]

Although prone to find reasons to delegate the assessment and management of sex offenders to specialized forensic services, the general psychiatrist will find it impossible to avoid them entirely. Minor varieties, such as 'flashers' (exhibitionists) or 'peeping Toms' (voyeurs), may be viewed by the courts as less serious, and the opinion of the generalist will still be appreciated, especially when specialist resources are scarce and an appointment for a forensic assessment could be long-delayed. It is important, therefore, for the general psychiatrist to have some understanding of this area in order to make appropriate decisions and recommendations.

Definitions of sexual offending

Put simply, a sexual offender is an individual whose sexual behaviour contravenes the law in a particular jurisdiction. The types of activities that may be proscribed vary considerably. Western countries are generally more tolerant though most societies provide sanctions for sexual activity involving children below the age of consent, non-consensual sexual acts, sexual relations with close family, and sexual interference with animals or corpses. There are also typically legal and other interventions where a person fears sexual harassment or assault, and where there has been abuse, or likelihood of abuse, in certain professional relationships. Typically, also, there is regulation of pornography or obscene material.

Relationship between sexual deviancy and sexual offending

There is some overlap of sexual offences with a medical diagnosis of a paraphilia. However, this is not a complete concurrence. Thus, not all paedophiles have molested a child and not all child molesters are paedophiles; many, perhaps most, men who sexually assault adult women are not sexually deviant or paraphilic at all. The psychiatric categories of paraphilia and their characteristics are described elsewhere in this volume (Chapter 4.11.3 by Fedoroff).

Types of sexual offender

The vast majority of sexual offenders are male though it is recognized that women may also commit similar crimes.[5] Male sexual offenders can be broadly divided into: child molesters, rapists, and non-contact sex offenders.

Child molesters

Typologies of this subgroup in part refer to the degree of paraphilic attraction (sexual deviancy). Thus, there has been a common categorization into 'fixated' and 'situational' or 'regressed' types.[6] With the 'fixated' type there is a permanent attraction to children typically dating from adolescence thus conforming to definitions of paedophilia in DSM-IV and ICD-10. Those attracted to males are more likely to repeat their offences with recidivism at least

twice as high as with those attracted to girls. The former tend to victimize boys aged 11–15 years old, whereas the latter molest girls of 8 to 10.

'Fixated' paedophiles tend to commit premeditated offences that often involve considerable planning. Manipulation and 'grooming' behaviour is used as a means of luring, even abducting, children into sexual activity, and they may gain the trust of the parents or other carers. They may appear to have an excellent rapport with the child victims and treat them kindly but their motive is primarily for the child to meet their own need for affection rather than the reverse. It is for this reason that 'needy' children are often selected as victims. Such offenders will typically profess their 'love' for children and convince themselves that their behaviour is not harmful. Other rationalizations, such as that were educating the child or introducing the child to sexual love in a caring way, are common.

'Regressed' or 'situational' child molesters are, according to this typology, attracted primarily to adult females and may be in a marital-type relationship at the time of their offence. They will often report feelings of personal inadequacy or low self-esteem and their offences are more typically spontaneous and occur in the context of a stressful life circumstance.

The 'regressed' child molester, in contrast to the 'fixated' type, is less inclined to 'groom' victims and their caregivers. Victims of this type may be older than those involved with the 'fixated' molester. Molestations however may begin before, and continue past, puberty.

Rapists

Once again, the typology expounded by Prentky, Knight, and colleagues can be useful.[6] In this scheme, offenders are differentiated through their apparent motivation by anger, power, or sadism. Most rapists act alone but individuals involved in 'gang' rape may represent several different types.

'Anger-motivated' rapists act out deviant fantasies of retaliation towards the victim, using violence as a means of expressing generalized anger, typically towards women though sometimes towards people in general. The intention of the attack is to humiliate and debase the victim who will typically have been selected randomly.

The 'power-motivated' rapist type has been further subdivided into the 'power-reassurance' and 'power-assertive' types. The former is plagued by doubts and insecurities about their own masculinity and sexual adequacy, often uses minimal force and may apologize to the victim or even seek a relationship afterwards. However, these victims may have been stalked and the attacks on them premeditated. The 'power-assertive' subtype is motivated by the desire to dominate women and has no doubts about his masculinity but, like the previous subtype, typically does not use gratuitous violence to subdue the victim.

'Sadistic' rapists, are the least common but perhaps most worrisome type, who derive sexual pleasure from inflicting hurt and suffering on their victim. Sometimes, however, this is difficult to differentiate from other types where pain, suffering and humiliation are the consequences, but not the primary motivation, for the attack.

Other types of sexual assault that do not involve penetration, and which represent related behaviours, include toucheurism and frotteurism. These involve, respectively, touching or grabbing strangers, typically females, in a way that provides him with sexual gratification. The former, in particular, may pass unnoticed by the victim, as attackers will typically touch or grab sexual areas such as breasts, buttocks, or crotch, in crowds and similar situations where the incident may be discounted as 'accidental'.

Non-contact sexual offences

This group includes 'peeping Toms', 'flashers', and indecent phone callers with corresponding psychiatric diagnoses of the paraphilias voyeurism, exhibitionism, and telephone scatalogia, respectively.

Peepers have a penchant to observe an unsuspecting female stranger undressing, or couples in the act of copulation. They may masturbate at the scene or later in private while recalling what they saw. Most voyeurs, like most sex offenders who are paraphilic, are aware of their deviant impulses while still adolescent but the behaviour may become chronic.

Exhibitionists derive sexual excitement from exposing their genitals to unsuspecting female strangers. The desired reaction is one of 'shock and awe' and indifference is a useful response, if the victim has the presence of mind. The perpetrator may masturbate at the time or later in private. Though some cases are particularly intractable, it is unusual to see an exhibitionist still active much past the age of about 40 years old.

The unsuspecting victims of obscene telephone callers are greeted with a barrage of sexually explicit commentary over the telephone. When this is a habitual practice, accompanied by or followed by masturbation, it is distinguishable from an isolated incident committed as a prank.

In all three of the above non-contact sexual offences, there is typically no desire to have further contact with the victim and, indeed, such a prospect often fills the offender with anxiety. However, a small minority of these offenders may later commit a more serious sexual offence, even sexually motivated homicide.[7] Atypical features, such as a desire to have personal contact with the victim, or the repeated selection of child victims, may be an early warning sign of such potential and warrants further, expert assessment.

In recent years, paralleling the advance of technology, offenders have been apprehended for using the Internet as a means of obtaining or distributing pornography (especially child pornography) or as a means of contacting children for sexual purposes. Possession of child pornography is an indicator of sexual interest in children and it is associated with self-reported sexual interest in children and with laboratory measurements of changes in penile volume or circumference.[8]

Assessment of sex offenders

In the context of court remanded cases, the usual pre-trial issues of fitness (competence) to stand trial and criminal responsibility will need to be addressed. Only a small proportion of all sex offenders demonstrate symptoms of psychotic mental disorder, particularly schizophrenia and depressive disorders. However, where those disorders are present it is necessary to explore how the symptoms, e.g. delusions or hallucinations, bear specifically on the sexual offending. It may be found that other psychiatric disorders are of greater importance and that a secondary diagnosis of mental retardation or personality disorder, or more particularly, a paraphilia, are more relevant with respect to the offending behaviour. More commonly the court will be interested in sentencing considerations, specifically the risk the offender presents to others, and whether and what medical treatment and/or other professional

interventions might be ordered or recommended by the court in order to reduce that risk.

Interview

The interviewer must assume that the subject will be at best guarded, or frankly hostile, in their response to the assessment process and it may be difficult to achieve the level of affinity more commonly experienced with non-psychotic general psychiatric patients. A non-judgemental approach is therefore preferable, regardless of the examiner's personal emotional reaction to the offender's behaviour. Few will have presented to the psychiatrist or psychologist voluntarily. Moreover, there is a distinct tendency among sex offenders to prevaricate or frankly lie. It is therefore essential for the examiner to have detailed information about the act or acts that are alleged to have occurred. It has already been noted that denials, rationalizations, distortions, and minimization are the norm with sex offenders. It is not, however, the assessor's function to judge the issue, in particular when guilt has not been determined by the court.

It is obviously unhelpful, in terms of gaining rapport or obtaining additional information, to accuse the subject of dishonesty. Rather, a sympathetic approach suggesting that sometimes people have difficulty accepting unpleasant aspects of themselves, may be more helpful. Another approach is to invite the subject to explore why the victim made the accusations if they are untrue and whether they can accept, if not the whole, then some part of the allegations against them.

It is also important to remember that, while the individual may have been accused of one type of deviant sexual behaviour, other types may have occurred and been undetected. Paraphilic disorders tend not to occur as it were in pure culture but rather in association with other paraphilias, typically at least two or three. Thus, for example, an exhibitionist may have been reported for exposing specifically to children rather than adults and this will suggest an additional diagnosis of paedophilia. Or, an individual who has been convicted of rubbing himself against women in public places may also have made obscene phone calls and harbour fantasies of rape.

In the case of child molesters it is important to consider how he gained access to his victims. Exploration of the methods of 'grooming' is important in understanding ways to assist the offender to avoid risky situations in the future.

In terms of the overall assessment, identification of psychopathologies outside the domain of sexual deviation is important. The presence of psychosis will have an implication for the type of treatment to be recommended, even if it is not common among sex offenders. It seems likely that some such sex offenders' behaviour is dismissed as a function of the psychosis and underlying paraphilic disorder may easily be discounted or not considered at all. More commonly it will be personality disorders or traits, alcohol or substance abuse, mild-to-moderate depression, and anxiety disorders, rather than major mental illness that will be noted. Attention to these will be an important part of any subsequent treatment or management strategy.

Psychometric testing

Psychometric testing may contribute additional information an overall assessment, in particular where the subject is not forthcoming in a personal interview. There are a number of general personality assessment instruments available. These include the well known and widely used *Minnessota Multiphasic Personality Inventory* (*MMPI*), the *Millon Clinical Multiaxial Inventory* (*MCMI III*), and the *Personality Assessment Inventory* (*PAI*). All have been extensively used in offender populations including sex offenders, and common profiles have been identified. Although none can specifically identify a sex offender, information concerning impulsivity, denial, judgement, and general psychopathology may be very useful.

There are also a number of psychological tests that have been specifically designed for the assessment of sex offenders. These include the *Multiphasic Sex Inventory II* (*MSI-II*) and the revised version of the *Clarke Sex History Questionnaire for Males* (*SHQ-R*). The *MSI-II* is designed to measure the sexual characteristics of an adult male (though there is a female version too) alleged to have committed a sexual offence or sexual misconduct, including those who deny the allegations. Though standardized in the United States on a large sample of sex offenders it is more widely used. It consists of 560 true/false questions and the completed questionnaire must be sent to the developers for computerized scoring and interpretation. The SHQ-R consists of 508 questions and the completed questionnaire may again be sent away for scoring and an interpretive report returned.

Laboratory testing

As sex offenders are prone to lie and distort their self-report of deviant interest and behaviours a more objective method of assessment has long been pursued. One of the earliest to be developed was the use of the penile plethysmograph (PPG, or phallometry) to measure changes in response to erotic stimulation. This method involves measuring changes in the size of the penis while presenting the subject with carefully selected images, both still and moving, of both sexes and different age groups, and audiotaped descriptions of various sexual activities. There are certainly problems with PPG testing, including the standardization of stimulus materials used, and some offenders are either able to learn to suppress their physiological responses or masturbate before the testing in order to render themselves unresponsive.[8] Nonetheless, the PPG, more commonly using a circumferential device, or the volumetric method, is extensively used in assessments of sex offenders (for critical review, see Ref.[10]).

Particularly in the United States, the PPG has come under attack as, in addition to standardization and reliability issues, it uses pictures of children whose consent or that of their parents was never obtained. The computer-generated images have been developed to attempt to obviate this concern. More recently the use of virtual reality of computer-generated images has also been used experimentally. Other, less intrusive methods of assessment are also being adopted, including the Abel Screen[11] which measures time spent viewing non-nude images.

Mention must also be made of polygraphy with sex offenders as it has been used extensively in many parts of the United States and to some extent in the United Kingdom. The subject is asked questions relating to their sexual interests and activities while their pulse, respiration, and skin conductance are measured. However, research in the area is generally weak and, despite its widespread use and perception of usefulness, particularly for monitoring sex offenders in the community, the method is controversial (for review, see Ref.[8]).

Assessment of risk in sex offenders

The courts will often wish to have a professional opinion regarding the risk an offender presents to re-offend, and it is important to have an understanding of the factors that will contribute to this (see Table 11.8.1).

Risk has been divided into:

1 Static risk, i.e. involving those factors which cannot change, such as the offender's age, sex, or number of previous criminal convictions and

2 Dynamic risk, i.e. involving those which potentially could change, either as a result of treatment or some other intervention, or simply by the passage of time. This can be further subdivided into relatively stable, though nonetheless potentially changeable, factors such as sexual preferences or negative attitudes, and acute factors, such as access to victims, reversion to substance use, and active mental illness.

These are important in estimating an offender's risk to re-offend. Static factors have received the most scientific study and it is chiefly that these have been incorporated into various actuarial or statistical instruments that have been developed in the past several years, based on follow-up studies of samples of sex offenders. Among these instruments the Rapid Risk Assessment of Sex Offender Recidivism (RRASOR), the STATIC-99,[12] and Sex Offender Risk Appraisal Guide (SORAG),[13] are the most widely used. An assessor intending to use any of these instruments needs to be thoroughly familiar with the literature on the topic and to have participated in training workshops that are given at conferences from time to time. Useful though these tools can be, too heavy reliance upon them is no substitute for a full understanding of how these were constructed and their limitations.[14]

Table 11.8.1 Predictors of sexual offence recidivism

Type of risk factor	Predictor
Static	Prior sex offence Prior non-sexual offences Prior non-contact sex offences Prior treatment dropout Any boy victims Any unrelated victims Any stranger victims Early age of onset Young age of offender Minimal cohabitation history Childhood behaviour problems Separation from parents as a child Antisocial personality disorder Prior violation of conditional release
Stable dynamic	Sexual preferences, children Sexual preferences rape/violence Sexually entitled attitudes Pro-child molester attitudes Pro-rape attitudes Lack of adult love partner Emotional loneliness Lifestyle impulsivity Ineffective problem-solving skills Callous and unemotional attitudes Aggressive, hostile, and suspicious Negative social influences
Acute dynamic	Access to potential victims Substance abuse Sexual preoccupation Emotional collapse Collapse of social support system Rejection of supervision Acute mental illness

(Reproduced from Webster, C.D. and Huxler, S.J. (2007) Violence risk: assessment and management. Copyright 2007, John Wiley & Sons, Inc.)

Treatment issues

It is rare for paraphilic individuals to present for treatment in order to prevent themselves from becoming a sex offender. Though some paraphilias are not usually associated with criminal charges (e.g. transvestitic fetishism) others, such as paedophilia, are more likely to present through the courts or probation and parole services *after* an offence has been committed. It is important to realize that sexual preferences are highly resistant if not impossible to change. The most that can be expected with sex offenders who have deviant sexual preferences (paraphilias) is to help them learn to control their behaviour and to recognize that their propensity will always remain in the background, much as alcoholics are advised to consider themselves always vulnerable to relapse.

Psychological treatments

Various psychological therapies have been attempted with sex offenders. Psychodynamically based individual and group treatments have been the most commonly used. It has become clearer more recently, however, that cognitive-based therapies (CBT) are the preferable approach to take, although techniques involving classical behaviour therapy, e.g. covert sensitization, are also sometimes used for specific purposes, such as creating aversion to deviant arousing images and replacement with non-deviant ones. Cognitive behaviour therapy itself involves helping the subject to develop strategies to alter their thought processes in order to avert their deviant behaviour, to improve their social skills, and to remedy their distorted beliefs and attitudes.

There is little or no evidence for the efficacy of psychological treatments prior to the introduction of CBT. Based on meta-analysis of 43 published studies, it has been shown that treatment programmes using this approach are associated with a reduction in overall recidivism rate from about 17 to 10 per cent.[15] Nonetheless, there is still considerable controversy over the effectiveness of formal sex offender treatment programmes.[16]

Medications

Various medications have been used to treat sex offenders. Based on empirical observations of animals, which become less sexually active following neutering, hormonal treatments that reduce testosterone levels have been extensively employed. All require careful discussion with the potential patient concerning side effects and it is important to obtain written, informed consent.

Oestrogens proved to be problematic because of the serious risk of thrombo-embolic complications and a safer alternative was found in Cyproterone acetate (Androcur), an anti-androgen, which is available in Europe, including the UK. Because this drug has not

been made available in North America, Medroxyprogesterone acetate (Provera) was introduced. Both however can still be responsible for minor side effects such as weight gain, tiredness, and gynaecomastia (with Cyperoterone, especially) but also more serious problems including thrombo-embolism and increase in blood sugar. More recently, leuteinizing hormone-releasing hormone (LHRH) agonists such as Leuprolide acetate (Lupron), Goserelin (Zoladex), and others, have been found useful as they produce almost total suppression of testosterone production such as would be seen following surgical castration. They tend, however, to be used mainly in very high-risk offenders or those who have failed with other drugs.[17] All of these hormone-affecting substances, though especially the LHRH agonists, have a tendency to leach calcium from the bones[18] and it is necessary to monitor carefully for this side effect and to administer antidotes including calcium supplements, vitamin D, and possibly biphosphonates.

The main problem with the hormonal treatments is their lack of acceptance by those who might potentially benefit. An alternative in the form of serotonin-reuptake inhibitors has therefore been better received though double blind trials are still lacking. They depress libido in about 50–60 per cent of cases though higher doses, such as are used with obsessive–compulsive disorder, are often necessary. There appears to be little basis on which to chose one SSRI over another, other than patient tolerance of side effects.[19]

Ethical problems

Several ethical issues have been mentioned above in passing. It is, however, worth emphasizing in conclusion that a disinterested professional demeanour is important when assessing sex offenders. No matter what his or her own private views, it is not the place of the clinician to decide on guilt or innocence or in any other way to pass judgement on the offender or alleged offender. Moreover, alienating the offender will present a further impediment to gaining information and to providing treatment when indicated and necessary.

It is important, at the assessment stage, to identify for the subject the nature of the evaluation, the role of the assessor, and the person or agency for whom they are acting, for example a Child Protection Service, a defence lawyer, or Crown prosecutor. This may determine the degree of cooperation but to not explain this fully and simply to present oneself as a 'doctor' in a helping role when the intention of the evaluation is solely to provide a risk assessment as opposed to treatment, for example would be unethical.

Certain assessment procedures, such as penile plethysmography and polygraphy (when used in a clinical setting) are particularly contentious. Though both may provide useful information, written and fully informed consent should always be obtained beforehand. PPG has been particularly criticized for the use of child images as sexual stimuli when the consent of neither the child nor its parents have been obtained, though the development of computer-generated images may avoid that particular objection.

Finally, when drugs that have been developed and marketed for other purposes are used to suppress sexual drive, the patient needs to be fully informed of the potential benefits as well as the risks involved but should not be denied complementary or alternative treatments should they decide not to expose themselves to the potential side effects.

Further information

Holmes, R.M. and Holmes, S.T. (2002). *Current perspectives on sex crimes*. Sage Publications, London.

Marshall, W.L., Fernandez, Y.M., Marshall, L.E., *et al.* (2006). *Sexual offender treatment: controversial issues*. John Wiley & Sons, Chichester, England.

References

1. World Health Organization. (2002). Other risks to health. In *The world health report 2002: reducing risk, promoting healthy life* (Chap. 4) @ http://www.who.int/whr/2002/chapter4/en/index9.html
2. Abel, G., Becker, J., Mittleman, M., *et al.* (1987). Self-reported sex crimes of nonincarcerated paraphiliacs. *Journal of Interpersonal Violence*, **2**, 3–25.
3. Spitzberg, B. (1999). An analysis of empirical estimates of sexual aggression victimization and perpetration. *Violence and Victims*, **14**, 241–60.
4. Romans, S., Martin, J., and Mullen, P. (1997). Childhood sexual abuse and later psychological problems: neither necessary, sufficient nor acting alone. *Criminal Behaviour and Mental Health*, **7**, 327–38.
5. Hislop, J. (2001). Female sex offenders: what therapists, law enforcement and child protective services need to know. Issues Press, Ravensdale, WA.
6. Prentky, R. and Burgess, A.W. (2000). *Forensic management of sexual offenders*. Kluwer Academic/Plenum, New York.
7. Ressler, R.K., Burgess, A.W., and Douglas, J. E. (1988). *Sexual homicide: Patterns & Motives*. D.C. Heath & Co., Toronto.
8. Seto, M. (2007). Pedophilia and sexual offending against children: theory, assessment, and intervention. American Psychological Association, Washington, DC.
9. Marshall, W.L. and Fernandez, Y.M. (2003). *Phallometric testing with sexual offenders*. Safer Society Press, Brandon, VT.
10. Laws, R. (2003). Penile plethysmography: will we ever get it right? In *Sexual deviance: issues and controversies* (Chap. 5) (eds. T. Ward, D.R. Laws, and S.M. Hudson). Sage Publications, London.
11. Abel, G., Huffman, J., Warberg, B., *et al.* (1998). Visual reaction time and plethysmography as measures of sexual interest in child molesters. *Sexual Abuse: A Journal of Research and Treatment*, **10**, 81–95.
12. Harris, A., Phenix, A., Hanson, K., (2003). *STATIC-99 coding rules revised—2003*. Solicitor General Canada, Ottawa, Ontario, Canada.
13. Quinsey, V.L., Harris, G.T., Rice, M.E., *et al.* (2006). *Violent offenders: appraising and managing risk*. American Psychological Association, Washington, D.C.
14. Webster, C.D. and Hucker, S.J. (2007). *Violence risk: assessment & management*. John Wiley & Sons, Chichester.
15. Hanson, K., Gordon, A., Harris, A., *et al.* (2002). First report of the collaborative outcome data project on the effectiveness of psychological treatment for sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, **14**, 169–92.
16. Laws, D. R. and O'Donohue, T. (eds.) (2008) *Sexual Deviance* (2nd ed). pp. 7–14. The Guilford Press: New York and London.
17. Hill, A., Briken, P., Kraus, C., *et al.* (2003). Differential pharmacological treatment of paraphilias and sex offenders. *International Journal of Offender Therapy and Comparative Criminology*, **47**, 407–21.
18. Grasswick, L.J. and Bradford, J.M. (2003). Osteoporosis associated with the treatment of paraphilias: a clinical review of 7 case reports. *Journal of Forensic Sciences*, **48**, 849–55.
19. Greenberg, D. and Bradford, J. (1997). Treatment of paraphilic disorders: a review of the role of selective serotonin reuptake inhibitors. *Sexual Abuse: A Journal of Research and Treatment*, **9**, 349–60.

11.9

Arson (fire-raising)

Herschel Prins

How great a matter a little fire kindleth!

New Testament, authorised version (Letter of James, 3:5)

The title of this chapter merits brief comment. Why is arson (fire-raising) a special problem, and why is there a dual title? Arson is a special problem because not only is it regarded as a very serious form of criminal behaviour, but because its detection can be very difficult. It is an offence that can be committed at 'one removed' by an offender, and it may sometimes involve unintended victims. Forensic psychiatrists will meet a number of arsonists in the course of their work but, increasingly, general psychiatrists are likely to come across them from time to time, for two reasons. *First*, the High Court in England and Wales has suggested that psychiatric reports are advisable in all cases where the motivation for the offence is unclear. *Second*, as will become clear from the writer's later comments, there has been a worrying increase in arson committed by young adults and children. Thus it is probable that child and adolescent psychiatrists are likely to come across cases more frequently. However, Soothill indicates a word of caution. He suggests that over-reliance upon psychiatric involvement may tend to 'medicalize' socially problematic behaviour.[1] The dual title that heads this chapter indicates the legal term used in the United Kingdom to describe acts of unlawful fire-raising. However, the term is not in universal use; readers will find that in other jurisdictions (notably in the North Americas) the terms are fire-raising, fire-setting, incendiarism, and in certain specific instances, pyromania and pathological fire-raising.[2] The term fire-raising is used in this chapter since the alternative terminologies can be subsumed under it.[3,4]

Brief historical context

The phenomenon of fire, its uses and misuses, has figured extensively in myth, legend, and literature. For example, Prometheus is said to have stolen fire from the Gods and the myth became the mainspring for much psychoanalytic theorizing about fire-raising behaviour.[3] There are numerous early historical references to incendiary mixtures and devices, including sketches for mortars by Leonardo da Vinci.[5] In the mid-nineteenth century, the medical profession became interested in the explanation of fire-raising behaviours; and subsequently, adherents of psychoanalysis proposed various complex and somewhat doubtful explanations for such conduct. In particular, they linked fire-raising behaviour to sexual disturbance of one kind or another. Although sexual problems do appear in the backgrounds of *some* recidivist fire-raisers, the importance of the links has, in the present writer's view, been somewhat overstated.[6] Having said this, it should perhaps be noted that the phenomenon of fire is not infrequently linked linguistically to aggression and sexuality. For example, we speak or write of 'white hot rage', 'heated arguments', 'inflamed passions', to have the 'hots' for a sexual partner. Language is the conveyor of cultural values and attitudes and can be a powerful force in influencing our modes of thinking and expression about the phenomenon of fire in its many manifestations.

The size of the problem

During the past two decades concern about the increase in fire-raising has been expressed worldwide. Table 11.9.1 gives the number of offenders convicted of arson in *England and Wales* for the years 1999–2003.

These figures only provide a partial picture, since it is often very difficult to establish whether a fire has been started deliberately (fires of doubtful origin or, if strongly suspected, malicious ignition). Happily, recent advances in forensic science have brought about improvements in the detection rate. A more reliable picture of the real size of the problem can be obtained from the United Kingdom *Fire Statistics*. The latest figures indicate a continuing worrying number of deliberate fires—some 91 200 in 2004 and, in particular, the number of attacks on vehicles and schools. 'Arson in vehicles . . . accounts for 60% of all deliberately set fires at 55,000 per year'. In 2004, in the United Kingdom, there were 840 school fires, slightly down on 896 in the previous year. Fatalities dropped from 117 in 2003 to 88 in 2004. Overall, the number of deliberate fires decreased by 21 per cent in 2004,[8] an encouraging trend. The cost of fires in

Table 11.9.1 Arson, number of offenders found guilty of arson in England and Wales for the period 1999–2003

1999	2000	2001	2002	2003
2475	2470	2644	2427	2501

(Extracted from Home Office, (2004). Criminal statistics, England and Wales, 2003, Cm 6361. TSO © Crown Copyright[7].)

purely monetary terms is considerable; for example, figures from the Association of British Insurers (ABI) indicate 'that the cost of commercial fire claims in 2005 was 791 million pounds'—a record.[9]

Legal aspects

Legal definitions of arson vary from country to country. In England and Wales, prior to 1971, arson was an offence at common law. Currently, it is dealt with under the Criminal Damage Act, 1971. Similar provisions apply in Northern Ireland. In Scotland, it is dealt with under various common law offences. Section 1 of the 1971 Act states:

1 A person who without lawful excuse destroys or damages any property belonging to another intending to destroy or damage such property or being reckless as to whether any such property would be destroyed or damaged shall be guilty of an offence.

2 A person who without lawful excuse destroys or damages any property, whether belonging to himself or another:

(a) intending to destroy or damage any property or being reckless as to whether any such property would be destroyed or damaged; and

(b) intending by the destruction or damage to endanger the life of another or being reckless as to whether the life of another would be thereby endangered; shall be guilty of an offence.

3 An offence committed under this section by destroying or damaging property by fire *shall be charged as arson* (emphasis added).

NOTE: Recklessness has recently been clarified by the High Court as follows: 'A person also acts recklessly within the meaning of Section 1 of the 1971 Act with respect to (i) a circumstance where he is aware of a risk that exists or will exist; (ii) a result when he is aware of a risk that will occur; and, it is in the circumstances known to him, unreasonable to take the risk'.[10]

The seriousness with which arson and endangering life is recorded is reflected in Section 4 of the Act where both are punishable by maximum penalties of life imprisonment.

Classification, motivation, and management

Classification

An early, large-scale attempt to classify fire-raisers was undertaken by Lewis and Yarnell.[11] For an account of this and other earlier studies see Prins.[12] Faulk proposed two useful broad groupings. *Group I* consisted of those cases in which the fire served as a means to an end (for example, revenge, fraud, or a plea for help); *Group II* consisted of those cases where the fire itself was the phenomenon of interest.[13] Some years ago the present author, together with two psychiatrist colleagues, examined the files of a group of 113 imprisoned arsonists being considered for parole.[14] From this small (and admittedly highly selective) sample a rudimentary classification was devised. (See Table 11.9.2). This has been used by others as starting points for their own and perhaps more sophisticated classifications. The present writer has more recently modified slightly this earlier classification.[3] Despite modification it can be seen to still have certain weaknesses since it collates the behavioural characteristics of fire-raisers, various types of fire-setters and their motivations.[15]

Table 11.9.2 Suggested classification of the motives of arsonists (fire-raisers)

(a) Arson committed for financial reward (insurance fraud, etc)
(b) Arson committed to conceal another crime (for example, burglary or homicide)
(c) Arson committed for political purposes (terrorist and associated activities)
(d) Self-immolation as a political gesture. (Not arson as such, but included here for completeness, see Prins[12])
(e) Arson committed for mixed motives (for example, during the phase of minor depression, as a cry for help, or as a result of abuse of alcohol or other drugs)
(f) Arson due to the presence of formal mental disorder (for example, severe affective disorder, schizophrenic illnesses, organic mental disorder, mental impairment (learning disability))
(g) Arson due to motives of revenge—against (i) an individual or individuals; (ii) against society or others more generally
(h) Arson committed as an attention-seeking act (but excluding motives set out under (e) above) and arson committed as a means of deriving sexual satisfaction and/or excitement (for example, some forms of pyromania)
(i) Vandalistic arson (by young adults and children)

Rix broadened our original classification to include attempts to gain rehousing, carelessness, 'anti-depressant' (to relieve depressed feelings), and 'proxy' (in which the offender had acted on behalf of another who had borne a grudge).[16] Barker, an experienced forensic psychiatrist, in a wide-ranging and meticulous study of the psychiatric aspects of fire-raising, suggested that future classifications need to be more sharply focussed, emphasizing that arson should be seen 'merely as a symptom' to be viewed in the context of the whole person, not only to delineate different 'syndromes' of arsonists but also to identify individual points of therapeutic intervention and future dangerousness.[15]

Recent work by Canter and Fritzon has carried this focus forward. They suggest four themes to arson. Two related to *expressive* acts; (a) those that are realized within the arsonist's own feelings, being analogous to suicide, and (b) those that are acted on objects, like the burning of symbolic buildings. The two others relate to *instrumental* acts; (c) those that are for personal indulgence, similar to personal revenge; and (d) those that have an object focus such as hiding evidence from a crime (emphasis added).[17] More recently Canter and Fritzon's work has been replicated successfully by Almond *et al.*[18] and Hakkanen *et al.*[19]

(a) Some general characteristics of fire-raisers

It is a reasonable generalization to state that fire-raisers appear to be mostly young adult males who have exhibited behavioural difficulties from an early age (see Kennedy *et al.*[20] for a systematic review of the literature on this aspect and Repo and Virkkunen[21]). A significant proportion of these youthful fire-raisers have problems of alcohol abuse and intelligence levels lower than average.[21,22] Females who commit repeated acts of fire-raising and show some degree of mental disorder and self-mutilating behaviours are more likely to be awarded a mental health disposal by the courts than their male counterparts (see Coid *et al.*[23]and Noblett and Nelson[24]).

Motivation

To conform to the requested word limit for this chapter, categories (a), (b), (c), and (d) in Table 11.9.2 are not considered here. Detailed discussion of these and illustrative case vignettes may be found in Prins.[12]

(a) Fire-raising committed for mixed and unclear reasons

These are cases in which it is difficult to ascribe a single specific motive and which cause significant problems in assessment. They are likely to include the presence of a degree of mild (reactive) depression which may lead the fire-raiser to direct anger at a spouse or partner; thus revenge may also play a part (see discussion below). This group may also include cases in which the fire-raising may be a disguised plea for help, or a reaction to sudden separation or bereavement; in a proportion of these cases alcohol appears to play a part.

(b) Fire-raising due to serious mental disorder

Functional psychoses, notably the schizophrenias, may play a part in some acts of fire-raising. Such offenders will most likely be detained in secure hospitals or units. Manic depressive psychosis features occasionally, a classic case being that of Jonathan Martin, the nineteenth century arsonist who set fire to York Minster.[12]

(c) Fire-raising associated with 'organic' disorders

Occasionally, brain tumours, injury, epilepsy, dementia or metabolic disturbance may play a part. For example, although the epilepsies are not commonly associated with fire-raising (or other serious crimes for that matter) one should always be on the lookout for the case in which the crime has been committed when the person appeared not to be in a state of clear consciousness or when onlookers were present. Examples of organic states and their relationship to fire-raising are provided elsewhere.[3,25–27] The relationship between learning disabilities and fire-raising is discussed by Prins[12] and Clare *et al.*[28]

(d) Fire-raising motivated by revenge

Those incidents motivated by revenge are potentially the most dangerous. Such offenders are like the monster in Mary Shelley's *Frankenstein* who said 'I am malicious because I am miserable'. These are the fire-raisers who have serious problems with their feelings of anger and frustration caused by real or imagined wrongs. In considering the links between motives of revenge, it is important to stress the hazards of trying to place motivations for fire-raising in discrete categories; the vengeful fire-raiser may show clear signs of identifiable mental illness (for example, delusional jealousy), may be learning disabled and/or physically impaired, or may not be diagnosable as 'ill' in any formal psychiatric sense.

(e) Pyromania

The diagnostic criteria for pyromania are set out in DSM-V(Rev) TSM (1994) on page 615.[29] The condition and its diagnosis may be said to be one of exclusion. It is dealt with in detail by McElroy in Chapter (4.13.1) of this volume. Perhaps its manifestation of excitement for those who show the disorder is best exemplified by the poet Walt Whitman in his *Poems of Joy* (1860).

> I hear the alarm at dead of night,
>
> I hear the bells—shouts!
>
> I pass the crowd—I run!
>
> The sight of flames maddens me with pleasure.

(f) Sexually motivated fire-raising

The possible connection between fire-raising and sexuality has already been referred to. The lack of frequent association should not blind clinicians and others to its possible existence in certain cases, or its similarity to sex offending. Fras puts it well—as follows: 'In its comparative, stereotyped sequence of mounting pressure . . . it resembles the sexual perversions, as it may parallel them in its imperviousness to treatment'.[30] It is not without significance that imprisoned fire-raisers appear to have more than their fair share of psychosexual difficulties and partnership problems. In Hurly and Monahan's Grendon Prison study, a large proportion of their sample reported difficulties in social relationships with women.[31]

Suggestions for management

The word 'Suggestions' is used to indicate that what follows is not intended to be prescriptive; it must be emphasized that no single form of management is likely to be effective. At the assessment stage it is vital to treat every case as singular. Any attempt at assessing the future risk of fire-setting must view the behaviour on the basis of all the facts (for example, full details of the index offence and antecedent history). One can then begin to take the rounded and long view as advocated by Scott in his seminal paper on dangerousness. Assessors of whatever discipline will gain much from absorbing Scott's balanced and insightful views.[32] Pointers to successful assessment and management may be summarized as follows:

- Distinguish the fraudulent fire-raiser. But note that the fire-setter who appears to be engaged upon a fraudulent insurance claim to 'rescue' a failing business may be suffering from an underlying depressive illness. The history-taking in such cases needs to be painstaking and searching. Do not 'run' with what appears to be the obvious explanation.
- Distinguish the politically motivated. But, note that some politically motivated fire-raisers may also have serious mental health problems. In an age when fear of terrorist attacks abounds and the remedies appear to be of the 'knee-jerk' variety, a cool head in the assessment process is essential.
- Distinguish the vandalistic and the differences between young people who set fires out of boredom or for 'kicks' from child fire-setters who are more likely to have seriously dysfunctional social backgrounds.[33,34]
- Distinguish those who are driven to set fires by clear evidence of mental disorder, notably functional psychosis, severe anti-social personality disorder (psychopathy), organic disorder, and learning disability.
- Distinguish those who appear to exhibit pyromania as defined in DSM-IV TSM(Rev).
- Distinguish those rarer cases in which sexual disorder (and in particular sexual dysfunction) may have played a significant role.
- Distinguish the vengeful. It is important to remember that feelings of vengefulness may persist over long periods of time and such fire-raisers may be adept at concealing their vengeful feelings. These fire-raisers have some features in common with the delusionally jealous (Othello-type syndrome).

Successful assessment and management (which ideally should be a 'seamless' process) needs to rest upon a multi-faceted and multi-team approach. An excellent example of such a multi-disciplinary approach may be found in Clare *et al.*[28] They describe their management of a case that necessitated an understanding of both physical and learning disability combined with a capacity to work intensively using eclectic behavioural techniques over a prolonged period of time. Despite minor setbacks, the offender-patient, who had been subject at one time to containment in a high security hospital, remained free of his long-standing fire-raising behaviour at 4-year follow-up. It would also be unwise to believe that psychoanalytically based psychotherapy had no place in the management of psychotic and seriously personality disordered fire-raisers. Cox described some very productive work with such patients in Broadmoor.[35] Social skills training of one kind or another has a very important part to play. Many fire-raisers (particularly the vengeful and those with a pyromania diagnosis) are socially inept and believe themselves to be misunderstood by society. Techniques aimed at improving their self-regard, self-image, and social competence can help to minimize recidivism.

Conclusion

Not only is arson (fire-raising) a very worrying offence for the reasons given, but it has shown an increase in recent years. Moreover, the 'profile' of those convicted of arson has shifted over the years with an increased proportion of female offenders. And, in a very important recent study, Soothill *et al.* showed at 20 year follow-up that the proportion of those reconvicted for arson had more than doubled. The authors conclude with the sobering observation that 'the situation in relation to arson has deteriorated significantly over the past 40 years'.[36] The causes of fire-raising are complex and attempts at classification have not been entirely successful. Viewing fire-raising as a 'symptom' appears to offer the best hope for more successful diagnosis and management. The brief survey in this chapter has merely touched upon the topic. The suggestions for further reading and the references should assist those who wish to pursue the topic in further depth.

Acknowledgement

My thanks to Mrs Janet Kirkwood for so ably preparing the final version of the manuscript.

Further information

Arson Prevention Bureau. The Bureau publishes the *Arson Intelligence Newsletter* at regular intervals. Available from Arson Prevention Bureau, 51 Gresham Street, London, EC2V 7HQ.

Enayati, J., Grann, M., Lubbe, S., *et al.* (2008). Psychiatric Morbidity in Arsonists referred for Forensic Psychiatric Assessment in Sweden. *The Journal of Forensic Psychiatry and Psychology,* **19**, 139–147.

Office of the Deputy Prime Minister (ODPM) (2006) *Fire Statistics* (*UK*) 2004 and *Fire Statistics Monitor* Issue 1/06. 28.2 ODPM. (These publications provide extensive statistical and descriptive data).

Repo, E. and Virkkunen, M. (1997). Outcomes in a sample of Finnish fire-setters. *Journal of Forensic Psychiatry*, **8**, 127–37.

Swinton, M. and Ahmed, A. (2001). Arsonists in maximum security. *Medicine, Science, and the Law*, **41**, 51–7.

References

1. Soothill, K. (1991). Arson. In *Principles and practice of forensic psychiatry* (eds. R. Bluglass and P. Bowden), pp. 779–86. Churchill Livingstone, London.
2. Barnett, W. and Spitzer, M. (1994). Pathological fire-setting: 1851–1991: a review. *Medicine, Science, and the Law*, **34**, 4–20.
3. Prins, H. (1994). *Fire-raising: its motivation and management*. Routledge, London.
4. Prins, H. (2005). *Offenders, deviants or patients*? (3rd edn) (Chapter 7). Routledge, London.
5. MacDonald, J.M. (1977). *Bombers and fire-setters*. Thomas, Springfield, IL.
6. Prins, H. (2001). Arson and sexuality. In *The encyclopaedia of criminology and deviant behaviour*, Vol. 3 (eds. C.D. Bryant, N. Davis, and G. Geis), pp. 11–5. Brunner-Routledge, New York.
7. Home Office. (2004). *Criminal statistics, England and Wales, 2003*. Cm6361. TSO.
8. Arson Prevention Bureau. (2006). *Arson intelligence newsletter*.(87).
9. Arson Prevention Bureau. (2006). *Arson intelligence newsletter*. (87), and *Fire Statistics, UK, 2004*.ODPM (2006), London.
10. R v. G., and another. (2003). UK HL50. Reported in *The Independent Law Report*. 22.10.03.
11. Lewis, N.D.C. and Yarnell, H. (1951). Pathological fire-setting (Pyromania). *Nervous and mental disease monograph 82*. Coolidge Foundation, New York.
12. Prins (3). Notably pp. 98–103.
13. Faulk, M. (1988). *Basic forensic psychiatry*. Blackwell Science, Oxford.
14. Prins, H., Tennent, G., and Trick, K. (1985). Motives for arson (fire-setting). *Medicine, Science, and the Law*, **25**, 275–8.
15. Barker, F. (1994). *Arson: a review of the psychiatric literature*. Oxford University Press, Oxford.
16. Rix, K.J.B. (1994). A psychiatric study of adult arsonists. *Medicine, Science, and the Law*, **34**, 21–34.
17. Canter, D. and Fritzon, K. (1998). Differentiating arsonists: a model of fire-setting actions and characteristics. *Legal and Criminological Psychology*, **3**, 73–96.
18. Almond, L., Duggan, L., Shine, J., *et al.* (2005). Test of the arson system model on an incarcerated population. *Psychology, Crime and Law*, **11**, 1–15.
19. Hakkanen, H., Puolakka, P., and Santilla, P. (2004). Crime scene actions and offender characteristics in arsons. *Legal and Criminological Psychology*, **2**, 197–214.
20. Kennedy, P.J., Vale, L.E., Khan, S.J., *et al.* (2006). Factors predicting recidivism in child and adolescent fire-setters. *Journal of Forensic Psychiatry and Psychology*, **17**, 151–64.
21. Repo, E. and Virkkunen, M. (1997). Young arsonists: history of conduct disorder, psychiatric diagnoses and criminal recidivism. *Journal of Forensic Psychiatry*, **8**, 311–20.
22. Taylor, J.L., Thorne, I., Robertson, A., *et al.* (2002). Evaluating a group intervention for convicted arsonists with mild and borderline intellectual disabilities. *Criminal Behaviour and Mental Health*, **12**, 282–93.
23. Coid, J., Wilkins, J., and Coid, B. (1999). Fire-setting, pyromania and self-mutilation in female remanded prisoners. *Journal of Forensic Psychiatry*, **10**, 119–30.
24. Noblett, S. and Nelson, B. (2001). A psycho-social approach to arson—a case-controlled study of female offenders. *Medicine, Science, and the Law*, **41**, 325–30.
25. Carpenter, P.K. and King, A.L. (1989). Epilepsy and arson. *The British Journal of Psychiatry*, **154**, 554–6.
26. Byrne, A. and Walsh, J.B. (1989). The epileptic arsonist. *The British Journal of Psychiatry*, **155**, 268.
27. Hurly, W. and Monahan, T.M. (1969). Arson: the criminal and the crime. *British Journal of Criminology*, **9**, 4–21.

28. Clare, I.C.H., Murphy, D., Cox, D., *et al*. (1992). Assessment and treatment of fire-setting: a single case investigation using a cognitive behavioural model. *Criminal Behaviour and Mental Health*, **2**, 253–68.
29. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (DSM-IV-4th edn). American Psychiatric Association, Washington, DC.
30. Fras, I. (1983). Fire-setting and its relationship to sexuality. In *Sexual dynamics of anti-social behaviour* (eds. L.D. Schlesinger and E. Revitch), pp. 192–203. Thomas, Springfield, IL.
31. Hurly and Monahan (27).
32. Scott, P.D. (1977). Assessing dangerousness in criminals. *The British Journal of Psychiatry*, **131**, 127–42.
33. Santilla, P., Hakkanen, H., Alison, L., *et al*. (2003). Juvenile fire-setters: crime scene actions and offender characteristics. *Legal and Criminological Psychology*, **8**, 1–20.
34. Perrin-Wallqvist, R. and Nerlander, T. (2003). Fire-setting and playing with fire during childhood and adolescence: interview studies of 18 year-old male draftees and 18-19 year-old female pupils. *Legal and Criminological Psychology*, **8**, 151–8.
35. Cox, M. (1979). Dynamic psychotherapy with sex offenders. In *Sexual deviation* (2nd ed) (ed. I. Rosen), pp. 306–50. Oxford University Press, Oxford.
36. Soothill, K., Ackerley, E., and Francis, B. (2004). The criminal careers of arsonists. *Medicine, Science, and the Law*, **44**, 27–40.

11.10

Stalking

Paul E. Mullen

What is stalking?

Stalking is now used to describe a problem behaviour characterized by repeatedly inflicting unwanted intrusions and/or communications on another in a manner which creates fear and/or significant distress.[1] The intrusions can involve, following, loitering nearby, maintaining surveillance, and making approaches. The communication can be via telephone (including SMS), letter, electronic mail, graffiti and notes attached, for example, to the victims' car. Stalking can be associated with a range of harassments which though not part of the core behaviours are all too frequent. These include, ordering goods and services on the victim's behalf (late night pizza's being a favourite) damaging property, spreading malicious rumours, vexatious complaints, threats, 'cyber terrorism', and assault.

There are two basic patterns to stalking.[2] The first involves repeated incursions predominantly in the form of approaches and following perpetrated most often by a stranger and lasting only a day or so. The second is characterized by a range of both communications and intrusions, is usually perpetrated by an ex-intimate or acquaintance, and lasts for weeks, months, or even years. The first type can be intense and distressing at the time but uncommonly culminates in a physical attack and though upsetting, rarely inflicts long-term psychological or social damage. The second type is associated not infrequently with psychological and social damage to the victim and will involve physical assaults in up to a third of victims.

The epidemiology of stalking

Estimates of the prevalence of stalking, as with any other phenomenon, will vary according to definition, sampling, method of enquiry, and the willingness of subjects to respond and respond frankly.[2–7] Reported lifetime rates of victimization for women are between 8 per cent and 22 per cent and for men between 2 per cent and 8 per cent. Most victims are female (70–80 per cent), most stalkers are male (80–85 per cent), with 20–25 per cent involving same gender stalking, typically male on male.

Cyberstalking

Cyberstalking has attracted considerable interest but few systematic studies. Even the definitions employed of cyberstalking vary widely.[8–10] As befits an online phenomenon much of the information about it is to be found on the Internet rather than in the more traditional sources of academic knowledge.

Sheridan and colleagues[11] in an important study concluded that cyberstalking was usually one more invasive technique for pursuing stalking rather than a distinct type of activity.

Cyberstalking can include the use of the Internet and SMS facilities to:

1 *Send repeated unwanted messages.*

2 *Order goods and services on the victim's behalf.*

3 *Publicizing private information of a potentially damaging or embarrassing nature.* Including circulating e-mails, placing information on the web containing personal details, and occasionally explicit sexual images.

4 *Spreading false information.* A wide range of misinformation can be spread via the Internet with the authors of these calumnies able, should they wish and have the necessary skills, to hide their identity.

5 *Information gathering online about a victim* can cover a wide range of material from addresses, employment histories, to financial details. There are even services for tracing people available online which can be utilized by stalkers whose victim has eluded them.

6 *Identity theft* goes beyond simply pretending to be the victim for the purposes of ordering goods or initiating contacts to an attempt to assume not just the name but the actual property and attributes of the victim.

7 *Encouraging others to harass the victim.* This can cover activities such as placing communication purporting to be from the victim on web likely to attract unwanted communications or attentions. The most egregious example involved a rejected stalker who posted personal advertisements in his ex-partner's name and giving her address which suggested she enjoyed being raped and solicited such attentions. Apparently six men actually came to her house in response to these provocations.[12]

Impact on victims

Stalking is both an act of violence in itself which causes psychological distress and social disruption, and is a harbinger of assault. Being stalked can produce a corrosive state of fear, arousal, and

helplessness. As with domestic violence for most victims it is not the blows which are the most destructive but living in a chronic state of intimidation and the expectation of imminent intrusion. In the study of Pathé and Mullen[13] the majority reported disruptive levels of anxiety with intrusive recollections of the stalking, sleep disturbance, lowered mood, with 25 per cent admitting considering suicide to escape the situation. A community study found increased rates of psychiatric morbidity and post-traumatic symptomatology amongst those stalked for more than 2 weeks but not amongst those who had experienced the briefer periods of harassment.[14] Dressing and colleagues[7,15] also document significant psychological and social disruption in response to being stalked with 56 per cent reporting agitation, 44 per cent increased anxiety, 41 per cent sleep problems, and 28 per cent increased depression.

Stalkers: classifications and typologies

Stalking, like most forms of complex human behaviour, can be the outcome of a wide range of psychological, social, and cultural influences. Some stalk in hope, some in anger, some in lust, some in ignorance, and many in mixtures of the above. In an attempt to advance the understanding of stalkers a range of typologies and classifications have been advanced.[16]

Classifying stalkers by the nature of their prior relationship with the victim has the advantages of simplicity and utility. The classification advanced by Mohandie and colleagues[17] represent the best supported by empirical evidence of such approaches to date. They divide stalkers into those with and without a prior relationship. Those with a prior relationship are subdivided into ex-intimates (ex-partners both long term and more casual) and acquaintances (including friends, colleagues, and professional contacts). Those without a prior relationship are subdivided into firstly 'public strangers' who were encountered through the media or in their public roles, and secondly into 'private strangers' encountered by chance in the interactions of everyday life. This classification's greatest utility is in predicting the risks of assault, with those with a prior intimate relationship constituting the highest risk group and those targeting public strangers the lowest. In their view the pursuers of public strangers are the most likely to be psychotic with those pursuing ex-intimates being relatively impervious to therapy but responsive to criminal sanctions.

The typology first developed by Mullen and colleagues[1,18] depends primarily of the context in which the stalking emerges and the motivations which initiated and sustained the behaviour. Its appeal has been primarily to clinicians managing stalkers and their victims.[19] There are five main types:

1 *The rejected* whose stalking begins in the context of the breakdown of a close relationship. The stalking is initially motivated either by the desire for reconciliation or to express the rage at rejection, with a mixture of both being quite common. The stalking is often sustained by the pursuit of the ex-partner becoming a substitute for the lost relationship with the satisfactions from intrusion and control replacing those of intimacy.

2 *The intimacy-seeker* who is pursuing love. The stalking begins in the context of a life bereft of intimacy and is motivated by the hope, or firm expectation, of obtaining a loving relationship with a stranger or casual acquaintance on whom they have fixed their amorous attentions. The pursuit is sustained in the face of indifference or outright rejection because better a love based on fantasy or delusion than no love at all.

3 *The incompetent suitor* who is pursuing a sexual encounter or friendship. This usually begins in the context of loneliness and is motivated by a desire to start some form of relationship with someone who has attracted their interest. This group often pursues intensely with multiple intrusions but rarely persists for more than a day or so, presumably because multiple rebuffs bring few rewards.

4 *The resentful*, whose stalking starts in the context of a grievance at being unjustly treated or humiliated. The initial motivation is revenge but this gives way to the satisfactions obtained from the sense of power over someone who has previously been experienced as an oppressor, or the representative of oppressors.

5 *The predatory*, which begins in the context of the desire to act out violent or sexual fantasies often of a sadistic or paedophilic nature. The initial motivation is to gain information about the movements of a potential victim (usually a stranger but occasionally an acquaintance). The stalking continues because of the satisfactions accruing not just from voyeurism but from the excitement and sense of power which comes from rehearsing the planned attack in fantasy whilst watching the future victim.

Each of the stalker types, hopefully with the exception of the predatory, has correlates in normal behaviour. When relationships break down one partner is often confused or distressed by the separation and seeks to understand, to reconcile, or to express anger. The incompetent suitor is kin to the awkward adolescent male and the socially inept adult who fails to traverse effectively the social minefields of courting or simply making acquaintance. The intimacy seeker is the adolescent crush and the enthusiastic fan writ large. Even the resentful is not far removed from some seekers after justice and those asserting their rights. In theory the boundary between persistent approaches as part of socially acceptable behaviour and the crime of stalking are difficult to pin down. In practice the distinction is rarely a problem. Stalkers are those who repeatedly force themselves on another person in a manner which creates obvious distress. It is the total disregard, or blindness to, the disturbance and often fear that their behaviour creates which distinguishes the stalker from their more normal counterparts. Sometimes the stalkers are so caught up in their own world they are oblivious to their effect on others. Sometimes they are blinded by delusion. Sometimes self-righteousness makes them indifferent. But sometimes they delight in the effect they produce in their victim.

Psychopathology of stalkers

Stalkers are rarely, if ever, drawn from the psychologically adequate or socially able of the world. The estimates of the proportion of stalkers whose behaviour is directly related to mental disorders varies according to where the researchers derived their sample. For example, Zona and his group[20] whose sample contained many who pursued Hollywood celebrities had a significant number with erotomanias and morbid infatuations.

In broad terms psychotic disorders are relatively frequent in the intimacy seeking group. In the resentful type it is the paranoid disorders which unsurprisingly predominate, though most are not

associated with frank delusion. The rejected often have problems around dependency, rigidity, control, and self-esteem with substance abuse and depressive states on occasion complicating the picture, but psychotic states are uncommon. The incompetent suitors are socially disabled sometimes by shyness, sometimes by narcissism, sometimes by intellectual limitations, sometimes by culture, sometimes by disorders such as Asperger's syndrome, rarely by psychosis, but always by interpersonal insensitivity or indifference. The predatory are sexually perverse and not infrequently have marked psychopathic traits, but again are rarely psychotic.

Attempts have been made to conceptualize stalking as a manifestation of obsessive–compulsive disorder. Stalkers are certainly often obsessive in the everyday sense of that word in their pursuit of the victim. They rarely however regard their behaviour as unjustified let alone irrational, and few see their persistence as senseless. They may resist the urges to stalk on occasion but for the most part devote themselves wholeheartedly to the pursuit. Anxiety is more likely to be generated by the fear of failure, or of consequences, than by not acting on their impulses to stalk. They may well spend many hours thinking about the object of their unwanted attentions, and in the resentful reliving the experiences of actual or supposed injustice, so in that sense they are ruminators. Personality traits of rigidity, rumination, and the overvaluing of order are not infrequently so marked in rejected and resentful stalkers as to justify a label of an obsessional personality. In short, the behaviour often has an obsessive quality but the state of mind rarely conforms to that found in obsessive–compulsive disorders.

Attachment theory has unsurprisingly been evoked to explain stalking. That stalkers as a group don't do interpersonal relationships very well is obvious. Evidence exists that insecure attachment styles predominate amongst rejected stalkers, the intimacy seekers may have the type of secure attachment style only sustainable by delusion, and the incompetent and resentful favour the dismissive style. This is useful in assessment and management but what connection it may have with any theory of early development is speculative and here as elsewhere more likely to be productive of mythologizing than good clinical practice.

The stalking of health professionals

Health care professionals have a heightened vulnerability to being stalked by their patients and clients.[21–24] The risk stems largely from resentful and disappointed patients but in part from lonely and disordered people who misconstrue sympathy and attention for romantic interest. While some stalking behaviours constitute little more than minor irritations, they may also ruin a clinician's career.

Sandberg *et al.*[25] studied an inpatient psychiatric service reporting 53 per cent of clinical staff had been stalked by patients. Galeazzi and colleagues[26] found 11 per cent of the mental health professionals in an Italian service had been stalked for lengthy periods by patients. Purcell and colleagues[14] surveyed a randomly selected sample of 1750 psychologists (73 per cent female). The lifetime prevalence of stalking by clients was 19.5 per cent with 8 per cent being stalked in the preceding 12 months. Most victims were working in direct client care (95 per cent) and experienced rather than new entrants to the profession. Stalkers fell predominantly into either the intimacy seeking (19 per cent) or resentful (42 per cent) types. Over 30 per cent of psychologists in this study were subjected to vexatious complaints by their stalker. The impact of complaints to professional boards, health ombudsmen, and other agencies of accountability can be devastating.

Too often in the past therapists who fell victim to stalking by patients had to bear the additional burden of implied or overt criticism from colleagues to the effect that, had they more adroitly managed the therapeutic encounter and the resultant transference, they would not now find themselves in this predicament. There should be no sympathy with blaming the victim, even if it comes in the guise of technical advice or supervision. Being stalked is a risk inherent in the therapeutic process. Our colleagues should be accorded support and help, if for no other reason than we do not know when it may be our turn to face the pursuit of the vengeful or lustful patient.

Risk assessment and risk management

Stalking came to prominence because it was regarded as a risk factor for violence. Subsequently it became clear that the damage inflicted on those who are stalked could also encompass significant social and psychological damage.[23]

Assessing and managing the stalker requires a primary focus on the risks they present to the victim. Nevertheless the risk that stalkers incur from their own behaviour also needs to be considered (Table 11.10.1). The conflict between the stalker's desires and the victim's interests are obvious, but they are at one in being at risk of damage from the stalking situation. There can be a tragic symmetry between the victim forced to live an increasingly restricted life in a state of constant fear, and the stalkers devoting all their time and resources to a damaging and ultimately self-defeating pursuit. The victim's and the perpetrator's lives can be laid waste. This is not to argue for equivalence between victim and perpetrator but merely to note they share the chance of disaster. A perspective which encompasses the risks to stalkers and victims has the advantage for health professionals of reducing the ethical dilemma when treating stalkers around whose interests one is serving, the patient's or their victim's. We help both to the extent that we contribute to stopping the stalking, or reducing its damaging consequences.

The empirical basis for evaluating risk in the stalking situation

Risk assessment in stalking situations is currently hampered by a paucity of either retrospective or prospective studies of representative samples. Clinicians do not, however, have the luxury of deferring action until such evidence emerges. They must, for the present, depend on integrating knowledge from stalking research, borrowing from the systematic studies of risk in other areas, and drawing on clinical experience.[27]

The risk of continued or recurrent stalking

The duration of stalking is longest for rejected stalkers pursuing ex-intimates and intimacy seekers, with the incompetents usually pursuing only briefly. Women, here as elsewhere, are more persistent than men.[6,13] Once stalking has continued for more than 2 weeks the chances are high that it will continue for months.

The risks of psychological and/or social damage

Female victims of stalking report a greater psychological impact than male victims.[5,13] Clinically, the distress and disruption to

Table 11.10.1 The stalker's clinical risk factors and future hazards specific to stalking situation

Risk factor	Management possibilities include
Clinical	
1. Mental state e.g. depression, delusional preoccupations	Active treatment usually involving pharmacotherapy
2. Substance abuse	Referral to a specialist substance abuse service where possible or to self-help groups like AA
3. Anger	Anger management remains a problematic area. Ideally those with anger control problems should be receiving special help independently of the broader management of their stalking
4. Attitudes towards, and beliefs about, the victim which sustain stalking	Appropriate legal interventions; CBT and focussed psychotherapies aimed at such areas as; abandoning love, accepting loss, confronting misperceptions
5. The conviction that they are right to engage in stalking	Enhancing victim empathy. Confronting false attributions using CBT
6. The refusal to engage in any therapy, or conform to legally imposed restrictions on access to the victim	Ultimately confronting the stalker with consequences (e.g. through breaching parole, referring back to court, etc.); employing motivational interviewing strategies to assist the stalker to appreciate the need for intervention
7. Social incompetence	Social skills training, therapies aimed at enhancing self-efficacy
8. Paraphilia	Sex offender program incorporating CBT +/− pharmacotherapy as indicated
Future hazards	
1. Likely future contact with the victim	Every effort should be made to enforce a total ban on direct contact or direct communications
2. Lack of a feasible set of plans for avoiding a recrudescence of stalking	Ensure structured plan around avoiding provocations and using protections re: stalking; CBT to assist the stalker to overcome the compulsion to stalk
3. That the underlying precipitants remain unresolved	Focussed psychotherapy aimed at the areas identified in the formulation; social skills training for the inept; assistance abandoning the relationship; the treatment of paraphilias using CBT +/− pharmacotherapy as indicated
4. Continuing instability of residence and/or employment	Assistance obtaining housing; career counselling; and active employment rehabilitation as indicated and appropriate
5. Continuing social isolation	Use of clubs, day centres, recreational counselling, domestic pets
6. Likely low level of compliance with legal restraints on contact with victim	Ensure knowledge of consequences of breaches and never collude—implicitly or explicitly—with avoiding those consequences
7. Likely low level of cooperation with any treatment programme	Use of compulsory community treatment orders either imposed by court or as part of mental health legislation

victims is usually most obvious in ex-intimates pursued by their rejected partners, perhaps because of the higher levels of violence and intimidation combined with the complexity, as well as the intensity, of feelings stirred up in this situation.[13] Psychological distress was higher amongst victims who were subjected to prolonged and repeated following and the experience of property theft or destruction.[2,13,28,29] The relationship between psychological impact and the experience of physical violence is less clear, despite its intuitive appeal.[29,30]

The risks of violence

(a) Prior relationship

Victims who have shared a prior intimate relationship with their stalker are at a high risk of physical violence.[1,4,13,20,31–35] Purcell and colleagues[6] for example reported in a random community sample that ex-intimates were the most likely to be attacked (56 per cent), followed by estranged relatives or previous friends (36 per cent), then casual acquaintances (16 per cent), work-related contacts (9 per cent), and finally strangers (8 per cent). Such findings should not be interpreted, however, as suggesting that victims of stalkers who are not ex-intimates are in little danger of physical violence. A chance of between 8 per cent and 36 per cent of being assaulted is no small risk.

(b) Threats

Between 30 per cent and 60 per cent of stalking victims are threatened.[1,20,31] In a community-based study 44 per cent of those threatened were subsequently assaulted and 73 per cent of victims assaulted by their stalker had previously been explicitly threatened.[6] In short threats predict violence and should be taken seriously.

(c) Mental disorder

Research has generally concluded that psychotic stalkers are less likely to be physically violent than their non-psychotic counterparts but the relationship to personality disorder remains unclear.[1,34,36,37]

(d) Substance abuse

Substance abuse is associated with violence in the stalking situation.[1,38–40]

(e) Prior offending and antisocial behaviour

The empirical data on the association between past criminal or violent behaviour and stalking violence is inconsistent, however the balance of the evidence favours such a relationship.[1,34,39,41]

(f) Demographic variables

The gender of stalkers has repeatedly been shown to have no impact on the prevalence of either threats or assault.[4–6,13,42]

(g) The nature of the stalking

Violence is predicted by escalating intrusiveness and intensity of the stalking behaviour. The strongest association is to physical intrusions into the victims house or place of work.[36]

The assessment process

Initial assessments of stalkers often occur in the context of pre-sentence or parole board evaluations. Victims may be encountered in a wider range of contexts, many seeking help from general rather than forensic mental health professionals. Stalkers usually lack insight into their behaviour and tend to deny, minimize, and rationalize their actions. Victims often minimize the experience of stalking and over-emphasize their own responsibility for the harassment, which should be of no surprise to anybody experienced in working with victims in other contexts. Conversely, the problem of false claims of stalking victimization cannot be entirely ignored.[43] This makes it essential to assess collateral information from such sources as witness statements, victim impact reports, judges' sentencing remarks, and professional to professional contacts, confidentiality allowing. Attempts to contact the victim when assessing the stalker, or the stalker when assessing the victim are, in our opinion, best avoided. However skilfully managed, such contacts tend to be experienced by the victim as the professional acting as an agent of the stalker, and by the stalker as support for their beliefs that this is a misunderstanding within a mutual relationship rather than a unilateral imposition of unwanted attentions.[27]

Management

The management of stalkers remains very much the province of forensic mental health professionals and even amongst them it is a specialist area. Basic approaches to identifying potentially remediable risks and their management however is presented in. Victims of stalking are however likely to be seen by a wide range of mental health professionals.

Reducing the impact on victims

Stalking victims will often present with significant problems with anxiety symptoms and depressed mood. The symptom complex of PTSD will be present to a greater or lesser extent in most victims of prolonged and intense stalking. Like many victim groups there may be a reluctance to disclose the details or even the existence of the traumatic experiences. As noted earlier self-blame is not infrequently part of the picture.

Given the high profile of stalking there are a number of disturbed people who claim they are being stalked as a way to express their distress, claim attention, or give form to their persecutory delusions.[24,43] This group, particularly if delusional, are often obvious given the flamboyant, implausible, and exuberant accounts of victimization. Care should be taken in dismissing claims of being stalked, however, as there are some very unpleasant stalkers out there some of whom stalk leaving few, if any, objective signs. False victims require help and treatment not rejection, but they require quite different treatment from actual victims.

Stalking victims need first and foremost good psychiatric care. Manage the stress symptoms, treat the depression, ameliorate the distress, and provide adequate support. Individual treatment is best initially but the use of groups for long-term support and treatment is worth considering.[44] What follows is a brief account of stalking specific interventions.

1 Informing others. When you are stalked it is essential to inform those you live with, work with, and are friends with. This performs three functions:
   (a) It allows you a 'reality check' on your fears that stalking is occurring.
   (b) It enables others to support you and equally important avoid inadvertently assisting the stalker.
   (c) It prevents those around you being put at risk by ignorance provoking the stalker thereby also falling victim.

3 Avoid contact and/or confrontation. All contacts or direct communications with a stalker risks reinforcing their behaviour. Confrontation and worse still violence, legitimizes, and encourages their violence. Once stalking is established it is usually too late to simply sort things out by having a meeting.

4 Documentation. The best protection for stalking victims lies in the criminal law. The police are more likely to respond appropriately if you can demonstrate that the behaviour is occurring and make it relatively easy for them to pursue the allegations. Keep the letters (dated), record the phone calls and retain the tapes, keep a diary noting approaches and where possible witnesses to those events, record verbatim any threats and obtain witnesses or photographs of any property damage. Not only will this be invaluable for a later prosecution it is part of the victim taking control back over their lives.

5 Restraining, non-molestation, apprehended violence and other such orders. Sometimes these civil orders do provide an effective means of reducing or stopping stalking. Police often advise their use, hopefully not just to avoid work. Some jurisdictions insist on obtaining such orders prior to considering criminal prosecution. Reservations have to be expressed, however, not only about their effectiveness (totally ineffective for intimacy seekers, relatively ineffective for the rejected and intermittently effectively for the resentful) but also about the level of insecurity and distress consequent on their breach and facing the frequent indifference of the police to such breaches.

6 Increased security. Simple cheap security measures such as good locks, movement triggered lighting outside the house, and securing the mail box may provide a degree of reassurance and a modicum of security.

Conclusions

Not a few psychiatrists, at least of my generation, have difficulties with the notion of problem behaviour like stalking being a proper subject for mental health concern. Psychiatry has traditionally been wary of concerning itself directly with criminal and antisocial behaviours.[45,46] The approach taken in this chapter was, in contrast, to define a pattern of behaviour destructive to the interests of perpetrator and victim and then to examine its origins, effects, and potential therapeutic management. That this is an enterprise with risks for the ethical integrity of psychiatry is undoubted. But recognizing that psychiatry can have a role in assessing and managing problem behaviours, without first performing obfuscating transformations into supposed mental disorders such as paraphilias

and impulse control disorders, allows a more clear sighted and effective approach to areas of human activity where our intervention can benefit both the actor and the wider community.

Further information

Mullen, P.E., Pathé, M., and Purcell, R. (2000). *Stalkers and their victims*. Cambridge University Press, Cambridge (new edition due 2008).

Pinals, D.A. (ed.) (2007). *Stalking psychiatric perspectives and practical approaches*. Oxford University Press, New York.

References

1. Mullen, P.E., Pathé, M., Purcell, R., *et al.* (1999). A study of stalkers. *The American Journal of Psychiatry*, **156**, 1244–9.
2. Purcell, R., Pathé, M., and Mullen, P.E. (2004). When do repeated intrusions become stalking? *The Journal of Forensic Psychiatry and Psychology*, **15**, 571–83.
3. Australian Bureau of Statistics. (1996). *Women's safety, Australia, 1996*. Commonwealth of Australia, Canberra.
4. Tjaden, P. and Thoennes, N. (1998). *Stalking in America: findings from the national violence against women survey*. National Institute of Justice and Centres for Disease Control and Prevention, Washington, DC.
5. Budd, T. and Mattinson, J. (2000). *Stalking: findings from the 1998 British crime survey*. Research Findings No. 129, Home Office Research Development and Statistics Directorate.
6. Purcell, R., Pathé, M., and Mullen, P.E. (2002). The prevalence and nature of stalking in the Australian community. *The Australian and New Zealand Journal of Psychiatry*, **36**, 114–20.
7. Dressing, H., Kuehner, C., and Gass, P. (2005). Lifetime prevalence and impact of stalking in a European population: epidemiological data from a middle-sized German city. *The British Journal of Psychiatry*, **187**, 168–72.
8. Bocij, P. and McFarlane, L. (2002). Online harassment: towards a definition of cyberstalking. *Prison Service Journal*, **139**, 31–8.
9. Bocij, P. and McFarlane, L. (2003). Cyberstalking: the technology of hate. *Police Journal*, **76**, 204–21.
10. Barak, A. (2005). Sexual harassment on the Internet. *Social Science Computer Review*, **23**, 77–92.
11. Sheridan, L.P., *et al.* (in press).Is cyberstalking different. *Psychology Crime and Law*.
12. Miller, G. and Maharaj, D. (1999). Chilling cyber-stalking case illustrates new breed of crime. *Los Angeles Times*.
13. Pathé, M. and Mullen, P.E. (1997). The impact of stalkers on their victims. *The British Journal of Psychiatry*, **170**, 12–17.
14. Purcell, R., Powell, M.B., and Mullen, P.E. (2005). Clients who stalk psychologists: prevalence, methods and motives. *Professional Psychology: Research and Practice*, **36**, 537–43.
15. Dressing, H., Gass, P., and Kuehner, C. (2007). What can we learn from the first community-based epidemiological study on stalking in Germany? *International Journal of Law and Psychiatry*, **30**, 10–7.
16. Spitzberg, B.H. and Cupach, W.R. (2003). What man pursuit? Obsessive relational intrusion and stalking related phenomena. *Aggression and Violent Behaviour*, **8**, 345–75.
17. Mohandie, K., Meloy, J.R., McGowan, M., *et al* (2006). The RECON typology of stalking: reliability and validity based upon a large sample of North American stalkers. *Journal of Forensic Sciences*, **51**, 147–155.
18. Mullen, P.E., Pathé, M., and Purcell, R. (2000). *Stalkers and their victims*. Cambridge University Press, Cambridge.
19. Pinals, D.A. (2007). *Stalking: psychiatric perspectives and practical applications*. APA Press, Washington, DC.
20. Zona, M.A., Sharma, K., and Lane, J. (1993). A comparative study of erotomanic and obsessional subjects in a forensic sample. *Journal of Forensic Sciences*, **38**, 894–903.
21. Sandberg, D.A., McNiel, D.E., and Binder, R.L. (1998). Characteristics of psychiatric inpatients who stalk, threaten, or harass hospital staff after discharge. *The American Journal of Psychiatry*, **155**, 1102–5.
22. Purcell, R., Pathé, M., and Mullen, P.E. (2001). A study of women who stalk. *The American Journal of Psychiatry*, **158**, 2056–60.
23. Mullen, P.E. and Purcell, R. (2007). Stalking of therapists. In *Severe personality disorders: major issues in everyday practice* (eds. B. van Luyn, S. Akhtar, and J. Livesley). Cambridge University Press, Cambridge.
24. Mullen, P.E., Pathé, M., and Purcell, R. (in preparation). *Stalkers and their victims* (2nd edn). Cambridge University Press, Cambridge.
25. Sandberg, D.A., McNiel, D.E., and Binder, R.L. (2002). Stalking, threatening and harassing behavior by psychiatric patients toward clinicians. *The Journal of the American Academy of Psychiatry and the Law*, **30**, 221–9.
26. Galeazzi, G.M., Elkins, K., and Curci, P. (2005). The stalking of mental health professionals by patients. *Psychiatric Services*, **56**, 137–8.
27. Mullen, P.E., Mackenzie, R., Olgoff, J.R.P., *et al.* (2006). Assessing and managing the risks in the stalking situation. *The Journal of the American Academy of Psychiatry and the Law*, **34**, 439–50.
28. Blaauw, E., Winkel, F.W., Arensman, E., *et al.* (2002). The toll of stalking: the relationship between features of stalking and psychopathology of victims. *Journal of Interpersonal Violence*, **17**, 50–63.
29. Kamphuis, J.H., Emmelkamp, P.M.G., and Bartak, A. (2003). Individual differences in post-traumatic stress following post-intimate stalking: stalking severity and psychosocial variables. *The British Journal of Clinical Psychology*, **42**, 145–56.
30. McEwan, T., Mullen, P.E., and Purcell, R. (2007). Identifying risk factors in stalking: a review of current research. *International Journal of Law and Psychiatry*, **30**, 1–9.
31. Harmon, R.B., Rosner, R., and Owens, H. (1998). Sex and violence in forensic population of obsessional harassers. *Psychology, Public Policy and Law*, **4**, 236–49.
32. Meloy, J.R. (1998). The psychology of stalking. In *The psychology of stalking: clinical and forensic perspectives* (ed. J.R. Meloy), pp. 2–23. Academic Press, San Diego.
33. Meloy, J.R. (1999). Stalking: an old behavior, a new crime. *Forensic Psychiatry*, **22**, 85–99.
34. Meloy, J.R., Davis, B., and Lovette, J. (2001). Risk factors for violence among stalkers. *Journal of Threat Assessment*, **1**, 3–16.
35. Sheridan, L., Davies, G.M., and Boon, J.C. (2001). Stalking: perceptions and prevalence. *Journal of Interpersonal Violence*, **16**, 151–67.
36. Farnham, F.R., James, D.V., and Cantrell, P. (2000). Association between violence, psychosis, and relationship to victim in stalkers. *Lancet*, **355**, 199.
37. Kienlen, K.K., Birmingham, D.L., Solberg, K.B., *et al.* (1997). A comparative study of psychotic and nonpsychotic stalking. *The Journal of the American Academy of Psychiatry and the Law*, **25**, 317–34.
38. Rosenfeld, B. and Harmon, R. (2002). Factors associated with violence in stalking and obsessional harassment cases. *Criminal Justice and Behavior*, **29**, 671–91.
39. Brewster, M.P. (2000). Stalking by former intimates: verbal threats and other predictors of physical violence. *Violence and Victims*, **15**, 41–54.
40. Roberts, K.A. (2005). Women's experience of violence during stalking by former romantic partners: factors predictive of stalking violence. *Violence Against Women*, **11**, 89–114.
41. Palarea, R.E., Zona, M.A., Lane, J.C., *et al.* (1999). The dangerous nature of intimate relationship stalking: threats, violence, and

associated risk factors. *Behavioral Sciences & the Law*, **17**, 269–83.
42. Meloy, J.R. and Boyd, C. (2003). Female stalkers and their victims. *The Journal of the American Academy of Psychiatry and the Law*, **31**, 211–19.
43. Pathé, M., Mullen, P.E., and Purcell, R. (1999). Stalking: false claims of victimisation. *The British Journal of Psychiatry*, **174**, 170–3.
44. Pathé, M. (2002). *Surviving stalking*. Cambridge University Press, Cambridge.
45. Clare, A. (1997). The disease concept in psychiatry. In *Essentials of postgraduate psychiatry* (eds. R. Murray, P. Hill, and P. McGuffin), pp. 41–52. Cambridge University Press, Cambridge.
46. Lewis, A. (1955). Health as a social concept. *The British Journal of Sociology*, **4**, 109–24.

11.11

Querulous behaviour: vexatious litigation, abnormally persistent complaining and petitioning

Paul E. Mullen

> Querulantenwahn (Ger.) A form of so called paranoia in which there exists in a patient an insuppressible and fanatic craving for going to law in order to get redress for some wrong which he believes done to him. Individuals who fall victim to this disorder are always strongly predisposed . . .. extremely egotistical . . . know everything better . . . differs from other forms of paranoia in so far as the wrong may not be quite imaginary . . . the more he fails the more he becomes convinced that enormous wrong is being done to him . . . neglects his family and his business . . . going down the road to ruin.[1]

The above quote neatly summarizes classical psychiatry's view of querulous, or litigious, insanity as a form of paranoia. A problematic form, however, in that the querulousness was usually based on a genuine grievance and was often regarded as developing on the basis of predispositions rooted in the sufferer's personality.[2–4] As to treatment Krafft Ebing[3] notes the 'necessary and beneficent (effects of the) appointment of a guardian and commitment to an asylum' but regretted that this 'takes place unfortunately only after they have used up their property, insulted the courts, and disturbed public order' (p. 395).

Psychiatries interest in the querulous (from the Latin to mutter and to mumble) waned rapidly in the latter half of the twentieth century. The diagnosis was appealed to less and less and the literature largely fell silent.[5] In part the disappearance of querulousness, and even the querulous patient, from the realms of psychiatry paralleled the decline of paranoia as a diagnostic entity. In part it reflected psychiatries increasing reluctance to play the role of social regulator. Probably most importantly the emerging culture of individual rights made pathologizing complainants potentially disastrous as it could deprive them of access to the major social mechanisms for obtaining justice.[6] Psychiatry lost interest in the querulant, however, at the very time that the 'culture of complaint' drew more and more vulnerable people into the systems of complaint management. Agencies of accountability, which range from Ombudsmen's offices, via registration boards, to complaints departments, are now almost all faced with the problems created by a small group of people pursuing grievances with a persistence and insistence out of all proportion to the substantive nature of their claim. It is estimated that 20–30 per cent of the resources of these agencies are being consumed by less than 1 per cent of unusually persistent complainants. In the civil and family courts the number of interminable cases being pursued, often by unrepresented litigants, escalates year by year. Last, but not the least, querulants pursuing what they regard as their rights through repeated petitions and intrusive approaches to politicians and heads of state distract protection services from more substantial threats.[7–9]

Clinical features

The querulant pursue their vision of justice through litigation in the court, through petitions to the powerful, and finally through the various agencies of accountability. In practice all three avenues are often explored. In the nineteenth and early twentieth century it was the civil courts in which these dramas were usually played out. Today the main burden falls on the complaints organizations.

It is not easy to distinguish the querulant from the difficult complainant or even from social reformers and victims of gross injustice. A simple typology may assist:

1 *Normal complainants* are aggrieved seeking compensation, reparation, or just an apology. They will accept conciliation and reasonable solutions, though they may become persistent and insistent if provoked by inefficiency or injustice.

2 *Difficult complainants* also seek compensation and reparation but often want in addition retribution. They tend from the outset to anger, to seeing themselves as the victim of others intentional malevolence, and to resist all solutions but their own. Eventually, however, they will settle for the best deal they can obtain.

3 *Altruistic reformers* who pursue goals of social progress via the courts, petitions, and complaints. They sacrifice their personal interests in pursuit of better outcomes for others. Though they may have a political agenda which is sectarian (e.g. antigenetically

modified foods, fathers rights) they do not have idiosyncratic and personalized objectives.

4 *Fraudsters* who knowingly pursue false or grossly exaggerated claims.

5 *The mentally ill* whose claims are driven by delusional preoccupations frequently bizarre in nature which reflect underlying disorders often of a schizophrenic type.

6 *The querulous* who seek personal vindication in addition to compensation, reparation, and retribution. They are on a quest for justice which becomes totally preoccupying. Unlike reformers, and most of the difficult, there is an obvious discrepancy between the provoking event and the importance attached to it by the querulous. They appear to seek not resolution but continuation of the conflict. They lay waste to their social and economic functioning.

The querulous are usually males who first become embroiled in complaining and claiming in their fourth or fifth decade. Premorbidly they were often able to function reasonably well. They rarely have criminal records or prior psychiatric contact, and substance abuse is not prominent. Many had relationships but by the time they reach psychiatrists they have usually alienated their family and friends. Querulants are often disappointed people who feel their qualities have been ignored and left unrewarded. Their pursuit of justice offers an opportunity to vindicate their lives and obtain the public recognition so long denied. Their personalities tend to have the traits of self-absorption, suspiciousness, and obsessionality combined with an enviable capacity for persistence.

Clinically they typically present as energized, garrulous individuals eager to convince you of the merits of their case. There is an enthusiasm which can seem almost manic but unlike the manic they are totally focussed and almost impossible to distract from their narrative of injustice. They may come with bags overflowing with documents testifying to their misplaced scholarship. If challenged they usually become patronizing as they pedantically refute all objections, to their complete satisfaction. Alternatively they may become menacing and overtly threatening.

Communications from querulants were noted by Lester and colleagues[10] to be often characterized by:

- Multiple methods of emphasis including, underlining, highlighting (often in multiple colours), and capitalization
- The generous use of inverted comas, exclamation marks, and question marks
- Numerous foot and marginal notes
- The use of attachments, often extensively annotated, some potentially pertinent (e.g. letters received, copies of legislation), others of less obvious relevance (e.g. Magna Carta, UN Declaration of Human Rights)
- And many many pages.

The content of communications may also be unusual sometimes containing:

- Legal, medical, and other terms used frequently but often incorrectly
- Repeated rhetorical questions
- A curious combination of rambling repetitiveness with pedantry (more difficult to describe than recognize)
- Veiled threats to harm themselves or others if their wishes are not granted or
- Exaggerated politeness and attempts to ingratiate.

Clinical assessment

The querulous can only be adequately assessed by considering the development over time of their behaviour as well as their state of mind. In an interview they may present as merely overenthusiastic and over hopeful pursuing their legitimate rights with at worst a degree of fanaticism. It is the unfolding of their story which reveals the damage they are suffering, and they have inflicted on those around them.

Case history (representing an anonamized conflation of several cases)

A man in his late forties made a complaint to the local bank manager over the manner in which mortgage documents had been prepared. There were grounds for legitimate concern as irregularities had occurred, though of a minor nature and of a kind which might have been expected to be to his advantage. His initial complaint to the bank was rejected. He appealed unsuccessfully to the banking ombudsman. He stopped paying the mortgage and initiated civil action. Over the next 5 years he pursued his complaints with the human rights commissioner, the securities exchange commissioner, consumer rights organizations, via further civil litigation, and petitions to Parliament and the Queen. The foreclosure on the mortgage intensified the complaints and litigation. Finally he made a series of bomb threats leading to his prosecution and referral. When assessed he was righteously indignant, believing he had been right to take extreme action to bring attention to an injustice which had destroyed him and his family and threatened the very economic fabric of the nation. He firmly believed he was owed millions in punitive damages, and that when he inevitably prevailed this would bring down the transnational banking corporation which owned his particular branch office. He regarded himself as a whistle blower who would be publicly recognized as one of the major social reformers of his generation. The changes over time in the grievance, the agents, his state of mind, social situation, beliefs, and aims are presented schematically in Table 11.11.1

Traditionally psychiatry has attempted to distinguish between deluded querulants, who are in the business of mental health services, and the non-deluded, who are not. Unfortunately for this approach the querulous present a formidable phenomenological challenge. They advance their ideas plausibly making apparently rational connections between the underlying grievance, which is almost always based on some actual injustice, and their current claims and complaints. Unlike many deluded patients, their beliefs do not usually seem to arise either on the basis of some difficult to understand interpretation of an event, or from an idiosyncratic insight into reality. The querulous offer a detailed and apparently logical account of the emergence of their grievances and the progress of their quest for justice. Reasonable that is if taken in cross section but not when considered over time when there emerge gross discrepancies between the supposed initiating cause and subsequent behaviour. The persuasive presentation can obscure the essential absurdity of the quest and distract attention from the chaos they have created for themselves and those around them. The temptation is to normalize the clinical presentation but this is

Table 11.11.1 A case of querulous behaviour: the changes over a 5-year period in the various domains illustrating the descent from the reasonable if over hopeful and oversensitive to the unrealistic and unrealizable

Grievance	*Agents*
Errors in mortgage documents	Bank's accountant
Potential financial loss	Plus manager
Actual financial loss	Plus senior management
Victim of major fraud and theft	Plus banking ombudsman
System wide corruption	Plus lawyers and judges
A campaign of financial corruption threatening the nations economic stability	Plus wife
	Plus various public agencies
	Plus police
	Plus prime minister
	Plus secret services
State of mind	*Social situation*
Rigid discontented man obsessional traits but articulate and ambitious	Moderately successful small businessman, married, two children, but experiencing financial pressures and marital problems
Increasingly fixated on grievance	Business begins to fail as all his attention moves to grievance
Pursuit of justice subordinates all other concerns	Marriage breaks down
Increasingly convinced he is being persecuted and spied upon	Alienates few friends he had
He is a man of destiny fighting forces of national and international corruption	Bankruptcy
	Living alone
	Destitute
Beliefs	*Aims*
That order and due process are the bedrock of civilization	Compensation for malfeasance (the sum rapidly escalating over the years)
Bad things happen not by chance but because of carelessness or malevolence	Reparation—a return of his house mortgage free
That he had never received the recognition and rewards he deserved	Retribution—punitive damages, sackings, and criminal prosecutions
He was destined for greatness	Vindication by public recognition as a whistle blower who had reformed the banking system

(Reproduced from P. E. Mullen and G. Lester, Vexatious litigants and unusually persistent complainants and petitioners: from querulous paranoia to querulous behaviour, *Behavioural Sciences and the Law*, **24**, 333–49, copyright 2006, John Wiley & Sons, Inc.)

to ignore both the peculiarity of their behaviour and beliefs, as well as the devastation they have wrought on their own lives. Sometimes the querulous are obviously deluded, sometimes they appear to inhibit that borderline which is captured in such terms as overvalued ideas and delusion like ideas. Debates over the phenomenological niceties should not, however, distract from recognizing the pathological nature of such querulousness.

The querulous are sometimes regarded as obsessional. The level of preoccupation, the ruminative quality of their thinking, and the pedantic attention to the minutiae of their case, all suggest obsession. Certainly most, if not all, querulants have obsessional personality traits. But the querulant does not regard their core beliefs and the behaviour as absurd or absurdly insistent. Quite the reverse they know they are right and are totally identified with their ideas. The querulous therefore may be regarded as obsessive or fixated but not as having an obsessional disorder.[6]

Management

Our courts and agencies of accountability are designed to deliver conciliation, arbitration, reparation, and compensation, but rarely retribution, except in the exceptional case of punitive damages, and never personal vindication. The querulous seek above all personal vindication and retribution so from the outset are doomed to fail.

One view is that it is the failures of the courts and complaints organization that drive claimants to become querulous. Charles Dickens articulates this in his great novel Bleak House arguing the courts 'give to monied might the means abundantly of wearying out the right. . . (which) so overthrows the brain and breaks the heart to leave its worn out lunatics in every madhouse'. Lester and colleagues[10] in their study of unusually persistent complainants failed to document any significant differences between the manner in which the complaints had been dealt with in those whom became querulous and those who did not. This suggests a role in the pathogenesis for vulnerability not just reacting to provocations.

The impact of querulousness can be reduced by improved recognition and improved management practices in courts and agencies of accountability.[6] Psychiatrists currently only tend to be involved after the situation has reached the stage of the querulant either becoming seriously depressed or being charged with threats, violence, or contempt.

The literature on the therapeutic management of the querulous is both small and predominantly discouraging.[11] Ungvari,[12] however, reported successful treatment using pimozide. Our own experience is that relatively low doses of atypical antipsychotics are helpful though the response is slow in coming often taking months before there is obvious improvement. The first problem is attaining some semblance of a therapeutic alliance with the patient. This requires avoiding being caught up in discussions of the rights and wrongs of their quest. The focus should be on the price they and their family are paying for the pursuit.[6] Interestingly some of those who come on orders from the court which mandate treatment will accept medication and other therapeutic interventions as they wish to make clear they abide by the law. Paradoxically they can

be ultra compliant patients. A number have continued voluntarily in treatment after the end of the order though they never acknowledge either that they were in error or in need of treatment because of their querulousness. What changes is the involvement in the querulous ideas, the degree of preoccupations, and the behaviour, but the core belief that they were right never wavers. Querulous behaviour appears to be sustained by a range of cognitive distortions including:

- Those who do not fully support their cause are enemies.
- Any lack of progress is the product of malevolent interference from someone.
- Any compromise is humiliating defeat.
- The grievance is the defining moment of their lives.
- That because they are in the right the outcomes they seek must be not only possible but necessary.

These distortions are open to challenge and amelioration if not completely overcoming. In theory the cognitive therapy approaches advocated for the delusions should also be of value.[13,14] The problem with the therapeutic management of querulous behaviour is that we have no trial of treatment or even much beyond case reports. This reflects widespread prejudice that the querulous are not the business of mental health and even if they are they are untreatable. Hopefully if this neglect is overcome and querulous behaviour is once more recognized as a legitimate concern for mental health professionals then systematic studies of therapy will follow. For the present the querulous destroy the fabric of their lives as well as creating distress and occasionally damage to those around them.

Further information

Charles Dickens. *Bleak House*—Multiple Editions.

Mullen, P.E. and Lester, G. (2006). Vexatious litigants and unusually persistent complainants and petitioners: from querulous paranoia to querulous behaviour. *Behavioural Sciences and the Law*, **24**, 333–49.

Douglas, M. (1992). *Risk and blame, essays in cultural theory*. Routledge, London.

References

1. Hack Tuke, D. (1892). *A dictionary of psychological medicine*, pp. 1060–1. Churchill, London.
2. Jaspers, K. (1923). *General psychopathology* (trans. J. Hoenig and M.W. Hamilton 1963). Manchester University Press, Manchester.
3. Krafft Ebing, R. (1905). *Textbook of insanity* (trans. C.G. Chaddock), pp. 397–9. Davies Company, Philadelphia.
4. Kraepelin, E. (1904). *Lectures in clinical psychiatry* (trans. and ed. T. Johnstone). Bailliere, Tindall and Cox, London.
5. Caduff, F. (1995). Querulanz-ein verschwindendes psychopatholgisches Verhaltensmuster? *Fortschritte der Neurologie-Psychiatrie*, **63**, 504–10.
6. Mullen, P.E. and Lester, G. (2006). Vexatious litigants and unusually persistent complainants and petitioners: from querulous paranoia to querulous behaviour. *Behavioural Sciences and the Law*, **24**, 333–49.
7. James, D.V., Mullen, P.E., Meloy, J.R., *et al.* (2007). The role of mental disorder in attacks on European politicians 1990–2004. *Acta Psychiatrica Scandinavica*, **116**(5), 334–44.
8. Mullen, P.E., James, D.V., Meloy, J.R., *et al.* (in press). The fixated and the pursuit of public figures. *Journal of Forensic Psychiatry and Psychology*.
9. Poole, S. (2000). *The politics of regicide in England, 1760–1850: troublesome subjects*. Manchester University Press, Manchester & New York.
10. Lester, G., Wilson, B., Griffin, L., *et al.* (2004). Unusually persistent complainants. *The British Journal of Psychiatry*, **184,** 352–6.
11. Dietrich von, H. (1968). *Der Querulant Munchener Medizinische Wochenschrift*; 14, June 110 Jahrgang, pp. 1445–50.
12. Ungvari, G.S. (1993). Successful treatment of litigious paranoia with pimozide. *Canadian Journal of Psychiatry*, **38**, 4–8.
13. Chadwick, P., Birchwood, M., and Trower, P. (1996). *Cognitive therapy for delusions, voices and paranoia*. John Wiley and Sons, Chichester.
14. Bentall, R. and Kinderman, P. (1994). Cognitive processes and delusional beliefs: attributions and the self behaviour. *Research and Therapy*, **32**, 331–41.

11.12

Domestic violence

Gillian C. Mezey

Introduction

Over the last decade, the issue of domestic violence has been transformed from a position of 'selective inattention' to becoming a high-priority social and public health issue.[1] Although it is now recognized that experiences of domestic violence are associated with adverse mental as well as physical health outcomes for the victim, this has not always been the case. Early psychiatric writings tended to attribute responsibility for violent relationships, to the masochistic traits of women who are drawn to and then fail to separate, from abusive and violent partners. During the 1980s and 1990s, however, the perception of victims of domestic violence, or 'battered women' began to change, towards an understanding that the responsibility for domestic abuse lies with the wife beater, rather than the wife beater's wife.

This change came about as a result of several factors; first, effective lobbying by the feminist movement which put the issue of domestic violence firmly on the political agenda; second, the influence of a number of researchers who began to conceptualize the psychological and behavioural problems seen in victims as the consequence, rather than the cause of, domestic violence, for example through the identification of a specific 'Battered Woman Syndrome',[2] and finally the introduction of Posttraumatic Stress Disorder as a distinct psychiatric diagnosis in 1980 (APA, 1980).

Definition

Domestic violence is currently defined in the United Kingdom as 'Any incident of threatening behaviour, violence or abuse (psychological, physical, sexual, financial or emotional) between adults who are, or have been in an intimate relationship'.[3] Most cases of domestic violence involve the abuse of a woman by her male partner, however, domestic violence may also involve other family members, same sex partners, and the abuse of men by women partners.

Epidemiology

The estimated prevalence of domestic violence varies, according to the definition being used and the population surveyed. Based on an analysis of 48 population-based studies from around the world, the prevalence of domestic violence is between 10 and 69 per cent over a lifetime and between 3 and 52 per cent in the past year.[4] In the United Kingdom, around 16 million incidents of domestic violence are recorded annually, with a lifetime prevalence of 21 per cent lifetime and 4 per cent in the past year for women and 10 per cent lifetime and 2 per cent in the past year, for men.[5] Women who report domestic abuse consistently report experiencing more incidents and more injuries and being more fearful of their partner, than male victims.[5]

Aetiological factors

Domestic violence arises out of a complex interplay of personal, situational, and socio-cultural factors. Poverty, alcohol use, low academic achievement, being single, separated or divorced, and witnessing or experiencing violence as a child are the most important individual risk factors for domestic violence.[4] Poverty operates at an individual and societal level, in that it places increased stress on the individual and the family system and also acts as a marker for a number of other social conditions (e.g. low education, overcrowding) that combine to increase the risk of domestic violence. Alcohol use by the perpetrator also significantly increases the risk of domestic violence.[4,5] Women appear to be at particularly high risk of domestic violence during pregnancy and the immediate post-partum period.[6] Experiences of domestic abuse in childhood and adolescence are associated with an increased risk of perpetration, for men[4] and re-victimization, for women[7] in adult relationships. Inequality and a power imbalance within the relationship[8] and in the wider society,[9] also increases the risk of domestic violence. Higher rates are found in societies where men have economic and decision-making power in the household, where women do not have easy access to divorce, and where there is a high level of public acceptance of men's right to discipline their wives.[4,9]

Mental health effects of domestic violence

Domestic violence is associated with a range of adverse physical and psychological health and social outcomes.[1,4,5] In extreme cases, the violence results in the victim's death. Between 40 and 70 per cent women victims of homicide are killed by a current or former spouse or partner, compared with between 4 per cent and 8 per cent male victims of homicide.[4]

Women who are abused by their partners over many years, often find it extremely difficult to leave their partner and, even if they do leave, many women end up returning to the family home. There are a number of psychological and social reasons why many women find it difficult to separate emotionally, as well as physically, from their violent partners. Learned helplessness,[2] the progressive erosion of confidence and self esteem cased by the abuse, feelings of guilt, shame and isolation, make it difficult for victims to assert themselves, seek help, or even contemplate the possibility of an existence separate from their partner. For some women the abuse appears, paradoxically, to strengthen their emotional ties to the perpetrator, so-called 'traumatic bonding'.[10] Women who try to leave abusive partners, experience significant economic and social hardship, including lack of accommodation, inadequate financial support, and difficulties in caring for the children, as well as a disruption caused to their family and social support networks. These difficulties, as well as the fear of being tracked down by their partner, are often so daunting, that many women end up returning to their homes, choosing an existence that is familiar to them, in preference to precarious survival elsewhere. It is clear that physical separation does not always end the violence; many separations result in an escalation of threatening and violent behaviour, including stalking[11] and the risk of domestic homicide is greatest around the time, or shortly after, separation.

Domestic abuse is associated with increased rates of depression, suicidality, Posttraumatic Stress Disorder, alcohol and substance misuse, and dependence in victims of abuse, compared with the general population.[12] The more severe and chronic the abuse, the greater the impact on the victim's mental health and symptoms of depression and hypervigilance may persist, even after separation.[2]

Where present, symptoms of mental illness must be treated appropriately. However, offering victims a non-judgemental sympathetic response, providing information about options available to them, about the risks associated with staying, or going and providing information about community resources, including refugees and counselling facilities are likely to be more important for the victims than any specific 'treatment'.[13]

Risk assessment

In the absence of direct questioning by a health professional, which is conducted in a sensitive and non-judgemental way, women are unlikely to spontaneously disclose experiences of domestic abuse in the context of a health consultation.[5,13] Unless domestic abuse is identified, then the risk cannot be properly assessed or communicated to the victim. Screening for domestic abuse in health settings appears to be acceptable to women and is also effective in terms of increasing rates of identification.[13]

The frequency, severity, and chronicity of the abuse must be taken into account in assessing and managing risk. Browne's study,[14] comparing battered women who had killed their partner with battered women who had not killed, identified the following risk factors for domestic homicide: frequency of violent assaults; presence and severity of injuries; alcohol intoxication or substance misuse in the perpetrator, threats to kill, sexual violence, suicidal ideation (in the victim), and access to a weapon. Campbell[15] also identified the following risk factors in a study of 220 cases in which women had been killed by violent partners: unemployment (male), choking, abuse during pregnancy, threats or harm to the child, the presence of a stepchild in the home, and separation. Morbid jealousy in the perpetrator and stalking behaviours are also more common in lethal, compared with non-lethal cases of domestic violence. Risk assessment must be carried out in all cases of domestic violence and communicated to the victim, to allow her to make informed and safe choices.

Confidentiality and domestic violence

The primary consideration for the health professional should be the safety of the victim and any affected children. In general, the victims consent should always be sought prior to sharing information with others. A breach of confidentiality, without the victim's knowledge, or consent, could increase the risk she faces and may discourage further disclosures to health professionals. The legal framework varies in different countries but in the United Kingdom, health professionals should be guided by the Data Protection Act (1998) and by principles on information sharing and good practice, as set out within GMC[16] and Royal College of Psychiatrists[17] guidance. The Crime and Disorder Act (1998) allows for information to be passed on, in the absence of consent, in cases where the courts request information about a specific case or if the health professional judges there to be a significant risk of harm to the woman, her children, or to someone else if that information is not passed on. The health professional should always inform the woman if they intend to breach confidentiality, they should properly record their reasons for doing so and they should only pass on the minimum information required to achieve their objective. Further specific guidance on sharing information in the context of domestic violence has recently been provided by the Home Office.[18]

Perpetrators of domestic violence

Studies of men who abuse their partners are difficult to interpret, largely because of the difficulties in identifying and recruiting men who are representative of the population as a whole. There is little evidence that domestic abuse is primarily attributable to underlying mental illness in the perpetrator. However, personality profiles of perpetrators indicate high rates of personality disorder, as well as alcohol and substance misuse and jealousy. There are two main personality profiles described in a perpetrator of domestic violence. The borderline/emotionally dependent type tends to confine their violence to within the family, they tend to be extremely insecure, jealous, and dependent on their partner and the violence is often precipitated by actual or threatened separation. The antisocial/narcissistic offender, is violent both within the family and outside and their violence is often associated with alcohol and drug misuse and high rates of criminality.[19,20]

Group treatment programmes, aimed at changing the behaviour of men who batter, have had a degree of success in reducing violence.[21,22] However the evidence is limited and results are hard to interpret, given the fact that many men, arguably those who are at highest risk and who most need behaviour changing, are least likely to seek help or to remain in treatment. Treatment appears to be most effective when men are mandated through the Courts. Elements of such programmes include: encouraging men to take personal responsibility for their violent behaviour; increasing their awareness of the dynamics involved in the use of violence in relationships; challenging attitudes and beliefs around the use of

violence in relationships between men and women; and developing skills for relating non-violently to others. Perpetrators of domestic abuse are encouraged to accept responsibility for the violence and to consider the consequences of their violence and the gains and losses it entails. Some perpetrators may need to learn particular skills to manage situations where they may previously have resorted to violence. It is important that any treatment programme of perpetrators monitors the effectiveness of the programme through maintaining constant contact with female partners in order to confirm whether participation is having the desired effect and, more importantly, is not endangering them further.

Further information

Roberts, G., Hegarty, K., and Feder, G. (2006). *Intimate partner abuse and health professionals. New approaches to domestic violence*. Churchill Livingstone Elsevier, London.

Taket, A. (2004). *Tackling domestic violence: the role of health professionals*. Home Office Development and Practice Report, No. 32. Home Office, London.

www.womensaid.org.uk

www.refuge.org.uk

www.victimsupport.org.uk

References

1. British Medical Association. (1998). *Domestic violence: a health care issue?* BMA, London, UK.
2. Walker, L.E. (1979). *The battered women*. Harpers and Row, New York.
3. Home Office. (2005). *Domestic violence: a national report*. Home Office, London, UK.
4. World Health Organization. (2002). *World report on violence and health*. World Health Organization, Geneva.
5. Walby, S. and Allen, J. (2004). *Domestic violence, sexual assault and stalking: findings from the British crime survey* (ed. Home Office Research DaSD), p. 276. Home Office, London, UK.
6. Lewis, G. and Drife, J. (2004). *Why mothers die 2000-2002. Confidential enquiry into maternal and child health*. RCOG, London, UK.
7. Coid, J., Petruckevitch, A., Feder, G., *et al.* (2001). Relationship between childhood sexual and physical abuse and risk of re-victimisation in women: a cross sectional survey. *Lancet*, **358**, 450–4.
8. Jewkes, R. (2002). Intimate partner violence: causes and prevention. *Lancet*, **359**, 1423–9.
9. Levinson, D. (1989). *Family violence in cross cultural perspective*. Sage, Thousand Oaks, CA.
10. Dutton, D. and Painter, S. (1983). Traumatic bonding: the development of emotional attachments in battered women and other relationships of intermittent abuse. *Victimology: An International Journal*, **6**, 139–55.
11. Mullen, P.E., Mackenzie, R., Ogloff, J.R., *et al.* (2006). Assessing and managing the risks in the stalking situation. *The Journal of the American Academy of Psychiatry and the Law*, **34**, 439–50.
12. Golding, J. (1999). Intimate partner violence as a risk factor for mental disorders: a meta analysis. *Journal of Family Violence*, **14**, 99–132.
13. Department of Health. (2005). *Responding to domestic abuse: a handbook for health professionals*. Department of Health, London.
14. Browne, A. (1987). *When battered women kill*. The Free Press, New York.
15. Campbell, J.C., Webster, D., Koziol-McLain, J., *et al.* (2003). Risk factors for femicide in abusive relationships: results from a multisite case control study. *American Journal of Public Health*, **93**, 1089–97.
16. General Medical Council. (2004). *Confidentiality: protecting and providing information*. General Medical Council, London.
17. Royal College of Psychiatrists. (2000). *Good psychiatric practice: confidentiality*. CR85. Royal College of Psychiatry, London.
18. Douglas, N., Lilley, S., Kooper, L., *et al.* (2004). *Safety and justice: sharing personal information in the context of domestic violence: an overview*. Home Office Development and Practice Report No. 30. Home Office, London.
19. Gilchrist, E., Johnson, R., Takriti, R., *et al.* (2003). *Domestic violence offenders: characteristics and offending needs*. Home Office, London, UK.
20. Holtzworth-Munroe, A. and Stuart, G.L. (1994). Typologies of male batterers: three subtypes and the differences among them. *Psychological Bulletin*, **116**, 476–97.
21. Gondolf, E.W. (2000). Re-assault at 30 months after batterer programme intake. *International Journal of Offender Therapy and Comparative Criminology*, **44**, 111–28.
22. Morran, D. and Wilson, M. (1994). Confronting domestic violence: an innovative criminal justice response in Scotland. In *Penal theory and practice: tradition and innovation in criminal justice* (ed. P. Dobash). Manchester University Press, Manchester.

11.13

The impact of criminal victimization

Gillian C. Mezey and Ian Robbins

Epidemiology

The prevalence of crime depends on the methodological approach that is adopted, the questions being asked and the population being surveyed. Crime figures are also affected by the willingness, or unwillingness of individuals to declare themselves as victims. This is particularly the case with 'sensitive' crimes, such as domestic and sexual violence. Not surprisingly, the self completion phases of the British Crime Surveys have been more effective in identifying such experiences than standard survey methodology. In addition, not all violence is necessarily recognized as a crime, and similarly not all crimes are necessarily defined as such, by their victims. Occasionally, violence may occur by mutual consent.

In the UK the British Crime Survey is regarded as the most reliable and comprehensive data source on criminal victimization, providing information about the extent of crime, as well as trends in the frequency and patterns of crime and changes in public attitudes to crime over the years. The British Crime Survey for 2005–2006 found there were approximately 10.9 million crimes against adults living in private households.[1] Not all individuals in the population are at equal risk. Most crime differentially targets and damages individuals who are poor, disempowered, and marginalized within society. The risk of victimization is highest for divorced single or separated individuals between the ages of 16–24. Men are at greater risk of experiencing violent crime than women, except in the categories of domestic violence and sexual assault, where women are more at risk.[2] Women are more vulnerable to domestic violence and younger people are more at risk of crime than elderly people. Alcohol is also involved in a significant number of violent offending both offenders and victims having been found to be inebriated or to have recently consumed alcohol at the time of the offence.[3] Severe mental illness also appears to be a risk factor for crime victimization. In her study of 936 patients with severe mental illness living in the community, Teplin[4] found that over one quarter of them had experienced violent crime in the past year, a rate more than 11 times that of the general population, even after controlling for socio-demographic variables.

Criminal victimization can have profound psychological and emotional effects, with the impact of violent (including sexual) victimization being greater than property or non violent offending. The experience of crime and the perception of crime as possible or probable, also has an impact on the individual's fear of crime, and on their lifestyle. Women and the elderly are most fearful of crime, even though, in reality, young men are at greater risk. This may be because of the greater perceived adverse consequences of victimization and greater vulnerability in women and elderly people.[5]

General effects of victimization

The experience of victimization can leave the individual feeling 'diminished, pushed down, exploited and invaded'.[6] Victims are often describe feeling stigmatized and isolated and unable to communicate their distress or feelings of vulnerability. Although friends and relatives may initially be supportive, such support may begin to fall away if the victim fails to recover within a reasonable period of time.

Social support and gender are important predictors of psychiatric problems, including PTSD, in crime victims.[7] Andrews *et al.*[8] found that, women were more likely to receive negative responses from family and friends following violent crime and also to have higher rates of PTSD at 6 months follow up. The benefits of positive social support and impact of negative social responses were greater for women victims than men. Negative support was predictive of PTSD in both men and women, although this effect was more pronounced in women.

Immediate and short term effects of crime

High rates of dissociative symptoms are reported by victims of violent and sexual assault in the form of numbing, reduced awareness, derealization, and depersonalization, at the time of the crime.[9,10] Dissociation may have immediate survival value, in terms of reducing the victim's sense of immediate threat and minimizing the pain. However, it may also interfere with the individual's longer-term recovery. Peri-traumatic dissociation at the time of the offence has been found to predict post-traumatic stress disorder development in women victims of violent and sexual assault.

Dissociative symptoms are part of the diagnostic criteria for Acute Stress Disorder which is a risk factor for the development of Posttraumatic Stress Disorder. Brewin *et al.*[11] found that a diagnosis of Acute Stress Disorder at one month post trauma, predicted 83 per cent of post-traumatic stress disorder cases at 6 months follow-up. Acute Stress Disorder may be experienced during or

immediately after a trauma and should resolve within four weeks of the conclusion of the traumatic event.

Long term psychological effects

Most crime victims are able to resume normal functioning and health, following a transient state of disequilibrium, without the need for medical or psychological intervention. However, some victims go on to develop chronic and persistent psychological or psychiatric problems.[12] This may include increased rates of depression, anxiety and substance misuse. Recovery following criminal victimization is largely dependent on how the victim processes and makes sense of what has happened, whether the act can be accommodated into an existing frame of reference or whether the experience is so overwhelming and outside ordinary everyday experience as to render them incapable of reaching some kind of resolution.

Amongst crime victims, victims of violent crime have higher rates of psychological disturbance than victims of property crime who, in turn, have higher rates of disturbance that non victims.[13,14,15] Perception of life threat, physical injury and completed rape are associated with particularly high rates of PTSD.[14,16,17]

Many of the psychological responses exhibited by victims and witnesses of crime fit within a post-traumatic stress disorder framework. In a study of 391 women victims, 27 per cent of all crime victims developed post-traumatic stress disorder.[18] Although most victims of crime show substantial improvement up to 9 months after the offence, very little spontaneous recovery occurs thereafter[19,20] and for some victims the effects are profound and long lasting.[18]

Physical health effects of crime

In general, people who have experienced crime have a poorer perception of their physical health and physical functioning and experience more chronic medical conditions than non victims. Physical and sexual assault are associated with increased cigarette consumption, alcohol and other drug abuse, self neglect, risky sexual behaviour, and eating disorders.[21] Shepherd and Farrington[22] have suggested that a young man from a deprived urban area may suffer 60 years of incapacity as a result of injury, reduced quality of life, and self esteem. Increased crime rates are found in poorer areas, which means that the negative impact of crime on physical health may be difficult to disentangle from the negative impact of poverty and deprivation on physical health.

Responses following specific criminal acts

(a) Murder

The act of murder has profound effects, not just on the individual victim, but also on the friends, family members and acquaintances who are left behind and who are sometimes referred to as the 'secondary victims.' The act of murder is shocking in its finality and irrevocability, and the responses of survivors are both qualitatively and quantitatively different from the normal grieving process.[23,24] Rock[25] has suggested that it is not just the death itself, but the manner of death and its social meaning that is so devastating for those who are left behind. Unlike a 'natural' death, survivors are unprepared for their loss, there can be no anticipatory mourning, no reconciliation, and no proper leave taking. Many survivors describe feelings of stigmatization, isolation, shame, and betrayal, but feel unable to communicate their distress or to connect emotionally with fellow beings. They often feel marginalized by the criminal justice system, with little access to information and they are burdened with having to cope with the inevitable, legalistic bureaucracy and the practical demands of life during a period of acute distress and emotional turmoil. In cases where the perpetrator and the victim are members of the same family, the survivors may experience particularly intense feelings of guilt and conflicting emotions.

The effects of violent traumatic bereavement on the secondary victims include physical health problems, cognitive impairment, and psychological effects, including posttraumatic stress disorder, depression, phobic avoidance, and impaired work and social functioning.[26,27] Female gender and losing a child predict worse psychiatric outcome.[27] Survivors of homicide, tend to manifest both trauma symptoms and symptoms of grief, phenomena, with either predominating or appearing intermittently.[28] This has lead to a proposed new diagnostic category of 'traumatic grief', which contains two core components; trauma and of loss.[29]

(b) Rape and sexual assault

The definition of rape varies across countries and between states within countries. In the UK, prior to 1994, the definition of rape was restricted to penile penetration of the vagina, with other forms of non consensual penetrative sex being defined as indecent assault. However, in 1994, the definition of rape was extended to include non consensual anal intercourse, thereby recognizing male rape victims for the first time. The 2003 Sexual Offences Amendment Act further broadened the definition of rape to include penetration of the mouth as well as penetration of the vagina or anus by the penis. It also introduced three new measures on the issue of consent: first, that a person can only consent to sexual relations if they have the freedom and capacity to make that choice, second, that all the circumstances at the time of the offence must be considered in determining whether the defendant is reasonable in believing the complainant consented and third, that individuals will be considered most unlikely to have agreed to sexual activity if they were subject to threats or fear of serious harm, if they were unconscious, drugged or abducted, or if they were unable to communicate because of a physical disability.

In the majority of cases of rape, the perpetrator is known to the victim and in many cases, the rapist is the current or former husband or partner.[30] In spite of the seriousness of the offence, the British Crime Survey[30] found that only 60 per cent women who had been subjected to rape or serious sexual assault had told anyone about it and only around one in seven cases had been reported to the police. Reasons women gave for not reporting include: fear of reprisals, fear of public identification, fear of appearing in Court and having to give evidence and lack of confidence in the legal system.[31] There is some evidence suggesting that women who proceed with prosecution following rape do worse, in terms of social adjustment and self esteem at one year follow up than women who decide not to proceed.[32] Whether this is because Court proceedings delay or slow down the process of psychological recovery following rape, or because the legal process and particularly the experience of being cross examined in Court, represents a form of 'secondary traumatization', is not entirely clear. If women do proceed with the Criminal Justice process, however,

preparing them for the experience, providing them with appropriate information beforehand and giving them the opportunity to exercise choice, can help to offset the potentially de-stabilizing and distressing impact of criminal proceedings. Ultimately, the attitude taken by the police and the way the victim feels they have been treated appears to be more important in determining their psychological adjustment and satisfaction with the process, than the actual verdict.[32]

About one-third of women who report rape develop long term psychological and social problems. These effects tend to be more severe and chronic than following non sexual violence.[33,34] Rape trauma syndrome was first described in the 1970's[35] and was subsequently superseded by Posttraumatic Stress Disorder. Posttraumatic stress symptoms are generally present in the days and weeks following the assault, but then spontaneously resolve in the majority of cases. For some victims, however, the condition may become chronic and persist for many years, if left untreated.[36] Higher rates of depression, suicidal ideation, generalized and phobic anxiety, alcohol and drug dependence and sexual dysfunction as well as physical health problems are also found in rape victims compared with non rape victims.[37–41]. Women who have been raped often describe problems in relationships, with excessive dependence, inability to trust and loss of confidence and self esteem. Similar responses have been described with male victims of sexual assault.[42]

The characteristics of sexual assault that predict long term mental health problems are: being the victim of a completed rape, being injured and the perception of a threat to life.[14] Other predictors of long term disturbance include; prior psychological and social problems previous victimization, particularly childhood abuse past psychiatric illness, drug or alcohol misuse and lack of a supportive network.

Psychiatric treatment may be required for individuals who develop serious psychological problems following a sexual assault. It is probably inappropriate to embark on psychiatric treatment too early, because of the natural tendency for symptoms to resolve spontaneously in the weeks and months following the assault. Unless symptoms have resolved by 6 months, however, they are unlikely to resolve spontaneously thereafter without some form of psychological intervention or psychiatric treatment (see Chapter 4.6.2)

There is no evidence that counselling is effective in alleviating short term distress or in preventing the development of long term psychiatric disability in rape victims. Indeed recent studies have suggested that counselling may even be harmful, if carried out by inadequately trained and supervized individuals.[43] Many of the key organizations working with victims e.g. Victim Support, emphasize their role as supporters and befrienders and they provide both practical assistance, such as accompanying victims to identification parades, helping with paperwork and compensation claims, as well as offering emotional support following the assault and through any subsequent criminal proceedings. Rape victims are most likely to benefit from services that are co-ordinated, integrated and streamlined.[43] An example of this in the UK has been the development of Sexual Assault Referral Centres (SARCs), which provide a 'one stop shop' of medical, counselling, legal and forensic services for victims of sexual assault. Evaluation of these centres and in particular the benefits for victims and the Criminal Justice System, is ongoing.

(c) Burglary and robbery

The effects of burglary are generally less severe and long lasting than following violent and 'contact' crimes although some victims may develop chronic mental health problems, including Posttraumatic Stress Disorder.[18] Repeat victimization is especially common in the case of burglary and second or subsequent burglaries are more likely to have a greater long term impact. Repeated experiences of burglary may lead to victims of burglary taking additional security precautions or even moving house in order to restore a sense of safety and control over their lives.[44] Individuals living in areas of poverty and deprivation and single parent families are most vulnerable to burglary. Robbery, unlike burglary involves not only direct contact between the victim and the perpetrator, but also implies a degree of life threat and is therefore more likely to precipitate post-traumatic psychiatric illness, including post-traumatic stress disorder.[45]

(d) Workplace violence

Violent assaults in the workplace have increased in frequency and severity in recent years.[46–48] Budd *et al.* found that working at night was the only significant factor which predicted its occurrence. There was a clear relationship between workplace violence and increased job stress, reduced job satisfaction, increased likelihood of looking for a new job, as well as bringing weapons to work Kopel & Friedman[49] found that police officers witnessing incidents of violence reported intrusive thoughts and images, and used avoidance to deal with the intrusive phenomena. Whilst recognizing that avoidance is a feature of post-traumatic stress disorder, they also suggest that denial and avoidance are part of the culture in male-dominated occupations such as law enforcement agencies. Miller-Burke *et al.*[50] found that most employees who had experienced robberies had multiple adverse consequences. Psychological functioning, physical wellbeing, social and occupational functioning were all impaired. Having been involved in more than one incident was associated with more severe outcomes.

In the United States Hewitt and Levin[51] point to the high rates of occupational assaults among health-care workers. Whilst fatal workplace assaults are more likely to involve males in the course of robberies, women were more likely to be involved in non-fatal workplace assaults, with health-care workers being most affected. The rate for health and social care workers was 10 times that of private non-health care industries. Williams[52] found that 26 per cent of nurses reported physical assault while at work. This rate of assault is similar to that found by O'Connell and Bury[53] who carried out a survey of all general practitioners in the Eastern Health Board of Ireland. They had a 98 per cent response rate which revealed that 21 per cent of general practitioners had experienced violence or aggression although in only 7 per cent of the incidents reported did it result in injury. Not surprisingly there is a strong inverse relationship between workplace assaults and job satisfaction.

(e) Child and adolescent victims of crime

Children and young people are especially vulnerable to crime and victimization, in particular by people they know and are dependent on, although stranger violence is a small but important problem.[54] The 1992 British Crime Survey[55,56] looked at the victimization of young people away from the home and reported that about one-third of 12 to 15 year olds had been assaulted in the last 6 to 8 months. This would compare with a rate of around 1 per cent in

adults. Most of these assaults happened at or near school and were committed by a sole perpetrator who was known to the victim. The survey also identified high rates of fear of crime amongst adolescents, with girls being more fearful than boys. A range of mental health problems have been identified following victimization within the child and adolescent population.[57]

Hate crimes

The defining characteristics of a hate crime is that the individual victim is targetted because of bias or prejudice, based on their actual or perceived social grouping or ethnicity, sexual orientation, religion or political orientation. Herek *et al.*[58] suggest that a significant proportion of lesbians and gay men who had been assaulted believed that their sexual orientation had been a motivating factor. One-quarter of men and one-fifth of women said that they had experienced victimization because of their sexual orientation. When compared to other crime victims, they had significantly more symptoms of depression, anxiety and PTSD as well as more crime related fears and a lower sense of mastery. Rose & Mechanic[59] reported that 73 per cent of lesbians and gay men had experienced at least one homophobic attack. Victims of homophobic violence had more PTSD symptoms than did victims of other homophobic crimes or non victims although there was no significant difference in rates of depression. They also found that homophobic sexual assaults were more likely to involve known assailants, and multiple perpetrators and more likely to be repeated.

Hate crimes tend to be under-reported to the police, the psychological effects tend to be long lasting, and have a negative impact not only on the individual victim, but also on the community. It has been argued that hate crimes require policy directed to addressing the causes rather than simply dealing with the needs of individual victims. Black and ethnic minority individuals appear to be particularly vulnerable to criminal victimization and in a proportion of these cases the offence is considered to be 'racially motivated'. [60,61]. Around one in six of all incidents of criminal victimization against Asians and African-Caribbeans are considered to be racially motivated[61] and, regardless of the type of offence, ethnic minority victims tend to report higher levels of worry about crime then white victims.[60] Fear of crime appears to be particularly high amongst Bangladeshi and Pakistani individuals, who are also most likely to describe their victimization as racially motivated. It has been suggested that victims who perceive their victimization to be racially motivated, experience more serious and persistent psychological effects that individuals who did not consider racism as a motivation.

(g) Terrorist crimes

A number of studies have looked at the impact of terrorist attacks which, whilst relatively infrequent, in comparison to other crimes, have a considerable political and social impact. Whalley and Brewin[62] reviewed the mental health effects of terrorist attacks on the direct victims, as well as on the general (non affected) population. In the general population, symptoms of distress and stress are high in the first few days following an incident, but tend to resolve spontaneously so that, although PTSD is found in between 11 per cent-13 per cent of the general population, in the first six weeks following a single terrorist attack, this falls to below 3 per cent after 8 weeks. Factors such as previous adversity, pre-existing mental health issues and membership of minority groups may increase vulnerability to the impact of terrorism. PTSD is the most common psychiatric disorder in direct victims of terrorist attacks, followed by depression although other problems such as traumatic grief, panic disorder, phobias, generalized anxiety and substance misuse have also been reported.[63] Most studies have reported that between 30–40 per cent of those closest to the site of the attack, will develop a psychiatric disorder within 2 years. However, the majority of direct victims have no contact with mental health professionals. In the aftermath of the September 11th attacks in New York only around one-quarter to one-third of those suffering from PTSD were in receipt of treatment. Whalley and Brewin[64] suggest that a focused outreach approach such as a 'screen and treat' programme may be needed to identify those with significant impairment and help them to access evidence based treatments.

The extent of physical injury during a terrorist attack is the best predicator of post-traumatic stress disorder rates both in the short term[65] and many years after the event,[66] although it does not predict the development of a depressive illness. Studies which have looked at the impact of shootings tend, by their very nature to be small scale, but have found significant levels of distress and high rates of post-traumatic stress disorder and other psychiatric disorders.[67] Being held hostage has also been related to high levels of distress both in victims and their families.[68] Where captivity has occurred, there may be strong attachment and paradoxical gratitude towards the captors with positive emotions including compassion and identification with the terrorist's values, often described as the Stockholm syndrome.

Support services and treatment interventions

There are so few culturally accepted rituals of support for victims of crime, that it often becomes the task of the therapist to normalize the process. In the United Kingdom victim support schemes offer practical assistance, for example accompanying the victim to identification parades and court hearings, completing Criminal Injuries Compensation Board forms, as well as providing support and reassurance following crime. Referrals to the schemes are generally made by the police, but are occasionally accepted from involved professionals or from the victims themselves. However, victims of serious crimes, such as sexual assault or physical violence, and the families of murder victims may develop psychiatric illness, which requires referral to mental health services for specialist treatment.

Treatment approaches are drawn from a variety of paradigms including cognitive behavioural, psychodynamic, psychosocial, and pharmacological treatments, and are often trauma focused in general rather than being specific to problems associated with criminal victimization. Ochberg (6) categorizes them into two main approaches, the first focuses on previous personality and attributes mental health problems and difficulties in adjustment following crime to pre-existing unresolved issues and weaknesses, rather than to the traumatic events. The second approach focuses more on the events themselves, the individual strengths and coping styles of the victim and setting realistic achievable goals.

In the UK the 1998 Crime and Disorder Act placed an obligation on the NHS to work in partnership with the Police and Local Authorities in dealing with the consequences of crime. Many people who present to the NHS for treatment following violent crime

do not present to the Police or criminal justice agencies. As many of the services for crime victims are organized by or accessed through the criminal justice system these people are likely to miss out on services. There is a clear need for a direct relationship between the NHS and criminal justice agencies in the way in which treatment is provided and its relationship to the criminal justice processes but as yet, other than in a small number of specific projects it does not happen.

Often immediately following victimization help is offered and family and friends rally round. This may not be the time when most people need formal psychiatric help. The majority of people post assault experience sympotms of PTSD and other psychiatric problems but also the majority improve spontaneously even in the case of serious crimes such as rape. The question comes as to when to intervene. The National Institute for Clinical Excellence (NICE, 2006) recommend a position of watchful waiting and if in the first month(s) symptoms are not improving or indeed may be increasing then treatment is indicated otherwise waiting and monitoring to identify who will not spontaneously improve is the approach of choice.

Previous work had already shown that timing of appropriate services is important. Shepherd *et al.*[62] have shown that while accident and assault victims have similar levels of depression and anxiety in the immediate aftermath three months later the assault victims have higher levels of symptoms. In the immediate aftermath of criminal victimization it has been suggested that having the opportunity to talk may reduce later symptoms Where there have been specific RCTs of immediate intervention they have not shown themselves to be particularly effective. Rose, Brewin, Andrews and Kirk[63] compared an education intervention with education plus psychological debriefing or an assessment only condition in a randomized controlled trial. They found that while all groups improved over time there were no significant differences between groups. This is hardly surprising when considering the result of Rose, Bisson & Wessely[64] who carried out a systematic review of RCTs examining the impact of debriefing. They found 11 RCTs of which six showed no effect at all, three had a positive effect and two had a negative outcome.

This has led to a belief that there is no point in any immediate interventions but this ignores the contribution made by organizations such as Victim Support or Rape Crisis centres. They do not offer formal therapy but rather give information, offer the opportunity to ventilate emotions, practical help and assistance in making compensation claims or participating in the criminal justice system. Shepherd and Bisson[65] who make the case for more integration between health services and other agencies describe the two models of Victim Support interventions which are initiated within the NHS and address the needs of victims which would otherwise be overlooked.

In terms of formal psychological therapies and in the longer term Cognitive Behavioural Therapy has been used with great effectiveness. Ehlers & Clark[66] reviewed the use of CBT in early interventions. They concluded that CBT was superior to supportive counselling but that brief CBT in the first month showed no superiority over repeated assessments. Where longer programmes ie 16 session were offered in the first 4 months they found that CBT was superior to supportive counselling, repeated assessment or no intervention at all.

There have been a number of other empirical studies demonstrating the effectiveness of structured cognitive behavioural approaches to treating established PTSD, which are discussed further in Chapter 4.6.2

Conclusion

The impact of crime on the individual victim is profound but is frequently underestimated by mental health professionals. Wide-ranging personal, social, and economic consequences could be prevented if a range of appropriate interventions were available. Most post traumatic stress treatment programmes in the United Kingdom have developed in response to specific disasters, which may not be relevant to or as effective with crime victims. In order to provide appropriate treatment to crime victims, mental health professionals need to recognize the importance of active interagency liaison with the police, the courts and with voluntary organizations such as victim support schemes. Crime victims tend to be relatively invisible and disempowered; they are less likely to be supported by active campaigning groups than survivors of major disasters and, because of associated feelings of shame and stigmatization, they may be reluctant to claim their entitlement to proper care and treatment. The fact that their plight is often used as a political football is likely to reinforce feelings of helplessness and insecurity. Given its prevalence, crime represents both an ordinary and an extraordinary event; it is likely to affect everyone at some point in their lives and the fact that most crime victims recover from the experience should not deprive those who need it, to proper care.

Further information

Garcia-Moreno, C., Jansen, H., Ellsberg, M., *et al.* (2005). *Multi-country Study on Women's Health and Domestic Violence against Women* WHO, Geneva.

Krug, E., Dahlberg, L., Mercy, J., *et al.* (2002). *World Report on violence and health*. WHO, Geneva.

For general information about victims, policy, statistics and support provision www.homeoffice.gov.uk/crime-victims

Hearing the Relatives of Murder and Manslaughter Victims. Consultation. D.C.A.;London. Sept 2005 www.dcs.gov/manslaughter/manslaughter.pdf

Domestic Violence, Crime and Victims Act (2004) www.opsi.gov.uk/ACTS/acts 2004/20040028.htm

References

1. Walker, A., Kershaw, C., Nicholas, S. (2006). *Crime in England and Wales*. London, Home Office.
2. Kessler, R.C., Sonnega, A., Bromet, E., *et al.* (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry*, **52**(12), 1048–60.
3. Mirrlees-Black, C., Budd, T., Partridge, S., *et al.* (1998). *The 1998 British Crime Survey England and Wales*. London, HMSO.
4. Teplin, L.A., McClelland, G.M., Abram, K.M., *et al.* (2005). Crime victimization in adults with severe mental illness: comparison with the National Crime Victimization Survey. *Archives of General Psychiatry*, **62**(8), 911–21.
5. Maxfield, M.G. (1984). Fear of crime in England and Wales. London, HMSO.
6. Ochberg, F.M. (1988). Post traumatic therapy and victims of violence. In *Post traumatic therapy and victims of violence* (ed. F.M. Ochberg), Brunner-Mazel, New York, 3–19.
7. Brewin, C.R., Andrews, B., Valentine, J.D. (2000). Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. *Journal of Consulting and Clinical Psychology*; **68**(5), 748–66.

8. Andrews, B., Brewin, C.R. and Rose, S. (2003). Gender, social support, and PTSD in victims of violent crime. *Journal of Traumatic Stress*, **16**(4), 421–7.
9. Dancu, C.V., Riggs, D.S., Hearst-Ikeda, D., *et al.* (1996). Dissociative experiences and post traumatic stress disorder among female victims of criminal assault and rape. *Journal of Trauma Stress*, **9,** 253–67.
10. Foa, E.B., Riggs, D.S. and Gershuny, B.S. (1995). Arousal, numbing, and intrusion: symptom structure of PTSD following assault. *American Journal of Psychiatry*, **152**(1), 116–20.
11. Brewin, C.R., Andrews, B., Rose, S., *et al.* (1999). Acute stress disorder and posttraumatic stress disorder in victims of violent crime [see comments]. *American Journal of Psychiatry*; **156**(3), 360–6.
12. Norris, F.H. and Kaniasty, K. (1994). Psychological distress following criminal victimization in the general population: cross-sectional, longitudinal, and prospective analyses. *Journal of Consulting and Clinical Psychology*, **62**(1), 111–23.
13. Kilpatrick, D.G., Saunders, B.E., Amick-McMullan, A.A., *et al.* (1989). Victim and Crime Factors Associated with the Development of Crime-Related Post-Traumatic Stress Disorder. *Behavior Therapy*, **20,**199–214.
14. Ullman, S.E., Siegel, J.M. (1994). Predictors of Exposure to Traumatic Events and Posttraumatic Stress Sequelae. *Journal of Community Psychology*, **22,** 328–38.
15. Breslau, N., Davis, G.C., Andreski, P., *et al.* (1991). Traumatic events and posttraumatic stress disorder in an urban population of young adults. *Archives of General Psychiatry*, **48**(3), 216–22.
16. Norris, F.H. (1992). Epidemiology of trauma: frequency and impact of different potentially traumatic events on different demographic groups. *Journal of Consulting and Clinical Psychology*, **60**(3), 409–18.
17. Kilpatrick, D.G., Saunders, B.E., Veronen, L.J., *et al.* (1987). Criminal victimisation: lifetime prevalence, reporting to police and psychological impact. *Crime and Delinquency*, **33,** 479–489.
18. Boudreaux, E., Kilpatrick, D.G., Resnick, H.S., *et al.* (1998). Criminal victimization, posttraumatic stress disorder, and comorbid psychopathology among a community sample of women. *Journal of Traumatic Stress*, **11**(4), 665–78.
19. Norris, F.H. and Kaniasty, K. (1994). Psychological distress following criminal victimization in the general population: cross-sectional, longitudinal, and prospective analyses. *Journal of Consulting and Clinical Psychology*, **62**(1), 111–23.
20. Resnick, H.S., Acierno, R. and Kilpatrick, D.G. (1997). Health impact of interpersonal violence. 2: Medical and mental health outcomes. *Behavioral Medicine*, **23**(2), 65–78.
21. Shepherd, J.P. and Farrington, D.P. (1993). Assault as a public health problem: discussion paper. *Journal of the Royal Society of Medicine*, **86**(2), 89–92.
22. Murray-Parkes C. (1993). Psychiatric problems following bereavement by homicide. *British Journal of Psychiatry*, **162**, 49–54.
23. Mezey, G., Evans, C., Hobdell, K. (2002). Families of homicide victims: Psychiatric responses and help-seeking. *Psychology & Psychotherapy: Theory, Research & Practice*, **75**(1), 65–75.
24. Rock, P. (1998). After homicide: practical and political responess to breavement. Clarendon Press, Oxford.
25. Mezey, G., Evans, C. and Hobdell, K. (2002). Families of homicide victims: Psychiatric responses and help-seeking. *Psychology & Psychotherapy: Theory, Research & Practice*, **75**(1), 65–75.
26. Murphy, S.A., Braun, T., Tillery, L., *et al.* (1999). PTSD among bereaved parents following the violent deaths of their 12 to 28 year old children: A longitudinal prospective analysis. *Journal of Traumatic Stress*, **12,** 273–91.
27. Raphael, B. (1997). The Interaction of Trauma and Grief. In *Psychological Trauma - a Developmental Approach* (eds. D. Black, M. Newmay, J. Hains-Hendriks and G. Mezey). Gaskell, London.
28. Prigerson, H.G., Shear, M.K., Frank, E., *et al.* (1997). Traumatic grief: a case of loss-induced trauma. *American Journal of Psychiatry*, **154**(7), 1003–9.
29. Walby, S., Allen, J. (2004). Domestic Violence, Sexual Assault and Stalking: Findings from the British Crime Survey. London, Home Office Research, Development & Statistics Directorate.
30. Campbell, R., Wasco, S.M., Ahrens, C.E., *et al.* (2001). Preventing the 'second rape' - Rape survivors' experiences with community service providers. *Journal of Interpersonal Violence*, **16**(12), 1239–59.
31. Cluss, P.A., Boughton, J., Frank, E., *et al.* (1983). The Rape Victim - Psychological Correlates of Participation in the Legal Process. *Criminal Justice and Behavior*, **10**(3), 342–57.
32. Boudreaux, E., Kilpatrick, D.G., Resnick, H.S., *et al.* (1998). Criminal victimization, posttraumatic stress disorder, and comorbid psychopathology among a community sample of women. *Journal of Traumatic Stress*, **11**(4), 665–78.
33. Resnick, H.S., Kilpatrick, D.G., Dansky, B.S., *et al.* (1993). Prevalence of civilian trauma and posttraumatic stress disorder in a representative national sample of women. *Journal of Consulting and Clinical Psychology*, **61**(6), 984–91.
34. Burgess, A.W. and Holmstrom, L.L. (1974). Rape trauma syndrome. *American Journal of Psychiatry*, **131**(9), 981– 6.
35. Kilpatrick, D.G., Best, C.L., Veronen, L.J., *et al.* (1985). Mental health correlates of criminal victimization: a random community survey. *Journal of Consulting and Clinical Psychology*, **53**(6), 866–73.
36. Kilpatrick, D.G., Best, C.L., Veronen, L.J., *et al.* (1985). Mental health correlates of criminal victimization: a random community survey. *Journal of Consulting and Clinical Psychology*, **53**(6), 866–73.
37. Ellis, E.M., Atkeson, B.M. and Calhoun, K.S. (1981). An assessment of long-term reaction to rape. *Journal of Abnormal Psychology*, **90**(3), 263–6.
38. Atkeson, B.M., Calhoun, K.S., Resick, P.A., *et al.* (1982). Victims of rape: repeated assessment of depressive symptoms. *Journal of Consulting and Clinical Psychology*, **50**(1), 96–102.
39. Santiago, J.M., McCall-Perez, F., Gorcey, M., *et al.* (1985). Long-term psychological effects of rape in 35 rape victims. *American Journal of Psychiatry*, **142**(11), 1338–40.
40. Waigandt, A., Wallace, D.L., Phelps, L., *et al.* (1900). The Impact of Sexual Assault on Physical Health Status. *Journal of Traumatic Stress*, **3**(1), 93–101.
41. Coxell, A., King, M., Mezey, G., *et al.* (1999). Lifetime prevalence, characteristics, and associated problems of non-consensual sex in men: cross sectional survey. *British Medical Journal*, **318**(7187):846–50.
42. Campbell, R. and Ahrens, C.E. (1998). Innovative community services for rape victims: An application of multiple case study methodology. *American Journal of Community Psychology*, **26**(4), 537–71.
43. Maguire, M. (1982). Burglary in a dwelling. Heinemann, London.
44. Gale, J.A., Coupe, T. (2005). The Behavioural, Emotional and Psychological Effects of Street Robbery on Victims. *International Review of Victimology*, **12**(1), 1–22.
45. Kamphuis, J.H. and Emmelkamp, P.M. (1998). Crime related traumas: psychological distress in victims of bank robbery. *Journal of Anxiety Disorders*, **12**, 199–208.
46. Eisele, G.R., Watkins, J.P. and Matthews, K.O. (1998). Workplace violence at government sites. *American Journal of Industrial Medicine*, **33**(5), 485–92.
47. Budd, J.W., Arvey, R.D. and Lawless, P. (1996). Correlates and consequences of workplace violence. *Journal of Occupational Health Psychology*, **1**(2), 197–210.
48. Kopel, H. and Friedman, M. (1997). Posttraumatic symptoms in South African police exposed to violence. *Journal of Traumatic Stress*, **10**(2), 307–17.
49. Miller-Burke, J., Attridge, M. and Fass, P.M. (1999). Impact of traumatic events and organizational response. A study of bank robberies. *Journal of Occupational and Environmental Medicine*, **41**(2), 73–83.
50. Hewitt, J.B. and Levin, P.F. (1997).Violence in the workplace. *Annual Review of Nursing Resarch*, **15,** 81–99.

51. Williams, M.F. (1996). Violence and sexual harassment: impact on registered nurses in the workplace. *American Association of Hospital Nurses Journal*, **44**(2), 73–7.
52. O'Connell, P. and Bury, G. (1997). Assaults against general practitioners in Ireland. *Family Medicine*, **29**(5), 340–3.
53. Finkelhor, D. (1995). The victimization of children: a developmental perspective. *American Journal of Orthopsychiatry*, **65**(2), 177–93.
54. Mirrlees-Black, C. and Aye Maung, N. (1994). Fear of crime: findings from the 1992 British Crime Survey. Research findings No. 9. London, Home Office Research and Statistics Directorate.
55. Aye Maung, N. (1995). *Young people, victimisation and the police: summary findings*. HMSO, London.
56. Boney-McCoy, S., Finkelhor, D. (1995). Psychosocial sequelae of violent victimization in a national youth sample. *Journal of Consulting and Clinical Psychology*, **63**(5), 726–36.
57. Herek, G.M., Gillis, J.R. and Cogan, J.C. (1999). Psychological sequelae of hatecrime victimization among lesbian, gay, and bisexual adults. *Journal of Consulting and Clinical Psychology*, **67**(6), 945–51.
58. Rose, S.M. and Mechanic, M.B. (2002). Psychological Distress. Crime features and help seeking behaviours related to homophobic bias incidents. *American Behavioural Scientist*, **46**(1), 14–26.
59. Clark, I. and Leven, T. (2002). The 2000 Scottish Crime Survey: analysis of the ethnic minority booster sample. Edinburgh, UK, Scottish Executive Central Research Unit.
60. Fitzgerald, M. and Hale, C. (1996). Ethnic minorities: victimization and racial harasssment. In: Home Office Research and Statistics Directorate, editor. *Home Office Research Study*, **154**. London.
61. Whalley, M.G., Brewin, C.R. (2007). Mental health following terrorist attacks. *British Journal of Psychiatry*, **190**, 94–6.
62. Abenheim, L., Dab, W., Salmis, L.R. (1992). Study of civilian victims of terrorist attacks. *Journal of Clinical Epidemiology*, **45**, 103–9.
63. Desivilya, H.S., Gal, R. and Ayalon, O. (1996). Extent of victimization, traumatic stress symptoms, and adjustment of terrorist assault survivors: a longterm follow-up. *Journal of Traumatic Stress*, **9**(4), 881–9.
64. Trappler, B., and Friedman, S. (1996). Posttraumatic stress disorder in survivors of the Brooklyn Bridge shooting. *American Journal of Psychiatry*, **153**(5), 705–7.
65. van der Ploeg, H.M. and Kliejn, W.C. (1989). Being held hostage in The Netherlands: a study of the long-term after effects. *Journal of Traumatic Stress*, **1,** 153–9.
66. Shepherd, J.P. and Bisson, J.I. (2004). Towards integrated health care: a model for assault victims. *British Journal of Psychiatry*, **184**, 3–4.
67. Ehlers, A. and Clark, (2004). Review of the impact of early CBT programmes in the treatment of trauma. *Biological Psychiatry*, **3**(9), 817–26.

11.14

Assessing and managing the risks of violence towards others

Paul E. Mullen and James R. P. Ogloff

'Prediction is very difficult, especially about the future'

Niels Bohr (1885–1962)

Introduction

Assessing and managing the risk of our patients being violent towards others now occupies a prominent position in virtually all forms of mental health practice, but it remains a contentious area. At the highest level researchers, psychometricians, and statisticians argue about almost every aspect, even whether anything useful can be said about individual outcomes rather than group indicators. At the next level an industry flourishes of selling training, and risk assessment tinstruments, to those who then appear as experts in a wide range of mental health and criminal justice contexts. On the ground, almost everyone in mental health is drawn into filling out purpose-designed forms and complying with protocols, most of little or no demonstrated validity. This chapter is intended to make clinicians aware of both the possibilities and limitations of existing approaches to the assessments of risk. Given that there is no reason for mental health professionals to evaluate risk without gaining information to manage it, this chapter will also address the management of risk for aggression and violence.

Constructing risk

A critical analysis will be attempted of how risk has come to be constructed in our society and how this is impacting on mental health and criminal justice. When an approach is adopted which attempts to reveal the foundations and historical evolution of a widely accepted social construct, like risk, there is a danger that it will appear overly sceptical or even mocking. It is important to emphasize at the outset that:

1 The assessment of the probability of patients behaving in ways damaging to others and the management of that risk is a legitimate clinical activity.

2 That attributions of levels of risk to a patient occurs in a social and cultural context and is inescapably a construct.

The discourses around the dangerousness of the mentally ill have gradually been replaced by those of risk. This change is usually presented as a product of the progress of knowledge and the improved conceptualization of that knowledge.[1–4] The language of danger transmuted into the language of risk also emphasizes the probabilistic nature of risk assessment. We are not now and probably never will be in a position to be able to determine with certainty who will or will not engage in a violent act. Relying on a range of empirically supported risk factors, though, we can makes a reasoned determination of the extent to which those we are assessing share factors that have been found in others to relate to an increased level of risk. Risk embodies the interaction of a range of factors, which are not necessarily dangerous in themselves, such as age, gender, marital status, ethnicity, employment status and, of course, mental disorder.[5] Risk factors can be any variables which are statistically associated with a future violent episode or event. There is no assumption of causality linking the predictor to the predicted.

Risk assessment came relatively late to the mental health field. Not until 'harm to others' or 'undue risk' became criteria for involuntary hospitalization and forensic detention did 'dangerousness' of patients assume the spotlight. The focus on risk first surfaced in Western Societies in the 1970s in the context of concerns about damage being inflicted on individuals by the actions, or inactions, of corporate and governmental agencies. These concerns fed the emergence of widely based environmental movements as well as an escalating number of class actions and individually driven litigations. Under the banner of risk a new blaming system emerged of which Douglas[6] writes 'we are . . . almost ready to treat every death as chargeable to someone's account, every accident as caused by someone's criminal negligence, every sickness a threatened prosecution. Whose fault? is the first question . . . then what damages? what compensation? what restitution . . .' (pp. 15–16). One response to this culture of blame has been the emergence of what O'Malley[7] refers to as a new prudentialism in which individuals, professionals and corporations, increasingly held responsibility for the impact of their actions on others, resort to risk management strategies in which risk is assessed, managed, insured against, and where possible removed.

Psychiatrists and psychologists are among those who have become caught up in the 'culture of blame'. Any damaging or distressing occurrence which is experienced by, or caused by, someone who is, or has been, a patient of the mental health services, is transformed into a preventable tragedy for which professionals are

to be held responsible. Rose[8] suggests the new imperatives of risk assessment and risk management operate to establish mechanisms to control mental health professionals which through standards, audits and enquiries not only regulate professionals but hold them personally responsible for unwanted outcomes. Douglas (6) argues 'probability analysis arrives at politics in the form of a word 'risk' . . . the word gets its connection with probability squeezed out of it and put to the same primitive political uses as any term for 'danger" (p. 48). Risk assessment and risk management are concepts which have the potential to shift blame towards clinicians who have failed to follow procedure and away from managers who fulfilled their responsibilities by ensuring correct protocols were in place, irrespective of the possibility of the realistic application of such protocols. The language of risk can also shift the focus from politicians who determine resources and establish systems of care to those who fail to identify and manage risk in the individual case. Perhaps most importantly the increasing centrality of risk assessment potentially creates a vision of the mentally disordered as primarily embodiments of varying degrees of risk and the mental health services as agents in controlling and obviating the supposed danger to the community.

Assessing dangerousness used to be the almost exclusive province of the forensic mental health professional.[9,10] It was a marginal activity based on arcane knowledge and assumed wisdom that only experience could provide.[11,12] Risk assessment and management in contrast have become central to current mental health practice in almost all its guises. It has become among the most important activities defining professional competence. Understanding the cultural, legal, and political roots of the increasing hegemony exerted by the rhetoric of risk over psychiatric practice may demystify, but does not free the professional from the imperatives of operating effectively in this new environment. It seems so obvious, as to be self evident, that mental health professionals are expected to consider the probability that their patient will act in a destructive manner and to act to prevent such harms. But the self evident is often the ideological unconsidered. It is not obvious that a mental health professional's primary responsibility is to the wider community rather than their patient. It is not obvious that it is possible to effectively predict such risks as they apply in the individual case. It is not at all obvious how we should act in the face of a prediction of risk, and it is certainly not obvious that such concerns should be a major determinant of our approach to patients.

Words are rarely innocent. Risk is not the same as probability, for risk implies a degree of danger. Even those deemed to be at 'low risk' still appear to present some degree of danger. No one, it seems, is considered risk free. Risk management is not the same as harm minimization, for it promises a prevention of unwanted outcomes. Psychiatry deals with disorders which have both substantial morbidities and mortalities. Good management may reduce but cannot, in our present state of knowledge, prevent all such morbidity and mortality. Furthermore, reducing morbidity and mortality long term may only be possible at the price of accepting an increased probability of mortality in the short term. Suicide in prison, for example, can be prevented by isolating and observing vulnerable inmates in transparent plastic bubbles, bereft of features from which suspension is possible, or by the simple expedient of chaining them hand and foot to a bed (both strategies are in use today). If the only good is preventing self harm such draconian measures acquire currency irrespective of the psychological damage and abuse of basic human dignity involved. Moreover, as the aim has become removing all risk of suicide in prison, more and more vulnerable prisoners are being subjected to such restrictions and in many jurisdictions 'witch hunts' now regularly follow every death across jurisdictions. We cannot force the genie of risk back into the lamp. Mental health professionals will continue to be made publicly accountable. Professional self regulation is being replaced by statutory regulation and the ravages of civil litigation.[13] As Rose[8] notes we will be forced to play a central role 'in the strategy of reducing risk and minimizing harm under threat of sanction and within the disciplines imposed by a plethora of practices of blame' (p. 18).

Contemporary approaches to risk assessment

It is no longer possible for mental health professionals to distance themselves from the process of risk assessment. Throughout the 1970s and 1980s mental health professionals were almost of one voice in proclaiming both their inability to predict dangerousness and the basic pacifivity of the mentally disordered.[14] Despite this public stance dangerousness criteria came to dominate civil commitment, with courts simply ignoring arguments that the prediction of dangerousness was beyond the ken of psychiatrists and psychologists.[15] In addition liability based on failures to predict dangerousness was established in landmark cases including those arising out of Poddar's killing of Tarasoff[16] and Hinkley's attempt to assassinate President Reagan.[17] Last, but not least, the criminal courts increasingly suborned mental health professionals into predicting dangerousness in the pursuit of such sentences as preventive detention and death.[15,18] In the US and the UK the emergence of the language of risk at the end of the 1980's was in part a recognition of the centrality assumed by predicting and managing the potential for violence perceived by the public to reside in the mentally disordered. In the UK the public inquiry into the killing of Jonathan Zito by the psychiatrically disordered Christopher Clunis set a pattern for future homicide enquires.[19] Failures of risk assessment and management by individual practitioners, together with inadequacies of communication and service provision, were identified as major contributors to 'avoidable' killings.[19] Such developments set risk assessment at the very centre of the mental health agenda.

Given the challenges of risk assessment, what is the current state of knowledge and can it be of assistance to clinicians? The limited research available on 'dangerousness prediction' conducted in the 1970s showed that psychiatrists and psychologists had unacceptably low levels of accuracy in predicting which patients would go on to be violent in the future.[20] It was found, perhaps not surprisingly, that psychiatrists, psychologists, and release decision-makers tended to make conservative decisions that suggested that people were at risk for dangerousness or violence when, in fact, they were not. Similar findings have been obtained over time. For example, Belfrage[21] found that clinicians found 90 per cent of a group of 640 offenders sentenced to psychiatric treatment in Sweden to be at 'risk of severe criminality;' when, in fact, only 50 per cent went on to commit any kind of crime.

Reasons given for the errors made in risk prediction include clinicians' confusion and lack of knowledge of valid risk markers

and risk factors. In addition, as with other areas of decision-making, even if clinicians do have a reasonable understanding of risk factors, it is difficult to systematically consider them and to put together a risk appraisal in a systematic way. Advances in risk assessment have included the identification of an expanded range of predictor variables relevant to violence. Most important among these are those variables that are subject to change (i.e., they can change over time and they can be influenced by treatment or other intervention). Generally speaking, risk assessment variables can be classed as 'static' (i.e., those that cannot be changed) and 'dynamic' (i.e., those that can change over time). Actuarial risk schemes, which will be discussed later in this chapter, are based upon variables that were measured from the past. These historic variables generally could not change over time. For example, if one began being violent as a young person, that fact will not change over time. Dynamic variables, in contrast, are subject to change over time, sometimes rapidly. These variables include such things as state of mind, situational factors, attitudes, plans, support, etc. Effective risk assessment must take into account both static and dynamic variables; however, risk management generally requires an understanding of the dynamic risk variables. Contemporary approaches to risk assessment and management take into account both static and dynamic variables, thus considering an individual's past, present, and future risk factors that might affect the likelihood of him or her becoming violent.

There has been considerable progress since Monahan[20] first reported that psychologists and psychiatrists were essentially unable to predict risk to any acceptable extent. Current research shows that risk assessment approaches provide a level of accuracy that now far exceeds chance.[22]

The limits on mental health professionals' engagement in risk assessment

There are two legitimate perspectives on risk assessment:-

1 The clinician whose work involves considering patients' levels of risk as part of a process whose purpose is primarily to improve the management of patients.

2 The forensic evaluator for whom risk assessment is a tool to improve the reliability of opinions provided to courts and tribunals charged with making decisions about an individual.

Forensic mental health professionals in the US and Canada tend to see their role primarily as evaluator. It is not uncommon for forensic psychiatrists and psychologists to primarily conduct assessments for the courts as their means of employment. This reflects the generally accepted separation of the assessment and treatment roles. In stark contrast in the UK and Northern Europe the clinician perspective dominates even for forensic specialists. Relatively few forensic clinicians would operate as court evaluators alone, though some like their US colleagues would try and separate the roles.

This chapter is written for mental health professionals in a wide range of contexts, not for forensic specialists. As a result the emphasis on this chapter will be on the clinical perspective. North American readers should, however, keep in mind that this perspective is not shared by many specialists in the forensic field there who would consider it not just problematic but ethically questionable to mix the assessment and treatment roles.

Boundaries need to be drawn around when, where and for what purpose, mental health professionals can ethically engage in assessing the probability of an individual committing violent or criminal acts. There are somewhat different constraints operating on clinicians than for those carrying out evaluations for the purpose of preparing reports for decision-making bodies like courts and tribunals.

The ethical and practical constraints on a clinician assessing risk entirely in service of effective treatment include:

1 Ensuring the assessment serves the interests of the patient in terms of improving management and protecting them from acts which will damage their interests. The seriously mentally ill when they become violent all too often target those who support and care for them thus destroying the relationships critical to their own social survival. In addition, criminal and violent acts of the sort usually associated with the mental disordered rarely, if ever, brings anything but increased problems for the patient. Thus, reducing the likelihood that the patient will engage in criminal and violent acts serves not only the public interest but also the patient's interests.

2 Mental health variables (which include psychological variables and personality traits) are a prominent feature of the individual's clinical picture and are also of potential relevance to the probability of future damaging behaviours.

3 Avoiding providing greater emphasis to risk than is necessary given the totality of the patients needs and vulnerabilities.

4 The assessment should wherever possible connect potential risk to those factors whose amelioration will reduce that risk. Ultimately a health professional can only justify engagement in risk evaluation if they lead to better outcomes for the evaluated as well as the community. Risk assessment finds its ultimate justification in risk management.

5 Avoiding using risk and risk assessment to disqualify patients from access to the treatments they require. Increased risk requires increased therapeutic enthusiasm, not rejection.[23]

6 That any concerns raised by a risk assessment are shared with the patient and the proposed management strategies explained. Even unwelcome restrictions are resented less if the reason for their imposition is explained.

7 Retaining an awareness of the limitations of the predictive power of risk assessments and the need to ensure proportionality between the risk actually apprehended and any imposed remedy.

8 Ensuring a level of professional competence adequate to the task.

9 Making use, where possible, of the skills and knowledge of the multidisciplinary team in the assessment.

Those obligations on those engaging in risk assessment as part of an evaluation for a decision-making body are even more onerous and should include:-

1 The patient consents to the examination in the knowledge of the nature of the assessment, the purposes to which it may be proffered, and the limitations on confidentiality that may apply.

2 A reasonable body of empirical evidence exists to guide the risk assessment including, where possible, empirically validated structured risk assessment measures.

3 The risk assessment is conducted in consideration of the legal parameters governing the decision-making body (e.g., criteria to be considered for change of orders or release for forensic psychiatric patients) while realizing that the legal questions to be answered never parallel the clinical or evaluative results mental health professionals can reasonably provide (e.g., there is no clinical parallel to legal criteria such as 'undue risk').

4 The assessment is based on a careful analysis of the relevant characteristics of the particular individual which in all but exceptional circumstances have been obtained in part by a direct examination of the individual.

5 The risks are expressed in terms of probabilities (not attributions of dangerousness) with clear admissions of the fallibility and potential variability in the prediction. The problem should be acknowledged of employing risk factors derived from studies on populations from different cultures and contexts. After all nobody would use risk data from Los Angeles to calculate the car insurance premium for a driver in Dublin or Oslo without considerable caution.

6 Account is taken not just of the probability of damaging behaviour but the nature and severity of such conduct. Proportionality needs to be maintained between what is predicted and the response. It is all too easy to employ methods which establish increased risks of a wide range of unwanted behaviours only to find them used to justify draconian and punitive responses which would only be acceptable in the face of the imminence of serious violence.

7 The confidence and certainty with which any prediction is formulated to take account of the implications for the person being assessed. Risk predictions may be offered in terms of probabilities but they will almost always be used to justify all or nothing decisions.

8 That personal and professional integrity is strictly maintained. This is no simple matter when the evaluator is either in the pay of the patient, or of those whose interests are not necessarily those of the patient.

The potential conflicts generated when acting for a patient as both a professional evaluator and a clinician are so considerable that some experts argue that such a dual role is inherently unethical. They argue in preparing reports for decision-making bodies the interaction ceases to be that of health professional and patient, and becomes entirely that of expert, or 'forensisist', and evaluated. The expert's obligation is then not to the evaluated but to their own professional competence and the rules governing the process the report will serve (e.g., criminal court, family court, mental health tribunal).

In our view even in an encounter between a health professional and a person purely for the purposes of an evaluation for a court there remains obligations to the person as patient [patient from the Latin for to suffer, who in this instance suffers the intrusions of an examiner who cannot for the evaluated entirely cast off the guise of physician or healer]. The solution is to learn to live with the contradictions and accept the dialectic between responsibilities to the patient and obligations to the agencies of society. The result of accepting the duality of the role inherent in assessing a patient's risk of harming others does exclude participation in a process that could increase the risk to the patient of fatal (e.g., death penalty evaluations) or serious harm (e.g., sexual predator laws whose sole purpose is justify a process of prolonging incarceration beyond the expiry of a sentence). It does however legitimate having the dual role of clinician and evaluator in some circumstances. In fact those who totally eschew ever taking on such a dual role, we believe, are at risk of deluding themselves that they can caste off the mantle of clinician for the patient, and become a socially neutral, objective, observer and reporter.

The dialectic between the demands of a professional responsible to the health of your patient and the demands of professional integrity and honesty owed courts and tribunals is almost always possible to resolve. To fail to accept engagement with the conflicts which in reality usually exist for evaluators and clinicians reporting to external authorities is in our opinion an act of self deception in which you become an agent either of patient, or authority, and no longer an autonomous responsible professional.

Some situations make nonsense of the above considerations, notably death penalty hearings. One of us (J.O) has extensive experience of working with those on death row, the other (P.M) only a comparatively slight acquaintance. We are both of the opinion, however, that it is impossible to honestly discharge your responsibilities as clinician, as evaluator, or as decent human being, in such circumstances.

Risk assessment approaches

There are five basic approaches to evaluating the risk of violence:-

1 Probability models based on established risk factors. The risk factors can be derived actuarially from studies of particular populations (e.g. Violence Risk Appraisal Guide (VRAG) and Static 99) or rationally ascertained from the risk literature (e.g. Historical Clinical Risk 20 (HCR-20)).

2 Clinical experience based on recognizing previously encountered (personally or in the literature) patterns associated with future violence. The clinical approach is largely relevant to the avoidance of obvious errors like discharging morbidly jealous men who are threatening to kill their partner.

3 A mixture of 1 and 2 where the risk assessment instrument is employed to guide the appraisal of risk factors and clinical judgment is applied to balance idiographic information with the nomothetic variables as in the structured professional judgment approach of which the HCR-20 is the prime example.[22]

4 The strictly idiographic approach which employs individual profiles of violent offenders to detect those on a similar pathway to attack. The idiographic approach is employed to evaluate the risks of rare events, such as attempts to assassinate a head of state, and has little application in general mental health.

5 A plethora of local risk assessment tools have sprung up. Sometimes it seems as if every psychiatric service, probation/parole service, prison, and security consultant have their own unique sheet of questions which are supposed to establish the future probability of whatever particular piece of nastiness currently concerns the organization. These ad hoc parochial risk assessment protocols have no evidentiary basis or psychometric integrity (even if they incorporate aspects of other properly constituted instruments). It is far better to validate existing

empirically supported measures for use in a particular setting and with a particular population.

The core of risk assessment is the systematic application of probability models usually incorporated in standardized instruments.

The utility of risk assessment

Even given a firmly based knowledge of those factors which in populations may increase or decrease the probability of violent behaviour, there remain theoretical and practical limitations on effective prediction in the individual case. If factors were identified which occurred only in the violence prone and never in the pacific and, if present, were in every case the harbinger of future attack, then the power of the predictive paradigm would be independent of both the frequency of future violence in the population, and of variables which effect the factors' expression in the individual case. In the real world the sensitivity (the accuracy with which the outcome is predicted) and the specificity (the extent to which only those who will act in the predicted way are identified) fall short of 100 per cent. This being so the less common the future behaviour in the population the less specific will become predictions. Equally, the more complex the influences affecting the expression of the identified predisposition the less the sensitivity of the predictive paradigm.

Say we develop a predictive paradigm of 70 per cent sensitivity and 95 per cent specificity (which is feasible). If we set an acceptable level for the practical use of such an instrument that it will not unfairly restrict, or stigmatize, more than 1 person for every 4 correctly designated as candidates for future violence, then the base rate for violence in the population of interest would have to exceed 20 per cent. Even if we accepted one error for each correct designation (which, if the outcome of ascertainment were incarceration or other significant curtailment of basic freedoms, few would regard as ethically defensible) it still requires a base rate of higher than 6 per cent.

This hopefully makes clear that if predictions of the probability of future violence are to be used to significantly restrict the patient's freedoms the base rate of the behaviour must be reasonably high in the group under consideration. Equally, if measures such as long term institutionalization or compulsory community treatment with restrictions on residency and movement are contemplated the degree of violence apprehended must be commensurately damaging. It would be difficult to justify such interventions if what is at issue is embarrassing, or even fear inducing behaviour, which does not involve either assault occasioning injury or gross intimidation.

Statistical approaches to analysing predictive efficiency in populations with low base rates for the target behaviour exist, the most commonly employed being derived from a measure developed for use with radar or signal detection systems, the Receiver Operating Characteristic (ROC) curve.[24, 25] The ROC curve is a graph of true positives (sensitivity) along the y-axis and false positives (1-specifity) along the x-axis. With respect to violence, these correspond to patients who were predicted to be violent those who were predicted not to be violent, respectively. The line running from the lower left corner of the graph to the top right indicates chance prediction, where true positives are equal to false positives (i.e., for every patient predicted to be a violent who in fact becomes violent another patient who was predicted to be violent does not become violent). A curve above this line indicates that, in this case, recidivism is being predicted at rates above chance. The Area Under the Curve (AUC), which is represented as a proportion of the graph that falls under the curve, reflects the proportion of true positives over false positive (e.g. an AUC of 70 would indicate that 70 per cent of those predicted to be violent in fact do become violent). The ROC analyses allow the accuracy of predictions to be established independent of variations in base rate.

In many western nations between 5 per cent and 10 per cent of all homicides and more than 5 per cent of serious crimes of violence are committed by those with a schizophrenic syndrome. The annual risks of a person with a schizophrenic syndrome committing a homicide is however, in the region of 1 in 10,000 and for a crime of violence about 1 in 150.[26] This is because serious violence is, the media notwithstanding, an uncommon event. Though fear inducing behaviours occur with distressing frequency among the seriously mentally disordered the inflicting of serious injury is measured in annual risks of below 1 per cent.[27,28,29] This suggests that risk assessment instruments will not be relevant to predicting serious violence in those with a schizophrenic syndrome. Underlying the homicide enquiries in the UK, and the litigation, particularly in the US, however is the assumption that they can be and will be.

Given their particularly low base-rate of occurrence, attempting to predict who will commit serious acts of violence or murder will inevitably be accompanied by vast numbers of false accusations. Further, in reality we often trade off outcome variables, thus avoiding over ready resort to civil commitment may improve the chances of establishing a long term therapeutic alliance, which in turn may reduce long term risks, albeit at the price of tolerating a higher degree of risk in the short term. Avoiding all probability of any patient committing a future act of violence would involve the use of widespread coercion and move mental health professionals into increasingly custodial and controlling roles.

An alternative argument deserves consideration. In those with a schizophrenic syndrome, for example, it may well be feasible to identify the 10 per cent who will perpetrate 90 per cent of all the future fear inducing and violent acts. In this 10 per cent may be included nearly all of the far smaller number who will commit potentially lethal or seriously injurious acts. In effectively identifying the 10 per cent and managing them appropriately then the risks to the community of damage, including the small chance of serious damage, will be reduced. The majority of those so identified will be a nuisance who can occasion fear, and who may push, punch or kick others.[30] Effectively identifying and managing all patients in the high risk group will lower the overall risk to the community while minimizing the deprivation to liberty of those in the low risk group. This risk group management approach is not perfect and does not increase the ability to identify a particular individual who may commit a heinous act, but it does allow effective management of those at higher risk.

The management ethically and pragmatically justified, must retain some semblance of proportionality between the apprehended insult and the impact of the proposed preventive strategy. In practice this obliges us in most clinical situations not to resort to increased coercion, let alone preventive detention, but to focus attention on

greater support and more active follow up and treatment in the community with a more ready resort to admission during exacerbations of symptoms or social conflict.

Risk assessment instruments

The last 15 years or so has been marked by a wave of enthusiastic advocacy specifically for the benefits of so called actuarial risk assessment instruments (e.g.,[31,32,33]). The advocates of actuarial risk assessment claim directly, or by implication, to be able to identify the likelihood that specific individuals will progress to various forms of interpersonal violence. These risk assessment tools are based mainly on retrospective, though occasionally prospective, studies of specific populations, such as discharged patients and released offenders. Actuarial approaches have the advantages of:-

1 Multiple variables delineating level of risk.

2 Designed to move from group data to individual attribution.

3 Realized in simple objective reproducible rating scales which minimize individual clinician's discretion and therefore responsibility.

4 Focus attention on 'high risk' individuals (principle of targeted resources).

5 Provides protection to clinicians and managers in event of disaster. Nobody can be blamed for the failures of 'science'.

Actuarial approaches are not without their problems. The results, for example of commonly used risk assessment instruments like the VRAG and the Static 99 change little if at all over time and with circumstances. Clearly whatever the Static 99 may indicate the risks of committing further rapes in a fit 25 year old is unlikely to remain the same as he ages and acquires disabilities.[34] Actuarial approaches also almost inevitably revolve around a limited number of variables which exclude uncommon though potential critical factors. Thus, for example, morbid jealousy is not sufficiently common to emerge as an actuarially established risk factor despite studies on such cases indicating a very high probability of significant violence. The structured clinical judgement approach allows the incorporation of such potential modifiers[22]. Actuarial instruments are developed on specific samples constituted in particular places at particular times. This can lead to problems with generalisability and equally important idiosyncratic and false attributions. For example, in both the VRAG[31] and the Classification of Violence Risk (COVR)[35] the schizophrenic syndrome emerges as a protective, or at best a neutral, factor with regard to the risk of violence, despite being associated in many other studies with far higher rates of violence than occurs in the general population.[26] This is because in the VRAG the rate in those with schizophrenia was compared to other offenders, some of whom had severe personality disorders, and as expected it was lower. In the MacArthur study, from which derives the Classification of Violence Risk (COVR),[35] substance abuse was treated as an independent confounder and the rates for the select few who had a schizophrenic syndrome, but were not substance abusers, were used to calculate the level of risk for schizophrenia.

The use of the term actuarial links these approaches to the well established actuarial methods familiar from the insurance industry. In the insurance industry actuaries usually generate their risk groups on the basis of samples numbered in the thousands, with the occasional in the tens of thousands. In the mental health fields the samples are usually measured in hundreds, with the occasional topping the thousand mark.[1] Actuaries in the insurance industry work exclusively with group based predictions. Your car insurance will depend on such variables as the type of car you drive, your age, your gender, your prior driving record, and where you live. That determines the group you fall into for the purposes of costing the policy you request. The actuary is not interested in what happens to each individual only whether the whole group for whom policies have been drawn costs sufficiently less in claims than is received in policy payments to produce the required profit margin. The actuarial method is not designed to assign levels of risk to individuals but to groups, though the cost of an individual's policy will be determined by the risk group into which their policy falls. The precision of the group prediction is determined by the size of the sample and the frequency with which the event of concern occurs. The rarer the outcome of interest (e.g., murder or plane crashes) the larger has to be the sample, and to some extent the commoner the event (minor assaults or fender benders) the smaller the sample. Irrespective of the sample on which the risks have been established as you try to make finer and finer distinctions involving smaller and smaller subdivisions within the original sample so the confidence that can be placed in the estimate decreases. Paradoxically the more intense and detailed the analysis of the original group the less reliably can the derived risk algorithm be applied to those outside the group.[19] The MacArthur study on which the COVR is based exemplifies this problem. The smallest unit is obviously the individual group member and here the inherent variability of the risk prediction will be at its highest.

To take the example of two of the most widely used and best established actuarial risk instruments. The VRAG claims to identify nine groups, known by the unfortunate term 'bins', with a probability of future violent recidivism varying from 0 per cent to 100 per cent.[31, 32] The analysis of Hart and colleagues[36] demonstrates that rather than nine statistically separable groups there are only three statistically distinct groups. The Static 99 claims to separate sex offenders into 7 groups with recidivism rates varying from less than 10 per cent to greater than 50 per cent.[37] Statistically however only two separable groups are generated one with offending rates between 4 per cent and 25 per cent and one between 30 per cent and 60 per cent.[36] In short even ignoring the all important problem of attributing a group risk to an individual within the grouping the Static 99 identifies recidivism with at best 2 in 3, and at worst 1 in 3, chance of accuracy. In this really good enough to damn a person to indefinite incarceration or extended imprisonment?

Unfortunately, unlike other areas that rely on actuarial approaches to decision-making, there has not been a concerted effort in the area of violence risk assessment to pool the results of various studies to obtain the large samples necessary to reduce the variance which would thereby reduce the broad values on the confidence intervals.

[1] There are notable exceptions, such as the Level of Supervision/Case Management Inventory which has accrued normative data on more than 35,000 prisons and almost 80,000 people under community corrections supervision [Andrews, D. A., Bonta, J. L., & Wormith, J. S. (2005). Level of Service/Case Management Inventory (LS/CMI). Toronto: Multihealth Systems.].

As noted in Footnote 1 above, there are some exceptions such as the LS/CMI. Doubtless, over time researchers will pool results to determine the extent to which confidence intervals can decrease, thereby increasing the predictive utility of instruments.

The big question which hangs over the use of risk assessment instruments is the extent to which it is possible, or acceptable, to make attributions about an individual's future behaviour on the basis of their sharing characteristics with those in a group with a known level of risk for such behaviour. In medicine we are so used to using probabilities to dictate our actions that there are dangers in failing to recognize the problems. An 18 year old woman presents with a history of severe central abdominal pain moving to the right iliac fossa associated with anorexia. The probability in those with similar symptoms of an inflamed appendix may be 30 per cent but 100 per cent of the patients are advised to have surgery. Probability is used here to provide advice entirely in the interest of the patient. If the advice is refused the patient is not forced into surgery but is observed and managed non-surgically. Compare this with advice to a court or tribunal considering extending the incarceration of an offender or patient. The offender may have similar characteristics to those with a better than 50 per cent (or even 80 per cent) chance of re-offending and to attribute the group risk to the individual may benefit the community. It will however almost certainly disadvantage the individual who will not be the one accepting or rejecting the advice. It is fallacious to argue that making such attributions from group membership to individual risk is acceptable because it reduces the errors of false positives. It is not possible to benefit the so called low risk without disadvantaging the rest. Even with particularly strong measure of risk assessment, the false positive rates still hover around 25 per cent.

The only strong defence of attributing risk to an individual by virtue of their group membership, which is the essence of prejudice, is that it has better outcomes than not using this approach. But what are better outcomes? Better outcomes could be reduced to the accounting of true and false attribution. On this basis actuarial risk assessment instruments perform better than the best guesses of most experts. In the case of possible appendicitis given the benefits of early versus delayed intervention a high number of false positives is tolerable to avoid even one false negative. When the outcome is imposing further incarceration even the price of one false positive for every true positive, which the Static 99 and VRAG might offer may raise moral and legal qualms. Particularly as given both the true and the false positive will be incapacitated making it impossible to ever know which was which. There is no quick corrective feedback in the world of assessing the risk of future criminality, unlike appendicitis where the pathologist will soon tell you if you were right or wrong. Further when supposedly low risk individuals are discharged or released and re-offend, this will reinforce any tendency to err on the side of caution and incarceration.

Statistical decision theory can support using group membership to attribute levels of probability to individual group members. Though even here sceptics might suggest that the efforts put into Bayesian approaches to make the best of limited data sets might better be expended in enlarging such data bases. What in any case the approach cannot do is decide on utility and moral propriety. It is only relatively recently that the criminal justice system has moved to making the prediction of future offending the dominant issue in determining sentences, parole, and the extensions of sentences (which in and of itself is an entirely new phenomenon). Similarly the emphasis on the risks psychiatric patients present to others is a relatively recent preoccupation for mental health services. The development of standardized risk assessment instruments not only serves these changes but sustains them. The sexual predator legislation in the US, the DSPD and indefinite sentencing provisions in the UK, and the extended supervision/imprisonment laws in Australasia, are all dependent on mental health professionals providing them a veneer of science and objectivity through risk assessments.[38, 39] While providing the courts with information regarding the extent to which the individual's characteristics place them in a general level of risk we by implication give our assent and support to that process. We place a hope for community benefit before an inevitable disadvantage to those with whom we have engaged professionally.

The current state of the science of risk prediction some might suggest delivers only a limited improvement in the decision making process. If they are correct what, if anything, remains in the risk assessment literature of practical value? The short answer is that the establishing of risk factors remains of inestimable value, not in placing labels on individuals but in identifying how to reduce the probability of future violent and criminal behaviours. The standardized risk assessment approaches such as the HCR-20 and the VRAG allow individuals to be assigned to broad levels of high, low or medium probability of future violence. It can be argued whether or not the confidence that can be placed in such attributions justify imposing extended incarceration. They are however of undoubted value in assigning priority for management interventions.

The structured professional judgement measures, such as the HCR-20, also add considerable value by identifying risk factors, such as substance abuse, specific personality vulnerabilities, and the like allowing targeted interventions to reduce the chances of engaging in violence and offending. The exposure of the current limitations in the project of identifying dangerous individuals is bad news for courts, parole boards and governments looking for short cuts to community safety. But for mental health professionals it is merely a reminder that our business is managing patients to reduce the risks for them and for others, not trying to separate the dangerous goats from the mostly harmless sheep.

The limitations on the ability to accurately predict risk also emphasize that in all circumstances where psychiatrists and psychologists are being asked to provide opinions of risk of future violence to courts or other decision-making bodies, they should provide the courts with the information that they can. This would include, for example, the general risk level in which the risk assessment tools would place the patient but it would also extend to an anamnestic consideration of factors in the patient's case that are known to increase or decrease the level of risk. In the end, though, the clinicians need to be clear to state the limitations of their findings and to, of course, leave the ultimate decision of whether the patient meets the criteria of the statute in question to the legal decision makers.

Practicalities of risk prediction and management

Risk assessment and management can be conceptualized as a four stage process:

1 An evaluation of general level of risk and priority for active intervention. In practice this can rarely go further than saying they are in a high, medium, or low, priority or risk category.

2 Identify current risk factors and future hazards which are both potentially remediable and causally relate to increased chances of violence.

3 Develop management strategies to reduce or remove the deleterious influence of these factors.

4 Evaluate the effectiveness of the interventions in reducing subsequent violence.

Each of the above steps will be discussed in turn below.

Evaluating level of risk

Such evaluations should start simple, consider the context, apply clinical and common sense, then potentially progress to the use of a standardized assessment instrument. In both the mentally disordered and general population the majority of violent and criminal behaviour is committed by young males with histories of repeated antisocial behaviour dating back to conduct disorder in childhood. This is often combined with anomie, that lawless disaffection with society in general and contempt for rules and authority in particular. Impulsiveness is concerning[40] but so is the more muted feckless disregard for consequences found in some with a schizophrenic syndrome. Unemployment, living in a high crime neighbourhood, and antisocial peers all add substantially to risk. Similarly substance abuse is a robust maker for the risk of criminality in the disordered and non disordered alike. Specific to the seriously mentally disordered is a refusal to recognize they are ill, and resistance to complying with treatment with often an antagonism to professionals; the 'fuck off and leave me alone' syndrome.

Clinicians must maintain their common sense. A range of situations must be taken seriously. The angry and threatening who can tell you exactly what they plan to do to their supposed antagonist, the frightened who see no alternative to a pre-emptive strike, those who have prepared for violence (weapons, surveillance, put their affairs in order), the actively suicidal who have nothing to lose but still care enough to take a final revenge, and commonest of all those making threats in a manner which creates fear and concern either in the potential victim or those who have been privy to the threats.

Beyond normative data and risk factors, there exist a number of high risk situations and syndromes particular to psychiatry. These include:-

1 Morbid jealousy

2 Some misidentification syndromes

3 Depressed suicidal mothers of young children.

4 Delusional systems focussing on specific individuals believed to present for the patient a serious threat, or a malevolent impediment to their central project.

5 Some stalking situations

6 Confusional states be they toxic or related to dementia or other cerebral impairments (this is the source of a significant proportion of assaults on health staff)

Patients presenting any of the above characteristics require careful assessment and management to reduce the likelihood that risk will eventuate.

When there is an indication that a patient may be at an elevated level of risk for violence, or when the clinician is doing a risk assessment specifically, there is much to be said for becoming familiar with the use of standardized risk assessment instruments. They can direct attention to important areas that require consideration and provide a structure for both gathering and evaluating relevant information. The choice is between the actuarial (e.g., VRAG, Static 99 and COVR) and those employing structured professional judgement (e.g., HCR 20). The former generate fixed risk levels, the latter in their nature are open to further modification in the light of clinical and common sense. Pure actuarial instruments (e.g. Static 99 and COVR) may appeal to the less experienced given their relative simplicity of administration. Their disadvantage even for the expert is, however, that they can function like black boxes which generate evaluations without the user being necessarily aware of how the instrument was constituted and the limitations that should attend its use. Moreover, they have been validated for use in fairly limited contexts and countries. As a result, clinicians may put undue weight on the results the instruments produce, without knowing how the results might vary in the context in which they are working.

Structured professional judgement is more transparent, if more demanding of the user. Where the actuarial measures can be used by relatively poorly trained individuals or in some contexts clerical staff, the structured professional judgment measures must be employed by those with considerable expertize. We favour the structured professional judgement approach of the HCR 20 but readers need to understand that actuarial instruments are perhaps the dominant forms of risk assessment instrumentation, due in no small part to their reliance on static variables and ease of administration. The HCR 20 is designed for use with general and forensic psychiatric patients and it has been found to have utility in general offender populations. Those whose work brings them into contact with specific groups such as perpetrators of domestic violence, child molesters or stalkers should consider using risk assessment approaches developed specifically for these populations (e.g., the Spousal Assault Risk Assessment (SARA),[41,42] Sexual Violence Risk-20 (SVR-20),[43] Risk for Sexual Violence Protocol (RSVP).[44,45]

The end point of a risk analysis in most cases will allow individuals to be placed into one of three somewhat arbitrary groups which encompass both likelihood and potential level of damage:-

i) High Risk – the individual presents a significant risk of committing a seriously damaging act of violence within a reasonable timeframe (less than a year or so). It is generally impossible to quantify the numerical probability for 'high risk' as it will vary across instruments and in different situations from approximately 30 per cent (sexual offending) to 80 per cent (the occurrence of general violence). Management of the risk is required immediately and the level of risk should be re-evaluated periodically depending on the extent to which the individual's personal situation and dynamic risk factors may vary. When the potential violence is of a particularly horrendous nature (e.g. potentially lethal) the timeframe can reasonably be extended to encompass a number of years. In, for example, some predatory child molesters and some who have killed from morbid jealousy, the chances of re-offending may be substantial and continue virtually throughout the offender's life span.

ii) Moderate Risk – the individual presents a real risk of committing a damaging act which might inflict minor injuries and/or significant fear and distress within a year or so. This group also requires their level of risk to be managed though, because the severity of behaviour is less extreme and the time period

perhaps less imminent, the extent of management is less intensive or restrictive. Also in this group can be placed those with a more remote (less than 30 per cent) but not inconsequential (above 5 per cent which in practice is the limit of reliable detection) risk of serious violence. Examples include those at lower risk of committing acts of a more serious nature (e.g., those making viable and tangible threats of engaging in behaviour which would lead to harm).

iii) Low Risk – Individuals who do not present a real risk of harming a third party. Their level of risk and potential for severe risk is low. They do not require any risk management plan beyond normal care and there is no need to re-evaluate the risk at any time in the reasonably foreseeable future. In essence, this category includes everyone not in the High or Moderate categories.

A final re-emphasis. Even using standardized instruments the final evaluation has to take into account factors beyond the figures generated by the black box. Risk evaluation is about formulation not simply calculations.

Identifying factors relevant to decreasing risk

The primarily clinical purpose of risk assessment should be risk management. Risk management is about identifying those factors which mediate the increased risk and modifying them to decrease risk. The structured professional judgment instruments and a good working knowledge of the risk literature will assist the clinician to identify and understand relevant risk factors. As noted above, identification of the dynamic risk factors points directed to potential management approaches. A simplified schema of the mediators and moderators which link having a schizophrenic syndrome to violent behaviour is presented in Fig. 11.14.1 with the basic approach to breaking or attenuating those links represented in Fig. 11.14.2.

The management strategies in high risk groups take us directly to good clinical practice with the addition of few specific approaches. The basic approach to management of high risk individuals can be summarized:

1 Substance abuse. This claims first place not because it is necessarily the most important risk factor but because the presence of the significant abuse of alcohol or drugs can disrupt all other management approaches.

2 Psychopathology. Obtaining adequate control of the delusions, affective disturbances, and hallucinations which predispose in some cases to offending behaviours is the second management imperative. Again without adequate control of the psychopathology little progress is likely in more targeted treatment modalities. High risk groups are often reluctant to comply either with medication or psychological management. This may force the use of compulsory treatment in an inpatient situation for sufficient time for the patient to begin to experience the benefits of amelioration of their active psychosis. Extended admissions (4–12 weeks) also provide an opportunity to establish trusting therapeutic relationships, though equally they can be productive of resentment and even more marked resistance to treatment. Second generation antipsychotics are preferable in this group as they are better tolerated and less likely to further impair cognitive and particularly frontal deficits. Depot medication is useful initially though this currently restricts considerably the choice of medication at least until more second generation antipsychotics come available in depot format. Clozapine can be particularly effective in this group but the level of cooperation required usually prevents its use at least in the early phase of management.

3 Social Circumstances. Discharging high risk patients to disorganized or casual accommodation in high crime neighbourhoods virtually guarantees offending behaviours. Similarly the drift back into contact with substance abusing peer groups increases risk and further disrupts management. Unemployment or just

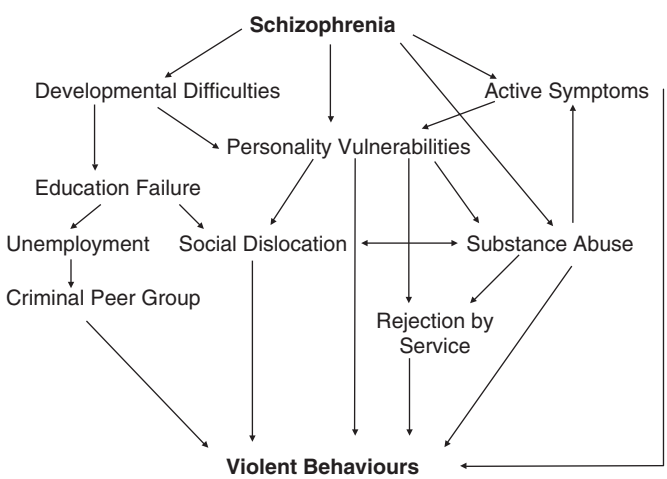


Fig. 11.14.1 A simplified schema of the mediators and moderatory between having a schizophrenic syndrome and behaving violently are illustrated. The very complexity of the nexus between illness and violence offers multiple opportunities for intervening to break the links. [Reproduced from Mullen, P.E., (2006). Schizophrenia and violence: from correlations to preventative strategies. *Advances in Psychiatric Treatment*, **12**, 239–48, copyright 2006, The Royal College of Psychiatrists.]

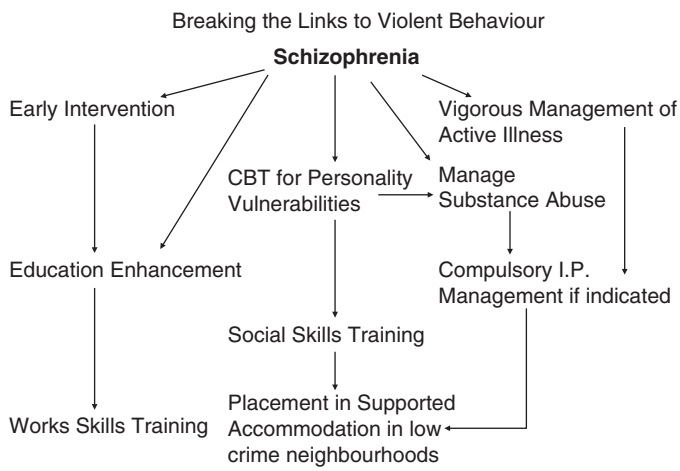


Fig. 11.14.2 Illustrates some of the interventions which could reduce the strength of the association between having schizophrenia and behaving violently. All interventions depend on accepting that it is the services duty to manage both the violence which can emerge from schizophrenia, and those with schizophrenia who are also substance abusing, delinquent and objecting. [Reproduced from Mullen, P.E., (2006). Schizophrenia and violence: from correlations to preventative strategies. *Advances in Psychiatric Treatment*, **12**, 239–48, copyright 2006, The Royal College of Psychiatrists.]

the lack of structure to their days increases risk. An absence of non offending, non substance use satisfactions leaves patients vulnerable to offending. The high risk groups need stable accommodation in low crime neighbourhoods, with active support, and structured recreation with later more long term educational and work related activities. Assisting them into contact with non deviant social groups (via sport, voluntary work, employment, or hobbies) is essential to establish social rewards to replace the pleasures of substance use and crime.

4 Insight. Or more concretely the acceptance of the need to change their attitudes and behaviours which support criminal and violent behaviours is a requirement for more targeted interventions. The stages of change model[46] combined with motivational interviewing[47] can assist in beginning the process. Ultimately it is through establishing a trusting relationship with the therapist and ideally the treatment team that commitment to change and maintaining change is obtained.[48]

5 Personality Traits. Personality disorders may or may not respond to treatment but the traits out of which they have been constructed are mostly open to modification. The objective is not to transform a suspicious, manipulative, insensitive, self absorbed thug into a paragon of the social virtues. It is simply to ameliorate those traits which predispose to antisocial behaviour. Targeted CBT offers a range of options. Wong and Hare's[49] guide to managing psychopathic traits offers a useful source of guidance for developing such programs.

6 Victim Empathy. Assisting the high risk group to understand the impact of their behaviour is essential. Sadly this is often best approached through sensitizing them to the harm they bring on themselves rather than through victim empathy programs, but both are worth attempting.

7 Common Sense and Prudential Wisdom. High risk patients with a schizophrenic syndrome not infrequently lack the mundane capacity to foresee the obvious outcomes of their behaviour. This produces a feckless foolishness. Instilling prudential wisdom in those impaired in this manner is a matter of slow progress in structured interactions which focus on their actual behaviours and the enhancement of the capacity to modify those behaviours in function of their longer term outcomes.

Conclusions

There are significant pressures and obligations—both legal and professional—for mental health professionals to give prominence to risk assessment and management with their patients (clients). The extent to which risk assessment will be more or less prominent will depend largely upon the nature of the professionals work, and the environment in which it occurs. Caring for those with high prevalence disorders in private practice will have little to worry about. For those responsible for patients who are acutely psychotic and of course those working in forensic contexts, the importance of risk assessment will be more significant.

We are at an important crossroads in our level of knowledge about both violence among psychiatric patients and about risk assessment. The fact that the risk for violence among psychiatric patients is not insignificant does mean that risk assessment and management is a legitimate activity for mental health professionals.[45] As such, psychiatrists and psychologists must become familiar with the risk assessment literature and emerging technologies. With appropriate levels of knowledge and training, clinicians can master this complex area in a way that can serve to satisfy their professional and legal obligations. Only with a good understanding of the field can we be protected from the risk of either blithely neglecting the violence risk our patients pose or of becoming so risk averse so as to unnecessarily and arbitrarily restrict our patients' liberty.

In the area of risk assessment, the framing of the main research questions and the articulation of the resulting data is increasingly in terms of actuarial risks and the generation of standardized questionnaires which will generate predictive scores. The technology of risk assessments could become one of the primary mediators of the relationship between the professional and the mentally disordered person. This will radically alter how the patient and their disorders and disabilities are revealed to us. We must not allow the technological focus on risk to replace the importance of the patient's personal and social context. We must not allow it to objectify them and their disorder as an embodiment of a quantum of 'riskiness.' To this end, there is an argument to occasionally separate the formal risk assessment task from that of patient treatment. The roles are blended—as they should be—in most cases, such as caring for voluntary patients and those involuntarily committed for brief periods. When the question of the patient's level of risk is that which will determine their liberty over the long term, however, it is less tenable for a treating clinician to maintain a productive therapeutic relationship with patients while holding the reins on their liberty (e.g., consideration of reviews of indeterminate dispositions, evaluations of risk for sentencing purposes and parole decision-making).

Technology is about performance and control, it is about domination, and the objects of technological manipulation are just that, objects [see[50, 51]]. To the extent that technological approaches to risk assessment come to dominate clinical practice, whatever benefit they may bring, the price will be reframing the clinician's view of their patients as potentially dangerous things. Risk assessment forms part of a major shift in psychiatric practice and theory away from individually based engagements between clinicians and uniquely troubled individuals to a world of standardized best practice. Instruments direct diagnosis, diagnosis determines which system of treatment is to be applied and risk assessments enable us to prevent damage to, or by, the objects of our professional responsibilities. Efficient, effective, properly evaluated performances of mandated procedures becomes the definition not only of the normative but the ethical.

This chapter began with a quotation from a nobel prize winner, it will end with the story of another. In reading the account provided by Nasser of the schizophrenic illness of the Nobel Laureate John Nash images emerge of the complex interactions between illness and the sufferer's humanity, life and even genius.[52] Nash had a devastating mental illness, he was as a result of his illness compulsorily treated and even at times regarded as dangerous. Nash finally entered a stable remission in which he could once more work as a mathematician. This was without the continuing aid of medication or any other form of mental health ministrations. What has this to do with risk assessment and management? Everything. The outcome of an illness such as one of the schizophrenias in an individual case remains enormously difficult to predict. We must, as mental health professionals, act on our estimates of

future probabilities. We should struggle to make our risk assessments and risk management strategies as effective as possible. But in the end we should remain modest about our capacities to perform such predictive and preventive functions and not loose curiosity about what really delivers Nash and many others from insanity and even dangerousness.

Further information

Maden, A. (2007). *Treating Violence: A guide to risk management in mental health*. Oxford University Press, New York.

Mossman, D. (2007). Critique of Pure Risk Assessment or, Kant Meets Tarasoff. *University of Cincinnati Law Review*, **75**, 523–609.

Douglas, K., and Skeem, J. (2005). Violence risk assessment: Getting specific about being dynamic. Psychology, *Public Policy and Law*, **11**, 347–83.

Ogloff, J. R. P. (2006). The Psychopathy/Antisocial Personality Disorder Conundrum. *Australian and New Zealand Journal of Psychiatry,* **40**, 519–28.

References

1. Monahan, J. (1988). Risk Assessment of Violence Among the Mentally Disordered: Generating Useful Knowledge. *International Journal of Law and Psychiatry,* **11**, 249–57.
2. Snowden, P. (1997). Practical aspects of clinical risk assessment and management. *British Journal of Psychiatry,* **170** (suppl. 32), 32–4.
3. Gunn, J. (1996). Lets get serious about dangerousness. *Criminal Behaviour and Mental Health*, Supplement: 51–64.
4. Kraemer, H., Kazdin, A., Offord, D., *et al.* (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry,* **54**, 337–41.
5. Castel, R. (1991). From dangerousness to risk. In *The Foucault Effect; Studies in Governmentality* (eds. G. Burchell, C. Gordon, P. Miller). Hemel Hempstead, Harester Wheatsheaf.
6. Douglas, M. (1992). *Risk and Blame: Essays in Cultural Theory*. Routledge, London.
7. O'Mally, P. (1991). Risk, Power and Crime prevention. *Economy & Society,* **21**, 252–75.
8. Rose, N. (1996). Psychiatry as a political science: advanced liberalism and the administration of risk. *History of the Human Sciences,* **9**, 1–23.
9. Scott, P. D. (1977). Assessing dangerousness in criminals. *British Journal of Psychiatry*, **131**, 127–42
10. Stürup, G.K. (1968). Will this man be dangerous? In *The Mentally Abnormal Offender* (eds. A. de Reuk and R. Porter), pp. 5–18. Churchill, London,
11. Steadman, H.J. and Cocozza, J.J. (1975). We can't predict who is dangerous. *Psychology Today,* **8**, 22–35.
12. Shaw, S.H. (1973). The dangerousness of dangerousness. *Medicine Science and Law,* **13**, 269–71.
13. Carstensen, P.C. (1994). The evolving duty of mental health professionals to third parties: a doctrinal and institutional examination. *International Journal of Law and Psychiatry,* **17**, 1–42.
14. Cocozza, J. and Steadman, H. (1976). The failure of psychiatric predictions of dangerousness: Clear and convincing evidence. *Rutgers Law Review,* **29**, 1084–1101.
15. *Barefoot v. Estelle*, 463 U.S. 880, (1983).
16. *Tarasoff v. Regents of University of California*, 551 P.2d 334 (1976).
17. *United States v. Hinckley*, 672 F.2d 115 (D.C. Cir. 1982).
18. *Kansas v. Hendricks*, **521** U.S. 346, (1997).
19. Maden, A. (2007). *Treating Violence: A guide to risk management in mental health*. Oxford University Press, New York.
20. Monahan, J. (1981). *The Clinical Prediction of Violent Behaviour*. US Government Printing Office, Washington, DC.
21. Belfrage, H. (1998). Implementing the HCR-20 scheme for risk assessment in a forensic psychiatric hospital: Integrating research and clinical practice. *Journal of Forensic Psychiatry,* **9**, 328–38.
22. Douglas, K., Ogloff, J. and Hart, S. (2003). Evaluation of the structured professional judgment model of violence risk assessment among forensic psychiatric patients. *Psychiatric Services,* **54**, 1372–79.
23. Andrews, D.A. and Bonta, J. (2003). *The Psychology of Criminal Conduct* (3rd). Anderson Publishing Co., Ohio.
24. Rice, M.E. and Harris, G.T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology,* **63**, 737–48.
25. Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology,* **62**, 783–92.
26. Mullen, P.E. (2006). Schizophrenia and Violence: from correlations to preventative strategies. *Advances in Psychiatric Treatment* **12**, 239–248.
27. Wallace, C., Mullen, P.E., Burgess, P., *et al.* (1998). Serious criminal offending and mental disorder: Case linkage study. *British Journal of Psychiatry*, **172**, 477–484.
28. Lindqvist, P. and Allebeck, P. (1990). Schizophrenia and Crime. A Longitudinal Follow-up of 644 Schizophrenics in Stockholm. *British Journal of Psychiatry*, **157**, 345–50.
29. Hafner, H. and Böker, W. (1982). *Crimes of Violence by Mentally Abnormal Offenders* (trans. H. Marshall). Cambridge University Press, Cambridge.
30. Douglas, K., Ogloff, J., Nicholls, T., *et al.* (1999). Assessing risk for violence among psychiatric patients: The HCR-20 Violence Risk Assessment Scheme and the Psychopathy Checklist: Screening Version. *Journal of Consulting and Clinical Psychology,* **67**, 917–30.
31. Quinsey, V.L., Harris, G.T., Rice, M.E., *et al.* (1998). *Violent Offenders: Appraising and Managing Risk*. American Psychological Association, Washington, DC.
32. Quinsey, V.L., Rice, M.E., Harris, G.T., *et al.* (2004). *Violent offenders: Appraising and managing risk (2nd ed.)*. American Psychological Association, Washington, DC.
33. Silver, E., Smith, W.R. and Banks, S. (2000). Constructing actuarial devices for predicting recidivism: A comparison of methods. *Criminal Justice and Behavior*, **27**, 733–64.
34. Barbaree, H.E., Langton, C.M. and Blanchard, R. (2007). Predicting recidivism in sex offenders using the VRAG and SORAG: The contribution of age-at-release. *International Journal of Forensic Mental Health,* **6**, 29–46.
35. Monahan, J., Steadman, H.J. *et al.* (2005). *Classification of Violence Risk*. Psychological Assessment Resources, Inc., Lutz, FL.
36. Hart, S.D., Michie, C. and Cooke, D.J. (2007). Precision of actuarial risk assessment instruments. Evaluating the 'margins of error' of group v. individual predictions of violence. *The British Journal of Psychiatry* **190** (Suppl. 49) s60–s65.
37. Hanson, R.K. and Thornton, D.M. (1999). *Static 99: Improving Actuarial Risk Assessments for Sex Offenders*. Ottawa Public Works and Government Services, Canada.
38. Mercado, C.C. and Ogloff, J.R.P. (2007). Risk and the preventive detention of sex offenders in Australia and the United States. *International Journal of Law and Psychiatry,* **30**, 49–59.
39. Wood, M. and Ogloff, J.R.P. (2006). Victoria's Serious Sex Offenders Monitoring Act: Implications for the accuracy of sex offender risk assessment. *Psychiatry, Psychology and the Law,* **13**(2), 182–98.
40. Enticott, P., Ogloff, J.R.P. and Bradshaw, J. (in press). Response inhibition and impulsivity in schizophrenia. *Psychiatry Research*.
41. Kropp, P.R., Hart, S.D., Webster, C.D., *et al.* (1994). *Manual of the Spousal Assault Risk Assessment Guide (2nd ed)*. British Columbia Institute on Family Violence, Vancouver, Canada.
42. Kropp, P.R., Hart, S.D., Webster, C.D., *et al.* (2000). The Spousal Assault Risk Assessment Guide (SARA) guide: Reliability and validity in adult male offenders. *Law and Human Behaviour,* **24 (1)**, 101–18.

43. Boer, D.P., Hart, S.D., Kropp, P.R., *et al.* (1997). *Sexual violence risk-20: Professional guidelines for assessing risk of sexual violence.* British Columbia Institute on Family Violence and Mental Health, Law, and Policy Institute, Simon Fraser University; Vancouver, British Columbia.
44. Hart, S.D., Kropp, P.R. and Laws, D.R. (2003). *The risk for sexual violence protocol (RSVP): Structured professional guidelines for assessing risk of sexual violence*. Mental Health, Law, and Policy Institute, Simon Fraser University and British Columbia Institute on Family Violence, Vancouver, British Columbia.
45. Mullen, P.E., MacKenzie, R., Ogloff, J.R.P., *et al.* (2006). Assessing and managing the risks in the stalking situation. *Journal of the American Academy of Psychiatry and the Law,* **34**, 439–50.
46. Prochaska, J.O. and Di Clemente, C.C. (1992). Stages of change in the modification of problem behaviours. *Progress in Behavioural Modification,* **28**, 183–218.
47. Miller, W.R. and Rollnick, S. (1991). *Motivational Interviewing: Preparing People to Change Addictive Behaviour*. Guilford, New York.
48. Douglas, K.S., Webster, C.D., Hart, S.D., *et al.* (2001). *HCR-20 Violence Risk Management Companion Guide*. Simon Fraser University, British Columbia.
49. Wong, S. and Hare, R.D. (2005). *Guidelines for a Psychopathy Treatment Program*. MHS, Toronto.
50. Heidegger, M. (1993). The Question Concerning Technology. In: *Basic Writings*. Harper, San Francisco, 297–352.
51. Lyotard, J.F. (1979). *The Post Modern Condition: A Report on Knowledge*. Trans. G Bennington and B Massumi. University of Minnesota Press, Minneapolis.
52. Nasser, A. (1998). *A Beautiful Mind*. New York: Simon and Schuster

11.15

The expert witness in the Criminal Court: assessment, reports, and testimony

John O'Grady

As an expert witness in the Criminal Court, the psychiatrist ceases to be simply a doctor as a psychiatrist's report and testimony addresses issues on the boundary between law and psychiatry. The law is not primarily concerned with the welfare of the defendant. Criminal law is concerned with justice, fact finding, and the attribution of guilt whilst psychiatry concerns itself with the welfare of the individual, their mental disorder, and its treatment. This chapter will explore the legal framework for expert reports and testimony, standards for such work, the particular ethical dilemmas of this work and provide practical guidance on preparation of reports and testimony.

This chapter draws upon previously published work by the author.[1,2] Expert evidence cannot be understood except in reference to a particular legal jurisdiction. For this chapter the legal system in the United Kingdom is chosen but the general principles will apply to all jurisdictions. Issues specific to Civil and Family courts will not be discussed.

The psychiatric expert witness[3–5]

Witnesses in court can only give evidence of facts they personally perceived and not evidence of their opinion. It is for the court to draw inferences from the testimony of witnesses. The opinion of an expert witness is an exclusion to this general rule because courts need the assistance of experts to consider issues beyond their knowledge.

Lawton L.J in R v Turner established a 'common knowledge' rule governing expert evidence as follows:

> An expert opinion is admissible to furnish the Courts with scientific information which is likely to be outside the experience and knowledge of a Judge or a jury. If on the proven fact, a Judge or jury can form their own conclusions without help, then the opinion of an expert is unnecessary. In such a case if it given dressed up in scientific jargon it may make judgement more difficult. The facts that an expert witness has impressive scientific qualifications does not by that fact alone make his opinion on matters of human nature and behaviour within the limits of normality any more helpful than that of the jurors themselves; but there is a danger that they may think it does … jurors do not need psychiatrists to tell them how ordinary folk who are not suffering from any mental illness are likely to react to the stresses and strains of life.

This seems to limit psychiatric evidence to recognized mental disorder. However, expert advice is allowed which is 'outside the experience and knowledge of a Judge or jury'. The abnormal/normal dichotomy is not a rule of law but guidance. Courts have allowed evidence on a variety of conditions which would not normally be thought of as established mental disorder, for example 'Battered Women's syndrome'.

Particular problems arise for the Court in respect of borderline conditions falling short of recognized mental disorder. Here admissibility will usually be determined by the court's judgement as to whether the expert evidence addresses matters outside the experience or knowledge of a Judge or jury. Generally courts seek to limit evidence to established abnormal conditions.[3,5] Courts have problems with evidence that utilizes leading edge or novel theory or diagnosis. The Court will require evidence that the novel theory or diagnosis is sufficiently organized or recognized to be accepted as a reliable body of knowledge by the profession. The expert will need to demonstrate that acceptance through reference to scientific literature.[3,5]

For medical experts, the Courts are able to establish expertise by reference to qualification and training. Nevertheless, the expert must be able to demonstrate that they have the requisite expertise in a particular case. For example, a psychiatrist trained in general adult psychiatry may not be an expert in a case concerning a person with moderate to severe learning disability.

Immunity from suit

Lawyers and experts enjoy immunity from suit (civil litigation) for their professional work in court.[3] For medical expert witnesses this includes their report and any oral evidence presented in court. That immunity does not extend to immunity from report to the doctor's regulatory body. It does not extend to subsequent actions such as duty of confidentiality in respect of disclosure of reports to third parties. The judgement in the case of GMC v Meadows lays out the legal and public policy arguments for immunity from suit but with regulation by their professional regulatory body.

Reliability of expert testimony

Courts need to assure themselves that an expert witness's evidence is reliable. This creates an obvious problem as the very reason an expert is giving evidence is that they have expertise which the judge or jury does not possess.

Courts have utilized three broad approaches to this problem. The first is to examine the scientific validity of evidence. The second is to devise standards for expert evidence and the third is to regulate experts through formal accreditation systems.

Scrutiny of scientific evidence

The landmark case is that of Daubert v Merrell Dow Pharmaceuticals in the United States courts. That ruling established stringent criteria to judge reliability to include that the technique, body of knowledge, or theory can be tested, has been subjected to peer review and publication, has a known rate of error, is subject to maintenance of standards and controls and is generally accepted by the scientific community. This judgement is problematic for a number of reasons,[6] not least because the court does not have that expert's specific knowledge but nevertheless has to make a scientific judgement on the reliability of that expert's evidence. Judgements are unlikely to be value free determinations and there is a risk that the admissibility of evidence could be distorted by policy considerations or interfere with the use of leading edge science in the court. These considerations have made United Kingdom courts reluctant to introduce a 'Daubert' type test but there is pressure to do so[4,7] with public concern about miscarriages of justice linked to expert evidence.

Regulatory rules for expert witnesses

Courts have defined the standards expected of an expert witness; the landmark case being the judgement of Cresswell J in The Ikarian Reefer. Court judgements have been used to draw up formal rules governing civil family and criminal courts; for UK Criminal courts, the relevant rules are contained in Part 33, Criminal Procedure Rules.[8] The common features are listed in Box 11.15.1 below. The understandable anxiety of the court to ensure experts adhere to these stringent standards may have the unfortunate effect of deterring psychiatrists from providing occasional expert reports for criminal courts.

Box 11.15.1 Common features of regulatory rules for civil family and criminal courts.

Expert Reports should contain:

1. Details of academic and professional qualifications together with experience and accreditation relevant to the opinions expressed in the report (usually as a summary in the introduction with more detail within and Appendix).
2. A statement of the range and extent of expertise together with limitations upon that expertise, particularly declaring when a particular issue is outside his expertise.
3. A statement setting out the substance of all instructions received together with listing all materials provided and considered, upon which the opinion is based.
4. Where there is a range of opinion on matters dealt with in the report, a summary range of opinion together with reasons for the experts preferred opinion (see section below on Report Writing).
5. A declaration of any facts, materials, or investigations which might bear upon or be made against the expert opinion.
6. Extracts of literature or any other material upon which the scientific evidence is based.
7. A statement of which facts are within the expert's own knowledge and which are assumed.**
8. Where an opinion is qualified, a statement to that effect.
9. A statement that the expert has complied with his or her duty to the court to provide independent assistance by way of objective unbiased opinion in matters within his or her expertise.
10. A statement that the expert will inform all parties, including the court, in the event that his or her opinion changes on any material issue.
11. A declaration of truth.

** Courts distinguish true and assumed facts. The only facts the psychiatric expert will routinely know to be true are the results of examination and results of tests or investigations. All other facts will be assumed to be true.[9]

Accreditation

In the United Kingdom, there are a number of organizations to accredit expert witnesses utilizing some combination of direct scrutiny of reports and references from legal teams. None so far have addressed the specific needs of the expert psychiatrist in court. They provide the court with some measure of an expert witness's expertise in legal matters over and above what comes from their basic professional qualification. Critics have pointed out[10] that once registered accreditation is unlikely to pick up poor practice, as experts with years of experience, but not necessarily competence, are unlikely to be refused accreditation. Accreditation is unlikely to prevent expert straying outside their area of expertise. These schemes have not as yet been able to deal effectively with problems of accrediting inexperienced but competent aspiring expert witnesses. To be effective they may require codes of discipline with the attendant danger of attracting vexatious complaints. The Royal College of Psychiatrists in the United Kingdom utilizes a competency based training framework together with standards for continuing professional development (CPD) following training to promote a high standard in medico-legal work. Evidence of completion of such training and CPD is likely to be the most effective way of demonstrating credibility as an expert witness.

Ethics

Dual role

Stone[11] used the term 'dual role' to describe the psychiatrist in the legal context. In Stone's view the role of the clinician and medical examiner for Court are irreconcilable. The evaluee/patient is unable to distinguish the role of the medical examiner as a Court expert from that of personal physician. This result is an inability to protect

themselves from inadvertent disclosure that might adversely affect the outcome in Court. He argued that clinicians cannot help using their therapeutic skills to engage the patient in disclosure. The dual role arises from the use made of the resulting psychiatric evidence for non-welfare purposes. Appelbaum[12] argued that the dual role of psychiatric experts in Court is best managed by understanding that psychiatrists operate outside the medical framework when they undertake forensic Court work and their practice is not governed by the ethical principles underpinning medical practice (beneficence and non-maleficence). Instead he argued that psychiatric experts should operate from a perspective of justice ethics employing ethical principles of objective truth finding and respect for the person (termed autonomy and truthfulness).

If this solution to the Dual Role dilemma were accepted, it would mean that the psychiatrist should not have a welfare/treatment role in respect of the person under evaluation. In the United Kingdom this is untenable[13] primarily because of Mental Health law which provides for diversion to the health system as a sentence following a finding of guilt (Hospital order). Weinstock *et al.*[14] have argued in the United States legal and clinical context the Appelbaum solution[12] is, untenable in that legal context as psychiatrists routinely have conflicting responsibilities thrust upon them where legal or other requirements may take precedence over patient welfare.

Reports addressing sentencing in the United Kingdom place the psychiatrist in a Dual Role position. This is because the psychiatric opinion can result in two outcomes for the evaluee

1 A welfare disposal under Mental Health legislation.

2 Potentially greater restrictions on the defendant, including an indeterminate life sentence where there is expert evidence on mental disorder but no recommendation for a welfare disposal.

Statutes that introduce indeterminate life sentences for public protection based upon assessed future risk of re-offending cause particular problems.[1,13,15] English Courts have, through case law and practice, sought psychiatric evidence when they consider defendants may have mental disorder and where the Court is considering an indeterminate life sentence.[1] Psychiatric evidence on risk will be central to the expert's evidence. The Court may have two options, a Hospital Order in suitable cases or an indeterminate life sentence. The psychiatrists does not have 'a priori' advance knowledge of what the outcome might be in a particular case thus routinely placing the psychiatrist into a dual role in respect of the evaluation.

Calvedeno[16] has pointed out that even where a welfare disposal is recommended, medical evidence in respect of special restrictions to a Hospital Order may lead to lengthy periods in hospital justified not by the need for treatment but by psychiatric judgement on risk in the future. Similar arguments apply to reports to Mental Health Tribunals for patients detained under mental health legislation.

One solution to the dual role is to act only where there is a realistic prospect of benefit to the patient. This leads some psychiatrists to only undertake work for defence teams. In the author's view, this is unethical as it lends itself to bias and deprives one side in the adversarial process of high quality experts.

A theory of mixed duties to address dual role conflict

Doctors are members of society and as citizens have responsibilities, prior to responsibilities as a doctor. The narrow domain of medical ethics does not remove from doctors their duty to consider the interests and rights of other people and to consider the distribution of benefits and risks. Beauchamp[17] proposes augmenting traditional medical ethics with principles he terms justice and respect for autonomy. On that basis O'Grady[1] suggests a framework of mixed duties for expert witnesses in court to address their 'dual role' conflict (see Box 11.15.2).

This approach implicitly requires the psychiatrist to work within a framework of conflicting duties where ethical judgements must balance the welfare of the evaluee against the rights of others and society's legitimate interest in protection from risk.

The psychiatric ethical expert in court is then the one who 'feels the tension' inherent in a dual role and is painfully aware of the conflicting demands of different ethical imperatives.

Risk assessment

Sentencing where public protection is a central issue poses particular difficulties for the psychiatrist as risk assessment becomes central to the court's decisions. Actuarial risk assessments can be particularly dangerous in the legal context. Mullen[18] argues 'The margins of error in every actual or conceivable risk assessment instrument are so wide at the individual level that their use in sentencing, or any form of detention, is unethical'. Whilst acknowledging the significant limitations of risk prediction at the individual level, the Court may nevertheless legitimately argue that evaluation of risk associated with mental disorder is an area falling outside the 'common knowledge' of Judge and Jury. Therefore the court must rely upon psychiatrist's opinion on risk and mental disorder as the psychiatrist is the only witness with the necessary expertise. Using a structured risk assessment methodology may go someway to ensure accuracy, objectivity, and truthfulness. A clear role for the psychiatric expert is to ensure the Court is provided with informed scientific evidence on the limitations of risk assessment and particularly the limitations of utilizing structured or actuarial risk instrument at the individual level[18,19]).

Structure of reports

Receiving instructions

The psychiatrist should understand the legal question to be addressed; where necessary standard text should be consulted (see

Box 11.15.2 Ethical principles to address Dual Role responsibilities

1 Medical ethics:
- Non-maleficence
- Beneficence

2 Justice ethics:
- Truthfulness (objective and subjective)
- Respect for autonomy
- Respect for the human rights of others (balancing the distribution of benefits and risks for the patient and society)

recommended reading). The psychiatrist must ensure that they have the necessary expertise to address the issues for the Court. Trainees must ensure that they are supervised by a suitably qualified senior and disclose this to the instructing party (including disclosure of the supervisor's appointment and qualifications).

The psychiatric expert must ensure that they can meet the needs of the Court as regards timescale for the report and understand that they can be compelled by the Court to give oral evidence; for example when they are on leave. The Court will not do so if the doctor has in advance disclosed dates when they are unavailable. Where there are fees to be paid, the letter of acceptance should state the contractual conditions for accepting instructions.

Rules of evidence in all legislations impose on the expert witness an overriding duty to the court outside of the duty owed to the party instructing them. The psychiatric expert witness has to develop a working relationship with the legal team instructing them but simultaneously discharge their overall duty to the court. One way of conceptualizing the relationship to the instructing party is as a 'consultant' to the legal team, educating them in the meaning of psychiatric findings.[20] Nevertheless it is naive to believe that expert will not be subject to overt or subtle influence by the instructing side.

Psychiatric reports should comply with relevant court rules for example Part 33 of the Criminal procedure rules for England.[8] Notes and documents must be retained for a sufficient period (undefined but at least to until last date for appeal) and disclosed to other experts in the case.[20] The expert should have appropriate indemnity insurance.

The interview

If the defendant is to be visited in prison, arrangements should be made well in advance and comply with the requirements of the institution.

At the beginning of the interview, the examining psychiatrist should explain carefully to the defendant the nature of the doctor's dual role, the limits of confidentiality in producing a medico-legal report and that the Court will have full disclosure of all material known to the report writer (no off record material). It is prudent to obtain a signed record of this discussion and to include it as part of the introduction to the report. Whenever possible, an informant should be interviewed; by telephone if necessary.

Structure of the report

(a) Declarations and introduction

The first section of the report should lay out the instructions received and what was done in order to produce the report. The dates and duration of interviews should be stated, including interviews with informants. For British Criminal Courts, Part 33 of the Criminal Procedure Rules[8] requires certain declarations and statements at the beginning of the report (see Box 11.7.1 above). A section on limitations to the report should be included to record matters such as documents not disclosed or unavailability of informants and state the impact on the expert's opinion.

(b) The facts

The middle section of the report should record briefly the facts upon which the opinion is based and should avoid interpretation which is the proper function of the opinion section. In psychiatric reports, the only facts that are within the psychiatrists own knowledge are likely to be those based on the findings of mental state examination. All other facts are assumed. If structured tests are utilized, they may also constitute facts within the psychiatrists own knowledge.

(c) Opinion

The role of the psychiatric expert is to provide an opinion on mental disorder and its implications for the matters before the Court. The opinion section should then start with a description of the defendant's mental disorder. If there is no evidence of mental disorder, then the privileged exception accorded to psychiatric experts no longer applies (see earlier section). The features that lead to a diagnosis of mental disorder should be described, avoiding jargon, and including mental state findings, so that others can understand how the opinion is reached. The diagnosis should be clearly stated using a recognized classification which for British psychiatrists will be the International Classification of Mental and Behavioural Disorders. Where a condition is described which is not part of such recognized classification systems or where 'leading edge' scientific findings are used to support the opinion that should be justified by disclosure (as an appendix to the report) of relevant literature to support the expert's opinion.

The second stage of the opinion is to translate the psychiatric findings into the legal language employed by the Courts. Terms such as 'diminished responsibility', 'insanity', or 'automatism' have precise legal definitions and the report should address how the psychiatric findings translate to the legal definitions employed by the Court.

Usually there will be a range of opinion and the psychiatrist should indicate the range, giving due weight to alternative opinion before recording the reasons for their own opinion. More than one legal issue may have to be considered together with the range of opinion on each separate legal issue. One helpful mental model is to consider the range of opinion that might be given by other experts if the case were presented to a psychiatric case conference.[20]

At point of sentencing where the court has concerns about public safety, the psychiatrist will be expected to provide an opinion on risk linked to the defendant's mental disorder. The ethical issues arising from that expectation should be thoroughly understood (see section above). It is usual to express a range of opinion (the case conference model) and given reasons for the expert's own opinion.

Where recommendations for a disposal under specific Acts are included, the precise wording of the relevant section of those Acts should be employed.

(i) Opinion on the ultimate issue

There is a common law injunction against a witness expressing an opinion upon the ultimate issue to be determined by the court. Many questions put to psychiatric experts test this rule to its limits. In this author's opinion, psychiatric experts should provide the court with objective evidence upon the mental state in and around the time of an alleged offence but stop short of expressing an opinion on the ultimate issue unless specifically instructed to do so by the presiding Judge.

Confidentiality

Medico-legal work undertaken by psychiatrists is governed by the same rule of confidentiality as applied to other clinical work.

Reports cannot be disclosed to a third party without the consent of the body commissioning the report. Psychiatric reports do not form part of a person's NHS medical record except by the express consent of the individual or their legal representative. Defence solicitors can exercise a right not to disclose a report to Court. Failure to comply with rules of confidentiality can lead to civil action or report to a professional regulatory body.

There may be circumstances where a psychiatrist believes that it is necessary to divulge confidential information to a Court without the evaluee's consent. This could arise:

(a) Where the evaluee refuses to cooperate with the preparation of a report.

or

(b) Where a report is not disclosed but the psychiatrist believes that disclosure is in the public interest.

A psychiatrist who believes that the evaluee is not cooperating with the preparation of a report because of mental illness has a duty to consider whether the evaluee's mental illness could interfere with a fair trial (for example, fitness to plead or lack of consideration of a mental health disposal). The psychiatrist must then make a judgement whether it is in the best interests of the evaluee for sufficient information to be provided to the Court to alert them of the doctor's concerns. Such disclosure will almost certainly be justified in the interests of a fair trial and justice. The doctor will also have a duty to consider whether steps ought to be taken to undertake a Mental Health Act assessment (British law) to enable transfer to a hospital for medical treatment.

The other situation where a breach of confidentiality may be justified is where a report is not disclosed but the report writer believes that the Court ought to consider the report's findings on potential risk to the public. In B.W. v Edgell and Rv Crozier, the Court held that the doctor's duty of confidence did not prevent a psychiatrist from taking steps to communicate the grounds of concern to the court. The strong public interest in disclosure to prevent a court from making decisions based upon inadequate information was held to override the psychiatrist's duty of confidentiality. Where a doctor is considering disclosure in these circumstances, advice should be sought from an experienced colleague, case law, and regulatory body guidance consulted and the doctor should seek advice from their indemnity insurer.

Appearing in court

Advice on practical matters concerning a Court appearance is beyond the scope of this chapter but guidance is available (see recommended reading). Those undertaking regular expert work should consider courses which prepare them for appearance in Court and should understand the legal framework for giving oral evidence in court.[21]

The cardinal rule when giving oral evidence in Court is that although called by one party, the expert witness is not giving evidence for that party's side but is under a duty to provide fair and impartial evidence to the Court even where this conflicts with the interests of the party calling them.

In Criminal Courts, the defendant gives evidence before expert evidence is heard. Experts are allowed, unlike witnesses of fact, to sit in Court and hear the evidence of other witnesses before they, themselves, give evidence. An expert can be called by any interested party in proceedings.

When calling an expert witness, the advocate must elicit the following[21]:

1 The expert's qualifications: If the report has been prepared according to criminal procedure rules, the report will contain a biography setting out the qualifications and experience of the witness. It will then be usual for the advocate to lead this part of the evidence by reference to the biography supplied in the report. It will be perfectly permissible for the other side to call into question the expert's qualifications. This should be met politely by outlining the reasons why the expert believes they have the requisite qualifications and experience to answer the questions posed in instructions.

2 Disclosure of the expert's report: The report will have been pre-read by the Judge and it is usual for the examiner to refer to relevant sections of the report. A report with numbered paragraphs is easier for the Court to follow. The jury will not have read the report and will not usually be given sight of the report. Their knowledge of the expert's report will come from submissions made by either side and through the Judge's questioning and summing up.

3 Advocates are under a duty to challenge disputed evidence. Thus where more than one expert opinion is provided and they differ; the expert must expect their opinion to be disputed. The expert must resist pressure from one party to deviate from or express greater certainty about an opinion they have reached in the written report.

4 An expert witness may be cross-examined as a hostile witness if there is good reason to suppose that they are not telling the truth. Thankfully this is extremely rare. The possibility of deliberate or inadvertent bias must, however, always be considered.

Further information

Expert Witness Institute. (2006). *Experts in the civil courts*. Oxford University Press, Oxford.

Hodgkinson, T. and James, M. (2007). *Expert evidence: law and practice* (2nd edn). Sweet and Maxwell, London (detailed legal textbook on law relating to expert witnesses).

Holburn, C.J., Bond, C., Solon, M., *et al.* (2000). *Healthcare professionals as witnesses to the court*, Greenwich Medical Media, London (practical advice on reports and appearing in court).

Omerod, D., and Hooper, A. (2008) *Blackstone's Criminal Practice 2009*. Oxford University Press, Oxford (standard textbook on criminal law, frequently revised).

Weinstock, R. (2003). Part 1. History and practice of forensic psychiatry. 1-94. In *Principles and practice of forensic psychiatry* (2nd edn) (ed. R. Rosner), pp. 56–72. Arnold, London.(for a USA perspective)

See also references 9 and 20 below

References

1. O'Grady, J.C. (2002). Psychiatric evidence and sentencing: ethical dilemmas. *Criminal Behaviour and Mental Health*, **12**, 179–84.
2. O'Grady, J.C. (2002). Report writing for the criminal court. *Psychiatry*, **6**, 11.
3. Murphy, P. and Phillips, J. (2007). F10 expert opinion evidence. In *Blackstone's criminal practice 2007*, Section F10, pp. 2449–63. Oxford University Press, Oxford.
4. Ormerod, D. and Roberts, A. (2006). The admissibility of expert evidence. In *Witness testimony* (eds. A. Heaton - Armstrong,

E. Shepherd, G. Gudjonsson, and D. Wolchover), Chap. 22, pp. 401–24, Oxford University Press, Oxford.
5. Hodgkinson, T. and James, M. (2007). Introduction: the principles and development of expert evidence. In *Expert evidence: law and practice* (eds. T. Hodgkinson and M. James), Chap. 1, pp. 1–43, Sweet and Maxwell, London.
6. Sales, B.D. and Shaman, D.W. (2007). Science, experts and law: reflections on the past and the future. In *Expert psychological testimony for the courts* (eds. M. Costanzo, D. Krauss, and K. Pezdek), Chap. 2, pp. 9–30, Lawrence Erlbaum Associates, Mahwah, New Jersey.
7. Hodgkinson, T. and James, M. (2007). Expert evidence: the future. In *Expert evidence: law and practice* (eds. T. Hodgkinson and M. James), Chap. 28, pp. 645–9, Sweet and Maxwell, London.
8. Ministry of Justice. (2005). Part 33: expert evidence. In *Criminal procedure rules*. Ministry of Justice, London. Available at: http://www.dca.gov.uk/procedurerules.htm
9. Rix, K.J.B. (2008). The psychiatrist as expert witness: Part 1. General principles and civil cases. *Advances in psychiatric treatment*.
10. Burn, S. and Thompson, B. (2006). Single joint expert. In *Experts in the civil courts* (ed. L. Blom-Cooper), Chap. 5, pp. 57–76. Oxford University Press, Oxford.
11. Stone, A.A. (1984). The ethical boundaries of forensic psychiatry - a view from the Ivory tower. *Bulletin of the American Academy of Psychiatry and the Law*, **12**, 209–19.
12. Appelbaum, P.S. (1997). A theory of ethics for forensic psychiatry. *Journal of the American Academy of Psychiatry and the Law*, **25**, 233–47.
13. Royal College of Psychiatrists. (2004). *The psychiatrist, courts and sentencing: the impact of extended sentencing on the ethical framework of forensic psychiatry College Report 129*. Royal College of Psychiatrists, London. Available at:http://www.rcpsych.ac.uk/publications/collegereports.aspx
14. Weinstock, R., Leong, G.B., and Silva, A. (2003). Ethical guidelines. In *Principles and practice of forensic psychiatry* (2nd edn) (ed. R. Rosner), pp. 56–72. Arnold, London.
15. Feeley, M. and Simon, J. (1992). The new penology: notes on an emerging strategy of corrections and its implications. *Criminology*, **30**, 449–74.
16. Calvedeno, M. (1999). The psychiatrist as Gaoler. *Journal of Forensic Psychiatry*, **10**, 525–37.
17. Beauchamp, T.L. (1999). The philosophical basis of psychiatric ethics. In *Psychiatric ethics* (3rd edn) (eds. L.S. Bhoc, P. Chadoff, and S.A. Green), pp. 25–47, Oxford University Press, Oxford.
18. Mullen, P.E. (2007). Dangerous and severe personality disorder and in need of treatment. *The British Journal of Psychiatry*, **190**, s3–7.
19. Hart, S.D., Michie, C., and Looke, D.J. (2007). Precision of actuarial risk assessment instruments: evaluating the 'Managing of Error' of group v individual predictions of violence. *The British Journal of Psychiatry*, **190**, s51–9.
20. Rix, K.J.B. (2008). The psychiatrist as expert witness: part 2. Criminal cases and Royal College of Psychiatrist's guidance. *Advances in Psychiatric Treatment*, **14**, 109–14.
21. Hodgkinson, T. and James, M. (2007). The expert witness at trial. In *Expert evidence: law and practice* (eds. T. Hodgkinson and M. James), Chap. 6, pp. 163–82. Sweet and Maxwell, London.

Case law

R v Turner [1975] Q.B. 834 and 841
GMC v Meadows [2006] EWCA Civ 1390
Daubert v Merrell Dow Pharmaceuticals Inc [1993] 509 US 575
Ikarian Reefer [1993] 2 Lloyds Rep 68 at 81
R v Edgell [1990] Cr App ch 359
R v Crozier [1990–1991] 12 Cr App(s)206

11.16

Managing offenders with psychiatric disorders in general psychiatric services

James R. P. Ogloff

It has been shown that the prevalence of mental illness among those in the criminal justice system is significantly greater than that found in the general community.[1,2] As presented in Chapter 11.4, for example, the per capita rate of psychotic illness in prisons is approximately 10 times greater than that found in the general community. Tragically, relatively few services exist that provide continuity of mental health care between gaols and the community.[3] This produces a situation where individuals whose mental illness may have been identified and treated in gaol find themselves without services in the community. Typically, only when in crisis do they find their way into general psychiatric services either in community settings or in hospital. This situation has produced considerable stress on already taxed mental health services.[4]

Given the prevalence of offence histories among psychiatric patients, it is important for mental health professionals to be aware of the unique issues—and myths—that accompany patients with offence histories. At the outset it is important to emphasize that the duty of mental health services is to address mental health issues. That ought to be the focus of mental health services. As this chapter makes clear, though, for some patients, there is a relationship between the mental illness and offending and by addressing the mental illness, the risk of re-offending might well be reduced. Moreover, many of the ancillary issues that lead to relapse and destability in psychiatric patients also may lead to offending. Addressing these issues will both help provide long-term stability for patients and will help reduce their risk of offending. As a result, there is a need for general mental health services to acquire expertize to identify and manage patients with offending histories.[5]

This chapter will provide information about the relative risk of offending among psychiatric patients and the relationship (or lack thereof) of inpatient aggression and community-based violence and offending. A framework will be provided for assessing and treating patients with offending histories and issues using a typology of mentally ill offenders. The role of forensic mental health services in bolstering general psychiatric services, and in sometimes providing primary care for mentally ill offenders, will also be discussed.

How many patients have criminal histories?

Surprisingly little research exists that investigates the number of patients entering general psychiatric services who have an offence history. For reasons having to do with privacy, lack of perceived relevance, and professional reluctance, general mental health services do not consistently obtain reliable information regarding patients' offence histories. This is often the case even when the patient has a current community-based corrections order. The two following studies can help shed light on the question of how many general psychiatric patients have histories of criminal offending.

In a study that was conducted to investigate the post-discharge violence of psychiatric patients and the predictive validity of risk assessment measures among almost 193 involuntarily committed psychiatric patients in British Columbia, Canada who were discharged to the community, Douglas, Ogloff, Nicholls, and Grant[6] obtained official criminal histories for all patients who had ever been arrested or convicted of any criminal offence. The vast majority of patients had prior psychiatric hospitalizations (n=184, 95 per cent). Informally, members of the hospital staff were asked what percent of patients they believed had a prior criminal history. Staff, including psychiatrists, estimated that a very small percent of patients would have been arrested or convicted of offences—less than 20 per cent. The review of criminal histories, however, showed that 64 per cent (n=123) of patients had previous arrests or convictions for any type of criminal offence, including 40 per cent (n=78) who had been arrested or convicted of violent offences.

In an Australian study based upon Victorian samples of cohorts of patients with schizophrenia, Wallace and colleagues have found that almost 22 per cent of patients with schizophrenia have a history of offending at some point in their lives.[7] Moreover, eight percent of patients with schizophrenia had a criminal conviction for a violent offence. These percentages increased three-and-four fold when the patients with schizophrenia also had a known substance abuse problem. In a recent study, Hodgins and Muller-Isberner[5] found that one quarter of patients discharged from a general mental health service had a criminal record.

While it is difficult to know exactly how many psychiatric patients across different services have committed offences, the point that may be drawn from the above research suggests that many patients have offence histories—likely more than mental health professionals would expect. The starting point of the chapter, therefore, is that while most psychiatric patients will not have violent criminal histories, many will have offence histories, including the commission of violent offences. Moreover, many more patients will have exhibited violent behaviour that did not lead to arrest or conviction. Therefore, even if they do not realize it, all psychiatrists and other mental health professionals have experience working with patients who have offence histories.

What leads mentally ill people to offend?

Although the reasons that anyone—including psychiatric patients—offends are myriad and complex, a typology of mentally ill offenders is helpful for understanding the reasons they offend.[1] There are three general categories of people with mental illness who offend; understanding the general mentally disordered offender type will enable clinicians in general psychiatric services to provide appropriate treatment. The first, and smallest group, includes those psychiatric patients for whom a necessary and sufficient cause of their offending is the presence of their mental illness and the symptoms the illnesses produce. The second group includes patients who do not offend because of their mental illnesses, *per se*, but due to the concomitant social difficulties that all too often accompany mental illness. The final general group of offenders with mental illness include those patients whose offending occurs irrespective of their mental illness. Each of these groups will be described below.

Patients who offend because of their mental illness

This group is likely the smallest of the three groups. This group includes people who may not be criminally responsible because, as a result of their mental illnesses, they do not know what they are doing, or do not appreciate that what they are doing is wrong. Their offences occur as a direct result of the mental illness. But for the mental illness and the presence of symptoms which led to the patient's offending behaviour, the crime would not have occurred. Their mental illness is both a necessary and sufficient explanation for their offence. They only offend when they are acutely unwell and the offence behaviour is a product of their mental illness (e.g. acting on delusions or hallucinations). Depending upon the jurisdiction in which they reside, they may be found not guilty by reason of insanity or mental illness. They most likely will be housed in secure hospitals rather than prisons following legal adjudication. Typically the illnesses that are present in people who fall into this category are psychosis or serious affective disorders accompanied by psychosis. Many jurisdictions that retain some form of insanity defence specifically exclude the use of the defence by those with antisocial or dissocial personality disorder.

Patients who offend as a result of the sequelae of mental illness

The second general group of psychiatric patients who offend comprises hose whose mental illnesses are a necessary but not sufficient explanation for their offending. It is by far the largest group of psychiatric patients who offend. As is typical for many patients with serious mental illnesses, these patients begin to spiral downward socially as a result of their mental illnesses. They can become estranged from family and pro-social support networks. Their lives become unstable; housing, basic needs, and their need for non-judgmental personal support may go unmet. They may end up being accepted by groups of people who are themselves unstable. They often resort to engaging in illicit drug abuse. These social factors contribute to their resultant offending. While their mental illness may be a catalyst in the course of events that lead to the offending, the mental illness itself is not the direct cause of the offending. Had they not had a mental illness, they likely would not have begun offending. However, by the time they develop offending behaviour, their lives have become so disorganized and their maladaptive coping and survival strategies have become so entrenched as to make the reversal of these processes difficult over the long-term. Psychiatric treatment, while a necessary starting point, will not be sufficient alone to eliminate the offending behaviour.

Patients who offend despite their mental illness

The final group of patients are those who would offend irrespective of the fact that they have a mental illness. Although not as large a group as the one above, many more patients who offend fall into this category than into the first. The fact that they have a mental illness is neither a necessary or sufficient explanation for their offending. Patients in this group are typically characterized by early onset antisocial and illegal behaviour. They differ from other mentally ill offenders by having a pervasive and stable pattern of offending regardless of their mental state.[5] This behaviour almost always precedes the onset of mental illness. While people with a psychopathic or dissocial personality disorder will be included in this group, most of the people in the group will not be so disordered. It is important to acknowledge, though, that the broad range of people that may fall into this group, including the psychopaths, may well develop psychiatric illnesses. We must avoid the tendency to deny this group proper services or to acknowledge their mental illnesses. These patients' mental illnesses may well exacerbate their offending or lead to unusual offending; however, even when they are asymptomatic they may continue to offend.

Aren't psychiatric patients with offence histories unusually burdensome or too dangerous for mental health services?

The perception all too often still exists that patients with offence histories are unusually burdensome or even too dangerous to be seen by general mental health services. While there are doubtless patients, largely those drawn from the third group above, who are burdensome and even dangerous, in the main patients with offence histories are nether unduly burdensome nor dangerous. For example, in a recent prospective study of violence among discharged general and forensic psychiatric patients, Doyle and Dolan[8] found no

[1] Readers are referred to Chapter 11.3.1 (Associations between Psychiatric Disorder and Offending by Thomson and Darjee) for additional information regarding the relationship between mental illness and offending.

significant differences in post-discharge violence rates (both official and unofficial) between patient groups in the UK. Ogloff and colleagues have obtained similar results from separate studies of post-discharge violence among samples of general and forensic psychiatric patients.[6]

One of the concerns expressed in general psychiatric services about patients with offence histories is the risk for aggression and violence they might present during hospitalization. It is often assumed that if a patient has an offence history, particularly one marked by aggression, that the patient will be more likely to be aggressive in hospital. Research suggests, however, that this may not be the case. It is true that over the entire period of hospitalization patients who have more psychopathic traits might have higher rates of aggressive incidents.[9] Research shows that in fact there is no significant relationship, at least for forensic psychiatric patients, between aggression in hospital, aggressive behaviour preceding admission, or violent recidivism.[10]

Analyses of what leads to aggressive behaviour by psychiatric patients suggests that dynamic (highly changeable) factors are responsible and that a functional analysis of inpatient aggression shows that rarely are the acts related to general patient aggression or purely to the patients' mental state.[11] Rather than assuming that patients with forensic histories will be any more or less aggressive than other patients, recent instruments have found useful in assessing patients risk for inpatient aggression.[12] Such instruments should be employed.

Assessment of psychiatric patients who offend

Prior to commencing ongoing mental health care to patients with offence histories, it is important that a comprehensive assessment be conducted, preferably by a psychiatrist or clinical psychologist with expertize and experience in forensic mental health. In some jurisdictions, mental health services may be able to draw upon forensic mental health services to obtain secondary assessments of the patients to assist with assessment and treatment planning.[13] The assessment must address three major components: mental health, substance use, and the presence of criminogenic factors (i.e. factors that increase the likelihood that the patient will re-offend).

First, a thorough mental health assessment is required that includes both a review of the patient's current mental state as well as their psychiatric history. Although seemingly straight-forward, this can be difficult with some patients who have offence histories. All too often now we see young people, usually males, whose mental illnesses are only identified upon admission to gaol or prison.[3] As such, it may be difficult to obtain reliable information about the genesis and onset of these people's mental illnesses.

The second component that must be considered is whether the patients have a substance use or dependence disorder and what role substances have on their mental illness and offending. In mental health generally,[14] and in patients with offence histories in particular,[15] high percentages of patients are substance abusers. Ogloff and colleagues found that 74 per cent of patients in the secure forensic hospital in Victoria, Australia had a lifetime history of substance abuse or dependence. The presence of a substance use disorder is a key risk factor in determining which patients will re-offend (or have a relapse of their illness for that matter). Unfortunately, very often substance use disorders, and their effects on patients, are overlooked in the routine assessment and treatment of patients with mental illnesses.

The final area that must be considered in a comprehensive assessment is the presence of so-called 'criminogenic factors' present in the patient's case. This concept is part of a contemporary well accepted and supported theory of offending known as the Psychology of Criminal Conduct, which was developed by Andrews and Bonta in the 1980s and it has been refined over time.[16] It is a theory concerned with individual differences and variability in criminal behaviour, making it a particularly useful guide for both assessing the risk of reoffending and planning rehabilitation attempts. This emphasizes the complexity of criminal behaviour, thereby acknowledging the contributions of social context, biology, and psychopathology. Criminogenic factors are the subset of dynamic (changeable) risk factors that have been found to relate directly to a risk for re-offending. They are therefore modifiable characteristics, whereby a change in the risk factor equates with a change in the risk of re-offending. These are factors that can affect patients with mental illness just as they can affect people with no mental illness who offend. Examples include having friends who are criminals, developing pro-criminal attitudes, having an anti-social personality, having limited problem-solving skills, and having difficulties controlling anger and hostility (Ogloff & Davis, 2004).[17]

To assist in assessing Andrews and Bonta[18] have developed the *Level of Service Inventory, Revised* (LSI-R), which assesses the presence of criminogenic factors as the basis for offender assessment and treatment.[1] The LSI-R consists of 54-items "grouped into the following domains or sub-components (with the number of items in parentheses): Criminal History (10); Education/Employment (10); Financial (2); Family/Marital (4); Accommodation (3); Companions (5); Alcohol/Drug Problems (9); Emotional/Personal (5); and Attitudes/Orientation(4).[18] While developed for general criminal populations, the LSI-R has been found very useful for assessing the presence of criminogenic factors and general needs of psychiatric patients with offending histories. Recent research findings show that a screening version of the LSI-R reliably identifies risk factors for patients in forensic psychiatric services.[19]

The presence of antisocial personality or dissocial personality is a criminogenic factor that must be considered in the assessment of mentally ill offenders. Unfortunately, antisocial personality disorder as it is defined by the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision[20] is vastly over-represented in psychiatric populations due to the nature of the criteria for the disorder which are essentially the presence of criminality.[21] Thus, great care must be taken to ensure that the diagnosis of antisocial personality does not rely solely on the fact that the patient has a history of offence behaviour. Instruments designed to reliably measure the presence of psychopathy, such as the *Psychopathy Checklist*, can be useful for assessing the aspects of personality and behaviour that comprise psychopathy.[21,22]

Following the assessment of each of the above components, it is necessary to develop a formulation that considers where the

[1] There is a revised version of the LSI-R which includes a section for case management planning, the Level of Supervision/Case Management Inventory and a version for young offenders (Youth Level of Service/Case Management Inventory; Hoge, Andrews, & Leshied, 2002).

patient's mental illness factors into their offending. Drawing on the three typologies of offenders with mental illnesses outlined above, the clinician can determine which category best describes the patient. Because the typologies are general, there will be overlap in characteristics for some patients. In addition to understanding the factors that help explain a patient's offending, the typologies are very important for determining what treatment and management strategies can be most effective for the patient.

The treatment of mentally ill offenders by general mental health services

The typology of mentally ill offenders will be revisited below with respect to the mental health and related services they require to assist with their treatment and management. The three-prong assessment strategy briefly described above will be helpful to identify the range of treatment needs the patient has. To be clear, the primary and even sole purpose of general mental health services is to treat patients' mental illnesses. However, it is useful to consider the presence of an offence history as an indication of the patient's functional impairment. Depending on the group into which the patient falls, the relative efficacy of mental health services alone varies in the extent to which it will satisfactorily address their mental health and offence issues. Most often, particularly for the latter two groups below, ancillary services and forensic mental health services will be required to help ensure the patient's long-term stability and to reduce the likelihood of offending.

Patients who offend because of their mental illness

Although this group of patients may commit horrendous acts, it is as likely that they engage in nuisance offences. Despite the particular type of offending behaviour in which they engage, perhaps surprisingly, their management by mental health services is oftentimes less complex than is the case for the other two groups of offenders with mental illnesses to be discussed below. Generally speaking, the treatment that this group requires is conventional mental health care. As it is the case that the primary cause of their offending behaviour is the mental illness, and the symptoms that it produces, addressing their mental health needs can serve to eliminate the offending. General mental health services are generally well equipped to deal with these patients, though they may feel reluctant to do so. Very often, patients will respond to medication and with supervision their mental state will begin to improve. If treatment in the cases is complex, it will often be because of the mental illness itself. For example, the patients may have chronic psychosis which is refractory to psychiatric treatment.

While the patient's mental illness is the main cause of the offending behaviour in this category, related issues will need to be addressed to stabilize the patient over the long term and to further reduce the patient's risk of re-offending. The LSI-R, noted above, will be particularly helpful in identifying such issues. Common issues include substance abuse, life skills, housing, financial support, and personal support. Services to address these issues will need to be organized to effect long-term psychiatric and behavioural stability.

Patients who offend as a result of the sequelae of mental illness

Just as the complexity of the reason this group offends is greater than with the first group, the treatment they require to stabilize is also more complex. This group is characterized by general disorganization and social damage. As such conventional mental health services alone will have relatively limited effect on patients' mental state and stability over the long-term. Even if psychiatric treatment is effective in the short-term, patients in this group will be likely to return to a chaotic life which eventually may include a return to offending. Nonetheless, the treatment of these patients' psychiatric illnesses is the central component of their care.

The comprehensive assessment approach outlined above will be particularly useful in determining the range of issues beyond mental illness that affects the patients and contributes to their offending behaviour. In particular, the areas of concern identified by the LSI-R are particularly important for informing intervention need. For example, if employment issues, financial issues, accommodation needs and alcohol/drug problems are revealed in the assessment, these issues, in addition to the patient's mental illness will need to be addressed. Not only will addressing these issues satisfactorily lead to a reduction in the patient's risk of re-offending, but it will assist with ensuring stability in mental state over the long-term.

Generally speaking, the greater the number of criminogenic factors that arise from the assessment, the more intensive treatment will need to be to ensure long-term stability. To the extent that needs arise that cannot be addressed directly by the mental health service, these services will need to be sought from appropriate providers in the community. This is where effective case management and service brokerage is critical. All too often psychiatric patients revolve in and out of general mental health services (and the criminal justice system); all the while their underlying needs are not identified or addressed. The vast majority of offenders with mental illnesses can be properly treated and managed by general mental health services if only their related needs and issues can be addressed. Moreover, given that there is a relationship between these patients' mental illnesses and their offending, addressing the mental illness and related matters can help lead to a reduction in offending, although that will not be the purpose of providing them with general mental health services.

Patients who offend despite their mental illness

As with all of the categories of offenders, this group will still require comprehensive mental health care; however, the mental health care will be essentially futile in reducing the patient's proclivity for offending. It is important to note that mental health services still have an obligation to treat these patients' mental illnesses, but addressing their offending issues will be beyond the scope of care or even the expertize of general mental health services. Moreover, addressing the ancillary issues that arise to affect their mental illnesses will be less likely to reduce their offending risk than would be the case for either of the other two groups above. Where possible, patients in this group should be seen by forensic mental health services and they will be candidates for offender rehabilitation programmes offered by contemporary correctional services (in prisons or in the community).

It is important to note that the cautious approach advised is not intended to dissuade services from providing adequate psychiatric care, but to recognize their limitations and to reduce the sense of failure and frustration that occurs when treating patients with such an offence pattern. Despite the nihilism that sometimes exists, there is an ever expanding corpus of firm empirical support to

show that offender rehabilitation can help reduce recidivism,[17,23] however, such services are beyond the scope of general mental health services. Over time, of course, as these patients' offending issues may be successfully addressed, they will be appropriately cared for by mental health services just as other people are. Once an offender not always an offender!

The role and support of forensic mental health services

The thorny question of when forensic mental health services, as opposed to general mental health services, ought to be responsible for a patient's psychiatric care is difficult and depends much on the jurisdiction, the relevant legislation, policies, and practices that are in place. It is never the case that all patients who offend require, or should have access to, forensic mental health services. For the most part, the goal should be to maintain patients in general mental health services to ensure continuity and normality of care. Realistically, though, given the relatively high rate of psychiatric patients who offend or who have offence histories, the level of knowledge and awareness of offence issues among general mental health professionals must increase.

It is still the case in many settings that the mere mention that a patient has a forensic history raises angst and concern about the capacity of general mental health services to care for the individual. This is most often nonsensical, particularly because a relatively high percentage of psychiatric patients in general services have an offence history—whether or not it is known by the service. Ideally, general mental health services should adequately address patients' mental health and ancillary issues to an extent that would actually prevent patients in the first two groups of mentally ill offenders from offending. Very often we in forensic mental health services see patients who had contact with general mental health care yet their problems were exacerbated, they deteriorated and went on to offend. Oftentimes inadequate assessment and identification of the patients' needs is at the core of the shortcoming of their care.

Realistically, despite the oftentimes excellent care provided to psychiatric patients, forensic mental health services will be required. In the first instance, patients found not criminally responsible or who require involuntary treatment during incarceration should be provided service in appropriate forensic mental health facilities. Psychiatric patients with complex presentations, including myriad criminogenic issues, likely will require at least secondary consultation from a forensic mental health service.[13] The best models that exist internationally have forensic mental health services take responsibility for the complex offence-related cases initially with the introduction and eventual transition to general mental health services over time once the patient's offending issues are addressed and stability is realized.[24,25] Continuing general care is the preferred modality of care except in the most difficult cases.

Conclusions

General psychiatry has an important role in providing care to all patients, and under most circumstances this includes patients with offending histories and issues. Given the relatively high rates of offending among psychiatric patients, whether they realize it or not, general services have considerable experience with this patient group. All too often, though, general mental health clinicians do not have adequate training or expertize to systematically assess patients to determine what factors have lead to their offending. The typology of patients who offend presented in this chapter can prove useful for determining the factors that must be addressed to treat the patient. Moreover, an indication will be made as well when assistance may be required by specialist mental health services.

Further information

Hodgins, S., and Muller-Isberner, R. (2004). Preventing crime by people with schizophrenic disorders: The role of psychiatric services. *British Journal of Psychiaty,* **185**, 245–50.

Ogloff, J. R. P., and Daffern, M. (2006). The Dynamic Appraisal of Situational Aggression: An instrument to assess risk for imminent aggression in psychiatric inpatients. *Behavioral Sciences and the Law,* **24**, 799–813.

Ogloff, J. R. P., and Davis, M. R. (2004). Advances in offender assessment and rehabilitation: Contributions of the risk-needs-responsivity approach. *Psychology, Crime and Law,* **10**, 229–42.

References

1. Brugha, T., Singleton, N., Meltzer, H., *et al.* (2005). Psychosis in the community and in prisons: A report from the British National Survey of Psychiatric Comorbidity'. *American Journal of Psychiatry*; **162**, 774–80.
2. Fazel, S., and Danesh, J. (2002). Serious mental disorder in 23,000 prisoners: A systematic review of 62 surveys. *Lancet*, **359**, 545–50.
3. Ogloff, J.R.P. (2002). Identifying and accommodating the needs of mentally ill people in gaols and prisons. *Psychiatry, Psychology, and Law*, **9**, 1–33.
4. Torrey, E.F. (1992). Criminalizing the seriously mentally ill: the abuse of jails as mental hospitals. Washington DC: Report of the National Alliance for the Mentally Ill and Public Citizen's Health Research Group.
5. Hodgins, S. and Muller-Isberner, R. (2004).Preventing crime by people with schizophrenic disorders: The role of psychiatric services. *British Journal of Psychiatry*, **185**, 245–50.
6. Douglas, K.S., Ogloff, J.R.P., Nicholls, T., *et al.* (1999). Assessing risk for violence among psychiatric patients: The HCR-20 Violence Risk Assessment Scheme and the Psychopathy Checklist: Screening Version. *Journal of Consulting and Clinical Psychology*, **67**, 917–30.
7. Wallace, C., Mullen, P.E., Burgess, P., *et al.* (2004).Criminal offending in schizophrenia over a 25-year period marked by deinstitutionalization and increasing prevalence of comorbid substance use disorders. *American Journal of Psychiatry*, **161**, 716–27.
8. Doyle, M., and Dolan, M. (2006). Predicting community violence from patients discharged by mental health services. *British Journal of Psychiatry*, **189**, 520–6.
9. Nicholls, T.L., Ogloff, J.R.P., Douglas, K.S., *et al.* (2004). Assessing risk for violence among male and female civil psychiatric patients: the HCR-20, PCL:SV, and McNiel & Binder's screening measure. *Behavioral Sciences and the Law*, **22**, 127–58.
10. Daffern, M., Ogloff, J.R.P., Ferguson M., *et al.* (2007). Appropriate treatment targets or products of a demanding environment? The relationship between aggression in a forensic psychiatric hospital with aggressive behaviour preceding admission and violent recidivism. *Psychology, Crime and Law*, **13**, 431–41.
11. Daffern, M., Howells, K., Ogloff, J., *et al.* (2007). The interaction between individual characteristics and the function of aggression in forensic psychiatric inpatients. *Psychiatry, Psychology and Law*, **14**, 17–25.

12. Ogloff, J.R.P., and Daffern, M. (2006). The Dynamic Appraisal of Situational Aggression: An instrument to assess risk for imminent aggression in psychiatric inpatients. *Behavioral Sciences and the Law*, **24**, 799–813.
13. Warren, L.J., MacKenzie, R., Mullen, P.E., *et al.* (2005). The problem behaviour model: The development of a stalkers clinic and a threateners clinic. *Behavioral Sciences and the Law*, **23**, 387–97.
14. Mueser, K.T., Yarnold, P.R., Rosenberg, S.D., *et al.* (2000). Substance use disorder in hospitalized severely mentally ill psychiatric patients: prevalence, correlates, and subgroups. *Schizophrenia Bulletin*, **26**, 179–92.
15. Ogloff, J.R.P., Lemphers, A., Dwyer, C., *et al.* (2004).Dual Diagnosis in an Australian forensic psychiatric hospital: Prevalence and implications for services. *Behavioral Sciences and the Law*, **22**, 543–62.
16. Andrews, D.A., and Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Anderson Publishing Co., Cincinnati, OH.
17. Ogloff, J.R.P., and Davis, M.R. (2004). Advances in offender assessment and rehabilitation: Contributions of the risk-needs-responsivity approach. *Psychology, Crime and Law*, **10**, 229–42.
18. Andrews, D.A., and Bonta, J. (1995). *The level of service inventory – revised*. Multi-Health Systems, Toronto.
19. Ferguson, A. M., and Ogloff, J.R.P. (in press). Predicting recidivism in an Australian mentally disordered offender population with and without comorbid substance abuse using the LSI-R:SV. *Criminal Justice and Behavior.*
20. American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th edn.), Text Revision. Author, Washington DC.
21. Ogloff, J.R.P. (2006).The Psychopathy/Antisocial Personality Disorder conundrum. Australian and New Zealand *Journal of Psychiatry*, **40**, 519–28.
22. Hare, R.D. (2003). *Manual for the Hare Psychopathy Checklist* (2nd ed.). Revised. Multi-Health Systems, Toronto, Ontario.
23. McGuire, J. (2002). Criminal sanctions versus psychologically-based interventions with offenders: A comparative empirical analysis. *Psychology, Crime, and Law*, **8**, 183–208.
24. Simpson, A.I.F., Jones, R.M., Evans, C., *et al.* (2006). Outcome of patients rehabilitated through a New Zealand forensic psychiatry service: A 7.5 year retrospective study. *Behavioral Sciences and the Law*, **24**, 833–43.
25. Skipworth, J., and Humberstone, V. (2002). Community forensic psychiatry: Restoring some sanity to forensic psychiatric rehabilitation. *Acta Psychiatrica Scandinavica*, **106**, 47–53.

11.17

Management of offenders with mental disorder in specialist forensic mental health services

Pamela J. Taylor and Emma Dunn

Philosophy and theoretical models

Specialist forensic mental health (fmh) services are for people with serious mental disorders and grave offending behaviour who tend to be rejected from mainstream services. Although often triggered by single high profile cases, these specialist services are among the best planned and commissioned services in psychiatry, founded in evidence of need, risk and efficacy of interventions. They are grounded in a multidisciplinary clinical perspective and often have integrated academic units. They interface both with other clinical services and with the criminal justice service. Good relationships with the local community are vital for establishment and growth.

Mentally disordered offenders have been sources of tension between services at least since the early 19th century. In Britain, the Lunacy Commission argued that it was 'highly objectionable' that offender patients should be detained in a general lunatic hospital, while an 1807 parliamentary select committee noted that 'to confine lunatics in a common Gaol is equally destructive of all possibility of the recovery of the insane and the comfort of other prisoners'.[1] High security hospitals followed about 50 years later.

Funding and commissioning of fmh services worldwide continue to follow oscillations between considered responses to changes in the structure and availability of general services, and responses to single notorious cases. For England and Wales, the Butler Report[2] considered the then increasing gaps in service provision as psychiatric services shifted from mainly institutional to mainly community care. It was the most thoughtful and powerful driver of modern forensic mental health services in England and Wales, presaging the arrival of medium security hospital units. The contrasting path, of case driven service development, is illustrated by the so-called 'Dangerous and Severe Personality Disorder' (DSPD) services,[3] driven by the inquiry into the care and treatment of Michael Stone[4] following his conviction for two homicides and an attempted murder, occurring soon after his discharge from a psychiatric hospital. In the USA, mandated community mental health law (Kendra's Law) and treatment programmes in New York followed a subway killing by a psychotic man,[5] while legislative change necessary for specialist service development in Japan followed a school massacre.[6] Concern that single cases make poor law – and poor health reforms—is tempered by the mutual commitment of government agencies and of practitioners to keep offender patient services under review.

The Department of Health and Home Office report, for England,[7] is a good example of such review, proposing five principles for secure healthcare provision:

i) quality of care and proper attention to individual needs;

ii) community rather than institutional care where possible;

iii) security no greater than justified by the danger presented—to self or others;

iv) maximization of rehabilitation and chances of sustaining an independent life in the longer term;

v) proximity to the patient's own home or family if s/he has them.

It is arguable that only the fifth principle requires an evidence base. Intra-familial violence may contribute to mental health and/or behavioural difficulties, and most violence by people with mental illness occurs within their close social circle.[8] Nevertheless, people have attachments, and the fifth principle is retained for that reason—*and* as a convenient way of anchoring responsibilities for services. The first four principles embody the medical ethic of maximizing autonomy and anticipated the Human Rights Act 1998, which gives effect to the rights under the European Convention to *liberty and security of person* (article 5) and prohibits *degrading treatment* (article 3). The Act also emphasizes proportionally—if it is necessary to breach a right, that breach should not go further than necessary.

Organizational models often founder in a clash between the needs of service users and providers. The ideal is a fully integrated services in which service users move freely between forensic and general mental health services, according to need.[9] General services, however, tend to have to focus on crisis management, and the greater the tensions between specialist and general services the greater the likelihood of ever-longer periods of residency in a physically secure hospital[10] or default to parallel service delivery

reasons for this include reduced availability of non-secure psychiatric beds, perverse incentives in funding in which those in highest security may be funded centrally with minimal local funding burden. Also government caution may dictate that once detained in medium or higher security, every individual must progress stepwise through lower levels of security before returning to the community. Evidence does not support the notion that this stepwise route reduces criminal recidivism,[10] but it has led to the growth of additional tiers of specialist security provision, including 'low security' hospital units and forensic community mental health teams.

The international context

There is insufficient space to explore the international context in any detail. Laws on criminal responsibility, criminal justice and mental health vary between European countries[11] and elsewhere, but many underlying clinical principles are shared. Most countries acknowledge some association between mental disorder and offending behaviours, but there is variation in how this influences prosecution, fitness to plead and stand trial, the extent to which mentally disordered offenders may be regarded as wholly or partially without responsibility for a criminal act, and the extent to which they are treated in mainstream or specialist secure health services, or in prisons, albeit with some health service input.

Our international research group (SWANZDSAICS) drawn from culturally distinct jurisdictions across five continents (Sweden, Wales, the Australian state of Victoria, New Zealand, Denmark, the South African province of Western Cape, Japan, the Canadian province of Quebec, and Scotland) finds a shared therapeutic philosophy in managing offenders with psychosis, but struggles to be therapeutic with sex offenders or people with personality disorder.[12] Other countries however, most notably the USA, seek to separate the business of psychiatry in the courts from any therapeutic endeavour with mentally disordered offenders.[13]

The nature of security

Secure psychiatric hospitals have two overarching aims: improving health and delivering safety for patients and others. In secure hospitals, patients' autonomy is limited in a number of important ways: they may not be allowed to leave the hospital at all, may be confined to a particular area within the hospital, and/or treatment may be enforced. Although these restrictions are undoubtedly at least partly in the interests of the patients themselves, they are commonly also in the interests of others.

The elements of security

Security in a clinical setting is made up of four main elements: physical, procedural, and relational security, *and* treatment. Treatment, including (re)habilitation, becomes vital to safety and security whenever a clear pathway can be shown between a mental disorder and offending behaviour.

Physical security refers to the qualities of buildings: the nature of perimeter walls and internal structures and functions. High security requires at least one high and distinct perimeter wall or fence clear of the main hospital building. In medium-security, the walls of the building alone generally provide the main perimeter, with high fences only surrounding exercise areas which are not entirely within the main building. All specialist security hospitals provide for staff and visitor entry through an 'airlock', using independent locking systems, the external one generally controlled by dedicated security or administrative staff. Ideally, clinical staff contributing to building design, which should ensure good sightlines throughout, while allowing residents a sense of privacy. Each patient has his/her own room, and ideally holds a key to it (with a staff over-ride potential). This enhances his/her safety and sense of personal security, and also the safety of property. In high-security units, cameras may be used for continuous monitoring. The environment should be pleasant, enabling both patients and staff to feel comfortable; small frustrations often trigger violence.

Procedural security provides for a formal set of checks for factors thought to be associated with risk of harm by patients. This includes minimizing patient access to weapons, fire-setting materials, or potentially disinhibiting substances, and preventing absconding. In high security, any communication with the outside world may be monitored; at lower levels of security, such monitoring is determined case by case. Procedures should also guard against potential harm to each patient. Some measures used to prevent or control violence may have 'side effects'. Time-out and seclusion may be necessary, but can be provocative and open to abuse. Physical restraint may sometimes be essential, but if done incorrectly or brutally may damage the possibility of a therapeutic relationship, physically harm or even kill the patient. In the UK, procedures for such measures are subject to guidance both from professional bodies[14] and legislative Codes of Practice (e.g.[15]).

There is insufficient space to detail the extensive range of procedures for ensuring security, so a couple of examples—searching and screening of contacts—must suffice. Searches of the person and of the environment are conducted mainly to minimize access to drugs and weapons. The level of unit security dictates the nature, extent and frequency of searches. In English high security hospitals, no-one is trusted. Staff, professional visitors and social visitors are all searched on entry; many items—such as mobile/cell phones are forbidden anywhere in the hospital. Patients may be searched randomly, but also when moving between areas in the hospital or if there are particular grounds for suspecting they have secreted something that could become a weapon, or acquired drugs. At any security level, patient rooms and other areas may be searched—similarly, randomly or on specific grounds; in high security, patients' possessions are routinely restricted in quantity to facilitate searching. For all such occasions, however, procedures incorporate measures which reflect concern for the individual being searched. Patients must be informed of searches (immediately beforehand if randomly timed) and invited to observe.

Screening of contact with visitors is multifaceted. Visitors may be enticed into aiding absconsion, or be irresponsible in their 'gifts' for the patient; apparently innocuous items may be fashioned into weapons, and they may be under pressure to bring drugs, perhaps disguised in food. There may also be risk of harm to visitors. Telephone calls, mail and personal visits may all be observed, but only in accordance with written procedures. Policies pertaining to visits will refer to classes of visits—for example visits by specific individuals who may threaten or be under threat, or by children, Such visits must be supervised by specifically trained staff.

Relational security skills are founded in therapeutic approaches and, with specific treatments, form the core of hospital security, clearly demarcating hospitals from prisons. It lies in extensive knowledge of each patient, accurate empathy and highly developed

capacities for communicating and working in a clinical team. At best, it not only provides immediate safety and the milieu for change, but it may also facilitate lightening of physical and procedural securities. Effectiveness, however, is reliant on sufficient numbers of adequately trained staff.

Relational security may, however, create anxiety in hospital managers and their political masters, partly because it is more difficult to understand as security than locks and walls, but also due to the perception that its corruption is possible and difficult to predict. Over time, staff may be vulnerable to potentially counter-therapeutic change.[16] Strategies to ensure maintenance of clinical integrity therefore include personal supervision and appraisal, peer review and audit of team- and hospital-wide practice. Access to psychodynamic psychotherapists is not only, or even primarily, for the patients, but also for the staff and the institution.[17]

Treatment as security targets the link between symptoms of mental disorder, most obvious for psychotic symptoms, and criminal or risky behaviour.[18] In contrast to prisons, secure hospitals generally select residents for their treatability. It seems simple then—specific treatment with antipsychotic medication for people with psychosis should bring safety—but matters are rarely so straightforward. Multiple diagnoses are common: at least 25 per cent of offender patients with psychosis have personality disorders established before onset of their psychotic illness, and many abuse alcohol and/or other drugs at levels to qualify for a diagnosis.[18] Over 25 years, an increasing proportion of English high security hospital patients were found to have substance misuse disorders,[18] especially affecting the psychosis-personality disorder group. Substance misuse not fully meeting diagnostic criteria is also common. In the short term, specific treatment for psychosis combined with preventing access to substances of abuse can restore safety. For longer term success and safety, specific treatments aimed at substance misuse are best integrated as part of the overall treatment package,[19] although this is still not common practice (e.g. UK,[20] and Sweden,[21]). This may partly explain the counterintuitive finding[22] that, in the UK at least, there is a preference for admitting people with 'pure' psychosis to medium security hospitals, even though substance misusing people with psychosis would be regarded as a higher risk group (e.g.[23]).

Hospital security in practice

In England and Wales, *high security hospitals* are mandated in statute for those patients who pose an imminent risk of serious harm to 'the public'. Public safety is often construed merely as removing dangerous persons from the community and strictly confining them. During the early stages of treatment adequate protection for staff and other residents is also vital. As patients become apparently safe, it is important that they can be tested out before discharge to the community, but government restrictions often make this difficult, so that patients may arrive at lower security levels ill-prepared for the new challenges; this may be dangerous in itself.

Medium security can be defined only in relation to high and low security services above and below it. The range, quality and quantity of low secure and open provision varies from place to place, so that providers of some medium secure service may have to retain patients longer than other providers and/or provide more parallel services.

The most constant aspect of medium security services is the in-patient unit. 'The aim is for the building design to support the nursing staff, who are the main security barriers.'[24] Procedural security tends to be less stringent than in high security, but relational and treatment security are generally comparable. Nurse: patient ratios should be at least 2:1, and staff from a full range of other clinical disciplines should be available.

Low security relies almost entirely on relational security, but units generally have a locked door, and some procedures relevant to security, such as monitoring of substance misuse.

Service structure

Planning principles

'*Pyramid planning*' applies to specialist fmh services: the most intensive, specialized and secure services, most of which constrain patients' liberty and are the most costly, should be the smallest and at the top of the pyramid. At the base of the pyramid, and greatest in number, should be services linked directly to the local community. Trends in service provision in England and Wales reflect this model, with high secure bed numbers falling, and an increasing number of medium secure services (in mid-2007, 800 and 3 500 respectively for 55 million general population). Low security service provision is also increasing.

Private healthcare facilities. It is generally regarded as unethical to profit from the indefinite institutional detention of people. In Japan, where private provision of mental healthcare was the norm, recognition of service gaps for mentally disordered offenders resulted in the provision of publicly funded, purpose designed facilities for such patients.[6] In the UK, independent provision supplements a shortfall in specialist NHS facilities, providing just under half of all forensic beds. There are several reasons for this, not least the greater flexibility in service planning enjoyed by independent than public sector providers. Nevertheless, all places for detained patients in the UK are publicly funded and independent service providers are subject to the same levels of scrutiny as NHS facilities.

'*Super-specialist*' *provision* is necessary to provide specific treatments and the appropriate milieu for people with some special needs, such as female offender-patients, who almost invariably have suffered prolonged and serious neglect and abuse through childhood. Approaches founded in attachment theory[25] or trauma based work[26] go some way towards meeting their needs. People with a learning disability also benefit from dedicated services,[27] as may children and adolescents.

While there are advantages to specialist developments, most groups needing them tend to be small, and hence there are few units and it is difficult for residents to retain close ties with home, e.g. women resident in the only English high secure provision may be more than 350 miles away from home.

Planning service capacity is difficult. It is often done from utilization figures, but these tend to underestimate requirements by failing to recognize unmet need. While most developed countries have shifted general psychiatric services away from inpatient provision to the community, they have also seen an upsurge in availability of legal and illicit mind-altering substances. This has the probable effect of raising violence levels not only among the mentally healthy, but also among those with mental disorders (e.g.[28]).

Prison populations are growing in many countries. In the USA and the UK indefinite sentences 'for public protection' have been implemented. In England and Wales, the Lord Chief Justice and the Chairmen of the Parole Board have projected an additional

12,000 prisoners under this sentence alone by 2012, all of whom will need some sort of 'treatment' before they can be released[29] (at the time of writing a legal challenge to the lawfulness of such sentences in the absence of sufficient treatment is underway). Surveys of mental disorder among UK prisoners(e.g.[30,31]) suggest that neither generic nor specialist mental health services have kept pace with need in prisons even before this new expansion.

Capacity planning must be dynamic, taking into account assumptions about other services, and socio-economic differences between communities. Statistical models may help, (e.g.[32]) but it is more difficult to allow for changes in sentencing policy.

The service users: assessment, management, treatment and rehabilitation

Principles of assessment

Assessment in forensic psychiatry is considered in Chapter 11.14. In some countries, such as the USA, assessments for the courts are the main tasks for forensic psychiatrists, with some specializing in just one type of assessment and few providing treatment services.

Pre-admission assessment is important for patients destined for a secure hospital. It informs the initial treatment plan, and ensures a safe case-mix and appropriate staffing. Patients admitted to a secure hospital must have a mental disorder which meets legislative criteria for detention. Security requirements are then based on assessment of both mental disorder and criminal/violent behaviour. External factors must also be considered, for example, assessment of victim vulnerability. If, at one extreme, the identity of patient A's potential victim and the circumstances in which s/he might be at risk are clear, and that putative victim can acknowledge and co-operate with safety strategies, then it may be possible to manage risk of harm with little recourse to physical security. If, however, patient B threatens named people who cannot understand, accept or help manage the threat, then physical security may initially be required for her, even though both patients are otherwise similar. A serious, more generalized threat may also mean that initial assessment must be in security.

Admission criteria

Broad agreement on criteria for admission to a secure hospital bed[33] does not necessarily lead to consistency in practice.[22] In England and Wales, only 40–50 per cent of people referred for a specialist secure hospital bed, whether in high[34] or medium security[22] are actually offered one. There are common reasons: i) that some disorders are not widely considered to be treatable, and ii) the high level of security requested is thought not to be necessary. In spite of powers in England and Wales to detain people with personality disorder for assessment or treatment, many medium security beds are effectively closed to them. Few are now admitted to high security hospitals, except to the new purpose designed 'DSPD' units. It seems to be only in the Netherlands that there is a preference for treating people with personality disorder in secure hospitals.[35] There are few data on which to judge the quality of such assessments, but one study showed that 85 per cent of those refused a high security bed remained outside high security during the following 12 months, without incident.[36]

There have been efforts to attain standard ratings of security need. Cohen and Eastman[37] generated a set known as the Admission Criteria to Secure Services Schedule (ACSeSS). Factors considered relevant to admission were (* indicates factors most used in determining admission):

- gravity of recent* and past violence;
- likely immediacy of violence;
- psychopathology/developing behaviour possibly predictive of violence;
- special pathology, for example PD or sexual offending;
- the likely longevity of risk of violence*;
- predictability of the gravity and immediacy of potential future violence.

Collins and Davies[38] developed the Secure Needs Assessment Profile (SNAP), designed for *clinicians* to operationalize clinical judgement, following the first three elements of the security matrix outlined above—need for perimeter, and procedural or relational security.

The practice of using structured aids for clinical assessment of risk of harm to others or to self is growing, with nursing staff finding them particularly attractive. The HCR-20 (20 item Historical and Clinical Risk Management instrument;[39] includes dynamic risk factors, which can both inform management and provide a baseline measure from which progress can be monitored.[40]

Progressing along the pathway to treatment.

When someone has one or more serious mental disorders and poses a serious risk of harm to others, assessment may become a lengthy process. The elements of this must in turn be evaluated and adjusted according to interim outcome measures. Figure 11.17.1 summarizes a typical pathway. The initial assessment is directed at establishing safety—either through advice to others, or through admission to a specialist setting. In the latter, the assessment merges seamlessly into engagement into treatment, directed first at the primary disorder, before working with underlying factors—such as experience of childhood trauma.

The concept of 'treatability'. Under mental health legislation for England and Wales, patients legally classified as having *psychopathic disorder* or *mental impairment* alone cannot be detained for treatment in hospital unless 'such treatment is likely to alleviate or prevent a deterioration' of the condition. This intended safeguard against improper detention has sometimes been used to reject 'difficult' patients. New mental health legislation does not distinquish disorder types and puts the burden on availability of treatment.

Treatability should not be confused with curability. Assessment of treatability should be made in full awareness of the assessor's own attitudes and fears, with primary consideration given to the patient's psychopathology, needs, insight, and preferences and motivation for treatment. The timing of such an assessment may also be crucial. If the consequence of no treatment is more-or-less certain, for example, imprisonment, and this will most likely cause deterioration, then this in itself could justify admission for hospital treatment.

Principles of treatment. First line treatment of mental disorder in a secure setting generally differs little from that used elsewhere in mental health services. It is important to be aware, however, that some treatments have never been systematically evaluated among people whose mental disorders are complicated by violence. For example,

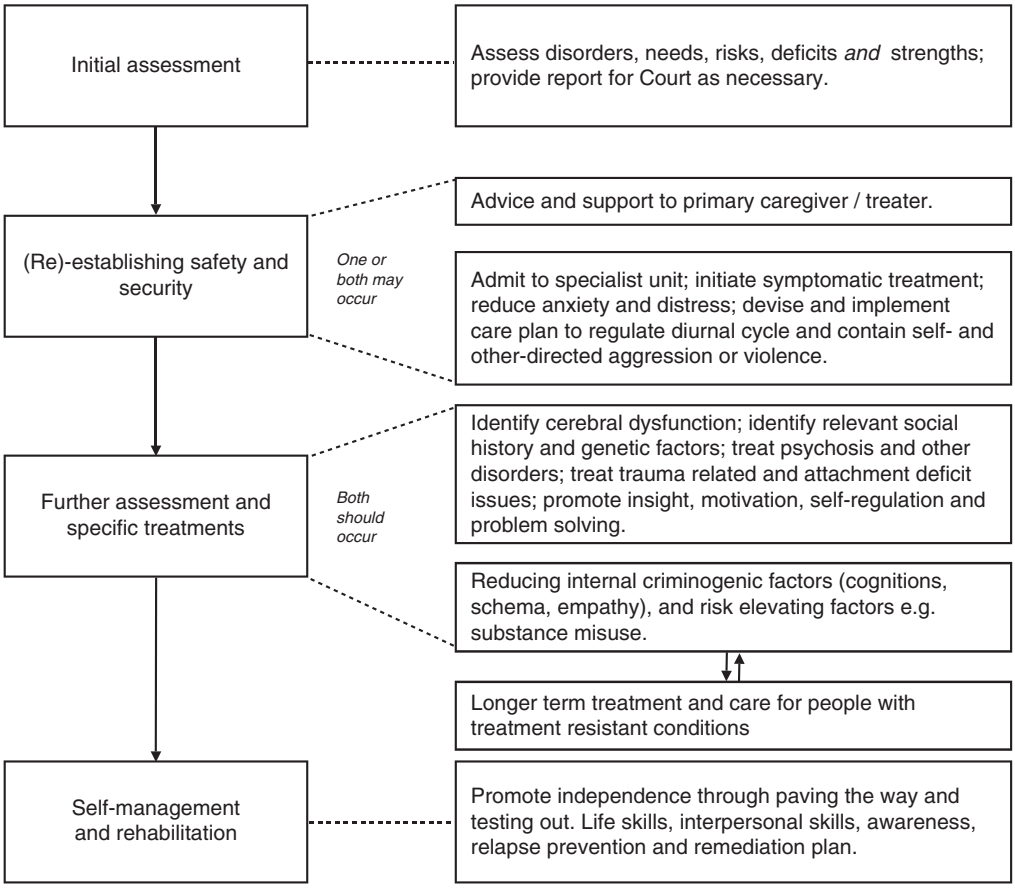


Fig. 11.17.1 Schematic representation of typical treatment and care pathway.

some consider that caution should be exercised in the use of cognitive behaviour treatment to alleviate psychotic symptoms.[41]

Delivery of treatment in secure conditions differs in some important ways from its delivery elsewhere. Intrusive interventions are not unique to forensic mental health services, but the length of application means that staff may need to engage in regular peer and external review to ensure their work remains effective and ethical. Areas for concern include the creation of a climate of compulsion and coercion, and the conflicting roles for psychiatrists using therapeutic skills for legal assessments and dangerousness predictions, as well as true therapeutic engagement.

The *inpatient environment* must be conducive to habilitation and rehabilitation. As well as treatment of specific disorders, the core concern of staff is aiding independent living. A sense of progress, evident in both the physical environment and staff attitudes, encourages hope and consolidates treatment gains. This sense of progress can be achieved in several ways:

- provision of graded living spaces in which patients can experience greater autonomy;
- increasing engagement in work, education and recreation;
- extending responsibility for personal and domestic care tasks;
- phased introduction of supervized activities outside the institutional boundaries, appropriate to the patient's mental state and public safety;
- consideration for and support of family and social relationships, and a basis for developing sustainable social networks. Intimate relationships constitute an important element of this aspect of the work.[42]

Successful return to the community depends on everyone having a sense of security. This will require suitable accommodation, basic financial provision and skills for managing finances, a sense of control over mental disorder, and sufficient mastery of local services to be able to find help in a crisis. Regular reviews of the service user, the services, their use of these services, and the goodness of fit between them, are essential. People regarded as posing a special risk of harm to others may be referred to a multi-agency public protection panel (MAPPP, [43, 44]; see also Chapter 11.16).

Direct involvement of the patient in setting goals for health and for safety creates in the patient a sense of ownership of resultant plans, and is likely to improve collaboration with them. The process of community re-entry is summarized in Fig. 11.17.2, drawing on a grounded theory of discharge from a secure hospital.[45] It lends itself to the development of outcome measures which are more appropriate for an offender *patient* than the almost exclusively used but over-simplistic measure of re-offending. The core concern is movement between pathological dependence and healthy independence, with staff taking an active role in facilitating progress in two phases—paving the way, when skills necessary to the attainment of independence are built up, and testing out, when

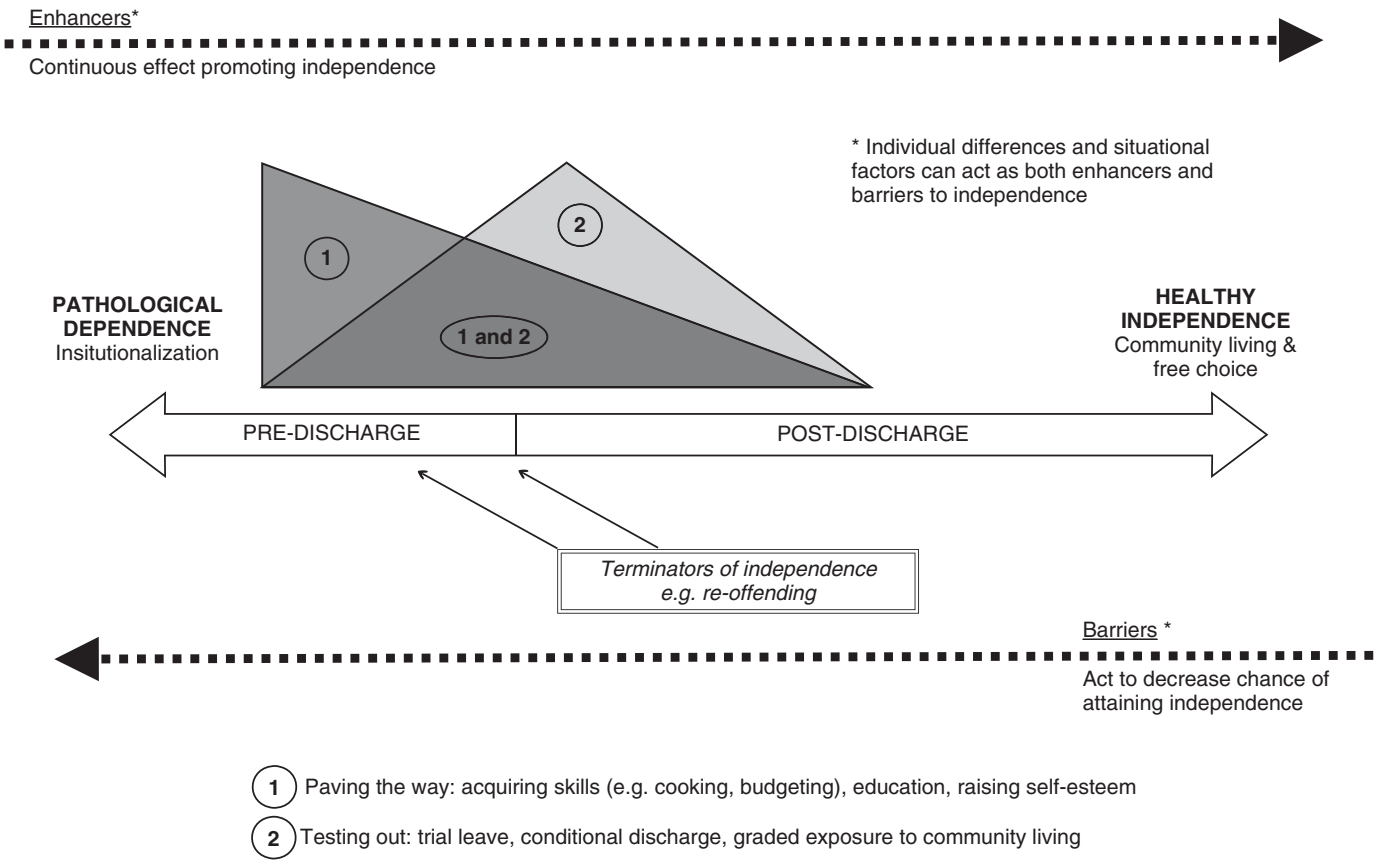


Fig. 11.17.2 Attaining healthy independence while recovering from serious mental disorder and offending: A model.

individuals are given opportunities to prove themselves in situations of increasing trust. Re-offending is important, but as a barrier to gaining or sustaining independence rather than the sole outcome indicator.

Assessing outcomes

Most studies of people who have been discharged from specialist security hospitals have serious limitations, which partly arise from political pressure to focus on the non-clinical outcome of re-offending and inadequate funding of the studies, which are costly and time-consuming. Results of existing studies are limited by:

1 Choice of outcome measures, which are almost solely re-offending, with some reference to mortality and hospital re-admission.

2 Lack of information about the nature of treatment. At best, there is information about length of stay in the institution. Some effects may be inferred, for example the complex model for treating personality disorder in Broadmoor Hospital in England,[46] but most studies simply treat secure hospitals or units as a 'black box'.

3 Social attitudes, substance use, policies/legislation, and provision of specialist services all change over time and between jurisdictions.

Further information

Home Office (2003) MAPPA Guidance. www.probation.homeoffice.gov.uk/pdf/MAPPA per cent20guidance.pdf Accessed 10.07.2007.

Jamieson, L., Taylor, P. J., and Gibson, B. (2006). From pathological dependence to healthy independent living: an emergent grounded theory of facilitating independent living. *The Grounded Theory Review* **6**, 79–107.

Royal College of Psychiatrists (2006) Psychiatrists and Multi-Agency Public Protection Arrangements. www.rcpsych.ac.uk/memebers/current_issues/publicprotection.aspx. Accessed 27.07.2007

References

1. Taylor, P. J., Grounds, A., and Snowden, P. (1993). Forensic psychiatry in the National Health Service of England and Wales. In *Forensic Psychiatry, Clinical, Legal and Ethical Issues* (eds. J. Gunn & P. J. Taylor), pp. 691–731. Butterworth Heinemann, Oxford.
2. Home Office, Department of Health and Social Security (1975). (Butler) *Report of the Committee on Mentally Abnormal Offenders*. HMSO, London. Cmnd 6244.
3. www.dspdprogramme.gov.uk
4. Francis, R., Higgins, J., Cassam, E., *et al.* (2006). Report of the independent inquiry into the care and treatment of Michael Stone. South East Coast Strategic Health Authority, Kent County Council, Kent Probation Area. www.kent.gov.uk/publications/council-anddemocracy/ michael-stone.htm Accessed 20.07.2007

5. Collins, G. (2005). Court-mandated psychiatric outpatient treatment in New York: doesn't this process invoke more care than controversy? *Criminal Behaviour and Mental Health*,**15**, 214–20.
6. Yoshikawa, K. and Taylor, P. J. (2003). New forensic mental health law in Japan. *Criminal Behaviour and Mental Health,* **13**, 225–8.
7. Department of Health, Home Office (1992) *Review of Health and Social Services for Mentally Disordered Offenders and Others Requiring Similar Services.* London: HMSO. Cm 2088.
8. Taylor, P. J. (2007). *Second Expert Paper: Mental illness and serious harm to others*. www.nfmhp.org.uk/expertpapers (accessed 15.07.2007)
9. Gunn, J. (1976). Management of the mentally abnormal offender: integrated or parallel. *Proceedings of the Royal Society of Medicine,* **70**, 877–80.
10. Jamieson, L. and Taylor, P. J. (2005). Patients leaving an English high security hospital. Do discharge cohorts and their progress change over time? *International Journal of Forensic Mental Health,* **4,** 59–75.
11. Salize, H.J. and Dreßing, H. (2005). Placement and Treatment of Mentally Ill Offenders - Legislation and Practice in EU Member States. Mannheim, Germany: Central Institute of Mental Health.
12. SWANZDSAJCS (Accepted for publication). Offenders with mental disorder on five continents: a comparison of approaches to treatment and demographics factors relevant to measurement of outcome. *International Journal of Forensic Mental Health.* (Corresponding author P.J. Taylor).
13. Appelbaum, P.S. (1997). A theory of ethics for forensic psychiatry. *Journal of the American Academy of Psychiatry and the Law,* **25**, 233–47.
14. Royal College of Psychiatrists (1995). *Strategies for the management for disturbed and violent patients in psychiatric units* (CR41-03/1995).
15. Department of Health & Welsh Office (1999). *Code of Practice for the Mental Health Act 1983*. London: The Stationery Office.
16. Moore, E., Yates, M., Mallendine, C., *et al.* (2002). Expressed emotion in relationships between staff and patients in forensic services: changes in relationship status at 12 month follow-up. *Legal and Criminological Psychology,* **7**, 203–18.
17. McGauley, G. and Humphrey, M. (2003). Contribution of forensic psychotherapy to the care of forensic patients. *Advances in Psychiatric Treatment,* **9**,117–24.
18. McMahon, C., Butwell, M., Taylor, P.J., *et al.* (2003). Changes in patterns of excessive alcohol consumption in 25 years of high security admissions from England and Wales. *Criminal Behaviour & Mental Health,* **13**, 17–30.
19. Mueser, K.T., Bellack, A.S., Blanchard, J.J., *et al.* (1992). Comorbidity of schizophrenia in substance abuse: implications for treatment. *Journal of Consultant Clinical Psychology,* **60**, 845–56.
20. Durand, M. A., Lelliott, P., Coyle, N., *et al.* (2005). Availability of treatment for substance misuse in medium secure psychiatric care in England: A national survey. *Journal of Forensic Psychiatry and Psychology,* **17**(4), 611–25.
21. Lindqvist, P. (2007). Mental disorder, substance misuse and violent behaviour – the Swedish experience of caring for the triply troubled. *Criminal Behaviour and Mental Health*, **17**, 242–9.
22. Meltzer, D., Tom, B.D.M., Brugha, T., *et al.* (2004). Access to medium secure psychiatric care in England and Wales: a national survey of admission assessments. *Journal of Forensic Psychiatry and Psychology,* **15**, 17–31.
23. Swanson, J.W., Holzer, C.E., Ganju, V.K., *et al.* (1990). Violence and psychiatric disorder in the community: evidence from the Epidemiologic Catchment Area surveys. *Hospital and Community Psychiatry,* **41**, 761–70.
24. Shaw, J., McKenna, J., Snowden, P., *et al.* (1998). The Northwest region. II: Patient characteristics in the research panels recommendation placement groups. *Journal of Forensic Psychiatry,* **5**, 107–22.
25. Barber, M., Short, J., Clarke-Moore, J., *et al.* (2006). A secure attachment model of care: meting the needs of women with mental health problems and anti-social behaviour. *Criminal Behaviour and Mental Health,* **16**, 3–10.
26. Mason, F.L. (2006). Services for women offenders with personality disorder: focus on a trauma-based approach to treatment. In *Personality Disorder and Serious Offending* (eds. C. Newrith, C. Meux, and P. J. Taylor), pp. 231–9. Hodder Arnold, London.
27. *Journal of Intellectual Disability*. (2002). Research Special Issue **46**:s1–85.
28. allace, C., Mullen, P.E., Burgess, P., *et al.* (2004). Criminal offending in schizophrenia over a 25-year period marked by deinstitutionalization and increasing prevalence of comorbid substance use disorders. *American Journal of Psychiatry,* **161**, 716–27.
29. Prison Reform Trust (2007) - Runaway sentence causes havoc – the indeterminate sentence for public protection. www.prisonreformtrust.org.uk/subsection.asp?id=1037.
30. Gunn, J., Maden, A., Swinton, M., *et al.* (1991). Treatment needs of prisoners with psychiatric disorders. *British Medical Journal,* **10**, 338–41.
31. Singleton N., Meltzer H., Gatward R., *et al.* (1998) *Psychiatric morbidity among prisoners in England and Wales.* London: The Stationery Office.
32. Coid, J., Kahtan, N., Cook, A., *et al.* (2001).Predicting admission rates to secure forensic psychiatry services. *Psychological Medicine,* **31**, 531–9.
33. Grounds, A., Gelsthorpe, L., Howes, M., *et al.* (2004). Access to medium secure psychiatric care in England and Wales: the qualitative study of admission decision making. *Journal of Forensic Psychiatry and Psychology,* **15**, 32–49.
34. Jamieson, E., Butwell, M., Taylor, P.J., *et al.* (2000). Trends in special (high security) hospitals. I: Referrals and admissions 1986-1995. *British Journal of Psychiatry*, **176**, 253–9.
35. McInerny, T. (2000). Dutch TBS forensic services: a personal view. *Criminal Behaviour & Mental Health,* **10**, 213–28.
36. Berry, A., Larkin, E., Taylor, P., *et al.* (2003). Referred to high secure care: determinants of a bed/offer admission and placement after one year. *Criminal Behaviour and Mental Health,* **13**(4), 310–20.
37. Cohen, A. and Eastman, N. (2000). *Assessing Forensic Mental Health Need. Policy, Theory and Practice*. London: Gaskell.
38. Collins, M. and Davies, S. (2005). The Security Needs Assessment Profile: A multidimensional approach to measuring security needs. *International Journal of Forensic Mental Health,* **4**(1), 39–52.
39. Douglas, K. S., Guy, L. S., Weir, J., *et al.* (2006). HCR-20 violence risk assessment scheme: Overview and annotated bibliography. Burnaby, Canada: Department of Psychology, Simon Fraser University. Available at http://www.sfu.ca/psyc/faculty/hart/Resources.htm.
40. Belfrage, H., Fransson, G., Strand, S., *et al.* (2000). Prediction of violence using the HCR-20: A prospective study in two maximum security correctional institutions. *Journal of Forensic Psychiatry,* **11**, 167–75.
41. Taylor, P.J. (2006). Delusional disorder and delusions: is there a risk of violence in social interactions about the core symptom? *Behavioral Sciences & the Law,* **24**, 313–32.
42. Taylor, P.J. and Swan, T. (eds.). (1999). *Couples in Care and Custody.* Butterworth-Heinemann: Oxford.
43. Home Office (2003) MAPPA Guidance. www.probation.homeoffice.gov.uk/pdf/MAPPA per cent20guidance.pdf Accessed 10.07.2007.
44. Royal College of Psychiatrists (2006) Psychiatrists and Multi-Agency Public Protection Arrangements. www.rcpsych.ac.uk/memebers/current_issues/publicprotection.aspx. Accessed 27.07.2007
45. Jamieson, L., Taylor, P. J., and Gibson, B. (2006). From pathological dependence to healthy independent living: an emergent grounded theory of facilitating independent living. *The Grounded Theory Review* **6**, 79–107.
46. Newrith, C., Meux, C., Taylor, P., *et al.* (2006). *Personality Disorder and Serious Offending - Hospital Treatment Models*. Hodder Arnold, London.

This page intentionally left blank

Index

Note: Page numbers in **bold** refer to main entries. Those in *italic* refer to figures and / or tables. Alphabetical order is word-by-word.

Plates for Chapter 2.3.1

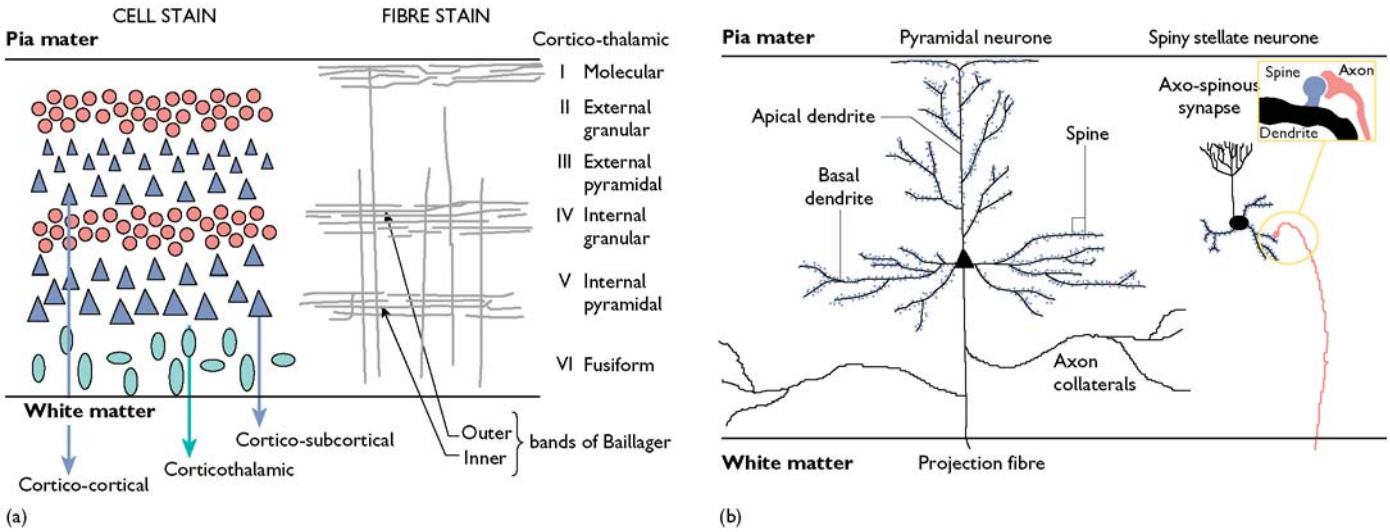


Plate 1 (a) Laminar structure of the cerebral neocortex; (b) excitatory amino acid using spiny neurones of the neocortex.

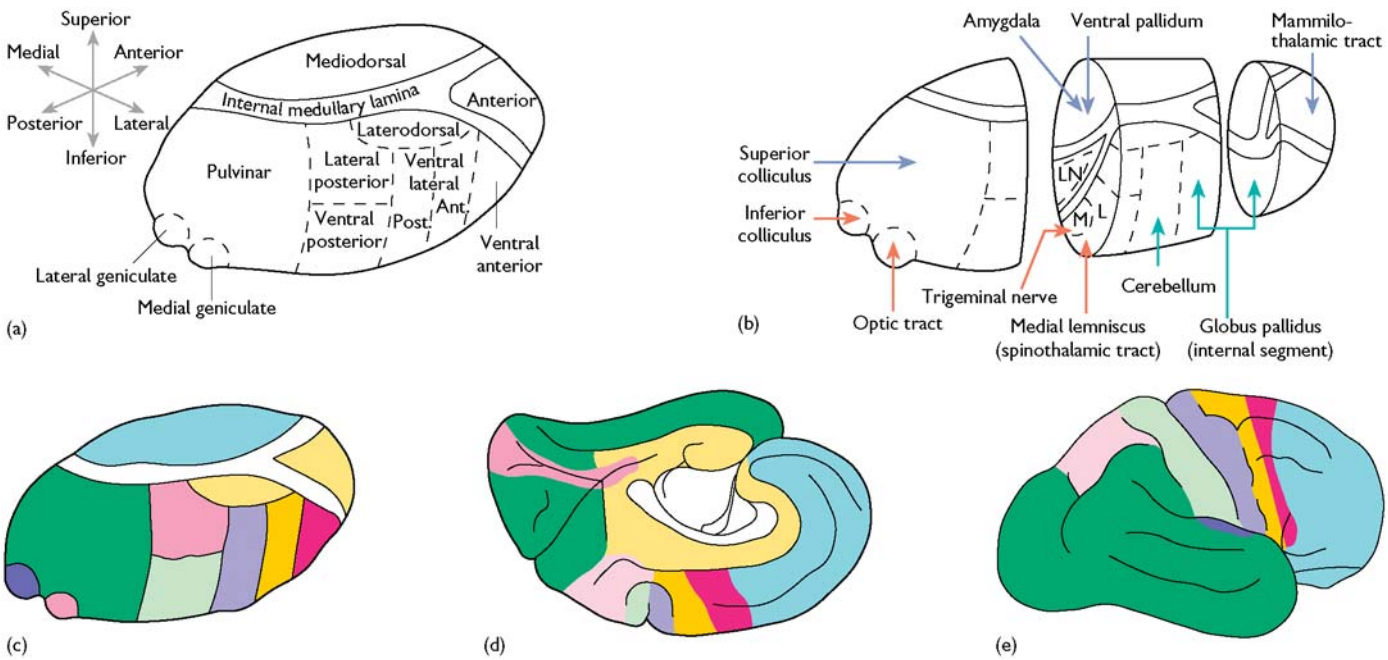


Plate 2 (a) The thalamus is shown as if removed from the brain at the top, with the individual major nuclei indicated. (b) Schematic diagram showing the isolated thalamus divided approximately at the middle of its anteroposterior extent; the arrows indicate known sources of major subcortical afferents to the individual named nuclei. (c) The diagram of the isolated thalamus shown in (a) with the individual main nuclei colour coded. (d), (e) Schematic diagrams of the cerebral surfaces (medial and lateral) showing the regions of cortex colour coded to correspond to the main thalamic nuclei with which they have major connections.

Plates for Chapter 2.3.6

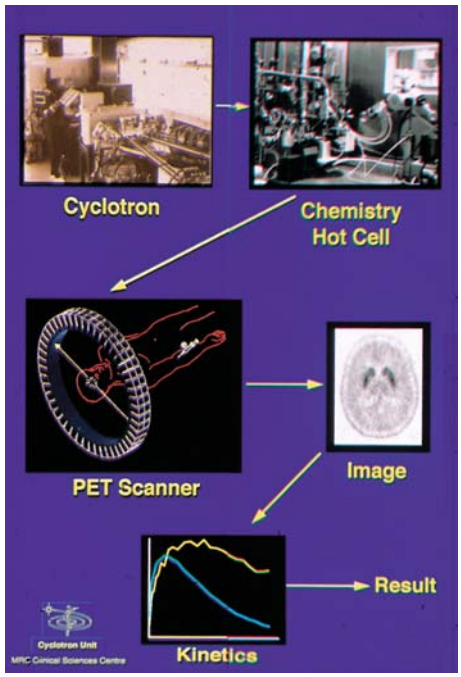


Plate 3 Steps in the production and use of PET radio-isotopes.

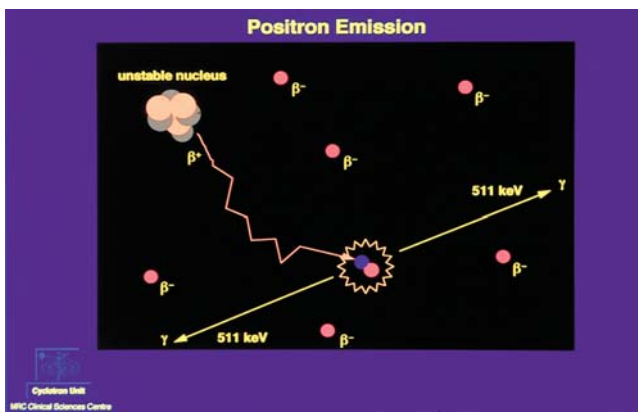


Plate 4 Principles of positron emission: β^- is an electron and β^+ is a positron. Two high-energy gamma rays (γ) are produced on annihilation of a positron by an electron

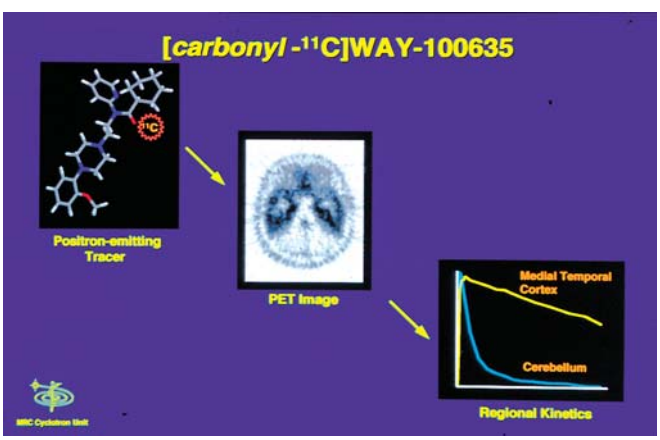


Plate 5 Use of a PET radiotracer to image 5-HT_{1A} receptors in the human brain.

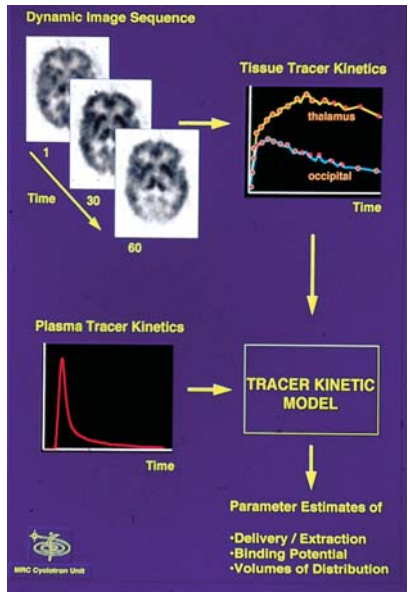


Plate 6 Steps in the data analysis of a PET radiotracer.

Plates for Chapter 2.3.7

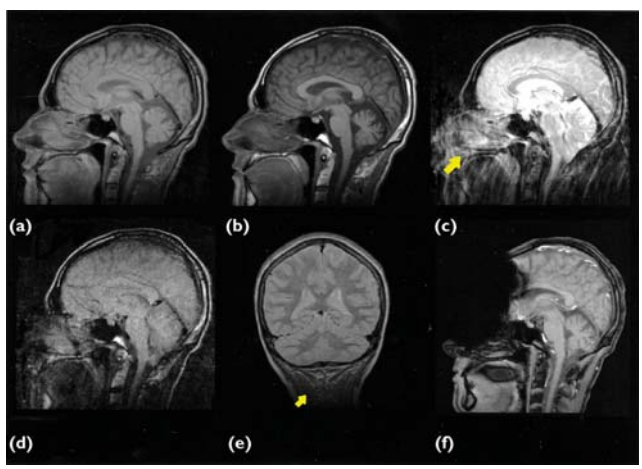


Plate 7 Spin echo MR images and artefacts. Top row, spin echo images: left, proton density image (TR = 2000 ms, TE = 20 ms); middle, T_1-weighted image (TR = 350 ms, TE = 20 ms); right, T_2-weighted image (TR = 2000 ms, TE = 90 ms). The arrow on the T_2-weighted image indicates blurring caused by movement (swallowing) during acquisition. Bottom row, examples of poor tissue contrast and susceptibility artefact: left, spin echo image showing poor contrast due to injudicious prescription of pulse sequence (TR = 350 ms, TE = 90 ms); middle, bulk susceptibility artefact; right, ferromagnetic susceptibility artefact caused by a metallic hairgrip.

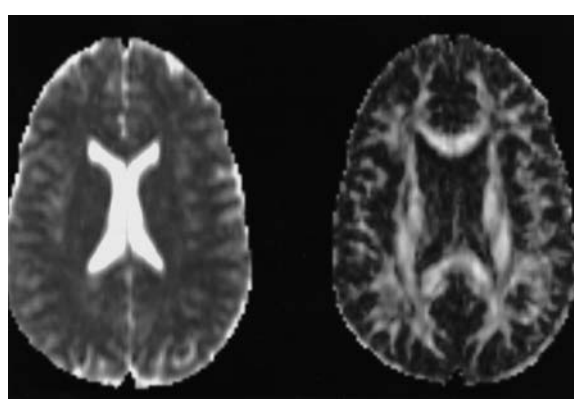


Plate 8 Diffusion-weighted MRI data can be used to generate maps of (left) the apparent diffusion coefficient (**ADC**) and (right) the anisotropy of diffusion. Diffusion of protons is most rapid and isotropic in cerebrospinal fluid, and least rapid and most anisotropic in white matter. White matter is clearly defined by relative hyperintensity in the anisotropy map.

Plates for Chapter 2.3.7 *(continued)*

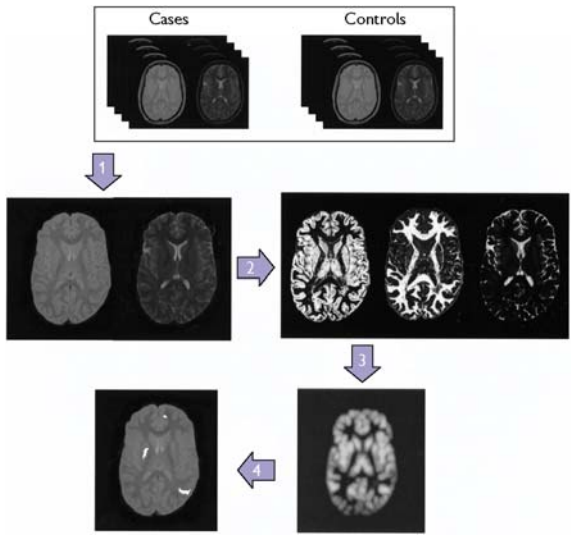


Plate 9 Steps in computerized image analysis. Dual-echo (fast spin echo) data are acquired from several cases and controls in a cross-sectionally designed study. Extracerebral tissue is removed (1) from each image before segmentation or tissue classification (2). Tissue-calssified images are registered with a template image in standard space (3) before hypothesis testing (4). Voxels or clusters, which demonstrate a significant difference in tissue class volume between groups, are colour coded.

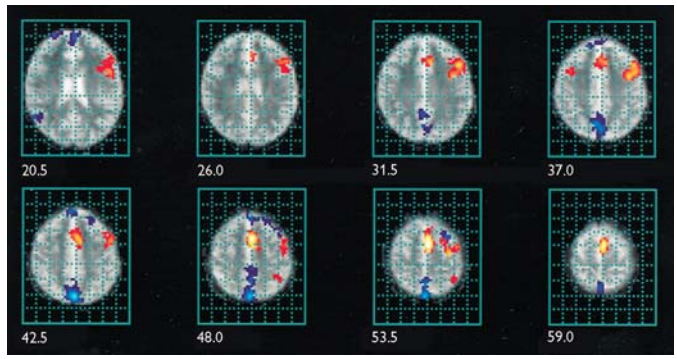


Plate 11 Activation map. Generically activated voxels are colour coded against a grey-scale background of gradient echoplanar imaging data. The grid respresents the standard Talairach-Tournoux Space;[12] z-coordinates for each slice are shown at the bottom left. Colour codes the timing and power of a periodic response to a covert verbal-fluency experiment. Blue voxels show increased magnetic responance signal during condition B (repeat a word covertly); light blue represents a greater power of response than dark blue. Red voxels show increased magnetic resonance signal during condition. A (generate a word beginning with a cue letter); yellow and orange represent a greater power of response than dark red. The voxel-wise probability of a false-positive error is $p = 0.0001$. The main areas activated during conditioin A are the dorsolateral prefrontal cortex, inferior frontal gyrus, and supplementary motor area; the main areas activated during condition B are the medial parietal cortex and posterior cingulate gyrus.

Plates for Chapter 2.3.8

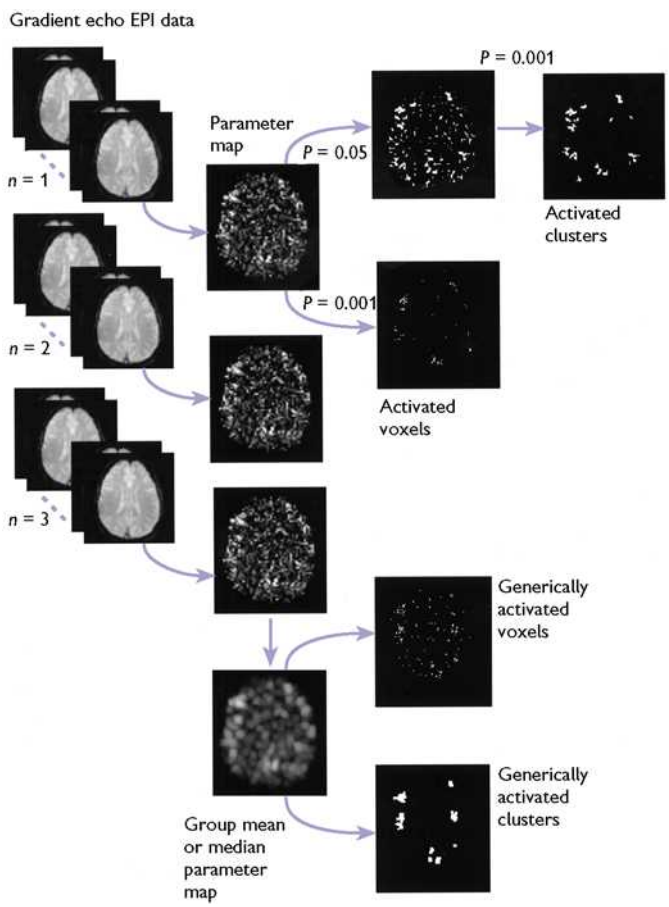


Plate 10 Steps in computerized data analysis. Gradient echoplanar imaging data have been acquired from three subjects under identical conditions. The time series at each voxel is analysed to estimate a measure or parameter of the experimental effect, which is represented as a parameter map. Significantly activated voxels or clusters can be identified in each individual image. Parameter maps can be averaged over individuals and generically activated voxels or clusters identified over the group of subjects. The power to detect activation is enhanced by cluster-level analysis and by combining data from several subjects.

Plate for Chapter 4.1.3

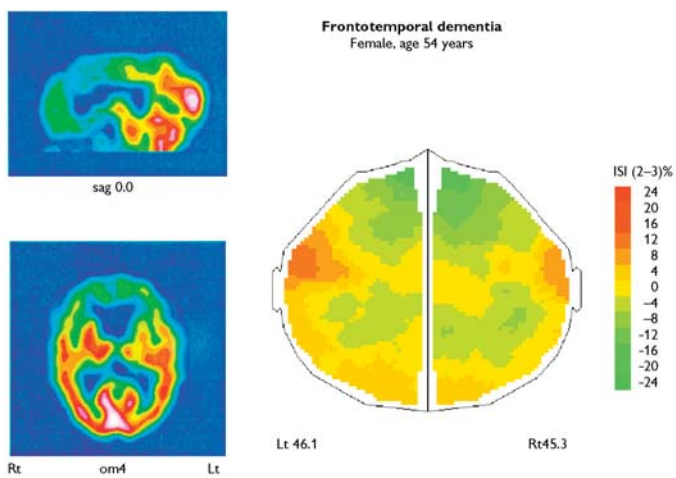


Plate 12 Regional cerebral blood flow (rCBF) measured using SPECT with exametazine (left) and the Xenon-133 inhalation method (right) in a 54-year-old female with clinical signs of FTD. The variation of regional cerebral blood flow is measured with xenon-133, above (red) or below (green) the average flow level, as indicated by the colour code. The patient showed the first signs of personality change, and stereotypy of speech and behaviour at the age of 48 years. EEG was normal, and CT and MRI showed slight frontal cortical atrophy. The regional cerebral blood flow measurement with xenon-133 showed a normal average flow level and marked bilateral, frontal flow decreases. The SPECT scan showed a severe perfusion deficit in the frontal and anterior cingulate cortex. (Courtesy of Department of Neurophysiology, University Hopital, Lund, Sweden.)

Plate for Chapter 4.2.1

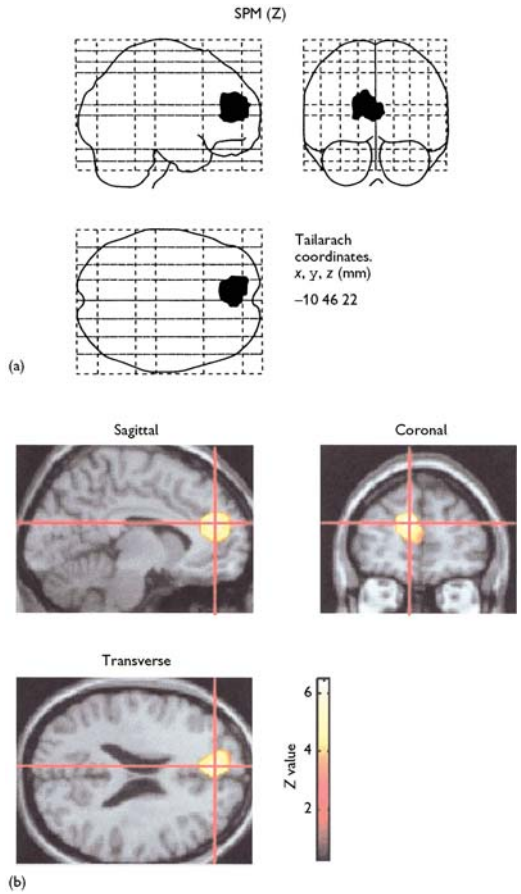


Plate 13 Area of activation during opiate craving: (a) [^{15}O]H_2O PET SPM image; (b) area of activation superimposed on magnetic resonance image.

Plate for Chapter 4.5.6

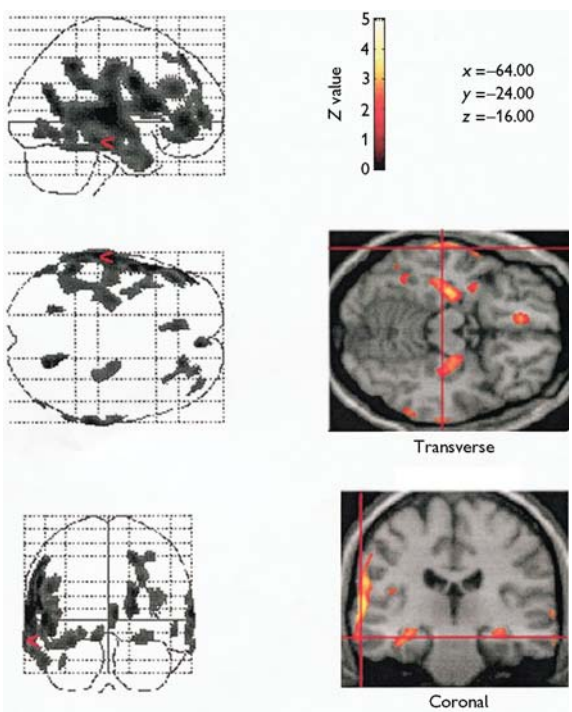


Plate 14 Statistical parametric maps ($p < 0.001$) for reductions in grey matter densities in subjects with chronic refractory depression compared with controls. Effects are controlled for age. (Reproduced with permission from PJ. Shah *et al.* (1998). Cortical grey matter reductions associated with treatment-resistant chronic unipolar depression: controlled MRI study. *British Journal of Psychiatry*, **72**, 527–32.)

Plates for Chapter 6.2.10.4

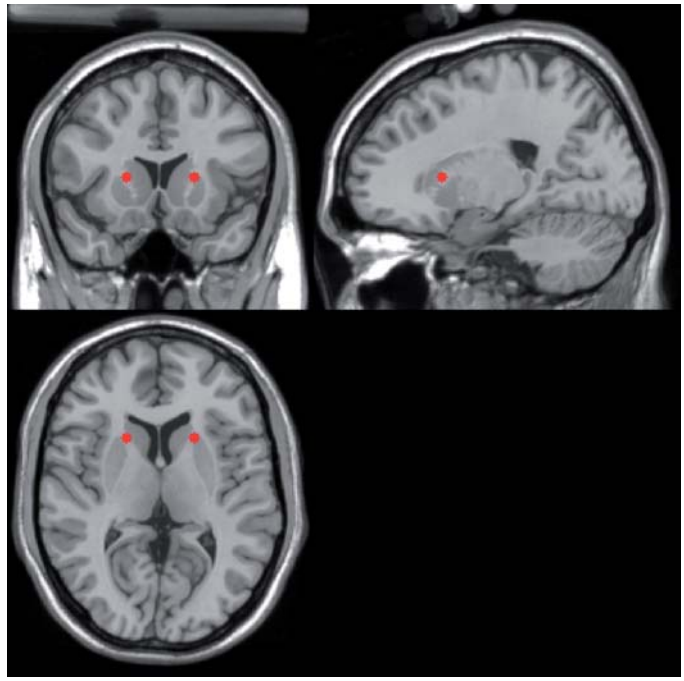


Plate 15 Typical locations of anterior capsulotomy lesions superimposed upon normalized T1 MRI scan. Lesions not to scale.

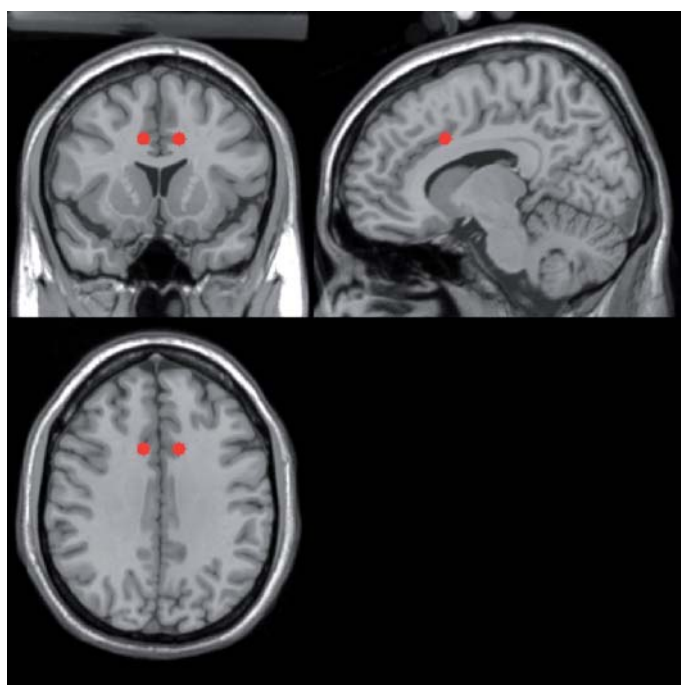


Plate 16 Typical locations of anterior cingulotomy lesions superimposed upon normalized T1 MRI scan. Lesions not to scale.